Agricultural Productivity, Structural Transformation, & Urbanization: Evidence from Indonesia's Transmigration Program*

Idaliya Grigoryeva (UC San Diego)

March 27, 2026

Abstract

In this paper, I study rural-to-urban spillovers, estimating the effect of rural agricultural productivity on urban population and employment growth, migration, and structural transformation in cities, utilizing the quasi-random allocation of migrants to destinations under the Transmigration Program in Indonesia. By exploiting variation in agricultural productivity across villages, I show that cities surrounded by more productive rural areas experience sectoral shifts from agriculture to manufacturing and services, with limited evidence of overall population and employment growth. The primary channel is likely the goods' trade: productive villages supply food to nearby towns, reducing the cost of living and stimulating demand for urban services, while farmers in successful villages tend to stay rather than migrate. As transmigration villages are settled, the areas around more productive villages develop into trade and agro-processing hubs and towns, as evidenced by stronger growth in trade and transportation services. I develop a three-sector spatial equilibrium model with non-homothetic (Stone-Geary) preferences, agglomeration externalities in services, and endogenous food prices to rationalize these findings and decompose the trade and migration channels. Urban structural transformation appears to result from these regional economic spillovers through higher agricultural output and trade rather than outflows of transmigrants from the villages themselves.

*I thank Sam Bazzi, Prashant Bharadwaj, Juan Pablo Chauvin, Fabian Eckert, Gaurav Khanna, Craig McIntosh, Paul Niehaus for their continued support, feedback, and guidance throughout all stages of this project. I am grateful to Titan Alon, Itzik Fadlon, Edward Glaeser, Gabriel Kreindler, Paul Novosad, Giovanni Peri, Tom Vogl, the participants at Lindau Nobel Laureate Meeting 2025, University of Hawaii Applied Seminar, the Cities and Development Workshop 2024, Migration Summer School 2024, IHS Trade workshop 2024, Southern California Graduate Conference in Applied Economics (SOCAE) 2024, visiting speakers at the UCSD development seminar series, the UCSD Applied and Development Seminars, and the UCSD Graduate Student Research Seminars for their helpful suggestions and questions. I thank Roberto Gonzalez-Tellez, Haicheng (Kevin) Xu, and the participants of the undergraduate-graduate research lab (URL) at UC San Diego for their outstanding research assistance.

This research was financially supported by the STEG PhD Research grant, several Institute for Humane Studies grants, and UC San Diego Economics Department Research & Travel grants. All remaining errors are my own.

1 Introduction

Understanding the nexus between rural agricultural productivity and urban growth is crucial for developing effective economic development strategies. Classical dual-economy models posit that improvements in agricultural productivity can fuel urbanization by freeing surplus rural labor and providing food to support a growing urban workforce (Harris & Todaro 1970; Lewis 1954). In these models, rising farm productivity gradually pushes underemployed rural workers toward cities in search of higher wages. In today's developing countries, however, various frictions—such as moving costs, housing constraints, and imperfect information—impede this migration process, leading to persistently large rural-urban income gaps (Young 2013; Lagakos 2020; Munshi & Rosenzweig 2016) and potentially sub-optimal lower levels of rural-to-urban migration (Lagakos et al. 2020; Bryan & Morten 2019). Subsequent theoretical work offers ambiguous predictions: Matsuyama (1992) proposes a framework where higher agricultural productivity can either stimulate or hinder industrialization depending on trade openness, while Glaeser (2014) extends the model to highlight how rural improvements affect modern-day urbanization more specifically and empirically tests it on a cross-country sample. These contrasting perspectives underscore the need for further empirical evidence on how boosting rural productivity affects structural transformation and urban growth in developing countries.

This paper investigates how agricultural productivity differences in rural areas influence the economic expansion of nearby urban areas. I leverage the quasi-random allocation mechanism of the Transmigration Program during the 1970s-1980s in Indonesia, which offered a free land lot and transportation expenses covered for migrants to relocate from the main islands of Java and Bali to less densely populated outer islands (Bazzi et al. 2016, 2019). At its peak between 1979 and 1988, the program resettled over two million people. In this setting, I identify and estimate the causal effect of variation in village-level productivity based on the transmigrant skills match on economic outcomes in nearby cities, including population and employment growth, structural transformation, and migration in cities.

I utilize a unique combination of administrative and remote sensing data on the distribution of urban areas, population, migration, and employment from Indonesia's census data and the Global Human Settlements layer derived from remote sensing data. Compiling these sources, I construct a rich dataset of 180 urban centers on the Outer Islands, aggregating urban district-level observations (kecamatans) to a consistent urban center definition that captures contiguous urban agglomerations. For the transmigration (TM) villages, I obtain information on the establishment year and the number of migrants initially settled from the Transmigration Census, as well as agro-climatic similarity — a proxy for village agricultural productivity — based on the approach in Bazzi et al. (2016).

With the haphazard rollout of the program during the peak time, there was no systematic or targeted matching of migrants to destinations and program participants had no say in where they end up, resulting in quasi-random assignment of migrants from different districts of origin to destination villages. Given the wide range of agro-climatic conditions across places of origin and destinations,

there is exogenous variation in agro-climatic similarity – the match quality between settlers' skills and local agricultural conditions, which has been shown to be a proxy of migrant skills and agricultural productivity (Bazzi et al. 2016). I further leverage this village-level agro-similarity along with the cost-distance estimate of travel time between each city-village pairs to construct a city-level exposure shock to the change in rural productivity, providing me with a plausible identification strategy to estimate the effects of agricultural productivity on urban development. Using the earliest available census data from 1980, I show that there was no apparent selection by village productivity in getting placed closer to specific types of urban areas: baseline urban population, employment, economic structure and migration rates are uncorrelated with surrounding village productivity at the time of village settlement.

My findings demonstrate a strong link between rural productivity and urban structural transformation, alongside muted population effects. I establish four empirical facts about rural productivity spillovers to cities. First, cities exposed to more productive transmigration villages show limited overall population and employment growth: effects are positive but statistically insignificant, suggesting that the productivity shock does not trigger a dramatic urban population explosion. Second, these cities experience substantial structural transformation: employment shifts away from agriculture toward manufacturing and services, with the service share growing significantly; within services, growth is concentrated in trade and transportation, consistent with agricultural surplus generating local market activity. Third, the evidence points to a goods-market mechanism operating through urban food markets: exposed cities reduce agricultural employment and local agricultural production while maintaining broadly stable food consumption, alongside a decline in the food spending share and rising housing prices, consistent with Stone-Geary preferences where households reallocate spending toward non-food items as the effective cost of food declines. Fourth, the compositional shift in urban demographics is driven primarily by in-migration from other regions rather than by the mechanical relocation of transmigrants out of villages.

In fact, transmigrants are more likely to remain and work in agriculture in more productive villages, consistent with the idea that productive rural areas retain incumbents while generating regional spillovers that attract additional workers into nearby towns. This suggests that productive rural areas not only create direct local opportunities but also generate spillover effects that draw other migrants to nearby urban centers. The central narrative is thus one of structural transformation without a dramatic urban population explosion: compositional shifts in employment and consumption patterns occur through trade and demand channels rather than through large-scale rural-to-urban migration.

In addition to the reduced-form evidence, I develop a quantitative spatial equilibrium model to interpret and quantify the mechanisms behind these findings. The model features multiple locations across the rural-urban continuum and three sectors — agriculture, manufacturing, and services — with non-homothetic Stone-Geary preferences that capture both subsistence food requirements and a transition from home-produced to market-purchased services as incomes rise. Agricultural output is tradable across locations subject to iceberg trade costs, while manufacturing is freely tradable and serves as the numeraire, and services are non-tradable and subject to agglomeration externalities, whereby

service-sector productivity rises with local population size. Workers make a nested location-sector choice governed by Fréchet taste shocks, where the location-choice elasticity θ_L controls migration responsiveness and the sector-choice elasticity θ_S governs within-location occupational sorting.

The structural framework serves two purposes. First, it provides a unifying lens to organize the empirical results, illustrating how rural productivity improvements transmit into urban outcomes in general equilibrium and rationalizing the full set of empirical facts: muted population growth (reflecting high migration frictions), declining agricultural employment shares (trade lowers urban food prices, shifting relative returns), rising service employment and housing prices (increased demand for non-tradables), stable non-agricultural wages (agglomeration offsets diminishing returns), and declining food spending shares (Stone-Geary non-homotheticity). Second, it enables counterfactual experiments beyond the reach of the reduced-form design, including shutting down inter-regional trade (quantifying the trade channel's contribution to structural transformation), shutting down migration (isolating the role of within-city labor reallocation), and alternative targeting of transmigration villages (assessing optimal spatial allocation of productivity shocks).

In this paper, I provide novel causal evidence of the role of agricultural productivity on rural-to-urban spillovers in a developing country. While prior studies have examined how agricultural booms affect local economies (e.g., Bustos et al. (2016) on soy expansion leading to local structural changes in Brazilian municipalities), most focus on structural change and growth in the same areas as the shock (Bustos et al. 2016; Nunn & Qian 2011; Gollin et al. 2016), rather than spillovers from rural areas to surrounding urban regions. By linking exogenous variation in village-level agricultural productivity to urban outcomes, this project sheds light on a new mechanism whereby improving rural productivity can catalyze nearby urbanization. A distinctive finding of this paper is that structural transformation occurs without large population inflows, suggesting that the trade and goods-market channels are quantitatively more important than the migration channel in this context. This contrasts with the classical Harris-Todaro narrative of rural push-driven urbanization and has direct policy implications: investments in rural productivity can transform nearby urban economies through market linkages without requiring (or producing) massive rural-to-urban migration. I also add to the literature on migration barriers and misallocation: consistent with Bryan & Morten (2019)'s findings of costly internal migration, my results suggest that the spillover benefits of agricultural productivity manifest largely through trade and demand linkages rather than labor flows, with migration primarily drawing workers from other provinces rather than from the transmigration villages themselves. More broadly, my findings speak to the debate on rural-urban duality and balanced growth across sectors, suggesting that investing in rural productivity need not come at the expense of urban growth. On the contrary, it can indirectly foster the expansion of cities and the diversification of employment. The model's explicit decomposition of trade versus migration channels — novel in this empirical context — provides a disciplined framework for quantifying each channel's contribution.

This paper contributes to several strands of literature. First, it speaks to work on the role of agriculture in development and structural transformation, which emphasizes that gains in farm productivity

can raise aggregate productivity and shift labor toward non-agricultural activities through general equilibrium channels (Matsuyama 1992; Gollin et al. 2002; Restuccia et al. 2008; Lagakos & Waugh 2013; Bustos et al. 2016; Nunn & Qian 2011). Second, it relates to the literature on urbanization and city growth in developing economies, which highlights the emergence and expansion of cities as a function of market access, sectoral composition, and agglomeration forces (Glaeser 2014; Vernon Henderson & Turner 2020; Bryan et al. 2020; Gollin et al. 2016; Jedwab & Vollrath 2015; Dingel et al. 2019). Third, the results inform research on internal migration frictions and spatial misallocation: when moving is costly and networks shape mobility, local shocks can generate spatially uneven population reallocation and re-sorting across sectors (Bryan & Morten 2019; Morten & Oliveira 2016; Munshi & Rosenzweig 2016; Lagakos 2020; Monte et al. 2018). Finally, by combining quasi-random within-program variation in migrant agricultural productivity with a geography-aware measure of urban exposure, this paper complements empirical work on rural economic geography and downstream linkages in Indonesia, including evidence that agricultural processing can induce non-agricultural employment and local urbanization (Edwards 2019), and extends the Transmigration literature beyond settlement-level outcomes to the surrounding urban system (Bazzi et al. 2016).

The rest of the paper is structured as follows. **??** provides background on Indonesia's regional development and the Transmigration Program. **??** describes the data. **??** presents the empirical strategy and **??** presents the main results on how agricultural productivity affects urbanization. Section 4 introduces the quantitative spatial equilibrium framework with three sectors, non-homothetic preferences, and agglomeration externalities. **??** proposes an alternative identification comparing cities with nearby villages that were settled vs. planned but not settled, offering suggestive evidence on the joint effect of the population and productivity shock. Section 5 concludes.

2 Setting and Data

Many development programs implemented by governments and often supported by external financing target rural areas: raising farm productivity, expanding cultivation, and improving rural livelihoods. These policies are typically evaluated using rural outcomes. But when goods and labor markets are linked across space, rural productivity improvements can also reshape nearby towns and cities. Higher agricultural output can lower the local cost of living (especially when staple foods are locally sourced), raise rural incomes and demand for non-agricultural goods and services, and alter migration incentives and urban labor supply. Whether these channels translate into urban growth and structural transformation is ultimately an empirical question.

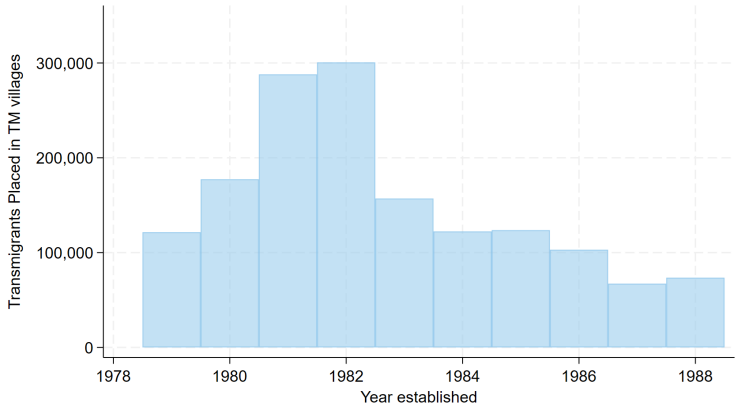
This paper studies these **rural-to-urban spillovers** by leveraging a policy that (i) generated a large and spatially dispersed rural shock, (ii) produced plausibly exogenous differences in rural productivity across treated locations, and (iii) can be linked to urban outcomes over several decades. Indonesia's Transmigration Program provides this setting. Although the program was designed as a rural resettlement and agricultural development policy, it created sharp, localized changes in rural population and agricultural production in the hinterlands of many towns. Crucially, the program's implementation

also produced quasi-random variation in **agricultural productivity** across newly established villages, which I use to study how rural productivity affects nearby urban areas.

2.1 Rural Resettlement Program with Quasi-Random Migrant Assignment

Indonesia's Transmigration (TM) Program was one of the largest resettlement initiatives in modern history. During the program's peak period between 1979 and 1988, the government resettled more than two million people from the densely populated Inner Islands (primarily Java and Bali) to newly created rural villages across the Outer Islands. Each participating household received a standardized settlement package, including land for farming, basic housing, and transportation to the destination, intended to facilitate the rapid creation of new agricultural communities. The program was explicitly rural: it aimed to reduce population pressure in origin regions and expand agricultural production and food security in receiving regions, rather than to target cities directly.

Figure 1: Population Trend in Transmigration Village Settlement by Year



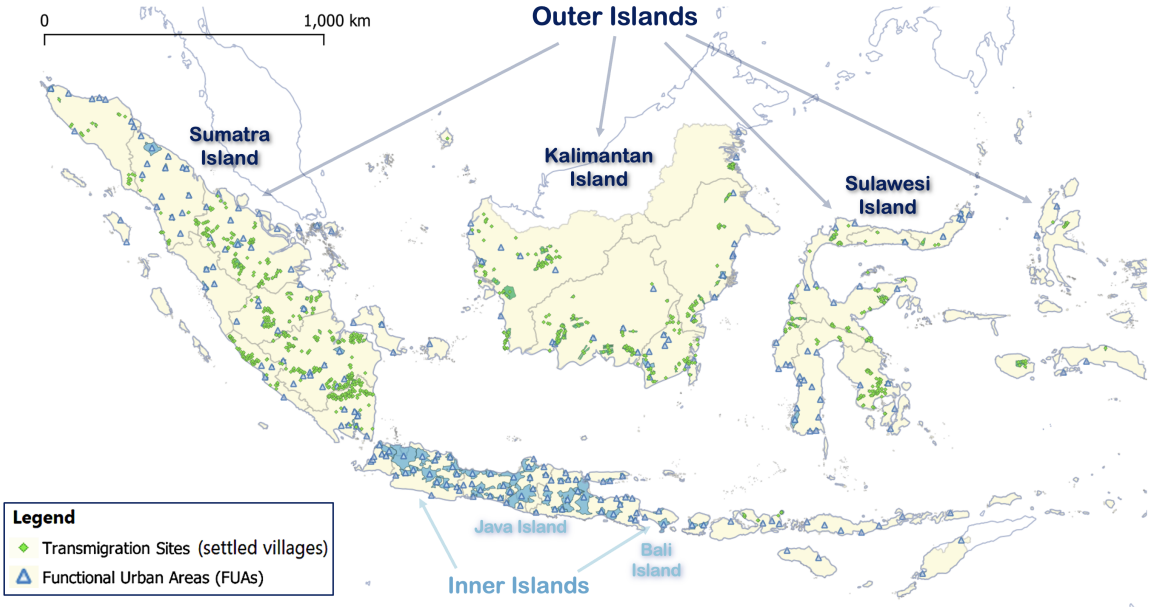
Source: Transmigration census data extracted from World Bank Transmigration Program reports.

Two additional features matter for interpreting the program as a shock. First, the rollout was large and concentrated in time. Settlement activity expanded rapidly in the early 1980s and then declined, with the program ending abruptly in 1988. Figure 1 shows that the years with the most transmigrants placed (and the highest corresponding population increase) were 1981-1982, when around 300,000 transmigrants were moving to about 150 new villages annually. Second, TM villages were geographically dispersed across the Outer Islands, generating substantial spatial heterogeneity in which towns were exposed to new rural settlements and how close those settlements were (See the map of transmigration sites and urban agglomerations in Figure 2).

2.2 Quasi-Random Assignment & Agricultural Productivity

For this project, the key feature of Indonesia's Transmigration Program is not only that it moved people, but that it generated quasi-exogenous differences in how productive newly settled villages became. Transmigrant households did not choose their destination villages. Instead, destinations were assigned through an administrative process shaped by logistical constraints and the sequential opening of settlement sites. In practice, transport schedules and arrival timing determined where families were sent, and

Figure 2: Transmigration villages and urban agglomerations



Notes: Green points indicate TM villages established during 1979–1988. Blue areas with markers indicate urban agglomerations. The TM villages are widely dispersed across the Outer Islands and often located within the broader market areas of existing towns. Source: Original map based on TM village data from Bazzi et al. (2016) and urban areas from Moreno-Monroy et al. (2021).

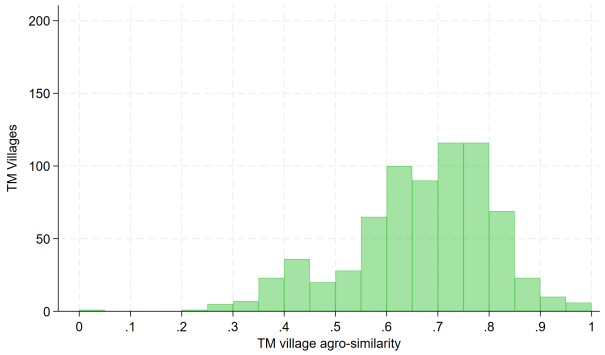
there was no systematic matching of migrants' origin characteristics to destination conditions (Bazzi et al. 2016). This institutional feature implies that, conditional on where the government chose to open a settlement site, the composition of settlers arriving to that site was plausibly as-good-as-random.

A central insight in Bazzi et al. (2016) is that this quasi-random assignment interacted with Indonesia's substantial agro-ecological heterogeneity to create persistent differences in village-level agricultural success. Migrants brought agricultural experience, crop knowledge, and farming practices suited to the environments in which they had previously worked. When destination environments resembled migrants' origins—along dimensions such as climate, soil, and terrain—migrants' skills were more transferable, facilitating faster learning and higher output. When environments differed sharply, settlers faced a steeper learning curve and lower productivity. To operationalize this idea, Bazzi et al. (2016) construct an *agro-climatic similarity* index that measures how closely a village's agro-ecological conditions match those of the migrants' origin districts (aggregating origin–destination distances using the village's origin composition). They show that this index is strongly predictive of agricultural performance and broader local development outcomes in transmigration villages, consistent with a "skill transferability" channel rather than selective placement of migrants into more promising sites (Bazzi et al. 2016). Importantly for my application, the components of the index are based on time-invariant (or slow-moving) agro-ecological attributes, and the assignment process makes the resulting variation in match quality plausibly orthogonal to unobserved determinants of later growth.

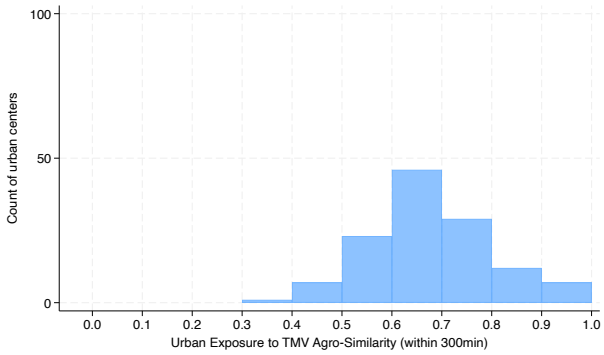
Building on this foundation, I use village-level agro-climatic similarity as a proxy for *relative agricultural productivity* across transmigration villages, and I repurpose it to study a different and previously under-measured outcome: how rural productivity shocks transmit to nearby urban areas. In particular, the empirical design in this paper uses the within-program variation emphasized by Bazzi et al. (2016) (differences in match quality across villages) and then exploits the fact that villages are spatially distributed around towns and cities. This allows me to construct city-level exposure to rural productivity shocks by aggregating surrounding villages' productivity (while separately controlling for the scale of the settlement shock), and to estimate how exogenous rural productivity differences shape urban population growth, employment, and structural change.

