The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City *

Melanie Morten †  Jaqueline Oliveira ‡
Stanford University and NBER Rhodes College

August 9, 2018

Abstract

A large body of literature studies how infrastructure facilitates the movement of traded goods. We ask whether infrastructure also facilitates the movement of labor. We use a general equilibrium trade model and rich spatial data to explore the impact of a large plausibly exogenous shock to highways in Brazil on both goods markets and labor markets. We find that the road improvement increased welfare by 13.3%, of which 95% was due to reduced trade costs and 5% to reduced migration costs. Nevertheless, costly migration is responsible for large spatial heterogeneity in the benefits of roads: the interquartile range of welfare improvement is 6%–23%, as opposed to uniform gains with perfect mobility.

Keywords: Internal migration, Brazil, Infrastructure, Roads
JEL Classification: J61, O18, O54

*We thank Treb Allen, Marcella Alsan, Ana Barufi, Gharad Bryan, Kate Casey, Arun Chandrasekhar, Rafael Dix-Carneiro, Dave Donaldson, Taryn Dinkelman, Pascaline Dupas, Fred Finan, Doireann Fitzgerald, Paul Gertler, Doug Gollin, Saum Jha, Brian Kovak, Kyle Mangum, Ana Minian, Mushfiq Mobarak, David McKenzie, Paula Pereda, Nancy Qian, Steve Redding, Mark Rosenzweig, and Sam Schulhofer-Wohl for discussions; and seminar participants at the NBER Economic Geography Workshop, 36th Meeting of the Brazilian Econometrics Society, Emory University, Harvard/MIT Development Workshop, 2014 NBER Summer Institute, 7th Migration and Development Conference, University of Calgary, WUSTL, Clemson University, University of Los Andes, University of Sao Paulo, Oxford University, University of Houston, UCL/LSE, Duke University, University of Memphis, IIIES, and 2016 Lacea/Lames for helpful comments. Anita Bhide, Caue Dobbin, and Devika Lakhoti provided excellent research assistance. Part of this work was completed while Morten was a visiting scholar at the Federal Reserve Bank of Minneapolis, and their hospitality is greatly acknowledged. This paper is a revised version of an earlier version of the paper, “Paving the way to development: costly migration and labor market integration,” which also circulated as “Migration, roads and labor market integration: Evidence from a planned capital city.” Any errors are our own.

†Email: memorten@stanford.edu
‡Email: oliveiraj@rhodes.edu
1 Introduction

Roads are built to facilitate travel. Most scholarship on the impact of roads on economic outcomes has focused on how roads facilitate the “travel” of goods\(^1\), while ignoring how roads and highways may also facilitate the travel of people and thus promote internal migration. This paper asks (i) whether roads enable migration; and (ii) if so, what share of the welfare gains from roads accrue due to trade market integration and what share accrue as a result of labor market integration.

Our first, and main, contribution is to document that roads promote migration. To that end, we need an empirical setting with two characteristics. First, roads are generally not built exogenously; instead, they are built to connect places where there may be high (or low, if the goal of policy is to stimulate) demand for either trade or migration. We therefore look for a place with plausibly exogenous variation in the location of the road network. Second, because roads also change prices (wages, goods prices, and rents), we need to observe gross migration flows and not simply net population changes. Gross flow data allow us to separate out the effect of roads on costs from the effect of roads on prices because we can undertake the analysis within origin-destination pair, hence controlling for a complete set of origin-year and destination-year fixed effects. In other words, we are able to estimating the elasticity of migration to roads while controlling for any effects of roads on wages or prices at either the destination or origin.

Brazil is an empirical setting that meets both requirements. In 1960, Brazil built a new capital city, Brasilia, and subsequently constructed a highway network radiating out from the new capital to connect to other state capitals. The network is known as the radial highways, and is depicted in Figure 1 as dashed red lines. The actual location of the radial highways and the choice of the cities connected in the way between Brasilia and a state capital could have been dictated by economic factors. We then construct a least-cost predicted highway system and use this predicted highway system as an instrument for the location of the actual road network\(^2\). The predicted highway system is presented in Figure 1 as gray lines. The instrument picks up the fact that many smaller locations along the road between Brasilia and a state capital get incidentally connected by the radial highways. We start by showing, using data from before and after the construction of the radial highways, that migration rates increased relatively more between states that were

---

\(^1\)See, for example, Michaels 2008; Banerjee et al. 2012; Duranton et al. 2014; Faber 2014; Ghani et al. 2014; Hornung 2014; Donaldson 2016; Donaldson and Hornbeck 2016.

\(^2\)The construction of this straight-line instrument follows Faber 2014 and Banerjee et al. 2012.
more connected as a result of the road expansion. We also show that trade increased relatively more between states that became more connected. We establish these results by estimating gravity equations that control flexibly for shocks at the destination-year, origin-year, and pair level, and so identify the elasticity of migration or trade to roads only from the within-pair change in travel time. We then turn to using rich individual-level migration data, collected in the national census from 1980, to show that the same results hold for migration measured at the sub-state level for the later period of 1980-2010. In these regressions, we are not able to control for pair fixed effects, and so the elasticity is identified by using cross-sectional, rather than temporal, variation in the location of roads. In both specifications we find a statistically significant elasticity of migration to roads. We take this as evidence that roads did facilitate migration.

It is reasonable, of course, to ask whether a one-time migration cost, which may be small relative to the present value of a higher future income stream, will affect the decision to migrate. We think of migration costs broadly to include both the fiscal cost of moving, as well as utility costs, such as being away from friends and family (Sjaastad, 1962). For migrants who return home to visit friends and family after migrating, migration costs also capture the flow costs of such return visits (with both a fiscal component and a time component). Migration costs can also capture any costs of not being able to consume the same types of goods as at home. Empirically, we use geocoded data on migration choices and show that people make migration decisions in a way that is consistent with it being more difficult to move to a place that is farther away from the origin.

Our second contribution is to decompose the welfare gains from road connectivity into migration and trade components. A spatial equilibrium framework with costly trade and migration underpins our quantitative analysis (Roback, 1982; Eaton and Kortum, 2002; Monte et al., 2018). We follow Donaldson and Hornbeck 2016 (DH) and define the equilibrium prices and labor in terms of “market access”. Like DH, prices and labor depend on a market access term, which we call “trade market access” (TMA). TMA measures a city’s ease of access to cheaply produced goods and increases when trade costs fall. Unlike DH, who consider perfect mobility of workers, in our framework workers may not be perfectly mobile due to the presence of migration costs. We define a term, “labor mar-

---

3 For example, Atkin (2016) documents that internal migrants in India pay a “caloric tax” to keep eating the types of food they eat in their origin states that are less available in the destination state.

4 To be precise, we study people who have changed their location of residence in the last five years. This is distinct from commuting, which is working in a different location than you live. Commuting flows in the data are small, with 92% of the sample working in the same municipality they live (a municipality, \( n = 3659 \), is a much smaller region than the 135 meso-regions we study in the main analysis).
ket access” (LMA), to capture this. LMA measures a city’s ease of access to workers and increases when migration costs fall.

We show that all the endogenous prices and quantities of the model can be summarized by the elasticity of each component to TMA and to LMA. These elasticities depend on three structural parameters: the elasticity of the price of goods to trade market access (a parameter that governs the incentives to specialize and trade), the elasticity of migration to real wages (a parameter that represents the heterogeneity in idiosyncratic tastes for cities), and the elasticity of housing prices to population. For example, as road connectivity improves, both TMA and LMA increase. An increase in LMA leads to a location receiving more migrants; this captures the effect of roads on migration (and hence labor) through lower migration costs. The effect of TMA on migration is ambiguous. On the one hand, higher TMA translates into lower tradable goods prices, and lower prices attract workers. On the other hand, higher TMA indirectly raises housing prices, because of the migrants attracted by lower wages, and these higher rents repel workers. The first effect dominates if the housing expenditure share is low or the housing elasticity multiplied by the trade elasticity is low. We derive similar intuitive expressions for rents, prices, wages, and welfare.

To estimate the three structural elasticities, we follow the approach of Diamond 2016 and decompose the fixed effect of the gravity equation, which has a structural interpretation as the bundle of wages, amenities, and prices at the destination. We construct instruments for wages, goods prices, and labor from labor demand shifters (Bartik, 1991), and these labor-demand shifters interacted with labor market access and trade market access. This procedure yields an estimate of the migration elasticity to wages of 1.9 and an estimate of the housing elasticity to population of 0.9. We set the trade elasticity to 4 following Simonovska and Waugh 2014.

With these elasticities in hand, we perform two counterfactual exercises in the model. First, we simulate in the model what would have happened if the road network had not been built. This is equivalent in our framework to increasing trade costs by 28% and increasing migration costs by 11%. We find that this counterfactual would have decreased welfare in Brazil by 13.3%. Of this reduction, 95% is due to the reduction in goods market integration and 5% due to the reduction in labor market integration. While we find that trade constitutes the dominant source – 95% – of gains from the roads, costly migration is responsible for large spatial heterogeneity in the benefits of roads. We estimate the interquartile range of welfare improvement is 6–23%, as opposed to uniform gains with
perfect mobility. The second exercise is to compute the welfare effects of a hypothetical road network if the capital had remained in Rio de Janeiro, instead of shifting to Brasilia. This counterfactual builds the road highway in different parts of the country, with relatively fewer roads in the center of the country as compared to the road network that was built extending from Brasilia. We find that keeping the capital in Rio de Janeiro would have reduced welfare by 4.8%. In this case, we find relatively little in the way of spatial heterogeneity in effects, with an interquartile range of welfare loss of 4.5–5.1%.

Our paper is related to several strands of the literature. A rich spatial literature has examined issues of migration, but has traditionally assumed that migration costs are purely due to preferences (Moretti, 2011; Diamond, 2016; Allen and Arkolakis, 2014; Redding, 2016) or that labor is immobile across space (Donaldson, 2016; Topalova, 2010). Recent papers have relaxed this assumption to consider costly migration or commuting (Caliendo et al., 2017; Monte et al., 2018; Tombe and Zhu, 2015; Fan, 2015). We follow this latter stream of the literature and write down a model with costly migration and costly trade. The focus of our paper, relative to other papers considering the joint determination of trade and migration, is to explicitly consider the role of roads in facilitating both migration and trade.5

Our work is also related to the migration literature. Many studies have focused on the responsiveness of migration to economic returns (Sahota, 1968; Harris and Todaro, 1970; Pessino, 1991; Tunali, 2000); other papers have found strong evidence, in partial equilibrium settings, of migration costs (Kennan and Walker, 2011; Bryan et al., 2014). We add to this literature by considering the responsiveness of migration to both costs and returns and by considering the general equilibrium effects of migration and trade from roads.6

Finally, our paper is related to a development literature that studies the allocation of resources. While the prior literature has looked at institutional barriers (Janvry et al., 2015) and insurance barriers (Banerjee and Newman, 1998; Munshi and Rosenzweig,

5A related stream of the literature studies costs in switching between sectors (Artuç et al., 2010; Dix-Carneiro and Kovak, 2014). The modeling framework in these papers has a very similar structure as to the costs of switching location.

6Chein and Assunção 2016 study the effect of migration on wages and use the construction of a road in the North of Brazil as an instrument for migration. While finding that roads do affect migration, the paper does not separate out the effects of roads on migration separately from the effects of roads on trade (and hence prices and wages). Bird and Straub 2015 also use the construction of Brasilia as an instrument to study the effect of road construction on regional GDP. Their study focuses on the effect of roads on GDP, and they do not examine the effect of roads on migration. Jayachandran 2006 studies how the general equilibrium pass-through of productivity shocks into wages is mitigated for areas that are more connected to other locations.
we focus on travel time as a barrier. If it is costly to move out of low income locations, labor may not be able to move to its most productive locations. If labor is not allocated most productively, then aggregate productivity may decrease, along similar lines to studies examining the misallocation of capital, such as Hsieh and Klenow 2009. Our paper presents evidence that, while these migration costs are in large part attributed to tastes, they can be considerably reduced by improving access to transportation infrastructure. This finding suggests there is a margin for policy makers to improve the allocation of labor across space.

Our paper does have several limitations. A key contribution of our paper is to calculate the relative effect of roads on goods and labor market migration. In order to make direct comparisons with other studies, we focus on a static model of migration. Recent work in trade has considered dynamic approaches to migration (Artuç et al., 2010; Caliendo et al., 2017); including a dynamic component in the model would allow an additional channel for long-run adjustment that we do not consider. Second, our model does not explicitly consider endogenous agglomeration or congestion forces (although we do allow for endogenous cost-of-living effects, which are similar to endogenous congestion forces, through the housing market). We make this assumption because our primary focus is on understanding the additional effects of roads on labor market integration, compared to the case where this is not considered.

The plan of the paper is as follows. In Section 2, we discuss the historical context that led to the construction of Brasilia and explain how we use this natural experiment to provide exogenous variation in the road network. We present the structural model in Section 3 and our estimation strategy in Section 4. We then highlight the decomposition of the effects of roads on goods and factor markets in Section 5. We briefly conclude in Section 6.

2 Do roads affect trade and migration?

This section asks whether roads affect the level of trade and migration between locations. We estimate gravity equations that show that places that are more closely connected by roads both have larger trade flows and larger migration flows. We first look at state-to-state flows. The advantage of the state-to-state data is that we have these data for before

---

7In the Brazilian context, Dix-Carneiro and Kovak 2017 find evidence that regional responses to tariff shocks may in fact be amplified over time. In our setting, we study the migration response to exogenous labor demand shocks and find evidence consistent with, albeit delayed, convergence after economic shocks.
and after the construction of the radial highways. We establish that, within pairs, migration increased more between places that became more connected by the road network. Next, we look at sub-state (meso region to meso region) migration flows. While more disaggregated, meso-to-meso flows are only available for the post-Brasilia period. We establish that, across pairs, migration is more common for locations that are more connected by roads.

2.1 Data

We start by considering how state-to-state flows of traded goods and of migration responded to the presence of roads. We source data on state-to-state trade and migration flows obtained from statistical yearbooks. Trade flow data correspond to the value of imports and are available annually, spanning the periods 1942–1949, 1967–1974, and 1999. We source state-to-state trade flows for the year 1999 from de Vasconcelos 2001; the other years are sourced from historical yearbooks as required. Migration flow data refer to the total number of people in each state who originated from other states. These data are available decennially from historical yearbooks for the period 1940 to 1980 and then from 1980-2010 from the Brazilian census. We fully describe the data in Appendix B.2.

