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

Melanie Morten †
Stanford University and NBER

Jaqueline Oliveira ‡
Rhodes College

March 6, 2023

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. Using a general equilibrium trade model and rich spatial data, we explore the impact of a large plausibly exogenous shock of highways in Brazil on both goods markets and labor markets. We find that the road improvement increased welfare by 2.8%, of which 76% was due to reduced trade costs and 24% to reduced migration costs. Costly migration is responsible for significant spatial heterogeneity in the benefits of roads: the range of welfare improvement is 1%–15%, as opposed to uniform gains with perfect mobility.

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

*We thank the editor, anonymous referees, and those who offered comments: 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. We additionally thank seminar participants at numerous seminars and conferences for their feedback. Anita Bhide, Caue Dobbin, Betsenat Gebrewold, Devika Lakhote, and Cecilia Moreira provided excellent research assistance. 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 movement of goods,[1] while ignoring how roads and highways may also facilitate the movement 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 take advantage of an empirical setting with two features. The first feature is plausibly exogenous variation in the location of the road network. This feature is key because roads are generally not built randomly; instead, they are built to connect places where there may be high (or low, if the policy goal is to stimulate) demand for either trade or migration. The second feature is availability of gross migration flow data, not simply net population changes. Under standard gravity model assumptions, gross flow data allow us to separate out the effect of roads on bilateral migration costs from the effect of roads on prices because we can undertake the analysis within origin-destination pair, hence controlling for origin-year and destination-year fixed effects. In other words, we are able to estimate the elasticity of migration to roads while controlling for wages or prices at either the destination or the origin.

Brazil is an empirical setting that meets both requirements. In 1960, Brazil built a new capital city, Brasilia, and began constructing a highway network radiating out from the new capital to connect to other state capitals. The roads radiating out from Brasilia are known as radial highways. The actual location of the radial highways and the choice of the cities connected on the highway between Brasilia and a state capital could have been dictated by economic factors. We 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 network therefore changes the travel time between states, potentially changing inter-state trade and migration flows. Using data from before and after the construction of the radial highway network, we show that travel times decreased and migration rates increased relatively more between states that were better connected as a result of the road expansion. We also show that trade increased

[1] See, for example, Michaels 2008; Banerjee et al. 2020; Duranton and Puga 2014; Faber 2014; Ghani et al. 2014; Donaldson 2018; Donaldson and Hornbeck 2016.

[2] The construction of this straight-line instrument follows Chandra and Thompson 2000 and Michaels 2008.

relatively more between these well-connected states. We establish these results by estimating gravity equations that control flexibly for shocks at the levels of destination-year, origin-year, and pair, thus identifying the elasticity of migration or trade to roads only from the within-pair change in travel time. We take this as evidence that roads facilitate both migration and trade.

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 a 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 later return home to visit friends and family after migrating, migration costs also capture the flow costs of such return visits (with both fiscal and time components). Migration costs can also capture any costs of not being able to consume the same types of goods as at home.[3] Empirically, we use geocoded data on migration choices and show that the way people make migration decisions is consistent with it being more difficult to move to a place 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). This framework has much in common with trade frameworks that consider how trade increases the market access of a location (Donaldson and Hornbeck, 2016). In our framework, roads facilitate improvements of economic activity through two channels. First, reductions in trade costs facilitate the movement of cheaper goods to a location. Second, reductions in migration costs facilitate the relocation of labor to higher-welfare (i.e., more productive or higher amenity) locations. The overall welfare effects of roads thus depend on the relative strength of the reduction in trade frictions and the reduction in migration frictions, combined with three additional 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.

To estimate the additional structural elasticities, we follow the approach of Diamond 2016 by decomposing the fixed effect of the gravity equation, structurally interpreted as the bundle of wages, amenities, and prices at the destination. We construct instru-

[3] For example, Atkin (2016) documents that internal migrants in India pay a "caloric tax" to keep eating the types of food they ate in their origin states when those foods are less available in destination states.

ments for wages, goods prices, and labor from labor demand shifters (Bartik, 1991) that interact with measures of access to labor markets (through migration) and goods markets (through trade). This procedure yields an estimate of the migration elasticity to wages of 4.5 and an estimate of the housing elasticity of 1.7. We set the trade elasticity to 4 following Simonovska and Waugh 2014.

With these elasticities in hand, we perform two counterfactual exercises. 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 9% and increasing migration costs by 8%. We find that this counterfactual would have decreased welfare in Brazil by 2.8%. Of this reduction, 76% would be due to the reduction in goods market integration and 24% attributable to reduced labor market integration.

What drives the large role of trade in explaining the overall welfare gain from road expansion? Empirically, from the gravity equation we see that trade flows are more responsive to changes in road travel time than migration is, suggesting that trade costs are more responsive to distance than migration costs. We confirm this intuition when we convert the gravity equations, dividing by either the trade or migration elasticity, to compute the implied reduction in costs. Our estimates imply that trade costs are approximately 20% more responsive to changes in travel time than are migration costs. However, we also show that even if the reduction in trade costs was equal to the reduction in migration costs, trade would still explain more of the welfare gains. In other words, trade explains a larger share of the welfare gains partly because we estimate that trade costs fell relatively more than migration costs after the construction of Brazil's road network, and partly because trade is a relatively larger share of the economy than migration at baseline.

Finally, while we find that trade represents 76% of gains from the roads, accounting for costly migration is important to understanding spatial heterogeneity in road benefits. Where models using free labor mobility would predict that the benefits of roads were equally distributed across space, we estimate the regional-level range of welfare at 1–15%.

Our second exercise is computing the welfare effects of a hypothetical road network had the capital remained in Rio de Janeiro instead of shifting to Brasilia. This counterfactual builds the road highway in different parts of Brazil, with relatively fewer roads in the center and northeast of the country in contrast to the road network that extended out from Brasilia. We find that the road network constructed around Brasilia increased welfare by 0.3% relative to the alternative network constructed around Rio de Janeiro.

Our paper relates to several strands of the literature. A rich spatial literature examining issues of migration 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, 2018; Topalova, 2010). Recent papers have relaxed this assumption to consider costly migration or commuting (Caliendo et al., 2019; Monte et al., 2018; Tombe and Zhu, 2019). We follow this latter approach, writing a model with costly migration and costly trade. Relative to other papers considering the joint determination of trade and migration, the focus of our work is to explicitly consider the role of roads in facilitating both migration and trade.[4]

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 of migration costs in partial equilibrium settings (Kennan and Walker, 2011; Bryan et al., 2014). We supplement this literature by considering both the responsiveness of migration to both costs and returns and the general equilibrium effects roads have on migration and trade.[5]

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, 2016) to migration, we focus on travel time as a barrier. Our model is one where labor is always optimally located given the costs of migrating. If migration costs reduce, migration increases. Our paper presents evidence that migration costs can be considerably reduced with improved access to transportation infrastructure, facilitating migration and increasing welfare.[6] This finding suggests there is latitude for policymakers to improve the allocation of labor across space through investments in infrastructure.

While a key contribution of our work is to calculate the relative effect of roads on

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

[5]Chein 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 2020 also use the construction of Brasilia as an instrument to study the effect of road construction on regional GDP. Their study does 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.

[6]Whether productivity increases after a reduction of migration costs will depend on the relationship between productivity and amenities. Because people maximize welfare, not just income, when choosing where to live, a reduction in migration costs can either increase, decrease, or leave productivity unchanged. Bryan and Morten (2019) consider the effect of reducing migration costs on productivity when individuals have comparative advantage for different locations.

goods and labor market integration, this paper has several limitations. In order to make direct comparisons with other studies, we use a static model of migration. Recent work in trade has considered dynamic approaches to migration (Artuç et al., 2010; Caliendo et al., 2019); including a dynamic component in the model would allow an additional channel for long-run adjustment, something that we do not consider.[7] In addition, our model does not explicitly attend to endogenous agglomeration or congestion forces (although we do allow for endogenous cost-of-living effects in the housing market, which are similar to endogenous congestion forces). This is because our primary focus is on understanding the additional effects of roads on labor market integration in contrast to cases where this is not considered.

The plan of the paper is as follows. In Section 2, we explain how we use a natural experiment to provide exogenous variation in the road network, including its basis in the historical context leading to the construction of Brasilia. We present our 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. Section 6 offers a brief conclusion.

2 Do roads affect trade and migration?

To answer this question empirically, we leverage bilateral migration and trade flow data before and after the building of Brazil's new capital city, Brasilia, comparing the differences in outcomes across state pairs more or less connected through a least-cost road network centered around the new capital. We establish that, within state pairs, migration and trade increased more between places that were better connected by the Brasilia-induced road network. This section lays out the empirical analysis in detail.

2.1 Data

We source historical data on state-to-state trade and migration by digitizing data reported in statistical yearbooks, and, for the period after 1970, census micro-data on migration. We consider the state as the unit of analysis, aggregating together those states with boundary changes between 1940 and 2000. The total number of spatial units is 21 after we account

[7] In 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.

for changes in administrative boundaries during the analysis period. Migration data, available decennially from 1940 to 2000, measures the total number of people living in each state, disaggregated by state of birth. The maximum possible number of state-state-year observations on bilateral migration is $21 \times 20 \times 7 years = 2,940$. We also use migration data from the Brazilian census measuring moves between mesoregions, a smaller geographic unit, over the last five years. The meso-level migration data are available for the years 1980-2000.

Trade data for each year spanning the periods 1942–1949, 1959–1974, 1985, and 1998–1999 are created from information on the yearly value of imports for each state, disaggregated by state of origin. Data for the years 1998 and 1999 are from de Vasconcelos 2001; for all other years they come from statistical yearbooks. The trade data are imperfect. Many state pairs are missing data because trade was not reported. Thus, while the maximum possible number of state-state-year observations on bilateral trade is $21 \times 20 \times 27 years = 11,340$, our final sample includes 8,960 pair-year observations if we include zero flows and 7,451 if we exclude zero flows.[8]

To measure bilateral travel time on the road we use geo-referenced maps of the Brazilian federal highway network from the Ministry of Transportation for the years 1960 through 2000. We keep only the segments in the maps that refer to paved federal highways. To recreate the maps for the years 1940 and 1950, we use information from historical sources to remove from the 1960 paved road network map those roads that did not exist in 1940 and 1950.[9] Once we have the road networks from 1940 to 2000, we run a fast marching algorithm to construct the traveling time between pairs of state centroids, making assumptions about the relative speed on a road compared to off the road. This follows the approach used in Allen and Arkolakis 2014.

