The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia*

Gharad Bryan †
London School of Economics

Melanie Morten ‡
Stanford University and NBER

May 9, 2018

Abstract

We estimate the aggregate productivity gains from reducing barriers to internal labor migration in Indonesia, accounting for worker selection and spatial differences in human capital. We distinguish between movement costs, which mean workers will only move if they expect higher wages, and amenity differences, which mean some locations must pay more to attract workers. We find modest but important aggregate impacts. We estimate a 22% increase in labor productivity from removing all barriers. Reducing migration costs to the U.S. level, a high-mobility benchmark, leads to a 7.1% productivity boost. These figures hide substantial heterogeneity. The origin population that benefits most sees an 104% increase in average earnings from a complete barrier removal, or a 25% gain from moving to the U.S. benchmark.

Keywords: Selection, Internal migration, Indonesia
JEL Classification: J61, O18, O53, R12, R23

*We thank the editor, three anonymous referees, as well as Ran Abramitzky, Abhijit Banerjee, Paco Buera, Rebecca Diamond, Dave Donaldson, Pascaline Dupas, Greg Fischer, Pete Klenow, David Lagakos, Ben Olken, Torsten Perrson, Steve Redding, Daniel Sturm, Adam Sziedl, Michael Waugh, and seminar audiences at the Stanford Institute for Theoretical Economics, Paris School of Economics, Namur, Helsinki, Berkeley, Yale, LSE, Columbia, the University of Chicago, Toronto, NBER SI 2015, Harvard/MIT, and the CEPR/PODER conference for helpful comments and suggestions. Chris Becker, Anita Bhide, Allan Hsiao, and Jay Lee provided outstanding research assistance. Any errors are our own.

†Email: g.t.bryan@lse.ac.uk

‡Email: memorten@stanford.edu

1 Introduction

Recent evidence suggests that a policy of encouraging internal labor migration could have large productivity effects in developing countries. On the macro side, Gollin et al. (2014) show that nonagricultural (urban) workers produce four times more than their agricultural (rural) counterparts. On the micro side, Bryan et al. (2014) show a 33% increase in consumption from experimentally induced seasonal migration. Neither of these results, however, is definitive: The experimental estimates apply only to seasonal migration, and to a specific part of Bangladesh. The macro estimates do not account for selection on unobservables (Young 2013), and only apply to movement between rural and urban areas.

This paper uses micro data from Indonesia to quantify the aggregate effect of increasing mobility. Two observations motivate our approach. First, migration could increase productivity if it: (1) allows individuals to sort into a location in which they are personally more productive (sorting); (2) allows more people to live in more productive locations (agglomeration); or (3) both.[1] Second, in the absence of constraints or amenity differentials, people will maximize their production; therefore, a policy that encourages migration will have no effect on output if there are no existing constraints on mobility.

We build a model in which workers have idiosyncratic location-specific productivity, and in which locations differ in their overall productivity. This setup allows for both sorting and agglomeration effects. Into this framework we incorporate two kinds of mobility constraints. Movement costs exist if workers must be paid higher wages to induce them to work away from home. Compensating wage differentials exist if workers must be paid higher wages to work in low-amenity locations. The result is a general equilibrium Roy model in which workers sort across locations that have heterogeneous amenities and productivities. The model is similar to that used by Hsieh et al. (2018); our approach also has close connections to the seminal work of Hsieh and Klenow (2009).[2] We use this

[1] We use the term agglomeration to encompass two mechanisms that are often separated in the literature: the first is more people living in locations with higher fundamental productivity, the second is the externalities that arise when more people live close to each other.

[2] Our framework also has much in common with recent quantitative models of economic geography such as Allen and Arkolakis (2014), Redding (2016) and Desmet et al. (2016). We also draw on important contributions studying commuting, e.g. Monte et al. 2018 and Ahlfeldt et al. (2015). Our framework is similar to that used in work by Tombe and Zhu (2015). Relative to that paper, we use more detailed micro

structural framework to quantify the change in aggregate productivity that would result from removing movement costs and/or equalizing amenity differentials. Like Hsieh and Klenow (2009) and Caselli (2005), we do not consider specific policies, but rather try to quantify the potential impacts of a set of policy options.

Our main contribution is combining this quantitative framework with rich micro data from Indonesia. The Indonesian data, which are unique in recording location of birth, current location and current earnings, allow for particularly transparent identification of key model parameters. For example, we are able to identify the key parameter that controls sorting from a simple linear regression of the origin-destination wage on the origin-destination migration share. Intuitively, across-destination/within-origin variation in migration rates can be used to estimate the strength of selection forces, but few datasets contain the information necessary to run this regression.

Before turning to our structural analysis, we document five motivational facts, which suggest both that movement costs and compensating differentials exist, and that selection is important in the data. Our rich micro data allow us to demonstrate these facts. In the case of movement costs, we first show that a gravity relationship holds in the data. A 10% reduction in the distance between two locations leads to a 7% increase in the proportion of migrants who flow between the two locations. We also show that people who live farther from their location of birth have higher wages. A doubling of distance leads to a 3% increase in average wages, suggesting that people need to be compensated to induce them to move away from home. In running these regressions, we think of distance as a proxy for movement costs, which may not capture all policy-relevant constraints. For compensating differentials, we show that workers in observably low-amenity locations receive higher wages. Selection effects also appear to be important in the data: the greater the share of people born in origin o that move to destination d, the lower their average wage. The elasticity of average wage with respect to share is approximately -0.04. Importantly, because our model is one in which movement costs reduce migration and lead to selection, we show that there is almost no effect of distance on average wages once the

data which enables us to directly estimate the extent of selection, and we are interested in a different set of questions.

proportion of the origin population at the destination is controlled for; proportion migrating is sufficient to account for the wage differences. All of these effects are predicted by our model. We also show that the same set of motivating facts hold for migration between states in the U.S.

To estimate the potential effects of policy, we turn to our structural model. When estimating the model, we treat both movement costs and amenity differentials as nonparametric objects to be inferred from the data. Movement costs are nonparametric in the sense that we estimate a separate cost for each origin-destination pair that is independent of distance or any other measure. Our measures of movement costs therefore capture a wide range of barriers. For example, language differences that reduce bilateral migration would be a movement cost. Amenities, following the tradition in urban economics, are estimated as a residual.The choice to treat movement costs and amenity differentials in this way reflects our view that amenities are hard to measure and distance is unlikely to capture all policy-relevant dimensions of movement costs.

Our model allows for straightforward quantification of the effects of reducing movement cost-driven, or amenity-driven, wage differentials. The intuition is straightforward. We first generate counterfactual population distributions by estimating where people would live if we removed their empirical tendency to stay at their place of birth and their tendency to avoid some locations that have high measured productivity. Next, we ask how productivity would change if people moved as suggested by our counterfactuals. Our model of selection implies that each additional migrant will earn less than the last; to account for this we need to understand how wages change as workers move. Since selection, in our model, is relative to location of birth, it is the average wages of people from a given origin who live in a given destination that matter. As noted above, our unique data, which captures both location of birth and current location of work, combined with an instrumental variables (IV) strategy inspired by our model, allows us to estimate the relevant elasticity.

Our results suggest moderate aggregate gains, but important heterogeneity. Removing all frictions is predicted to increase aggregate productivity by 22%. These gains are modest relative to the potential gains suggested by studies such as Gollin et al. (2014), but are in line with what one may expect from other microeconomic studies. For the people born in some locations, however, the results are much larger, with predicted gains peaking at 104%. We show, theoretically and empirically, that gains are larger for origins that have higher dispersion in average wages across destinations. Because complete barrier removal may be impossible, we also compute the gains from moving to the U.S. level of movement costs, which we see as a high-mobility benchmark. We predict an aggregate productivity boost of 7.1%, with the origin that gains most seeing a 25% increase. We conclude that, while migration that improves the static allocation of labor is unlikely to have very large productivity effects of the sort estimated, for example, by Hsieh and Klenow (2009), targeted policies may have big impacts on the lives of some communities.

Our paper differs from existing approaches in three ways. First, we consider region-to-region rather than rural-to-urban movement. Since Lewis (1954) and Harris and Todaro (1970), the development and migration literature has been dominated by rural to urban studies. In our setting this is potentially inappropriate. Figure 1 shows kernel density plots of the log of the average monthly wage, calculated at the sub-province (Indonesian regency) level and broken down by rural/urban status.[3] The figure highlights that while there is large variation between regencies, there is little overall variation between rural and urban locations. Table 1 shows that the majority of migration also occurs within category, rather than across category: between 75 and 85% of migration out of urban areas is to another urban area, and between 25 and 30% of migration out of a rural area is to another rural area. Focusing only on rural-urban migration misses the within-rural and the within-urban migrations.

Second, we focus on counterfactual estimates that predict the effect of removing constraints. While we can learn much from work documenting returns to past migration,[4] there are challenges moving from these estimates to predictions of future returns. On one hand, selection effects mean future migrants may earn less than past migrants; on

[3] We code regencies that have greater than median rural population share as rural, and the remaining regencies as urban. Appendix Figure 1 shows that the same patterns hold if we plot the distribution of individual, rather than regency average, wages.

[4] Recent work by Kleemans and Magruder (2017); Hicks et al. (2017); Beegle et al. (2011); Garlick et al. (2016) provide important estimates of the returns to, and impact of, past migrations in Indonesia, Kenya, Tanzania and South Africa.

the other hand, migration policies work by reducing constraints, and so will tend to encourage migration where past movement was minimal. Because of this, past returns may contain little information on the likely effects of future policies. For our analysis we directly estimate the impact of removing constraints. Our only use of past migration is to estimate the strength of selection effects. While this approach is similar to macroeconomic estimates based on productivity gaps (e.g. Gollin et al. 2014), it accounts for selection effects that are likely to be important.

Finally, we take account of general equilibrium (GE) effects. First, by incorporating sorting, we allow for aggregate productivity gains in the absence of large net populations flows. Second, we calibrate agglomeration, congestion, rental, and price elasticities using consensus estimates, and we then assess how our results depend on these parameters.

Our results are limited in three ways. First, we look only at static gains, leaving examination of dynamic effects for future work.[5] Second, when doing our counterfactuals we look only at the productivity impacts, and only at gains. We do not consider welfare effects of removing migration restrictions (which may be negative) and we do not consider the costs of policy. A full consideration of costs is difficult and can be avoided if benefits are small. Third, we do not consider specific policies, but rather provide estimates of the total gains that may be available. Our approach is similar, therefore, to the development accounting and macro misallocation literatures (Caselli, 2005; Hsieh and Klenow, 2009).

The paper starts by laying out five motivational facts. These facts strongly suggest that spatial labor markets in Indonesia are characterized by costs of movement, compensating wage differentials and selection on productivity. The facts imply the possibility of productivity gains from increased movement. We then provide a simple two-location example that explains how we quantify the possible gains. We follow this by briefly describing our formal model, discussing identification and estimation, and demonstrating that our structurally estimated parameters correlate sensibly with real world proxy measures. Finally, we present results from counterfactual exercises.

[5] There are several potential sources of dynamic gains. For example, migration costs may be endogenous (Carrington et al., 1996), firm openings may depend on the pool of available migrant labor, or both.

2 Data, Motivation, and Two-Location Example

2.1 Data

Our approach has specific data requirements. In our view, people will only migrate if their earnings increase enough to compensate them for living away from home (which we take to be their location of birth). We therefore need data that records an individual's location of birth, current location of work, and earnings. Our interest in aggregate returns implies that data have to be geographically representative. Because we want to nonparametrically estimate movement costs, the dataset must be large enough that it records flows between all pairs of locations. Data of this kind are available in very few locations, and Indonesia is the unique country that meets these specifications and has location recorded at a level below the equivalent of a state.

Our Indonesian data come from the 1995 SUPAS (Intercensal Population Survey) and from the 2011 and 2012 SUSENAS (National Socioeconomic Survey). These datasets record, for a large representative set of people, location of birth (origin o), current work location d (which could be the same as the origin), and monthly earnings (which we refer to as the wage). A limitation of these data is that they do not capture earnings for the self-employed. To understand the biases that this may introduce, we supplement the SUPAS/SUSENAS data with data from the Indonesia Family Life Survey (IFLS), a longitudinal survey. The IFLS has a much smaller sample and by design covers only 13 out of 25 Indonesian provinces, but it does collect more detailed information on incomes, including for the self-employed, and follows the same individuals over time. While we cannot use the IFLS data to estimate the structural model, we can use it to understand how key parameter estimates are affected by the limitations of the SUPAS/SUSENAS data. We also use data from the United States, both to show that our migration facts hold more generally and to generate a suitable counterfactual for a high-mobility economy. We use the 1990 5% Census sample and the 2010 American Community Survey, as these dates overlap most closely with our Indonesia dates. In all cases, we restrict the sample to be male heads-of-household between 15 and 65 years old.[6] Summary statistics for the

[6] This restrictions reduces our sample size in Indonesia from 419,760 to 187,065. We restrict to male head

Indonesian and the United States sample are given in Appendix Table 1. Summary statistics for the IFLS sample are given in Appendix Table 2. All wage variables are reported in monthly terms.

In the U.S., we have locations of birth and work recorded at the level of the state; in Indonesia, we have this information for the regency (and, aggregating up, at the province level).[7] Because of the census nature of our data, our measure of migration is permanent migration based on a repeated cross-section. This may miss people who have moved multiple times, or who have moved and returned home. To ascertain the scope of these issues, we look at detailed migration histories collected in the IFLS. A migration episode in the IFLS is defined as a move lasting at least six months. We find that multiple and return migration are not large issues in our context. As Appendix Table 3 shows, migration in Indonesia can be broadly characterized as one permanent migration episode, made in adulthood. Looking at male household heads, conditional on moving out of the birth province, 69% of all migrants make only one migration, 26% make two moves, and only 5% make three or more moves. Importantly, only 8% of migration is undertaken by people under the age of 16 and 50% of second migrations are made by people returning home. These numbers are broadly similar to those for the U.S., where Kennan and Walker 2011 find that the average male migrant makes 1.98 moves and 50.2% of movers move home.

We use the 2005 and 2011 Village Potential Statistics (PODES) datasets to get measures of amenity. These data are reported by a local leader and contain information on all locations, both urban and rural, in Indonesia. We collapse to the regency level, using population weights.

---

of households as our model is one in which migration is motivated by work, and women and children may migrate for a more diverse set of reasons. As we discuss below, the key parameter that drives our estimates of the gains from migration is the distribution of talent in the population. Reassuringly, estimates of this key parameter change little when we include both non-household heads as well as women. Tables available upon request.

[7] Regency is a second level administrated subdivision below a province and above a district. For all surveys, we drop the provinces of Papua and West Papua. We generate a set of regencies which have maintained constant geographical boundaries between 1995 and 2010. This primarily involves merging together regencies that were divided in 2001. This leaves us with a sample of 281 regencies. Later, for the structural estimates we aggregate regencies up to the level of province, of which there are 25.

2.2 Five Empirical Facts About Migration

From our data, we can calculate the proportion of people from each origin o that move to each destination d, which we denote π_{do}, as well as the average wage within origin destination pair, $\overline{wage}_{do}$. Using this data, we document five empirical facts about migration in Indonesia. We present these five facts at the regency level. For the later estimation of the model, we aggregate regencies into provinces.[8] We then show that these basic facts about migration also hold true in the U.S. sample.

Fact 1 (Gravity: Movement Costs Affects Location Choice)**.** *Controlling for origin and destination fixed effects, the share of people born in o who move to d is decreasing in the distance between o and d.*

To document Fact 1, we run a regression

$$\ln \pi_{dot} = \delta_{dt} + \delta_{ot} + \beta \ln dist_{do} + \epsilon_{dot}$$

where δ_{dt} and δ_{ot} are destination-year and origin-year fixed effects respectively and $dist_{do}$ is the straight distance between regency o and regency d.[9] The destination effect controls for any productivity or amenity differences across destinations, and the origin effect controls for the benefits of other possible locations from the perspective of those living at the origin (this term is similar to the multilateral resistance term in the trade literature).

We interpret distance as a proxy for movement costs, which we think include both the costs of travel as well as a broader set of concerns including cultural differences and language differences. The results are shown in Table 2 Column 1. We estimate that the elasticity of π_{do} with respect to $dist_{do}$ is negative, strongly significant, and sizeable. A 10% increase in distance leads to a 7% reduction in the proportion migrating. These results suggest that there are costs of moving people across space.

[8] The Indonesian results are also robust to aggregating to the province level (Appendix Table 4) and using the IFLS data (Appendix Table 5). We report our motivational facts at the regency level because it increases power. When we conduct our structural estimation we aggregate to the province level to reduce the number of zeros in the bilateral migration matrix. We discuss the IFLS results in more detail in Section 6.5.2 where we consider the robustness of our estimates.

[9] $dist_{do}$ is the straight line distance, in kilometers, between the centroid of regency o and the centroid of regency d. We have experimented with movement time, generated using Dijkstra's algorithm and assumptions about the time cost of different types of travel. This does not materially affect the results.

Fact 2 (Movement Costs Create Productivity Wedges)**.** *Controlling for origin and destination fixed effects, the average wage of people born in origin o and living in destination d is increasing in the distance between o and d.*

To establish Fact 2, we run the regression

$$\ln \overline{wage}_{dot} = \delta_{dt} + \delta_{ot} + \beta \ln dist_{do} + \epsilon_{dot}.$$

The results are shown in Table 2 Column 2. We estimate that the elasticity of the average wage with respect to distance is positive, strongly significant, and sizeable. A doubling of the distance between origin and destination leads to a 3% increase in the average wages. These impacts can be very large. For example, the straight line distance from Denpasar to Jakarta on the western tip of Java is about 1000km. On the other hand, the distance from Denpasar to Banyuwangi on the eastern tip of Java is about 100km. Our estimates suggest that the average wage of migrants from Denpasar to Jakarta will be 30% more than those to Banyuwangi.