To compute road distances and travel times, we obtain geo-referenced maps of the Brazilian road network from the Brazilian Ministry of Transportation decennially from 1960 through 2010. We compute travel time measures by setting the travel speed for three types of pixels (those without a road, those with a road, and those in a city), and then computing the least-cost path between an origin and a destination using the fast marching algorithm. We do this for two measures of geography: meso regions and states. This follows the approach used in Allen and Arkolakis 2014. We compute travel time on both the actual road network as well as on a predicted road network, which we explain in more detail below, as an instrument for the actual road network.

---

8For the state-level analysis, all the results are robust to assign the inter-state migration cost as the average of the inter-meso moves. We could potentially compute the travel times one level of geography lower, at the “AMC” (minimum comparable level) and then take aggregates for the inter-AMC costs to construct the inter-meso and inter-state costs. However, we have not pursued this as it would require running the fast marching algorithm 13.4 million times (there are 3659 AMCs) to compute this measure, compared with the current dimensionality of 135 meso regions. As a result, the travel cost measure may contain measurement error associated with whether certain sub-parts of the region are predominantly responsible for migration and trade flows, rather than the average midpoint of the region.
2.2 Gravity in trade and migration

During the first half of the twentieth century, roads were nearly nonexistent in Brazil. The few that existed were dirt or gravel roads and served a limited number of urban centers in the southeast. The construction of the new capital city—Brasilia—in the middle of the country was accompanied by the development of a new highway network to link the national capital to the other existing state capitals. Figure 1 shows the location of Brasilia in the epicenter of Brazil. We construct the travel time between origin \( o \) and destination \( d \) by running a least-cost algorithm based on both the actual roads and then on our predicted roads (described in further detail below).

Figure 2 shows the change in trade and migration for the two years adjacent to the construction of Brasilia. State pairs that had larger decreases in travel time had larger increases in migration (left panel) and larger increases in trade (right panel). This figure suggests that roads may facilitate both trade and migration. However, there may be several other explanations for this relationship: for example, if it were the case that roads were built between pairs that were expected to have high economic growth, then economic growth and not roads may be the driving force for higher trade and migration flows. It may also be the case that favorable labor market shocks happened to hit locations that were recently connected by the road, and this is what drives trade and migration. To address these concerns we propose an instrument for the location of the road network and then estimate a gravity model that can flexibly control for origin-year, destination-year, and pair level shocks.

2.3 Instrument for road location: location of new highway network

To generate an instrument for travel time, we use plausibly exogenous cross-sectional variation in the location of highways in Brazil generated by the construction of a planned capital city, Brasilia. Brasilia was constructed in 1960 in response to the long-standing issue of finding the ideal location for the country’s capital city.\(^9\) Starting with Brazil’s first Constitution in 1891, a 60 x 90 kilometer piece of land, the Quadrilátero Cruls, located close to the border between the states of Goias and Minas Gerais, had been allocated for the site of a future capital city. In 1922, the National Congress approved the creation of the new

---

\(^9\)Brazil is not alone in solving the capital-city location problem by constructing an entirely new city. Other countries that have employed this strategy include Australia (Canberra), Belize (Belmopan), Burma (Naypyidaw), India (New Delhi), Kazakhstan (Astana), Nigeria (Abuja), Pakistan (Islamabad), and the United States (Washington, D.C.).
capital within this site. However, between 1922 and 1946, there was little movement towards building the new capital city. In 1946, Eurico Dutra became president, and renewed debates over the site and construction of the new capital. From there things progressed rapidly: in 1955, the recently created Commission for the New Federal Capital finalized the area in which Brasilia would be placed; the president elected in 1956, Juscelino Kubitschek, created the Company for Urbanization of the New Capital, and the construction of Brasilia began immediately. After three years and ten months, Brasilia was officially inaugurated on April 21st, 1960.

2.3.1 Construction of the road network

The construction of Brasilia and the construction of the radial highway system occurred simultaneously. Before 1951, the few existing roads in Brazil were limited to the coastal areas of the Southeast and Northeast. The 1934 and 1944 national transportation plans (Planos Nacionais de Viacao, PNVs) were the first to mention a national highway system. The planned highway system was finalized in PNV 1956, once the location of the new capital had been decided. The final highway network connected Brasilia to the rest of the country. The roads run radially from Brasilia towards the country’s extremes in eight directions: north, northeast, east, southeast, south, southwest, west, and northwest.

2.3.2 Construction of the instrument

One possible instrument would be to construct straight lines between Brasilia and the cities connected to Brasilia by the radial highways and then use these straight lines as an instrument for the actual road network. One concern with this approach is that the terminal cities of the road network may have been chosen based on their economic attributes. We therefore proceed slightly differently. Our best understanding, based on reading the planning documents, was that the goal of the highway system was to connect the national capital to the state capitals in the eight directions mentioned earlier. We therefore divide Brazil into eight segments and predict, within each segment, the minimum path to connect Brasilia to all state capitals contained in that segment, following Faber 2014. We label

---

10Between 1951 and 1957, the Brasilia–Belo Horizonte line was laid down, connecting the soon-to-be new capital to the capital of the state of Minas Gerais. In the same period, parts of the Brasilia–Anapolis highway, a road that would link the new capital to the city of Sao Paulo, was initiated. There were also plans to build the 2,276 km-long Belem–Brasilia, or Transbrasiliana, highway, which would provide an overland route from the underpopulated Northern states to the demographic and industrial centers of the country located in the south.
the resulting network the Euclidean Minimum Spanning Tree (EMST). Figure 1 shows the EMST network, overlaid on the actual radial highway network.

2.3.3 Exogeneity of the road network

The key identifying assumption for the instrument is that the regions connected to the EMST network were similar to the non-connected regions, with respect to baseline characteristics at the time of construction in 1950. This assumption might not hold true if investors had begun development in certain anticipated connection sites. As we outline above, the historical context makes this unlikely because the city was completed very quickly once the government had decided on its final site. An additional concern would be if the road network simply formalized preexisting historical travel routes between cities, then any effects we attribute to the road network may instead be due to the effects of the initial travel routes and not the new roads. Here, however, since roads were being built between an entirely new city and existing state capitals, it is unlikely that the roads simply replaced preexisting travel routes. To test formally whether regions connected to the EMST network are similar to the non-connected regions we examine whether being on the EMST network predicts population and municipality GDP growth in the years before Brasilia, after controlling for distance to the coast, distance to the nearest state capital, and distance to Brasilia. Appendix Table 2 displays the results. We cannot reject the hypothesis that the municipalities that were on the EMST network are similar to those not on the EMST network.

2.4 IV regressions for gravity in trade and migration

We now turn to estimating the elasticity of migration and trade to roads. We run gravity regressions of the form, where a rich set of fixed effects allows us to identify the elasticity of interest, $\beta$:

$$M_{ot} = \gamma_{ot} + \gamma_{dt} + \gamma_{od} + \beta \log \text{travel time}_{odt} + \epsilon_{odt},$$ (1)

where $M_{ot}$ is either migration or trade flows. $\gamma_{ot}$ is an origin-year fixed effect, and controls for any common shocks at the origin. $\gamma_{dt}$ is a destination-year fixed effect, and controls for any common shocks at the destination. $\gamma_{od}$ is an origin-destination pair fixed effect, and controls for any time-invariant factors (such as distance or historical trade/migration networks). The parameter of interest is $\beta$, the elasticity of trade (or mi-
gration) to travel time. This parameter is identified by changes over time to the travel time between origin $o$ and destination $d$ due to the expansion of the road network.

We estimate Equation 1 by Poisson Pseudo Maximum Likelihood (PPML) (Silva and Tenreyro, 2006; Tenreyro, 2009). We first estimate the gravity equation using the actual travel times and not handling the potential endogeneity. Column (1) in Table 1 shows a specification where we control for the (log) distance between origin $o$ and destination $d$ and an indicator for whether the destination location is different from the origin location (dum migrant); column (2) shows the specification where we include the pair fixed effects, so that the coefficients on log distance and dum migrant are not identified. We find that the elasticity of migration to travel time is between -0.37 and -0.48. That is, larger migrant flows occur between places that are more connected compared with places that are less connected.

Road connectivity between bilateral pairs, however, is likely correlated with $\epsilon_{odt}$, introducing a potential endogeneity concern. This is so if, for example, highways are placed to integrate two localities with higher propensities to migrate and stimulate migration flows. In this case, OLS estimates will likely overstate $\beta$ and yield smaller negative impacts of travel time on migration costs. To address this concern, we use the instrument for travel time based on the EMST network. Due to the non-linearity of the Poisson estimating equation, we implement this by using a control function approach (Petrin and Train 2010). We estimate the first-stage equation:

$$\log(\text{travel time}_{odt}) = \delta_{ot} + \delta_{dt} + \delta_{od} + \varphi \log(\text{EMST travel time}_{odt}) + \nu_{odt}.$$  

We add the residuals from the first stage as an additional control in estimating Equation (1). Columns (3) and (4) in Table 1 show the results. We find evidence that the road network was built to connect places that had higher unobserved preferences to migrate, and so the point estimate of migration elasticity to roads is lower after addressing the endogeneity of road placement. We find an elasticity of -0.68 (Column 3) when we control for log distance, and an elasticity of -0.65 when we control for pair fixed effects (Column 4). These point estimates are not statistically different from one another. We show results both with and without pair fixed effects as we want to establish whether unobserved pair confounders are likely to be correlated with our predicted road network. This is impor-
tant because later in the analysis using meso-to-meso migration flows we will not be able to control for pair fixed effects. The fact that the point estimates in Column (3) and (4) are not statistically different from one another alleviates concerns that unobserved pair-level confounders are correlated with the placement of the road network.

We repeat the same exercise for trade flows in Table 2. We find an elasticity of trade to travel time of -4.0 after including the control function residual. Again, comparing the point estimate with and without the pair fixed effects (Column (3) and (4)), we do not find any evidence of unobserved pair-level confounders that are correlated with the placement of the road network.

2.5 Sub-state evidence

Starting in 1980, the Brazilian micro census collects migration data coded to the municipality level. This lets us look at whether the location of roads still affects migration decisions in this later period. To examine this, we run Equation 1 for migration flows, where now the unit of analysis is five-year flows between meso regions, rather than movements out of state of birth. One shortcoming of this data is that we do not observe meso-to-meso flows before and after the construction of the radial highways, so we cannot include pair fixed effects. Instead, we include the log distance between the origin and destination to proxy for time-invariant variables at the pair level. This involves an assumption that log travel time is exogenous conditional on distance, which we could not reject in the state-level data. We also include Dum migrant as control to allow for a fixed cost of moving.

The results are in Table 3. We find, in Column (3), an elasticity of migration to roads of -1.2. This is within the confidence interval for the elasticity measured at the state level. Columns (3) and (4) reestimate the model allowing for individuals to gain utility from living in their state of birth. While gross migration flow data are rarely available in population censuses, state of birth is often recorded and is used as a proxy variable for migration costs. As expected, people gain utility from living in their state of birth, with a coefficient of 4.51, but we still find a statistically significant negative elasticity to travel time.\footnote{In fact, the elasticity is exactly the same magnitude. This is expected: the Poisson model is a count model, where the dependent variable is the gross number of people moving to a location; the inclusion of the state-of-birth does not change the allocation of people and hence does not affect the estimate of the bilateral elasticity.}

One threat to identification is that, between 1950 and 1980, people may have had time to adjust to the road network and, as a result, while the road network may be exogenously located across geographical location, it is no longer exogenously distributed across indi-
viduals. We provide three pieces of evidence that mitigate this concern. First, even if people responded by moving closer to the road network, it is likely that they moved to the locations that were actually connected through the radial highways. We mitigate this concern by using a least-cost predicted highway (EMST road network) instead of the actual radial highways, so that we are always exploiting the variation from locations that were fortuitously connected to the road network. Second, the way that we estimate the costs of migrating are through “gravity-style” regressions, where we explain the bilateral flows of people migrating as a function of the (instrumented) travel time between the two locations, controlling for a full set of origin-year and destination-year fixed effects. If it is the case that the types of workers who live near the road highway are different than the types of workers who live away from the road network, then this effect will be controlled for by the origin-year or destination-year fixed effects. Lastly, the ideal analysis would be able to control for any pair-level effects, such as Euclidean distance or cultural and socioeconomic proximity, that may affect both roads and migration, and then study the change in migration as two places become more connected. Because the migration data are only available from 1980, our main specifications do not do this test. However, the fact that we found no evidence or pair-level effects when running the regressions at the state level help alleviate this concern.

Taken together, the results in this section imply that building more roads leads to a decrease in the cost of migrating and a decrease in the cost of trading goods. We now turn to the framework to formalize how road connectivity will determine migration and trade and therefore to decompose the relative effect of each channel.

3 Model developing trade and labor market access

This section describes the framework we use to quantify the effects of roads in Brazil following the construction of the new capital city in Brasilia. The goal of the model is to map migration and trade costs into a quantitative framework. The framework is based on a standard spatial equilibrium model, where trading goods is costly, as in Eaton and Kortum 2002, and there may be origin-destination costs of movement, such as in Monte et al. 2018. Given trade costs between origin $o$ and destination $d$ at time $t$, $\tau_{odt}$, migration costs $\kappa_{odt}$, exogenous productivities $A_{ot}$, exogenous amenities $B_{ot}$ and initial labor force, $L_{o,t-1}$,
the model endogenously determines wages $w_{ot}$, prices $P_{ot}$, rents $r_{ot}$ and labor $L_{o,t-1}$.\footnote{In the data, a period $t$ is the 10-year period between census rounds, except for the migration question, which asks about location five years earlier.}

In the model, agents optimally choose their location each period. Locations specialize in production based on the costs of importing goods from other locations. A location that can cheaply import goods will have a lower price level for goods. Agents pay a cost if they migrate. This cost has two components, one origin-destination specific and one idiosyncratic. To provide a cohesive framework to understand the model, we rewrite the model in terms of market access, following Donaldson and Hornbeck 2016. We have the standard market access term – which we call trade market access – and then a second term, labor market access, that measures market access for workers. We show that all equilibrium prices and quantities can be expressed as a function of TMA and LMA.

### 3.1 Trade flows

The production side of the economy is based on a standard model of a location producing goods based on its comparative advantage. Assume that there is a continuum of goods, $j \in [0, 1]$, which are produced in each location $o$ at time $t$ according to a linear production function, $Y_{ot}(j) = z_{ot}(j)L_{o,t-1}$. Goods shipped between origin $o$ and destination $d$ are subject to an iceberg cost, $\tau_{odt} \geq 1$. The cost to a consumer at destination $d$ of buying one unit of good $j$ produced at origin $o$ is $p_{odt}(j) = \frac{\tau_{odt}w_{ot}}{z_{ot}(j)}$.