2.2 Gravity in trade and migration

To estimate the migration and trade response to changes in road travel time, we use a standard gravity equation of the form:

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

[8]We differentiate between "true" zeros and missing data, as the difference is economically meaningful. We fully describe the data in Appendix C.1.

[9]See Appendix B.2 for details on sources.

where M_{odt} is either bilateral migration or trade, expressed in either logs or levels. γ_{ot} is an origin-year fixed effect that controls for any common shocks at the origin; γ_{dt} is a destination-year fixed effect that controls for any common shocks at the destination; γ_{od} is an origin-destination pair fixed effect controlling for any time-invariant heterogeneity such as distance, cultural proximity, and historical trade/migration networks. The parameter of interest is β, the elasticity of migration (or trade) to road travel time. Our hypothesis is that travel times increase migration and trade costs, thus reducing the movement of people and goods between pairs of locations. Therefore, we expect β to be negative and significant.

One concern with estimating Equation 1 directly is that road expansion between two pairs could be correlated with a shock to economic activity at the pair level. For example, if a state pair shares a common economic sector that experiences a negative shock, policymakers may counteract that shock by raising investment in infrastructure to promote regional growth and development. That same shock would discourage the movement of people and goods between locations sharing the sector. This would imply a positive association between road travel times and migration and trade that would lead us to understate the magnitude of the elasticities.

With this in mind, we discuss next our proposed estimation approach, which exploits the relocation of the federal capital city as a catalyst for the brand-new road network of radial highways connecting the new city in the interior of Brazil to the other parts of the country. We begin with a brief background on the process that led to the creation of the new city, Brasilia, and then describe the algorithms used to create the Brasilia-induced predicted road network, and, finally, calculate bilateral travel times on actual and predicted road maps.

2.3 Brasilia and the radial highways

Brasilia was conceived as a response to the long-standing issue of finding the ideal location for the country's capital city.[10] Brazil's first Constitution in 1891 situated a future capital city on a 60 x 90 kilometer piece of land, the *Quadrilatero Cruls*, located close to the border between the states of Goias and Minas Gerais. In 1922, the National Congress approved the creation of the new capital within this site but for more than two decades there

[10]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.).

was little movement toward building the city. In 1946, Eurico Dutra became president, and renewed debates over the site and construction of the new capital. Things began to move: in 1955, the recently created Commission for the New Federal Capital finalized the area in which Brasilia would be placed; once the new Brazilian president (Juscelino Kubitschek) was elected in 1956, the construction of Brasilia began immediately. After three years and ten months, Brasilia was officially inaugurated on April 21st, 1960.

The construction of Brasilia and work on the radial highway system occurred simultaneously. Before 1951, Brazil's few existing roads were limited to Brazil's northeast and southeast coastal areas. The 1934 and 1944 national transportation plans (*Planos Nacionais de Viação*, 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.[11] The final radial highway network connected Brasilia to the rest of the country. The roads run radially from Brasilia toward the country's extremes in eight directions: north, northeast, east, southeast, south, southwest, west, and northwest. Figure 1 shows the Brazilian highway network in 2000, separating radial highways from non-radial highways. In 1960, the total length of paved highways in Brazil was 10,890 km. Radial highways were approximately 20% (2,145 km) of the network. Over the next four decades the road network grew, increasing to 88,167 km in 2010. The radial highways also grew but their share of the total road network declined: at 8,352 km in 2010, they were just over 9% of the network. Appendix Figure 1 shows the spatial evolution of the paved road network from 1940 to 2010, marking radial highways in pink and non-radial highways in blue. Alongside the radial highways another important set of highways run north-south through the country. Along the eastern coast are the 4,800km BR-101, and the 4,486 km BR-116 (slightly inland from the coast). The Transbrasiliana highway, BR-153, runs north-south through the middle of the country. Appendix Figure 2 plots the total length of the highway for each year. Pink lines show the radial highways and the blue lines show the entire highway network including non-radial highways.

[11] Between 1951 and 1957, the Brasilia–Belo Horizonte line was laid down, connecting the soon-to-be new capital to the capital of Minas Gerais state. In the same period, parts of the Brasilia–Anapolis highway, a road that would link Brasilia to Sao Paulo, were laid. There were also plans to build the 2,276 km-long Belem–Brasilia, or *Transbrasiliana*, highway, as an overland route from the underpopulated northern states to the demographic and industrial centers of the country located in the south.

2.4 The Minimum Spanning Tree (MST) road network

We leverage the creation of Brasilia as a quasi-experiment that accelerated road connectivity growth for states closer to a network radiating out of the new capital *after* its actual inauguration in 1960. One approach would be to employ the actual radial highways to determine treatment intensity, that is, the reduction in road travel times induced by Brasilia. The main concern is that the exact location of the radial highways may have been chosen based on economic attributes. We propose instead to create a predicted road network centered around Brasilia on the assumption that the policymakers' goal was to choose the shortest path to connect the newly-built national capital to the state capitals in eight directions: the four cardinal directions and the four inter-cardinal directions. To that end, we proceed in four steps. First, we overlay the map of Brazil onto a circle centered at Brasilia's centroid. Second, we split the circle into eight pie slices of equal size so that the midpoint of each slice's arc is one cardinal or inter-cardinal direction (Appendix Figure 6 maps these slices). Third, within each slice we predict the shortest path to connect Brasilia, the starting point, to the state capitals contained in that slice. Finally, we form the predicted road network by joining the eight shortest paths that were calculated separately for each slice. We label the resulting network the Minimum Spanning Tree (MST).[12] Figure 1 shows the MST network, overlaid on the actual radial highway network.

2.5 Econometric analysis

For each state pair we create a time-invariant treatment intensity measure. We run the fast marching algorithm to calculate travel times on the MST network, capturing the exposure to the predicted highway, and then repeat the process to calculate travel times on an empty map, which creates a comparable no-roads travel time. We then calculate the log change in predicted travel time by subtracting the log MST travel time from the log empty travel time. The resulting variable, which we label the MST-induced reduction in travel time, approximates the predicted percent change in travel time spurred by the capital city relocation. We show in Appendix Table 1 that the predicted reduction in travel time is larger when paired states are farther apart; where either the origin or destination state is farther from Brasilia; and where the centroid center is farther from the state capital. These correlations are expected: mechanically, the travel route between two pairs is more likely to cross the radial highway network if the pairs are on opposite sides of the country

[12] Faber (2014) undertakes a similar approach in his study of Chinese highway construction.

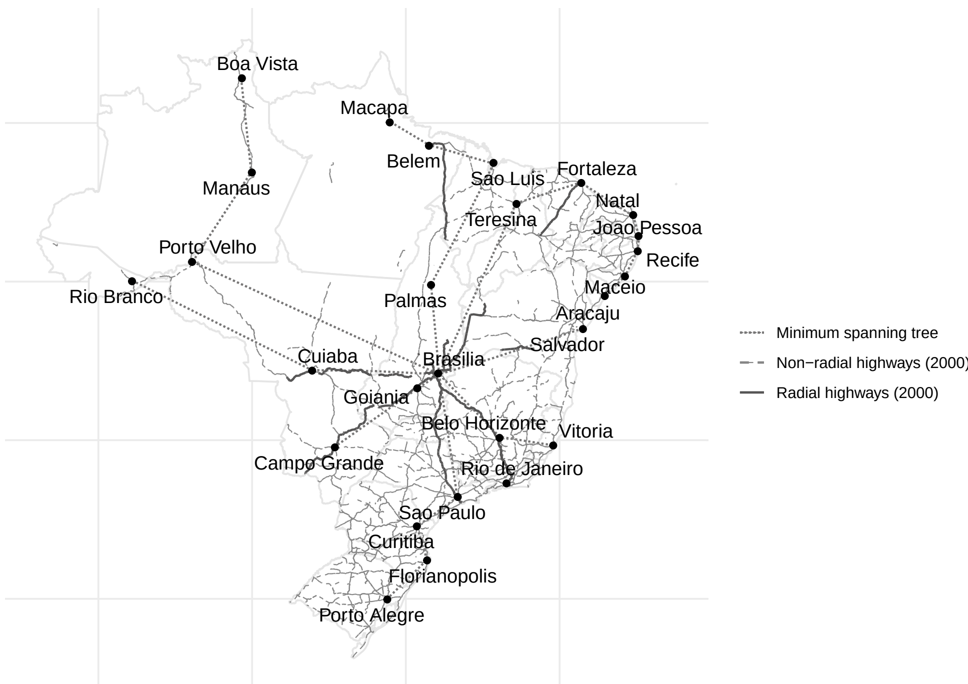


Figure 1: Map of straight-line instrument and radial highways

Notes: Figure is a map of the Brazilian road network indicating 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 MST road network. Consistent state boundaries appear in the background of the map. Source: authors' calculations based on maps obtained from the Brazilian Ministry of Transportation.

and pairs on the opposite side of the country are farther apart than neighboring pairs. Our specifications will always include a rich set of fixed effects—origin-year, destination-year, and pair levels—thus any time-invariant characteristics that are correlated with the exposure measure (such as distance between pairs or distance to Brasilia) will be absorbed by the fixed effects. Additionally, we confirm that there are no differential pre-trends for locations that get more exposure to the MST network, thus alleviating concerns that differential time trends explain our results.

We examine whether the main outcomes of interest—actual road travel times, bilateral migration and trade—responded to the MST-induced reduction in travel times, and, if so, whether the timing lines up with the introduction of Brasilia. To do this, we estimate the equation

$$\log \mathrm{y}_{odt} = \gamma_{ot} + \gamma_{dt} + \gamma_{od} + \alpha_t \text{MST-induced Reduction in TT}_{od} + v_{odt}, \tag{2}$$

where the dependent variable is either the travel time on the actual road network, migration, or trade between o and d in year t. MST-induced Reduction in TT_{od} is as described above; γ_{ot}, γ_{dt}, and γ_{od} are origin-year, destination-year, and origin-destination fixed effects; and v_{odt} is the idiosyncratic error term. We normalize bilateral trade flows by the total value of each year to avoid challenges posed by nominal values that change dramatically due to high inflation in Brazil between 1960 and 1990. We cluster standard errors at the pair level and weight the migration and trade regressions by the level of migration and (normalized) trade flows so that each individual migration move (or good transaction) has equal weight.