Figure 3: Variation in urban exposure to TM-village agricultural conditions

(a) Village-level agro-similarity (TM villages)



(b) City exposure to TMV agro-similarity

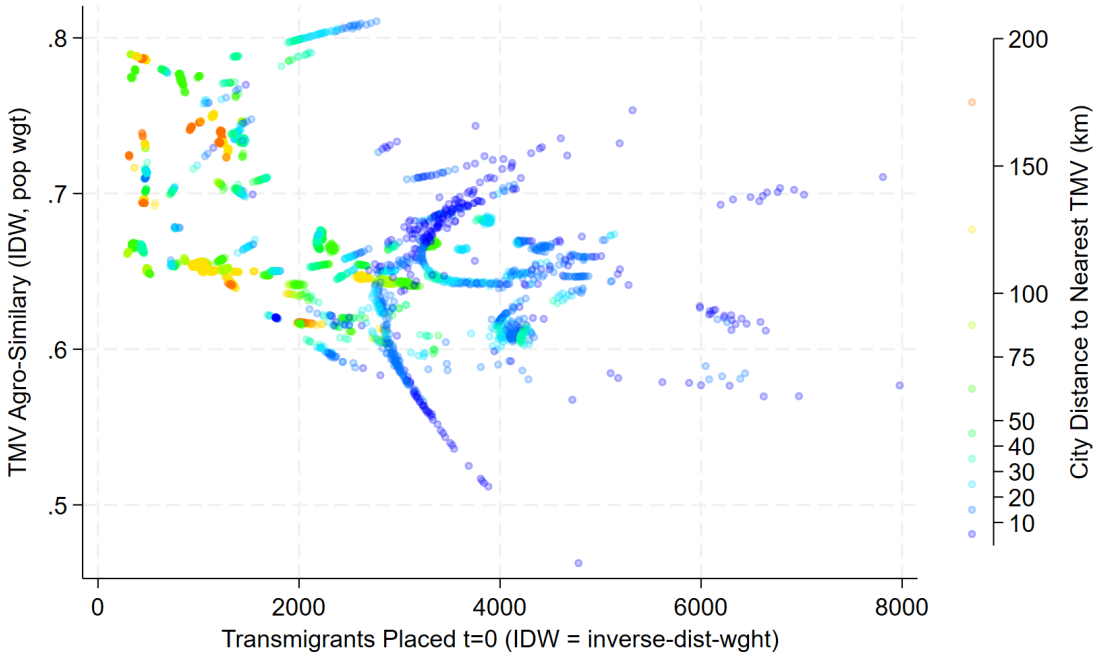


Notes: Urban agro-similarity exposure is calculated as the inverse-distance & population-weighted agro-similarity across villages (IDW).

Figure 3 summarizes heterogeneity in agro-similarity across TM villages and the implied exposure of urban districts to those agricultural conditions. Panel 3a shows the distribution of the village-level agro-similarity index, with 75% of TM villages reporting positive rice production in the analysis sample; empirically, a one-standard-deviation increase in agro-similarity is associated with an approximately 20% increase in rice productivity. Panel 3b maps these village fundamentals into an urban "exposure" measure using a population-weighted inverse-distance weighting (IDW) aggregation, a standard spatial weighting approach that places greater weight on closer and larger locations. This variation is economically relevant given the importance of rice in urban consumption: in the household survey data used here, rice accounts for more than 25% of urban household expenditures on agricultural goods (BPS Susenas 2011).

The program's geography implies that towns were exposed to different combinations of (i) the scale of nearby TM settlement and (ii) the productivity of those surrounding villages. Some towns are surrounded by TM villages with high agro-climatic match and thus higher agricultural productivity; others are surrounded by villages with low match; and many towns are exposed to few or no TM villages at all (See some of the variation in these key variables in Figure 4). This spatial structure allows me to compare urban outcomes across towns facing different rural productivity shocks while holding fixed overall program scale by controlling for the local intensity of TM settlement.

Figure 4: Aggregated Agro-Similarity Index & TMP Exposure



Note: The index quantifies the match between migrants' farming techniques and the new environment, influencing their potential productivity and integration success.

In the empirical analysis, I construct measures of each urban area's exposure to TM villages and to TM village agricultural productivity using distance-based (and later travel-time-based) weights. These exposure measures capture the idea that rural shocks matter more for cities that are closer and more connected to the affected hinterland. The next section describes the data used to measure village productivity and urban outcomes, and the construction of these city-level exposure variables.

To empirically investigate how rural productivity shocks transmit to urban areas, I now describe the construction of the analysis dataset, the key outcome measures, and the specific methodology for quantifying urban exposure to transmigration village agricultural conditions.

2.3 Data Sources

This section describes the construction of the analysis dataset and the key variables used in the empirical analysis. The unit of data collection is the urban district (*kecamatan*, administrative level 3). In the main analysis, I aggregate district-level observations to the urban center level, which I interpret as the effective local labor market. I combine (i) administrative population census data (1980, 2000, 2010), (ii) Transmigration program village records and replication files from prior work on Indonesia's Transmigration program (Bazzi et al. 2016, 2019), and (iii) geospatial layers used to define functional urban areas and to construct travel-time (cost-distance) measures between villages and cities.

The main data inputs are administrative microdata from three key census years and the Transmigration census, along with satellite-derived remote sensing data for constructing a travel-time distance measure.[1]

Population census (1980, 2000, 2010). I use the population census individual-level microdata (full count in 2000 and 2010 and the 5% population sample for 1980) to measure population, employment, sectoral employment composition, and migration patterns. The 1980 data are representative geographic identifiers down to the district level (kecamatan, level 3), and 2000 and 2010 data are representative at the village/community level (desa/kelurahan, level 4). For the later years, there are also level-4 spatial shapefiles available at the municipality level I use 1980 outcomes as baseline controls and construct post-policy outcomes primarily in 2010, with growth measures using 2000–2010.

Transmigration Census and replication files. I use the universe of settled Transmigration villages (TMVs), including settlement timing, village locations, and the number of individuals initially placed. I also use the village-level agro-climatic similarity index constructed in Bazzi et al. (2016).

Administrative boundary crosswalks. To harmonize district identifiers across census waves, I use Olken (2009) to link historical BPS *kecamatan* codes (1976–2002), which I use to match 1980 baseline measures to 2000 district identifiers. To align 2000 and 2010 geography, I construct an additional 2000–2010 crosswalk using spatial overlays of ADM4 municipality polygons (see subsection 2.9).

Geospatial remote-sensing data layers. I use the Global Human Settlement (GHS) functional urban area (FUA) polygons to define urban agglomerations. FUAs include commuting zones and are distinct from the tighter GHS Urban Center cores. I aslo use the gridded raster friction surface created by Weiss et al. (2018), combine it with maritime shipping routes to construct the Indonesia-specific friction surface for travel over land and over water, particularly relevant for Indonesia's geography. The friction surface corresponds to the transportation infrastructure in the 1990s based on the main data inputs in it, and given limited time-variant data on the road network expansion, I will use it as the most informative source to compute least-cost travel time between each transmigration village and each urban districts (see subsection 2.4 and subsection 2.6).

2.4 Urban district sample construction

The main analysis focuses on **urban districts on Outer Islands** in Indonesia. I exclude Java and Bali because there are no settled TM villages there in the program sample used here, and I exclude **Papua** because baseline (1980) census measures needed for pre-period controls are unavailable and spatial coverage is sparse in early years.

I define urban areas using a functional urban agglomeration overlap, rather than administrative status. I spatially match GHS functional urban area (FUA) and urban center (UC) polygons to ADM3 (*kecamatan*) boundaries and define the urban sample as districts that **overlap with a FUA or Urban**

[1] Further details can be found in the Appendix subsection A.1.

Center polygon. For the main analysis in this paper, I will focus on the urban-center level analysis, which includes a sample of nearly 500 urban districts that correspond to 180 urban centers in Outer Islands.

Aggregation to urban centers. In the main analysis, I aggregate district-level observations to the **urban center** level. An urban center is defined as a contiguous cluster of urban districts that form a single urban agglomeration, constructed by spatially grouping districts that overlap with the same functional urban area (FUA) polygon or that are spatially contiguous. This aggregation yields **180 urban centers** on the Outer Islands, each representing a local urban labor market composed of one or more contiguous urban districts. I aggregate census outcomes (population, employment, sectoral composition, migration) and the city-level exposure measures to the urban center level by summing total counts for extensive variables (population, employment) and taking appropriately weighted averages for shares and price indices. This approach reduces measurement error from arbitrary administrative boundaries, ensures that economically integrated districts within the same urban agglomeration are treated as a single observation, and better reflects the spatial scale at which urban labor markets and goods markets operate.

Main estimation sample and proximity cutoffs. After constructing the city-village travel-time matrix (described below), I restrict the preferred sample to urban centers whose **nearest TM village is within a travel-time cutoff**. The main results use a cutoff of **300 minutes**, and I report robustness using **180 minutes** (the results are not sensitive to variable cutoffs). This restriction ensures that the analysis is driven by urban centers with plausible economic linkage to nearby TM settlements and meaningful variation in the composition of TM villages around them.

2.5 Urban outcomes

I construct urban outcomes from census data at the district (*kecamatan*, ADM3) level. Outcomes fall into three groups: population and employment growth, sectoral composition, and migration-based measures of demographic change, aggregated for each city c from individual microdata.

Population and employment. I use total population and total employment levels from the 2000 and 2010 censuses. I analyze both (i) endline levels (e.g., $\ln Pop_{c,2010}$, $\ln Emp_{c,2010}$) and (ii) growth over 2000–2010 (e.g., $\Delta \ln Pop_{c,2000\to2010}$, $\Delta \ln Emp_{c,2000\to2010}$). Employment is measured as **total employed persons** (constructed from counts), consistent with the reduced-form urban growth outcomes in the main results. I do not winsorize the outcome variables in the baseline analysis.

Sectoral employment and shares. Using census sector classifications, I construct sectoral employment in levels and shares. In particular, I measure: (i) employment in agriculture, manufacturing, and services; and (ii) service-sector subcomponents used in the main tables (e.g., trade, transportation, and other services), expressed as shares of total employment. I report results both in levels (log counts) and in shares, depending on the specification.

Migration outcomes. Migration measures are constructed from individual-level census microdata (primarily 2010). For each individual observed in 2010, I observe the current district (ADM3) of residence and information on the **district of birth (ADM2)** and **district of residence (ADM2) five years prior**. Since place of birth and prior residence is observed at the higher ADM2 district level, all lifetime and 5-year migration measures are defined as regency-to-regency (ADM2) moves or province-to-province moves. For each urban district c, I construct the share of 2010 residents who: (i) were born in a different province (or a different district), (ii) were born in the same province (or the same district), (iii) were born in Java/Bali, (iv) lived in a different district five years earlier, (v) lived in a different province five years earlier, (vi) lived in Java/Bali five years earlier. These outcomes allow me to quantify whether urban growth is accompanied by in-migration and to characterize the origins of migrants.

Prices, consumption, wages, and production (Susenas 2002/2011; PODES 2003). From Susenas, I construct district-level food/crop price indices using unit values, a housing/rent price index as a proxy for non-tradable/service prices, consumption expenditures and expenditure shares, and sectoral wages (agriculture vs. non-agriculture). I also construct "real" wages by deflating nominal wages with the relevant local price indices. From PODES, I aggregate agricultural production (especially rice) to ADM3 to measure peri-urban agricultural production within urban districts. For a subset of outcomes, I additionally construct an indicative "rice surplus" measure (production minus consumption) combining PODES production with Susenas consumption aggregates.

2.6 Main explanatory variable: agro-climatic similarity

The main explanatory variable captures quasi-exogenous variation in the **agricultural productivity** of Transmigration villages arising from differences in the match between migrants' origin environments and destination environments. I use the agro-climatic similarity index developed in Bazzi et al. (2016), which was designed to measure **skill transferability** of agricultural migrants across space.

Village-Level Agro-Climatic Similarity. Following Bazzi et al. (2016), agro-climatic similarity in a TM village v aggregates the agro-climatic "distance" between each migrant's origin district and the destination village. Let x_d denote a vector of agro-climatic characteristics in origin district d (e.g., soil and terrain, rainfall and temperature moments, and stable geographic attributes such as distance to rivers/coast), and let x_v denote the corresponding vector for destination village v. Let π_{dv} be the share of migrants in village v born in district d. The village-level index is:

$$\text{Agroclimatic Similarity}_v = AS_v = (-1) \times \sum_{d=1}^{I} \pi_{dv}\, d(x_d, x_v), \tag{1}$$

where $d(\cdot)$ is an agro-climatic distance metric. The leading (-1) converts a dissimilarity measure into a similarity measure. Bazzi et al. (2016) show that this index predicts realized agricultural success of TM settlements (including higher productivity outcomes), motivating its use here as a proxy for the

intensity of the rural productivity shock. Figure 3 (Panel A) shows the variation in the agro-similarity across settled Transmigration villages.

Aggregate Village Productivity Across Villages Around Each City

To map the village-level productivity variation into an urban shock, I construct a city-level exposure measure that aggregates AS_v across nearby TM villages. The aggregation follows the inverse-distance weighting logic in **??**, with two key features: (i) villages closer to a city receive greater weight, and (ii) larger TM villages receive greater weight. The variation in the resulting aggregated index of agro-similarity is shown in Figure 3 (Panel B).

I take the village-level agro-similarity index AS_v (the village-level agro-climatic similarity index, see **??**) and aggregate it across villages around the city by taking a "spatial" average, i.e., taking an inverse-distance-weighted average of migrant productivity across village weighted by village population (essentially taking a spatial-average productivity of transmigrants) (See Equation 2). Thus, I account for the size of the village, as well as the village proximity to the city. Figure **??** illustrates the spatial matching of a sample urban agglomeration to nearby villages using a straight-line Euclidean distance. For each city, I calculate the **average village productivity** proxied by the spatial average of migrant agro-similarity across villages around the city:

$$TMAgroSimilarity_c^{IDW} = \frac{\sum_{v=1}^{V_c} \frac{1}{d_{cv}} N_v \cdot AS_v}{\sum_{v=1}^{V} \frac{1}{d_{cv}} N_v} \in [0,1] \tag{2}$$

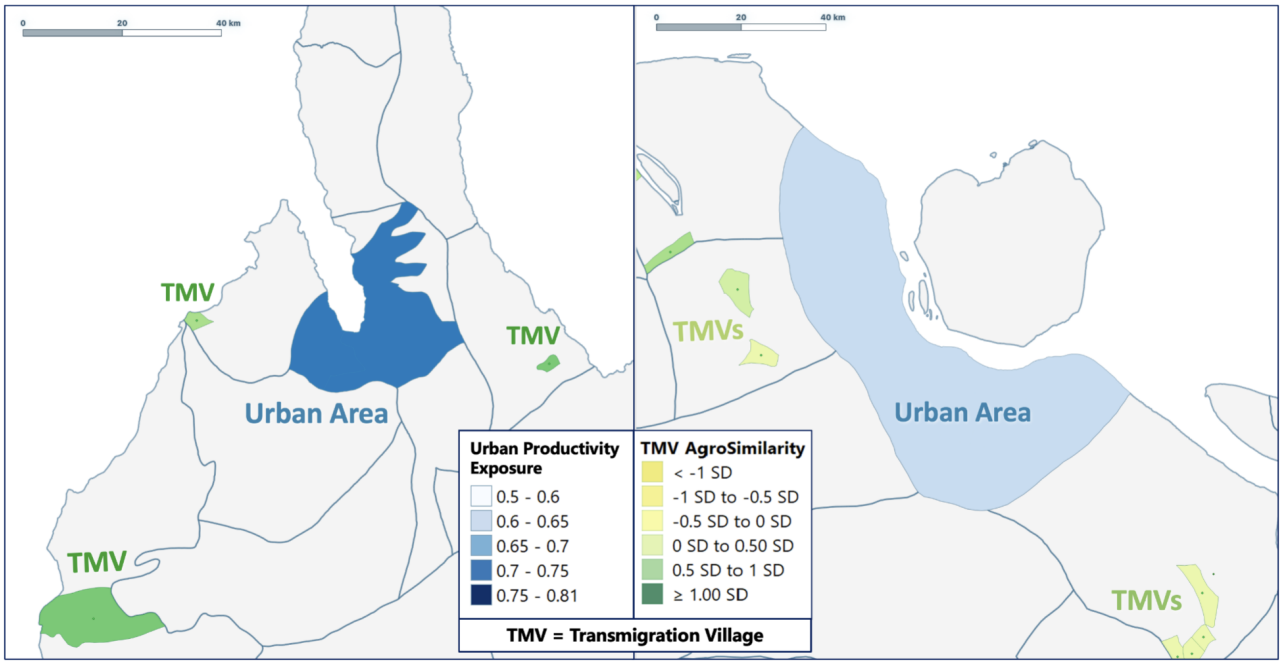
where $TMAgroSimilarity_c^{IDW}$ is the average transmigrant agro-similarity around each city c measured between $[0,1]$ aggregated from the transmigrant agro-similarity in every village AS_v, weighted by village population N_v and the inverse of distance d_cv for each city-village pair and scaled by the total number of transmigrants around each city. d_{cv} is the distance from the centroid of city municipality c to the centroid of village v, N_v is the number of transmigrants in village v and V_c is the set of all villages included in the calculation[2] around city c, AS_v is the average agroclimatic similarity across migrants in village v (adopted from Bazzi et al. (2016)), IDW stands for inverse distance weighting used for spatial averaging.

Maps in Figure 5 show two sample regions: both cities are surrounded by a number of Transmigration villages in their vicinity, but one receives a larger agricultural productivity shock (better match quality in the villages), and the other gets a smaller productivity shock as, on average, less productive villages surround it.

Winsorization and scaling. In the main regressions, I winsorize the constructed exposure measures (agro-similarity exposure and transmigrant count exposure) and the aggregated village controls at

[2]For computational simplicity reducing the dimensionality of the city-village pairs, I impose a cutoff for inclusion of 1,000 km. Note that any village population or productivity beyond 1,000km will already be coming into the estimation with a weight $< \frac{1}{1000}$, accounting for less than 0.1% of each aggregated measure.

Figure 5: A Sample Area of Urban Agglomerations & Nearby Transmigration Villages



Notes: Urban districts (kecatamans) in blue and TM village in green, each highlighted by productivity: urban productivity exposure in urban districts (in shades of blue) vs. transmigration village productivity (in yellow-green).

the 1st and 99th percentiles to reduce sensitivity to extreme observations. I report the agro-similarity exposure both in levels and standardized (z-score) form.

2.6.1 Cost-Distance Travel Time between Cities & Villages

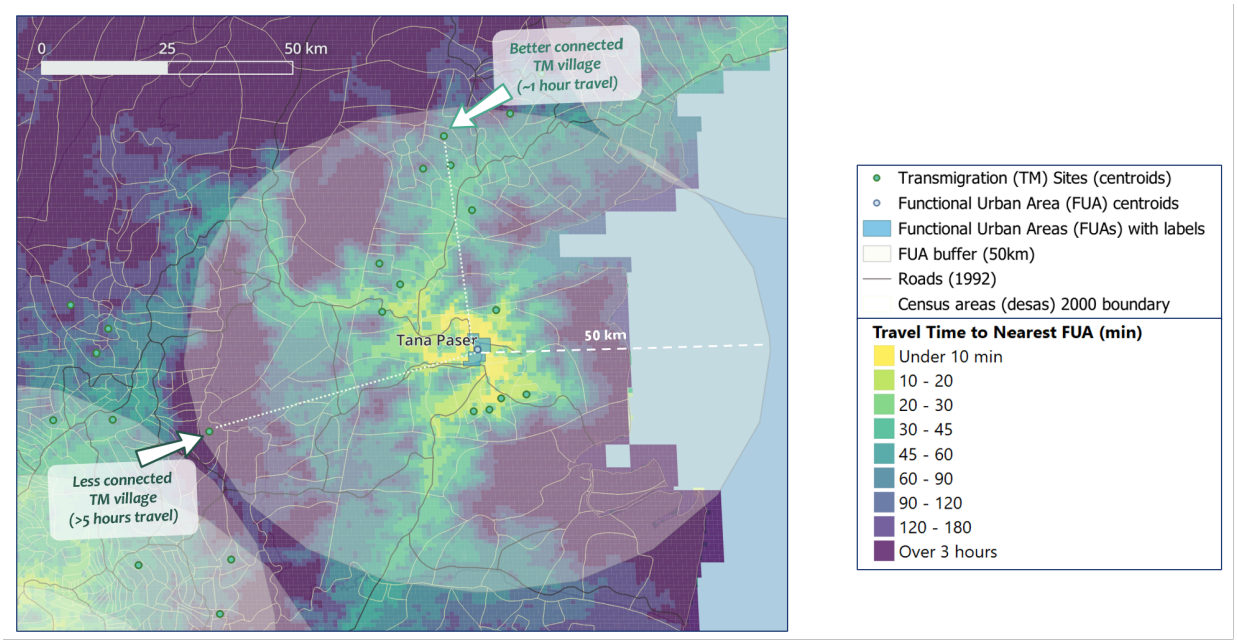
Crucially, I measure proximity using **cost-distance travel time** rather than straight-line distance. I compute the minimum travel time between each TM village v and each urban district c on a 1km grid friction surface. The friction surface is based on the GHS travel speed raster (built from road-network and terrain inputs) and I augment it by incorporating **marine travel links** to reflect Indonesia's archipelagic geography. This delivers a full city–village travel-time matrix τ_{cv}, which is used both to define the main estimation sample (nearest TMV within 180/300 minutes) and to weight villages in the exposure measures.

To calculate cost-distance, I use the travel speed friction surface from the previous work estimating travel time to cities (Weiss et al. 2018), currently provided as the raw raster from the. While this friction surface is representative of 2015, it uses multiple data sources from the early 1990s as inputs, including the road network and waterway data, which are closer to my baseline period and to currently available data.

The map in Figure 6 illustrates the geographic distribution of transmigration (TM) sites by travel time in a sample region in relation to one Functional Urban Areas (FUAs) Tana Paser. The map shows a clear variation in the connectivity of two sample villages, both within the 50km radius. The highlighted village in the south-west (SW) is over 3 hours away in travel time while the sample village

in the North, also around 45-48kn away in straight line distance is just about an hour away in travel time since it is located closer to the road network (road network displayed as of 1992).

Figure 6: Sample Area: Euclidean Distance vs. Travel-Time Cost-Distance Measure Comparison



Notes: The figure's background illustrates the variability in the travel time from each census area (desa) to the nearest Transmigration (TM) site, ranging from under 10 minutes (yellow) to over 3 hours (dark purple). TM sites are identified with their centroids in green. The map also depicts urban areas (FUAs) in blue with a 50km buffer to compare the matching approach using the Euclidean-distance buffer around the urban area boundary with a cost-distance estimate. The map also highlights roads as of 1992, FUA buffers within 50 km, and census area boundaries from the year 2000. Urban area labels are added based on the GHS data. Source: original calculations using the friction surface from Weiss et al. (2018).

As such, simple distance weighting would not completely resolve this, since we can also notice that within the buffer radius, some TM sites are effectively much closer to the city than others based on the travel time conditional on similar Euclidean distance (e.g., close to the edge of the buffer). Therefore, using weights by cost-distance will be a meaningful next step in creating the exposure shock variables for cities based on their composite exposure to nearby TM sites.

2.7 Transmigration Village Population

To separate the role of **productivity** from that of **scale**, I construct a second, purely quantitative exposure measure: the number of individuals initially placed in nearby TM villages. Village-level initial placement counts N_v come from the Transmigration Census. I then aggregate N_v to the urban district using the same travel-time weighting scheme as for agro-similarity exposure (see **??**). This controls for the possibility that cities surrounded by more TM settlements experience different population and demand dynamics independent of the productivity shock.