As we explain in more detail in our two location example below, this fact suggests that movement costs reduce productivity. To easily illustrate this, consider two locations d and d' that are identical except that d is closer to o. Fact 2 implies that those who choose to move to d' have higher average wages than those who choose to move to d. Under the hypotheses that the two destinations are identical, that workers are rational, and that workers are paid their marginal product, the only way that those in d' can have higher wages is if distance (movement costs) dissuaded the moves of some positive productivity movers, who would have earned less than the current average wage.

Fact 3 (Selection)**.** *Controlling for origin and destination fixed effects, the elasticity of average wages with respect to origin population share is negative.*

Fact 3 is documented by running the regression

$$\ln \overline{wage}_{dot} = \delta_{ot} + \delta_{dt} + \beta \ln \pi_{dot} + \epsilon_{dot}. \tag{1}$$

Estimates from this regression are presented in Table 2 Column 3. Our estimates, which are strongly statistically significant, show that the elasticity of average wages is negative. In Indonesia, a doubling of the share of people who migrate to a particular destination leads to a 4% decrease in average wages. This fact suggests selection on productivity. If workers are paid their marginal products, then, controlling for destination productivity, the only way that average wages can differ across destinations within origin is if the distribution of worker skills is a function of π_{do}. We show below that the coefficient on $\ln \pi_{dot}$ in this regression is the key parameter that measures the importance of selection and sorting in our model. This fact is subject to a potential endogeneity concern: any shock to productivity in destination d that differentially affects people from different origins o will tend to also alter π_{do}. Below, we use our full theoretical model to motivate an instrument to correct for this. Instrumentation changes the quantitative results, but does not alter the qualitative fact.

Fact 4 (Movement Costs Reduce Productivity by Reducing Selection)**.** *The elasticity of average wage to distance drops to almost zero after controlling for the fraction of the origin population that migrate.*

We document Fact 4 by running the regression

$$\ln \overline{wage}_{dot} = \delta_{ot} + \delta_{dt} + \beta \ln \pi_{dot} + \gamma \ln dist_{do} + \epsilon_{dot}. \tag{2}$$

Results are presented in Table 2 Column 4. The coefficient on $\ln \pi_{dt}$ changes little when the distance control is added, but the magnitude of the estimated distance effect, while still positive and statistically signficant, drops relative to the results in Column 2, falling to an economically insignificant size.

Facts 3 and 4 together suggest a framework where increasing movement costs, proxied here by distance, lead to a reduction in the proportion of people who move (Fact 1). This, in turn, leads to an increase in wages (Fact 2), but these wage effects are generated by a selection effect created by a reduced proportion moving (Facts 3 and 4). This is consistent with our discussion of Fact 2 and 3, where we assume that workers are paid their marginal productivity, so once destination and origin fixed effects are controlled for wage differences reflect selection. Importantly, Fact 4 suggests that our structural approach of estimating the impact of reducing movement costs using the elasticity of wage with respect to proportion moving will capture most of the effects of removing movement cost.

Fact 5 (Compensating Wage Differentials)**.** *Controlling for origin fixed effects, locations with higher amenities have lower wages.*

To document Fact 5 we run the regression

$$\ln \overline{wage}_{dot} = \delta_{ot} + \delta_{dt} + \beta amen_{dt} + \epsilon_{dot}$$

where $amen_{dt}$ is measured amenity in destination d at time t. To determine amenity, we take six different measures of amenity from the Indonesian PODES survey and convert to a single measure by taking the first principal component.[10] We then standardize to give this variable a zero mean and unit standard deviation. The results are shown in Table 2 Column 5. Our estimates imply that a 1 standard deviation increase in amenities leads to a 2.3% decrease in average wages. This is direct evidence that firms pay a compensating wage differential to attract workers to low amenity locations. Importantly, there is little endogeneity concern with the sign of this result. While one may be concerned that higher wage locations can afford higher amenities, this result goes in the opposite direction.

2.2.1 The basic facts also hold in the U.S. data

Table 3 shows that the main facts also hold for the U.S., when migration is defined as crossing a state border. We show evidence for the first four facts as we do not have a measure of amenity at the state level for the U.S. Starting with Column (1), we find evidence of a gravity equation for migration. Column (2) shows that wages in the destination are increasing in the distance measured. Column (3) shows that wages in the destination are decreasing in the share of people migrating, and Column (4) shows that the wage effect is

[10] We have two broad categories of amenities: amenities affecting services ("ease" amenities), such as the ease of reaching a hospital, and negative amenities affecting pollution ("pollution" amenities), such as the presence of water pollution in the last year. A full list of the amenities in the data are given in Appendix Table 6. For the motivating fact we use the "ease" amenities only because we are concerned that pollution is picking up economic output directly. We use the first principle component because we are interested in computing a unidimensional measure of amenities. We only require our measure to be a proxy measure for amenities.

driven by the share of people migrating, not the distance effect. This implies that the same framework can be used to interpret migration patterns in the U.S.: increasing movement costs, proxied here by distance, lead to a reduction in the proportion of people that move, which, because of selection effects, leads to an increase in wages.

2.3 An Example with Two Locations

In this section we briefly discuss a two-location version of our model. We highlight the mechanisms through which migration costs and amenity differentials reduces productivity. We also show how we estimate the productivity impacts of policies that reduce migration frictions. Because of the simplicity of the two-location model, we can give an intuitive graphical analysis.

We think of each work place, or destination d, as being characterized by a productivity w_d and amenity α_d. We also assume that each location produces different goods and that people's productivities depend on their location. In particular, we assume that the wage of person i living in destination d is $w_d s_{id}$, where s_{id} is the skill level of person i for location d. Total utility for person i, from location o who decides to live and work in destination d, is then $\alpha_d w_d s_{id}(1-\tau_{do})$, where τ_{do} is the cost that a person born in origin o pays to live in destination d. We refer to τ_{do} either as a movement cost or migration cost. We assume that $\tau_{do} \in [0,1]$, $\tau_{oo} = 0$ and $\tau_{do} = \tau_{od}$. In our empirical work we will back out α_d and τ_{do} as residuals, and so this way of writing the utility function normalizes the measure of amenities and movement costs relative to wages.

Figure 2 shows the distribution of skill (s_{id}) across two locations, which we call A and B; the figure is drawn from the perspective of people born in location B. If there were no frictions, people would live where their earnings, $w_d s_{id}$, are highest. As drawn, location A has the higher productivity, and all those above the ray OE, which has slope w_B/w_A, should move to location A (that is those in regions I, II, and III should migrate). If the two locations had equal productivity, those above the 45 degree line (in areas I and II) should move to maximize productivity.

With movement costs, people from B must be compensated for their move to A. This means that earnings in A are effectively less valuable, and only those above the line OC, which has slope $w_B/w_A(1-\tau_{AB})$ will choose to move. We can divide those born in location B into four groups. Those below ray OE (the dots in region IV) should not move, because their returns are highest to stay in B, and they do not. Those above OE and below the 45 degree line (the dots in region III) should move, because A has higher productivity than B. The higher productivity in A compensated these people for the fact that their comparative advantage lies in B. With movement costs, these people do not move. Those above the 45 degree line and below ray OC (the dots in region II) should move, for two reasons. First, they have a comparative advantage in location A. Second, A is a more productive location. Consider person x: she loses productivity equal to the distance xy because she has a comparative advantage in A but does not move, and an additional amount yz because A is more productive. These two channels mean that movement costs reduce productivity by reducing sorting, and by reducing agglomeration in high-productivity locations. Finally, those above OC in region I should move and they always do. In line with all models inspired by the work of Roy (1951), this figure shows that those with the most to gain will move first, and therefore suggests limits on the gains to promoting migration. It also highlights that most of the gains from migration are to be had by encouraging movement to places where costs are high, and so historical movements have been low.

Fact 2 and its interpretation can be seen in this diagram. As movement costs increase, fewer people move to A and the wages of those that move increase. This increase occurs because some people who would have been more productive in A now choose to stay in B.[11]

Amenities also move worker locations away from the productivity-maximizing allocation. With amenities, but no movement costs, people now maximize $\alpha_d w_d s_{id}$. The effect can be understood in the same diagram. With no movement costs and B having higher relative amenity, the ray OC would have slope $\frac{\alpha_B w_B}{\alpha_A w_A}$. The same effects – a lack of sorting

[11] This fact depends on the properties of the skill distribution. In the language of Lagakos and Waugh (2013), comparative and absolute advantage must be aligned. Appendix D discusses the relationship between comparative and absolute advantage in our framework. We find evidence consistent with comparative and absolute advantage being aligned. See also Adao (2016) for a discussion.

and too little agglomeration – are present, and, so long as the level of amenity in A differs from the level of amenity in B, productivity will not be maximized. The main difference between amenity differentials and movement costs is that movement costs will reduce migration relative to home, while amenity differentials reduce the number of people living in one location relative to the other.

It is worth noting that selection plays two roles in our model. On one hand, worker heterogeneity and selection are a source of gains. Movement costs, which stop workers from moving to their location of comparative advantage, reduce productivity. On the other hand, selection limits the potential gains from moving more workers to high-productivity locations. In the absence of selection on productivity, all workers who move will have the same wage, and so aggregate impacts of removing amenity differentials can be larger.

Our empirical task is to estimate the gain in productivity that would come from allocating people to their productivity-maximizing location. This problem can be separated into two parts. First, we estimate the movement response. This is equivalent to estimating how many people lie in the triangle OCE. This is conceptually straightforward. In the case where there are no productivity differences between locations, the productivity-maximizing choice is that half the people from B will stay in B and half will live in A. Second, we estimate how this movement will affect the average wages of the four groups in our data: those from A that move to B, those from B that live in A and those that stay in A or B. Functional form assumptions laid out below imply that average wages for these groups are a constant elasticity function of the fraction of the origin population that live in the destination. This elasticity is estimable given our data, which records origin and destination, and is shown in Fact 3 above. Because our data records the proportion of people from each origin who live in each destination π_{do}, and counterfactual population distributions can be expressed in the same way, this elasticity is sufficient to estimate the counterfactual aggregate productivity. In the next two sections, we lay out how these ideas extend to more than two locations, how to account for heterogeneous location productivities, and how we incorporate general equilibrium effects.

3 Model

In this section we present a static general equilibrium model of migration. The model is designed to be as simple as possible, we discuss a number of extensions and how they might affect the results in Appendix B. The model is an adaptation of the labor sorting model in Hsieh et al. (2018), which itself draws on Eaton and Kortum (2002). The model also has similarities with recent work on quantitative economic geography, particularly Allen and Arkolakis (2014), and quantitative urban economics, particularly Monte et al. 2018 and Ahlfeldt et al. (2015).[12]

The economy consists of N locations. Workers are born in a particular origin (o), draw a skill for each destination (d), and sort across destinations according to wages, amenity and migration costs. Migration costs are relative to the birth location. Wages and amenities are endogenous and adjust to ensure equilibrium. We first discuss how workers choose where to live and work taking wages and amenities as given, and then turn to production and general equilibrium determination of wages and amenities.

3.1 Utility and Sorting

L_o individuals are born in each origin o. Each person i receives a skill draw s_{id} for each possible work destination $d \in N$. It seems unlikely that this is literally true, what we have in mind is that people have different talents for different industries, and that different destinations have different represented industries. So, for example, a person who is very talented at data science would have a high draw for San Francisco, while someone with a talent for banking would have a relatively high draw for New York.[13]The individual also

[12]The urban models include a cost of commuting, which is conceptually similar to our treatment of movement costs. See Redding and Rossi-Hansberg (2017) for a review of work on quantitative spatial models.

[13]In fact, as noted by Lagakos and Waugh (2013), the assumption that talent is drawn from a Fréchet distribution is consistent with this interpretation. Hence, we can think of the assumption that s_d is drawn from a Fréchet distribution as being consistent with a richer setting in which individuals receive skill draws for a large number of industries in each destination, and choose the industry that maximizes their wage. The main challenge to this interpretation is that data limitations mean that we are forced to assume that talent draws for each destination are drawn from the same Fréchet distribution; we show in Appendix D that there is no evidence that the shape parameters differ by destination or origin, consistent with this assumption. Given this interpretation of the shock, migration frictions will include frictions that prevent people from moving industry, if that industry move requires migration.

receives a skill draw for her location of origin. Skill is drawn from a multivariate Fréchet distribution,

$$F(s_1, \ldots, s_N) = \exp\left\{ -\left[\sum_{d=1}^{N} s_d^{-\frac{\tilde{\theta}}{1-\rho}} \right]^{1-\rho} \right\},$$

which does not depend on the location of birth.[14] Here, $\tilde{\theta}$ measures the extent of skill dispersion or the importance of comparative advantage. As $\tilde{\theta}$ decreases, there is a greater difference between skills across locations. ρ measures the correlation in skills across locations. As ρ increases, individuals with a high draw in destination d are also likely to have a high draw for destination d'. The interpretation is that each different location has a different set of required skills. To the extent that $\tilde{\theta}$ is estimated to be high, locations do not differ greatly in their skill requirements. We allow for correlation between skill draws to allow for general talent, and the case in which talent is unidimensional is a limiting case as $\rho \to 1$. Throughout it is useful to work with $\theta = \tilde{\theta}/(1-\rho)$ rather than $\tilde{\theta}$.

Innate skills are combined with schooling in the location of origin to become human capital. Location d human capital for individual i born in location o is given by

$$h_{ido} = s_{id} q_o.$$

Throughout, we refer to q_o as the quality of schooling in o, but it likely reflects a broader set of factors that contribute to human capital. We consider the possibility of endogenous acquisition of human capital in Appendix B. The wage per effective unit of labor in destination d for someone from origin o is given by $w_d \epsilon_{do}^{w}$ where w_d is destination d productivity, and ϵ_{do}^{w} is a mean one log normally distributed error which captures any reason why people from origin o may be more productive in destination d (i.e., it is an origin-specific labor demand shifter in destination d). We assume that the error is observed by the individual before they migrate, and we introduce it because it allows for a meaningful discussion (in Section 4.1) of an intuitively important endogeneity issue: any unmeasured characteristic that increases productivity in destination d, will also increase movement to

[14] We later introduce a difference in skill by origin, q_o, the resulting model is isomorphic to one in which the scale parameter of the Fréchet parameter differs across locations. The important assumption is that $\tilde{\theta}$ does not differ by origin.

destination d. The wage for individual i from origin o is therefore

$$\text{wage}_{ido} = w_d \epsilon^w_{do} h_{ido} = w_d \epsilon_{do} s_{id} q_o.$$

Indirect utility for individual i from origin o living in destination d is given by

$$U_{ido} = \alpha_d \epsilon^\alpha_{do} (1 - \tau_{do}) w_d \epsilon^w_{do} s_{id} q_o \equiv \bar{w}_{do} s_{id}. \tag{3}$$

The term $w_d \epsilon_{do} q_o s_{id}$ captures consumption, which is equal to the wage.The term α_d measures the amenity of location d and captures the need for compensating differentials. Moving to a location with half the amenity level would be compensated for by a doubling of earnings. Amenities could include natural beauty, the availability of services, or rental rates.[15] The term ϵ^α_{do} is assumed to be mean zero and log-normally distributed; it captures differences in amenity that depend on location of origin. Again, this error term is observed by the individual before making the decision to move, and ensures that the model does not perfectly fit the data. The term τ_{do} captures the utility cost of living away from home (the origin o), and we refer to it as a moving cost. We assume that $\tau_{oo} = 0$, so that moving away from home to a destination d would require an individual to be compensated with $(1 - \tau_{do})$ times the income. For example, compared to consumption at the origin o, the same level of consumption at destination d may be less pleasurable as it is not undertaken with family and friends. We assume throughout that movement costs are symmetric, so that $\tau_{do} = \tau_{od}$. With this background, known results regarding the Fréchet distribution imply the following results.

First, let π_{do} be the portion of people from origin o who choose to work in destination d. We have

$$\pi_{do} = \frac{\tilde{w}^\theta_{do}}{\sum_{j=1}^N \tilde{w}^\theta_{jo}} \tag{4}$$

where $\tilde{w}_{do} = w_d \epsilon^w_{do} \alpha_d \epsilon^\alpha_{do} (1 - \tau_{do})$. Here $\tilde{w}_{do}$ measures the attractiveness of location d for someone from o. Equation (4) is the key sorting equation, and it asserts that sorting

[15] Much work in the tradition of Rosen (1979) and Roback (1982) separate out rents from other amenities. We discuss how to incorporate rents in Appendix B.3.

depends on relative returns, relative amenities and relative movement costs; it does not depend on the quality of human capital formation in the origin, q_o. That sorting does not depend on q_o is key to our exercise: we wish to distinguish between human capital or schooling effects that lead to higher production and human capital effects which are a barrier to migration. Barriers to migration coming from differences in human capital are, to the extent they are symmetric, captured in τ_{do}. To the extent that human capital differences are a barrier to migration but are not symmetric, they will be captured in ϵ_{do}^{w} and will not form part of our counterfactuals.

Second, we can use this characterization to determine the average skill of workers from o working in d by noting that

$$E(s_d \mid choose\ d) = \pi_{do}^{-\frac{1}{\theta}} \bar{\Gamma}, \tag{5}$$

where $\bar{\Gamma} = \Gamma\left(1 - \frac{1}{\theta(1-\rho)}\right)$ and $\Gamma(\cdot)$ is the Gamma function. This equation implies that the more people from o that move to d, the lower is their average skill. This is intuitive as it implies that there is less selection: the marginal migrant is drawn from further down the left tail of the talent distribution. Finally, we can work out the average wage in a particular location for people from a given origin:

$$\overline{wage}_{do} = w_d \epsilon_{do} q_o E(s_d \mid choose\ \ d) = w_d \epsilon_{do}^{w} q_o \pi_{do}^{-\frac{1}{\theta}} \bar{\Gamma}. \tag{6}$$

Equations (4) and (6) are our main estimating equations. Taking logs of these two equations also shows that the model is consistent with the motivating facts discussed earlier. Fact 1, gravity, is an estimate of equation (4), where distance is substituted for moving cost. Facts 2 and 5 come from (6), with π_{do} substituted from equation (4). Facts 3 and 4 come directly from (6).