We consider the time-varying effect of the MST-induced reduction in travel time on the actual road travel time. The coefficient α_t is plotted in panel (a) of Figure 2.[13] We normalize the coefficient on year 1950 (the last pre-Brasilia observation) to equal zero and plot the coefficient relative to 1950 for the years between 1940 and 2000. The first reassuring result is that the coefficient on 1940 is not statistically different from zero, suggesting that there are no pre-trends in road construction. We then find a sudden decrease in travel time from 1960 on for those pairs predicted to face the largest reduction in travel time. The coefficient on 1960, of -0.16, is interpreted as a 0.1 unit change in MST-induced travel time (i.e., approximately a 10 percentage point difference in the predicted percent reduction in travel time) leads to a 1.6% reduction in actual travel time.[14] We reach similar conclusions when we analyze the migration and trade data shown in panels (b) and (c). To facilitate the visualization of trade flows results, which are annual, we average the data over five years, leaving us with 2,913 state pair-year observations. Both panels (b) and (c) show a consistent pattern. Before Brasilia (i.e., relative to 1950) there is no difference in bilateral migration induced by the MST network, but after Brasilia there is a significant spike; for trade flows we see a slight downward trend before Brasilia and a significant increase in the years after. The corresponding coefficients on 1960 are, respectively, 0.4 and 1.4 for migration and trade, implying that, relative to the year 1950, there was a 4% increase in

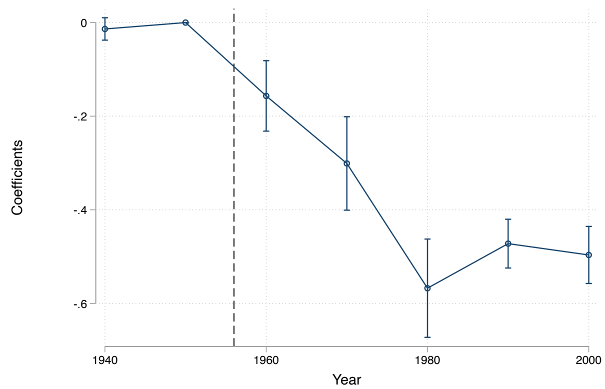
[13] Coefficient tables matching Figure 2 are reported in Appendix Table 2.

[14] Appendix Figure 3 shows the same pattern for meso-regions, a smaller geographic region than states.

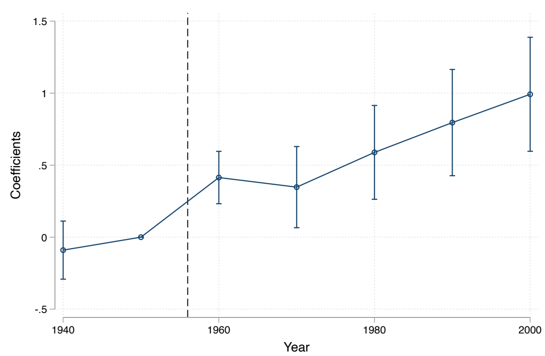
migration and a 14% increase in trade resulting from a 10 percentage point difference in MST-induced reduction in travel time.[15]

Figure 2: Time-varying effects of the predicted road network

(a) Travel time on roads



(b) Migration



(c) Trade

Notes: The graphs plot the estimated α_t coefficients from Equation 2 for (a) travel time on the actual road network, (b) migration, and (c) trade. The dashed line is the year 1956, which is when the radial highway system opened. The coefficient corresponding to the year 1950, which is the last year we have data available for the period before the radial highway system opened, is normalized to zero. For all other years, the coefficient measures the difference in the MST-induced changes in the outcome of interest between that year and the year 1950. Appendix Table 2 reports regression tables corresponding to the figures plotted here. To aid the visualization of trade patterns, we average the annual trade data over a five-year period. Standard errors are clustered at the pair level. Migration and trade regressions are weighted by the level of migration and (normalized) trade flows, respectively.

[15] A recent literature has studied the robustness of two-way fixed effect estimators. We show robustness by re-estimating the time-varying effects of the MST-induced reduction in travel times using the imputation estimator of Borusyak et al. 2021. The results are in Appendix Figure 4. The time patterns are very similar to our main results.

2.6 Gravity equation estimates

We now estimate Equation (1) using the actual travel times (OLS) and the travel times instrumented by the MST-induced reduction in travel times interacted with year dummies (omitting year 1950), and adding origin-year, destination-year, and pair fixed effects. The approach is one of "difference-in-differences IV," which we label IV for short. We cluster the standard errors at the unit of analysis, the pair level, to account for error serial correlation.[16] As earlier, we used normalized bilateral trade flows. (We present either the normalized trade or the logged normalized trade flows.) Finally, we weight the migration and trade regressions by the level of migration and (normalized) trade flows so that each individual migration move (or good transaction) has equal weight.

The results are presented in Table 1. Columns (1) and (2) show linear regression estimates where the dependent variable is bilateral migration in logs; columns (5) and (6) show the results of the same specification with log bilateral trade as the dependent variable. Starting with columns (2) and (4) in Table 1, the IV estimates imply an elasticity of migration to road travel time of 1.75 and an elasticity of trade to road travel time of 1.9. For both migration and trade we find that the magnitude of the IV elasticity exceeds the OLS elasticity (0.33 and 0.78, respectively). The greater magnitudes of the IV elasticities compared to the OLS ones suggest that road connectivity may have responded positively to negative economic shocks affecting pairs of locations, so that the OLS estimates likely understate the migration and trade responses to travel time reductions. The first stage regressions corresponding to the IV regressions are reported in Appendix Table 4.

We then estimate the elasticities when the dependent variables are in levels rather than logs. The log-transformation of the gravity equation in the linear IV estimates may raise concerns about the exclusion of zero flows. While zeros are not prevalent in the migration data (only one pair-year observation is excluded), they could pose a problem for estimating the trade elasticity to travel times since the trade data is much sparser. We estimate the model by Pseudo Poisson Maximum Likelihood. Columns (3) and (7) estimate the models directly for travel time; columns (4) and (8) use a control function approach where including the first-stage residuals in the estimating equation addresses the potential endogeneity of travel time.[17] The PPML control-function estimates imply

[16] The results are robust to clustering the standard errors by origin, destination, and year. However, because we have only 21 states, we do not cluster at the origin, destination, and year levels to avoid issues due to the small number of units. Appendix Table 3 shows three-way clustered standard errors for the migration regressions; we are unable to compute three-way clustered standard errors for the trade regressions due to sparse data.

[17] Atalay et al. 2019 use Monte Carlo simulations to demonstrate the good performance of the control

Table 1: Migration and Trade elasticities to travel time on roads

	Migration				Trade			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV	PPML	PPML	OLS	IV	PPML	PPML
Log Travel Time on Roads	-0.332	-1.753	-0.560	-2.123	-0.783	-1.890	-0.145	-1.963
	(0.116)***	(0.454)***	(0.124)***	(0.336)***	(0.155)***	(0.521)***	(0.204)	(0.437)***
N	2939	2939	2940	2940	7451	7451	8960	8960
F-stat	8.180	14.910			25.586	13.142		
First-stage F stat		22.086				11.217		
First-stage residuals control			no	yes			no	yes

Notes: An observation is a state-pair-year. **Log travel time on roads**: travel time between state o and state d centroids on the actual federal highway network in year t, calculated using a Fast-Marching Algorithm. Data available decennially from 1940-2000. Road data source: Brazilian Ministry of Transportation. Log travel time on roads is instrumented by the change in travel times on the predicted road network, relative to travel times on an empty map, interacted with year dummies (omitted year is 1950). The predicted road network is created by using a Minimum Spanning Tree (MST) algorithm aiming to connect Brasilia's centroid to those of all other existing state capitals within each one of the eight pie slices defined by the cardinal and intercardinal directions. **Migration**: Measure of migration is the stock of people born in state o living in destination state d in time t. Data is decennial covering 1940-2000. The maximum number of observations is 2940 (21 states*20 states*7 years). Data source: pre-1990: digitized from historical yearbooks. 1991-2000: microdata from Brazilian Population Census. **Trade**: Measure of trade value is the value of trade from state o to destination d in time t. Data is annual covering 1942-1949, 1959-1974, 1985 and 1998-1999. The maximum number of observations is 11340 (21 states*20 states*27 years) but some pairs are missing trade data because there was no trade between them or it was not reported. Data source: pre-1998: digitized from historical yearbooks. 1998-99: de Vasconcelos and de Oliveira (2006). All regressions include state origin-year, state destination-year, and state pair fixed effects. The dependent variables are in logs in columns (1), (2), (5), and (6), and in levels in columns (3), (4), (7), and (8). The Kleibergen-Paap F statistic (First-stage F stat) for weak identification is reported for IV estimates. Columns (3), (4), (7) and (8) present Poisson Pseudo Maximum Likelihood (PPML) estimates. Columns (4) and (8) add control for the first-stage regression residuals on the entire sample, including those with zero flows. Standard errors clustered at the pair level are reported in paretheses. OLS/IV regressions are weighted by the migration and normalized trade flows so that each individual migration move or good transaction has equal weight. Regressions are unweighted for the PPML regressions. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

a migration elasticity to travel time of -2.1, slightly larger than the IV estimate of -1.8, but not statistically so. The PPML control-function estimates imply a trade elasticity to travel time of -1.96, statistically indistinguishable from the linear IV estimate of -1.9. We conclude that our linear IV estimates do not appear biased due to the exclusion of zero flows.[18]

A final concern is that our results might be driven by increased economic activity instigated by the creation of a new city. Our regressions control for origin-year and destination-year fixed effects that should absorb any direct economic activity. We can approximate the thought experiment of excluding direct economic activity by considering whether the estimates are sensitive to dropping pairs that involve Goias, the state that houses Brasilia, where much of the new economic activity was concentrated. We show in Appendix Table 7 that the travel time elasticities are of the same magnitudes whether or not Goias is included in the sample, suggesting that the treatment effect is not induced by Brasilia alone.

With the estimated distance elasticities in hand, we now turn to understanding the economic interlinkages between locations through trade and migration.