If $z_{ot}$ takes the Frechet functional form, then the results in Eaton and Kortum 2002 show that the total trade flows, $X_{odt}$ from $o$ to $d$ are given by a gravity expression:

$$X_{odt} = V_{ot}\tau_{odt}^{-\theta}(CMA_{dt})^{-1}Y_{dt},$$

where $V_{ot} = A_{ot}w_{ot}^{-\theta}$ is the cost of producing goods in origin $o$, and, following Donaldson and Hornbeck 2016, consumer market access, $CMA$, is given by $CMA_{dt} = \sum_{o} V_{ot}\tau_{odt}^{-\theta}$. A related result from Eaton and Kortum 2002 is that the price index in location $d$ is given by $P_{dt}^{-\theta} = f(CMA_{dt})$. Consumer market access is higher when locations are better connected to low-cost producers of goods; as a result, they face lower prices for the goods they consume. The parameter $\theta$ governs the role of comparative advantage in shaping trade patterns. A lower $\theta$ means more heterogeneity in productivity and a stronger incentive to specialize and trade.
3.2 Migration flows

The utility of an individual $i$ from origin $o$ living at destination $d$ depends on (i) goods consumption, $C_{dt}$, (ii) housing consumption, $H_{dt}$, (iii) an individual-specific preference shock, $b_{dt}(i)$, and (iv) migration cost, $\kappa_{odt}$, according to $U_{odt}(i) = \frac{b_{dt}(i)}{\kappa_{odt}} \left( \frac{C_{dt}}{\kappa_{odt}} \right)^\alpha \left( \frac{H_{dt}}{1-\alpha} \right)^{1-\alpha}$. Migration costs also take an iceberg form, so that $\kappa_{odt} \geq 1$. We assume further that it is costless to stay in the origin ($\kappa_{oot} = 1$) and that migration costs are symmetric ($\kappa_{odt} = \kappa_{dot}$).

If we assume that taste shocks, $b_{dt}(i)$, are drawn from a Frechet distribution, then we get a gravity equation for the total migration flow, $M_{odt}$, between $o$ and $d$ at time $t$ that is symmetric to the case for trade:

$$M_{odt} = U_{dt} \kappa_{odt}^{-\epsilon} \left( WMA_{ot} \right)^{-1} L_{0,t-1},$$

where $U_{dt} = B_{dt} w_{dt} (P_{dt}^{\alpha} r_{dt}^{1-\alpha})^{-\epsilon}$ is the utility of living in location $d$, and $L_{0,t-1}$ is the beginning-of-period number of workers in origin $o$. We define worker market access, $WMA$, as $WMA_{ot} = \sum_d U_{dt} \kappa_{odt}^{-\epsilon}$. Worker market access measures the outside options of workers in location $o$. $WMA$ is higher for places that are closer to high-utility (higher amenities, higher wages, lower prices, and lower rents) destinations that workers can migrate to. The parameter $\epsilon$ (which we refer to as the migration elasticity to real wages) regulates the role of unobserved taste shocks in shaping migration patterns. A lower $\epsilon$ signifies more heterogeneity in idiosyncratic tastes for locations and a weaker response of migration flows to changes in economic returns of migration.

3.3 Determination of prices

The price of consumer goods, $P_{dt}$, is determined through the trade side of the model, with $P_{dt}^\theta = f(CMA_{dt})$, as noted above. The price of housing, $r_{dt}$, is determined through the housing market. Following Monte et al. 2018, we assume that all housing is owned by landlords who consume traded goods in the location where the housing is located.\footnote{This assumption will yield that total earnings in a location are equal to total expenditure, which is required for computing balanced trade later in the paper (the demand for traded goods is the sum of the goods share from labor and the expenditure on goods from landlords, yielding $P_{nt}C_{nt} = \alpha w_{nt} L_{nt} + (1-\alpha) w_{nt} L_{nt} = w_{nt} L_{nt}$. Redding 2016 assumes that housing revenue is redistributed as a lump sum to residents. This also yields the outcome that total income is the sum of labor income and residential expenditure; $v_{nt} L_{nt} = w_{nt} L_{nt} + (1-\alpha) v_{nt} L_{nt} \Rightarrow v_{nt} L_{nt} = \frac{w_{nt} L_{nt}}{\alpha}$. This would change the demand for both goods $p_{nt} C_{nt} = w_{nt} L_{nt}$ and housing $r_{nt} H_{nt} = (1-\alpha) \frac{w_{nt} L_{nt}}{\alpha}$ and would change the indirect utility term to be $U_{wnt} = b_{wnt} \frac{w_{nt}}{v_{nt}} \left( \frac{p_{nt} C_{nt}^{\alpha}}{\alpha P_{nt}^{\alpha} (w_{nt})^{1-\alpha}} \right)$.}
Following Diamond 2016, we model the price of housing depending on the underlying cost of producing housing units. The price of housing is determined by the marginal cost of constructing housing, which depend on construction costs, $CC_{dt}$, and land costs, $LC_{dt}$, as $p_{dt}^{\text{house}} = MC(CC_{dt}, LC_{dt})$.

Assuming a steady state equilibrium in the asset market, prices equal the discounted value of rents. Therefore, the housing equilibrium condition can be approximated by a linear expression:

$$\log r_{dt} = \log t + \log CC_{dt} + \eta \log HD_{dt},$$

where $HD_{dt} = (1 - \alpha)w_{dt}L_{dt}$,

and $\eta$ is the housing supply elasticity. The cost of housing increases when the demand for housing services increases, which can be due to either an increase in wages or an increase in the labor force.

### 3.4 Model equilibrium

We assume that trade is balanced between locations ($Y_{ot} = \sum d X_{odt}$). Substituting equation 2 into this expression yields $Y_{ot} = V_{ot}FMA_{ot}$, where, again following Donaldson and Hornbeck 2016, firm market access, $FMA$, is defined as $FMA_{ot} = \sum d \tau_{odt}^{-\theta} (CMA_{ot}1)^{-1} Y_{dt}$. Firm market access measures how easy it is for a destination to sell goods: $FMA$ is higher when a location is near destinations that have high incomes or destinations that are less competitive (have lower $CMA$). Donaldson and Hornbeck 2016 show that, if trade costs are symmetric, the two measures of market access relating to trade are the same (up to a constant). Therefore, we define trade market access, $TMA$, as $TMA_{ot} = FMA_{ot} = \rho CMA_{ot}$ for some constant $\rho$, so we can write income as a function of $TMA$:

$$Y_{ot} = V_{ot}TMA_{ot}.$$  (5)

We then derive the analogous expression for the labor side of the market. In our model, it is costly for labor to move. The total labor force in a destination is given by the sum of all in-migrants, $L_{dt} = \sum o M_{odt}$. Substituting equation 3 into this expression, we have $L_{dt} = U_{dt}EMA_{dt}$, where employer market access, $EMA$, is equal to $\sum o \kappa_{odt}^{-\varepsilon} (WMA_{ot}1)^{-1} L_{o,t-1}$. Employer market access measures how easy it is to attract workers to the location, and it is higher when a location is close to a location with a large population or a location
that lacks a very competitive labor market (low worker market access). We will define labor market access, \( LMA \), as \( EMA \) or \( TMA \), and so can write the labor force in a destination as a function of LMA:

\[
L_{dt} = U_{dt} LMA_{dt}.
\]  

(6)

These two expressions then allow us to solve directly for equilibrium labor and wages, using the fact that since labor is the only factor of production, \( Y_{dt} = w_{dt} L_{dt} \). To do so, we use equations 6 and 5, along with \( P_{dt}^{-\theta} = \kappa_p TMA_{dt} \), for some constant \( \kappa_p \), and \( r_{dt} = \kappa_r (Y_{dt})^\eta \) (from equation 4), to write:

\[
w_{dt} = \left( \kappa A_{dt}^{1+\epsilon(1-\alpha)} (B_{dt})^{-1} TMA_{dt}^{1-\epsilon(\frac{s}{\theta} - \eta(1-\alpha))} (LMA_{dt})^{-1} \right)^{1+\theta+\epsilon(1+\theta\eta(1-\alpha))}.
\]  

(7)

Equilibrium wages are a function of a constant term, \( \alpha_w \), and the two market access terms:

\[
\log w_{dt} = \alpha_w + \epsilon_{TMA} \log TMA_{dt} + \epsilon_{LMA} \log LMA_{dt},
\]

where the elasticity of wages to TMA, \( \epsilon_{TMA} \), is given by \( \epsilon_{TMA} = \frac{1-\epsilon(\frac{s}{\theta} - \eta(1-\alpha))}{x} \), and the elasticity of wages to labor market access, \( \epsilon_{LMA} \), is given by \( \epsilon_{LMA} = -\frac{1}{x} \), where \( x = 1 + \theta + \epsilon(1+\theta\eta(1-\alpha)) \), a positive number. LMA measures the total supply of workers into a location so that, all else constant, a place that has abundant labor pays lower wages, and so the elasticity of wages to LMA is always negative. The sign of the elasticity of wages to TMA is indeterminate. There are two opposing forces at play. On the one hand, places with higher TMA have lower tradable goods prices, and the lower the \( \theta \), the higher the fall in prices following increases in TMA. Lower tradable goods prices attract workers, which depresses wages. On the other hand, a higher TMA translates into higher housing prices, and the higher the \( \eta \), the larger the increase in rents from increases in TMA. These opposing forces are mediated by \( \epsilon \), the parameter that governs the responsiveness of migration to changes in economic returns.

The equilibrium labor allocation is also a function of the two market access terms:

\[
\log L_{dt} = \alpha_L + \epsilon_{TMA} \log TMA_{dt} + \epsilon_{LMA} \log LMA_{dt},
\]

where the elasticity of labor to TMA, \( \epsilon_{TMA} \), is equal to \( \epsilon_{TMA} = \frac{\epsilon(1+(\frac{s}{\theta} - \eta(1-\alpha)))}{x} \) and the elasticity of labor to LMA, \( \epsilon_{LMA} \), is equal to \( \epsilon_{LMA} = \frac{1+\theta}{x} \). Labor is always increasing in labor market access, which is consistent with the underlying definition of labor market access - a nearby pool of available workers. Whether labor is increasing on TMA depends
on the size of the response of tradable goods prices to TMA relative to the response of housing prices.

The equilibrium value of the housing stock is given by:

$$\log r_{dt} = \alpha^r_{dt} + \frac{1 + \epsilon}{x} \log TMA_{dt} + \eta^r \log LMA_{dt}.$$ 

Rents are increasing in both TMA and LMA, which measures the returns to an immobile factor increasing when market access expands.

The elasticity of prices is given by:

$$\log P_{dt} = \alpha^P_{dt} + \epsilon^P_{TMA} \log TMA_{dt},$$

where $$\epsilon^P_{TMA} = \frac{-1}{\theta},$$ and so prices always decrease in TMA.

Indirect utility, which is given by

$$U_{dt} = B_{dt} w^e_{dt} \left( \frac{p^x_{dt} r^1_{dt}}{\alpha dt} \right)^{-\epsilon},$$

is given by:

$$\log U_{dt} = \alpha^u_{dt} \epsilon^U_{LMA} \log LMA_{dt} + \epsilon^U_{TMA} \log TMA_{dt}.$$ 

The elasticity of utility to LMA is given by

$$\epsilon^U_{LMA} = -\epsilon \frac{(1+(1-\alpha)\eta \theta)}{x},$$

and is always negative. The elasticity of utility to TMA is the weighted sum of all the components of utility,

$$\epsilon^U_{TMA} = \epsilon \left[ \epsilon^w_{TMA} - \alpha \epsilon^P_{TMA} - (1-\alpha) \eta \left( \epsilon^w_{TMA} + \epsilon^L_{LMA} \right) \right].$$

The sign of the elasticity of utility to TMA depends on the relative effect of TMA compared with the cost of living (prices and rents).

### 3.4.1 Alternative models of trade and migration

The framework above nests three alternative models of spatial equilibrium: (a) the full model with costly trade and costly migration (preference costs and bilateral costs); (b) costly trade, with migration only a function of preference costs and not bilateral costs; and (c) the model with immobile labor. The expressions we derived above were for case (a).

In case (b) LMA is constant across all regions, as there is no origin-destination component of migration. The comparative statics therefore collapse to only the TMA terms. Case (c) is the case where LMA is equal to zero, because labor is immobile. The comparative statics collapse to the TMA terms, setting the migration elasticity to zero; for the elasticity of utility to TMA, the migration elasticity term is simply omitted.

The actual values of the elasticities depend on three structural parameters: the migration elasticity, $$\epsilon,$$ the trade elasticity, $$\theta,$$ and the housing elasticity, $$\eta.$$ The next section
discusses our estimation of these three parameters. Once we have estimated the parameters, we compute the implied elasticities to TMA and LMA for each endogenous variable and summarize these across the three models. This exercise is reported in Table 5.

4 Estimation

This section discusses the estimation, which follows directly from the structure, of the model just outlined. We have already estimated the parameters determining the migration and trade cost function in Section 2. This section discussed how we estimate the other key elasticities in our model: the migration elasticity to wages ($\epsilon$) and the elasticity of rents to population ($\eta$). We set the value of the trade elasticity, $\theta$, to standard values in the literature.

4.1 Data

Our main database is composed of $135 \times 135$ origin-destination pairs, where each location is a meso region. We have four years of data available for each pair (1980–2010, decennially), yielding a total of $135 \times 135 \times 4 = 72,900$ observations. For each origin-destination pair, we have information on gross migration flows, the Euclidean distance between pairs, bilateral travel times on the road network, and bilateral travel times on the EMST road network. For each origin/destination, we have data on employment, rents, and wages. Our sample of interest consists of males aged 20–65 who report non-zero earnings in their main occupations. The primary datasource is the individual data files from the Brazilian Census, 1970–2010, collected by the Brazilian Institute of Geography and Statistics (IBGE).

4.2 Estimation of migration, housing, and trade elasticities

The destination fixed effects in the migration gravity model have a structural interpretation as a bundle of wages, amenities, prices, and rents. We decompose these fixed effects and estimate three structural parameters: the migration elasticity, $\epsilon$, the elasticity of rental rates to income, $\eta$, and the trade elasticity, $\theta$.