One important implication of our modeling choices is worth noting. When we observe large average wage gaps between locations or sectors, it is tempting to think that there will be large productivity gains to moving people. Our model highlights two reasons why this may not be the case. First, the gaps may reflect selection, as in Young (2013). Second, those in low-productivity locations may simply have low human capital in total, captured by low q_o in our model. In our empirical work, we will estimate q_o, allowing for unobservable heterogeneity in the quality of human capital production.

3.2 Production and General Equilibrium

Each location is assumed to produce a differentiated good y_d. This output is produced by a large number of firms in each location that each produce an identical product according to a linear production technology. Profits for firm j in location d are given by

$$\Pi_{jd} = p_d A_d h_{jd} - w_{jd} h_{jd}$$

where A_d is labor productivity in location d, p_d is the price, which firms take as given, w_{jd} is the wage paid by firm j, and h_{jd} is the total amount of human capital employed by firm j. Firms compete for laborers by setting wages w_{jd}, which implies that in equilibrium $w_{jd} = w_d$ and $\Pi_{jd} = 0 \; \forall j$ and so

$$w_d = p_d A_d.$$

Total economy-wide production is given by the constant elasticity of substitution (CES) aggregate

$$Y = \left(\sum_{d=1}^{N} y_d^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$

where y_d is the total production in location d, and σ captures the degree of substitutability between products produced by different locations.[16] Prices p_d are determined by assuming a representative firm chooses y_d to maximize total economy output less the costs of production $\sum_d p_d y_d$.[17] This aggregate final good is costlessly traded across the country, and is chosen as the numeraire. Utility is linear in the consumption of the aggregate final good, leading to the utility function given in (3).

We allow productivity to be endogenous. Total output of good d depends on the

[16] If $\sigma \to \infty$ all products are perfect substitutes, so the case in which all locations produce the same good is a limit case of our model. An alternative specification would be to allow for locations to produce goods that are perfectly substitutable with a decreasing returns to scale production function. Hsieh and Moretti (2018) show that the two approaches are isomorphic.

[17] This implies that prices are determined by the equation $p_d = \left(\frac{Y}{q_d}\right)^{\frac{1}{\sigma}}$.

amount of human capital in location d according to the function

$$y_d = A_d H_d$$

where H_d is the total human capital (or effective labor units) available at location d and

$$A_d = \bar{A}_d H_d^{\gamma}$$

is the productivity of location d. In this formulation, $\bar{A}_d$ can be thought of as intrinsic productivity – an exogenous parameter – which may change over time. For example, New York may presently have high productivity due to its proximity to a port, but this may have been even more important 100 years ago. Current labor productivity, A_d, depends on intrinsic productivity and the total amount of human capital in location d, with γ parameterizing the extent of human capital spillovers, or productive agglomeration externalities.

Finally, amenity is also endogenously determined. We assume

$$\alpha_d = \overline{\alpha}_d \hat{L}_d^{\lambda}$$

where $\overline{\alpha}_d$ is baseline amenity; for example, natural beauty, λ is a measure of congestion effects and likely to be less than zero, and $\hat{L}_d$ is the (endogenously determined) population of location d.

It is important to note one key characteristic of the model. Dividing through (4) and (6), it is easy to show

$$\frac{\overline{wage}_{do}}{\overline{wage}_{d'o}} = \left(\frac{\alpha_{d'}}{\alpha_d}\right)\left(\frac{1-\tau_{d'o}}{1-\tau_{do}}\right).$$

Hence, within origin, there are no wage gaps (per unit of human capital) without frictions (or, if only migration frictions are removed, then there are no amenity-adjusted wage gaps).[18]There are two key assumptions that drive this result. The first assumption is that

[18]Note that this does not imply that average wages of people in a particular destination labor market are not affected by w_d. Average wages differ across origin, with people born in more productive locations having higher average wages.

comparative and absolute advantage are aligned. This leads to the fact that reducing frictions will lead to a convergence in wages. The second assumption is that the elasticity of wages to the proportion of the population (from an origin) is constant and is the same across all locations. In our model we assume a Fréchet distributional assumption which hard-bakes assumption (1), and then because we assume that shape parameters are constant across all locations, this leads to assumption (2). We discuss these points fully in Appendix D where we argue that it is not possible to reject these two assumptions in the data.

The fact that, within origin, there are no wage gaps without frictions means that we rule out the kind of behavior discussed in Young (2013), where selection alone drives wage gaps. Our model is somewhere between the work of Young (2013), in which selection is the sole driver of average wage differences, and the work of Gollin et al. (2014), where raw wage gaps are used to infer potential gains from movement.

Appendix B discusses how this basic model might be extended to account for dynamics, endogenous human capital formation, non-traded goods such as housing, and costly goods trade, and how these extensions would affect our results.

4 Identification and Estimation

In this section, we discuss how we identify and estimate the exogenous parameters of the model $\{\theta, \rho, q_o, w_d, \alpha_d, \tau_{do}\}$. We also note that, while they are important for the counterfactuals, we do not need to take a stand on the general equilibrium parameters (γ, λ and σ) for identification; we discuss their calibration below. We make several normalizations. First, as noted above, we assume that $\tau_{oo} = 0$ and $\tau_{do} = \tau_{od}$: movement costs are symmetric, and it is costless to live at home. Second, we normalize $\alpha_1 = 1$: because we do not observe utility levels, the only variation we have to identify α comes from people's relative preferences for locations. Third, we normalize $q_1 = 1$: we identify only relative qualities of human capital generation. This normalizes productivity w_d as well: the wage w_d is what would be earned by someone living at location d who was born in location 1 and who has a skill draw of 1. This means that any aggregate improvement in human capital generation would be captured in productivities, w, and changes in q would capture changes in the spatial allocation of human capital production possibilities. Appendix B discusses identification challenges that arise in a richer model.

4.1 Identification of Model Parameters

4.1.1 Fréchet parameters: $\{\theta, \rho\}$

Taking the log of equation (6), we have

$$\ln(\overline{wage}_{do}) = \underbrace{\ln(\bar{\Gamma}) + \ln(w_d)}_{\text{Destination fixed effect}} - \frac{1}{\theta}\ln(\pi_{do}) + \underbrace{\ln(q_o)}_{\text{Origin fixed effect}} + \ln\epsilon^{w}_{do}. \tag{7}$$

That is, after controlling for origin and destination fixed effects, the elasticity of the average wage with respect to the proportion of migrants identifies the Fréchet parameter θ. Intuitively, if people are very similar (or destinations differ little in their skill needs), then θ is high, so the marginal migrant is not greatly less skilled than the previous migrant, and the average wage will change little with movement. However, if dispersion in talent is large (or there are large differences in the skill needs in different destinations), then the marginal migrant is much less skilled than the previous migrant, and so their wage is significantly lower.

Inspection of equation (4) shows that the error term ϵ^{w}_{do} also enters the definition of π_{do}. This is intuitive; any random variation that means wages for those from origin o are relatively high in destination d will encourage migration between the two locations. This correlation between the error term and the regressor π_{do} creates an endogeneity problem that will lead us to underestimate the extent of selection by overestimating θ.

We address this endogeneity concern with an instrumental variables strategy motivated by our model. We wish to isolate the variation in π_{do} that is driven by variation in the relative amenity of d and productivity in other locations $\neg d$. The proportion of people from other origins $\neg o$ who migrate to destination d is affected by these factors, but not by the random error ϵ_{do}. The set of migration proportions $\{\pi_{d\neg o}\}$ are therefore valid instruments for π_{do}, although the first stage relationship between $\ln\pi_{d\neg o}$ and $\ln\pi_{do}$ is nonlinear. Therefore, we follow the advice of Angrist and Pischke (2009) and instrument $\ln \pi_{do}$ with the fitted value from a "zero stage" regression in which $\ln \pi_{do}$ is regressed on a polynomial in $\ln \pi_{d\neg o}$. Monte Carlo estimates based on a roughly calibrated version of our model, which we discuss in Appendix C, confirm that this strategy leads to unbiased estimates and suggests that there are few efficiency gains to increasing the polynomial beyond a quadratic.

To separate dispersion and correlation, we use a property of the Fréchet distribution which implies:

$$\frac{\text{var}(w_{do})}{(\overline{wage}_{do}))^2} = \frac{\Gamma\left(1 - \frac{2}{\theta(1-\rho)}\right)}{\left(\Gamma\left(1 - \frac{1}{\theta(1-\rho)}\right)\right)^2} - 1. \tag{8}$$

Using data on the distribution of wages, combined with the θ identified as above, this equation identifies ρ, the parameter defining the within-person correlation of skill. Intuitively, if there is little correlation in skill types, so that everyone has some destination in which he or she excels, then the within-destination origin pair wage variance will be low. If, in contrast, ρ is high, then people of many different skill levels will find the same location to be best, and so the variance in observed wages will be high relative to the mean.

4.1.2 Location characteristic affecting the wage: $\{w_d, q_o\}$

Considering again equation (7), with the estimates of ρ and θ in hand, we can identify w_d from the destination fixed effect by noting $\bar{\Gamma} = \Gamma\left(1 - \frac{1}{\theta(1-\rho)}\right)$, which is identified. We identify w_d in levels using the normalization that $q_1 = 1$. Intuitively, after controlling for selection through π_{do} and the quality of human capital through q_o, any differences in wages between locations must be driven by differences in productivity. The quality of the human capital environment q_o can be similarly determined. After controlling for productivity differences at the destination as well as selection, any differences in wages earned by people from different origins must be accounted for by the relative quality of human capital formation opportunities.

4.1.3 Characteristics affecting movement: $\{\tau_{do}, \alpha_d\}$

Taking the log of (4) gives a gravity equation

$$\ln(\pi_{do}) = \theta \ln(w_d) + \theta \ln(\alpha_d) + \theta \ln(1 - \tau_{do}) - \underbrace{\ln\left(\sum_j \tilde{w}_{jo}^{\theta}\right)}_{\text{origin fixed effect}} + \theta(\ln \epsilon_{do}^{w} + \ln \epsilon_{do}^{\alpha}). \quad (9)$$

This equation allows us to identify movement costs for each destination-origin pair that are nonparametric, in the sense that they are not functions of any other data. Intuitively, low movement could be caused by amenity differences, productivity differences, or movement costs. Among these, movement costs are the only force that leads *both* people from o to be unlikely to move to d *and* people from d to be unlikely to move to o. This intuition is confirmed by rearranging the gravity equation to give:

$$(\ln \pi_{do} - \ln \pi_{oo}) + (\ln \pi_{od} - \ln \pi_{dd}) = 2\theta \ln(1 - \tau_{do}) + \eta_{do},$$

where η_{do} is a zero mean shock specific to the locations d and o.[19] In this equation, τ_{do} can be separately identified from η_{do} under the assumption that movement costs are symmetric, so that $\tau_{do} = \tau_{od}$. We see that movement costs are high when people tend to stay at home, and given an estimate of θ (identified as above), we can use differences in movement relative to staying at home to identify τ_{do}.

The gravity equation also allows for identification of relative amenities. In the tradition of urban economics, these amenities are residuals, not a function of any other data. The multilateral resistance term, $\ln\left(\sum_j \tilde{w}_{jo}^{\theta}\right)$, is correlated with the error, but can be removed by differencing the equation. Given this, and having identified w_d, θ, and τ_{do}, the only unknown in (9) is α_d. Intuitively, amenities are separated from movement costs because, while movement costs lead people to stay at home, amenity differences lead to a systematic flow of people toward particular locations. A location d is identified to have a high relative amenity if there is a flow of people to d that cannot be accounted for by productivity differences, measured by w_d, or by propensity to stay at home, measured

[19] $\eta_{do} = \theta(\ln \epsilon_{do}^{w} + \ln \epsilon_{do}^{\alpha} - \ln \epsilon_{oo}^{w} - \ln \epsilon_{oo}^{\alpha} + \ln \epsilon_{od}^{w} + \ln \epsilon_{od}^{\alpha} - \ln \epsilon_{dd}^{\alpha} - \ln \epsilon_{dd}^{w})$.

by τ_{do}. We can only identify amenities up to a normalization because of the origin fixed effect in the equation.

4.2 Estimation

We estimate the model using Poisson pseudo-maximum likelihood (PPML). The PPML model has several advantages for estimating migration flows. First, because it estimates the level of migration, rather than the log, it can rationalize zero observed migration flows between locations. This is important because in our context, as in most studies of migration and trade flows, zero observed flows are common (Silva and Tenreyro, 2006). Second, the PPML model respects the general equilibrium adding-up constraints implicit in the model (Fally, 2015).

Our two estimating equations, Equations (7) and (9), are:

$$\ln(\overline{wage}_{do}) = \ln(\overline{\Gamma}) + \ln(w_d) - \frac{1}{\theta}\ln(\pi_{do}) + \ln(q_o) + \ln\epsilon^{w}_{do} \tag{7'}$$

$$\ln(\pi_{do}) = \theta\ln(w_d) + \theta\ln(\alpha_d) + \theta\ln(1-\tau_{do}) - \ln\left(\sum_j \tilde{w}^{\theta}_{jo}\right) + \theta(\ln\epsilon^{\alpha}_{do} + \ln\epsilon^{w}_{do}). \tag{9'}$$

The identification assumption to estimate Equations (7′) and (9′) by PPML is that the (level) error terms are mean one and are uncorrelated with the (exponentiated) regressors. As discussed above, we assume that the errors are mean one, and we deal with correlation with the regressors through IV and differencing strategies.

We proceed as follows. We first employ an IV procedure to estimate θ. We then take this estimate of θ and estimate the system of three equations (Equations 7′, 9′ and 8) using a generalized method of moments (GMM) estimator. In implementing the procedure, we drop observations with fewer than five observed migrants from the wage data. Although our estimation method rationalizes the presence of zero observed migration between any two locations, we are concerned about small sample sizes affecting the precision of wage estimates. We bootstrap this entire procedure to generate standard errors for our estimated values of θ and ρ.

5 Estimation Results

This section presents our parameter estimates. Our main goal is to show that our structurally estimated parameters correlate with proxy measures and so they appear to measure something real. We show estimates for both Indonesia and the U.S. We use our U.S. model to estimate U.S. level movement costs to generate a counterfactual for a high-mobility economy. Our preferred estimates of migration cost use no structure other than symmetry. We show that this nonparametric estimate correlates with observable characteristics such as distance.

As noted above, the estimates presented in this section do not require us to take a stand on the general equilibrium parameters $\{\sigma, \gamma, \lambda\}$. These parameters will, however, be important for the results of counterfactuals presented below. We discuss the calibration and robustness of these important parameters in Section 6.

5.1 Fréchet Parameters

Table 4 presents estimates of the distributional parameters for both Indonesia and the U.S. The shape parameter for the distribution of talent is determined by $\tilde{\theta} = \theta(1-\rho)$. We estimate $\tilde{\theta}$ equal to 2.7 for the U.S. and 3.2 for Indonesia. Our estimate of $\tilde{\theta} = 2.7$ compares with the estimate of 2 in Hsieh et al. (2018). Separating the value into the component due to the correlation within person, ρ, and the underlying distribution, θ, we find that talent is more correlated in the U.S. compared with Indonesia (a value of 0.9 compared with 0.7), and shows a less disperse distribution (a value of θ of 28 compared with 13). Appendix Figure 2 shows random draws from the estimated distributions for Indonesia and the U.S., where each axis is the productivity level for location 1 or 2. The figure shows that, taking into account both dimensions, the skill distribution is overall more dispersed in Indonesia than the U.S.

5.2 Migration Costs

We estimate substantial migration costs. Table 4 reports the mean value of τ_{do}. On average, migrants in Indonesia must be compensated with a 39% higher income, while Americans require a 15% gain. In this sense, the U.S. is a high-mobility country according to our estimates.

Migration costs, for both the United States and Indonesia, correlate with distance. Figure 4 plots estimated migration cost τ_{do} against the (log) of distance. Particularly striking is the much lower correlation between distance and movement costs in the U.S. The elasticity of cost to distance is 2% in the U.S., compared to 15% in Indonesia. Several mechanisms are possible causes. It may be that transportation is cheaper in the U.S. Alternatively, it may be that people in the U.S. are more welcoming of migrants from physically distant communities, or that the U.S. is more culturally homogenous.

Measured movement costs also correlate with measures of social distance. Using census data, we construct indices of religious and linguistic similarity.[20] Figure 3 plots the relationship between these indices and movement costs, after controlling for distance. There is no correlation between migration costs and religion, but migration costs are statistically significantly correlated with linguistic similarity.

5.3 Amenities

Estimated amenities correlate with measured amenities. The left panel of Figure 5 shows that estimated amenities are negatively correlated with the (standardized) first principal component of pollution amenities. The middle panel shows that estimated amenities correlate positively with the first principal component of health and market access amenities.[21] In line with the discussion of rents in Appendix B.3, the right panel shows that

[20] The index is constructed by calculating the probability that a person selected at random from the origin will have the same characteristic (religion or language) as a person selected at random from the destination. For example, if the origin is 50% Hindu and 50% Muslim, and the destination is 100% Hindu, then the religious similarity index would be 0.5. If the destination was also 50% Hindu and 50% Muslim, then the index would also be 0.5.