3 Model of migration and trade

The gravity equation estimates presented in the previous section help us to understand if migration responds to changes in travel time. But to learn more about the main drivers of migration responses to improved road access, as well as to quantify and decompose the welfare gains from the Brasilia-induced road network into a migration and trade component, we need a quantitative spatial model with trade frictions, as in Eaton and Kortum 2002, and origin-destination migration costs, such as in Monte et al. 2018. In this section we present the model. We lay out the key structural parameters that govern the

function approach in a fixed effect Poisson model with endogenous regressor. Our robustness checks include higher-order polynomial functions of the first-stage residuals when running the control function approach in Appendix Table 5; the results are stable.

[18]The estimated trade elasticity of -1.9 is slightly larger than that presented in other papers in the literature, which tend to find an elasticity of trade flows to distance of about 1 (see, e.g., Silva and Tenreyro (2006)). We do not estimate exactly the same parameter, however. We estimate the elasticity to the change in travel time induced by the predicted expansion of the road network rather than the trade elasticity to distance. To check whether our estimated elasticity is reasonable, we run a number of robustness tests reported in Appendix Table 6. Column (1) shows that the elasticity of log trade flows to distance is -1.5, itself slightly higher than previous estimates. The elasticity of distance to roads is between -1.0 and -1.6, depending on whether or not pair fixed effects are controlled for. Altogether, these tests reassure us that our estimate is reasonable.

movement of people and goods, and the equilibrium conditions that determine prices and allocations across locations.

3.1 Consumer decision

Conditional on living in location d, individuals consume a CES aggregate of tradable good varieties, which costs P_{dt}, and housing, which costs r_{dt}. Their preferences for goods and housing accord with a Cobb-Douglass utility function. They spend a constant share, $1 - \alpha$, of their labor income, w_{dt}, on the consumption aggregate and a share, α, on housing. They also gain amenity of B_{dt} of living in location d. Given those assumptions, the indirect utility, V_{dt}, of living in location d is proportional (up to a multiplicative constant) to:

$$V_{dt} \propto B_{dt} P_{dt}^{-(1-\alpha)} r_{dt}^{-\alpha} w_{dt} \tag{3}$$

3.2 Migration decision

Given V_{dt}, an individual living in location o chooses which destination d to live in. The cost of migrating from o to d is given by $\kappa_{odt} \geq 1$. Individual i receives an iid shock for each destination, $\xi_{dt}(i)$, and chooses where to live to maximize their utility:

$$\max_{d} V_{dt} \kappa_{odt}^{-1} \xi_{dt}(i)$$

We assume that $\xi_{dt}(i)$ is Fréchet-distributed, with shape parameter ϵ and scale parameter B_{dt}. The parameter ϵ governs the extent of the heterogeneity in location preferences across individuals (the higher ϵ, the lower the heterogeneity) and the parameter B_{dt} captures the average desirability of a location which, in our model, is equivalent to an exogenous amenity value. Then, well-known results imply that the share of individuals from o who choose to migrate to d is given by:

$$\pi_{odt}^{\text{mig}} = \frac{V_{dt}^{\epsilon} \kappa_{odt}^{-\epsilon}}{\Phi_{ot}} \tag{4}$$

where $\Phi_{ot} = \sum_{d'} V_{d't}^{\epsilon} \kappa_{od't}^{-\epsilon}$.

These results show that the total number of people in destination d in year t is given by the sum of the migrant in-flows:

$$\begin{aligned} L_{dt} &= \sum_{o} M_{odt} \\ &= V_{dt}^{\epsilon} \sum_{o} \kappa_{odt}^{-\epsilon} \Phi_{ot}^{-1} L_{ot}. \end{aligned} \tag{5}$$

It follows from the utility maximization problem solved by each worker that the labor supply to each destination location increases with its amenity value and wages net of cost of living (V_{dt}), and decreases with the origin-specific migration costs to the destination (κ_{odt}) and with all origins' access to other high-value destinations (captured by Φ_{ot}).

3.3 Price determination

Each location o can use labor to produce a continuum of goods indexed by j. The productivity of location o in producing good j is given by $\zeta_{ot}(j)$, which we assume is Fréchet-distributed with shape parameter θ and scale parameter A_{ot}. The scale parameter A_{ot} approximately captures the average productivity of a region. The parameter θ measures a location's heterogeneity in producing different goods and, therefore, regulates the role of comparative advantage in shaping trade patterns.

The origin-cost of producing one unit of j is:

$$c_{ot}(j) = \frac{w_{ot}}{\zeta_{ot}(j)}$$

where w_{ot} is the wage. Because of trade costs, $\tau_{odt} \geq 1$ units of the good must be shipped from o so that one unit arrives at d. Assuming perfect competition in the goods market, the price at destination d of one unit of good j from o is $p_{odt}(j) = \tau_{odt} c_{ot}(j)$. Consumers in destination d purchase good j from the cheapest origin, that is, they solve

$$\min_{o} w_{ot} \tau_{odt} \zeta_{ot}(j)^{-1}.$$

The probability that d imports good j and, by the law of large numbers, all goods from o—is given by:

$$\pi_{odt}^{\text{trade}} = \frac{A_{ot} w_{ot}^{-\theta} \tau_{odt}^{-\theta}}{\Theta_{dt}}.$$

where $\Theta_{dt} = \sum_{o'} A_{o't} w_{o't}^{-\theta} \tau_{o'dt}^{-\theta}$. It follows that the flow of goods between origin o and destination d is:

$$X_{odt} = A_{ot} w_{ot}^{-\theta} \tau_{odt}^{-\theta} \Theta_{dt}^{-1} Y_{dt},$$

where Y_{dt} is total income at the destination. Trade between o and d will increase with average productivity at the origin (A_{ot}) and it will decrease with input costs (w_{ot}), trade costs (τ_{odt}), and the destination's ease of access to other low-cost and/or high-productivity sources (embedded in Θ_{dt}). Standard results also yield a price index for consumption goods in d that is proportional to the denominator of the trade share, up to a constant:[19]

$$P_{dt} \propto \Theta_{dt}^{-\frac{1}{\theta}},$$

so that price levels are lower in places better connected to other high-productivity/low-cost locations.

3.4 Wage determination

We make a balanced trade assumption that total expenditure is equivalent to total income. Therefore, the income in location o is the sum of what it earns from exporting goods to other locations:

$$\begin{aligned} Y_{ot} &= \sum_d X_{odt} \\ &= A_{ot} w_{ot}^{-\theta} \sum_d \tau_{odt}^{-\theta} \Theta_{dt}^{-1} Y_{dt}. \end{aligned}$$

Thus, the income in a location is higher if the location is well connected (lower τ_{odt}) to other destinations and faces less competition from other sources (lower Θ_{dt}).

We further assume that landlords spend their rental income in the location where their house is located and so, despite each worker only spending $(1-\alpha)$ of their income on traded goods, total expenditure on traded goods remains equivalent to the wage multiplied by the number of people, $w_{dt} L_{dt}$.[20] We assume that output is produced with labor alone, so an accounting identity implies that total output in location o, Y_{ot}, is equivalent

[19] The constant is $\bar{\Gamma} = \Gamma(\frac{\theta-(\sigma-1)}{\theta})^{\frac{-1}{\sigma-1}}$, which also accounts for the elasticity of substitutability of goods in the consumer's utility for produced goods.

[20] Workers spend $(1-\alpha)\,(w_{dt}L_{dt})$ on consumption goods and landlords earn $\alpha\,(w_{dt}L_{dt})$ in rental income, all of which they spend on consumption goods where their house is located. Therefore, $P_{dt}C_{dt} = (1-\alpha)\,(w_{dt}L_{dt}) + \alpha\,(w_{dt}L_{dt}) = w_{dt}L_{dt}$.

to the total wage bill in o. Putting both sides together implies:

$$w_{ot}L_{ot} = \sum_{d} \pi_{odt}^{\text{trade}} w_{dt}L_{dt}. \quad \forall o \in \{1, .., N\} \tag{6}$$

This system of N equations with N unknown wages allows us to solve the equilibrium level of wages indirectly.

3.5 Rent determination

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 (MC) of construction, which includes the interest rate, ι_t, construction costs, CC_t, and land costs, LC_{dt}. Equilibrium rent is the discounted value of house prices:

$$r_{dt} = \iota_t MC(CC_t, LC_{dt}) \tag{7}$$

The cost of land is a function of the demand for housing. The demand for housing is determined by the total expenditure on housing:

$$HD_{dt} = \frac{\alpha w_{dt} L_{dt}}{r_{dt}}$$

As a result, the equilibrium price of housing in location d is given by:

$$\begin{aligned} r_{dt} &= v_t HD_{dt}^{\epsilon_r} \\ HD_{dt} &= \frac{\alpha w_{dt} L_{dt}}{r_{dt}} \end{aligned}$$

where ν_t is a measure of construction costs at time t (inclusive of interest), and ϵ_r is the inverse housing supply elasticity. Substituting the two equations together yields:

$$r_{dt} = (v_t)^{\frac{1}{1+\epsilon_r}} (\alpha w_{dt} L_{dt})^{\frac{\epsilon_r}{1+\epsilon_r}} \tag{8}$$

3.6 Welfare effects of changing trade and migration costs

The measure of welfare is the sum of utility – the indirect utility and the idiosyncratic shock – for each individual in the economy. The Fréchet model has a well-known result that average utility is proportional to the denominator of the migration probability,[21]

$$E(V_d \kappa_{od}^{-1} \xi_{od}(i) | \text{choose } d) = V_d \kappa_{od}^{-1} \bar{\Gamma} \pi_{od}^{\text{mig}\,\frac{-1}{\epsilon}}$$
$$= \bar{\Gamma} \Phi_{ot}^{\frac{1}{\epsilon}}, \tag{9}$$

which implies that the average utility gain is the same for all individuals who start in origin o, regardless of where they end up migrating.

To measure the economy-wide welfare impact that results from arbitrarily large changes in migration and trade frictions, we apply the "exact hat" method (henceforth, a "hat" denotes the change, $\widehat{X_t} = \frac{X_t}{X_{t-1}}$.) It then follows that:

$$\widehat{\text{welfare}}_t = \sum_o \pi_{o,t-1}^{\text{mig}} \widehat{\Phi_{ot}}^{\frac{1}{\epsilon}}, \tag{10}$$

that is, the welfare change is given by a weighted average of the change in economic opportunities available in each origin.