---

16Brazil has 137 meso regions; due to overlapping municipality boundaries, we combine two sets of meso regions to give us 135 spatial units consistently defined over time. Each meso region has 1.4 million inhabitants on average.
4.2.1 Migration and housing elasticites

The destination-specific component of indirect utility, $\delta^{\kappa}_{dt}$, estimated in the gravity equation, is a function of wages, rents, and prices:

$$\delta^{\kappa}_{dt} = \log B_{dt} + \epsilon(\log w_{dt} - \alpha \log P_{dt} - (1 - \alpha) \log r_{dt}).$$

We model the common amenity value of location $d$ at time $t$ as:

$$\log B_{dt} = b_d + b_{st}t + \xi_{dt},$$

where $b_d$ is time-invariant component, $b_{st}t$ captures state-specific time trends, and $\xi_{dt}$ is an error term. These assumptions yield the following estimating equation:

$$\delta^{\kappa}_{dt} = b_d + b_{st}t + \epsilon(\log w_{dt} - \alpha \log P_{dt} - (1 - \alpha) \log r_{dt}) + \xi_{dt}. \quad (8)$$

From the housing supply equation, we have:

$$\log r_{dt} = \eta_d + \eta_{st}t + \eta(\log w_{dt} + \log L_{dt}) + \psi_{dt}, \quad (9)$$

where $\eta_d$ accounts for time-invariant determinants of construction and land costs, $\eta_{st}$ captures state-specific time trends in these costs, and $\psi_{dt}$ is an error term. The two coefficients of interest are the elasticity of migration to real wages, $\epsilon$, and the elasticity of housing to income, $\eta$. We calibrate $(1 - \alpha) = 0.2$ using the share of rents on total expenditure drawn from the 2008–2009 Survey of Family Budget.\footnote{Data available at http://www.ipeadata.gov.br/}.

We face two major challenges when estimating $\epsilon$ and $\eta$. Firstly, wages, rents, and labor are endogenous. Secondly, we do not directly observe prices. We leverage the model structure, discussed below, to obtain an equation for the change in price index and to draw instruments for the endogenous variables.

4.2.2 Instruments

Instrument for wages

We construct Bartik shocks (Bartik 1991) to instrument for $\Delta \log w_{dt}$. The Bartik shock takes the national-level growth rate in employment for each industry, and constructs a location specific shock based on a baseline industry specialization of each location, using...
the employment composition in 1970.\footnote{One immediate challenge to this identification strategy is that our model applies to a single industry, yet the instrument uses the variation of multiple industries. \textit{Bartelme 2015} provides a consistent microfoundation for Bartik shocks in an aggregate gravity framework that justifies that use of the Bartik shock for models of our type.} Let the Bartik shock for location \(d\) be given by:

\[
\Delta \text{Bartik}_{dt} = \sum_{\text{ind}} (\Delta \log w_{\text{ind},-d,t}) \frac{L_{\text{ind},d,0}}{L_{d,0}},
\]

where \(\log w_{\text{ind},-d,t}\) is the average (log) wage in industry \(\text{ind}\) in year \(t\), excluding workers in location \(d\), and \(L_{\text{ind},d,0}/L_{d,0}\) is the baseline industry composition in location \(d\). The Bartik shocks utilize variation across space in the location of industry. We compute the Bartik shock using data at least 500km from the meso region of interest to avoid spatial correlation in labor demand shocks.\footnote{Results are stable over thresholds from 0–1000km; results available on request. We also check for evidence of serial correlation in the shocks by correlating contemporaneous Bartik shocks with lagged outcomes. We plot the effects of the current Bartik shock on current outcomes and on past outcomes in Appendix Figure 2. Bartik shocks (2010–1991) are strongly correlated with 2010–1991 changes in wages and rents, but not with 1991–1980 changes.}

**Instrument for prices**

We need to control for prices to control for differential trade market access across locations. To construct the relevant measure we follow the trade literature and compute the price index, which turns out to be simply trade-cost weighted wages, and then instrument this index with trade-cost weighted Bartik shocks. To be precise, the Eaton-Kortum model produces a convenient closed form for the price index, given by \(P_{dt} = \left(\kappa TMA_{1 dt}^{1}\right)^{-1}\). Log-linearizing the price index around wages and productivities (holding trade costs constant) yields \(\Delta \log P_{dt} = \sum_{o} \pi_{od,t-1} (\Delta \log w_{ot} - \frac{1}{\theta} \Delta \log A_{ot})\), where \(\pi_{od,t-1}\) is the share of imports in destination \(d\) from origin \(o\) in time \(t-1\).\footnote{For full details, see Appendix C.1.} We observe wages, but the underlying productivities are not observed. We therefore approximate the price index by \(\Delta \log \tilde{P}_{dt} = \sum_{o} \pi_{od,t-1} \Delta \log w_{ot}\), and then as the constructed price index is a function of endogenous wages, we instrument the sum of wages with the sum of Bartik shocks: \(\Delta Z_{P_{dt}} = \sum_{o} \pi_{od,t-1} \Delta \text{Bartik}_{ot}\), an inverse trade cost-weighted sum of all Bartik shocks of all other regions.\footnote{In the estimation, we also explore using the trade-share weighted sum of Bartik shocks, \(\Delta Z_{P_{dt}} = \sum_{o} \pi_{od,t-1} \Delta \text{Bartik}_{ot}\), where the shares are computed from the model. The results do not change substantially.} The instrument makes intuitive sense: if all of a location’s close neighbors are locations that have had large wage-growth, then the expected prices for imports
will also increase; as a result, the price index will increase. Figure 3b displays binscatter plots of the log-linearized price index (constructed using actual wages) against $\Delta Z_{P_{dt}}$, which reveals a strong association between these two variables. The coefficient from a regression of $\Delta \log \tilde{P}_{dt}$ on $\Delta Z_{P_{dt}}$ is 1.01 (s.e. = 0.07).

**Instrument for labor**

From Equation 6, we have that $L_{dt} = U_{dt} LMA_{dt}$. LMA itself is a function of WMA. Therefore, a good instrument for labor will be the change in WMA arising from the exogenous wage shocks of all the neighbors that are connected to location $d$. Log-linearizing worker market access yields $\log WMA_{dt} \sim \sum_o \mu_{od,t-1} (\epsilon \Delta \log w_{ot} - \epsilon \alpha \Delta \log P_{ot} - \epsilon (1 - \alpha) \Delta \log r_{ot})$, where $\mu_{od,t-1}$ is the share of people in $o$ that chose to migrate to $d$ in time $t-1$. We therefore use inverse-migration-cost-weighted sums of the neighboring wage shocks to construct the measure of predicted labor inflow, $\Delta Z_{L_{dt}} = \sum_o \hat{k}_{odt}^{-1} \Delta Bartik_{ot}$.\(^{22}\) Intuitively, this instrument measures the fact that a location that is well-connected to another high-wage location could easily send workers away if local wages drop. This means that local labor will be smaller when there are many other options for people to work. Figure 3c displays binscatter plots of the change in labor against $\Delta Z_{L_{dt}}$, which reveals a strong association between these two variables. The coefficient from a regression of $\Delta \log L_{dt}$ on $\Delta Z_{L_{dt}}$ is 1.89 (s.e. = 0.40).

Armed with this set of instruments, we set up the following system of estimating equations in first differences:

\[
\Delta \hat{\delta}^{k}_{dt} = b_{st} + \epsilon(\Delta \log w_{dt} - 0.8 \Delta \log P_{dt} - 0.2 \Delta \log r_{dt}) + \epsilon^{u}_{dt}, \quad (8')
\]

\[
\Delta \log r_{dt} = \eta_{st} + \eta(\Delta \log w_{dt} + \Delta \log L_{dt}) + \epsilon^{r}_{dt}, \quad (9')
\]

where $b_{st}$ and $\eta_{st}$ are state-year fixed effects.

Our identifying restrictions are:

\[
E(\Delta Z_{dt} \epsilon^{u}_{dt}) = 0,
\]

\[
E(\Delta Z_{dt} \epsilon^{r}_{dt}) = 0,
\]

where

\[
\Delta Z_{dt} \in \{ \Delta Bartik_{dt}, \Delta Z_{P_{dt}}, \Delta Z_{L_{dt}} \}.
\]

We estimate the system formed by equations 8 and 9 using a three-stage least squares

\(^{22}\) Again, we check that our results are robust to using the migration-share weighted sum of Bartik shocks, $\Delta Z_{L_{dt}} = \sum_o \mu_{od,t-1} \Delta Bartik_{ot}$. 

21
estimator, allowing for an arbitrary correlation structure between the error terms. We present the results in Table 4. Column (1) displays the results from our baseline model. We estimate a housing supply elasticity of 0.9 and a migration elasticity to wages of 1.91. For the US, Saiz 2010 estimates a population-weighted elasticity of housing supply of 1.75. The migration elasticity to wages is a parameter that has not been extensively estimated, but we can compare our result to those reported by others in the (predominantly US) literature. Monte et al. 2018 estimate an elasticity of 3.30 using commuters. Caliendo et al. 2017 estimate an elasticity of 0.2 for a 5-month frequency. Tombe and Zhu 2015 estimate an elasticity of 2.5 from Chinese data. Our estimate of 1.91 is therefore within the range of these estimates.

Next, we contrast the estimated elasticities we obtain to the elasticities we would obtain in a model without bilateral costs of migration, but where individuals still have taste shocks for different locations. Due to data limitations, many spatial equilibrium models are estimated based on population allocations rather than population flows, and the migration elasticity is then estimated from changes in population in response to economic shocks. However, this yields an identification problem: both high migration costs and a low migration elasticity to wage shocks are consistent with a low observed population response to wage shock. Precisely because migration costs stop some members of the population responding to wage shocks, the migration elasticity may be understated when only population data are used. We show that this is indeed the case for our data by re-estimating the elasticities without bilateral migration costs, in Column (2) of Table 4. Using exactly the same data and the same estimation approach, we estimate an elasticity of migration of 0.5, compared to the point estimate of 1.91 when we use the gross flow data.

23 We present the robustness over these values in Appendix Table 4.
24 Diamond 2016 estimates an elasticity of between 2 and 4, but she does not incorporate origin-destination costs of migration.
25 Without bilateral migration costs, and still assuming Frechet amenity shocks, the probability that an individual will choose to live in \( n \) at time \( t \) is

\[
\mu_{ot} = \frac{B_{ot} U_{ot}}{ \sum_d B_{dt} U_{dt} },
\]

where \( \mu_{ot} \) is the share of total population living in \( o \) at time \( t \).
4.2.3 Trade elasticity

Due to data constraints, we do not estimate $\theta$. Instead, we set $\theta$ exogenously at 4 following Simonovska and Waugh 2014.\(^\text{26}\)

4.3 Implications for elasticities to TMA and LMA

With the estimated elasticities in hand, we can now compute the elasticities of each of the endogenous variables in our model to TMA and LMA. This is reported in Table 5. We do this exercise for the three nested models: (i) the full model with costly trade and costly migration (preference costs and bilateral costs); (ii) costly trade, with migration only a function of preference costs and not bilateral costs; and (ii) the model with immobile labor. This allows an easy way to examine the sensitivity of the key endogenous variables in the model to the elasticities we estimate.\(^\text{27}\)

Starting with (i), our estimates imply that the elasticity of wages to TMA is 0.12 and the elasticity of wages to LMA is -0.12. The elasticity of wages to LMA is always negative; the sign of the elasticity of wages to TMA depends on the relative strength of the migration and trade elasticities (adjusting for housing elasticity and housing expenditure share); for our estimates, this elasticity is positive for all values of the trade elasticity above 1.1. Our estimate of the elasticity of labor to TMA is 0.60; this elasticity remains positive at the point estimates for all reasonable values of the trade elasticity. The elasticity of labor to LMA is always positive; we estimate it to be 0.24. The elasticity of prices to TMA is the inverse of the trade elasticity, -0.25. The elasticity of rents to LMA is always positive because the magnitude of the (positive) elasticity of labor to LMA is always large than the (negative) elasticity of wages to LMA; we estimate it to be 0.43. We estimate an elasticity of rents to TMA of 0.32. Finally, the overall utility of a location depends on the net effects on wages and then the cost of living; we find that wages are decreasing in LMA and the cost of living is increasing in LMA, so both effects lead utility to be decreasing in LMA, with an elasticity of -0.40. For TMA, we find that wages increase in TMA; the cost of living on net is decreasing in TMA (since traded goods are 80% of the expenditure share). The elasticity is 0.48. Column (ii) considers the model where there are only preference costs to mobility, but no origin-destination component. The elasticity of indirect utility to TMA

\(^\text{26}\)We have experimented with a parallel estimation strategy on our trade data as that we use on the migration data. However, we only have trade data at the state level, compared with migration data at the meso-level, and our estimation is under-powered. Results available upon request.

\(^\text{27}\)TMA and LMA measure two distinct concepts. The correlation between estimated TMA and LMA is -0.15.
is 0.16. For the model where labor is immobile, we find an elasticity of utility to TMA of 0.37. These two cases consider the two extreme cases – perfect mobility and no mobility – compared to our benchmark with imperfect migration and trade.

5 Decomposing the effects of roads

We are now ready to answer the question, “What is the relative contribution of improved roads to migration and trade?” To estimate the welfare gains, we take the estimated trade costs and migration costs, recompute them under counterfactual travel times, and then re-solve the model.

5.1 Calculating counterfactual migration and trade costs

To construct the migration and trade costs we use the elasticities estimated from our preferred estimates that we have available, the state-level estimates with pair fixed effects, and assume that the relationship between roads and costs estimated at the state level holds for migration and trade costs at the meso level. While this is a strong assumption, we note that the evidence we presented earlier in the paper does not contradict this assumption: the elasticity of migration to roads was not statistically different between Column (3) in Table 1, estimated off state data, and Column (3) in Table 3, estimated off meso data.28

The first counterfactual is the thought experiment of simply deleting all roads in Brazil. To construct the travel times between all meso regions for the no-roads scenario, we use a constant speed of traveling across all pixels. We then use our earlier gravity estimates to convert this travel time variable into migration and trade cost units. Table 6 shows the resulting counterfactual relative migration and trade costs. Compared to the scenario of no roads, the radial highways reduced trade costs by 28% (the 10th percentile is a 0% reduction; the 90th percentile a 46% reduction) and migration costs by 11% (0%/19% for the 10th/90th percentiles, respectively).

The second counterfactual considers the road network as if, instead of having a national highway system centered on Brasilia, the country had a national highway system...
centered on the old capital city of Rio de Janeiro. To compute this network, we calculate a minimum spanning tree that connects all state capitals to Rio de Janeiro in the least-distance way possible. A picture of the alternative network is given in Appendix Figure 1b.\textsuperscript{29} We rerun the fast marching algorithm on this network to generate the travel times under the Rio counterfactual. Table 6 shows that, compared with keeping the capital in Rio de Janeiro, the highway network that resulted as a result of shifting the capital in Brasilia reduced average trade costs by 8% and average migration costs by 3%. However, the reductions are heterogeneous: 10% of pairs experienced a trade cost increase of at least 23%, and 10% of pairs experienced a migration cost increase of at least 8%.