[21] Appendix Table 6 correlates the estimated amenities with observed amenities one-by-one. Each entry in the table is the regression coefficient from a separate regression of estimated amenities on each amenity. As we only have 25 estimated parameters, we do not expect individual signs to necessarily be statistically

measured amenities correlate negatively with average housing costs.

5.4 Quality of Human Capital Formation

Figure 6 shows that the estimate of q_o (educational quality) correlates with average educational attainment. This is to be expected if people choose to receive more schooling where there are higher returns to schooling.

6 Counterfactuals

We now turn to the policy question we posed at the start of the paper: would there be productivity gains from removing mobility constraints? We have in mind policies that change migration frictions only, and leave all other factors, for example trade costs, unchanged. To produce counterfactuals, we need to take a stance on the GE parameters. We set these using estimates from the literature, and then evaluate the robustness of our findings to different choices.

A large literature estimates the agglomeration parameter (γ). The literature is reviewed in Rosenthal and Strange (2004) and Combes and Gobillon (2015). Recent consensus estimates suggest a γ of between 0.01 and 0.02 for the developed world, although some studies (e.g. Greenstone et al. 2010) suggest much higher numbers. Estimates for developing countries are more sparse and suggest a γ up to 1. We present our main estimate for $\gamma = 0.05$, but also consider robustness for numbers between 0 and 0.08. We expect that spatial integration will have a greater impact when γ is high.

A much smaller literature attempts to estimate the congestion parameter λ. As noted in Appendix B, we can think of λ having a component that is due to the pure amenity spillover, λ_a, and a component that is due to endogenous changes in housing prices, λ_r. For the first term, the work in Albouy (2012) could be seen as suggesting that $\lambda_a = 0$ in

significant, but we note the general pattern in these results: overall, measures of pollution are negatively correlated with amenities; measures of health outbreaks such as malaria, tuberculosis, and vomiting are also negatively correlated with amenities, although access to health care facilities seems also to be negatively correlated; village lighting and commercial banks are positively correlated, and we see a mixed pattern for natural disasters such as flooding and earthquakes.

the U.S. In contrast, work by Combes and Gobillon (2015) suggests a λ_a of around -0.04. We take 0 as our starting point and consider various values in robustness exercises. For the housing elasticity, λ_r, there are fewer estimates available in low-income countries. In Appendix Table 7 we use rental data to estimate this for our sample and find a value of 0.25.[22] (For comparison, Saiz 2010 estimates this number to be 0.65 for the United States), which we take as a baseline value. Adjusting for the expenditure share of income on housing, which we take to be 0.3, this implies $\lambda = \lambda_\alpha - 0.3\lambda_r = -0.075$. We predict that as λ decreases (and congestion becomes more important), reducing frictions will have a smaller impact. It will be hard to move people into productive areas, even if movement costs are low.

Accurate estimates of the elasticity of substitution across regions are also hard to obtain. Allen and Arkolakis (2014) use a value of 8, while Bernard et al. (2003) find a value of 4. We use 8 for our main results and consider values between 4 and 8 in robustness tables. We expect that as σ increases, there will be larger benefits to spatial integration: a high elasticity of substitution means that the products from different locations become more substitutable, and so there are larger costs to low production of some goods.

6.1 Reducing Movement Costs

The first policy we consider is removing movement costs. On a practical basis, this might be achieved by policies such as migration subsidies (Bryan et al. 2014), migrant welcome centers, language training, and road building (Morten and Oliveira 2016). To estimate possible impacts, we scale our estimated costs by a reduction factor κ, yielding $\widetilde{(1-\tau)} = (1-\tau)^{1-\kappa}$, with $\kappa \in [0,1]$. When $\kappa = 0$ this corresponds to the baseline case we estimated. When $\kappa = 1$ this corresponds to removing migration costs entirely.[23] When we undertake these counterfactuals. we allow for α_d (the combination of natural amenities and rental

[22] The appendix table also employs the same identification strategy to directly estimate the spillover parameters for amenities and agglomeration. We get an estimate of 0.01 for the agglomeration parameter, 0.04 for the congestion parameter, and 0.25 for the housing price parameter. While caution should be taken with these estimates, as both tests are underpowered (and the amenity test has the incorrect sign), we see these results as being broadly consistent with our choice of baseline parameters of 0.05 for the agglomeration parameter and 0 for the amenities parameter.

[23] The average value of $\tau_{us} = 0.15$ and the average value of $\tau_{ind} = 0.39$, so the policy experiment of lowering migration costs in Indonesia to the U.S. level is equivalent to considering $1 - 0.61/0.85 = 0.28$.

prices) to adjust endogenously.

We find modest gains. We predict an 7.1% output gain from reducing migration costs to the U.S. level, and a 7.5% gain from reducing migration costs to zero. The U.S. is usually considered the archetype of a spatially mobile economy, so the 7.1% figure is probably the maximum attainable. These results are illustrated for a range of values of κ in Figure 7. This figure highlights an important implication of our model: productivity effects of reducing movement costs may be non-monotonic, and, more generally, productivity may decrease as movement costs fall. This can occur because reducing migration frictions can lead workers to move away from high-productivity-low-amenity locations toward low-productivity-high-amenity locations. Our estimates suggest that, in our setting, this negative impact of reducing movement costs does not occur till costs have been substantially reduced, to lower than U.S. levels.

These modest gains hide substantial heterogeneity across origin populations. While the average increase from eliminating all migration costs is 7.5%, the effect ranges from -18% to 68%.[24] That is, the people born in some provinces may see a 68% increase in their average wage $\sum_d \overline{wage}_{do}$. For a move to the U.S. benchmark, the gains range from -5% to 25%. The distribution of gains from complete removal is depicted in the first panel of Figure 8 and the U.S. benchmark is presented in Figure 9. We discuss what drives these heterogeneous results in Section 6.4 below.

As noted above, selection plays two roles in our model. On one hand, skill heterogeneity implies that there are gains from sorting. The greater the heterogeneity, the greater the return to sorting. On the other hand, if each additional migrant earns less than the last, selection will strongly reduce predicted gains from agglomeration. These two opposing mechanisms mean that ignoring selection could lead to us to either over- or underestimate policy gains. To understand the importance of selection, we recompute productivity changes, shutting down the selection margin.[25] Sorting is the main source of output gains

[24]Recall there is no restriction that reducing migration costs will lead to increases in output. Reducing migration costs may lead people to migrate away from high-productivity-low-amenity locations towards low-productivity-high-amenity ones. This is indeed what we see in these counterfactuals.

[25]We do this by setting the endogenous component of human capital equal to 1. This maps to a model where people are migrating based on preference shocks, such as is considered in Allen and Arkolakis 2014 and Redding 2016.

from removing migration costs. Column 1 in Table 5 shows that all estimated gains come from improving worker sorting (we estimate a 7.5% gain with sorting, compared to a 8% loss without sorting). Ignoring selection would lead us to underestimate the gains from removing movement costs.

6.2 Reducing Amenity Dispersion

We consider a counterfactual in which amenities are equalized across space. This could be the result of policies such as encouraging home building in high-demand locations, which would tend to equalize rental rates (Harari 2017 and Hsieh and Moretti 2018), and reducing pollution in high productivity cities and providing equal access to schooling and hospitals, which would tend to equalize natural amenities. In undertaking these counterfactuals we assume that it is possible to fully control endogenous changes in amenity and rents so that all locations are equally desirable places to live. Amenities are estimated to scale. As with movement costs, we rescale amenities by a reduction factor κ, yielding $\frac{\widetilde{\alpha}_i}{\alpha_1} = \left(\frac{\alpha_i}{\alpha_1}\right)^{1-\kappa}$, with $\kappa \in [0, 1]$. When $\kappa = 0$ this corresponds to the baseline case we estimated. When $\kappa = 1$ this corresponds to equalizing amenities across all locations.

Here we do not compute a U.S. benchmark; this is for two reasons. First, we believe that it is plausible to have zero amenity differentials: there is no obvious reason why some locations have to have fewer services and more pollution. Second, in line with the general argument in Hsieh and Moretti (2018), we find that the U.S. has greater amenity dispersion than Indonesia. Hsieh and Moretti (2018) argue that that restrictive housing policies lead to high rents in some very productive locations; this would show up in our estimates as high amenity dispersion.

We find that equalizing amenities would lead to an increase in output of 12.7%. These gains are illustrated in panel two of Figure 7. As with migration costs, we find substantial heterogeneity. Some origin locations receive wage gains of up to 88%. Again, we explore what drives this heterogeneity in Section 6.4 below.

As above, we ask how these results are affected by selection. We find, in Column 2 of Table 5, that, in contrast to migration costs, removing the selection margin has very little effect on predicted gains. That is, by ignoring selection, we overestimate the gains from agglomeration.

6.3 Reducing both Migration Costs and Amenities Differentials

Finally, we consider eliminating both barriers – migration costs and compensating differentials – simultaneously. These gains are illustrated in panel three of Figure 7. Doing so leads to a 21.7% output gain. The effect of reducing both barriers is slightly smaller than the sum of their independent effects, suggesting the policies are very mild substitutes. Under the policy of reducing all barriers to mobility, the origin that benefits the most would face wage increases of 104%. For this combined policy, accounting for selection is also important. Column 3 in Table 5 shows that, if we do not account for selection, we understate gains by 40%.

6.4 Understanding Heterogeneity

What explains why some origins gain more than others? The regions that gain the most will be those locations that have the largest ex ante frictions, that is, the locations that are isolated due to migration costs or because they have higher amenity or low house prices. Formally, we can derive an intuitive expression that shows which locations gain the most. In the absence of any frictions (amenity differentials or movement costs), wages should be equalized within origin. That is, average wages of people from origin o who live in destination d, $\overline{wage}_{od}$ should be the same as the wages of those from origin o who live in destination d', $\overline{wage}_{od'}$. Define $wage_o = \sum_d \overline{wage}_{od}$ as the observed earnings of all people from origin o, and define $\widehat{wage}_o = \sum_d \widehat{\overline{wage}}_{od}$, the counterfactual earnings if all distortions were removed. We can show that if price adjustments are ignored:

$$\frac{\widehat{wage}_o}{wage_o} = \frac{L_o \left(\sum_d \pi_{do} \left(\overline{wage}_{do} \right)^{\theta} \right)^{\frac{1}{\theta}}}{L_o \left(\sum_d \pi_{do} \overline{wage}_{do} \right).}$$

In words, the ratio of optimal wages to current wages is higher the greater the "dispersion" in averages wages across destinations. As in Hsieh and Klenow (2009), the equation makes clear what the appropriate measure of dispersion is: it is the geometric mean of wage calculated with respect to θ. This gives a simple data-driven measure of the locations that are likely to gain most.[26] We show in Appendix Figure 3 that the average wage gain at the origin is indeed increasing in this measure of the initial variance of wages at the origin.

6.5 Robustness

This section discusses four robustness exercises.

6.5.1 Agglomeration, congestion, and substitution parameters

The main results use our baseline parameters for the agglomeration, congestion, and substitution. We undertake robustness over these parameters. Results are reported in Appendix Tables 8 through 11. As expected, when agglomeration is high, congestion forces are low, and the elasticity of substitution is high, the gains to removing barriers to mobility increase. For the experiment of reducing both migration costs and amenities, our baseline estimate was an increase in output of 21.7%. The range of results in Appendix Table 11 is from 15.9% to 24.7%.

6.5.2 Self-employment versus wage work

As noted above, a limitation of the SUSENAS and SUPAS data is that it does not record earnings for the self-employed. This may be a source of bias in our estimates. For example, if migrants are more likely to engage in wage work, our estimates of the impact

[26]This measure does not take in to account general equilibrium effects through changes in p_d. If we wish to do this, we have to generalize the above equation slightly:

$$\frac{\widehat{wage}_o}{wage_o} = \frac{L_o \left(\sum_d \pi_{do} \left(\frac{p_d}{\hat{p}_d} \right)^{\theta} (\overline{wage}_{do})^{\theta} \right)^{\frac{1}{\theta}}}{L_o \left(\sum_d \pi_{do} \overline{wage}_{do} \right)}$$

where p_d is the distorted set of prices across destinations, and $\hat{p}_d$ is the undistorted set of prices. Consider an origin location that has equal wages everywhere, except one location d^*, where wages are very high, implying large distortions. Removing all distortions, this origin location will send many workers to d^*, but this process will tend to depress prices, and hence wages, in d^*. The term $\left(\frac{p_d}{\hat{p}_d} \right)^{\theta}$ accounts for this effect. While it is possible to derive a closed form solution for this number, it does not add clearly to the intuition.

of migration on average wages will include both a sample selection effect and a causal effect. This will tend to bias our estimates of the key selection parameter θ. To explore the importance of this issue, we make use of the IFLS data, which records income for both wage and self-employed work. We cannot use the IFLS for our main analysis because it is too small a sample, and we do not expect to get the same estimate of θ using the IFLS, because it is not representative of Indonesia as a whole. Nevertheless, we can use the IFLS to understand the likely direction of bias in our estimates.

Appendix Table 5 replicates our motivating facts using the IFLS sample. The top panel shows results for the sample as a whole, and the bottom panel shows results for wage workers only. The table shows that the elasticity of average wage to the share migrating (Column (3)), which is the inverse of the selection parameter θ, is larger for all individuals than for wage employees only. We learn two things from this exercise: (1) the motivational facts are qualitatively robust to including self-employed individuals, but (2) the dispersion of talent may be smaller for wage employees than self-employed individuals. The implied θ for all individuals is around half the size as for wage earners alone. If this is the case, then our exercise is likely a lower bound on the gains from removing migration barriers in Indonesia as a whole. We show this is indeed the direction of the bias in a robustness exercise in Appendix Table 12, where we simulate our model assuming that θ is half the size of the baseline estimate of θ. The smaller θ suggests total gains from removing migration barriers that are on the order of 23%, rather than the 22% from our baseline model.

6.5.3 Asymmetric migration costs

As discussed earlier, identification of the model relies on the assumption that movement costs are symmetric. One may be concerned that this is a strong assumption; for example, it would be reasonable to have a prior belief that moving from a small country town to a large metropolis may be a more costly move than the reverse. We are able to introduce asymmetry to a limited extent by parameterizing a deviation from symmetry. For example, we can assume that $\tau_{do} = \kappa\tau_{od}$ whenever d has a higher population than o, and we can then estimate κ. We show results from such an approach in Appendix Table 13, where it appears that this particular parameterization leads to an increase in the aggregate gains from reducing migration costs. We do not present these larger gains as our main results and wish to urge caution in interpretation. While this particular parameterization leads to an increase in predicted gains, there may be alternative parameterizations that could lead to a decrease in predicted gains. Because we cannot accommodate all possibilities, we simply note that the size of the gains are subject to uncertainty.

6.5.4 Human capital

As discussed in Appendix B.2, endogenous human capital acquisition is a concern for our counterfactuals. We show in the appendix that it is possible to incorporate endogenous human capital into the model; doing so transforms the key sorting parameter, θ, into $\theta(1-\eta)$, where η is the elasticity of human capital with respect to education spending. We take a pragmatic approach to addressing the concern about whether endogenous human capital acquisition would change our conclusions by computing a lower bound estimate. To do this, we calculate counterfactual average wages in each destination removing frictions *and* setting $\eta = 0$. Effectively this removes all existing education and any optimization response. We find in Appendix Table 14 that the aggregate gains of removing all frictions fall from 22% to 18% when we restrict eduction acquisition, with similar percentage drops in gains for the other counterfactual. It is important to note that, because we preclude young people who have never moved from increasing education in response to changes in frictions, this is an upper bound on the importance of endogenous human capital. Overall, the results suggest there are still substantial gains to removing migration barriers even in the presence of frictions that limit re-optimization of education.

7 Conclusion

Large spatial wage gaps and recent experimental evidence suggest there may be important productivity gains from encouraging internal migration in developing countries. We estimate the size of the aggregate gains in Indonesia. Our approach entails using movement data to identify constraints on migration, then using wage data to consider how removing these constraints would affect locational choices and wages, taking into account selection and GE effects.

We implement our approach using unique data from Indonesia, which records location of birth, current location, and current earnings. Combined with our model, this data allows for particularly transparent identification of key model parameters. In particular, we are able to identify the key distributional parameter that determines the importance of sorting from a simple linear regression of the origin-destination wage on the origin-destination migration share.

We find aggregate output gains that are small but important, on the order of 20%. These estimates hide a great deal of heterogeneity, with some more constrained areas seeing gains of over 100%. Failure to account for selection would lead to an underestimate of the gains; accounting for selection both reduces estimated gains to agglomerating workers in one location and allows for larger gains through improved sorting. We find that the latter effect dominates.

Future research could aim to deepen our understanding of the mechanisms through which migration affects productivity. Theoretical and macroeconomic research could concentrate on the dynamic effects of encouraging migration. Microeconomic experimental evidence on the extent and nature of selection among internal migrants, as well as the strength of comparative advantage effects, would also add to our understanding. Experimental research along these lines is currently taking place as part of the broad research agenda motivated by Bryan et al. (2014) and related work, including this project.

References

Adao, Rodrigo, "Worker Heterogeneity, Wage Inequality, and International Trade: Theory and Evidence from Brazil," 2016.

Ahlfeldt, Gabriel, Stephen Redding, Daniel Sturm, and Nikolaus Wolf, "The Economics of Density: Evidence from the Berlin Wall," *Econometrica*, 2015, *83* (6), 2127–2189.

Albouy, David, "Are Big Cities Bad Places to Live? Estimating Quality of Life across Metropolitan Areas," 2012.

Allen, Treb and Costas Arkolakis, "Trade and the Topography of the Spatial Economy," *Quarterly Journal of Economics*, 2014, *129* (3), 1085–1140.

Angrist, Joshua and Jorn-Steffen Pischke, *Mostly Harmless Econometrics*, Princeton University Press, 2009.