The change in economic opportunities in location o, $\widehat{\Phi_{ot}}$, is:

$$\widehat{\Phi_{ot}} = \sum_d \pi_{odt-1}^{\text{mig}} \widehat{V_d}^{\epsilon} \widehat{\kappa_{od}}^{-\epsilon}, \tag{11}$$

which is a weighted average of the change in the value of the destinations they travel to and the change in migration costs to get to those destinations.

The change in the value of each destination (assuming that amenities are exogenous and do not change) is a function of wages, prices, and rents at the destination as follows:

$$\widehat{V_{dt}} = \widehat{P_{dt}}^{-(1-\alpha)} \widehat{r_{dt}}^{-\alpha} \widehat{w_{dt}} \tag{12}$$

and the change in the price index is given by:[22]

$$\widehat{P_{dt}} = \left(\sum_{o'} \pi_{o'dt-1}^{\text{trade}} \hat{A}_{o't} \hat{w}_{o't}^{-\theta} \hat{\tau}_{o'dt}^{-\theta} \right)^{-\frac{1}{\theta}}. \tag{13}$$

Welfare rises when migration costs (κ_{od}) fall or when there is an increase in destination

[21]This step requires the fact that the expected value of the unobserved shock for someone who migrates from o to d is given by $E(\xi(i)_{od} | \text{choose } d) = \bar{\Gamma} \pi_{odt}^{\text{mig}\,\frac{-1}{\epsilon}}$.

[22]Appendix D.1 presents the derivation of the "exact hat" expression for the change in price.

value (V_d). The value of the destination increases more when prices (P_{dt}) fall more, when rents do not increase as much, and when wages increase. Prices fall more when trade costs fall more, especially to locations which have low prices (either high productivity or low wages).

Overall, the key elasticities driving the welfare gains are the elasticity of migration and trade costs to distance and the elasticities of migration and trade. The elasticities of costs to distance control the magnitude of the size of the shock of building the road network. The trade and migration elasticities determine comparative advantage. A high trade elasticity will lead to large price reductions when roads are built, which will be a source of welfare gains. A high migration elasticity will lead to large migration responses when roads are built, which will be a source of welfare gains.

With this formulation of the changes in welfare due to road network expansion, we now turn to estimating the key elasticities needed to quantify the welfare gain.

4 Estimation

The goal of the quantitative exercise is to use estimates of the key elasticities in the model to make a quantitative analysis of the impact of Brazil's road network. This section discusses how we combine the earlier state-level estimates of the impact of roads on migration and trade with additional variation to estimate the migration elasticity to wages (ϵ) and the elasticity of rents to population (η). We set θ, the trade elasticity, exogenously at 4 following Simonovska and Waugh 2014.[23]

To estimate the migration and housing elasticities, we need data on wages and rents. Neither variable was collected in the historical data and so we use microdata from the Brazilian census between 1970 and 2010. The microdata enables analysis at the mesoregion, a finer geographical level than the state. Brazil has 137 mesoregions. Our sample of interest consists of males aged 20–65 who report non-zero earnings in their main occupations. We describe the census data and variable creation in more depth in Appendix Section C.2. We also present summary statistics in Appendix Table 18.

[23] We experimented with an estimation strategy on trade data parallel to the one we use on the migration data. However, with trade data only from the state level compared with meso-level migration data, our estimation is under-powered. Our point estimate is 0.014 (s.e. 0.1). See Appendix Table 8.

4.1 Scaling state-level migration costs to match meso-level migration

Section 2 uses state-level data to estimate the change in migration and trade costs in response to changes in travel times. We start by adjusting these cleanly-identified measure of the migration cost (computed on the bilateral state flows) to match the meso-level migration data. The microdata in the census asks people where they were living five years ago. We use that data to compute flows of migrants between mesoregions.[24]

In logs, the gravity equation (Equation 4) is:

$$\log \pi_{odt}^{\text{mig}} = \epsilon \log V_{dt} - \epsilon \log \kappa_{odt} - \log \Phi_{ot},$$

We parameterize the migration costs, κ_{odt}, as the following:

$$\log \kappa_{odt} = \beta_1 \log \text{distance}_{od} + \beta_2 \log \text{reduction in travel time with Brasilia.} \tag{14}$$

The equation above is closely related to Equation 1. Substitution yields that the parameter β estimated in Equation 1 corresponds to the migration elasticity multiplied by the elasticity of migration costs to travel time (i.e., $\beta = \epsilon\beta_2$). We did not estimate a distance coefficient in Equation 1 because distance is collinear with pair fixed effects. To recover $\epsilon\beta_1$, we therefore run the gravity equation on the meso-meso migration data without the pair fixed effects. Column (2) of Appendix Table 10 shows that we estimate a distance coefficient of -1.698.[25] We do a similar exercise to estimate the implied distance elasticity for trade. Without trade flows at the meso level, we run the trade regression on the state-state data after 1980, omitting the pair-level fixed effect. Appendix Table 11 shows that we estimate a trade distance coefficient of -1.187. As part of this estimation, we extract the destination-year fixed effect, $\widehat{\delta}_{dt} = \epsilon \log V_{dt}$, to decompose in the next step.

[24] The meso-meso migration measure differs from the state-level analysis on two dimensions. First, it is a flow (location compared to five years earlier) instead of a stock (living outside state of birth). Second, meso is a finer geographical level than state. Appendix Table 9 compares the migration response to various measures, such as distance, that we can compute in both datasets. We find remarkably similar elasticities to distance across geographical units (e.g., an elasticity of -1.35 in column (1) at the meso level and an elasticity of -1.30 in column (2) at the state level). We also find very similar elasticities across the measure of stock and flow migration: as might be expected, the long-run stock measure is slightly larger than the short-run flow measure, but not statistically so. Elasticities in the historical sample (1940-1980) are similar, although a little larger, than those of the more recent sample; e.g., an elasticity to distance of 1.95 (column 7) compared with 1.5 (column 3), which might be expected with improvements in transportation technology over the 70-year period. Overall, we find a surprisingly robust pattern of elasticities of migration to distance across definitions of migration, geographical unit, period of study, and sample.

[25] Column (2) of Appendix Table 11 shows the same exercise estimating the distance coefficient on state-level data for just the post-Brasilia period. The elasticity coefficients on distance are similar whether we consider state-level migration or meso-level migration.

4.2 Estimation of migration and housing elasticities

The destination-year fixed effects in the migration gravity model are structurally interpreted as a bundle of location-specific wages, amenities, prices, and rents. In particular, as Equation 3 shows, the destination fixed effect, $\hat{\delta}_{dt}$, is equivalent to:

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

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

$$\log B_{dt} = b_d + b_t + \xi_{dt},$$

where b_d is the time-invariant component for location d, b_t is the time effect, and ξ_{dt} is an error term. These assumptions yield the following estimating equation in differences:

$$\Delta\hat{\delta}_{dt} = -\epsilon \left(\alpha \Delta \log r_{dt} + (1-\alpha) \Delta \log P_{dt} \right) + \epsilon \Delta \log w_{dt} + \Delta \xi_{dt}. \tag{15}$$

From the housing supply equation (Equation 8), we have, also in differences:

$$\Delta \log r_{dt} = \eta (\Delta \log w_{dt} + \Delta \log L_{dt}) + \Delta \psi_{dt}, \tag{16}$$

where ψ_{dt} is an error term. The two coefficients of interest are the elasticity of migration to real wages, ϵ, and the elasticity of housing to income, η.

There are two challenges when estimating Equations 15 and 16. First, wages, rents, and prices are endogenous. We construct instruments, discussed below, to address these endogeneity concerns. Second, we do not observe the price index and so cannot directly estimate Equation 15. We leverage the model structure, discussed below, to simulate price data in place of observed data.

4.2.1 Instrument for wages

We devise an instrument for the change in wages by constructing a Bartik shock for wage growth (Bartik 1991).[26] The Bartik shock takes a weighted average of the national-level

[26] Adao et al. (2020); Goldsmith-Pinkham et al. (2020) and Borusyak et al. (2020) discuss the identification assumptions that are sufficient for shift-share instruments such as the Bartik shock to identify the parameter of interest. In our setting, we assume that national-level industry growth shocks are plausibly exogenous across industries and that unobserved heterogeneity in regional labor supply shocks is asymptotically ignorable.

growth rate in wages for each industry. The Bartik shock for location d in year t is:

$$\widehat{Z}_{wdt} = \sum_{\text{ind}} \frac{L_{ind,d,0}}{L_{d,0}} \widehat{w}_{ind,-d,t},$$

where $L_{\text{ind},d,0}/L_{d,0}$ is the baseline industry composition in location d and $\widehat{w}_{ind,-d,t}$ is the average wage change in industry ind in year t, excluding workers in location d. The Bartik shock utilizes variation across space in the location of industry. We compute the Bartik shock using data at least 200km from the mesoregion of interest to avoid spatial correlation in labor demand shocks.