### 5.2 Solving the model

We first solve the model to recover the underlying parameters. The first step is to use the destination-year fixed effect from the migration gravity equation, which has an interpretation as the bundle of rents, wages, amenities, and prices at a destination. We have data on rents and wages, and we construct data on the price level as discussed above. The unexplained portion of indirect utility is therefore the amenity level of a location.\textsuperscript{30} We therefore have values for all the initial determinants of migration and can solve the migration side of the model. Next, we construct the variables needed to solve for trade flows in the model. We observe wages, but we do not directly observe the productivity, $A_{ot}$, in each location. We solve for the unobserved productivity terms by imposing trade balance.\textsuperscript{31} We then have measures of all the variables determining trade flows and can solve the trade side of the model.

Then, we consider a shock to the system by introducing new trade costs and/or migration costs and find the new equilibrium. To do this, we take the initial value of rents,

\begin{align*}
29&\text{This alternative network bears some similarities to the actual network because the state capital of Goias, Goiania, is located relatively close to Brasilia. The main difference between the two networks is northeastern Brazil: in the current network, there is a road that runs from Brasilia to Fortaleza, passing through the northeast quadrant of the country. The predicted road network for Rio de Janeiro does not have this component.}

30&\text{Rents are determined by a location-specific intercept and a common elasticity to population. We have estimated the elasticity of population to rents, and so we use the initial labor and the elasticity to solve for the location-specific intercept. This intercept picks up any underlying characteristics of a location that are capitalized into housing prices. In addition to this intercept, which we hold constant through all counterfactuals, we have a distinct amenity term that captures any characteristics of a location that affect migration and are not captured in rents. The amenity term is also held constant in all counterfactuals.}

31&\text{Trade balance is the condition that } Y_{ot} = \Sigma_d X_{odt}, \text{ which implies that } w_{ot}L_{ot} = \Sigma_d A_{ot}w_{ot}^{-\theta}x_{otd}^{-\theta} (\Sigma_{o'} A_{o't}w_{o't}^{-\theta}x_{o'td}^{-\theta})^{-1} w_{ot}L_{ot}. \text{ This is a system of } N \text{ equations that, given an observed wage and labor, can be solved for the } N \text{ unobserved productivity terms, } A_{ot}; \text{ one for each location.}

25
wages, and prices. Then, given the exogenous productivities and exogenous amenities, we use the migration costs to construct an updated migration probability. The migration probabilities yield an update for the labor force in each location. Then, given the labor force and exogenous productivities, we use the trade balance equation to solve for wages. This step also yields prices, which are related to own-trade flows. We update rents by using the elasticity of rents to population. This yields an updated value of all of the endogenous values of the system (rents, prices, labor). The exogenous quantities (amenities, productivities) remain unchanged. We then use these updated values as the new starting values and compute the migration probabilities that would be consistent with these prices. We continue this way until the migration probability is stable. This gives the new equilibrium prices and quantities in the economy.

There is an important caveat to these counterfactual exercises. Our model is a static model of migration where individuals make a one-time migration decision. If individuals are forward-looking, it is reasonable to think that the decision of where to live today will also take expectations of future migration into account. We provide an extension of the model to the dynamic case in Appendix C.4. While our estimation strategy is robust to the existence of such dynamic considerations, as the continuation value is captured as part of the amenity term, during the counterfactuals we hold the amenity term fixed at its estimated value. As a result, our counterfactuals may underestimate the cumulative effect of roads by not accounting for the effect of repeated migration decisions.\footnote{Caliendo et al. (2017) consider this issue explicitly in their study of the dynamic effects for the US of increased Chinese import competition.}

5.3 95\% of the gains from improved infrastructure are due to reductions in trade costs

How much of the gain in improving road connectivity came through improvements in the ease of moving? Table 7 shows the equilibrium changes in trade, migration, and utility. The first panel considers the effects of Brasilia compared with no roads. Overall, welfare, normalized to units of a wage-equivalent gain in utility, is 13.3\% higher. The road network caused increase in trade of 24.7\% (measured as the average share of expenditure on non-self-produced goods) and an increase in migration of 11.7\%.

We next separate the welfare effects of roads into the piece caused by goods market integration and the piece caused by labor market integration. Column (2) of the table shows the equilibrium if only trade costs fell and migration costs stayed the same as the
baseline. Not surprisingly, if migration costs do not fall, we do not see as large an increase in migration (0.05%, compared with 11.7%), although migration still increases due to the indirect effects of trade on wages and prices. Overall, the welfare gain is 12.6%, 94% of the gain when both trade and migration costs fall. Column (3) repeats the exercise by changing migration costs only and keeping trade costs at the initial value. Migration increases by 11.3%, and trade falls by 0.02%. The welfare gain is equivalent to a 0.5% increase in welfare. The difference between the measure of welfare and the measure of utility is accounting for the unobserved preference shock. The unobserved preference shock is decreasing in the share of people migrating from an origin to a destination. If the increase in road connectivity causes an intensification of migration into a common destination, then the unobserved preference shock will decrease; if instead road connectivity allows people to migrate to where relatively few people were migrating, then the opposite happens.

The second panel of the table shows the effect of the Brasilia network compared with that of a hypothetical road network if the capital city had remained in Rio de Janeiro. This counterfactual varies both the location and the density of the roads. We find that the road network associated with the change in the capital to Brasilia led to an increase in welfare of 4.8%. Comparing Columns (2) and (3), we find that reducing trade costs only (and holding migration costs constant) would lead to a 4.6% increase in welfare whereas reducing migration costs (and holding trade costs constant) would lead to a 0.02% reduction in welfare.

5.4 It is important to account for bilateral migration costs to get the estimated gain from infrastructure correct

Although reduced trade costs explain 95% of the welfare gains we have reported, this does not mean that costly migration is not important. To demonstrate this point, Column (4) repeats the exercise, assuming that migration costs in the model only depend on preferences for location. The estimated effect of Brasilia is a location-utility gain of 7.2%, 33% lower than the location-utility gain of 11.5% when there is a role for both costly migration and costly trade. That is, in a model where migration also depends on access to infrastructure, the additional benefit raises the estimate of welfare above the baseline calculations. This has implications for any analysis involving optimal road investment: by omitting gains from labor market access, standard estimates understate the net benefits of labor market integration. Column (5) does the same exercise for the model where labor is immobile. We estimate a welfare gain of 11.1%; in this case, labor is immobile,
and so the location-utility matches the gain in welfare. The model with costly trade and migration generates a predicted welfare gain of 13.3%. Labor mobility is an additional channel through which roads affect market integration: not accounting for this leads to underestimates of the benefit of improving connectivity between locations.

5.5 Bilateral migration costs induce heterogeneity in the benefits of roads

Finally, we show that in a model with bilateral migration costs, utility is no longer equalized across space. This is a key equilibrium condition in standard economic geography models where labor can freely move (but people may have location preferences). Our model includes an equilibrium condition that, within origin, all people have the same expected realized utility. Bilateral migration costs induce heterogeneity across people from different origins, and so it is no longer the case that utility is equalized across all origins.\(^{33}\)

Figure 4 shows this graphically. In the model with preference shocks only, utility is equalized across space, with all people gaining an increase of 13.3% in utility from the road network. However, in the model with both migration costs and preference shocks, important spatial heterogeneity exists in the incidence of the shock. The average meso region gains 20% in utility, but the interquartile range of gains is 6%–23%.\(^ {34}\) That is, some regions are gaining much less than the average, while others are gaining much more than the average. If labor is immobile, then labor is not able to arbitrage at all, and so the interquartile ranges are larger. In this case, the average meso region facts a gain of 19%, with the range between 4-25%. Panel (c) shows that, as expected, the gains under costless migration are equally distributed, with every region receiving a 13% gain.

This result has direct implications for policy. Migration costs indicate the extent to which a location is “sticky”: people are differentially affected by any spatial investment depending on where they live. In many countries, including the US, governments invest resources to develop specific areas, including building infrastructure such as roads, but

---

\(^{33}\)The result is that expected utility is constant within origin, even though some may choose to migrate to a destination that has relatively higher wages. This is due to the unobserved taste shock. The taste shocks induce sorting across space – the people who are willing to go to places that do not look as appealing do so because they have a high unobserved preference for that location. Under most standard forms of this heterogeneity (notably, the Frechet and Gumbel distributions), the sorting effect exactly offsets the underlying heterogeneity, leaving utility constant across space. This is the same explanation for why the model without origin-destination migration costs will generate constant expected utility across space even though people live in different locations.

\(^{34}\)These numbers differ from the average gain of 13.3% shown in Table 7 because of weighting. The average meso region is not the average person.
also making broader investments to encourage job creation or economic growth. When migration is costly, there will be heterogeneity in the response to policy for both the regions that are directly affected as well as the regions that are indirectly affected.

6 Conclusion

A large body of literature has studied the effects of roads in facilitating trade. In this paper, we focus on the effects of roads in facilitating the movement of people. Our contribution is to empirically quantify the effects of a large road expansion on both the goods market and the labor market.

The large road expansion we study is the case of Brazil. Brazil relocated its capital city to the interior of the country in 1960 and subsequently built a large highway network connecting the new capital to the state capitals. We generate an instrument for road location based on the straight line connections between the new capital, Brasilia, and state capitals. We first document that states that became more connected by roads had increases in the movement of goods and people, compared with states that did not become more connected by roads. We then use this exogenous variation in migration and trade costs to estimate counterfactuals in a model of costly trade and costly migration (Eaton and Kortum, 2002; Monte et al., 2018).

We find that the road networks that connected Brasilia to the state capitals decreased migration costs by 11% and trade costs by 28%. Overall, these decreased costs increased welfare by 13.3%, of which 95% was the result of the reduction in trade costs and 5% was due to the reduction in migration costs. Although we find that the reduction in trade costs are the dominant source of welfare gains, we also show that it is important to account for costly migration for three reasons. First, without separating out migration costs from migration returns, researchers will underestimate the migration elasticity, a key input into many spatial equilibrium models. We estimate an elasticity of migration to wages of 1.9; with the same data, we estimate an elasticity of 0.5 without accounting for migration costs. Second, the overall estimate of the welfare benefits to improving roads depends on the assumption about how easily people can migrate. Not accounting for migration costs, including preference shocks, we estimate that the road expansion in Brazil increased welfare by 11.5%. Using the same data, we estimate a welfare gain of 7.2%, 37% lower, if we assume that the only friction to labor migration is due to heterogenous preferences, and a gain of 11.1%, 16% lower, if labor is immobile. Third, the spatial equi-
librium arbitrage condition, that expected utility is equalized across space, does not hold when migration is costly. Instead, an amended arbitrage equation, that expected utility is equalized within origin, holds in its place. As a result, we show that the spatial gains from any location-specific investment, such as the construction of new roads, depends on an individual’s origin. We find that the interquartile range of gains from the improved infrastructure is 6%–23%, compared to uniform gains in the absence of origin-destination costs of migrating.

Our paper shows an important role for infrastructure, to facilitate the movement of labor to where its return is highest, that has not been well studied. If labor is not allocated most productively, then aggregate productivity may decrease, along similar lines to studies examining the misallocation of capital (Hsieh and Klenow, 2009). Likewise, costs of adjustment of other mobile factors of production such as capital may also hinder the allocation of resources to where it would be most productive. The aggregate effects of this misallocation, particularly for developing countries where infrastructure is poor, is a potentially important mechanism to further explore.
References


Figures and Tables
Figure 1: Map of straight line instrument and radial highways

Notes: Figure shows Brasilia and the 26 state capitals. The map shows radial highways out of Brasilia and the straight line instrument for roads. The straight line shows the minimum spanning tree instrument between Brasilia and grouped state capitals. Background of map shows meso region boundaries. Source: Authors' calculations based on maps obtained from the Brazilian Ministry of Transportation.
Figure 2: Change in trade and migration against change in travel time

Notes: Source: Census data and authors’ calculations based on maps obtained from the Brazilian Ministry of Transportation.
Figure 3: Instruments for wages, prices, rents

Notes: Units of observations are meso regions. Source: Census data and authors’ calculations based on census data and maps obtained from the Brazilian Ministry of Transportation.
Notes: Figure shows estimated welfare gains by origin, comparing Brasilia roads vs no roads. Unit of analysis is the meso region. Panel (a) shows the estimated welfare is labor is immobile. Panel (b) shows estimated welfare for the model with costly migration and preference shocks. Panel (c) shows estimated welfare for the model with preference shocks only. Source: Authors’ calculations based on census data.

Figure 4: Heterogeneity of impact: Brasilia vs no roads
Table 1: State-level population flows

<table>
<thead>
<tr>
<th>Dep. variable: $M_{odt}$</th>
<th>Without CF</th>
<th>With CF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) b/se</td>
<td>(2) b/se</td>
</tr>
<tr>
<td>Log travel time</td>
<td>-0.482***</td>
<td>-0.366***</td>
</tr>
<tr>
<td></td>
<td>(0.143)</td>
<td>(0.135)</td>
</tr>
<tr>
<td>Dum migrant</td>
<td>4.438***</td>
<td>3.865***</td>
</tr>
<tr>
<td></td>
<td>(1.362)</td>
<td>(1.349)</td>
</tr>
<tr>
<td>Log distance</td>
<td>-1.324***</td>
<td>-1.169***</td>
</tr>
<tr>
<td></td>
<td>(0.221)</td>
<td>(0.222)</td>
</tr>
<tr>
<td>First stage residual</td>
<td>0.296</td>
<td>0.533**</td>
</tr>
<tr>
<td></td>
<td>(0.302)</td>
<td>(0.272)</td>
</tr>
</tbody>
</table>

| Orig x Year FE | yes     | yes     | yes     | yes     |
| Dest x Year FE| yes     | yes     | yes     | yes     |
| Pair FE        | no      | yes     | no      | yes     |
| No. state pairs| 4144    | 4144    | 4144    | 4144    |
| Mean migration rate | 0.12    | 0.12    | 0.12    | 0.12    |
| Mean distance migrated (km) | 894.9   | 894.9   | 894.9   | 894.9   |
| Mean traveltime migrated | 5.81    | 5.81    | 5.81    | 5.81    |

**Notes:** Table shows Poisson regressions. Data is 1940-2010. Measure of migration is the stock of people born in state $o$ living in destination state $d$ in time $t$. The maximum number of observations is 4232 (23 states*23 states*8 years). However, 88 pair-level observations are not defined in 1940, so the actual sample size is 4144. Data source pre-1980: digitized from historical yearbooks. Datasource post 1980: Brazilian census data. Standard errors clustered three-way (origin state, destination state, year).
### Table 2: State-level trade flows