To control for the overall number of migrants resettled around a given city c under the Transmigration Policy ($TransMigrants$), I calculate the inverse-distance-weighted number of transmigrants across villages around each city. In the main specification, I use the number of initial transmigrants

placed in each village as the inverse-distance-weighted average (IDW) of the village population at the time of the establishment of each village v (N_v), accounting for the distance to it d_{cv}:

$$TransMigrants_c^{IDW} = \sum_{v=1}^{V} \frac{1}{d_{cv}} N_v, \tag{3}$$

where $TransMigrants_c$ is the aggregated exposure measure to TM policy calculated as the total number of transmigrants in each village N_v around city c, inversely weighted by distance to the city, d_cv, which is the distance from the centroid of city municipality c to the centroid of village v.

2.8 Other controls aggregated from villages

I include a rich set of controls capturing agro-climatic and geographic features of TM villages to ensure that identifying variation in the agro-similarity exposure is not mechanically driven by underlying land quality or geography. These controls include the components used to construct the agro-climatic distance metric in Bazzi et al. (2016) (e.g., soil and terrain attributes; rainfall and temperature patterns; and stable geographic characteristics such as distances to rivers and coasts). I aggregate these village covariates to the urban district using the same travel-time weights used for the main exposure measure. I refer to this set as "agro-climatic TMV controls" in the tables.

2.9 City-level baseline controls

Finally, I include city-level baseline controls to account for heterogeneity in initial conditions and to improve precision. One important set of controls include **1980 baseline levels** of population and employment for each urban district, constructed from the 1980 census and matched forward to consistent ADM3 units.[3] Where relevant, I also include baseline sectoral composition measures and baseline migration measures to address the possibility that some cities were already on different trajectories prior to exposure to nearby TM settlements.

3 Empirical Strategy and Results

My model specification leverages the fact that Transmigrants were quasi-randomly assigned to destination villages, resulting in partially exogenous variation in their productivity in TMVs depending on agro-similarity with their districts of origin. This agro-similarity index reflects the alignment of migrants' agricultural skills with the conditions of their new locations compared to the agro-climatic conditions in their districts of origin, thus serving as a proxy for migrant productivity (following the detailed description of the agro-similarity measure in Section **??**). By comparing cities with variable productivity levels of nearby TMVs, I aim to disentangle the effects of rural productivity on urbanization.

[3] Papua is excluded as these baseline measures are unavailable.

3.1 Main Specification to Estimate the Effect of AgroSimilarity on Urban Growth

My main estimating equation is specified as follows with the quasi-exogenously determined measure of agricultural similarity across villages around each city, the total number of migrants in nearby villages, along with additional covariates:

$$Urban\ Outcome_c = \beta\ AgSimilarity_c^{IDW} + \gamma\ TransMigrants_c^{IDW} + C_c\eta + X_c^{IDW}\nu + \epsilon_c \qquad (4)$$

where $Urban\ Outcome_c$ is an endline urban outcome at the urban district level (e.g., population and employment in 2000 or 2010 – the 2 available census years)) in city c, $TransMigrants_c$ is the inverse-distance-weighted sum of migrants in Transmigration villages (spatial average), $AgSimilarity_c$ is the inverse-distance-weighted **and** village-population-weighted measure of agro-similarity (following Equation 2, measured between 0 and 1, will be shown as is and standardized relative to the mean on the full sample), C_c is the city's baseline characteristics (population, employment in 1980), X_c^{IDW} contains spatial-averaged agricultural suitability village characteristics (e.g., climate, soil, geography), incl. all characteristics used to construct the similarity measure, with the migrant count constructed in the following manner (as in Eq. (3)).

To account for the differential distribution of villages around cities, as some cities might have received a differential number of villages in their immediate proximity around them depending on the greenfield land availability during the site selection, I calculate the overall exposure measure for cities to allow for the effects of agro-similarity to be conditional on overall village population, identifying its causal effect more precisely. Note that the overall exposure is not necessarily exogenous since the selection of sites for village placement is not random, but as I demonstrated in the previous section, it is the matching of migrant skills to destinations that generates the quasi-random variation I am utilizing in the main identification. See a sample region of higher and lower-productivity villages and exposed cities in Figure 5.

The city-level TM policy exposure in Equation (3) accounts for the variable (and not necessarily exogenous) placement of villages relative to cities, such that the coefficient on this variable in the main regression does not capture a causal effect but instead controls for the overall number of migrants. Turning to the agro-similarity measure, the village-level data are aggregated using an Inverse-Distance Weighted (IDW) sum of population times average village productivity, which assigns higher weights to closer villages, assuming that similarly sized villages have a stronger impact on the city when located closer.

The final exposure index for each city is determined by the total TM population, with each village's contribution weighted by its inverse distance to the city. This is complemented by a weighted average of agro-similarity across villages, incorporating both population numbers and productivity potential into the city-level impact assessment. Together, these steps create a comprehensive index that reflects the intensity of policy exposure and its potential ramifications on urban growth and development.

Note that the interaction term (if included) would be composed of an interaction of two **continuous** variables: the continuous index of AgroSimilarity and the total number of migrants, which does not add a clearly interpretable coefficient. Thus, the interaction term will not be included as the third term in the main specification.

The inclusion of agricultural suitability controls X_{cs} implies that the quasi-exogenous variation is now primarily driven by the variability in migrant characteristics in districts of origin, which I claim are as-good-as-randomly assigned to specific destinations. The inclusion of island fixed effects is tested in one of the alternative specifications, but will not be included

The main outcome of interest is the urban population growth over the period 1980-2020. I can also test a later post-policy period separately given my 5-year panel dataset of remote-sensing-based population estimates and available 2000 and 2010 census data.

3.2 Urban Areas Sample

The unit of analysis is the urban center, with the data aggregated across urban districts (*kecamatans*, ADM3) overlapping with a given urban center as defined by the GHS Urban Center polygons. Crosswalks across years are made first at the district level: I construct a harmonized set of ADM3 identifiers across census waves, define urban districts using functional urban area (FUA) footprints, and then restrict the estimation sample to urban places plausibly connected to Transmigration villages (TMVs) using a travel-time (cost-distance) proximity criterion.

Geographic scope. The main analysis focuses on urban districts on Indonesia's **Outer Islands**. I exclude Java and Bali because there are no TM villages in the study sample in these provinces, and I exclude Papua because baseline (1980) census measures required for pre-period controls are unavailable and spatial coverage is sparse.

Harmonizing district identifiers over time. Administrative boundaries and codes change over time, so I express outcomes and baseline controls on a consistent set of ADM3 identifiers. First, I use the merged BPS *kecamatan* code crosswalk assembled by Olken (2009) (covering 1976–2002) to match 1980 census district identifiers to the 2000 ADM3 codes used as the reference geography in the main analysis. Second, I construct a 2000–2010 crosswalk using GIS spatial overlays of ADM4 municipality polygons: I identify 1:1 matches at the municipality level across vintages and then aggregate matched 2010 municipalities to the corresponding 2000 district identifiers.[4]

Defining urban districts using GHS Urban Centers. I define the urban sample using functional urban area (FUA) polygons from the Global Human Settlement (GHS) framework (Moreno-Monroy et al. 2021). I spatially overlay FUA polygons with ADM3 boundaries and include districts that overlap an FUA. To avoid including districts that only touch an FUA boundary, I exclude "border-

[4]This procedure ensures that 2010 outcomes can be consistently aggregated to the 2000 district system even when ADM3 units split or merge between 2000 and 2010.

only" overlaps.[5] The main analysis focuses on the Outer Islands and excludes Java and Bali (no TMVs in the study sample) as well as Papua (baseline 1980 measures are not available and spatial coverage is sparse).

Urban population share within ADM3. Even among ADM3 units that overlap an FUA, there is heterogeneity in the intensity of urbanization within the district. I construct a continuous measure of the *urban population share* for each district using the GHSL 1km population grid. For each ADM3 polygon c, I partition the polygon into two mutually exclusive areas: the portion that overlaps the FUA footprint ($c \cap FUA$) and the portion outside the FUA ($c \setminus FUA$). I then overlay each part with the GHSL population raster and sum population counts within each part:

$$Pop_c^{FUA} = \sum_{g \in (c \cap FUA)} Pop_g, \qquad Pop_c^{Total} = \sum_{g \in c} Pop_g,$$

where g indexes 1km grid cells. The urban population share is

$$UrbanShare_c = \frac{Pop_c^{FUA}}{Pop_c^{Total}}.$$

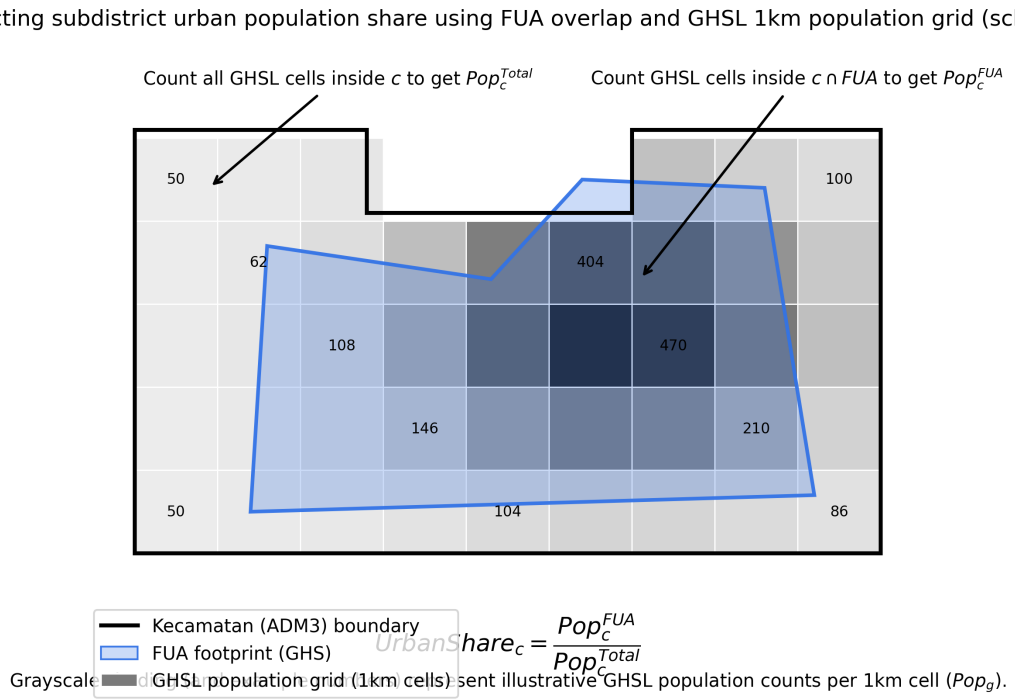
I compute terciles of $UrbanShare_c$ in the **1980 baseline sample** and reuse the same tercile cutoffs in later years to maintain comparability across time.

Proximity restriction to TM villages. After matching TM villages to urban districts using a travel-time matrix (constructed from a friction surface based on the GHS travel-speed raster augmented with marine travel links), I restrict attention to urban districts with plausible economic linkage to nearby TM settlements. Specifically, I define the nearest-village travel time for each city c as $\min_{v \in TMV} d_{cv}$, where d_{cv} is the least-cost travel time in minutes between the urban district and TM village centroids. The main estimation sample uses $\min_v d_{cv} \leq 300$ minutes, and I report robustness using $\min_v d_{cv} \leq 180$ minutes (and wider cutoffs in appendix robustness). Many urban districts fall outside this proximity cutoff and are excluded from the main sample; towns with zero nearby TMVs are not displayed in the agro-similarity index visualizations (the latter excluded from Figure 4, where the agro-similarity index requires at least one nearby TMV).

Urbanization Intensity and Sectoral Composition. Figure 8 and Figure 9 report sectoral employment shares in 1980 and 2010, respectively, within each tercile for urban districts whose nearest TMV is within 300 minutes. Two patterns stand out. First, in 1980, many "urban" ADM3 units in the Outer Islands remain highly agricultural: in the lowest-urbanization tercile ($UrbanShare_c < 60\%$), agriculture accounts for roughly 70% of employment, while services account for about 25%. Second, between 1980 and 2010, employment shifts toward services, especially in more urbanized districts: for

[5] Given problematic geometries in the BPS shapefile, I operationalize the bordering-polygon exclusion by including districts with a minimum overlap of 1% of the ADM3 polygon area. This threshold operationalizes the conceptual "overlap" criterion while excluding cases of border-only contact due to imperfect geometries; results are robust to alternative overlap cutoffs.

Figure 7: Schematic: Constructing district urban population share using FUA overlap and GHSL 1km population grids



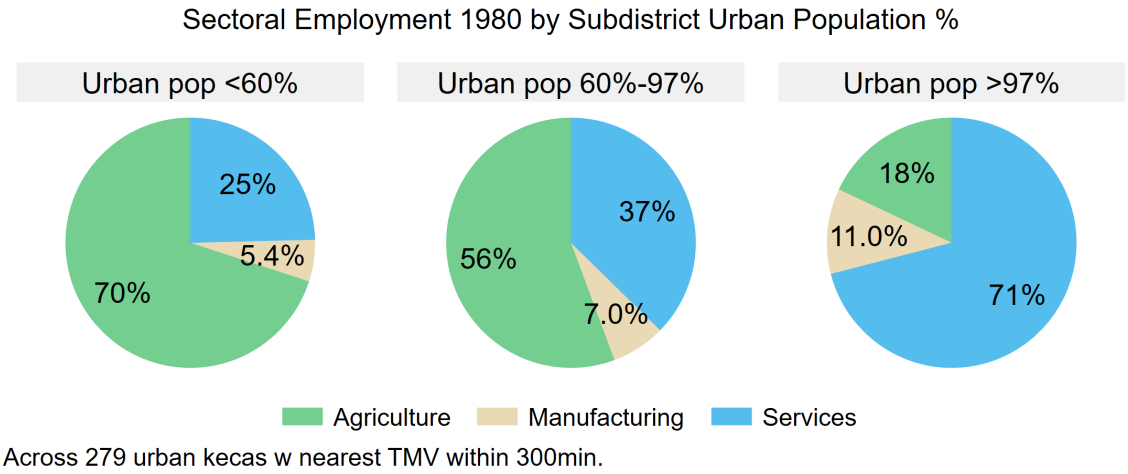
Notes: The figure illustrates the construction of $UrbanShare_c$ for a district c. Grayscale cells represent GHSL population counts per 1km grid cell (Pop_g), with darker shading indicating higher population. $Pop_c^{Total} = \sum_{g \in c} Pop_g$ sums all cells inside the district. $Pop_c^{FUA} = \sum_{g \in c \cap FUA} Pop_g$ sums cells inside the FUA-overlap portion of the district. The urban population share is $UrbanShare_c = Pop_c^{FUA} / Pop_c^{Total}$. Cell numbers are illustrative and shown in a subset of cells for readability.

example, in the middle tercile (urban population 60–97%), services rise from about 37% to 58% while agriculture falls from about 56% to 36%. These descriptive patterns motivate focusing on both (i) urban population and employment growth and (ii) sectoral transformation in the analysis.

3.3 Framework: Mechanisms Linking Rural Productivity to Urban Outcomes

Figure 10 provides a mechanism map linking a rural agricultural productivity increase to urban outcomes. Higher rural productivity raises food production, increasing agricultural output and lowering agricultural prices. Through trade integration, cheaper food reduces the urban cost of living and frees household budgets for non-food consumption, supporting service-sector expansion. In parallel, urban labor reallocates away from agriculture and toward services as relative returns change. Together, these goods-market and labor-market forces contribute to urban economic growth and structural transformation. The figure also highlights a competing migration channel: higher rural productivity raises village incomes, which may either relax liquidity constraints and facilitate migration, or strengthen incentives to remain in rural areas. In addition, migration into growing cities may originate disproportionately from other, less productive locations. Because these forces operate in opposite directions, the overall migration response is ambiguous and motivates empirical investigation.

Figure 8: Sectoral employment in 1980 by baseline district urban population share



Notes: Pie charts show employment shares in agriculture, manufacturing, and services in 1980 for urban districts (ADM3) grouped by terciles of baseline district urban population share ($UrbanShare_c$). Sample includes 279 urban districts whose nearest TM village is within 300 minutes of travel time.

Summary Statistics

Table 1 summarizes key characteristics of urban districts in Indonesia, highlighting significant population and employment growth between 1980 and 2000. While agriculture remained a dominant sector of employment in 2000, there is notable variation in sectoral composition, with services and manufacturing employment being relatively lower. Migration indicators show limited inter-provincial movement, but proximity to Transmigration Villages (TMVs) and agro-similarity metrics suggest potential rural-to-urban linkages influencing district outcomes.

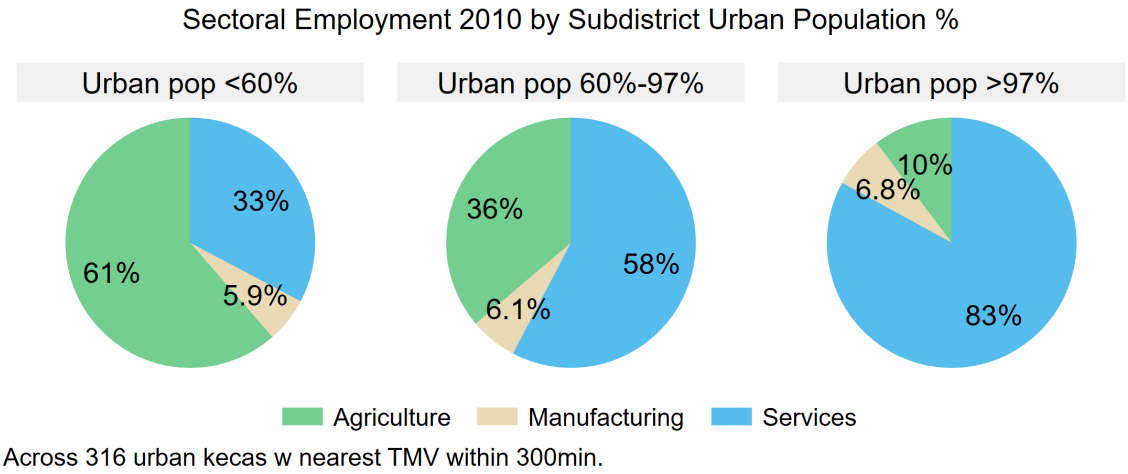
The summary statistics table includes the largest possible sample of urban districts that were successfully matched across census years (1980, 2000, 2010). Note that some of them are located relatively far from the nearest TMV (TMVs within 1,000 km are still included in the construction of the TMP exposure and the aggregated village productivity, but villages where the nearest TMV is over 1,000 km away are essentially excluded, with such distant villages contributing a negligible share to the weighted sum given the large number of closer villages with much higher weights).[6]

Placebo Control Test for Districts at Baseline

Table 2 provides a placebo test for the effect of TMV agro-similarity on baseline urban center characteristics in 1980. The observations are urban center-level aggregates of 1980 individual census records regressed on average agro-similarity across villages. The results show no significant relationship for most characteristics, such as population, employment, sectoral employment shares, (most) migration rates, population born in Java/Bali. This indicates that agro-similarity was not strongly associated

[6]

Figure 9: Sectoral employment in 2010 by baseline district urban population share



Notes: Pie charts show employment shares in agriculture, manufacturing, and services in 2010 for urban districts (ADM3) grouped by terciles of baseline district urban population share ($UrbanShare_c$). Sample includes 316 urban districts whose nearest TM village is within 300 minutes of travel time. The sample is larger as there are more urban subdistrics covered by the 2010 census, the final analysis sample will exclude those that are not matched to districts covered in the 1980 census.

with pre-existing urban center characteristics, suggesting that the observed effects on urban outcomes from the productivity shock are not driven by baseline differences in urban characteristics.

3.4 Main Reduced-Form Results

?? presents the baseline regression estimates examining how agricultural productivity in nearby Transmigration villages influences urban population, employment, and sectoral employment at the urban center level in 2010. Across the specifications, the coefficient on the TMV agro-similarity measure (standardized) is positive for population and total employment, though these effects are muted and often statistically insignificant. By contrast, the effects on sectoral composition are pronounced: agricultural employment declines significantly, while manufacturing and service employment increase substantially. This pattern is consistent with limited overall population response to the rural productivity shock alongside strong structural transformation — a key finding suggesting that the dominant channel operates through trade and goods markets rather than through large-scale migration. These estimates control for the size of the transmigrant population, baseline conditions in 1980 (population and/or employment), and a rich set of agro-climatic controls.