4.2.2 Instrument for labor

From Equation 5, labor is given by $L_{dt} = \sum_o V_{dt}^{\epsilon} \tau_{odt}^{-\epsilon} \Phi_{ot}^{-1} L_{ot}$. Using the exact hat method, the change in labor in destination d can be expressed as a direct term, measuring the change in the attractiveness of the location, and a labor market access term, measuring the change in opportunities for people in the origin locations who tend to migrate to destination d:[27]

$$\widehat{L}_{dt} = \underbrace{\widehat{V}_{dt}^{\epsilon}}_{\text{Direct effect}} \sum_o \underbrace{\pi_{odt-1}^{\text{mig}} \widehat{\kappa}_{odt}^{-\epsilon} \left(\sum_{d'} \pi_{od't-1}^{\text{mig}} \widehat{V}_{d't}^{\epsilon} \widehat{\tau}_{od't}^{-\epsilon} \right)^{-1} \widehat{L}_{ot}}_{\text{Origin labor market access effect}}.$$

Since we focus on the post-Brasilia period, we assume that there are no changes in migration costs, i.e., $\widehat{\kappa}_{odt} = 1$. We also abstract from exogenous population growth at the origin $\widehat{L}_{ot} = 1$. With these assumptions, labor change in destination d is a composite of the direct increase in the benefits of being in location d, given by $\widehat{V}_{dt}^{\epsilon}$, which we approximate with the change in wages, $\widehat{w}_{dt}^{\epsilon}$, and a market access term that shows that when migrants have easy access to other destinations with a large direct gain, they will be less likely to migrate to d. These simplifying assumptions result in the following wage equation:

$$\widehat{L}_{dt} = \widehat{w}_{dt}^{\epsilon} \sum_o \pi_{odt-1}^{\text{mig}} \left(\sum_{d'} \pi_{od't-1}^{\text{mig}} \widehat{w}_{d't}^{\epsilon} \right)^{-1}$$

We create an instrument, $\Delta Z_{L_{dt}}$, for this expression by a substitution of the Bartik

[27]See Appendix D.1 for details.

instrument in place of wages and setting $\epsilon = 1$:

$$\widehat{Z}_{L_{dt}} = \widehat{Z}_{wdt} \sum_{o} \pi^{\text{mig}}_{odt-1} \left(\sum_{d'} \pi^{\text{mig}}_{od't-1} \widehat{Z}_{wd't} \right)^{-1}$$

4.2.3 Price index

We do not observe the price level for each meso-region. We, therefore, use the model structure, embodied in Equation 13, to construct an imputed price change for each location d. To that end, we assume that in the post-Brasilia period there are no changes in trade costs i.e., $\widehat{\tau}_{odt} = 1$. We also assume the underlying fundamentals are the same i.e., $\hat{A}_{o't} = 1$, which yields an expression for the change in price as:

$$\widehat{P_{dt}} = \left(\sum_{o} \pi^{\text{trade}}_{odt-1} \hat{w}^{-\theta}_{ot} \right)^{\frac{-1}{\theta}},$$

where $\pi^{\text{trade}}_{odt-1}$ measures the share of imports in location d from location o. Thus, the price index in location d will decrease when wages in location o decrease. This effect is greater if location d tends to import a larger share of its goods from location o.

To operationalize this measure we need to observe the trade shares, $\pi^{\text{trade}}_{odt-1}$, at the meso-meso level. However, our data are at the state-state level. We therefore simulate the first part of the structural model to yield implied trade costs that match the observed wages. We then generate computed trade flows. Simulating trade flows requires only the value of the observed wage data and the trade elasticity; it does not depend on any other elasticity parameters that have yet to be estimated. As above, we generate an instrument for the imputed price index as above:

$$\widehat{Z}_{P_{dt}} = \left(\sum_{o} \pi^{\text{trade}}_{odt-1} \widehat{Z}^{-\theta}_{wot} \right)^{\frac{-1}{\theta}}.$$

4.2.4 Results

We set up the following system of estimating equations in first differences, calibrating the share of expenditure on housing, $(1 - \alpha)$, to 0.2 using the share of rents on total expenditure drawn from the 2008–2009 Survey of Family Budget.

$$\Delta\hat{\delta}^{\kappa}_{dt} = \epsilon(\Delta \log w_{dt} - 0.8\Delta \log P_{dt} - 0.2\Delta \log r_{dt}) + \varepsilon^{u}_{dt}, \tag{15'}$$

$$\Delta \log r_{dt} = \eta(\Delta \log w_{dt} + \Delta \log L_{dt}) + \varepsilon^{r}_{dt}, \tag{16'}$$

Our identifying restrictions are:

$$E(\Delta Z_{dt}\varepsilon^{u}_{dt}) = 0,$$
$$E(\Delta Z_{dt}\varepsilon^{r}_{dt}) = 0,$$

where:

$$\Delta Z_{dt} \in \{\widehat{Z}_{wdt}, \widehat{Z}_{P_{dt}}, \widehat{Z}_{L_{dt}}\}.$$

We estimate equations 15′ and 16′ using 2SLS, with results shown in Table 2. We cluster standard errors by the state. The regression is unweighted. Starting with columns (1) and (2), we present the estimate of the migration elasticity. We purge observed rents from observable quality differences by running the hedonic regressions (detailed in Appendix C.2.3). Our IV estimate in Column (2) yields a migration elasticity of 4.5. The first stage for the IV regression appears in Appendix Table 12. Appendix Table 13 runs a robustness test where we include the imputed price index. The estimated elasticity is higher, at 6.0. The parameter of migration elasticity to wages is a parameter that has not been extensively estimated but we can compare our result to those reported in the (predominantly US) literature. Monte et al. 2018 estimate an elasticity of 3.30 using commuters. Caliendo et al. 2019 estimate an elasticity of 0.2 for a 5-month frequency. Tombe and Zhu 2019 estimate an elasticity of 2.5 from Chinese data. Diamond 2016 estimates an elasticity of between 2 and 4, but she does not incorporate origin-destination costs of migration.

Next, we contrast the estimated elasticities we obtain to the elasticities we 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 base estimates on population allocations rather than population flows; 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 from responding to wage shocks, the exclusive use of population data may lead to an understated estimate of the migration elasticity. We show that this is indeed the case for our data by re-estimating the elasticities without bilateral migration costs[28] in column (2) of Table 2. Using the same data and the same estimation approach, we estimate an elasticity of migration of -0.08, not statistically different from zero, and smaller than our estimate of 4.8. This finding suggests that netting out migration costs is relevant to quantifying how migration flows are responsive to changes in the economic returns of migration.

In columns (5) and (6) we present estimates of the inverse housing elasticity. Our IV estimate in column (6) is an elasticity of rents to income of 0.4, which implies a housing supply elasticity of 1.7.[29] This number is very close to the population-weighted elasticity of 1.75 found for US metropolitan areas in Saiz 2010, despite the Brazilian context not being directly comparable to the US because of high levels of informality in the Brazilian housing market.[30]

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 the state-level estimates with pair fixed effects. We follow the procedure outlined in Section 4 to rescale these costs to the meso-meso level.

The first counterfactual is the thought experiment of deleting all roads in Brazil. To

[28]Without bilateral migration costs, but still assuming Fréchet preference shocks, the probability that an individual will choose to live in n at time t is given by the following expression, which no longer varies by origin o:

$$\pi_{odt} = \frac{V_{dt}^{\epsilon}}{\sum_{d'} V_{d't}^{\epsilon}} = \pi_{dt} \tag{17}$$

[29]From equation 8, the elasticity of rents to income identifies an elasticity of rents to income, η, that is related to the inverse housing supply elasticity, ϵ_r, by the equation $\eta = \frac{\epsilon_r}{1+\epsilon_r}$.

[30]Guedes et al. 2022 explore the heterogeneity in supply elasticities accounting for geographic constraints as in Saiz 2010 and also informal housing. They find an unweighted average housing supply elasticity of 1.09 for Brazilian metropolitan areas. Our estimates are larger, possibly because they are estimated for mesoregions, geographical units less land-constrained than their sample of metropolitan regions.

Table 2: Migration and housing elasticity

	Mig. elasticity		Mig elasticity - no cost		Housing elasticity	
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
Change in wages, adjusted for residualized rents	1.239	4.515	0.029	-0.076		
	(0.500)**	(0.763)***	(0.295)	(0.465)		
Change log imputed prices	0.197					
	(1.496)					
Change in log housing expenditure					-0.025	0.372
					(0.086)	(0.153)**
N	137	137	137	137	137	137
F stat	3.286	35.000	0.009	0.027	0.082	5.902
First-stage F stat		16.103		16.103		6.395

Notes: Unit of analysis is a meso-region and each observation is the 2010-1980 difference. The dependent variable in columns (1) and (2) is the 2010-1980 change in (log) indirect utility of the meso-region. The indirect utilities for years 2010 and 1980 are obtained as the set of (meso) destination-year fixed effects after estimating a gravity equation from meso-to-meso bilateral migration flows, accounting for bilateral migration costs (as a linear function of distance and travel times in logs) and (meso) origin-year fixed effects. The dependent variable in columns (3) and (4) are also the 2010-1980 change in indirect utilities from meso-to-meso flows, but assuming zero bilateral migration costs. The dependent variable in columns (5) and (6) is the change in (log) rental prices. Rents are calculated from census micro-data in each year as the meso-specific average after netting out housing characteristics such as number of rooms and bedrooms, electricity access, walls, roof, and floor quality, among others.Instruments for changes in wages and housing expenditure are meso-level Bartik shocks and a model-based measure of changes in labor as a function of Bartik shocks of all meso-regions weighted by migration costs. The Kleibergen-Paap F statistic (First-stage F stat) for weak identification is reported for IV estimates. Standard errors clustered at the state level. Regressions unweighted. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

construct the travel times between all mesoregions for the no-roads scenario, we use a constant travel speed across all pixels. We then use our earlier gravity estimates to convert this travel time variable into migration and trade cost units. Table 3 shows the resulting counterfactual relative migration and trade costs. Compared to the scenario of no roads, the highway network reduced trade costs by 9% (the 10th percentile is a 1% reduction; the 90th percentile a 16% reduction) and migration costs by 8% (0% and 13% for the 10th and 90th percentiles, respectively).

The second counterfactual considers the road network as if, instead of having a national highway system centered on Brasilia, that 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 panel b of Appendix Figure 6. We rerun the fast marching algorithm on this network to generate the travel times under the Rio counterfactual. Table 3 shows that the Brasilia-centered highway reduces average trade costs by 2% and average migration costs by 2%. However, the reductions are heterogeneous: 10% of pairs experienced a trade cost increase of at least 3% and 10% of pairs experienced a migration cost increase of at least 3%.

Table 3: Counterfactual migration and trade costs

	(1) Trade costs	(2) Migration costs
Relative cost: Brasilia road network vs. no-roads	0.91	0.92
	{0.84}	{0.87}
	{0.99}	{1.00}
Relative cost: Brasilia road networ vs Rio road network	0.98	0.98
	{0.92}	{0.94}
	{1.03}	{1.03}

Notes: Table reports the relative trade cost (column 1) and migration cost (column 2), relative to a baseline of 1. The first number is the mean value; the second and third numbers are the 10th and 90th percentiles. Estimates are computed by using the state-level gravity equation and dividing by the trade elasticity migration (4) or migration elasticity (4.515) respectively.

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 is interpreted as the bundle of rents, wages, amenities, and prices at a destination. We have data on rents and wages and, as discussed above, we construct simulated price data. The unexplained portion of indirect utility is therefore the amenity level of a location. We thus 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, as per Equation 6. We then have measures of all the variables determining trade flows and can solve the trade side of the model.