Beegle, K, Joachim de Weerdt, and Stefan Dercon, "Migration and Economic Mobility in Tanzania: Evidence from a Tracking Survey," *Review of Economics and Statistics*, 2011, *93* (3), 1010–1033.

Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum, "Plants and Productivity in International Trade," *American Economic Review*, 2003, *93* (4), 1268–1290.

Berry, Steven T, "Estimating Discrete-Choice Models of Product Differentiation," *The RAND Journal of Economics*, 1994, *25* (2), 242–262.

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak, "Under-investment in a Profitable Technology : The Case of Seasonal Migration in Bangladesh," *Econometrica*, 2014, *82* (5), 1671–1748.

Burstein, Ariel, Gordon Hanson, Lin Tian, and Jonathan Vogel, "Tradability and the Labor-Market Impact of Immigration: Theory and Evidence from the U.S.," 2017.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro, "Trade and Labor Market Dynamics: General Equilibrium Analyis of the China Trade Shock," 2017.

Carrington, W, E Detragiache, and Tara Vishwanath, "Migration with endogenous moving costs," *The American Economic Review*, 1996, *86* (4), 909–930.

Caselli, Francesco, "Accounting for Cross-Country Income Differences," in "Handbook of Economic Growth," Vol. 1, Elsevier, 2005, pp. 679–741.

Combes, Pierre-Philippe and Laurent Gobillon, "The Empirics of Agglomeration Economies," in Gilles Duranton and Will Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5 2015, pp. 247–348.

Desmet, Klaus, Dávid Krisztián Nagy, and E Rossi-Hansberg, "The Geography of Development: Evaluating Migration Restrictions and Coastal Flooding," *Journal of Political Economy (forthcoming)*, 2016, pp. 1–54.

Eaton, Jonathan and Samuel Kortum, “Technology, Geography, and Trade,” *Econometrica*, 2002, *70* (5), 1741–1779.

Fally, Thibault, “Structural Gravity and Fixed Effects,” *Journal of International Economics*, 2015, *97* (1), 76–85.

Garlick, Julia, Murrary Leibbrandt, and James Levinsohn, “Individual Migration and Household Incomes,” *NBER Working Paper*, 2016.

Gollin, Douglas, David Lagakos, and Michael Waugh, “The Agricultural Productivity Gap in Developing Countries,” *The Quarterly Journal of Economics*, 2014, *129* (2), 939–993.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti, “Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings,” *Journal of Political Economy*, 2010.

Harari, Mariaflavia, “Cities in Bad Shape: Urban Geometry in India,” 2017.

Harris, John and Michael Todaro, “Migration, Unemployment and Develpment: A Two-Sector Analysis,” *The American Economic Review*, 1970, *60* (1), 126–142.

Heston, A, R Summers, and B Aten, “Penn World Table Version 6.2,” 2006.

Hicks, Joan Hamory, Marieke Kleemans, Nicholas Li, and Edward Miguel, “Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata,” 2017.

Hsieh, Chang-Tai and Enrico Moretti, “Housing Constraints and Spatial Misallocation,” *American Economic Journal: Macroeconomics (forthcoming)*, 2018.

_ and Peter J Klenow, “Misallocation and Manufacturing TFP in China and India,” *Quarterly Journal of Economics*, 2009, *124* (4), 1403–1448.

_ , Erik Hurst, Charles Jones, and Peter J Klenow, “The Allocation of Talent and US Economic Growth,” 2018.

Kennan, J and J Walker, “The Effect of Expected Income on Individual Migration Decisions,” *Econometrica*, 2011, *79* (1), 211–251.

Kleemans, Marieke and Jeremy Magruder, “Labor markets changes in response to immigration: evidence from internal migration driven by weather shocks,” *The Economic Journal (forthcoming)*, 2017.

Lagakos, David and Michael Waugh, “Selection, Agriculture, and Cross-Country Productivity Differences,” *American Economic Review*, 2013, *103* (2), 948–980.

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

Monte, Ferdinando, Stephen Redding, and Esteban Rossi-Hansberg, "Commuting, Migration, and Local Employment Elasticities," *American Economic Review (forthcoming)*, 2018, pp. 1–57.

Morten, Melanie and Jaqueline Oliveira, "The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City," 2016.

Redding, Stephen, "Goods Trade, Factor Mobility and Welfare," *Journal of International Economics*, 2016, *101*, 148–167.

— and Esteban Rossi-Hansberg, "Quantitative Spatial Economics," *Annual Review of Economics*, 2017.

Roback, J, "Wages, Rents, and the Quality of Life," *The Journal of Political Economy*, 1982, *90* (6), 1257–1278.

Rosen, Sherwin, "Wage-Based Indexes of Urban Quality of Life," in P Mieszkowski and M Straszheim, eds., *Current Issues in Urban Economics*, 1979.

Rosenthal, Stuart S. and William C. Strange, "Evidence on the Nature and Sources of Agglomeration Economies," in "Handbook of Regional and Urban Economics," Vol. 4 2004, pp. 2119–2171.

Saiz, Albert, "The Geographic Determinants of Housing Supply," *Quarterly Journal of Economics*, 2010, *125* (3), 1253–1296.

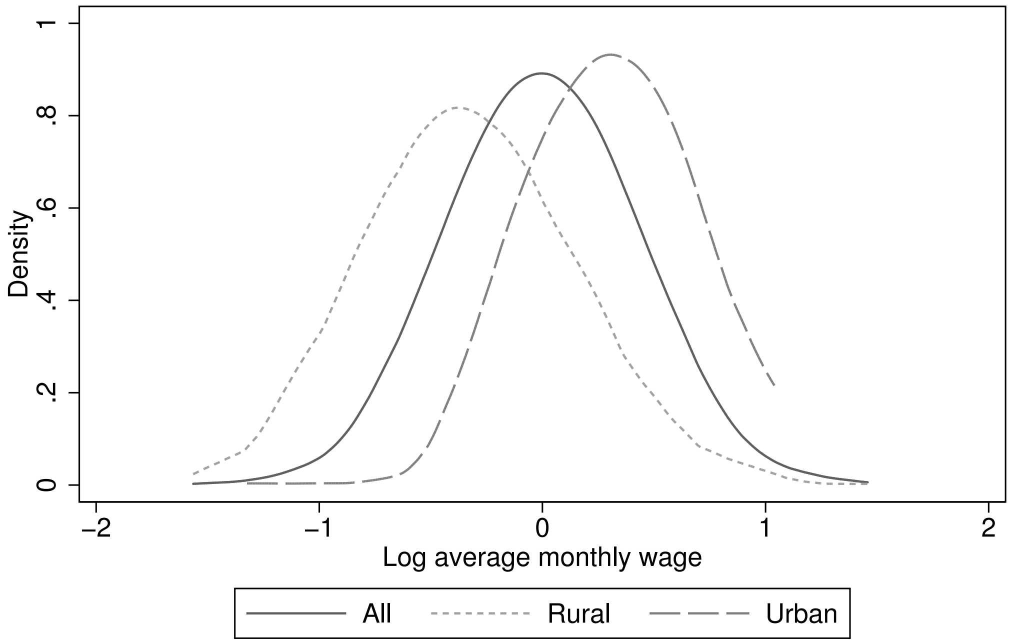
Silva, J M C Santos and Silvana Tenreyro, "The Log of Gravity," *The Review of Economics and Statistics*, 2006, *88* (November), 641–658.

The World Bank, "World Development Indicators Database," 2014.

Tombe, Trevor and Xiadong Zhu, "Trade Liberalization, Internal Migration and Regional Income Differences: Evidence from China," 2015.

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

— , "Structural Transformation, the Mismeasurement of Productivity Growth, and the Cost Disease of Services," *American Economic Review*, 2014, *104* (11), 3635–3667.



Log average monthly wage is demeaned of year fixed effects. Unit of observation is the regency.
Regency is defined as either rural or urban to match the national share of rural.
Source: 1995 SUPAS, 2011 SUSENAS; 2012 SUSENAS.

Figure 1: Distribution of wages: regency level

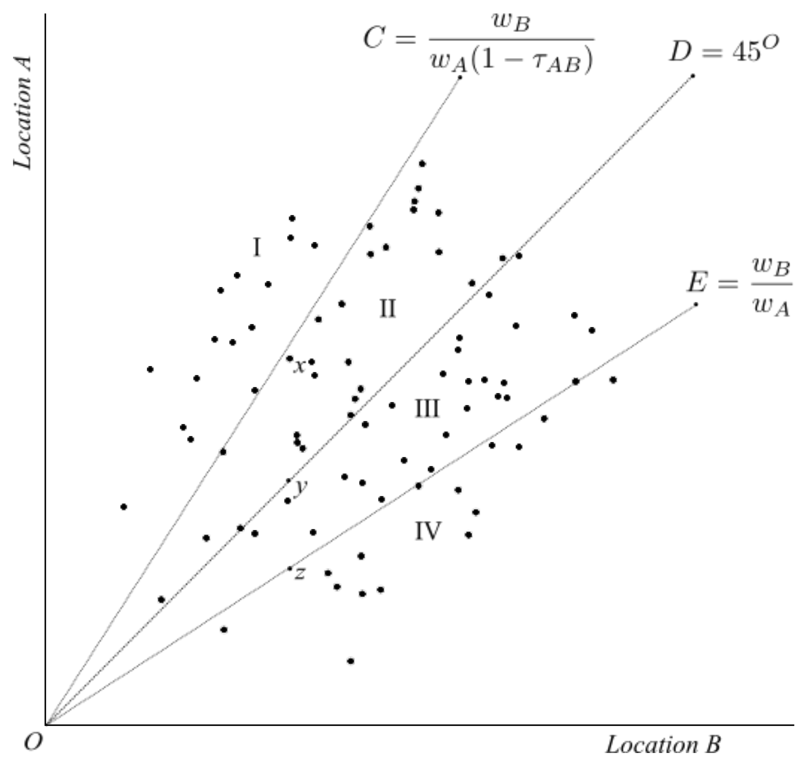
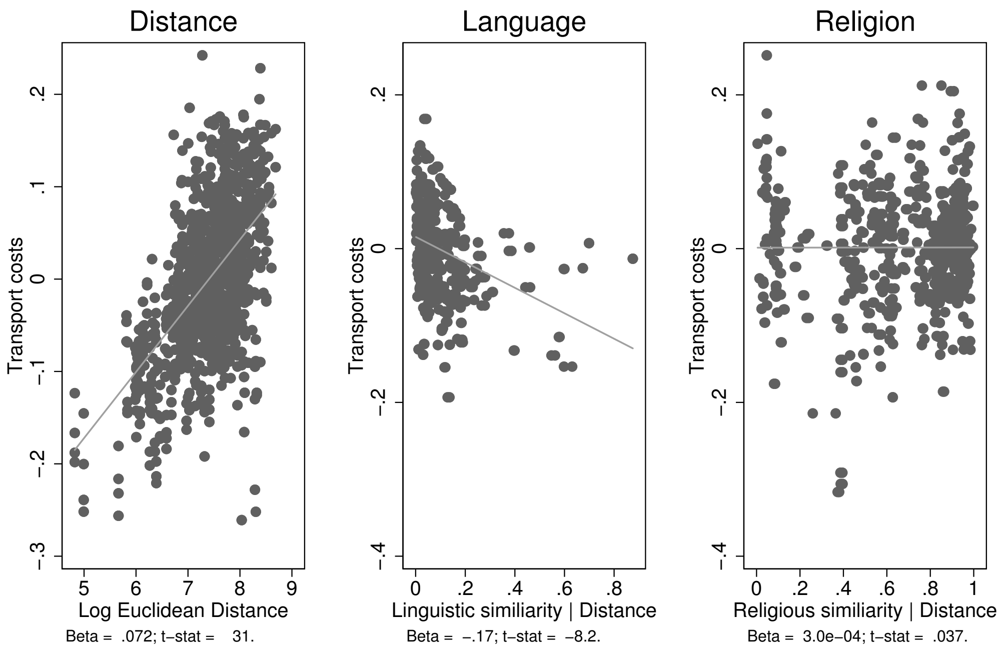


Figure 2: Productivity and Location Choices of People Born in B



Data, demeaned by year, from 1995, 2011, 2012.
Graph shows intensive margin transport costs (less than upper bound).

Figure 3: Correlates of iceberg costs in Indonesia

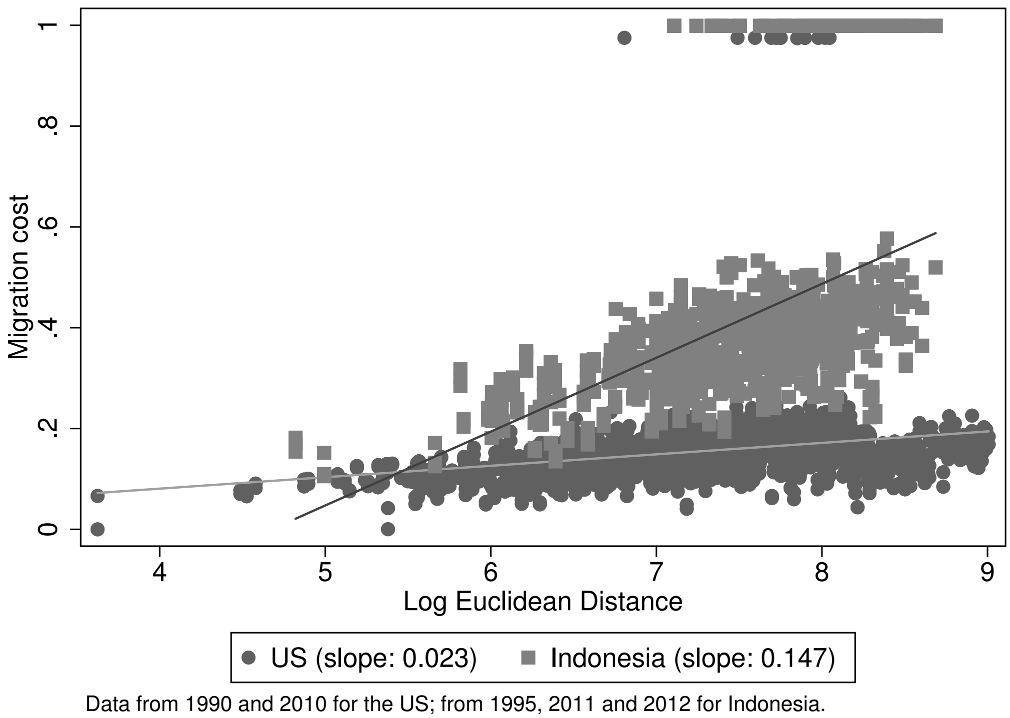


Figure 4: Relationship between iceberg costs and distance in Indonesia and the United States