Evidence of Structural Transformation as the Channel

In Table 4, the focus shifts to employment shares rather than levels. The share of service employment rises, though the urban center-level estimate does not reach conventional significance levels Table 4. At the finer FUA-district level (Table 6), the service share effect is larger and statistically significant. This finding implies that, as rural areas become more productive, the composition of

Table 1: Urban District Summary Statistics

	mean	sd	p50	min	max	count
Urban District population (1980)	49,374	35,432	40,740	20	237,900	461
Urban District population (2000)	76,716	55,956	60,574	1,347	426,205	461
Total employed (1980)	13,903	9,764	11,800	0	56,100	461
Total employed (2000)	32,395	24,081	26,139	616	254,110	461
City Distance to Nearest TMV (km)	85	87	58	1	534	462
TMV Agro-Similarity (std z-score, pop IDW)	-0.02	1.08	-0.41	-2.34	2.49	462
Total Transmigrants (IDW)	2,081	1,316	1,701	138	6,742	462
Employed % in ag (2000)	0.466	0.270	0.510	0.012	0.941	461
Employed % in manuf (2000)	0.056	0.064	0.036	0.001	0.510	461
Employed % in serv (2000)	0.352	0.209	0.298	0.025	0.914	461
% born on Java/Bali (proxy transmigrants) (1980)	0.063	0.101	0.019	0.000	0.613	457
% born on Java/Bali (proxy transmigrants) (2000)	0.056	0.070	0.025	0.000	0.429	461
% born in a diff province (2000)	0.113	0.122	0.065	0.003	0.646	461
% living in a diff province 5y ago (2000)	0.034	0.037	0.021	0.000	0.398	461

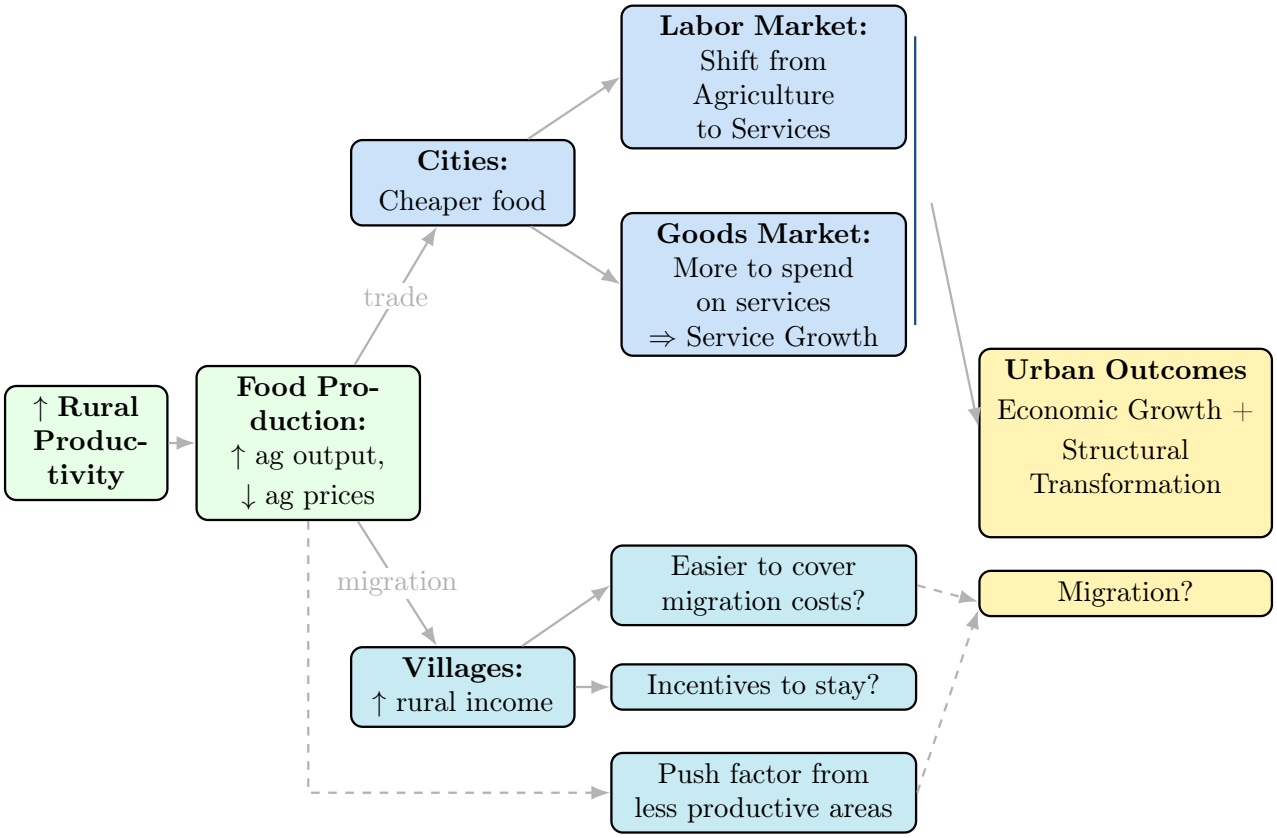
Census 1980 provides a 5% sample of population, currently aggregated to total population comparable levels without weights. IDW = inverse-distance weighted

Table 2: Placebo Baseline Test: Village Productivity Across 1980 urban districts

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	pop	emp	Ln Emp: ag	manuf	serv	% Migr JvBl 5yr	% Migr prov 5yr
Standardized values of	0.267	0.123	-0.227	0.333	0.382	0.008	0.016
AgSimilarity (IDW)	(0.291)	(0.292)	(0.334)	(0.307)	(0.284)	(0.009)	(0.011)
Transmigrant count (IDW)	✓	✓	✓	✓	✓	✓	✓
Agro-Clim. TMV controls	✓	✓	✓	✓	✓	✓	✓
Mean dep var	8.382	7.134	6.606	4.031	5.846	0.025	0.044
Std dev dep var	1.02	0.99	1.01	1.55	1.36	0.04	0.05
Observations	122	122	122	119	122	122	122

Notes: Outcomes in cols from baseline year 1980 across urban districts (kecatamans/adm3), aggregated to urban centers. Nearest TMV within 300min. Spatial Conley SEs (dist cutoff 200km). TMV = Transmigration Villages, IDW = inverse-distance weighted, FUA = Functional Urban Area (from GHS = Global Human Settlements), PC = Population Census. Agro-climatic TMV controls include a rich set of soil, climatic, and geogr. controls used to construct agro-similarity. *** 1%, ** 5%, * 10% significant.

Figure 10: Mechanism Map



Notes: A rural productivity shock increases food production and lowers agricultural prices, generating urban spillovers through cheaper food, labor reallocation from agriculture to services, and increased demand for services. Rural income changes may also affect migration through offsetting liquidity and incentive-to-stay forces, making the net migration response ambiguous.

nearby urban labor markets tilts toward services, consistent with broader service-driven structural transformation and urbanization.

3.5 Contracting Urban Agriculture: Food Market & Agricultural Employment

A key channel linking rural productivity to urban outcomes is the urban food market. Higher rural productivity can raise agricultural supply and improve market access, affecting the delivered price of food and the real income of urban consumers. Using household survey data on unit values and expenditures, I test whether urban centers exposed to higher rural productivity have different food prices and consumption patterns.

The results show that exposed cities do not experience a collapse in food consumption; instead, food consumption and total food spending are broadly stable. At the same time, food becomes relatively less expensive (or less salient) compared to non-food: the relative price of agricultural goods shows a positive but imprecise point estimate (Figure 11), while the food spending share declines significantly,

Table 3: Effects on Urban Pop & Employment in 2010 (aggregated to urban centers)

	Population + Employment			Sectoral Employment (Ln)			
	(1) ln pop	(2) ln employed	(3) unemp rate	(4) agric	(5) manuf	(6) constr	(7) serv
Standardized values of AgSimilarity (IDW)	0.043 (0.027)	0.041 (0.033)	0.003 (0.006)	-0.097* (0.049)	0.112*** (0.039)	0.075 (0.061)	0.074* (0.045)
Transmigrant count (IDW)	✓	✓	✓	✓	✓	✓	✓
Baseline Pop 1980	✓	✓	✓	✓	✓	✓	✓
Urban share 1980	✓	✓	✓	✓	✓	✓	✓
Total Empl 1980	✓	✓	✓	✓	✓	✓	✓
Agro-Clim. TMV controls	✓	✓	✓	✓	✓	✓	✓
Mean dep var	11.90	10.99	0.12	9.90	7.58	8.06	10.19
Std dev dep var	0.84	0.82	0.04	0.85	1.32	1.10	0.94
Observations	116	116	116	116	116	116	116

Notes: Outcomes in cols in 2010 across urban districts (kecatamans/adm3), aggregated to urban centers. Nearest TMV within 300min. Spatial Conley SEs (dist cutoff 200km). TMV = Transmigration Villages, IDW = inverse-distance weighted, FUA = Functional Urban Area (from GHS = Global Human Settlements), PC = Population Census. Agro-climatic TMV controls include a rich set of soil, climatic, and geogr. controls used to construct agro-similarity. *** 1%, ** 5%, * 10% significant.

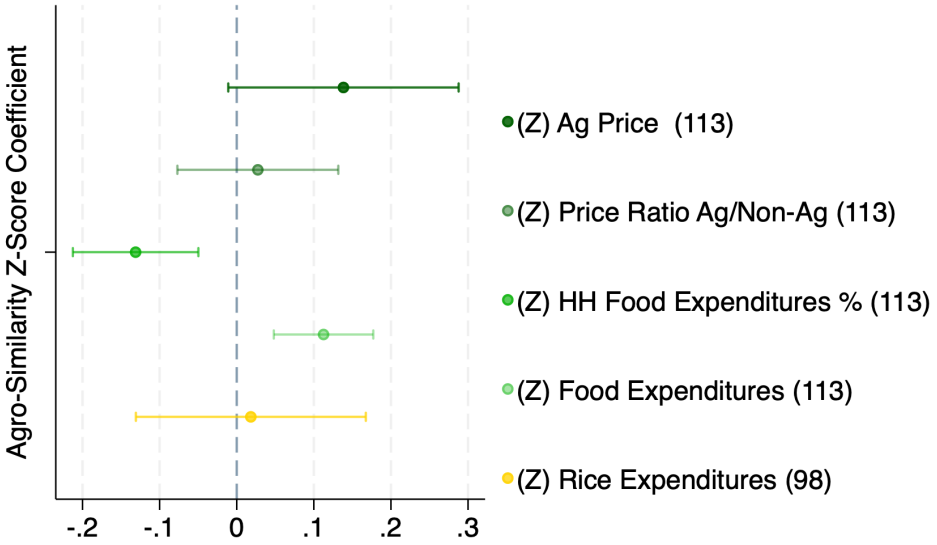
Table 4: Effects on Urban Pop & Employment in 2010 (aggregated to urban centers)

	Population + Employment			Sectoral Empl Shares			
	(1) ln pop	(2) ln employed	(3) unemp rate	(4) agric %	(5) manuf %	(6) constr	(7) serv %
Standardized values of AgSimilarity (IDW)	0.043 (0.027)	0.041 (0.033)	0.003 (0.006)	-0.029** (0.012)	0.002 (0.003)	0.002 (0.004)	0.020 (0.013)
Transmigrant count (IDW)	✓	✓	✓	✓	✓	✓	✓
Baseline Pop 1980	✓	✓	✓	✓	✓	✓	✓
Urban share 1980	✓	✓	✓	✓	✓	✓	✓
Total Empl 1980	✓	✓	✓	✓	✓	✓	✓
Agro-Clim. TMV controls	✓	✓	✓	✓	✓	✓	✓
Mean dep var	11.896	10.988	0.122	0.389	0.045	0.061	0.473
Std dev dep var	0.839	0.822	0.040	0.187	0.040	0.028	0.140
Observations	116	116	116	116	116	116	116

Notes: Outcomes in cols in 2010 across urban districts (kecatamans/adm3), aggregated to urban centers. Nearest TMV within 300min. Spatial Conley SEs (dist cutoff 200km). TMV = Transmigration Villages, IDW = inverse-distance weighted, FUA = Functional Urban Area (from GHS = Global Human Settlements), PC = Population Census. Agro-climatic TMV controls include a rich set of soil, climatic, and geogr. controls used to construct agro-similarity. *** 1%, ** 5%, * 10% significant.

consistent with a Stone-Geary mechanism where households reallocate marginal spending toward non-food items as real income rises.

Figure 11: Urban food prices and consumption patterns



Notes: Outcomes in standardized z-scores across urban centers (UCs) (aggregated from districts) (Susenas 2011).
Urban districts (kecamatans/adm3) based on the 2000 district boundaries.
Sample size in parentheses next to outcomes. Controlling for TMV agro-climatic features,
.
Nearest TMV within 300min. 90% CI. Spatial Conley SEs (dist cutoff 200km).

City-level rice production also falls, and measures of rice surplus (production minus consumption) decline (Figure 12), consistent with the idea that cities rely more on rural suppliers rather than producing staples locally. Interestingly, agricultural wages in cities can rise even as agricultural employment shrinks, which is consistent with a standard marginal-product logic: when fewer workers remain in city farming (and/or land is reallocated toward non-agricultural uses), the marginal product of the remaining agricultural labor can increase.

3.6 Services expand with rising non-tradable prices and wages

The previous subsections suggest that exposed cities reallocate away from agriculture and toward services. Here I test whether this expansion is accompanied by changes in non-tradable prices and wages that are consistent with rising local demand and labor absorption.

Using household survey data, I find that non-food expenditures rise in exposed urban centers. At the same time, the price of local non-tradables (proxied by a housing price index) increases strongly. This is consistent with a population inflow and increased demand for locally supplied goods and services, which must clear in local markets. Finally, non-agricultural wages rise modestly, suggesting that labor demand shifts outward in expanding urban service markets and, in equilibrium, agglomeration forces may partially offset the wage-depressing effect of an increased labor supply, consistent with the model predictions.

Figure 12: Urban Agricultural Sector: Goods & Labor Market



Notes: Standardized outcomes from Susenas 2011 across urban districts.
Sample: urban districts (kecamatans/adm3) based on the 2000 district boundaries.
Sample size in parentheses next to outcomes. Controlling for TMV agro-climatic features,
tm_migrants_000_IDW ln_employed_1980 ln_pop_1980 empl_agric_pct_1980 urban_pop_pct_1980 .
Nearest TMV within 300min. 90% CI. Spatial Conley SEs (dist cutoff 200km).

Table 5: Urban Service Sector

	Employment		Prices & Spending		Wages		
	(1) ln serv	(2) serv %	(3) ln nonfood expend	(4) price housing	(5) ln w nonag	(6) w n-ag/w Ag	(7) ln w
TMV Agro-Similarity (std	0.107**	0.018	0.066***	0.146*	0.038	-0.014	0.039*
z-score, pop IDW)	(0.049)	(0.014)	(0.025)	(0.079)	(0.026)	(0.047)	(0.020)
Transmigrant count (IDW)	✓	✓	✓	✓	✓	✓	✓
Base pop 1980 (Ln)	✓	✓	✓	✓	✓	✓	✓
Base empl 1980 (Ln)	✓	✓	✓	✓	✓	✓	✓
Agro-Clim. TMV controls	✓	✓	✓	✓	✓	✓	✓
Mean dep var	10.190	0.473	14.137	1.897	14.115	1.429	14.029
Std dev dep var	0.940	0.140	0.367	0.952	0.276	0.601	0.308
Observations	116	116	113	113	113	113	113

Notes: Outcomes in cols in 2010 across urban districts (kecatamans/adm3), aggregated to urban centers. Nearest TMV within 300min. Spatial Conley SEs (dist cutoff 200km). Agro-climatic TMV controls include a rich set of soil, climatic, and geogr. controls used to construct agro-similarity. *** 1%, ** 5%, * 10% significant.

Table 6 separates the service growth into the largest service types: trade, transportation, and higher-skilled services. The service employment growth (both in overall employment and their shares) is primarily driven by growth in trade and transportation, overall consistent with the productive TM villages driving the growth in local transportation hubs and trading centers from increased agricultural output.

Table 6: Urban Employment by Service Type 2010 (aggregated to urban centers)

	All (Ln)	Service Types (Ln)			All (%)	Service Types (% of All Empl)		
	(1) Ln Serv	(2) Trade	(3) Transport	(4) Skilled	(5) % Serv	(6) Trade	(7) Transport	(8) Skilled
TMV Agro-Similarity	0.079*	0.069	0.103***	0.086	0.020	0.004	0.004**	0.004
(standardized, w/i 300min)	(0.046)	(0.047)	(0.040)	(0.052)	(0.015)	(0.005)	(0.002)	(0.003)
Transmigrant count (IDW)	✓	✓	✓	✓	✓	✓	✓	✓
Baseline Pop 1980	✓	✓	✓	✓	✓	✓	✓	✓
Urban Pop % 1980	✓	✓	✓	✓	✓	✓	✓	✓
Baseline Emp 1980	✓	✓	✓	✓	✓	✓	✓	✓
Ln Serv Emp 1980		✓	✓	✓	✓	✓	✓	✓
Agro-Clim. TMV controls	✓	✓	✓	✓	✓	✓	✓	✓
Mean dep var	10.190	9.193	7.935	8.491	0.473	0.174	0.053	0.087
Std dev dep var	0.940	0.962	1.027	0.914	0.140	0.050	0.024	0.028
Observations	116	116	116	116	116	116	116	116

Notes: Outcomes in cols in 2010 across urban districts (kecatamans/adm3), aggregated to urban centers. For service shares, all shares are % of total employment across all sectors. Nearest TMV within 300min. Spatial Conley SEs (dist cutoff 200km). TMV = Transmigration Villages, IDW = inverse-distance weighted, FUA = Functional Urban Area (from GHS = Global Human Settlements), PC = Population Census. Agro-climatic TMV controls include a rich set of soil, climatic, and geogr. controls used to construct agro-similarity. *** 1%, ** 5%, * 10% significant.

These results collectively underscore the importance of considering local agricultural conditions when assessing the drivers of urban growth, and are consistent with the trade and goods-market channels emphasized in the spatial equilibrium model.

3.7 Urban In-Migration while Productive Villages Retain Transmigrants

Finally, I examine whether city growth reflects migration responses, and from where migrants originate. Urban centers exposed to higher rural productivity have higher shares of residents born outside the province and higher recent migration rates (e.g., five-year migration), indicating that much of the population growth comes from net in-migration. The composition evidence suggests that these migrants are not primarily the original transmigrants mechanically relocating out of settlement villages; rather, urban centers draw newcomers from other districts and provinces, including rural origins.

One important consideration in explaining such persistent migration rates of all types is whether such growth stems from transmigrants abandoning their placement villages and moving to nearby cities. Alternatively, productive villages could incentivize movers to stay in newly settled villages, making a good living, while rural-to-urban migrants might come from surrounding areas and not from the transmigration villages themselves. Figure 13 shows the relationship between agro-climatic similarity and the percentage of transmigrants who stayed in the Transmigration villages (TMVs) as of 2000. Overall, a positive relationship indicates that transmigrants are more likely to stay in villages with higher agro-climatic similarity to their origin areas, yet there is evidence of non-linearity (See Figure 13).

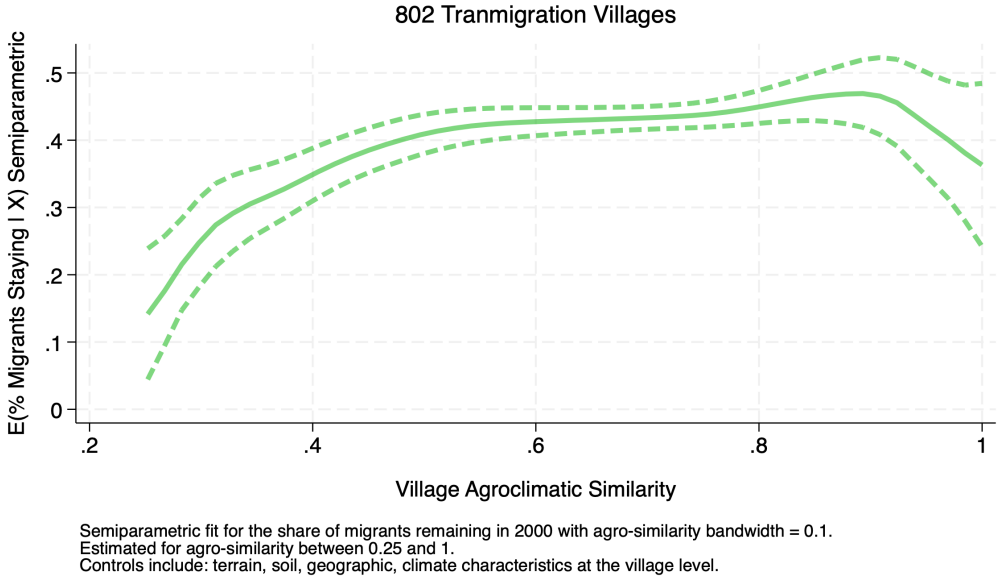
Table 7: Urban Migration Rates 2010 (aggregated to urban centers, TMV min 300min)

	Population			Transmigrants	Migrants since Birth	
	(1)	(2)	(3)	(4)	(5)	(6)
	Ln Pop	Born Same Prov: #	%	% Born Java/Bali	%: Born Diff Prov	Diff Distr
Standardized values of	0.043	75497.839***	-0.026***	0.005	0.026***	0.037***
Ag-Similarity (IDW)	(0.027)	(25843.182)	(0.008)	(0.003)	(0.008)	(0.014)
Transmigrant count (IDW)	✓	✓	✓	✓	✓	✓
Base pop 1980 (Ln)	✓	✓	✓	✓	✓	✓
Urban pop % 1980	✓	✓	✓	✓	✓	✓
Base empl 1980 (Ln)	✓	✓	✓	✓	✓	✓
Agro-Clim. TMV controls	✓	✓	✓	✓	✓	✓
Mean dep var	11.896	1.95e+05	0.842	0.069	0.158	0.261
Std dev dep var	0.839	3.44e+05	0.131	0.064	0.131	0.141
Observations	116	116	116	116	116	116

Notes: Outcomes in cols in 2010 across urban districts (kecatamans/adm3), aggregated to urban centers. Outcomes for 'other' province refer to other than current province and not Java or Bali. Rur-urb migrant share is the % of all migrants who were born in a rural district, urb-urb share is the % of all migrants who were born in an urban district (birth district status defined based on over 50% of pop in 1980 census identified as urban or rural). Nearest TMV within 300min. Spatial Conley SEs (dist cutoff 200km). TMV = Transmigration Villages, IDW = inverse-distance weighted, FUA = Functional Urban Area (from GHS = Global Human Settlements), PC = Population Census. Agro-climatic TMV controls include a rich set of soil, climatic, and geogr. controls used to construct agro-similarity. *** 1%, ** 5%, * 10% significant.

This suggests that higher agro-climatic similarity enhances the productivity and attractiveness of these villages, leading to lower out-migration rates overall, with potential evidence of some more out-migration (potentially intergenerational) in most productive villages.

Figure 13: Transmigrants More Likely to Stay in More Productive Villages



Lastly, Table 7 provides further insights by decomposing the urban population into local natives (born in the same province) and migrants (born in Java/Bali or in other provinces). Higher agro-similarity is associated with a larger absolute number and share of inter-provincial migrants, accompanied by a reduction in the share of locally born individuals (without an apparent reduction in overall number, however). In other words, urban centers surrounded by more productive rural areas grow primarily through the inflow of non-local residents (both transmigrants and those from other provinces), highlighting the spillover effects of rural productivity on urban demographic composition.

3.8 Interpretation of Mechanisms & Model Design

The empirical findings suggest several mechanisms by which a rise in rural agricultural productivity can stimulate urbanization and structural change. First, higher farm productivity increases the local food supply and can lower the price of food in the region. In a spatial equilibrium sense, cheaper or more abundant food acts as a positive shock to the real income of urban consumers and reduces the cost of living in nearby cities, as captured by the Stone-Geary framework where declining food prices raise supernumerary income and shift expenditure toward services. Second, greater agricultural output can spur demand for non-agricultural goods and services, creating employment multiplier effects in nearby towns. When transmigrant farmers enjoy higher incomes from bountiful harvests, they spend more on services, manufactured goods, and urban amenities. The especially strong growth observed in the trade and transportation subsectors of high-productivity areas supports this mechanism and is consistent with agricultural surplus generating local market activity through trade and agro-processing linkages.

Third, there may be production-side complementarities between productive agriculture and urban economic activity. A cluster of highly productive farms could support the growth of trade and transportation hubs in nearby cities (for example, wholesale markets, logistics centers, and distribution networks), as well as improved infrastructure that lowers input costs for urban firms. The model's agglomeration externality in services ($A_j^S = \bar{A}_j^S \cdot N_j^\eta$) captures this channel.

Notably, the muted population effects (F1, F2) alongside strong compositional shifts (F3–F6) suggest that the trade channel dominates the migration channel in this context. Direct migration of transmigrant farmers to cities was not the main driver since more prosperous farmers are more likely to stay on their land. This is consistent with Bryan & Morten (2019)'s observation of high moving costs in Indonesia, and the model rationalizes this through a low location-choice elasticity θ_L. The quantitative spatial equilibrium framework developed in Section 4 provides a disciplined way to decompose these channels through counterfactual exercises, including shutting down trade and shutting down migration.

4 Quantitative Spatial Equilibrium Framework

This section develops a quantitative spatial equilibrium model to (i) provide a structural interpretation of the reduced-form effects of nearby rural agricultural productivity on urban population and employment, and (ii) discipline counterfactuals that cannot be recovered from the reduced form alone. The motivating empirical facts are that urban centers exposed to more productive transmigration villages exhibit (a) muted population and employment growth, (b) employment growth concentrated in non-agricultural activities (especially trade and transport services), (c) declining agricultural employment shares, (d) rising housing prices and declining food spending shares, (e) stable non-agricultural wages, and (f) in-migration primarily from other provinces rather than from transmigration villages themselves.