Next, 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 first take the initial value of rents, wages, and prices. Then, given the exogenous productivities and exogenous amenities, we use the migration costs to construct updated migration probabilities, which 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 updates all the endogenous values of the system (rents, prices, labor). The exogenous quantities (amenities, productivities) remain unchanged. These updated values become the new starting values, so we compute the migration probabilities that would be consistent with these prices. We continue this way until the migration probability is stable. This gives the economy's new equilibrium prices and quantities.

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 their 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 D.2. While our estimation strategy is robust for the existence of such dynamic considerations, since 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.[31]

5.3 76% 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? Column (1) of Table 4 shows the equilibrium changes in trade, migration, and utility. The first panel considers the effects of Brasilia compared with no roads. Overall, welfare is 2.8% higher. The road network caused an increase in trade of 20.3% (measured as the average share of expenditure on non-self-produced goods) and an increase in migration of 20.6%.

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 trade costs alone fell and migration costs stayed at the baseline. At that migration cost level, a reduction of trade costs causes a small increase in migration, 1%. Overall, the welfare gain is 2.1% or 76% 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 19.1% and trade flows are essentially unchanged. Changing migration costs alone results in a 0.7% increase in welfare.

The second panel of the table shows the effect of the Brasilia network compared with that of a hypothetical road network had the capital city had remained in Rio de Janeiro. This counterfactual varies both the location and the density of the roads. We find that

[31] Caliendo et al. (2019) consider this issue explicitly in their study of the dynamic effects for the US of increased Chinese import competition.

the road network associated with moving the capital to Brasilia led to a slightly larger – 0.2% – welfare effect relative to a counterfactual network that did not explicitly target the capital cities to be connected to Brasilia. The counterfactual road network in Appendix Figure 6 suggests that this is probably because the connections along the very densely populated coast are similar under both networks.

What drives the large role of trade costs in explaining the overall welfare gain? As discussed in Equation 11, both the magnitude of the shock as well as the trade elasticity and migration elasticity figure in determining the magnitude of welfare gains. In the gravity equation, we estimate a larger elasticity of trade to roads, 1.9, than migration to roads, 1.75. As noted in Section 4.1, the gravity elasticity is a combination of the migration (trade) elasticity and the elasticity of migration (trade) costs to roads. The implied elasticity of trade costs to roads is therefore $1.9/4 = 0.5$ (using a trade elasticity of 4), and the implied elasticity of migration costs to roads is $1.75/4.52 = 0.4$ (using the estimated migration elasticity). The difference in these numbers is why the the road network led to a 9% reduction in trade costs compared to an 8% reduction in migration costs.

To isolate whether it is the larger shock to trade costs than migration costs that is driving the larger share of welfare gains accruing to trade, we run a robustness test in Appendix Table 14 where we equalize the change in the migration and trade costs by equalizing the relevant distance elasticities and overall elasticities. We find that if we equalize the reduction in trade and migration costs, the share of welfare gains accruing to trade is smaller – either 71% or 72% – compared to the baseline value of 76%. However, even with equal-sized shocks, trade still explains the majority of the welfare gains from improving roads. This suggests that part of the reason trade explains a larger share of the welfare gains is because trade makes up a larger share of the economy. In other words, trade explains a larger share of the welfare gains partly because we estimate that trade costs fell relatively more than migration costs after the construction of Brazil's road network, and partly because trade is a relatively larger share of the economy than migration at baseline.

We explore robustness over our estimated elasticities to evaluate our finding that trade is the dominant source of gains after the expansion of the road network. Appendix Table 15 shows robustness over the elasticity of trade (migration) costs to roads and the trade (migration) elasticity. The table provides two take-aways. First, even if we reduce the elasticity of trade costs to roads by 50%, trade still accounts for 57-86% of the gains from the road construction, depending on the value of the trade elasticity. If we double the elasticity of trade costs to roads, then trade accounts for 61-91% of the welfare gains from the road network. Second, a higher trade elasticity leads to both a larger increase in welfare and to trade explaining a larger share of welfare gains. Halving the trade elasticity to 2 would imply that trade only accounts for 74% of the gains instead of 76%; doubling the trade elasticity would lead trade to account for 79% of the gains. Increasing the trade elasticity also boosts the overall welfare gains of trade because a larger trade elasticity reflects larger productivity gains from increased comparative advantage facilitated by trade cost reductions. The second half of the table shows that a similar intuition holds for the migration cost elasticity and the migration elasticity: increasing the responsiveness of migration costs to distance expands the welfare gain from decreasing travel time and decreases the share of the welfare gains accruing to trade; increasing the migration elasticity increases the welfare gains of the network and decreases the share due to trade.

Table 4: GE counterfactuals

	Overall	Decomposition		Free mobility	Labor immobile
	(1) Both	(2) Trade cost only	(3) Mig cost only	(4) Trade cost only	(5) Trade cost only
Experiment 1: Brasilia road network (vs. no roads)					
Change in trade	1.203	1.196	1.005	1.184	1.195
Change in migration	1.206	1.010	1.191	1.000	.
Change in welfare	1.028	1.021	1.007	1.024	1.021
Experiment 2: Brasilia road network (vs. Rio road network)					
Change in trade	1.004	1.004	0.998	1.021	1.007
Change in migration	1.012	0.997	1.012	1.000	.
Change in welfare	1.002	1.002	1.000	1.004	1.002

Notes: Table shows the relative change. All values are relative to a baseline value of 1. The numbers are computed by simulating the structural model under different scenarios. Column (1) shows the effect of changing both migration and trade costs. Column (2) holds migration costs at the baseline level and reduces trade costs only. Column (3) holds trade costs at the baseline level and reduces migration costs only. Column (4) simulates a model where there are no origin-destination costs of migration. Column (5) simulates a model where labor is immobile.

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

Although reduced trade costs explain 76% of the welfare gains we have reported, this does not mean that costly migration is insignificant. To demonstrate this point, Column (4) repeats the counterfactual exercise, assuming that migration costs in the model depend only upon preferences for location, without an origin-destination component. In that case, the simulated gain from the Brasilia-induced highway system is a location-utility gain of 2.4%, 14% lower than the location-utility gain of 2.8% when both costly migration and costly trade have a role. 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 a model where labor is immobile. We estimate a welfare gain of 2.1%, 25% lower than the full effect. Labor mobility is an additional channel through which roads affect market integration. Not accounting for this leads to underestimating the benefit of improving connectivity between locations.

5.5 Bilateral migration costs induce heterogeneity in the benefits of roads

Finally, we show that a model with bilateral migration costs, unlike a model with free labor mobility, does not equalize utility across space. The equalization of utility across space is a key equilibrium condition in standard economic geography models where labor can freely move and only location preferences produce friction in mobility. 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.

Figure 3 shows the spatial distribution of welfare change. Panel (a) shows the baseline model where both migration and trade are costly. The change in welfare ranges from 1% to 15%, with the mean mesoregion experiencing a 4% increase in welfare. Panel (b) considers the case where labor can move freely. In this case, welfare is equalized across space with all mesoregions receiving a 2% increase in welfare. Panel (c) considers the case where labor is immobile. In this case, labor cannot arbitrage at all and so the range of outcomes is larger, from 0% to 16%.

The finding that welfare is not equalized across space under costly migration has direct implications for policy. Migration costs index the extent to which a location is "sticky": people are differentially affected by any spatial investment depending on where they live. In many countries, governments invest resources to develop specific infrastructures such as roads but also make broader investments to encourage job creation or economic growth. When migration is costly, there will be heterogeneity in the response to policy for both directly and indirectly affected regions.

6 Conclusion

Studies of the effects of roads on facilitating trade has generated a large body of literature. In this paper, we focus equally on the effects of roads on 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, where relocating the capital city to the interior of the country in 1960 spurred the building of a large highway network connecting the new capital to the state capitals. We generate an instrument for road location based on straight-line connections between the new capital, Brasilia, and state capitals. We first document that states better connected by roads experienced increased movement of goods and people compared to states not as well connected. 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).

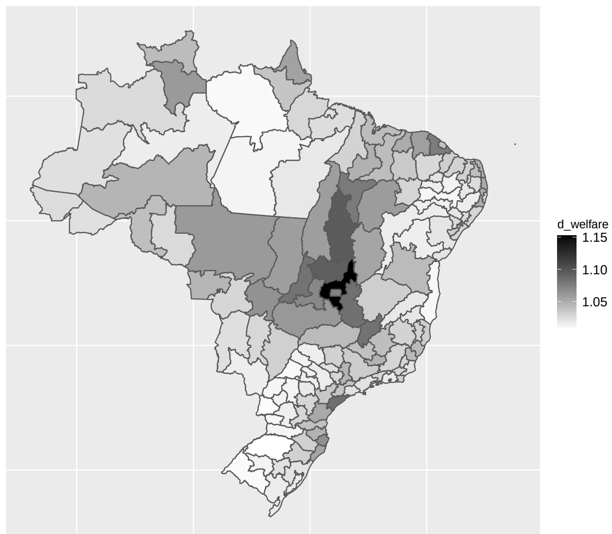
We find that the road networks connecting Brasilia to the state capitals decreased migration costs by 8% and trade costs by 9%. Overall, these decreased costs increased welfare by 2.8%, of which 76% was the result of reduction in trade costs and 24% was due to reduction in migration costs.

Although we find that trade cost reduction is the dominant source of welfare gains, we also show three reasons why it is important to account for costly migration. First, without separating out migration costs from migration returns, researchers will underestimate the migration elasticity, a key input for many spatial equilibrium models. We estimate an elasticity of migration to wages of 4.5; with the same data but without accounting for migration costs, we estimate an economically impossible negative (and statistically insignificant) migration elasticity. Second, the overall estimate of the welfare benefits of improving roads depends on assumptions about how easily people can migrate. A model that ties migration costs only to preference shocks does not account for the reduction in migration costs induced by the road network and will understate the welfare gain of building roads. Using the same data, we estimate a welfare gain of 2.4%, 14% lower, assuming that the only friction to labor migration is due to heterogeneous preferences; if labor is immobile, we estimate a 25% lower gain. Third, the spatial equilibrium arbitrage condition, that expected utility is equalized across space, does not hold when migration is costly. Instead, what does hold is an amended arbitrage equation that expected utility is equalized within origin. 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 range of regional gains from the improved infrastructure is 1%–15%, compared to uniform gains in the absence of origin-destination costs of migrating.