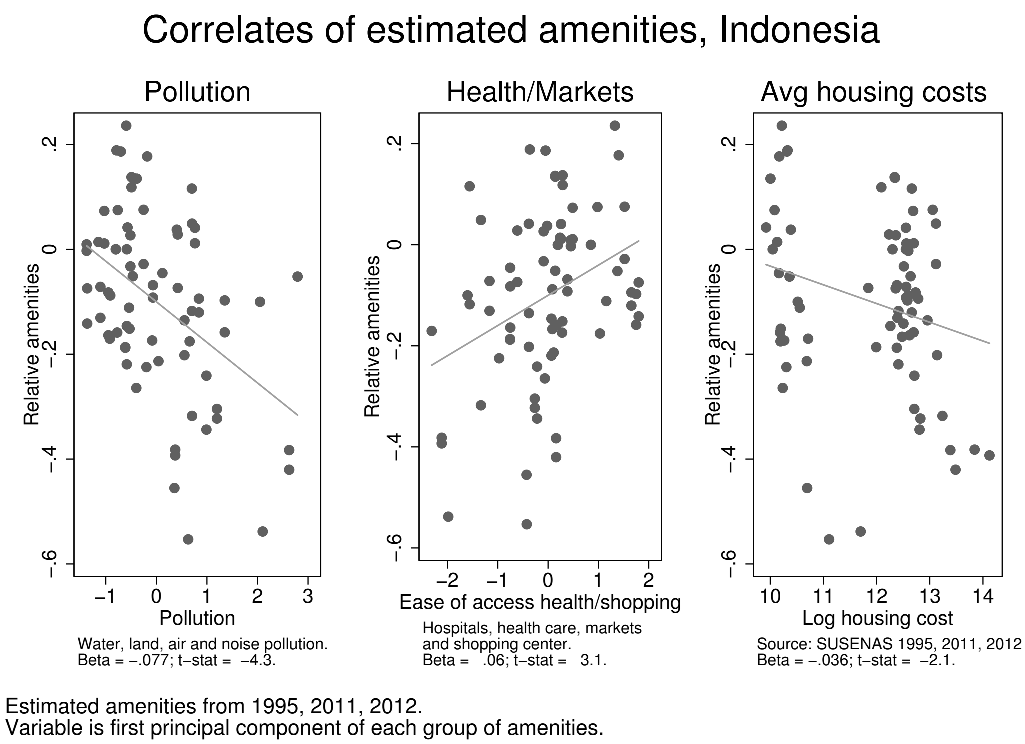
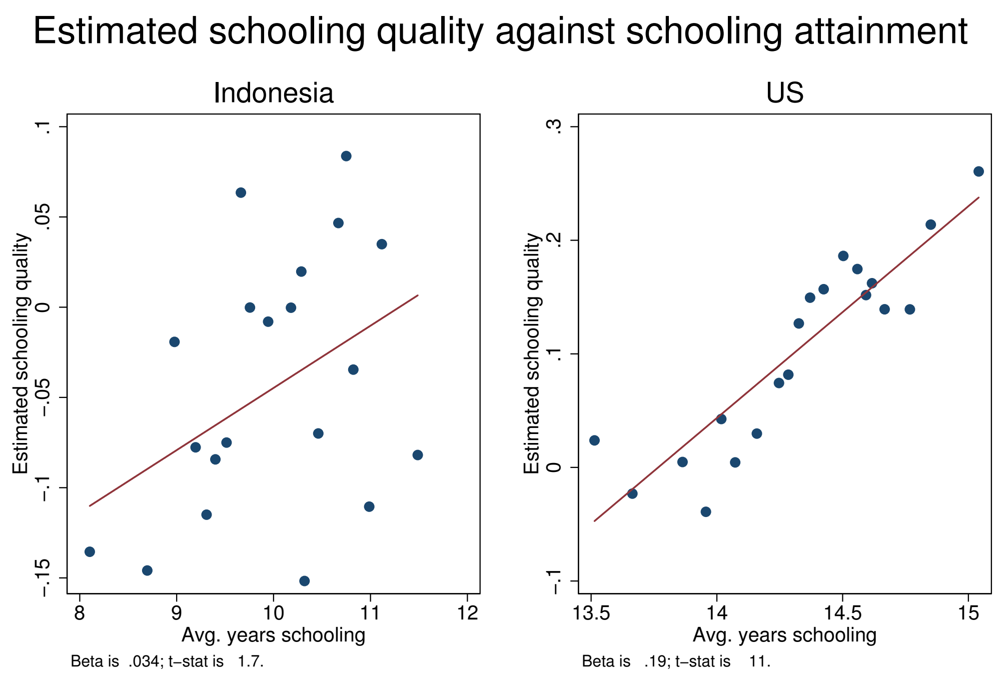
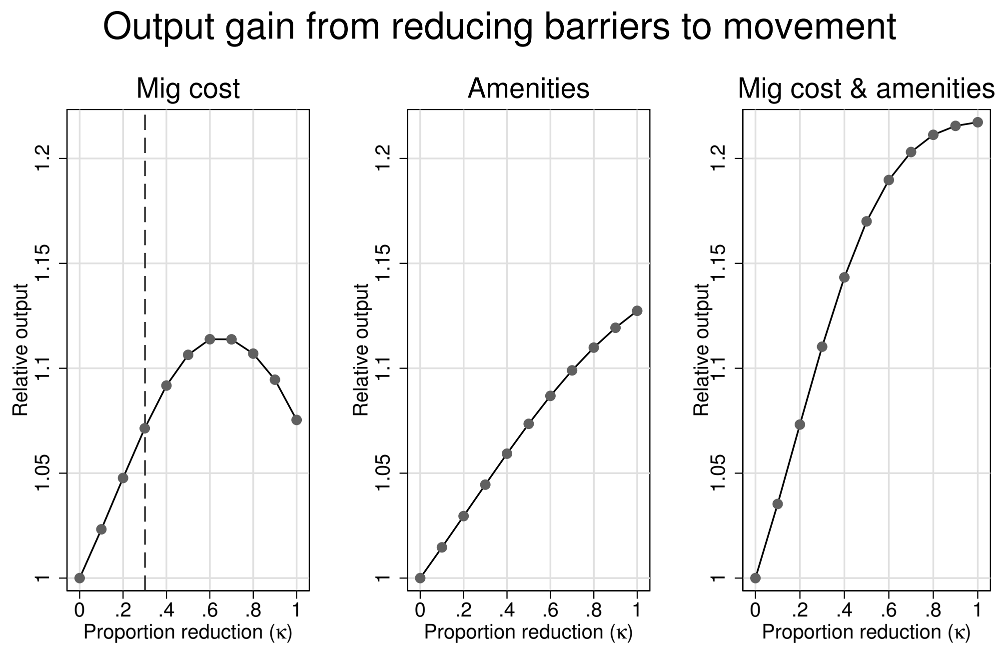


Figure 5: Estimated amenities against measured amenities



Data is: 1990, 2010 for the US. 1995, 2011, 2012 for Indonesia.
Figure shows a binned scatterplot. Year effects are removed from both graphs.

Figure 6: Schooling quality positively correlated with attainment



Data is average across 1995, 2011, 2012 for Indonesia.
The proportion reduction, κ, is defined in the text.
The red dashed line shows the US−level of migration costs.

Figure 7: Output gains from reducing barriers to movement

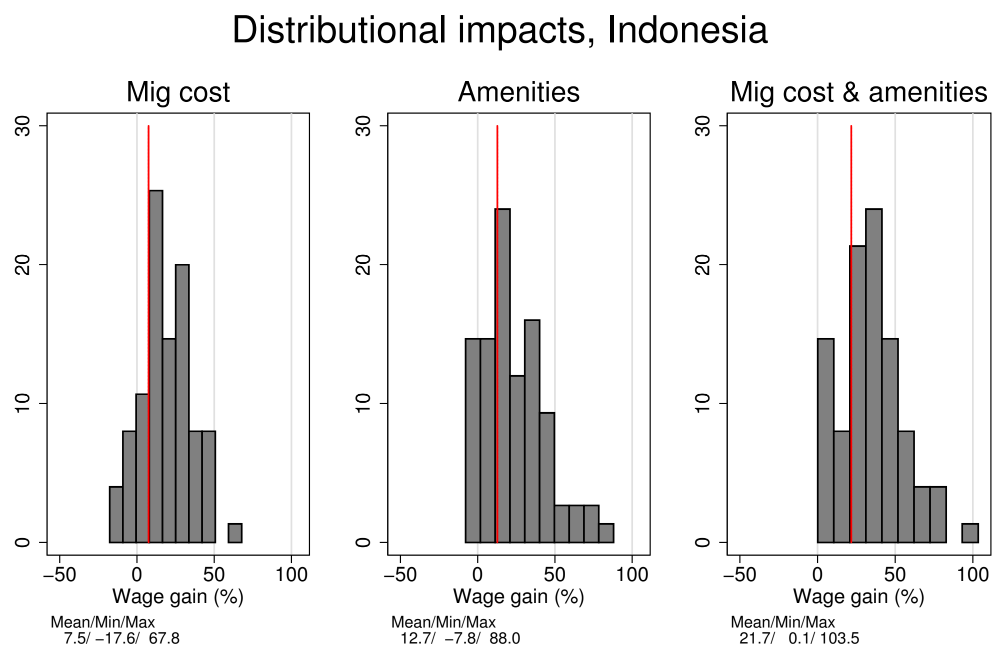


Figure shows average wage gain. The unit of observation is an origin−year.
National average (weighted by population) shown in red line.
Shows a reduction of costs of 100%. Data is 1995, 2011, 2012.

Figure 8: Distributional effects of fully reducing barriers to migration

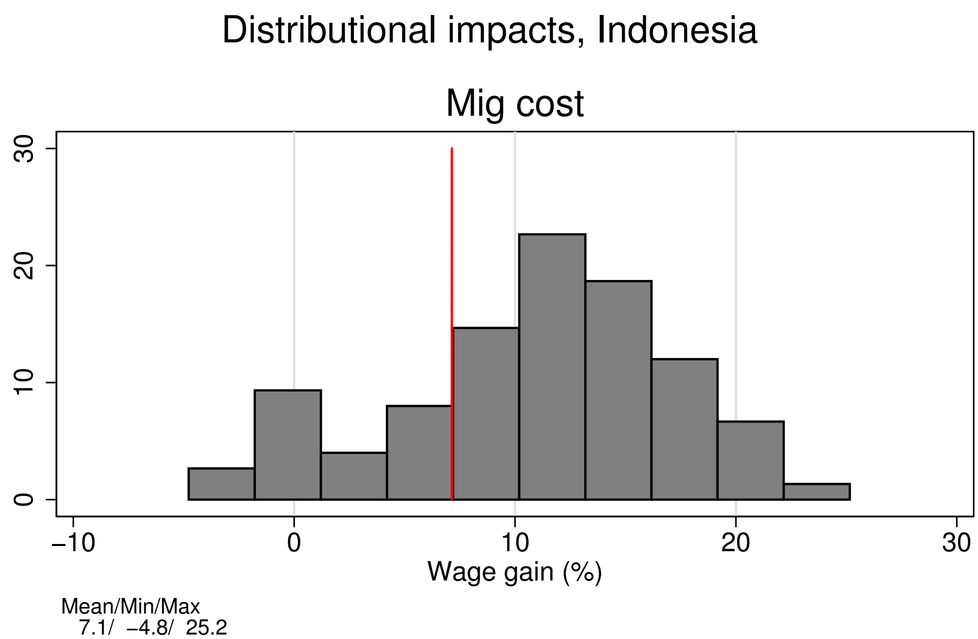


Figure shows average wage gain. The unit of observation is an origin−year.
National average (weighted by population) shown in red line.
Shows a reduction of costs of 30%. Data is 1995, 2011, 2012.

Figure 9: Distributional effects of reducing migration costs to US level

Table 1: Migration rates by origin, Indonesia

	Rural	Urban	All
1995			
Migration rate	32.3	35.8	33.7
Moves within category	31.1	74.6	49.4
2011			
Migration rate	38.7	33.7	35.7
Moves within category	24.4	84.2	58.7
2012			
Migration rate	38.9	34.1	35.8
Moves within category	25.4	83.8	60.7

Notes: Data source: 1995 Supas; 2011 Susenas; 2012 Susenas. Migration is measured as living in a regency other than the birth regency. Regencies are classified as rural or urban based on the share of their population that report being rural; we choose the cutoff to classify the regency as rural to match the national urbanization rate for each year.

Table 2: Five facts about migration in Indonesia

	Movement costs		Selection		Compensating Diff.
	(1)	(2)	(3)	(4)	(5)
Dep. variable	$\log \pi_{odt}$	$\log w_{odt}$	$\log w_{odt}$	$\log w_{odt}$	$\log w_{odt}$
Log distance	-0.717	0.029		0.007	
	(0.009)***	(0.001)***		(0.002)***	
Log share migrating			-0.039	-0.031	
			(0.001)***	(0.003)***	
Amenities					-0.023
					(0.010)**
Destination x Year FE	yes	yes	yes	yes	no
Origin x Year FE	yes	yes	yes	yes	yes
Destination FE					yes
No. of individuals	187065	186763	186763	186763	185357
No. of region pairs	25540	25244	25244	25244	25050

Notes: $\log \pi_{odt}$ is the log of the share of population migrating from o to d in year t. $\log w_{odt}$ is the log of the average wage of migrants from origin o in destination d in time t. An observation is an origin-destination regency pair. Datasource: 1995 SUSENAS, 2011 SUSENAS, 2012 SUSENAS. Amenity measure is the standardized value of the first principal component. Two-way clustering of standard errors at the origin-year and destination-year reported in Columns (1)-(4). Clustered standard errors, at the level of the origin-year, reported in Column (5). Number of observations changes between columns because not all pairs with positive migration flows have observed wages.

Table 3: Four facts about migration in the U.S.

	Movement costs		Selection	
	(1)	(2)	(3)	(4)
Dep. variable	log π_{odt}	log w_{odt}	log w_{odt}	log w_{odt}
Log distance	-0.553	0.023		-0.004
	(0.018)***	(0.002)***		(0.004)
Log share migrating			-0.043	-0.050
			(0.003)***	(0.006)***
Destination x Year FE	yes	yes	yes	yes
Origin x Year FE	yes	yes	yes	yes
Destination FE				
No. of individuals	2294054	2294046	2294046	2294046
No. of region pairs	5084	5076	5076	5076

Notes: log π_{odt} is the log of the share of population migrating from o to d in year t. log w_{odt} is the log of the average wage of migrants from origin o in destination d in time t. An observation is an origin-destination state pair. Datasource: 1990 Census, 2010 ACS. Two-way clustering of standard errors at the origin-year and destination-year reported in Columns (1)-(4). Number of observations changes between columns because not all pairs with positive migration flows have observed wages.

Table 4: Estimated Frechet parameters

	(1)	(2)
	Indonesia	U.S.
ρ (correlation)	0.74***	0.90***
	(0.029)	(0.015)
θ (dispersion)	12.5***	27.6***
	(1.36)	(3.29)
$\tilde{\theta} = \theta(1-\rho)$	3.18	2.69
Mean migration cost (iceberg)	0.39	0.15

Notes: Source: estimates from structural estimation of model. Transport costs estimated non-parametrically. Bootstrapped standard errors reported.

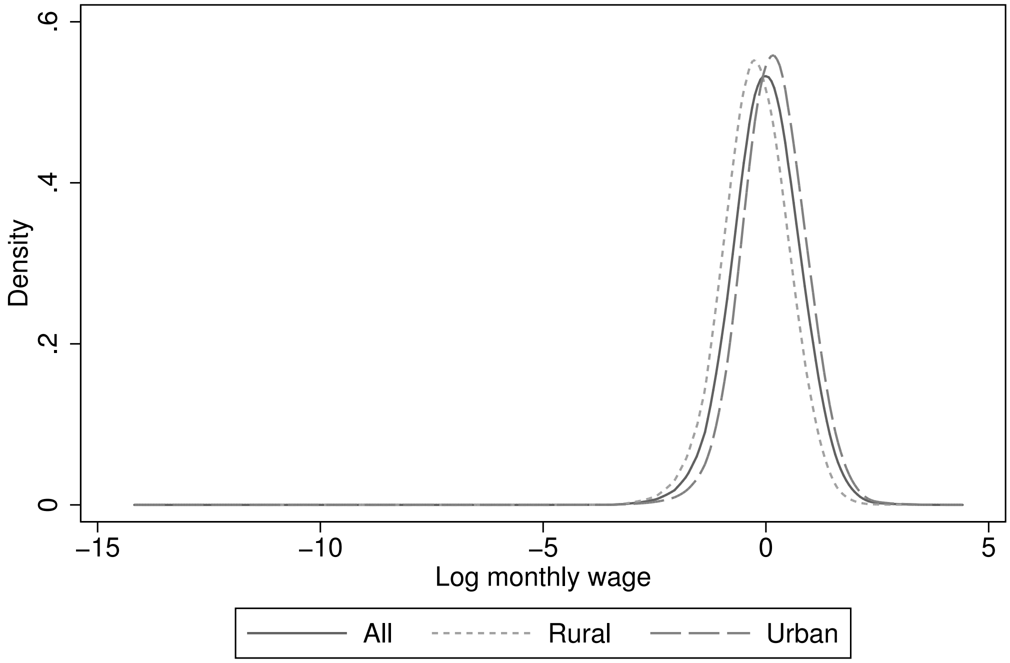
Table 5: Output gain from reducing migration barriers

	(1) Mig costs	(2) Amenities	(3) Mig costs, amenities
Baseline	1.075	1.127	1.217
No selection	0.914	1.127	1.133

Notes: Table shows the output gain from removing the barrier completely. Data is 1995, 2011, 2012 for Indonesia. No selection recalculates the output gain shutting down the role for comparative advantage.

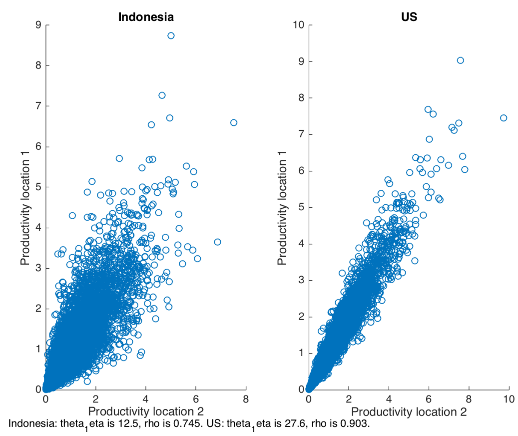
Online Appendix

A Tables and Figures

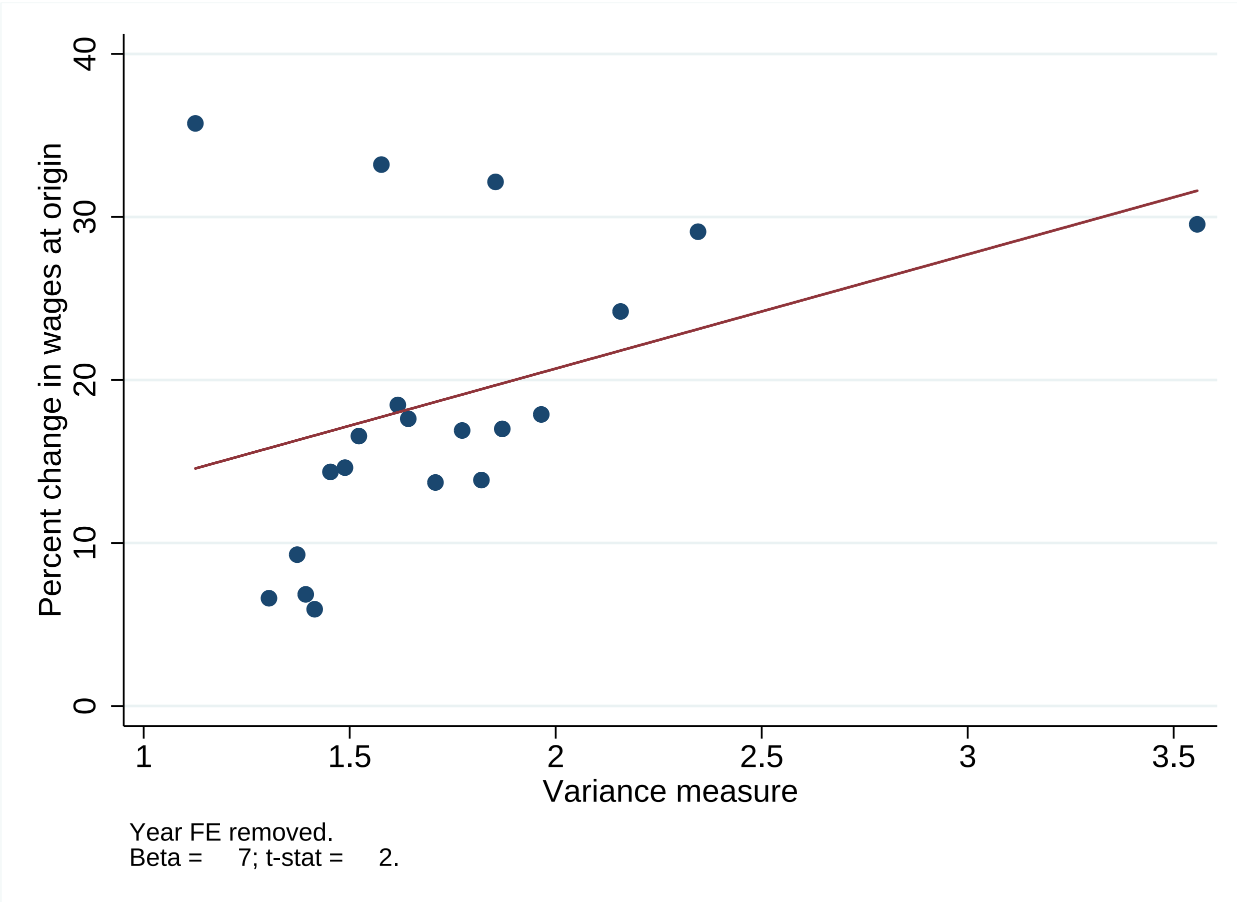


Log monthly wage is demeaned of year fixed effects. Unit of observation is the individual.
Source: 1995 SUPAS, 2011 SUSENAS; 2012 SUSENAS.