The model organizes these patterns through three mechanisms. First, a *price/trade channel*: higher agricultural productivity in nearby villages lowers the delivered price of food in cities, reducing the cost of living, raising real income, and shifting consumer expenditure toward non-food items. Second, a *sectoral reallocation channel*: within each location, workers sort across agriculture, manufacturing, and services as relative sectoral returns change in response to the food price decline. Third, a *migration channel*: spatial differences in real income induce population reallocation across locations, but high migration frictions (captured by a low location-choice elasticity) generate only muted population responses.

4.1 Urban-Center-Level Empirical Facts

Table 8 presents the full set of empirical facts from the urban center-level analysis and their mapping to the model.

Table 8: Empirical facts and model coverage

#	**Fact**	**Coef (SE)**
F1	Muted city pop growth	0.043 (0.027)
F2	Muted city employment growth	0.041 (0.033)
F3	Ag employment share ↓	−0.029 (0.012)**
F4	Ln ag employment ↓	−0.097 (0.049)*
F5	Ln manuf employment ↑	0.112 (0.039)***
F6	Ln service employment ↑	0.074 (0.045)*
F8	Non-food spending ↑	0.066 (0.021)***
F9	Housing price ↑	0.146 (0.079)*
F10	Rice production ↓	−0.315 (0.136)**
F12	Non-ag wage: ns	0.038 (0.026)
F15	Food spending share ↓	Negative, sig.
F17	Village retains ag workers	Positive
F19	Migration from other provinces	See Table 7
F20	Service growth in trade & transport	See Table 6
F21	Non-ag/ag wage ratio unchanged	−0.014 (0.047)

4.2 Economic Environment

The economy consists of many locations $j = 1, \ldots, J$ and three broad sectors:

- **Agriculture** (A): uses land and labor; tradable output
- **Manufacturing** (M): uses labor (and possibly capital); tradable output
- **Services** (S): uses labor and a fixed local factor (land/infrastructure); **non-tradable** output

Some locations are primarily urban (cities), some primarily rural (villages/TMV areas), and some are "other rural" areas. Workers of total measure $\bar{L}$ choose a location j and sector s, supplying one unit of labor inelastically.

The model includes the following assumptions:

1. Agriculture uses land T_j that varies across space, capturing yield differences driven by agro-climate and transmigration productivity shocks
2. **Preferences are non-homothetic** (generalized Stone-Geary) between agricultural, manufacturing, and service goods, featuring subsistence in both food ($\bar{a} > 0$) and home-produced services ($\bar{s} > 0$)
3. Agricultural and manufacturing goods are tradable subject to iceberg costs $\tau_{kj}^{(s)}$ obeying a power law in cost-distance; services are non-tradable
4. No fixed costs of firm entry and constant-returns or decreasing-returns production $\Rightarrow$ zero economic profits; land accrues as Ricardian rents to absentee landlords
5. **Agglomeration externalities** in the service sector: service-sector TFP is increasing in local population ($A_j^S = \bar{A}_j^S \, N_j^{\eta}$, $\eta > 0$)

4.3 Preferences: Generalized Stone-Geary

Following Kongsamut, Rebello, and Xie (2001) and Buera and Kaboski (2012), preferences feature non-homotheticity in both the agricultural good and the service good:

$$U(c_A, c_M, c_S) \;=\; (c_A - \bar{a})^{\psi_A} \;\cdot\; c_M^{\psi_M} \;\cdot\; (c_S + \bar{s})^{\psi_S} \tag{5}$$

where $\bar{a} > 0$ is the subsistence food requirement (minimum food consumption), $\bar{s} > 0$ captures home production of services (at low incomes, households produce services at home; as income rises, they substitute toward market services), $\psi_A + \psi_M + \psi_S = 1$ are the expenditure weights on discretionary consumption, and $c_A \geq \bar{a}$ is required (participation constraint).

Budget constraint:

Table 9: Key notation

Symbol	Interpretation
J	number of locations (baseline calibration $J \approx 116$)
$\mathcal{S} = \{A, M, S\}$	sector set: Agriculture, Manufacturing, Services
L_{js}	labor employed in location j sector s
$N_j = \sum_s L_{js}$	total population in location j
T_j	land input in agriculture in location j
H_j	fixed local factor in services (land for housing/commercial space, infrastructure)
A_j^A, A_j^M, A_j^S	sector-specific TFP in location j
$\bar{a}$	subsistence food requirement (Stone-Geary)
$\bar{s}$	home production of services (Stone-Geary)
ψ_A, ψ_M, ψ_S	expenditure shares on discretionary consumption ($\psi_A+\psi_M+\psi_S = 1$)
β	labor share in agriculture ($0 < \beta < 1$)
γ_M	labor share in manufacturing ($0 < \gamma_M \leq 1$)
γ_S	labor share in services ($0 < \gamma_S < 1$)
η	agglomeration elasticity in services
τ_{kj}^A	iceberg cost shipping agricultural goods from k to j
δ	trade cost elasticity w.r.t. cost-distance
$\varepsilon_{js}(i)$	idiosyncratic sector preference shock for person i, drawn Fréchet with shape θ_S
$\upsilon_j(i)$	idiosyncratic location preference shock, drawn Fréchet with shape θ_L
θ_S, θ_L	Fréchet shape parameters (sector choice; location choice)
Φ_{js}	sectoral amenity in location j, sector s
P_j	ideal price index in location j
W_j	wage index / option value of living in location j
π_j, $\pi_{s\|j}$	choice probabilities (location; sector conditional on location)

$$p_A^j \, c_A + p_M \, c_M + p_S^j \, c_S = w_{js} \tag{6}$$

where $p_M = 1$ is the manufacturing price (numeraire, nationally tradable), and p_A^j, p_S^j are location-specific.

Demand functions (from constrained optimization):

$$c_A = \bar{a} + \frac{\psi_A}{p_A^j} \, \tilde{w}_{js} \tag{7}$$

$$c_M = \frac{\psi_M}{p_M} \, \tilde{w}_{js} \tag{8}$$

$$c_S = -\bar{s} + \frac{\psi_S}{p_S^j} \, \tilde{w}_{js} \tag{9}$$

I assume that in equilibrium, all workers earn sufficient income to remain at an interior solution with $c_S > 0$, i.e., $\tilde{w}_{js}$ is large enough that $\frac{\psi_S}{p_S^j}\tilde{w}_{js} > \bar{s}$ for all occupied location-sector pairs.

where income beyond subsistence needs (adjusted for home production) is:

$$\boxed{\tilde{w}_{js} \;\equiv\; w_{js} - p_A^j\,\bar{a} + p_S^j\,\bar{s}} \tag{10}$$

Note that because wages w_{js} vary across sectors, supernumerary income also varies by sector and is indexed accordingly as $\tilde{w}_{js}$.

All discretionary spending is proportional to $\tilde{w}_{js}$, allocated with shares ψ_A, ψ_M, ψ_S.

Expenditure shares:

$$s_A^j = \frac{p_A^j c_A}{w_{js}} = \frac{p_A^j \bar{a} + \psi_A\,\tilde{w}_{js}}{w_{js}} \tag{11}$$

$$s_M^j = \frac{\psi_M\,\tilde{w}_{js}}{w_{js}} \tag{12}$$

$$s_S^j = \frac{-p_S^j \bar{s} + \psi_S\,\tilde{w}_{js}}{w_{js}} \tag{13}$$

Key properties: As income rises, the food spending share s_A^j falls (food is a necessity), matching **F15**, while the service spending share s_S^j rises (services are a luxury), matching **F8**. The manufacturing share s_M^j is roughly constant for moderate income changes. This non-homotheticity — absent in standard Cobb-Douglas preferences — is essential for matching the observed expenditure patterns in response to the productivity shock.

Indirect utility:

$$V_{js} = \frac{\tilde{w}_{js}}{(p_A^j)^{\psi_A} \cdot (p_M)^{\psi_M} \cdot (p_S^j)^{\psi_S}} = \frac{\tilde{w}_{js}}{P_j} \tag{14}$$

where the price index is:

$$P_j \;=\; (p_A^j)^{\psi_A} \cdot (p_M)^{\psi_M} \cdot (p_S^j)^{\psi_S} \tag{15}$$

4.4 Production and Wages

Each worker supplies one unit of labor inelastically. Total labor $\bar{L}$ is fixed. Land T_j is immobile and used only in farming. H_j is a fixed local factor in services (housing/commercial land, infrastructure).

Technology:

- **Agriculture:**

$$Y_{jA} \;=\; (A_j^A\, T_j)^{1-\beta}\,(L_{jA})^{\beta} \qquad 0<\beta<1 \tag{16}$$

Diminishing returns to farm labor via fixed land T_j. The term A_j^A varies across space and captures the agro-climatic productivity shock from the transmigration program.

- **Manufacturing:**

$$Y_{jM} \;=\; A_j^M\,(L_{jM})^{\gamma_M} \qquad 0<\gamma_M\leq 1 \tag{17}$$

In the baseline, set $\gamma_M = 1$ (CRS). Manufacturing output is tradable at $p_M = 1$ everywhere.

- **Services:**

$$Y_{jS} \;=\; A_j^S\,(H_j)^{1-\gamma_S}\,(L_{jS})^{\gamma_S} \qquad 0<\gamma_S<1 \tag{18}$$

With **agglomeration** ($\eta > 0$):

$$A_j^S \;=\; \bar{A}_j^S \cdot {N_j}^{\eta} \tag{19}$$

As city population N_j grows, service-sector productivity rises, partially offsetting diminishing returns and helping explain why service wages do not fall despite labor inflows (**F12**).

Wages (from profit maximization under perfect competition):

For **agriculture**:

$$w_{jA}^{\star} \;=\; p_A^j\,\beta\,\frac{(A_j^A\,T_j)^{1-\beta}}{(L_{jA})^{1-\beta}} \tag{20}$$

For **manufacturing**, with CRS ($\gamma_M = 1$) and tradable output at $p_M = 1$:

$$w_{jM}^{\star} \;=\; A_j^M \tag{21}$$

Manufacturing wages are pinned by local TFP. This makes manufacturing employment the residual absorber.

For **services**:

$$w_{jS}^{\star} \;=\; p_S^j\,\gamma_S\,\bar{A}_j^S\,{N_j}^{\eta}\,H_j^{1-\gamma_S}\,(L_{jS})^{\gamma_S-1} \tag{22}$$

4.5 Trade and Prices

Agricultural good: tradable across locations with iceberg costs. Let $\tau_{kj}^{A} \geq 1$ denote the cost of shipping agriculture from k to j. The local food price is determined by the cheapest supplier:

$$p_A^j \;=\; \min_k \{\tau_{kj}^{A} \cdot p_A^k\} \tag{23}$$

In calibration: $\tau_{kj}^{A} = d_{kj}^{\delta}$ where d_{kj} is the cost-distance travel time and $\delta > 0$ is the trade cost elasticity.

Manufacturing good: tradable at zero (or low) cost $\Rightarrow p_M = 1$ everywhere (numeraire).

Service good: non-tradable $\Rightarrow p_S^j$ determined by local market clearing.

4.6 Location and Sector Choice

Following the quantitative spatial economics literature (Allen and Arkolakis 2014; Redding 2016; Tombe and Zhu 2019), workers make a nested location-sector choice with idiosyncratic taste shocks.

Stage 2 — Sector choice within location j:

Worker i in location j draws sector-specific taste shock $\varepsilon_{js}(i)$ from Fréchet with shape $\theta_S > 0$. The probability that a worker in j chooses sector s is:

$$\pi_{s|j} \;=\; \frac{\left(\Phi_{js} \cdot w_{js}\right)^{\theta_S}}{\sum_{s' \in \{A,M,S\}} \left(\Phi_{js'} \cdot w_{js'}\right)^{\theta_S}} \tag{24}$$

where Φ_{js} is the amenity/attractiveness of sector s in location j.

The sector-level inclusive value (option value of living in j):

$$W_j \;=\; \left[\sum_{s' \in \{A,M,S\}} \left(\Phi_{js'} \cdot w_{js'}\right)^{\theta_S}\right]^{1/\theta_S} \tag{25}$$

Note on supernumerary income and sector choice. Under the generalized Stone-Geary preferences, indirect utility is proportional to supernumerary income $\tilde{w}_{js}$ rather than nominal wage w_{js} (Equation 14). The theoretically consistent formulation would replace w_{js} with $\tilde{w}_{js}$ in the Fréchet sector-choice probabilities above. In the calibration, I use the nominal wage specification as a tractable approximation, which is standard in quantitative spatial models with non-homothetic preferences where the supernumerary adjustment is small relative to total income for most workers in equilibrium.

Stage 1 — Location choice:

Worker i draws location taste shock $\upsilon_j(i)$ from Fréchet with shape $\theta_L > 0$. The probability of choosing location j:

$$\pi_j = \frac{(W_j/P_j)^{\theta_L}}{\sum_{k=1}^{J}(W_k/P_k)^{\theta_L}} \tag{26}$$

where P_j is the price index from (15).

Population and sectoral labor:

$$N_j = \bar{L} \cdot \pi_j \tag{27}$$

$$L_{js} = N_j \cdot \pi_{s|j} \tag{28}$$

A low value of θ_L yields muted population responses to real income differences, consistent with **F1** and **F2**. A higher θ_S implies workers are more responsive to wage differences across sectors, facilitating within-location structural transformation.

4.7 Equilibrium

A **spatial equilibrium** is a set of prices $\{p_A^j, p_S^j\}_{j=1}^J$, wages $\{w_{jA}, w_{jM}, w_{jS}\}_{j=1}^J$, labor allocations $\{L_{jA}, L_{jM}, L_{jS}\}_{j=1}^J$, and trade flows such that: (1) workers maximize expected utility over location-sector bundles; (2) firms maximize profits; (3) labor and goods markets clear; (4) trade is consistent (no-arbitrage in agricultural prices); and (5) no worker can improve expected utility by changing location or sector.

Key equilibrium conditions:

The log ratio of agricultural to service employment in location j:

$$\ln \frac{L_{jA}}{L_{jS}} = \theta_S \left[\ln \frac{\Phi_{jA}}{\Phi_{jS}} + \ln \frac{w_{jA}}{w_{jS}}\right] \tag{E1}$$

Similarly for manufacturing to service employment:

$$\ln \frac{L_{jM}}{L_{jS}} = \theta_S \left[\ln \frac{\Phi_{jM}}{\Phi_{jS}} + \ln \frac{w_{jM}}{w_{jS}}\right] \tag{E2}$$

Population distribution across locations (relative to a reference location r):

$$\ln \frac{N_j}{N_r} = \theta_L \left[\ln \frac{W_j}{P_j} - \ln \frac{W_r}{P_r}\right] \tag{E3}$$

Service market clearing and service price:

$$\boxed{p_S^j \;=\; \frac{N_j\,\psi_S\,\tilde{w}_j}{A_j^S\,H_j^{1-\gamma_S}\,(L_{jS})^{\gamma_S} \;+\; N_j\,\bar{s}}} \tag{E4}$$

Manufacturing employment is the residual:

$$L_{jM} = N_j - L_{jA} - L_{jS} \tag{29}$$

Key comparative static: An increase in A_v^A (village agricultural TFP) $\longrightarrow$ through trade, lowers p_A^j for nearby cities $\longrightarrow$ the food price term in (E1) falls $\longrightarrow$ the log ratio L_{jA}/L_{jS} falls $\longrightarrow$ **agricultural employment share in the city declines** (matching **F3, F4**). Simultaneously, lower food prices raise supernumerary income $\tilde{w}_j$, increasing demand for services $\longrightarrow$ p_S^j rises (matching **F9**), service employment expands (matching **F6**), and the food spending share declines (matching **F15**).

Table 10: Equilibrium conditions summary

#	**Condition**	**Determines**
E1	Sectoral log-odds (A vs S)	L_{jA}/L_{jS}
E2	Sectoral log-odds (M vs S)	L_{jM}/L_{jS}
E3	Location population shares	N_j
E4	Service market clearing	p_S^j
(20)	Ag wage = MPL	w_{jA}
(21)	Manuf wage = TFP	w_{jM}
(22)	Service wage = MPL	w_{jS}
(23)	No-arbitrage in ag prices	p_A^j
(29)	$L_{jA} + L_{jM} + L_{jS} = N_j$	L_{jM} (residual)

4.8 Estimating Equations

The equilibrium conditions yield reduced-form estimating equations that map directly to the main empirical specifications. Each equation corresponds to a structural relationship in the model.

Sectoral employment share. From (E1) and the equilibrium wages, the log ratio of agricultural to non-agricultural employment depends on relative prices and productivities, motivating:

$$\underbrace{\text{empl_ag_pct}_j}_{\approx L_{jA}/N_j} = \beta_0 + \beta_1 \underbrace{\text{AgSim}_j^{IDW}}_{\propto A_j^{A,\text{nearby}}} + \beta_2\,\text{TM_count}_j + \beta_3\,X_j + \varepsilon_j \tag{30}$$

The model predicts $\beta_1 < 0$: higher AgSim for nearby villages $\rightarrow$ lower p_A^j (through trade) $\rightarrow$ lower w_{jA}/w_{jS} $\rightarrow$ lower L_{jA}/L_{jS} (matching **F3**).

Population. From (E3):

$$\ln \text{pop}_j = \beta_0 + \beta_1 \, \text{AgSim}_j^{IDW} + \beta_2 \, \text{TM_count}_j + \beta_3 \, X_j + \varepsilon_j \tag{31}$$

The coefficient β_1 captures how AgSim affects the city's real wage index W_j/P_j. The model predicts this is positive but small if θ_L is low (high migration frictions), matching **F1**.

Service price. From (E4), taking logs:

$$\ln \text{housing_price}_j = \beta_0 + \beta_1 \, \text{AgSim}_j + \ldots$$

predicts $\beta_1 > 0$ (matching **F9**).

Service wage and agglomeration. From (22), the effect of AgSim on w_{jS} is ambiguous: p_S^j rises (pushing wage up), L_{jS} rises with $\gamma_S < 1$ (pushing wage down via diminishing returns), and N_j rises with $\eta > 0$ (pushing wage up via agglomeration). If these roughly cancel, the non-agricultural wage effect is small and insignificant, matching **F12**.

Food expenditure share. From (11), a decline in p_A^j and/or a rise in w_j reduces the food spending share s_A^j, matching **F15**. The magnitude depends on $\bar{a}$ relative to w_j.

4.9 Agglomeration: Why Service Wages Do Not Fall

Without agglomeration ($\eta = 0$), an increase in L_{jS} would lower w_{jS} (diminishing returns). This would predict that the agricultural productivity shock leads to more service workers and lower service wages — contradicting **F12**.

With agglomeration ($\eta > 0$), the inflow of workers raises TFP:

$$\frac{\partial \ln w_{jS}}{\partial \ln N_j} = \underbrace{\eta}_{>0,\ \text{agglom.}} + \underbrace{(\gamma_S - 1)}_{<0,\ \text{dim. ret.}} \cdot \underbrace{\frac{\partial \ln L_{jS}}{\partial \ln N_j}}_{\approx 1} + \underbrace{\frac{\partial \ln p_S^j}{\partial \ln N_j}}_{>0} \tag{32}$$

The first term ($\eta > 0$) and third term (rising prices) push wages up, while the second term ($\gamma_S - 1 < 0$) pushes wages down. For moderate η, these approximately cancel, yielding a stable nominal service wage consistent with **F12**.

4.10 Counterfactual Exercises

The calibrated model enables three counterfactual exercises that decompose the channels underlying the reduced-form results.

Counterfactual 1: Shut down trade (autarky). Set $\tau_{kj}^A \to \infty$ for all $k \neq j$. Each location must be food self-sufficient. Without trade, cities cannot import food and must maintain a large agricultural sector, resulting in less structural transformation. The difference from the baseline equilibrium measures the **trade channel's contribution** to urban structural change.

Counterfactual 2: Shut down migration. Fix $N_j = N_j^{\text{baseline}}$ for all j while allowing trade. All structural transformation comes from the price/trade channel plus within-city labor reallocation. The difference from the full equilibrium measures the **migration channel's contribution**.

Counterfactual 3: Alternative TMV targeting. Reassign A_v^A across locations (e.g., all high-productivity villages near the largest cities, or uniform AgSim everywhere). This addresses policy questions about optimal spatial targeting of agricultural extension programs.

5 Conclusion

This paper studies how rural agricultural productivity shapes nearby urban development in a setting where rural shocks are large, spatially dispersed, and plausibly exogenous. I leverage Indonesia's Transmigration Program and the quasi-random variation in migrants' agro-climatic skill match, previously captured by the agro-climatic similarity measure (proxy for agricultural productivity) across newly settled villages (Bazzi et al. 2016). I then construct a city-level exposure to surrounding rural productivity by aggregating village-level agro-similarity using travel-time-based connectivity weights across 180 urban centers on the Outer Islands, while separately controlling for the scale of nearby settlement.

The reduced-form evidence establishes four key facts. First, cities exposed to more productive transmigration villages show **muted overall population and employment growth**: the effects are positive but statistically insignificant, indicating that the productivity shock does not trigger a dramatic urban population explosion. Second, these cities experience pronounced **structural transformation**: employment shifts away from peri-urban agriculture toward manufacturing and services, with service employment shares rising significantly and growth concentrated in trade and transportation subsectors. Third, the results point to a **goods-market mechanism**: exposed cities reduce local agricultural production and agricultural employment while maintaining stable food consumption, and the food spending share declines relative to non-food spending, consistent with Stone-Geary preferences where households reallocate expenditure as the effective cost of food falls. At the same time, housing prices rise, consistent with increased demand for non-tradable services. Fourth, **urban growth is driven by in-migration** from other provinces and rural districts rather than by the mechanical relocation of transmigrants out of their settlement villages. Transmigrants are in fact more likely to remain in villages where their agro-climatic match is higher, suggesting that the urbanization effects operate through regional market linkages rather than a story of farmers leaving the land.

To interpret and quantify these patterns, I develop a three-sector quantitative spatial equilibrium model featuring non-homothetic Stone-Geary preferences (with subsistence food requirements and home-produced services), agglomeration externalities in the service sector, and endogenous food prices determined through inter-regional trade. The model rationalizes the full set of empirical facts through three mechanisms: a trade channel (cheaper food lowers the cost of living and stimulates demand for services), a sectoral reallocation channel (workers shift from agriculture to services as relative returns change), and a migration channel (modest, reflecting high moving costs in the Indonesian archipelagic

context). The model's nested Fréchet structure yields transparent estimating equations that map directly to the reduced-form regressions, and the agglomeration externality in services explains why non-agricultural wages remain stable despite expanding service employment.