<table>
<thead>
<tr>
<th>Dep. variable: $M_{odt}$</th>
<th>Without CF (1)</th>
<th>(2)</th>
<th>With CF (3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b/se</td>
<td>b/se</td>
<td>b/se</td>
<td>b/se</td>
</tr>
<tr>
<td>Log travel time</td>
<td>-2.497***</td>
<td>-0.585</td>
<td>-4.142***</td>
<td>-4.019***</td>
</tr>
<tr>
<td></td>
<td>(0.582)</td>
<td>(0.436)</td>
<td>(0.668)</td>
<td>(0.767)</td>
</tr>
<tr>
<td>Log distance</td>
<td>0.756***</td>
<td>1.255***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.220)</td>
<td>(0.175)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First stage residual</td>
<td>2.271*</td>
<td>5.171***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.236)</td>
<td>(0.928)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

|                          | yes | yes | yes | yes |
| Orig x Year FE           |     |     |     |     |
| Dest x Year FE           | yes | yes | yes | yes |
| Pair FE                  | no  | yes | no  | yes |
| No. state pairs          | 5975 | 5908 | 5975 | 5908 |

Table 3: Meso-level migration flows

<table>
<thead>
<tr>
<th></th>
<th>Without CF</th>
<th>With CF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Dep. variable: $M_{odt}$</td>
<td>b/se</td>
<td>b/se</td>
</tr>
<tr>
<td>Log travel time</td>
<td>0.18***</td>
<td>0.18***</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Dum migrant</td>
<td>2.26***</td>
<td>2.26***</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.10)</td>
</tr>
<tr>
<td>Log bilateral distance</td>
<td>-1.51***</td>
<td>-1.51***</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>Living in state of birth</td>
<td>4.51***</td>
<td>4.51***</td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td>(0.045)</td>
</tr>
<tr>
<td>First-stage residual</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

No. meso pairs       | 72,900 | 1,968,300 | 72,900 | 1,968,300 |
No. individuals      | 17,477,479 | 17,477,479 | 17,477,479 | 17,477,479 |
Mean migration rate  | 0.071  | 0.071  | 0.071  | 0.071  |
Mean distance migrated (km) | 678.0  | 678.0  | 678.0  | 678.0  |
Mean traveltime migrated | 16.3  | 16.3  | 16.3  | 16.3  |

Notes: $M_{odt}$ is the number of people from origin $o$ in time $t$ who move to destination $d$. All pairs involving state capitals are dropped. The maximum number of observations in Cols (1) and (2) is 46,656 (135 origin regions - 27 state capitals = 108 × 108 destination regions × 4 years). The maximum number of observations in the specifications which control for state of birth (Cols (3) and (4)) is 1,259,712 (108 × 108 × 4 × 27 states (26 states and the federal district of Brasilia)). Robust standard errors reported. Source: Brazilian Census data, 1980-2010.
### Table 4: Structural coefficient estimates

<table>
<thead>
<tr>
<th></th>
<th>(1) Mig costs and pref shocks</th>
<th>(2) Pref shocks only</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b/se</td>
<td>b/se</td>
</tr>
<tr>
<td>Eqn 1: $\Delta \log rent_{dt}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change in log labor</td>
<td>0.90***</td>
<td>0.90***</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Change in log wages</td>
<td>0.90***</td>
<td>0.90***</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Eqn 2: $\Delta \delta_{dt}$ (change in log indirect utility)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change in log wages</td>
<td>1.91***</td>
<td>0.50*</td>
</tr>
<tr>
<td></td>
<td>(0.53)</td>
<td>(0.26)</td>
</tr>
<tr>
<td>Change in log rents</td>
<td>-0.38***</td>
<td>-0.10*</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.053)</td>
</tr>
<tr>
<td>Change in log prices</td>
<td>-1.53***</td>
<td>-0.40*</td>
</tr>
<tr>
<td></td>
<td>(0.43)</td>
<td>(0.21)</td>
</tr>
<tr>
<td>State-Year FE</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>N</td>
<td>264</td>
<td>264</td>
</tr>
</tbody>
</table>

**Notes:** Estimated using 2010-1991 and 1991-1980 differences. Coefficients calculated using three-stage least squares. Equation 1 (rents) uses population-weighted market access and (own-wage) bartik shocks as instruments for changes in labor income. Equation 2 (wages) uses population-weighted market access and bartik-based price index. We impose three constraints on the system of equations: (i) that the coefficient of wage and rents is equal in Equation 1; (ii) that the coefficient of rents is the negative housing expenditure share (20%) times the wage coefficient in Equation 2; (iii) that the coefficient of prices is the negative of 1 - housing expenditure share (i.e., 80%) times the wage coefficient in Equation 2. Source: Brazilian Census data, 1980-2010.
Table 5: Equilibrium market access elasticities

<table>
<thead>
<tr>
<th></th>
<th>(1) Full model</th>
<th>(2) Pref costs only</th>
<th>(3) No migration</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wages</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labor market access</td>
<td>-0.121</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trade market access</td>
<td>0.116</td>
<td>0.169</td>
<td>0.200</td>
</tr>
<tr>
<td><strong>Labor</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labor market access</td>
<td>0.236</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trade market access</td>
<td>0.603</td>
<td>0.853</td>
<td></td>
</tr>
<tr>
<td><strong>Prices</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trade market access</td>
<td>-0.250</td>
<td>-0.250</td>
<td>-0.250</td>
</tr>
<tr>
<td><strong>Rents</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labor market access</td>
<td>0.432</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trade market access</td>
<td>0.315</td>
<td>0.231</td>
<td>0.179</td>
</tr>
<tr>
<td><strong>Indirect utility</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labor market access</td>
<td>-0.397</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trade market access</td>
<td>0.484</td>
<td>0.162</td>
<td>0.364</td>
</tr>
</tbody>
</table>

Notes: Table computes the implied elasticities using the formulae in the paper and estimates of the underlying parameters. Column (1) shows the base model with costly migration and costly trade. Column (2) shows a model with costless migration (preference shocks only) and costly trade. Column (3) shows a model with no migration.

Table 6: Counterfactual migration and trade costs

<table>
<thead>
<tr>
<th></th>
<th>(1) Trade costs mean, p10, p90</th>
<th>(2) Migration costs mean, p10, p90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative cost: Brasilia vs no Brasilia</td>
<td>0.72  {0.54} {1.00}</td>
<td>0.89 {0.81} {1.00}</td>
</tr>
<tr>
<td>Relative cost: Brasilia vs Rio</td>
<td>0.92  {0.77} {1.01}</td>
<td>0.97 {0.92} {1.00}</td>
</tr>
</tbody>
</table>

Notes: Conversion to cost assumes trade elasticity of 4 and migration elasticity of 1.92.
Table 7: Welfare gains of market integration

<table>
<thead>
<tr>
<th></th>
<th>Costly migration &amp; pref shocks</th>
<th>Pref shocks only</th>
<th>No mig</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Both</td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Migration rate</td>
<td>1.117</td>
<td>1.005</td>
<td>1.113</td>
</tr>
<tr>
<td>Trade share</td>
<td>1.247</td>
<td>1.247</td>
<td>0.998</td>
</tr>
<tr>
<td>Wage-eq. utility ((V_i))</td>
<td>1.115</td>
<td>1.110</td>
<td>1.006</td>
</tr>
<tr>
<td>Wage-eq. welfare ((V_{ij}\epsilon_{ij}))</td>
<td>1.133</td>
<td>1.126</td>
<td>1.005</td>
</tr>
</tbody>
</table>

Experiment 1: Brasilia compared to no roads

| Migration rate                   | 1.038                          | 0.998            | 1.041  | 1.001  |        |
| Trade share                      | 1.062                          | 1.063            | 0.999  | 1.075  | 1.059  |
| Wage-eq. utility \((V_i)\)       | 1.040                          | 1.037            | 1.003  | 1.027  | 1.038  |
| Wage-eq. welfare \((V_{ij}\epsilon_{ij})\) | 1.048                          | 1.046            | 1.002  | 1.045  | 1.038  |

Experiment 2: Brasilia compared to Rio MST

Notes: Table shows the effect of the counterfactual experiments. All values are relative to a baseline value of 1. Cols (1)-(3) are from the model with both costly migration and unobserved preference shocks. Col (4) sets the origin-destination component of migration cost to zero but retains the unobserved preference shock. Col (5) considers the case where labor is immobile. Utility is the indirect utility of each location. Welfare is indirect utility, migration cost, and the mean preference shock. Source: Authors’ calculations from census data.
A  Appendix Figures and Tables
Appendix Figure 1: Counterfactual EMST network for Rio de Janeiro
Appendix Figure 2: Bartik shocks (2010-1991): contemporaneous and lagged changes in wages and rents

Notes: Each dot is a meso region. The graphs at the top plot 2010-1991 Bartik shocks and 2010-1991 changes in wages and rents; the graphs at the bottom plot 2010-1991 Bartik shocks and 1991-1980 changes. Source: Authors’ calculations based on census data.
### Appendix Table 1: Summary statistics, by census year

<table>
<thead>
<tr>
<th></th>
<th>Mean/sd</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>35.8</td>
<td>36.4</td>
<td>36.8</td>
<td>37.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(11.7)</td>
<td>(11.5)</td>
<td>(11.3)</td>
<td>(11.7)</td>
<td></td>
</tr>
<tr>
<td>Years schooling</td>
<td>3.80</td>
<td>5.14</td>
<td>6.35</td>
<td>8.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.97)</td>
<td>(4.36)</td>
<td>(4.37)</td>
<td>(4.37)</td>
<td></td>
</tr>
<tr>
<td><strong>Employment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equiv. wage (all)</td>
<td>7.32</td>
<td>5.94</td>
<td>7.96</td>
<td>9.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(17.8)</td>
<td>(14.1)</td>
<td>(29.3)</td>
<td>(66.9)</td>
<td></td>
</tr>
<tr>
<td>Equiv. wage (employee only)</td>
<td>6.88</td>
<td>5.62</td>
<td>6.76</td>
<td>8.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(11.2)</td>
<td>(11.3)</td>
<td>(14.9)</td>
<td>(25.8)</td>
<td></td>
</tr>
<tr>
<td>Share pop. who are employees</td>
<td>0.65</td>
<td>0.62</td>
<td>0.64</td>
<td>0.70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.48)</td>
<td>(0.49)</td>
<td>(0.48)</td>
<td>(0.46)</td>
<td></td>
</tr>
<tr>
<td>Working in agriculture</td>
<td>0.31</td>
<td>0.30</td>
<td>0.23</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.46)</td>
<td>(0.46)</td>
<td>(0.42)</td>
<td>(0.40)</td>
<td></td>
</tr>
<tr>
<td>Working in manufacturing</td>
<td>0.18</td>
<td>0.16</td>
<td>0.15</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.39)</td>
<td>(0.36)</td>
<td>(0.35)</td>
<td>(0.35)</td>
<td></td>
</tr>
<tr>
<td><strong>Housing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean rent</td>
<td>314.1</td>
<td>309.9</td>
<td>342.7</td>
<td>342.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(360.9)</td>
<td>(350.9)</td>
<td>(309.7)</td>
<td>(309.7)</td>
<td></td>
</tr>
<tr>
<td>Share paying rent</td>
<td>0.22</td>
<td>0.15</td>
<td>0.13</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.42)</td>
<td>(0.35)</td>
<td>(0.34)</td>
<td>(0.37)</td>
<td></td>
</tr>
<tr>
<td><strong>Migration</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Municipality migration rate</td>
<td>0.15</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.36)</td>
<td>(0.32)</td>
<td>(0.31)</td>
<td>(0.31)</td>
<td></td>
</tr>
<tr>
<td>Meso-region migration rate</td>
<td>0.095</td>
<td>0.070</td>
<td>0.061</td>
<td>0.053</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.29)</td>
<td>(0.25)</td>
<td>(0.24)</td>
<td>(0.22)</td>
<td></td>
</tr>
<tr>
<td>Missing previous meso</td>
<td>0.014</td>
<td>0.0045</td>
<td>0.0070</td>
<td>0.0045</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.067)</td>
<td>(0.084)</td>
<td>(0.067)</td>
<td></td>
</tr>
</tbody>
</table>
| **Notes:** Summary statistics calculated from Census microdata. Sample is 20-65 year old males with non-zero earnings in main occupation, pooling 1980, 1991, 2000 and 2010. Young defined as below median age. Low skilled defined as below median years of schooling. Financial values in year 2000 Brazilian reals (BRL). 1USD =2.3 BRL.
Appendix Table 2: EMST road network and before-Brasilia population and GDP growth

<table>
<thead>
<tr>
<th></th>
<th>Population growth</th>
<th>GDP growth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td></td>
<td>All b/se</td>
<td>Drop capitals b/se</td>
</tr>
<tr>
<td>On EMST road network</td>
<td>0.024 (0.017)</td>
<td>0.024 (0.018)</td>
</tr>
<tr>
<td>Avg. dep. var</td>
<td>0.14</td>
<td>0.14</td>
</tr>
<tr>
<td>N. of obs</td>
<td>950</td>
<td>924</td>
</tr>
</tbody>
</table>

**Notes:** Units of observations are the minimum comparable areas (AMC) as of 1920. There are 950 amcs. Population and GDP growth from 1940 to 1950. Shortest distances are computed in GIS using the centroids of the geographic units. Cols (1) and (3) include all AMCs. Cols (2) and (4) drop AMCs which contain a state capital. Additional controls: capital city indicator, (log) distance to Brasilia, (log) distance to nearest capital city, and state fixed effects. Standard errors clustered at the state level. Source: IBGE and IPEA Data.
## Appendix Table 3: First stage estimates, meso-level migration flows

<table>
<thead>
<tr>
<th>Dep var: log road traveltime</th>
<th>(1) b/se</th>
<th>(2) b/se</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument x 1980</td>
<td>0.049***</td>
<td>0.049***</td>
</tr>
<tr>
<td></td>
<td>(0.0038)</td>
<td>(0.00072)</td>
</tr>
<tr>
<td>Instrument x 1991</td>
<td>0.029***</td>
<td>0.029***</td>
</tr>
<tr>
<td></td>
<td>(0.0038)</td>
<td>(0.00072)</td>
</tr>
<tr>
<td>Instrument x 2000</td>
<td>0.0065*</td>
<td>0.0065***</td>
</tr>
<tr>
<td></td>
<td>(0.0038)</td>
<td>(0.00072)</td>
</tr>
<tr>
<td>Instrument x 2010</td>
<td>0.019***</td>
<td>0.019***</td>
</tr>
<tr>
<td></td>
<td>(0.0038)</td>
<td>(0.00072)</td>
</tr>
<tr>
<td>Dum migrant</td>
<td>-1.67***</td>
<td>-1.67***</td>
</tr>
<tr>
<td></td>
<td>(0.0088)</td>
<td>(0.0017)</td>
</tr>
<tr>
<td>Log bilateral distance</td>
<td>0.46***</td>
<td>0.46***</td>
</tr>
<tr>
<td></td>
<td>(0.0013)</td>
<td>(0.00026)</td>
</tr>
<tr>
<td>Living in state of birth</td>
<td>-3.7e-18</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.00055)</td>
</tr>
<tr>
<td>No. meso pairs</td>
<td>72,900</td>
<td>1,968,300</td>
</tr>
</tbody>
</table>