Appendix Figure 1: Distribution of wages: individual level



Appendix Figure 2: Simulated Frechet Distribution



Appendix Figure 3: Variance in origin wages

Appendix Table 1: Summary statistics for Indonesia and US sample

	Indonesia				United States
	(1)	(2)	(3)	(4)	(5)
	1995	2011	2012	1990	2010
Demographic					
Average age	38.58	39.63	39.86	40.78	44.30
Average age (mig)	38.43	39.31	39.71	41.26	44.95
Years school	8.01	9.85	10.12	13.48	15.24
Years school (mig)	10.17	11.12	11.33	14.03	15.61
High education	0.82	0.92	0.93	0.85	0.93
Financial					
Monthly non-zero wage	129.62	188.06	198.72	4424.07	4985.62
Monthly non-zero wage (mig)	174.82	236.25	246.36	4829.80	5667.14
GDP per capita	1328.48	3177.26	3337.69	39982.11	48377.39
Migration					
Share migrating	0.34	0.36	0.36	0.40	0.41
Share low educ migrate	0.15	0.19	0.19	0.31	0.30
Share high educ migrate	0.39	0.37	0.37	0.42	0.46
Share migrating (prov)	0.18	0.22	0.23		
Number of obs	59,006	68,510	69,330	1,958,123	336,033
Sum of sample weights	13,454,136	17,324,774	17,736,228	39,251,152	32,404,124

Notes: Sample is male head of household, between 15 and 65 years old. Migration defined at the regency level in Indonesia, and at the state level in the US. Low education is 3 years of schooling or less in Indonesia; 12 years or less in the US. Data source: Indonesia: 1995 SUPAS, 2011 SUSENAS, 2012 SUSENAS. US: 1990 ACS and 2010 ACS. GDP data from the World Bank Development Indicators Database. All financial values reported in 2010 USD.

Appendix Table 2: Summary statistics for IFLS sample

	(1) 1993	(2) 1997	(3) 2000	(4) 2007
Demographic				
Average age	42.27	43.04	42.40	43.41
Average age (mig)	41.98	42.87	42.05	43.40
Years school	5.18	5.83	7.27	7.93
Years school (mig)	6.42	6.92	8.69	8.74
High education	0.73	0.79	0.83	0.87
Financial				
Monthly non-zero wage	94.05	111.51	114.82	128.11
Monthly non-zero wage (mig)	125.77	138.41	142.07	155.09
GDP per capita	1181.78	1552.12	1662.29	2276.16
Migration				
Share migrating	0.32	0.35	0.39	0.47
Share low educ migrate	0.22	0.23	0.27	0.35
Share high educ migrate	0.36	0.38	0.41	0.49
Share migrating (prov)	0.14	0.16	0.24	0.23
Number of obs	5,496	5,625	7,709	9,976
Sum of sample weights	5,501	5,952	8,430	10,810

Notes: Sample is male head of household, between 15 and 65 years old. Migration defined at the regency level. Low education is 3 years of schooling or less in Indonesia. Data source: 1993, 1997, 2000 and 2007 IFLS surveys. GDP data from the World Bank Development Indicators Database. All financial values reported in 2010 USD.

Appendix Table 3: Migration histories

	Entire sample		Male HOH	
	Regency mean	Province mean	Regency mean	Province mean
Whole sample				
Current migrant	0.40	0.14	0.45	0.18
Ever migrated	0.45	0.18	0.48	0.21
People who have migrated at least once				
Migrate once	0.55	0.64	0.60	0.69
Migrate twice	0.32	0.30	0.29	0.26
Migrate three times or more	0.12	0.06	0.12	0.05
Number moves	1.61	1.44	1.56	1.38
Migrant who has returned home	0.12	0.19	0.08	0.14
First migration before age 16	0.16	0.17	0.05	0.08
Num. individuals	37612	37612	9425	9425

Notes: Data source: IFLS rounds 1-4. Entire sample is all adults in sample who report a birth location. Male HOH is people who are ever heads of households over the sample. Migration history collected for all moves after age 12.

Appendix Table 4: Five facts about migration in Indonesia, province-level

	Movement costs		Selection		Compensating Diff.
	(1)	(2)	(3)	(4)	(5)
Dep. variable	$\log \pi_{odt}$	$\log w_{odt}$	$\log w_{odt}$	$\log w_{odt}$	$\log w_{odt}$
Log distance	-0.606	0.041		0.019	
	(0.013)***	(0.003)***		(0.009)**	
Log share migrating			-0.066	-0.036	
			(0.005)***	(0.012)***	
Amenities					-0.059
					(0.074)
Destination x Year FE	yes	yes	yes	yes	no
Origin x Year FE	yes	yes	yes	yes	yes
Destination FE					yes
No. of individuals	196846	196838	196838	196838	196838
No. of region pairs	1452	1444	1444	1444	1444

Notes: $\log \pi_{odt}$ is the log of the share of population migrating from o to d in year t. $\log w_{odt}$ is the log of the average wage of migrants from origin o in destination d in time t. An observation is an origin-destination province pair. Datasource: 1995 SUSENAS, 2011 SUSENAS, 2012 SUSENAS. Amenity measure is the standardized value of the first principal component. Two-way clustering of standard errors at the origin-year and destination-year reported in Columns (1)-(4). Clustered standard errors, at the level of the origin-year, reported in Column (5). Number of observations changes between columns because not all pairs with positive migration flows have observed wages.

Appendix Table 5: Four facts about migration in Indonesia, IFLS data

	Movement costs		Selection	
	(1)	(2)	(3)	(4)
Dep. variable	$\log \pi_{odt}$	$\log w_{odt}$	$\log w_{odt}$	$\log w_{odt}$
All individuals				
Log distance	-0.571	0.041		-0.044
	(0.018)***	(0.006)***		(0.017)*
Log share migrating			-0.079	-0.150
			(0.010)***	(0.028)***
Destination x Year FE	yes	yes	yes	yes
Origin x Year FE	yes	yes	yes	yes
No. of individuals	26820	26732	26732	26732
No. of region pairs	613	546	546	546
Wage employees only				
Log distance	-0.527	0.026		0.000
	(0.015)***	(0.004)***		(0.013)
Log share migrating			-0.049	-0.049
			(0.008)***	(0.024)
Destination x Year FE	yes	yes	yes	yes
Origin x Year FE	yes	yes	yes	yes
No. of individuals	12241	12223	12223	12223
No. of region pairs	497	479	479	479

Notes: $\log \pi_{odt}$ is the log of the share of population migrating from o to d in year t. $\log w_{odt}$ is the log of the average wage of migrants from origin o in destination d in time t. An observation is an origin-destination province pair. Datasource: 1993, 1997, 2000, 2007 IFLS. Two-way clustering of standard errors at the origin-year and destination-year reported in Columns (1)-(4). Number of observations changes between columns because not all pairs with positive migration flows have observed wages.

Appendix Table 6: Correlation of estimated amenities with data

	(1) 1995 b/se	(2) 2011 b/se	(3) 2012 b/se
Water pollution (past year)	-2.23***	-0.96**	-0.91**
	(0.86)	(0.42)	(0.38)
Land pollution (past year)	1.57	-5.15***	-4.91***
	(2.73)	(1.71)	(1.52)
Air pollution (past year)	-0.77***	0.16	-0.24
	(0.28)	(0.55)	(0.50)
Noise pollution (past year)	-2.46		
	(1.98)		
Main road village lighting	0.39	0.68*	0.68**
	(0.71)	(0.38)	(0.34)
Has movie theater	-10.6	-33.9	-77.1**
	(7.35)	(41.0)	(34.1)
Ease of reaching hospital	0.31**	0.10	0.27***
	(0.13)	(0.13)	(0.10)
Ease of reaching puskesmas/other health facility	0.44*	0.057	0.34**
	(0.26)	(0.20)	(0.17)
Ease of reaching market with permanent building	0.31**		
	(0.15)		
Ease of reaching shopping complex	0.37***		
	(0.13)		
Flooding		-0.46	-0.076
		(0.49)	(0.45)
Earthquake		-0.039	-0.073
		(0.21)	(0.19)
Whirlwind/tornado/hurricane		0.61	-0.20
		(0.41)	(0.38)
Drought		-0.52	0.53
		(1.20)	(1.09)
Outbreak (last year): Vomiting/diarrhea		-0.85	-0.81
		(0.72)	(0.65)
Outbreak (last year): Malaria		0.48	0.74**
		(0.42)	(0.36)
Outbreak (last year): Bird flu (1 case is considered an outbreak)		-10.2	-10.6
		(8.64)	(7.76)
Outbreak (last year): Tuberculosis		-0.59	-1.72*
		(1.13)	(0.97)

Notes: Data source: 1996 and 2011 PODES data and estimates from model. Table shows the regression coefficient between the estimated amenity value and the amenity measure given in each row. 1996 PODES data are correlated with the model estimates for 1995; 2011 PODES data are correlated with model estimates for 2011 and 2012.

Appendix Table 7: Estimating spillover parameters

	Log wage		Log amenity		Log rents	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	OLS	IV	OLS	IV
Log population	0.028	0.012	0.004	0.042	0.228	0.252
	(0.083)	(0.210)	(0.085)	(0.215)	(0.098)**	(0.247)
N	75	75	75	75	75	75
r2	0.89	0.89	0.76	0.76	0.99	0.99
F	0.11	0.00	0.00	0.04	5.38	1.04

Notes: Table shows OLS and 2SLS estimates using the model-based instrument for labor to estimate the strength of the agglomeration and congestion parameters. Province and year FE included in all specifications.

Appendix Table 8: Robustness: effect of migration cost on growth, Indonesia

	(1) Sub. elasticity = 4	(2) Sub. elasticity= 6	(3) Sub. elasticity = 8
Productivity spillover = 0			
$\lambda = 0$	1.113	1.107	1.098
$\lambda = .02$	1.112	1.102	1.086
$\lambda = -.02$	1.114	1.110	1.104
Productivity spillover = 0.05			
$\lambda = 0$	1.117	1.103	1.075
$\lambda = .02$	1.115	1.090	1.032
$\lambda = -.02$	1.118	1.110	1.096
Productivity spillover = 0.08			
$\lambda = 0$	1.119	1.094	1.033
$\lambda = .02$	1.115	1.069	0.929
$\lambda = -.02$	1.121	1.107	1.078

Notes: Table shows the effect of reducing migration cost to zero on labor output. Table shows different combinations of amenity (λ) and productivity spillovers, for different values of substitution parameter. Calculated for model with selection.

Appendix Table 9: Robustness: effect of migration cost and housing on growth, Indonesia

	(1) Sub. elasticity = 4	(2) Sub. elasticity= 6	(3) Sub. elasticity = 8
Productivity spillover = 0			
$\lambda = 0$	1.115	1.114	1.112
$\lambda = .02$	1.115	1.113	1.111
$\lambda = -.02$	1.115	1.114	1.113
Productivity spillover = 0.05			
$\lambda = 0$	1.121	1.118	1.116
$\lambda = .02$	1.120	1.117	1.112
$\lambda = -.02$	1.121	1.119	1.118
Productivity spillover = 0.08			
$\lambda = 0$	1.124	1.121	1.117
$\lambda = .02$	1.123	1.118	1.111
$\lambda = -.02$	1.124	1.122	1.120

Notes: Table shows the effect of reducing migration cost and housing to zero on labor output. Table shows different combinations of amenity (λ) and productivity spillovers, for different values of substitution parameter. Calculated for model with selection.

Appendix Table 10: Robustness: effect of amenities on growth, Indonesia

	(1) Sub. elasticity = 4	(2) Sub. elasticity= 6	(3) Sub. elasticity = 8
Productivity spillover = 0			
$\lambda = 0$	1.063	1.083	1.098
$\lambda = .02$	1.066	1.087	1.103
$\lambda = -.02$	1.059	1.079	1.093
Productivity spillover = 0.05			
$\lambda = 0$	1.074	1.103	1.127
$\lambda = .02$	1.078	1.108	1.133
$\lambda = -.02$	1.070	1.098	1.121
Productivity spillover = 0.08			
$\lambda = 0$	1.082	1.119	1.152
$\lambda = .02$	1.086	1.125	1.158
$\lambda = -.02$	1.078	1.114	1.145

Notes: Table shows the effect of reducing amenities to zero on labor output. Table shows different combinations of amenity (λ) and productivity spillovers, for different values of substitution parameter. Calculated for model with selection.

Appendix Table 11: Robustness: effect of migration cost and amenities on growth, Indonesia

	(1) Sub. elasticity = 4	(2) Sub. elasticity= 6	(3) Sub. elasticity = 8
Productivity spillover = 0			
$\lambda = 0$	1.163	1.177	1.189
$\lambda = .02$	1.166	1.182	1.193
$\lambda = -.02$	1.159	1.173	1.184
Productivity spillover = 0.05			
$\lambda = 0$	1.177	1.199	1.217
$\lambda = .02$	1.182	1.204	1.223
$\lambda = -.02$	1.173	1.194	1.211
Productivity spillover = 0.08			
$\lambda = 0$	1.187	1.215	1.240
$\lambda = .02$	1.192	1.221	1.247
$\lambda = -.02$	1.183	1.209	1.233

Notes: Table shows the effect of reducing migration cost and amenities to zero on labor output. Table shows different combinations of amenity (λ) and productivity spillovers, for different values of substitution parameter. Calculated for model with selection.

Appendix Table 12: Output gain from reducing migration barriers: robustness over theta

	(1) Mig costs	(2) Amenities	(3) Mig costs, amenities
Estimated θ			
Baseline	1.075	1.127	1.217
No selection	0.914	1.127	1.133
Set $\hat{\theta} = 0.5\theta$			
Baseline	1.115	1.120	1.231
No selection	0.923	1.120	1.127

Notes: Table shows the output gain from removing the barrier completely. Data is 1995, 2011, 2012 for Indonesia. No selection recalculates the output gain shutting down the role for comparative advantage.

Appendix Table 13: Output gain from reducing migration barriers: robustness

	(1) Mig costs	(2) Amenities	(3) Mig costs, amenities
Baseline			
Baseline	1.075	1.127	1.217
No selection	0.914	1.127	1.133
Migration costs parametric fn. of distance			
Baseline	1.094	1.122	1.219
No selection	0.935	1.112	1.131
Migration costs parametric fn. of distance, lg pop interaction			
Baseline	1.170	1.146	1.303
No selection	0.949	1.108	1.156

Notes: Table shows the output gain from removing the barrier completely. Data is 1995, 2011, 2012 for Indonesia. No selection recalculates the output gain shutting down the role for comparative advantage.

Appendix Table 14: Output gain from reducing migration barriers: robustness over education

	(1) Mig costs	(2) Amenities	(3) Mig costs, amenities
Baseline (education fully adjusts)			
Baseline	1.075	1.127	1.217
No selection	0.914	1.127	1.133
Not allowing education to adjust			
Baseline	1.045	1.095	1.183
No selection	0.919	1.128	1.137

Notes: Table shows the output gain from removing the barrier completely. Data is 1995, 2011, 2012 for Indonesia. No selection recalculates the output gain shutting down the role for comparative advantage.

B Model Extensions and Interpretation

This section discusses several extensions to the model: dynamics; endogenous human capital; non-traded goods; and costly trade.

B.1 Dynamics

We aim to understand whether improving the static allocation of workers would lead to increased productivity. Consequently, we rule out potential dynamic gains and leave quantifying these for future work. Some dynamic issues, however, could be a source of bias in our estimates, and we discuss those in this section.

To understand how to map our model to a dynamic setting, suppose that time is discrete and indexed by t. Our preferred interpretation is as follows. In each period t, every individual observes current wages, movement costs, and amenities, and decides where to live for the next period. She moves, and her labor for period t is consumed. The same process occurs again in period $t+1$, and on forever. If the parameters of the model are fairly stable, most people will move only one time in their working life. This interpretation can be linked to static trade models if one considers the migration decision as the choice to move to the destination for one period of work.