These findings contribute to a core debate in development and urban economics about whether raising agricultural productivity accelerates or delays urbanization and structural change. In this setting, higher rural productivity appears to complement urban growth: it supports expansion of nearby towns and reallocation into non-agricultural employment through trade and market linkages, while migration frictions limit the direct population response. A key implication is that investments in rural productivity can foster structural transformation in nearby urban economies without requiring massive rural-to-urban migration, challenging the classical Harris-Todaro push narrative and suggesting that trade-based spillovers may be quantitatively more important than labor mobility in contexts with high internal moving costs.

Ongoing work will use the calibrated model to evaluate counterfactual policies — including the effects of reducing internal migration costs, shutting down inter-regional food trade, and alternative spatial targeting of transmigration settlements — to assess which policy levers can amplify (or dampen) the urban spillovers from rural productivity improvements.

References

Bazzi, S., Gaduh, A., Rothenberg, A. D., & Wong, M. (2016). Skill Transferability, Migration, and Development: Evidence from Population Resettlement in Indonesia. *American Economic Review*, 106(9), 2658–98.

Bazzi, S., Gaduh, A., Rothenberg, A. D., & Wong, M. (2019). Unity in diversity? How intergroup contact can foster nation building. *American Economic Review*, 109(11), 3978–4025.

Bryan, G., Glaeser, E., & Tsivanidis, N. (2020). Cities in the Developing World. *Annual Review of Economics*, 12, 273–297.

Bryan, G. & Morten, M. (2019). No Title. *Journal of Political Economy*, 127(5), 2229–2268.

Bustos, P., Caprettini, B., & Ponticelli, J. (2016). Agricultural productivity and structural transformation: Evidence from Brazil. *American Economic Review*, 106(6), 1320–1365.

Dingel, J. I., Miscio, A., & Davis, D. R. (2019). Cities, Lights, and Skills in Developing Economies. *Journal of Urban Economics*, (pp. 103174).

Edwards, R. B. (2019). Spillovers from agricultural processing.

Glaeser, E. L. (2014). A world of cities: The causes and consequences of urbanization in poorer countries. *Journal of the European Economic Association*, 12(5), 1154–1199.

Gollin, D., Jedwab, R., Vollrath, D., Vollrath, D., & Vollrath, D. (2016). Urbanization with and without industrialization. *Journal of Economic Growth*, 21(1), 35–70.

Gollin, D., Parente, S., & Rogerson, R. (2002). The Role of Agriculture in Development. *American Economic Review*, 92(2), 160–164.

Harris, J. R. & Todaro, M. P. (1970). Migration, Unemployment and Development: A Two-Sector Analysis. *American Economic Review*, 60(1), 126–142.

Jedwab, R. & Vollrath, D. (2015). Urbanization without growth in historical perspective. *Explorations in Economic History*, 58, 1–21.

Lagakos, D. (2020). Urban-Rural Gaps in the Developing World: Does Internal Migration Offer Opportunities? *Journal of Economic Perspectives*, 34(3), 174–192.

Lagakos, D., Marshall, S., Mobarak, A. M., Vernot, C., & Waugh, M. E. (2020). Migration costs and observational returns to migration in the developing world. *NBER Working Paper Series*, 113, 138–154.

Lagakos, D. & Waugh, M. E. (2013). Selection, agriculture, and cross-country productivity differences. *American Economic Review*, 103(2), 948–980.

Lewis, W. A. (1954). Economic Development with Unlimited Supplies of Labour. *The Manchester School*, 22(2), 139–191.

Matsuyama, K. (1992). Agricultural productivity, comparative advantage, and economic growth. *Journal of Economic Theory*, 58(2), 317–334.

Monte, F., Redding, S. J., & Rossi-Hansberg, E. (2018). Commuting, migration, and local employment elasticities.

Moreno-Monroy, A. I., Schiavina, M., & Veneri, P. (2021). Metropolitan areas in the world. Delineation and population trends. *Journal of Urban Economics*, 125.

Morten, M. & Oliveira, J. (2016). Paving the Way to Development: Costly Migration and Labor Market Integration. *NBER Working Paper*.

Munshi, K. & Rosenzweig, M. (2016). Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap. *American Economic Review*, 106(1), 46–98.

Nunn, N. & Qian, N. (2011). The Potato's Contribution to Population and Urbanization: Evidence From A Historical Experiment. *The Quarterly Journal of Economics*, 126(2), 593–650.

Olken, B. A. (2009). Merged Indonesian Kecamatan Codes. *National Bureau of Economic Research Electronic Dataset*.

Restuccia, D., Yang, D. T., & Zhu, X. (2008). Agriculture and aggregate productivity: A quantitative cross-country analysis. *Journal of Monetary Economics*, 55(2), 234–250.

Vernon Henderson, J. & Turner, M. A. (2020). Urbanization in the Developing World: Too Early or Too Slow? *Journal of Economic Perspectives*, 34(3), 150–73.

Weiss, D. J., Nelson, A., Gibson, H. S., Temperley, W., Peedell, S., Lieber, A., Hancher, M., Poyart, E., Belchior, S., Fullman, N., Mappin, B., Dalrymple, U., Rozier, J., Lucas, T. C., Howes, R. E., Tusting, L. S., Kang, S. Y., Cameron, E., Bisanzio, D., Battle, K. E., Bhatt, S., & Gething, P. W. (2018). A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, 553(7688), 333–336.

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

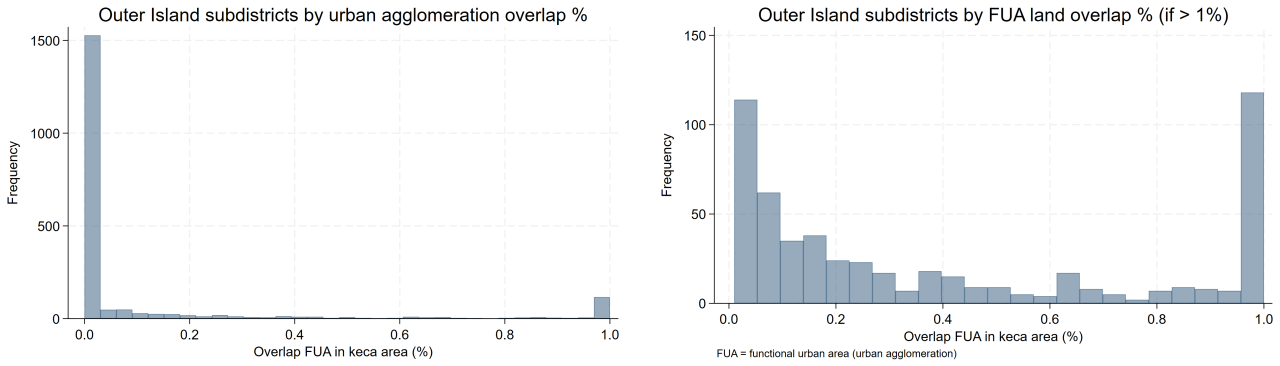
A Appendix

A.1 Data Sources

Outcomes of Interest	Geography	Available Data Details
Population, Employment, and Migration from the Population Census (1980, 2010):		
Sectoral employment	Cities + Villages	Individual employment in agriculture, manufacturing, services
Population	Cities + Villages	Population level/growth; total employment level/growth
Migration (by flow and origin)	Cities + Villages	Migration rates (by birth region and region 5 years ago)
Prices, Consumption, & Wages from Susenas HH Survey (2002, 2011):		
Prices (ag vs. non-ag)	Cities	HH Consumption module: Unit values by food item, crop/food price indices, housing price index (*proxy for non-tradables' price*)
Consumption expenditures	Cities	Consumption by food (quantity & value) and non-food item (value), expenditure shares (food and non-food)
Wages (ag vs. non-ag/services)	Cities	Median wage by worker employment sector (real wages calculated by dividing by the corresponding price index)
Agricultural Production from PODES Village Survey (2003):		
Agricultural production	Cities + Villages	Production by crop/item, yield per land area

A.2 Urban Sampling

Figure A1: Outer Island Subdistrict by FUA land area overlap



(a) All Outer Island subdistrict

(b) Outer Island subdistricts with >1% FUA land overlap

Notes:

Scatterplot:

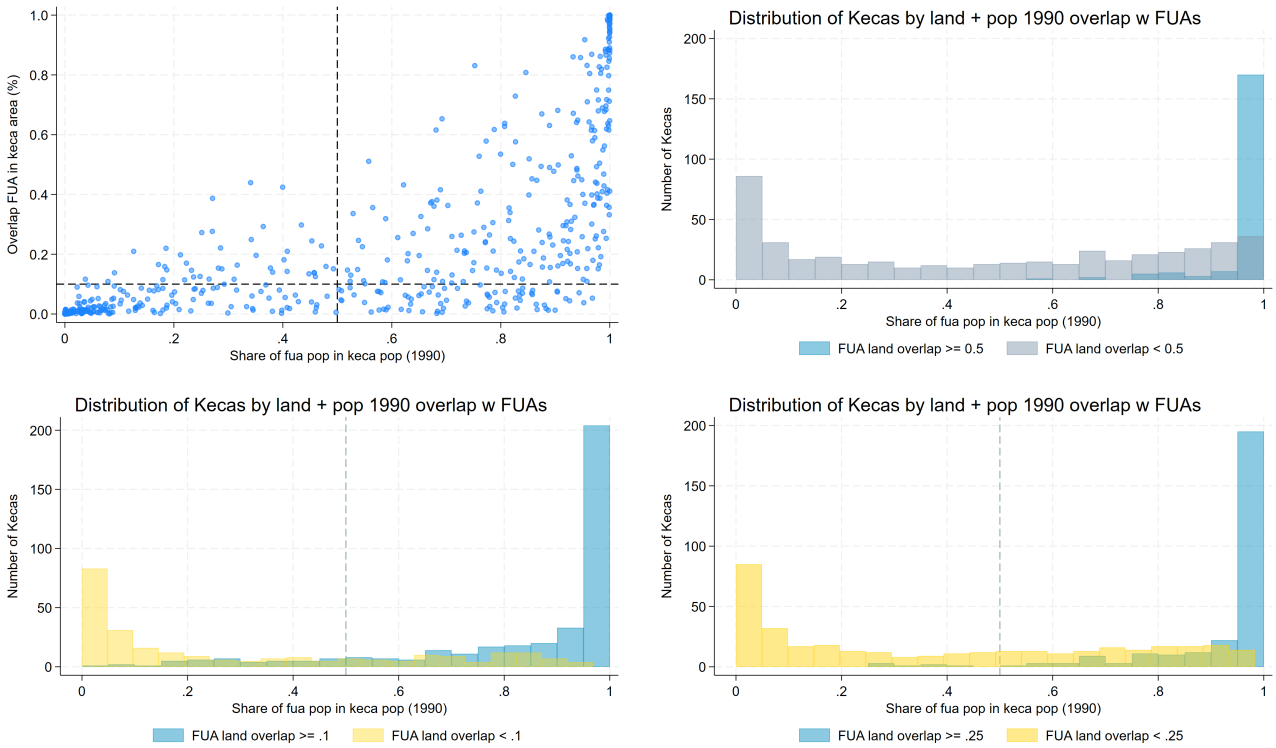
Overall pattern makes sense, but no natural cutoff Land cutoff less meaningful (high pop overlap in some even when land overlap < 0.1)

FUA Area vs. Pop overlap

Cutoff: (1) Take those where pop overlap in 1990 > 0.5 (most people in FUA) (2) Does it make sense to include those with lower pop overlap (and lower land)?

Next: Check the keca counts by both land and area Pop overlap > 0.5 would result in almost all large land overlap kecas + some of low-land-overlap kecas Robustness: can take pop overlap > 0.25 Need to get the distance for all kecas w FUA pop overlap > 0.2?

Figure A2: Title



Notes:

A.2.1 FUA-Keca pop and land shares

Finding share of FUA urban pop in keca pop (using 1990 as GHS baseline)

Overlap with urban desas (admin)

- 205 districts with perfect overlap with urban desas (=1) (administratively urban)

- Perfectly overlapping, so no need for separate pop calculations

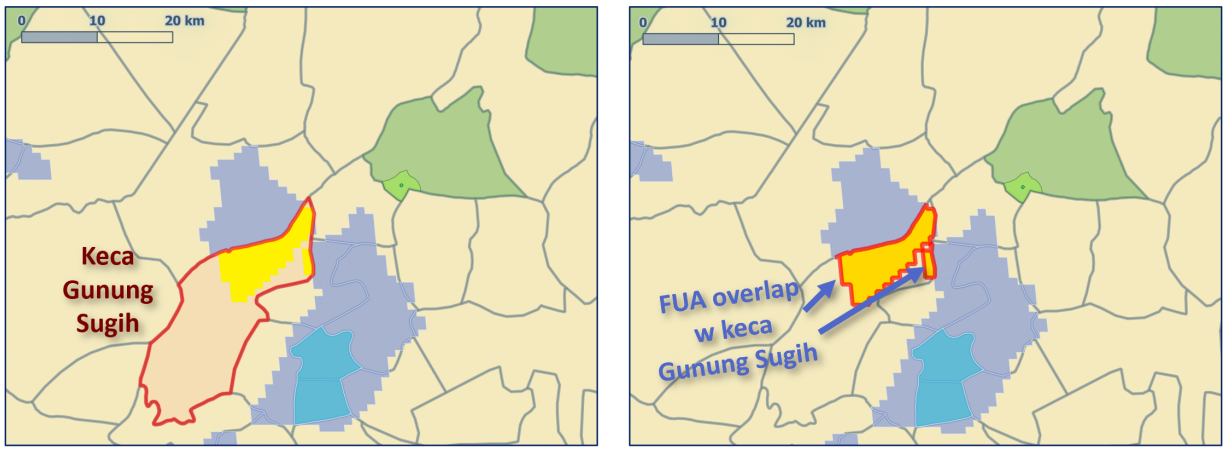
Overlap with FUAs (GHS)

Keca Gunung Sugih (1805040)

- Keca overlaps with parts of 2 different FUAs

- Perfectly overlapping, so no need for separate pop calculations

Figure A3: FUA-Keca Overlap Sample



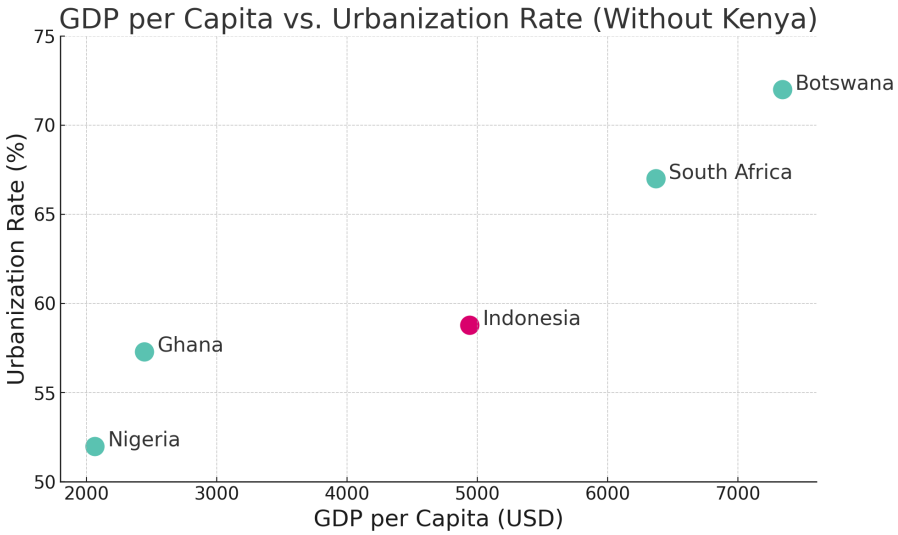
- Aggregate pop overlapping w FUAs within each keca

(sum) count	Freq.	Percent
1	1,979	92.00
2	170	7.90
3	2	0.09
Total	2,151	100.00

Table A1: Frequency table for `fua_polyg_count`

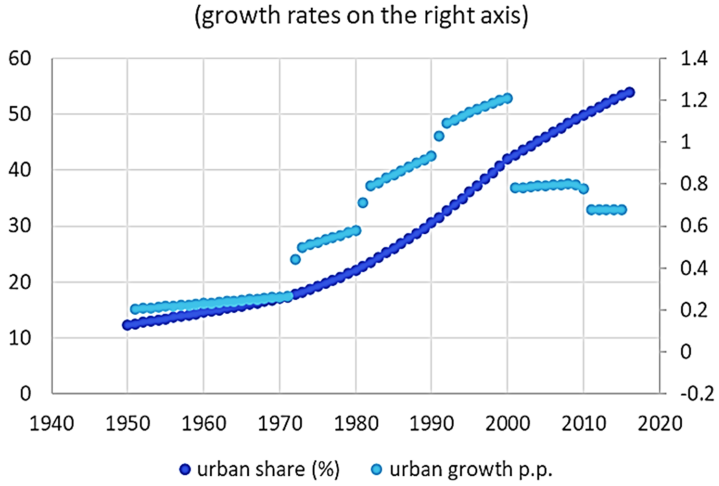
A.3 Urban Descriptives

Figure A4: GDP Per Capita Vs. Urbanization in Indonesia & Select Countries in Sub-Saharan Africa



Source: World Bank Development Indicators (2023).

Figure A5: Population Trend in Transmigration Village Settlement by Year

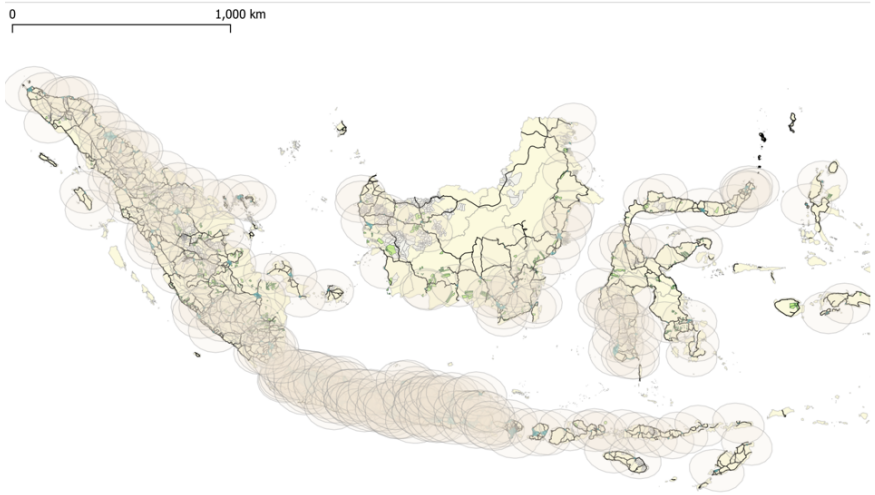


Source: Our World in Data.

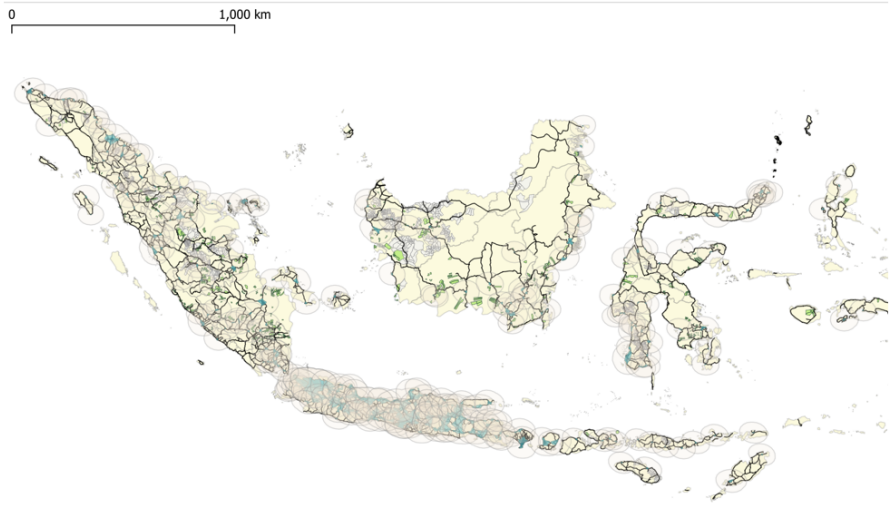
A.4 Identification & Empirical Design

Figure A6: Buffer Comparison

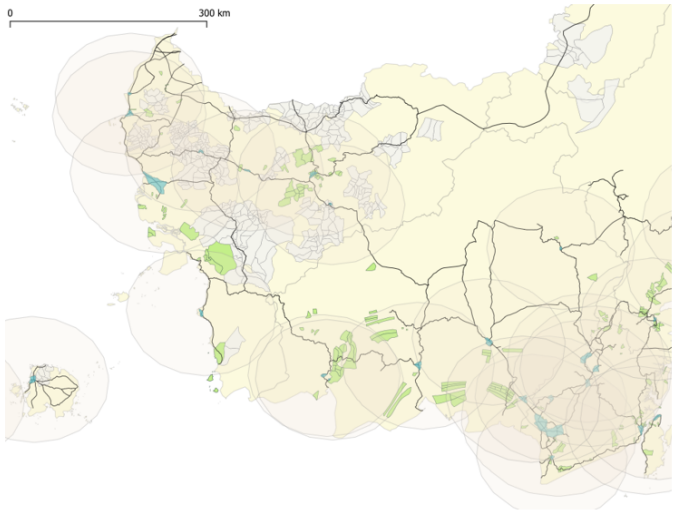
Panel A. 100 km – Buffers too wide



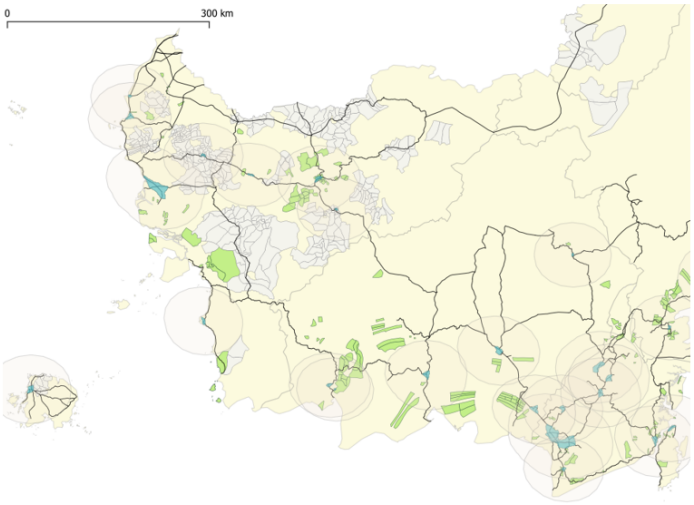
Panel B. 50 km – Buffers are more meaningful



Panel C. Select Zoomed-In Area with 100km Buffer (NE Kalimantan)