**Notes:** Instrument is the log traveltime on the MST network between origin \( k \) and destination \( n \). The maximum number of observations in Cols (1) and (2) is 46,656 (135 origin regions \( \times 27 \) state capitals \( \times 108 \) destination regions \( \times 4 \) years). The maximum number of observations in the specifications which control for state of birth (Cols (3) and (4)) is 1,259,712 (108 \( \times 108 \times 4 \times 27 \) states (26 states and the federal district of Brasilia)). Source: Brazilian Census data, 1980-2010.
Appendix Table 4: Structural coefficient estimates: robustness

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
<td>b/se</td>
<td>b/se</td>
<td>No cost</td>
<td>b/se</td>
<td>No cost</td>
<td>b/se</td>
<td>No cost</td>
<td>b/se</td>
</tr>
<tr>
<td>Cost/birth</td>
<td>b/se</td>
<td>b/se</td>
<td>No cost/birth</td>
<td>b/se</td>
<td>No cost/birth</td>
<td>b/se</td>
<td>No cost/birth</td>
<td>b/se</td>
</tr>
<tr>
<td>Eqn 1: ( \Delta \log rent_{dt} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change in log labor</td>
<td>0.27***</td>
<td>0.27***</td>
<td>0.54***</td>
<td>0.54***</td>
<td>0.90***</td>
<td>0.90***</td>
<td>0.90***</td>
<td>0.90***</td>
</tr>
<tr>
<td></td>
<td>(0.050)</td>
<td>(0.050)</td>
<td>(0.050)</td>
<td>(0.050)</td>
<td>(0.14)</td>
<td>(0.14)</td>
<td>(0.14)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Change in log wages</td>
<td>0.27***</td>
<td>0.27***</td>
<td>0.54***</td>
<td>0.54***</td>
<td>0.90***</td>
<td>0.90***</td>
<td>0.90***</td>
<td>0.90***</td>
</tr>
<tr>
<td></td>
<td>(0.050)</td>
<td>(0.050)</td>
<td>(0.050)</td>
<td>(0.050)</td>
<td>(0.14)</td>
<td>(0.14)</td>
<td>(0.14)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Eqn 2: ( \Delta \hat{\delta}<em>{kt</em>{dt}} ) (change in log indirect utility)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change in log wages</td>
<td>0.36*</td>
<td>0.36*</td>
<td>-0.079</td>
<td>-0.079</td>
<td>1.92***</td>
<td>1.91***</td>
<td>0.50*</td>
<td>0.50*</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.19)</td>
<td>(0.094)</td>
<td>(0.094)</td>
<td>(0.53)</td>
<td>(0.53)</td>
<td>(0.26)</td>
<td>(0.26)</td>
</tr>
<tr>
<td>Change in log rents</td>
<td>-0.072*</td>
<td>-0.071*</td>
<td>0.016</td>
<td>0.016</td>
<td>-0.38***</td>
<td>-0.38***</td>
<td>-0.10*</td>
<td>-0.10*</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td>(0.037)</td>
<td>(0.019)</td>
<td>(0.019)</td>
<td>(0.11)</td>
<td>(0.11)</td>
<td>(0.053)</td>
<td>(0.053)</td>
</tr>
<tr>
<td>Change in log prices</td>
<td>-0.29*</td>
<td>-0.29*</td>
<td>0.063</td>
<td>0.063</td>
<td>-1.53***</td>
<td>-1.53***</td>
<td>-0.40*</td>
<td>-0.40*</td>
</tr>
<tr>
<td></td>
<td>(0.15)</td>
<td>(0.15)</td>
<td>(0.075)</td>
<td>(0.075)</td>
<td>(0.43)</td>
<td>(0.43)</td>
<td>(0.21)</td>
<td>(0.21)</td>
</tr>
<tr>
<td>State-Year FE</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>p value</td>
<td>264</td>
<td>264</td>
<td>264</td>
<td>264</td>
<td>264</td>
<td>264</td>
<td>264</td>
<td>264</td>
</tr>
</tbody>
</table>

Notes: Estimated using 2010-1991 and 1991-1980 differences. Coefficients calculated using three-stage least squares. Equation 1 (rents) uses population-weighted market access and (own-wage) bartik shocks as instruments for changes in labor income. Equation 2 (wages) uses population-weighted market access and bartik-based price index. We impose three constraints on the system of equations: (i) that the coefficient of wage and rents is equal in Equation 1; (ii) that the coefficient of rents is the negative housing expenditure share (20%) times the wage coefficient in Equation 2; (iii) that the coefficient of prices is the negative of 1 - housing expenditure share (i.e., 80%) times the wage coefficient in Equation 2. Source: Brazilian Census data, 1980-2010.
B Data Appendix

B.1 Municipality Level Database

B.1.1 Geographic Units

Municipality boundaries change over time. In order to analyze the same geographical area, we use data aggregated to two geographical regions. The first are the minimum comparable areas (areas minimas comparaveis) constructed by the Institute of Applied Economic Research in Brazil. We refer to these units as AMCs, or municipalities for short hand. There are 3659 AMCs in Brazil in the period 1970 to 2000. The second unit of analysis are meso-regions. Meso-regions are statistical regions constructed by the Brazilian Institute of Geography and Statistics (IBGE). The 3659 were grouped into 137 meso-regions; we merge two of these meso regions together because of overlapping municipality boundaries. This leaves us with a final sample of 135 regions.

We construct a regional database of migration, wages and roads at the municipality level between 1970–2010. Summary statistics for the regional database are presented in Appendix Table 1). The primary datasource is the individual data files from the Brazilian Census, 1970–2010, collected by the Brazilian Institute of Geography and Statistics (IBGE). Our sample of interest is males aged 20–65 who report non-zero earnings in their main occupation. All nominal variables are converted into constant 2010 prices; the exchange rate between the USD and Real is approximately 1 USD = 2.3 BRL. 35

B.1.2 Employment and wages

Wage data are sourced from the census. The census asks both the average earnings per month in the main occupation,36 as well as the usual hours worked. We use earnings from main occupation and the hours worked to construct an equivalent hourly wage rate. This wage rate is 7.6 BRL on average. This average wage matches well to GDP estimates. Assuming a

---

35 We constructed a modified consumer price index that accounts for changes in the Brazilian currency that occurred within the period under analysis. All nominal variables were converted to 2010 BRL. See http://www.ipeadata.gov.br/ for the factors of conversion for the Brazilian currency.

36 The exception is 1970, where only total earnings, rather than earnings in the main occupation, is asked.
standard 2000 hour work year, the annual wage of 7.3 BRL in 1970 and 9.0 BRL in 2010 would be equivalent to annual incomes of $3000 and $7800. The per capita GDP figures for Brazil are $2400 in 1970 and $5600 in 2010 (World Development Indicators).

Nearly 65% of the population report being employees rather than self-employed. The share working in agriculture is about 26%. The high proportion of self-employed people, particularly in agriculture, may generate concerns that the wage we compute does not accurately reflect actual income. To check for this issue, we use detailed municipality level agriculture input and output data collected in agricultural censuses to show that self-reported income in the population census is highly correlated at the municipality level with agricultural profits (For details, see Appendix B.6).

B.1.3 Migration

The current location of the individual is coded to the municipality level. From 1980, location 5 years ago is also coded to the municipality level. We are able to match the previous location at the municipality for 99.2% of the population (96% of people who report living in a different municipality 5 years ago.)\(^{37}\) The inter-municipality migration rate is 12% in our sample. Of these moves, 60% (i.e. a migration rate of 7.2%) were between meso-regions.

A focus of our paper is to examine the spatial equilibrium of migration in a model with many locations. This is important because internal migration is more complex than simply rural-urban migration: using these data, 16% of all migrants are rural-rural migrants in 1980; 41% are urban-urban migrants; 35% are rural-urban; and 6.8% are urban-rural (numbers not reported in table but available upon request from authors). Our spatial model will capture the heterogeneity in migrant destination by studying the locational choice over \(d\) locations.

\(^{37}\)For the other 4%, the location is given at the state, not municipality, level. Fewer than 0.05% of the population report living abroad 5 years previously, so we ignore international migration.
B.1.4 Rental prices

To convert nominal wages into real wages, we need to construct measures of the cost of living across space. Unfortunately, consumer price data is not collected at the municipality level. We instead construct costs of living using the best data sources available: a consumer price index collected at 10 cities in Brazil, and housing prices collected in the population census.

The national consumer price index is a data series collected by IBGE for 10 locations across Brazil. For each AMC, we merge to the closest price collection point. In the analysis, we will make an adjustment sourced from equations linking the ability of a region to trade with other regions and source cheaper products to generate a measure of the change in price indices. Second, we use rental rates from the population census. Approximately 17% of our sample report paying rent to live in their accommodation.

For rental rates we use census data on the rents paid for housing. The mean rental rate for one bedroom is 321 BRL a month, equivalent to 42 hours of work at the mean wage. We show in Appendix B.7 that rental rates are positively correlated with the relative price index. 17% of the population report paying rents for their housing. While this may seem low, the equivalent number for US houses in 2005 is 24%. We run the estimation under several different definitions of the rental variables, including hedonic pricing to impute the cost of non-rented units, and find that the results are robust across definitions of the housing cost variable.

B.2 Historical state-to-state trade and migration flows

We draw state-to-state trade flow data from the statistical yearbooks produced by the IBGE. These data are available annually, spanning the periods 1942–1949 and 1967–1974. The yearbooks report the value of total exports of each state to other states across the country. The data was sourced by the Technical Council of Economics and Finance.

Data on state-to-state migration flows are also available from the statistical yearbooks on a decennial basis for the years 1940-1980. The books report the number of residents in all states by state of birth. Therefore, we are able to construct these flows for origin of birth. The data come from the decennial Censuses conducted by the IBGE.
For the year 1999, interstate bilateral trade flow data are derived from information on state tax on the movement of goods and services (*Imposto sobre Circulacao de Mercadorias e Servicos*). We use the study produced by de Vasconcelos (2001) as a data source.

### B.3 Road data

Our geographic data come from two sources. We obtained vector-based maps from the highway network for the period 1940 to 2000 from the Brazilian Ministry of Transportation. These maps were constructed based on statistical yearbooks from the Ministry’s Planning Agency, previously known as GEIPOT. We used the ArcGIS software to georeference the maps to match real-world geographic data. The geographic coordinate system applied to the maps is the SIRGAS 2000.

The second source of data is the IBGE, which provides municipality boundaries maps in digital format. We use the municipality boundaries from 2000 and apply the crosswalk that maps the municipalities that existed in 2000 into AMCs. We then aggregate the AMCs up to meso regions. Because of the overlap with AMC boundaries, we need to combine two sets of two meso regions, creating 135 adjusted meso regions. This will be our primary unit of analysis for the spatial component. 38

Similar to the road data, we applied the coordinate system SIRGAS 2000 to the AMC and meso-region boundaries. Finally, in order to compute geographic distances in kilometers, we projected the maps using the Brazil Mercator projection.

### B.4 Bilateral road travel time

To construct measures of the distance between origin-destination pairs taking into account the actual road coverage, we use the fast marching algorithm, following the approach used in Allen and Arkolakis (2014). The fast marching algorithm finds the solution to the Eikonal equation used to characterize the propagation of wave fronts. The algorithm uses a search

---

38The crosswalk file can be obtained from http://www.ipeadata.gov.br/. For cases where is overlap between the AMC and the meso region, we assign the AMC to the meso region which has the largest number of 2010 component municipalities. We then group together Madeira-Guapore and Leste Rondoniense (both in Rondonia) and Sul de Roraima with Norte de Roraima (two meso regions in Roraima).
pattern for grid points in computing the arrival times (distances) that is similar to the Dijkstra shortest path algorithm (Hassouna and Farag (2007)). However, because the fast marching algorithm is applied to a continuous graph, it reduces the grid bias and generates more accurate bilateral distances.

First, we generate a picture of the road network and the location of the 135 meso-regions. This picture is converted into pixels, and a travel speed is assigned to each pixel. Pixels corresponding to a paved road are assigned a travel speed of 100, whereas pixels outside the road network are assigned a travel speed of 10. Essentially, this algorithm finds the shortest route, traveling on roads, between two locations, with the minimum off-road traveled to connect a region without a road to the road network. The outcome is a 135x135 matrix where each entry corresponds to the fastest arrival time between an origin-destination pair. We undertake the same exercise for our predicted highway system (the EMST network) to find an instrument for the actual bilateral cost using the travel time on the exogenous road network.\footnote{See Section 2 for details on the EMST network.}

\subsection*{B.5 Minimum Spanning Tree network}

We use ArcGIS to compute the EMST network. First, we use the latitude-longitude coordinates to create point features representing the location of Brasilia and the 26 state capitals. Next, we divide the country into 8 exogenous slices, and consider the optimal network connecting the cities within each slice. We do this to avoid exogenous choice in which capital cities were connected to Brasilia. We proceed by creating an imaginary pie sliced into eight parts and centered around Brasilia. We form eight 45 degree slices. Then, we classify the 26 state capitals into eight groups, according to the location of their bearing with respect to Brasilia. We use the Spanning Tree Tool, in Arcmap, to find the minimum spanning tree connecting the states in each of the eight groups.\footnote{The tool uses Prim’s algorithm to design the euclidean minimum spanning tree.}
B.6 Self-reported agricultural income

In the Brazilian census, between 50–70% of the sample who report working in agriculture are self-employed rather than employees. Self-reported income in censuses may not accurately reflect agricultural wage income for at least three reasons: i) self-reported income may be revenues, rather than income; ii) it may contain payments to both labor as well as other factors of production such as capital; or iii) it may be more accurately provided at the household, rather than the individual, level (Lagakos et al. (2018)). In this section we use data from Brazilian agricultural censuses and present evidence that, despite the potential problems, agriculture self-employment income as measured in population censuses highly correlates with agriculture profits computed from agricultural censuses. In addition, we run the reduced form analysis in the paper both including and excluding non-employees, and results are robust.