Given this interpretation, our estimates may be biased if individuals do not maximize their static amenity-adjusted wages, net of movement costs. One possibility is the channel emphasized in Caliendo et al. (2017). They model workers who face a cost of moving away from their current location and therefore chose the destination to maximize a combination of static values and the option value of future moves. Dynamic investment in human capital raises a separate but similar concern. People may migrate to a specific location to invest in human capital and be paid temporarily below their market wage (e.g., going to grad school). Alternatively, some dynamic choices may lead to temporary migration for high short-term wages (e.g., an unpleasant seasonal fishing job in Alaska to raise money for grad school). If choices of these types are prevalent, we will record the incorrect wages for these individuals, potentially biasing our estimates.

Based on the IFLS migration histories, discussed in Section 2.1 and Appendix Table 3 we concluded that migration in Indonesia is best characterized by one migration event, made in adulthood, and hence that these dynamic concerns are unlikely to be a substantial source of bias.

B.2 Endogenous Human Capital

Our base model treats education in a particularly simple way. All people from a particular origin receive the same education q_o, this level of education is exogenous and independent of movement frictions, and education is equally valuable in all destinations. This set of assumptions may bias our estimates of important parameters (for example, θ the selection parameter) and/or affect our counterfactuals. We discuss each possibility in turn.

One way to understand biases is to experiment with different ways of incorporating education. One approach is to assume that

$$h_{id} = s_{id} q_o e^{\eta},$$

where e is the chosen level of education and η measures the elasticity of human capital (and hence wage) to a change in education. If we combine this production function with the strong assumption that workers make their human capital decisions with the aim of maximizing wages, rather than utility, solving

$$\max_e \left(w_d e^w_{do} s_{id} q_o e^{\eta} - \eta e \right),$$

then equation (6) becomes

$$\overline{wage}_{do} = w_d^{\frac{1}{1-\eta}} \epsilon^w_{do} q_o^{\frac{1}{1-\eta}} \pi_{do}^{-\frac{1}{\theta(1-\eta)}} \bar{\Gamma}.$$

This way of modelling education preserves the key identification requirement (discussed above) that π_{do} is sufficient to summarize the impact of amenities and movement costs on average wages. Hence, identification of the model proceeds as as in the main text, with slightly different interpretations of the parameters. For example, suppose that differences in movement costs lead to $\pi_{do} > \pi_{d'o}$. In the extended model with endogenous human capital, this will lead to $\overline{wage}_{do} < \overline{wage}_{d'o}$ for two reasons. First, as in the baseline model, those that move to d' are more selected. Second, because those who move to d' are more selected, they are also more educated. That is, education creates a complementarity which increases the impact of selection. Below we concentrate our efforts on estimating the elasticity of average wage to π_{do}, if education is not endogenous (as in the base model) this elasticity captures only the selection effect, if education is endogenous it captures a composite effect. That is, endogenous human capital modeled in this way does not create bias, but does change interpretation.

If we relax the assumption that education is chosen to maximize wage, assuming instead that workers try to maximize utility, we no longer maintain the structure that τ_{do} and α_d enter the wage equation only through π_{do}. This would mean that our method of identification is no longer valid. Fortunately, we have already presented some evidence that π_{do} is sufficient. Fact 4 above shows that, after controlling for π_{do}, distance (our proxy for movement costs) no longer correlates with $\overline{wage}_{do}$. Given this evidence, we conclude that endogenous education is unlikely to be a major source of bias in our estimates, noting only that the interpretation of the estimated θ is different if education choice is added to the model.

The model incorporating endogenous education also highlights a concern for our counterfactuals. If dynamics are thought of as discussed in Section B.1, then it becomes clear that we are assuming that all individuals, even the elderly, are able to optimize their human capital choices in response to changes in movement costs. In the model presented above, human capital investment will increase as movement costs decrease because workers are better able to match to their comparative advantage and human capital is complementary to skill. This issue is difficult to deal with while retaining the structure of the model. We address this issue quantitatively in a discussion in Section 6.5, using the model of endogenous human capital presented above. In this section, we set the value of η to zero and recompute the model. To do this, we need an initial estimate of η. We use the approach in Hsieh et al. (2018) and calculate this by dividing the share of GDP spent on education in Indonesia, which is 2.4% (The World Bank, 2014) by the share of labor income in GDP, 0.45 (Heston et al., 2006). This yields $\eta = 0.05$.[27]

B.3 Non-traded Goods

Prices of nontradeables, particularly housing, are likely to be an endogenous function of population. In our baseline model, nontradeable prices will be captured in the amenity term, α_d, and λ, which measures the extent to which amenities change with population. When we discuss calibration of λ below, we will be particularly concerned about endogenous house prices, so we briefly outline here how α_d can be decomposed into rents and other amenities.

Following Monte et al. (2018), let utility be a Cobb-Douglas function of the consumption aggregate and housing, with β_h measuring the share of expenditure on housing. The indirect utility function (3) then becomes

$$U_{ido} = \frac{\tilde{\alpha}_d}{r_d^{\beta_h}} \epsilon_{do}^{\alpha} (1 - \tau_{do}) w_d \epsilon_{do}^{w} s_{id} q_o \equiv \bar{w}_{do} s_{id}$$

where r_d is the price of housing in location d and $\tilde{\alpha}_d$ is a measure of all other amenities in that location. When we estimate amenities below, we will recover $\log \alpha_d = \log \tilde{\alpha}_d - \beta_h \log r_d$. The endogenous effect of population on α_d can also be decomposed. If $\tilde{\alpha}_d = \overline{\alpha}_d \hat{L}_d^{\lambda_a}$ and $r_d = \hat{L}_d^{\lambda_r}$ then $\alpha_d = \overline{\alpha}_d \hat{L}_d^{\lambda_\alpha - \beta_h \lambda_r}$. We separately calibrate λ_α and λ_r when undertaking counterfactuals in Section 6.

B.4 Costly Trade

Goods trade within Indonesia is surely costly, and incorporating trade costs could change our results. The main source of concern is not with estimation – differences in consumer goods prices caused by costly trade will be captured in the amenity term – but with counterfactuals, which may change if endogenous changes in the price of traded goods were accounted for.

There are two main sources of interaction. First, many policies that decrease movement costs would also decrease trade costs. For example, road widening would likely decrease the cost of moving goods and also decrease the amount of time required to visit relatives. We specifically *do not* want to capture this interaction between trade and migration costs. Our purpose in writing the paper is to understand whether migration frictions are an independent cause of low productivity.

Second, changes in movement costs will interact with (trade) market access. Burstein et al. (2017) discusses one example. They show that an industry's ability to accept new

[27] Hsieh et al. (2018) find a value of 0.1 for the U.S. This comes from an expenditure share on education of 6.6%, divided by the share of labor income in GDP of 0.64.

labor without price (or wage) decreases depends on trade costs. Incorporating this channel would add an additional dimension of heterogeneity across destinations. While we do not model this channel, σ (the elasticity of substitution) limits the extent to which one destination can accept migrants, and acts in a similar manner. A second interaction occurs if reducing migration frictions leads to an increase in productivity of one destination d, with a knock-on effect in other destinations that have greater access to d, which now produces lower priced goods. We note that our counterfactuals should be interpreted with this caveat.

C IV Approach and Monte Carlo Simulation

This appendix illustrates the IV approach discussed in Section 4. The estimating equation, equation (7), is:

$$\ln(\overline{wage}_{do}) = \underbrace{\ln(\bar{\Gamma}) + \ln(w_d)}_{\text{Destination fixed effect}} - \frac{1}{\theta}\ln(\pi_{do}) + \underbrace{\ln(q_o)}_{\text{Origin fixed effect}} + \ln \epsilon^w_{do}, \tag{10}$$

where the migration rate, π_{do}, is defined in equation (4) as:

$$\pi_{do} = \frac{w_d \epsilon^w_{do} \alpha_d \epsilon^\alpha_{do}(1-\tau_{do})}{\sum_{j=1}^J w_j \epsilon^w_{jo} \alpha_j \epsilon^\alpha_{jo}(1-\tau_{jo})}.$$

There is an endogeneity problem because the shock, ϵ^w_{do} appears in both the migration rate and the wage. As a result, OLS estimates are biased, with the implication that θ will be overestimated.

We propose an IV strategy that uses the common utility of the destination, estimated using shares from all other origins o', to provide a "pull" factor for the destination. This approach has links to the IV estimation of demand systems in IO, such as that in Berry (1994). We do the following three steps:

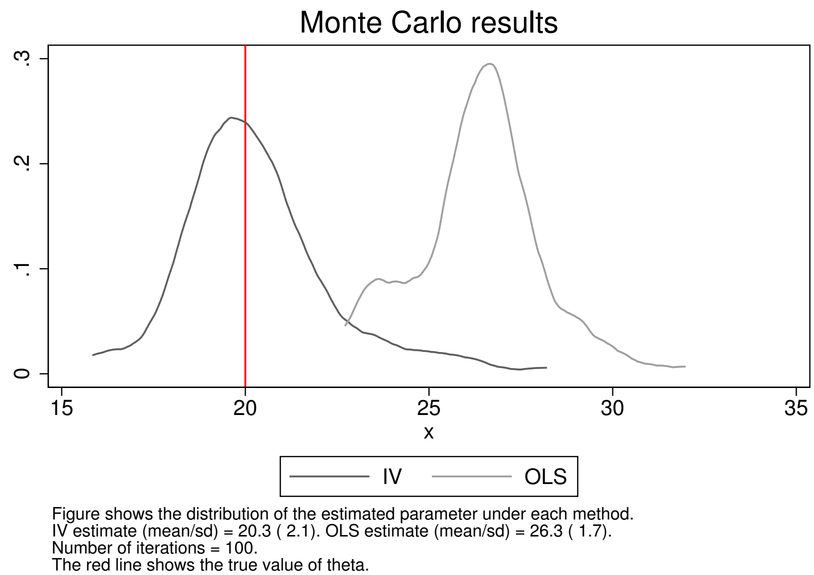
1. We construct a vector z^1_{od} consisting of all the flows and squared flows to a destination (dropping the dependent variable flows π_{od}): $\{\log \pi_{1d}, ..., \log \pi_{jd}, \log \pi^2_{1d}, ..., \log \pi^2_{jd}\}_{\forall j \neq o}$. We then run a "zero stage" regression and predict the fitted values $\widehat{\log \pi}_{od} = \hat{\beta} z^1_{od}$ from a linear regression.

2. We then run the first stage regression, where we regress $\log \pi_{ij}$ on $\widehat{\log \pi_{ij}}$ and a full set of origin-year and destination-year fixed effects.

3. We then run the second stage regression, where we regress $\log \text{wage}_{ij}$ on $\log \pi_{ij}$, instrumenting for $\log \pi_{ij}$ with $\widehat{\log \pi_{ij}}$ and a full set of origin-year and destination-year fixed effects, and then controlling for the full set of origin-year and destination-year FE.

This approach follows Angrist and Pischke (2009).

Appendix Figure 4 shows the results of a simulation of our model. As predicted, the OLS estimate of θ is upward bias. The IV estimator corrects this bias and is not statistically different than the true value. Appendix Table 16 shows that our IV estimate of θ is robust to several different definitions of the shocks used in the construction of the instrument.



Appendix Figure 4: Monte Carlo Simulations

Appendix Table 15: First stage regression for migration flows

Dep var: Log prob	(1) Baseline	(2) Drop large	(3) Drop close
Predicted log probability	9.151	6.340	7.140
	(0.241)***	(0.232)***	(0.244)***
Destination X Year FE	yes	yes	yes
Origin X Year FE	yes	yes	yes
N	1452	1452	1452
r2	0.75	0.66	0.68
F	1446.16	748.63	858.38

Notes: Table shows first-stage regression of migration probability on predicted probability. The unit of observation is a province-to-province-year flow. Column (2) drops the largest 5% of flows when constructing the instrument. Column (3) drops observations within 500 km of a location when constructing the instrument.

Appendix Table 16: Robustness: alternative instruments

	(1) Baseline	(2) Dropping large regions	(3) Dropping close regions
θ (dispersion)	12.2***	14.1***	13.5***
	(1.26)	(2.00)	(1.81)

Notes: Table shows the estimate of theta from the baseline IV specification in Colunmn (1). Column (2) drops the top 5% of observations when constructing the instrument. Column (3) drops locations closer than 500km when constructing the instrument. Bootstrapped standard errors reported.

D Can the Data Reject our Distributional Assumptions?

Our assumption of a multivariate Fréchet (Gumbel copula with Fréchet marginals) has important advantages. In particular, the distribution is closed under maximization, and so we get closed form solutions and transparent identification of key parameters. We consider this transparent identification one of the key contributions of the paper. Other copulas, such as the Frank copula considered by Lagakos and Waugh (2013), are not closed under maximization and therefore do not lead to closed form solution.

Nevertheless, it is important to explore whether the data supports the strong structure imposed by our distributional choice. The multivariate Fréchet implies two important assumptions. First, absolute and comparative advantage are aligned.[28] This implies that removing barriers to migration will reduce the difference between average wages of people born in the same origin, and living in different destinations. Second, the Fréchet distribution, combined with the assumption that the distributional parameter θ does not differ by origin, implies that average wages are a constant elasticity function of the proportion of people who migrate. This in turn implies that removing all barriers to migration will lead to complete convergence of wages within origin, with only the origin fixed effect q_o leading to variation across origins (and will lead to complete convergence in wages per unit of human capital across origins). Hence, selection alone cannot generate average wage differences in our model. This is in strong distinction to the model presented in Young (2013).

We first look for evidence supporting the assumption that comparative and absolute advantage are aligned. Young (2014) argues that if, on average, the elasticity of the average wage with respect to the proportion of the population in that sector is negative, then comparative advantage is on average aligned with absolute advantage. This is shown in Table 2: the elasticity of average wage to proportion migrating is negative, hence θ (the negative inverse of the elasticity) is positive. As pointed out by Young (2014), it is probably inappropriate to disaggregate these measures too much, Nevertheless, we present results broken down by destination and origin in Appendix Table 17. These results show that, although the estimates are very noisy, in all but 5 cases in Column (2) (6 cases for Column (3)), the elasticity is positive. So we feel comfortable concluding that comparative advantage is aligned with absolute advantage in the data.

The second assumption is that the elasticity of wages to population shares is constant across all locations. The above exercise implies that we cannot reject that the shape parameters are the same, although the test has little power. A second implication is that, in the absence of amenity differences, the wage gap should be equal to zero at when movement costs are zero. One way to empirically test this is to look at whether wage differences drop to zero as distance (a proxy for movement costs) drops to zero. Appendix Table 18 below shows a regression of the log wage gap on the log distance. (Note that we do not include amenities here because they will on average cancel out due to symmetry.)

[28] To see this, note that the average skill, conditional on selecting into location d, is given by $E(\epsilon_d|\text{choose } d) = \Gamma\pi_{od}^{\frac{-1}{\theta}}$. The unconditional mean of the Frechet distribution is given by $E(\epsilon_d) = \Gamma$. Therefore, $\frac{E(\epsilon_d|\text{choose } d)}{E(\epsilon_d)} = \pi_{od}^{\frac{-1}{\theta}}$ which is always greater than 1, because $\theta \geq 2$ in order for the variance of the Frechet distribution to be defined.

The intercept is not statistically significant, although imprecisely estimated. That is, we see no evidence of the presence of a wage gap at a hypothetical zero distance.

Taken together, we do not find any empirical evidence to suggest that our empirical specification fails to match the data. There is no strong evidence to suggest that the distributional assumption we make would lead to incorrect inference on either the amenities or movement costs.

Appendix Table 17: Robustness: does theta vary by destination or origin?

	(1) Pooled	(2) Destination-specific	(3) Origin-specific
θ (dispersion)	12.2***		
	(1.26)		
θ, region 1		20.8	5.48
		(96.8)	(141.9)
θ, region 2		17.7	6.11
		(71.2)	(25.6)
θ, region 3		-103.7	6.23
		(981.1)	(49.4)
θ, region 4		-5.31	25.8
		(83.3)	(67.4)
θ, region 5		22.3	11.9
		(72.1)	(78.8)
θ, region 6		8.18	-7.15
		(12.6)	(80.2)
θ, region 7		-2.82	4.85
		(83.2)	(120.2)
θ, region 8		45.6	2.02
		(375.8)	(18.9)
θ, region 9		16.4	8.90
		(107.5)	(24.0)
θ, region 10		-21.6	9.41
		(134.2)	(14.0)
θ, region 11		7.37	46.9
		(26.2)	(338.2)
θ, region 12		9.62	0.023
		(18.5)	(65.6)
θ, region 13		17.4	-34537.4
		(82.9)	(345255.5)
θ, region 14		27.2	-302.4
		(267.1)	(3053.5)
θ, region 15		8.63*	10.9
		(4.43)	(98.2)
θ, region 16		10.6	3.46
		(61.2)	(31.3)
θ, region 17		12.9	678.1
		(77.2)	(6746.0)
θ, region 18		-2.38	99.2
		(57.3)	(894.9)
θ, region 19		8.14	26.4
		(9.98)	(163.7)
θ, region 20		4.79	-24.5
		(4.09)	(122.4)
θ, region 21		15.4	15.0
		(42.1)	(183.8)
θ, region 22		5.01	-13.8
		(92.5)	(77.2)
θ, region 23		8.79	3.23
		(22.1)	(19.9)
θ, region 24		80.9	-6.93
		(740.0)	(15.5)
θ, region 25		12.6	2.95
		(40.7)	(47.2)
P value from F-test		1.00	1.00

Notes: Table shows the estimate of theta from the pooled data, and then the specification where this it is estimated region-by-region. The p value is the p value from the F test that all coefficients are equal. IV estimates.

Appendix Table 18: Wage ratio and distance

Dependent variable: log $\frac{wage\_do}{wage\_oo}$	(1) b/se	(2) b/se
Log distance +1	0.032	0.032
	(0.018)*	(0.018)*
Constant	0.042	0.067
	(0.135)	(0.136)
Year FE	No	Yes

Notes: Table testing hypothesis that the wage gap is zero when transport costs are zero.