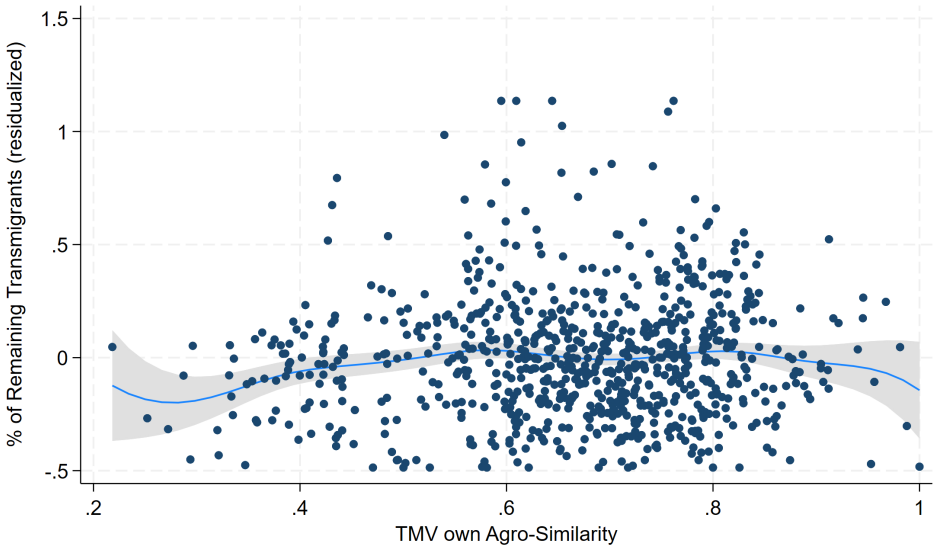


Panel D. Select Zoomed-In Area with 50km Buffer (NE Kalimantan)



A.5 Village-Level Analysis

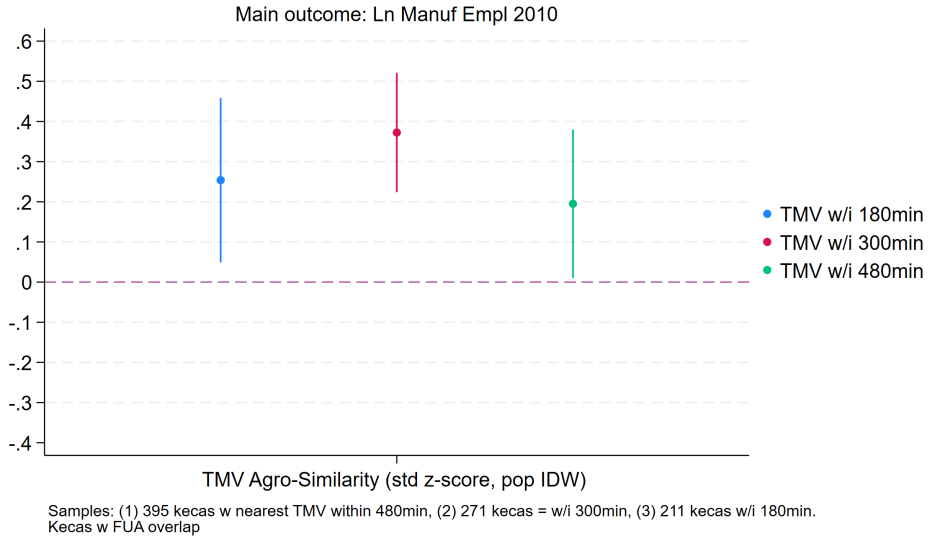
Figure A7: Non-Linearity in Out-Migration & Village Productivity

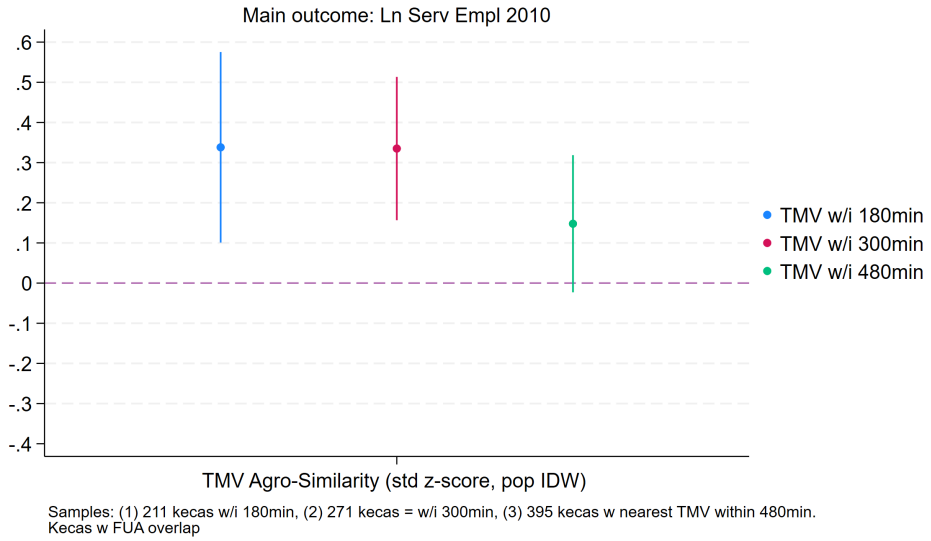


Notes: a scatter plot of residuals from the regression of the share of remaining transmigrants on agro-climatic controls and the number of initially placed individuals, plotted against village-level agro-similarity measure, across 803 villages excluding several outliers in residuals.

A.6 Robustness Results

Figure A8: Robustness: Effect by Industry by TMV Proximity





B Appendix: Detailed Derivations

[Note: These detailed derivations are presented for a slighty different, simpler 2-sector model, the updating to the extended three-sector Stone-Geary model in the main text is currently underway.

B.1 Deriving HH Utility Maximization

Each individual lives in one location $j \in \{1, \ldots, J\}$ and (after arrival) chooses a sector $s \in \{A, N\}$ to maximize their utility. Preferences over the two consumption goods are Cobb–Douglas with shares ψ and $1 - \psi$ and include a **location–sector amenity** B_{js}. Idiosyncratic tastes are i.i.d. drawn from a Frechét distribution.

$$U_{js}(c_A, c_N, \varepsilon_{js}, \upsilon_j) \;=\; \left[c_A^{\psi}\, c_N^{1-\psi}\right] B_{js}\, e^{\varepsilon_{js}}\, e^{\upsilon_j} \qquad (0 < \psi < 1) \tag{33}$$

where c_A and c_N are consumption quantities of the agricultural (A) and non-agricultural (N) goods, B_{js} captures location-sector-specific amenities (climate, networks, etc.), ε_{js} is the *sector preference shock* (idiosyncratic comparative advantage for sector s in location j), assumed $\varepsilon_{js} \overset{\text{i.i.d.}}{\sim} Gumbel(0, 1/\theta_S)$; υ_j is the *location preference shock* (idiosyncratic attachment to j), $\upsilon_j \overset{\text{i.i.d.}}{\sim} Gumbel(0, 1/\theta_L)$ (the two shocks enter multiplicatively as exponentials).

Step 1: take logs of the utility function

Taking natural logs of (33) yields:

$$\ln U_{js} \;=\; \psi \ln c_A \;+\; (1-\psi)\ln c_N \;+\; \ln B_{js} \;+\; \varepsilon_{js} \;+\; \upsilon_j \tag{34}$$

Important note: $\ln B_{js}$ is a *constant* with respect to the consumption choice (because B_{js} is predetermined), but it *does* enter the location-sector decision later on.

Step 2: set up and solve the consumption maximization

$$\max_{c_A, c_N} \; \psi \ln c_A + (1-\psi)\ln c_N + \ln B_{js} \quad \text{s.t.} \quad p_{j,A}c_A + p_{j,N}c_N = w_{js}$$

Because $\ln B_{js}$ is constant, the FOCs are *identical* to the standard Cobb-Douglas problem.

Define the Lagrangian:

$$\mathcal{L} = \psi \ln c_A + (1-\psi)\ln c_N + \lambda\big[w_{js} - p_{j,A}c_A - p_{j,N}c_N\big]$$

First-order conditions:

$$\begin{aligned} \partial_{c_A}: \; \frac{\psi}{c_A} - \lambda p_{j,A} = 0 \qquad &\Longrightarrow \quad c_A = \frac{\psi w_{js}}{p_{j,A}} \\ \partial_{c_N}: \; \frac{1-\psi}{c_N} - \lambda p_{j,N} = 0 \qquad &\Longrightarrow \quad c_N = \frac{(1-\psi) w_{js}}{p_{j,N}} \end{aligned}$$

We get the above from the following derivations:

$$\partial_{c_A}\mathcal{L}: \quad \frac{\psi}{c_A} - \lambda\, p_{j,A} = 0 \qquad \Longrightarrow \quad \lambda = \frac{\psi}{p_{j,A}\, c_A} \tag{35}$$

$$\partial_{c_N}\mathcal{L}: \quad \frac{1-\psi}{c_N} - \lambda\, p_{j,N} = 0 \qquad \Longrightarrow \quad \lambda = \frac{1-\psi}{p_{j,N}\, c_N} \tag{36}$$

Equating the RHS of the two expressions for λ, we get:

$$\frac{\psi}{p_{j,A}\, c_A} = \frac{1-\psi}{p_{j,N}\, c_N}$$

Rearranging gives:

$$\frac{c_N}{c_A} = \frac{1-\psi}{\psi}\,\frac{p_{j,A}}{p_{j,N}}$$

which is the Cobb–Douglas **marginal-rate-of-substitution** condition.

Step 3: derive consumption for each good

Use the budget constraint $p_{j,A}c_A + p_{j,N}c_N = w_{js}$ to solve for c_A and c_N, substitute this ratio into the budget constraint:

$$\text{Budget:} \qquad p_{j,A}\, c_A \;+\; p_{j,N}\, c_N = w_{js} \tag{37}$$

$$\text{Insert } c_N = c_A\,\frac{1-\psi}{\psi}\,\frac{p_{j,A}}{p_{j,N}} \quad \text{(from the MRS)} \tag{38}$$

$$p_{j,A}\, c_A \;+\; p_{j,N}\left[c_A\,\frac{1-\psi}{\psi}\,\frac{p_{j,A}}{p_{j,N}}\right] = w_{js} \tag{39}$$

$$\left[p_{j,A} + \frac{1-\psi}{\psi}\,p_{j,A}\right] c_A = w_{js} \tag{40}$$

$$p_{j,A}\left[\,\psi^{-1}\right]\, c_A = w_{js} \tag{41}$$

$$c_A^{\star} = \;\psi\,\frac{w_{js}}{p_{j,A}} \tag{42}$$

Plug in $c_A^{\star}$ into the equation for c_N to get $c_N^{\star}$:

$$c_N^{\star} = (1-\psi)\frac{w_{js}}{p_{j,N}}$$

Step 4: plug back to obtain indirect utility

Insert $c_A^\star, c_N^\star$ into (**??**):

$$\begin{aligned} u_{js}^\star &= \psi \ln\Big[\psi \tfrac{w_{js}}{p_{j,A}}\Big] + (1-\psi)\ln\Big[(1-\psi)\tfrac{w_{js}}{p_{j,N}}\Big] + \ln B_{js} + \varepsilon_{js} \\ &= \big[\psi + (1-\psi)\big]\ln w_{js} \;-\; \psi \ln p_{j,A} \;-\; (1-\psi)\ln p_{j,N} + \ln B_{js} \\ &\quad + \underbrace{\big[\psi \ln \psi + (1-\psi)\ln(1-\psi)\big]}_{\text{constant}} + \varepsilon_{js} \\ &= \ln w_{js} \;-\; \psi \ln p_{j,A} \;-\; (1-\psi)\ln p_{j,N} + \ln B_{js} + \gamma + \varepsilon_{js} \end{aligned}$$

We can drop the constant γ since utility is an ordinal function, so the **indirect utility** used in the migration–sector choice:

$$V_{js} \;=\; \ln w_{js} \;-\; \psi \ln p_{j,A} \;-\; (1-\psi)\ln p_{j,N} \;+\; \ln B_{js} \tag{43}$$

The taste shock ε_{js} will be added later when computing choice probabilities. We can interpret each term in indirect utility (43):

$$\underbrace{\ln w_{js}}_{\text{nominal income}} \quad - \quad \underbrace{\big[\psi \ln p_{j,A} + (1-\psi)\ln p_{j,N}\big]}_{\text{cost-of-living (CES price index)}} \quad + \quad \underbrace{\ln B_{js}}_{\text{location–sector amenity}}$$

Hence:

- Higher nominal wage $\uparrow \Rightarrow$ higher utility
- Higher food price $\uparrow$ or service price $\uparrow \Rightarrow$ lower utility, weighted by budget shares
- A bundle-specific amenity shifts utility *additively* – it does *not* affect the optimal consumption rule but does tilt migration/sector choice probabilities

Note: Can't set prices $p_{j,N}$ as a numeraire

Since $p_{j,N}$ differs across locations (e.g., the good is non-tradable services or has iceberg trade costs), I cannot set $p_{j,N} = 1$ everywhere. Instead, I will leave $p_{j,N}$ explicit as an unknown object to be estimated. Alternatively, I can pick one benchmark location $\bar{\jmath}$ and normalize $p_{\bar{\jmath},N} = 1$, expressing other prices relative to that location. I will estimate these with a proxy for these local service prices (e.g., CPI or cost-of-living indices).

B.2 Deriving the Conditional Sector Probability $\pi_{s|j}$

We show step by step how to derive why $\pi_{s|j}$, the probability that a worker already located in j chooses sector $s \in \{A, N\}$.

[leftmargin=*,label=**Step 2:**,itemsep=1.3em]

1. **Write utility in logarithmic form**

   From (33), after substituting the optimal consumption bundle (**??**), total utility is

$$U_{js} = B_{js}\frac{w_{js}}{P_j}\, e^{\varepsilon_{js}}$$

   with $\varepsilon_{js} \overset{\text{i.i.d.}}{\sim} \text{Gumbel}(0, 1/\theta_S)$. Taking natural logs (which preserves ordering across alternatives) gives

$$\ln U_{js} = \ln\Big(B_{js}\frac{w_{js}}{P_j}\Big) + \varepsilon_{js}$$

2. **Apply the Extreme-Value logit formula**

   For any set of alternatives whose deterministic utilities are $\delta_{js} = \ln\big(B_{js}w_{js}/P_j\big)$ and whose shocks are i.i.d. Gumbel with scale $1/\theta_S$, the choice probability for alternative s equals

$$\Pr\big[s \text{ chosen} \mid j\big] = \frac{\exp\big(\theta_S\,\delta_{js}\big)}{\sum_{s'\in\{A,N\}} \exp\big(\theta_S\,\delta_{js'}\big)}$$

3. **Replace δ_{js} with observable terms**

   Insert $\delta_{js} = \ln\Big(B_{js}\frac{w_{js}}{P_j}\Big)$ and exponentiate:

$$\exp\big(\theta_S\,\delta_{js}\big) = \Big(B_{js}\frac{w_{js}}{P_j}\Big)^{\theta_S}$$

   Doing the same for each s' in the denominator yields exactly (44), i.e.

$$\pi_{s|j} = \frac{\Big(B_{js}\frac{w_{js}}{P_j}\Big)^{\theta_S}}{\sum_{s'\in\{A,N\}} \Big(B_{js'}\frac{w_{js'}}{P_j}\Big)^{\theta_S}}$$

B.3 Derive Wage Index from Location-Sector Choice

Below each step explains *why* we perform a transformation and explicitly references the equations introduced in **??**. All Shock distributions are Extreme-Value Type I, which guarantees the closed-form "logit" probabilities.

[leftmargin=*,label=**Step 3:**,itemsep=1.4em]

1. **Take the conditional sector probability $\pi_{s|j}$**

Given location j, a worker compares only the two sector utilities U_{jA} and U_{jN} in (**??**). Because the shock scale is $1/\theta_S$, the Extreme-Value property implies

$$\pi_{s|j} = \frac{\left(B_{js}\frac{w_{js}}{P_j}\right)^{\theta_S}}{\left(B_{jA}\frac{w_{jA}}{P_j}\right)^{\theta_S} + \left(B_{jN}\frac{w_{jN}}{P_j}\right)^{\theta_S}} \tag{44}$$

Rewriting the denominator as the sum over the two sectors yields:

$$\pi_{s|j} = \frac{\left(B_{js}\frac{w_{js}}{P_j}\right)^{\theta_S}}{\sum_{s' \in \{A,N\}} \left(B_{js'}\frac{w_{js'}}{P_j}\right)^{\theta_S}} \tag{45}$$

Based on the derivations from subsection B.2.

2. **Construct the sector-inclusive value W_j**

The sector-inclusive value W_j summarizes, for each location j, the *expected maximum utility* a worker can obtain after drawing sector-specific shocks and choosing the better sector. Note: it is not exactly a wage, but rather a measure of indirect utility. We can think of it as a wage index capturing how attractive location j is once a worker can freely pick whichever sector pays off best after their idiosyncratic shocks are realized. Below I derive this indirect-utility wage index W_j, using the conditional sector probability derived above in (44).

(a) **Log utilities**

Write the two sector utilities in log form from (44):

$$\ln U_{jA} = \ln\!\left(B_{jA}\frac{w_{jA}}{P_j}\right) + \varepsilon_{jA}, \qquad \ln U_{jN} = \ln\!\left(B_{jN}\frac{w_{jN}}{P_j}\right) + \varepsilon_{jN} \tag{46}$$

where $\varepsilon_{jA}, \varepsilon_{jN} \overset{\text{i.i.d.}}{\sim} \text{Gumbel}(0, 1/\theta_S)$.

(b) **Expected log of the maximum**

Each individual will chose the sector that gives him/her a higher utility, choosing the max between U_{jA} and U_{jN}. Thus, for 2 i.i.d. Gumbel shocks with scale $1/\theta_S$, an individual expected utility is:

$$\mathbb{E}\Big[\max\{\ln U_{jA}, \ln U_{jN}\}\Big] = \frac{1}{\theta_S}\ln\Big[\left(B_{jA}\tfrac{w_{jA}}{P_j}\right)^{\theta_S} + \left(B_{jN}\tfrac{w_{jN}}{P_j}\right)^{\theta_S}\Big] \tag{47}$$

The CDF of the maximum of independent Gumbel draws is the product of individual CDFs. Taking the expectation of this maximum yields the standard log-sum-exp closed form.[7] The dispersion parameter $1/\theta_S$ scales the resulting expression.

(c) **Exponentiate to return to levels**

Undo the log-form by exponentiating back to level form ($exp(a\,ln\,b) = e^{a\,ln\,b} = (e^{ln\,b})^a = b^a$):

$$\begin{aligned} W_j &= \exp\Big\{\mathbb{E}\left[\max(\ln U_{jA}, \ln U_{jN})\right]\Big\} \\ &= \left[\left(B_{jA}\tfrac{w_{jA}}{P_j}\right)^{\theta_S} + \left(B_{jN}\tfrac{w_{jN}}{P_j}\right)^{\theta_S}\right]^{1/\theta_S} \\ &= \left[\sum_{s'\in\{A,N\}} \left(B_{js'}\frac{w_{js'}}{P_j}\right)^{\theta_S}\right]^{1/\theta_S} \end{aligned} \tag{test}$$

Equation (44) normalizes by that sum so sector probabilities add to one, while the last step raises the sum above to the power $1/\theta_S$ to convert the expected *log* maximum back to a multiplicative index used in the upper-level location choice.

Workers who have moved to location j still face uncertainty: they do not know whether their sector-specific shock ε_{jA} will make agriculture more attractive than non-agriculture (or vice-versa). Before the shock is realized they care about the expected maximum utility they can attain after choosing the better of the two sectors. The value of that ex-ante expectation is the sector-inclusive value W_j. It is a value that summarizes how good it is, on average, to live in j once the individual picks their preferred sector.

Thus, W_j is a single deterministic location-specific wage index, which will enter the upper-level location choice:

$$W_j = \left[\sum_{s'\in\{A,N\}} \left(B_{js'}\frac{w_{js'}}{P_j}\right)^{\theta_S}\right]^{1/\theta_S} \tag{48}$$

B.4 Derive Location Choice π_j

[label=**Step 4:**, leftmargin=*, itemsep=1.2em]

1. **Write sector-level utility in log form and isolate the deterministic part**

[7] This result follows from the standard closed-form expectation of the maximum of independent Gumbel random variables.