Starting in 1970, the Agriculture Censuses were collected every five years. The agriculture census allows us to measure agricultural income accurately, as it covers the universe of agricultural production unities, regardless of their size, output level, or location. It is worth mentioning that home gardens were not considered as agricultural unities for the purpose of data collection. Nonetheless, we believe that we only miss some of the production for own consumption of those who work mainly outside the agricultural sector.

We obtain the series of agriculture revenues and expenses at the AMC level from IPEADATA. Agriculture revenue comprises proceeds from the sale of agricultural products, including final goods produced inside the agricultural unities, as well as revenues from the rental of land and livestock and services rendered to third parts. Agriculture expenses include expenses with wages, rents, other inputs, and operational expenses. Our benchmark measure of AMC-level agricultural income is agricultural profits, as measured by the difference between revenues and expenses. We used the years 1975, 1980, and 1996, which are the closest to the population census years (1970, 1980, and 1991).

Appendix Figure 3 displays the scatterplot of the agriculture (log) profits obtained from the agriculture census against the agriculture self-employment

---

41 The share of the population working in labor force declines from 46% in 1970 to 22% in 2010.
42 The agricultural censuses include unities located in urban areas.
(log) income computed from the population census. The two income measures are positively correlated. The R-squared from regressing the level of agriculture self-employment income on the level of agriculture profits indicates is 0.80.

Appendix Figure 3: Comparison: population and agricultural censuses
Appendix Table 5: Correlation of CPI with other measures of cost of living

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dep var: Relative price index</td>
<td>b/se</td>
<td>b/se</td>
</tr>
<tr>
<td>Mean agricultural prices (producer)</td>
<td>0.026**</td>
<td>(0.0097)</td>
</tr>
<tr>
<td>Mean rental rate</td>
<td>0.0037</td>
<td>(0.015)</td>
</tr>
<tr>
<td>Year FE</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>22</td>
<td>40</td>
</tr>
</tbody>
</table>

Notes: Each observation is a municipality-year. The CPI is collected at 10 locations in Brazil. For each year, we normalize the mean of the index to 1, so the index measures spatial variation in the cost of living. Agriculture prices are available for 1980, 1991 and 2000. Rents are available for 1970, 1980, 1991 and 2010. Standard errors clustered at the municipality level.

B.7 Cost of living

Consumer prices are only collected at 10 cities in Brazil. In this section, we show how the prices correlate with two measures of the cost of living: mean rental rates from the population census, and producer prices at the municipality level computed from the agricultural census. The dependent variable is the price index, normalized each year to have value 1. The agricultural price index is a weighted average of the prices of the 4 main agricultural crops (soy, sugarcane, coffee, and corn), sourced from the agricultural census. The rental rate is the mean rental rate per bedroom, sourced from the population census. Appendix Table 5 shows that both are positively correlated with the relative price index, although the small sample size means that the rental rate is not statistically significant.43

43 Additionally, the CPI is only collected in cities; as a result, there is less variation in rental rates that in the entire sample. The variable of (log) rental rates in the municipalities included in the CPI sample is 0.49, compared with a variance of 0.84 across all municipalities.
C Theoretical derivations

C.1 Log-linearization of TMA, WMA

\[
TMA_{ot} = \sum_o A_{ot} \omega_{ot}^{-\theta} \tau_{odt}^{-\theta} \\
WMA_{ot} = \sum_d B_{dt} \omega_{dt}^\epsilon \left( P_{dt}^{\alpha} r_{dt}^{1-\alpha} \right)^{-\epsilon} \kappa_{odt}^{-\epsilon}
\]

We take a first-order Taylor expansion around \( (A_{o,t-1}, \omega_{o,t-1}) \), and use the fact that \( \frac{A_{o,t-1} \omega_{o,t-1}^{-\theta} \tau_{odt}^{-\theta}}{\sum_o A_{o,t-1} \omega_{o,t-1}^{-\theta} \tau_{odt}^{-\theta}} = \pi_{od,t-1} \), the share of \( o \)'s imports from \( o \) in period \( t-1 \). This log-linearization assumes that migration costs and trade costs are constant.

\[
\log(TMA_{ot}) = \log \left( \sum_o A_{ot} \omega_{ot}^{-\theta} \tau_{odt}^{-\theta} \right) \\
\frac{TMA_{ot} - TMA_{o,t-1}}{TMA_{o,t-1}} = \sum_o \pi_{od,t-1} \left( \frac{A_{ot} - A_{o,t-1}}{A_{o,t-1}} - \theta \frac{\omega_{ot} - \omega_{o,t-1}}{\omega_{o,t-1}} \right) \\
\Delta \log TMA_{ot} = \sum_o \pi_{od,t-1} (\Delta \log A_{ot} - \theta \Delta \log \omega_{ot})
\]

And we can do the same thing for WMA, where we log-linearize around \( (\omega_{d,t-1}, P_{d,t-1}, r_{d,t-1}) \), and using the fact that \( \frac{B_{d,t-1} \omega_{d,t-1}^\epsilon \left( P_{d,t-1}^{\alpha} r_{d,t-1}^{1-\alpha} \right)^{-\epsilon} \kappa_{od,t-1}^{-\epsilon}}{\sum_d B_{d,t-1} \omega_{d,t-1}^\epsilon \left( P_{d,t-1}^{\alpha} r_{d,t-1}^{1-\alpha} \right)^{-\epsilon} \kappa_{od,t-1}^{-\epsilon}} \) is the migration probability, \( \mu_{od,t-1} \):

\[
\log(WMA_{ot}) = \log \left( \sum_d B_{dt} \omega_{dt}^\epsilon \left( P_{dt}^{\alpha} r_{dt}^{1-\alpha} \right)^{-\epsilon} \kappa_{odt}^{-\epsilon} \right) \\
\frac{WMA_{ot} - WMA_{o,t-1}}{WMA_{o,t-1}} = \sum_d \epsilon \mu_{od,t-1} \left( \frac{\omega_{dt} - \omega_{d,t-1}}{\omega_{d,t-1}} - \alpha \frac{P_{dt} - P_{d,t-1}}{P_{d,t-1}} - (1 - \alpha) \frac{r_{dt} - r_{d,t-1}}{r_{d,t-1}} \right) \\
\Delta \log WMA_{ot} = \sum_d \epsilon \mu_{od,t-1} (\Delta \log \omega_{dt} - \alpha \Delta \log P_{dt} - (1 - \alpha) \Delta \log r_{dt})
\]

C.2 Log-linearization of the price index

The price index is location \( d \) is given by \( P_{dt}^{-\theta} = \kappa_{p} TMA_{dt} \). Therefore, we can apply the log-linearization for TMA above, which yields:
\[
\Delta \log P_{dt} = \sum_o \pi_{od,t-1} \left( \Delta \log w_{ot} - \frac{1}{\theta} \Delta \log A_{ot} \right)
\]

### C.3 Decomposition of migration costs

From the assumption of Frechet distribution

\[
\mu_{odt} = \frac{(V_{dt}/\kappa_{odt})^e}{\Phi_{ot}},
\]

where \(\mu_{odt}\) is the probability that a worker will migrate from \(o\) to \(d\) at time \(t\), \(V_{dt} = \frac{B_{dt}^{1/e}w_{dt}}{p_{dt}^{1/e}}\) and \(\Phi_{ot} = \sum_{s \in N}(V_{ot}/\kappa_{ost})^e\). Additionally

\[
E(b_{dt}) = \Gamma \left(1 - \frac{1}{e}\right) = \bar{\Gamma}.
\]

This implies that expected utility is the same across all destinations for people from the same origin

\[
E\left(\frac{V_{dt}b_{dt}}{\kappa_{odt}} | \text{choose } d\right) = \bar{\Gamma} \frac{V_{dt}}{\kappa_{odt}} \mu_{odt}^{-1/e}
\]

\[
= \bar{\Gamma} \frac{V_{dt}}{\kappa_{odt}} \left(\frac{(V_{dt}/\kappa_{odt})^e}{\Phi_{ot}}\right)^{-1/e}
\]

\[
= \bar{\Gamma} \Phi_{ot}^{1/e}.
\]

The unconditional expected utility from staying in \(o\) at time \(t\) is

\[
E\left(\frac{V_{ot}b_{ot}}{\kappa_{oot}} \right) = \frac{V_{ot}}{\kappa_{oot}} E(b_{ot}) = \frac{V_{ot}}{\kappa_{oot}} \bar{\Gamma}.
\]

Additionally,

\[
E\left(\frac{V_{ot}b_{ot}}{\kappa_{oot}} \right) = \mu_{oot} E\left(\frac{V_{ot}b_{ot}}{\kappa_{oot}} | \text{choose } o\right) + (1 - \mu_{oot}) E\left(\frac{V_{ot}b_{ot}}{\kappa_{oot}} | \text{don't choose } o\right)
\]

\[
\frac{V_{ot}}{\kappa_{oot}} \bar{\Gamma} = \mu_{oot} \bar{\Gamma} \Phi_{k}^{1/e} + (1 - \mu_{oot}) E\left(\frac{V_{ot}b_{ot}}{\kappa_{oot}} | \text{don't choose } o\right)
\]

Therefore

\[
E\left(\frac{V_{ot}b_{ot}}{\kappa_{oot}} | \text{don't choose } o\right) = \frac{1}{(1 - \mu_{oot})} \frac{V_{ot}}{\kappa_{oot}} \bar{\Gamma} - \frac{\mu_{oot}}{(1 - \mu_{oot})} \bar{\Gamma} \Phi_{ot}^{1/e}.
\]
Remember that
\[ \mu_{oot} = \left( \frac{V_{ot}}{\kappa_{oot}} \right)^{\epsilon} \Rightarrow \frac{V_{ot}}{\kappa_{oot}} = \left( \mu_{oot} \Phi_{ot} \right)^{\frac{1}{\epsilon}} \]

Then,
\[
E\left( \frac{V_{ot} b_{ot}}{\kappa_{oot}} \mid \text{don’t choose } o \right) = \frac{\mu_{oot}^{\frac{1}{\epsilon}}}{(1 - \mu_{oot})} \Gamma \Phi_{ot}^{\frac{1}{\epsilon}} - \frac{\mu_{oot}}{(1 - \mu_{oot})} \Gamma \Phi_{ot}^{\frac{1}{\epsilon}}
\]

\[= \Gamma \Phi_{ot}^{\frac{1}{\epsilon}} \left[ \frac{\mu_{oot}^{\frac{1}{\epsilon}} - \mu_{oot}}{(1 - \mu_{oot})} \right] \]

Now we can decompose the relative gain from migrating to \(d\) from \(o\) in time \(t\)

\[
\frac{E\left( \frac{V_{dt} b_{dt}}{\kappa_{oot}} \mid \text{choose } d \right)}{E\left( \frac{V_{ot} b_{ot}}{\kappa_{oot}} \mid \text{don’t choose } o \right)} = \frac{\Gamma \Phi_{ot}^{\frac{1}{\epsilon}}}{\Gamma \Phi_{ot}^{\frac{1}{\epsilon}}} \left[ \frac{\mu_{oot}^{\frac{1}{\epsilon}} - \mu_{oot}}{(1 - \mu_{oot})} \right]
\]

\[
\frac{V_{dt}}{V_{ot}} \frac{\kappa_{oot}}{\kappa_{oot}} \frac{E\left( b_{dt} \mid \text{choose } d \right)}{E\left( b_{ot} \mid \text{don’t choose } o \right)} = \frac{1 - \mu_{oot}}{\mu_{oot}^{\frac{1}{\epsilon}} - \mu_{oot}}
\]

So, the total migration cost is given by:

\[
\frac{\kappa_{oot} E\left( b_{ot} \mid \text{don’t choose } o \right)}{\kappa_{oot} E\left( b_{dt} \mid \text{choose } d \right)} = \frac{V_{dt} \mu_{oot}^{\frac{1}{\epsilon}} - \mu_{oot}}{V_{ot} 1 - \mu_{oot}}
\]

**C.4 Dynamics**

One other important component of the migration decision may be dynamic in nature: a location has benefits both today, but also in the future, given that the individual should be expecting to re-optimize location in the following period. While the main focus of our analysis is through the lens of a static model, it is easy to extend the model to incorporate dynamics. Our estimation strategy is robust to the presence of a dynamic component of utility (the "continuation value", below). However, our counterfactuals do not account for any dynamic benefits of roads. In that sense, our counterfactuals are an underestimate of the cumulative effect of roads through repeated migration decisions.
Following \textsuperscript{44} at a given time $t$, the (log) utility flow that worker $i$ living in $o$ and moving to $d$ enjoys is:

$$
\log U_{odt}(i) \equiv \hat{U}_{odt}(i) = \tilde{B}_{dt} + \tilde{\omega}_{dt} + \tilde{b}_{dt}(i) - \kappa_{odt},
$$

where $\tilde{B}_{dt} = \log B_{dt}$, $\tilde{\omega}_{dt} = \log \omega_{dt} - \alpha \log P_{dt} - (1 - \alpha) \log r_{dt}$ and $\tilde{b}_{dt}(i) = \log b_{dt}(i)$. Since $b_{dt}(i)$ has a Frechet distribution, $\tilde{b}_{dt}(i)$ has a Gumbel distribution. In a dynamic model, workers also take into account the expected future value of living in the destination $d$ given their information set at time $t$, $\beta E_t V_{d,t+1}$. Therefore, the gross flow of workers that migrate from $o$ to $d$ at time $t$ is given by:

$$
M_{odt} = \frac{\exp(\tilde{B}_{dt} + \tilde{\omega}_{dt} + \beta E_t V_{d,t+1} - \kappa_{odt})}{\sum_{s \in N} \exp(\tilde{B}_{st} + \tilde{\omega}_{st} + \beta E_t V_{s,t+1} - \kappa_{ost})} \times L_{ot}. \quad (11)
$$

The estimating equation from this model is exactly the same as Equation (??). The only difference from the original model is the continuation value $\beta E_t V_{d,t+1}$. The continuation value is isomorphic to an amenity value of location $d$ and is included in the fixed effect terms, denoted as $\hat{\delta}_d^\kappa$:

$$
\log M_{odt} = \hat{\delta}_d^\kappa + \hat{\delta}_o^\kappa - \epsilon \mu \log(\text{road travel time}_{odt}) + \epsilon_{odt}^\kappa.
$$

\textsuperscript{44} adopts this methodology to study the dynamic effects in the US from increased trade with China. \textsuperscript{45} assumes that a worker starting in location $o$ at time $t$ enjoys the real wage at the origin, $\tilde{\omega}_{ot}$ and then pays the cost of migrating to the destination $d$. Unlike them, we assume that the worker who chooses to migrate from $o$ to $d$ at time $t$ enjoys the real wages at the destination, $\tilde{\omega}_{dt}$.