We start from the indirect utility (level form) for bundle (j, s)

$$V_{js} \equiv \frac{B_{js} w_{js}}{P_j} \qquad \Longleftrightarrow \qquad \ln U_{js} = \underbrace{\ln\!\Big(\tfrac{B_{js} w_{js}}{P_j}\Big)}_{\delta_{js}} + \varepsilon_{js}$$

Intuition: The term δ_{js} collects all *deterministic* components: amenity B_{js}, nominal wage w_{js}, and the cost-of-living index P_j. The shock ε_{js} is the idiosyncratic sector preference within location j (assumed Gumbel with scale $1/\theta_S$). Writing utility in logs makes the Extreme-Value algebra tractable

2. **Take the expected maximum over sectors inside j**

A worker located in j will choose the sector that yields the higher realized utility, i.e. $\max_{s\in\{A,N\}} \ln U_{js}$. Before drawing ε_{js}, she cares about the *expected* value of that maximum. For i.i.d. Gumbel$(0, 1/\theta_S)$ shocks, the well-known log-sum-exp formula gives

$$\mathbb{E}\Big[\max_{s\in\{A,N\}} \ln U_{js}\Big] = \frac{1}{\theta_S} \ln\!\Bigg(\sum_{s\in\{A,N\}} e^{\theta_S \delta_{js}}\Bigg)$$

Rearrange:

$$\theta_S \, \mathbb{E}[\max_s \ln U_{js}] = \ln\!\left(\sum_s e^{\theta_S \delta_{js}}\right) \tag{49}$$

Exponentiate both sides:

$$\exp\!\Big(\theta_S \, \mathbb{E}[\max_s \ln U_{js}]\Big) = \sum_s e^{\theta_S \, \delta_{js}} \tag{50}$$

Raise both to the power $1/\theta_S$:

$$\exp\!\Big(\mathbb{E}[\max_s \ln U_{js}]\Big) = \left[\sum_s e^{\theta_S \delta_{js}}\right]^{1/\theta_S} \tag{51}$$

Define W_j :

$$W_j = \left[\sum_s e^{\theta_S \, \delta_{js}}\right]^{1/\theta_S} = \left[\sum_s \left(\frac{B_{js} w_{js}}{P_j}\right)^{\theta_S}\right]^{1/\theta_S} \tag{52}$$

$$W_j \equiv \left[\sum_{s \in \{A,N\}} \left(\frac{B_{js} w_{js}}{P_j} \right)^{\theta_S} \right]^{1/\theta_S}$$

Intuition: The "integration over shocks" step transforms a hard max into a smooth aggregator. The parameter θ_S controls how similar the two sectors must be for a worker to be indifferent: higher θ_S means less dispersion and a sharper max; lower θ_S means more dispersion and smoother substitution

Intuition: W_j is a single deterministic index that captures *how attractive location j is once the worker can pick her preferred sector.* It is a CES-like aggregator of "real wage $\times$ amenity" terms across the two sectors, with elasticity governed by θ_S

3. **Recover the log form if needed**

   Taking logs of W_j gives the expected deterministic utility from the *best* sector in j

$$\ln W_j = \frac{1}{\theta_S} \ln \left(\sum_{s \in \{A,N\}} e^{\theta_S \delta_{js}} \right)$$

   Intuition: This form is convenient when nesting another Extreme-Value layer (location choice). It mirrors the sector step but aggregates "sector-inclusive values" across locations instead of sectors

4. **Use W_j in the outer (location) nest**

   At the location stage, the random utility is

$$\ln \widetilde{U}_j = \ln W_j + \upsilon_j \qquad \upsilon_j \sim \text{Gumbel}(0, 1/\theta_L)$$

   Intuition: Once sectors are collapsed to W_j, the location problem "sees" only this scalar measure of j's attractiveness, plus a new idiosyncratic location shock. Applying the same logit logic across j delivers the location probability

5. Set up the probability of living in location j:

   An individual lives in j if j provides a higher utility than any other location. Thus, the probability of being in location j is:

$$\pi_j \equiv \Pr\left[\ln \widetilde{U}_j \geq \ln \widetilde{U}_k \ \forall k\right], \qquad \ln \widetilde{U}_j = \ln W_j + \upsilon_j, \qquad \upsilon_j \overset{i.i.d.}{\sim} \text{Gumbel}(0, 1/\theta_L)$$

6. Condition on the location shock υ_j and integrate it out

$$\pi_j = \int_{-\infty}^{\infty} \Pr\big[\ln \widetilde{U}_j \geq \ln \widetilde{U}_k \ \forall k \mid \upsilon_j = u\big] \ f_\upsilon(u)\, du \tag{53}$$

$$= \int_{-\infty}^{\infty} \underbrace{\prod_{k \neq j} F_\upsilon\big(u + \ln W_j - \ln W_k\big)}_{\text{independence across } k} \ f_\upsilon(u)\, du \tag{54}$$

7. Plug in the Gumbel CDF/PDF

   The Gumbel CDF and PDF (scale $1/\theta_L$) are

$$F_\upsilon(x) = \exp\big[-e^{-\theta_L x}\big], \qquad f_\upsilon(x) = \theta_L e^{-\theta_L x} \exp\big[-e^{-\theta_L x}\big]$$

   Hence the integrand in (54) becomes

$$\prod_{k \neq j} \exp\big[-e^{-\theta_L (u + \ln W_j - \ln W_k)}\big] \cdot \theta_L e^{-\theta_L u} \exp\big[-e^{-\theta_L u}\big]$$

   Factor out $e^{-\theta_L u}$ terms. Define

$$S_j \equiv \sum_{k \neq j} \left(\frac{W_k}{W_j}\right)^{\theta_L}$$

   Then the product simplifies to

$$\exp\Big[-e^{-\theta_L u} \sum_{k \neq j} \left(\frac{W_k}{W_j}\right)^{\theta_L}\Big] \exp\big[-e^{-\theta_L u}\big] \ = \ \exp\big[-e^{-\theta_L u}(1 + S_j)\big]$$

8. Change of variables

   Let $z = e^{-\theta_L u}$ so that $dz = -\theta_L e^{-\theta_L u} du$. As $u \to -\infty$, $z \to \infty$; as $u \to +\infty$, $z \to 0$. Thus (53) becomes

$$\pi_j = \int_{-\infty}^{\infty} \theta_L e^{-\theta_L u} \exp\big[-z(1 + S_j)\big]\, du \tag{55}$$

$$= \int_{\infty}^{0} \exp\big[-z(1 + S_j)\big]\, (-dz) = \int_0^{\infty} e^{-z(1+S_j)}\, dz \tag{56}$$

9. Evaluate the elementary integral

$$\int_0^{\infty} e^{-z(1+S_j)}\, dz = \frac{1}{1 + S_j} = \frac{1}{1 + \sum_{k \neq j} (W_k / W_j)^{\theta_L}} = \frac{W_j^{\theta_L}}{\sum_k W_k^{\theta_L}}$$

10. Final expression

Thus, the probability of being in location j is defined as a function of wage indices across locations and the θ_L elasticity from the shocks:

$$\pi_j = \frac{W_j^{\theta_L}}{\sum_{k=1}^{J} W_k^{\theta_L}} \tag{57}$$

Equations (44)-(57) link wages, prices, amenities, and taste parameters (θ_S, θ_L) to the labor allocation that feeds the goods and labor market-clearing conditions in the spatial general equilibrium model

C Appendix: Difference-in-Differences Analysis

This appendix presents a supplementary difference-in-differences analysis that benchmarks the overall settlement effect by comparing cities near settled transmigration villages with cities near planned but unsettled development areas. This exercise provides a complementary perspective to the main agro-similarity design but should be interpreted as preliminary: it relies on Euclidean distance buffers rather than the cost-distance measures used in the main analysis, and the results are sensitive to the chosen spatial radius.

C.1 DiD Identification Strategy

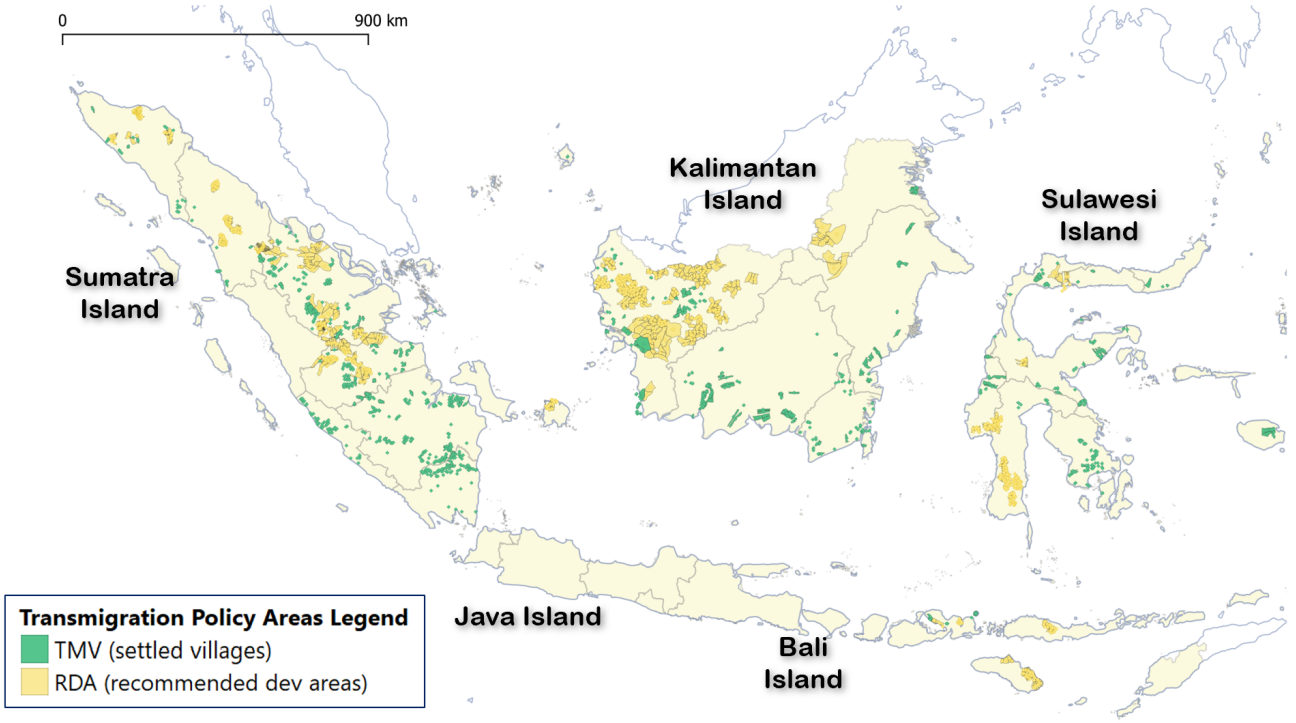
This is a preliminary version of an alternative identification using the settled vs. planned development areas for Transmigration villages. Currently, these are suggested results on overall village placement effect, which I plan to later incorporate into the model estimation and counterfactuals analysis later.

To further our understanding of the effects of a rural population shock on nearby urban growth, I provide an alternative specification in this section which will allow me to separately evaluate the effect of rural population growth in aggregate rather than agricultural productivity on urban growth. Here, I use a difference-in-difference specification, comparing cities near settled TMVs with those near planned but unsettled development areas (referred to as "recommended development areas" (RDAs) or "planned development areas" (PDAs) (See their distribution relative to the TMVs in Figure 9. The relevant assumption relies on the parallel trends in urban growth between cities near settled villages prior to settlement compared to cities near planned areas that had never been settled, driven by the fact that unsettled sites had a high likelihood of being settled had it not been for the abrupt cancellation of the program due to declining oil revenues.

The main idea behind the empirical analysis I present below is anchored in the hypothesis that further urban growth is enabled by rural population growth where excess rural population is absorbed by the cities, resembling the Lewis dual economy (Lewis 1954; Harris & Todaro 1970). This framework is

structured around the classical models of urbanization, where rising agricultural productivity, stemming from technological advancements and increased rural population, is posited as a catalyst for urban expansion. In the setting of Indonesia's rural resettlement policy, I empirically validate whether the increase in rural population and agricultural productivity under the Transmigration program has indeed spurred urban growth in the Outer Islands, thereby testing the classical theories of urbanization in a lower-middle-income country.

Figure 9: TM Populated Sites & Recommended Development Areas (PDA) (not settled)

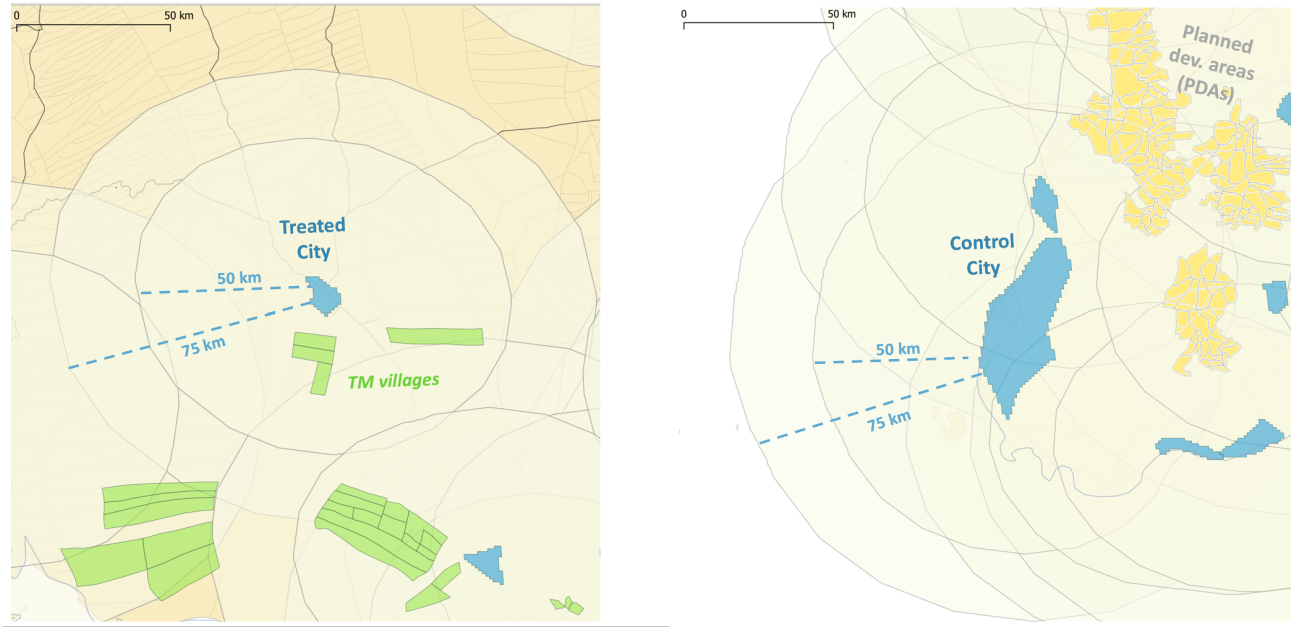


Notes: original map based TM Census data extracted from Bazzi et al. (2016) replication files.

The Harris-Todaro (HT) model, which claims that migration decisions are driven by the expected income differentials between urban and rural areas, rather than actual wage differences, incorporates the probability of obtaining urban employment and the relative wages in each sector (Harris & Todaro 1970). It draws on earlier work on the dual (agrarian and industrial) economy by Lewis (Lewis 1954), which centers on the structural transformation of an economy from agrarian to industrial. It suggests that surplus labor from the traditional (usually agricultural) sector is absorbed by the modern (industrial) sector, facilitating economic development without increasing wages until surplus labor is exhausted.

To investigate how population growth in rural areas affects urbanization in neighboring areas, I propose to compare cities based on their exposure to the rural shock from TM village settlement by matching cities to villages based on distance (see Figure 10). I identify control cities as those located near the planned development areas (PDAs) that were assigned but not settled due to the abrupt cancellation of the program. After demonstrating that baseline geographic characteristics are balanced, I compare the urban growth in treated vs. control cities in several specifications.

Figure 10: Zoomed-In Sample of a Treated vs. a Control City)



Notes: an example of a treated city (with a settled village and no planned area nearby) and a control city (a planned area, but no settled transmigration village nearby).

Using the individual-level census data and the grid-cell data on population density and urbanization level, I identify cities that had TM villages in their vicinity within a certain radius (treatment) and those that did not (controls). Control cities include those that have planned development areas (PDA) within the same radius around them and those that have neither PDAs nor settled TM villages.

To identify the initial set of cities to be split into treatment and control, I am proposing two avenues. First, I use the 2015 functional urban areas (FUAs) (based on the GHS data) and, within those, compared cities that had settled TM villages nearby and those that didn't – the advantage is that in this case, only comprehensively identified urban areas (as of 2015) are utilized, while the disadvantage is in relying on outcome data (post TM program) to determine the sample (cleaner and more straightforward, but with a minor endogeneity concern). Next, I plan to use the 1975 degree of urbanization layer (GHS 2022) and identify semi-urban areas and urban clusters and consider them to be local urban centers – a data-driven approach without relying on a pre-set list of administratively-defined cities, which might also enable me to investigate emergence of cities.

Utilizing a buffer radius for matching cities to TM villages, while straightforward, introduces complexities in actual connectivity and accessibility, influenced by factors like terrain and transportation infrastructure. Ideally, the study would use pre-TM road network maps from the 1970s to accurately determine the locations of TM villages in relation to existing roads and railways. In situations where historical maps are unavailable, alternative measures of geographic distance might provide a more accurate representation than simple linear distances. The paper anticipates presenting preliminary results based on these factors, starting with the most basic specification.

Therefore, my main preferred specification for urban growth in city c in district d is as follows:

$$\ln \text{Pop}_{ct} = \alpha + \gamma \text{ Post}_t + \kappa \text{ Treat}_c \times \text{Post}_t + \eta_c + \varepsilon_{ct} \tag{58}$$

if near TMV OR near PDA (flexible radius 50km/100km)

Since the specification includes city fixed effects η_c, the time-invariant treatment indicator Treat_c is absorbed and omitted from the equation.

where $Treat_c$ is an indicator for city c for being located near a TMV: it is equal to 1 for "treated" cities that had a settled TMV in their vicinity during the TM policy and equal to 0 for those that had a PDA located within the study radius but no settled TMV. $Post_t$ is an indicator equal to 1 for periods in the panel after 1985 (by the time when most TMs had been settled) and their interaction term of $Treat \times Post$ identifying the treatment effect of interest (κ), and η_c are city fixed effects. "If near TM OR near PDA" corresponds to the same sample selections as before: only including cities that had a TMV or a PDA, while excluding cities that had neither or that had both in their vicinity.

In the event study specification (Figure 16), treatment timing is defined dynamically as the year the first TMV was settled in each city's vicinity, allowing for staggered adoption.

In Figure 10, we see how several reasonable distances (e.g., 50km vs. 75km) might represent the city-village connections. Figure A6 shows that looking at an area up to 100km around cities encompasses separate island shores or, at times, even neighboring islands. It appears that this may create noise in the estimates, and we might be interested in looking at a smaller radius that captures more local changes. Second, comparing panels C and D, we observe that a smaller radius around the city also enables us to identify more unique local trends, with more cities having only a matched village or only a PDA, but not both. I will show the robustness of the main results across different distance bands (25km, 50km, 75km) and will implement the cost-distance travel time measures later to capture a more realistic connectivity measure, robust to island geographies.

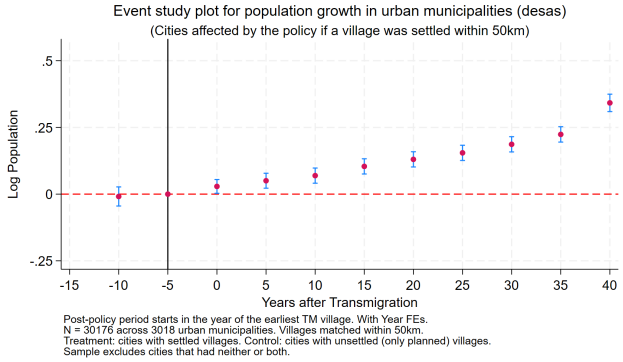
C.2 Preliminary DiD Results

This section presents results comparing the urban areas that received transmigration villages (TMV) in the vicinity vs urban areas that had the Planned Development Area (PDA) in the vicinity, but no TMV. The results below correspond to the simple version of the distance measure matching cities and villages by proximity. The results below are based on the simple Euclidean distance and buffer-zone matching that I implemented (see Figure 10), without yet accounting for the road network.

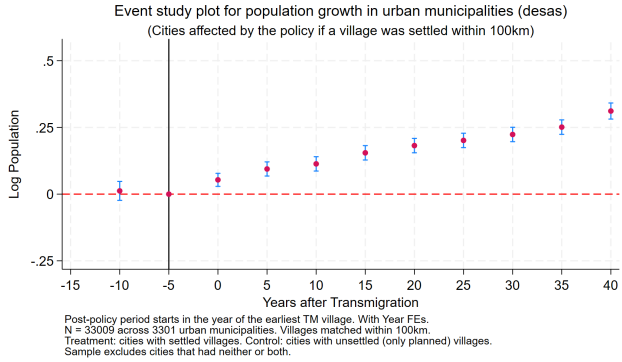
Figure 11 shows that the parallel trends assumption is reasonable and that the positive effects are delayed until later periods. Table 3 shows the preliminary results for this specification where the effect of interest is positive, but not statistically significant across the three tested radia (50km/75km/100km).

Figure 11: Event Study Plot for Urban Municipalities

(a) Matching within 50km



(b) Matching within 100km

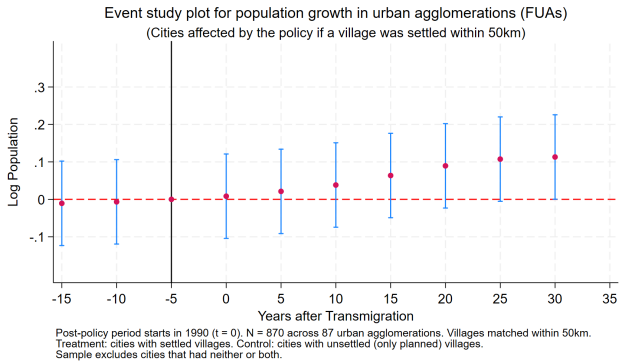


Notes: Here, the first year of policy is as the year of the first TMV settled in the vicinity. It allows for a more robust estimate of the treated year.

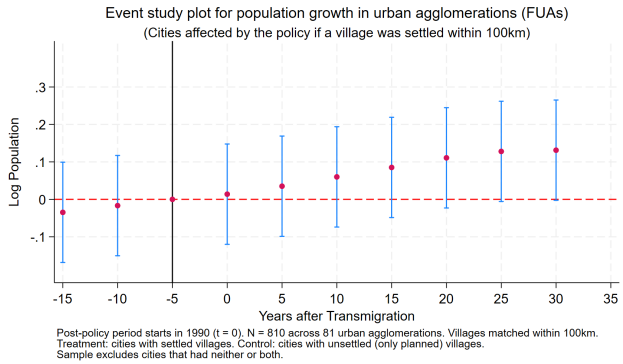
The estimates are less precise when looking at FUA agglomerations in aggregates since there are fewer observations left (Figure 12).

Figure 12: Event Study Plots for Urban Agglomerations (FUAs)

(a) Matching within 50km



(b) Matching within 100km



Notes: Here, the first year of policy is identified as 1990 allowing me to look at 1975-1985 pre-periods.

For the difference-in-difference specification, Table 2 shows that, on average, cities near TM villages grew slightly faster after the village settlement compared to those where the villages were not settled due to the policy cancellation. Part of the challenge in obtaining a sufficiently large sample size is that I currently exclude cities that had both a TMV and a PDA in the vicinity, which I plan to test in an alternative specification in the future. In the current specification, the coefficient on the interaction term is imprecisely estimated, potentially also driven by the small sample of cities in this panel, which motivated my next step of looking at smaller units of individual municipalities in the larger agglomerations (desa-equivalent areas) that are underlying the large urban agglomerations underlying the current sample. More precise estimates currently in progress that incorporate the road network data might help identify a cleaner causal effect.

Note: In columns 4–6, the time-invariant main effect of treatment is mechanically absorbed by city fixed effects; the reported coefficients for the treatment indicator should be interpreted with caution.

Table 2: Effect of Proximity to a Transmigration Village on City Population Growth in 1975-2020

	(1) LnPop-25km	(2) LnPop-50km	(3) LnPop-75km	(4) LnPop-25km	(5) LnPop-50km	(6) LnPop-75km
Post=1	0.312***	0.316**	0.306**	0.312***	0.316***	0.306***
	(0.092)	(0.125)	(0.142)	(0.045)	(0.049)	(0.058)
Treated (25km)=1	-0.103			1.020***		
	(0.115)			(0.050)		
Treated (50km)=1		-0.111			0.826***	
		(0.131)			(0.046)	
Treated (75km)=1			-0.364**			0.777***
			(0.143)			(0.053)
Treated (25km)=1	0.103			0.103		
× Post=1	(0.137)			(0.072)		
Treated (50km)=1		0.069			0.069	
× Post=1		(0.155)			(0.066)	
Treated (75km)=1			0.139			0.139*
× Post=1			(0.169)			(0.076)
City FE				✓	✓	✓
Mean dep var	11.58	11.64	11.63	11.58	11.64	11.63
Std err dep var	0.95	1.01	1.04	0.95	1.01	1.04
Observations	840	870	870	840	870	870
R^2	0.03	0.02	0.04	0.93	0.95	0.94
F stat	9.53	8.28	13.59	.	.	.

Notes: The dep. var. is ln pop, first treated year is 1990, based on exposure to settled TM villages within a radius. City sample matched to either a TM village (TMV) or a planned dev. area (PDA) (not both) within the radius. Exposure radius in column titles. Robust s.e. or clustered s.e. with city FEs. *** 1%, ** 5%, * 10% significant.