Our paper demonstrates a meaningful but understudied relationship: infrastructure plays an important role in facilitating the movement of labor to where its return is highest. If roads facilitate the allocation of labor, this is an important benefit to quantify. Likewise, costs of adjusting of other mobile factors of production, such as capital, may also hinder allocating resources to where they would be most productive. The aggregate effects of how to best allocate factors of production, particularly for developing countries where infrastructure is poor, is a potentially important mechanism to further explore.

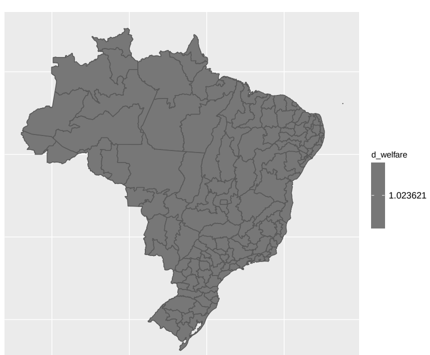
Figure 3: Spatial heterogeneity in welfare impacts of roads

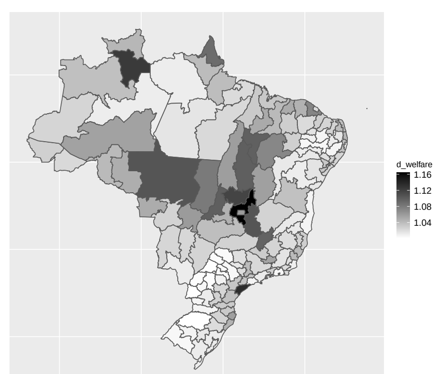
(a) Baseline: costly trade and migration



(b) Costly trade; free migration

(c) Costly trade; no migration





Notes: Figure shows the spatial distribution of the change in welfare for initial inhabitants of each region. Map (a) shows the welfare gain for the model with costly trade and costly migration. Map (b) shows the welfare gain for the model with no origin-destination costs for migration or trade. In this case, welfare is equalized across space, so all regions have the same welfare gain. Map (c) shows the welfare gain with origin-destination trade costs and labor immobility. Source: authors' calculations from simulating the model under different scenarios.

References

Adao, Rodrigo, Michal Kolesar, and Eduardo Morales, "Shift-Share Designs: Theory and Inference," *Quarterly Journal of Economics*, 2020, pp. 1949–2010.

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

Artuç, E, S Chaudhuri, and J McLaren, "Trade Shocks and Labor Adjustment: A Structural Empirical Approach," *The American Economic Review*, 2010.

Atalay, Enghin, Ali Hortaçsu, Mary Jialin Li, and Chad Syverson, "How Wide is the Firm Border?," *Quarterly Journal of Economics*, 2019, *134* (4), 1845–1882.

Atkin, David, "The Caloric Costs of Culture: Evidence from Indian Migrants," *American Economic Review*, 2016, *106* (4), 1144–81.

Banerjee, Abhijit and A Newman, "Information, the Dual Economy, and Development," *Review of Economic Studies*, 1998, *65* (4), 631–653.

_ , Esther Duflo, and Nancy Qian, "On the Road: Access to Transportation Infrastructure and Economic Growth in China," *Journal of Development Economics*, 2020, *145* (102442), 1–36.

Bartik, Timothy J, "Who Benefits from State and Local Economic Development Policies?," *Books from Upjohn Press*, 1991.

Bird, Julia and Stephane Straub, "Road Access and the Spatial Pattern of Long-term Local Development in Brazil," *World Development*, 2020, *127.*

Borusyak, Kirill, Peter Hull, and Xavier Jaravel, "Quasi-Experimental Shift-Share Research Designs," *Review of Economic Studies*, 2020, *89* (1), 181–213.

_ , Xavier Jaravel, and Jann Spiess, "Revisiting Event Study Designs: Robust and Efficienct Estimation," 2021.

Bryan, Gharad and Melanie Morten, "The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia," *Journal of Political Economy*, 2019, *127* (5), 2229–2268.

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

Caliendo, Lorenzo and Fernando Parro, "Estimates of the Trade and Welfare Effects of NAFTA," *The Review of Economic Studies*, 2015, *82* (1), 1–44.

_ , Maximiliano Dvorkin, and Fernando Parro, "Trade and Labor Market Dynamics: General Equilibrium Analyis of the China Trade Shock," *Econometrica*, 2019, *87* (3), 741–835.

Chandra, Amitabh and Eric Thompson, "Does public infrastructure affect economic activity? Evidence from the rural interstate highway system," *Regional Science and Urban Economics*, 2000, *30*, 457–490.

Chein, Flavia and Juliano Assunção, "How does Emigration affect Labor Markets? Evidence from Road Construction in Brazil," *Brazilian Review of Econometrics*, 2016, *36* (2), 157–184.

de Vasconcelos, Jose Romeu, "Matriz do Fluxo de Comércio Interestadual de Bens e Serviços no Brasil – 1999," *IPEA Discussion Paper*, 2001.

Diamond, Rebecca, "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000," *American Economic Review*, 2016, *106* (3), 479–524.

Dix-Carneiro, Rafael, "Trade Liberalization and Labor Market Dynamics," *Econometrica*, 2014, *82* (3), 825–885.

_ and Brian K Kovak, "Trade Liberalization and Regional Dynamics," *American Economic Review*, 2017, *107* (10).

Donaldson, D, "Railroads and the Raj: Estimating the Economic Impact of Transportation Infrastructure," *American Economic Review*, 2018, *108* (4-5), 899–934.

Donaldson, Dave and Richard Hornbeck, "Railroads and American Economic Growth: A "Market Access" Approach," *The Quarterly Journal of Economics*, 2016.

Duranton, Gilles and Diego Puga, *The Growth of Cities*, Vol. 2, Elsevier B.V., 2014.

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

Ehrl, Philipp, "Minimum comparable areas for the period 1872–2010: An aggregation of Brazilian municipalities," *Estudos Economicos*, 2017, *47* (1), 215–229.

Faber, Benjamin, "Trade Integration, Market Size, and Inudstrialization: Evidence from China's National Trunk Highway System," *Review of Economic Studies*, 2014, *81* (3), 1046–1070.

Ghani, Ejaz, Arti Grover Goswami, and William R Kerr, "Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing," *The Economic Journal*, 2014, *126* (March), 317–357.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, "Bartik Instruments: What, When, Why, and How," *American Economic Review*, 2020, *110* (8), 2586–2624.

Guedes, Ricardo, Felipe Iachan, and Marcelo Sant'Anna, "Housing Supply in the Presence of Informality," 2022.

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

Hassouna, M S and A A Farag, "MultiStencils Fast Marching Methods: A Highly Accurate Solution to the Eikonal Equation on Cartesian Domains," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2007, *29* (9), 1563–1574.

Janvry, Alain De, Kyle Emerick, Marco Gonzalez-Navarro, and Elisabeth Sadoulet, "Delinking Land Rights from Land Use : Certification and Migration in Mexico," *American Economic Review*, 2015, *105* (10), 3125–3149.

Jayachandran, Seema, "Selling Labor Low: Wage Responses to Productivity Shocks in Developing Countries," *Journal of Political Economy*, 2006, *114* (3), 538–575.

Kennan, John and James R. Walker, "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 2011, *79* (1), 211–251.

Michaels, G, "The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System," *The Review of Economics and Statistics*, 2008, *90* (4), 683–701.

Monte, Ferdinando, Stephen Redding, and Esteban Rossi-Hansberg, "Commuting, Migration, and Local Employment Elasticities," *American Economic Review*, 2018, *108* (12), 3855–3890.

Moretti, Enrico, "Local Labor Markets," *Handbook of Labor Economics*, 2011.

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

Pessino, C, "Sequential Migration Theory and Evidence from Peru," *Journal of Development Economics*, 1991, *36*, 55–87.

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

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

Sahota, Gian, "An Economic Analysis of Internal Migration in Brazil," *The Journal of Political Economy*, 1968, *76* (2), 218–245.

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

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

Simonovska, Ina and Michael Waugh, "Elasticity of Trade: Estimates and Evidence," *Journal of International Economics*, 2014, *92* (1), 34–50.

Sjaastad, L, "The Costs and Returns of Human Migration," *The Journal of Political Economy*, 1962, *70* (5), 80–93.

Tombe, Trevor and Xiadong Zhu, "Trade, Migration and Productivity: A Quantitative Analysis of China," *American Economic Review*, 2019, *109* (5), 1843–1872.

Topalova, P, "Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India," *American Economic Journal: Applied Economics*, 2010, *2* (4), 1–41.

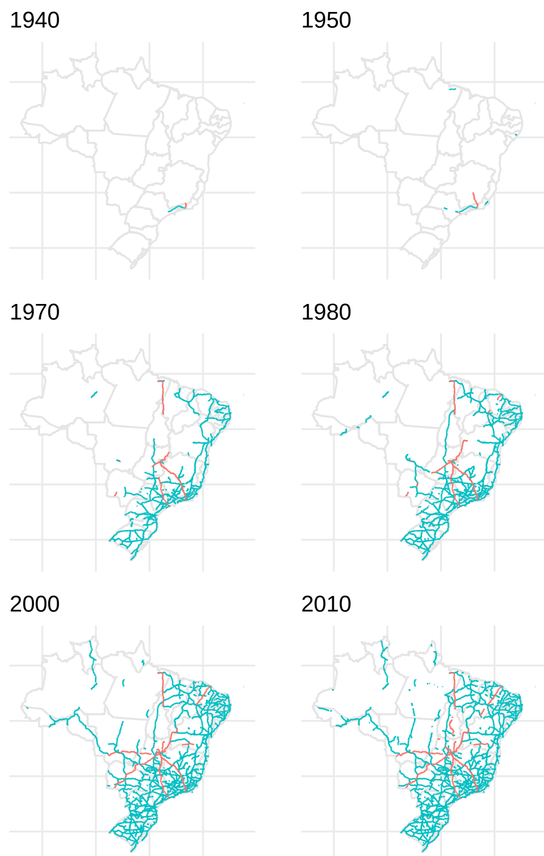
Tunali, Insan, "Rationality of Migration," *International Economic Review*, 2000, *41* (4), 893–920.

Appendix

Table of Contents

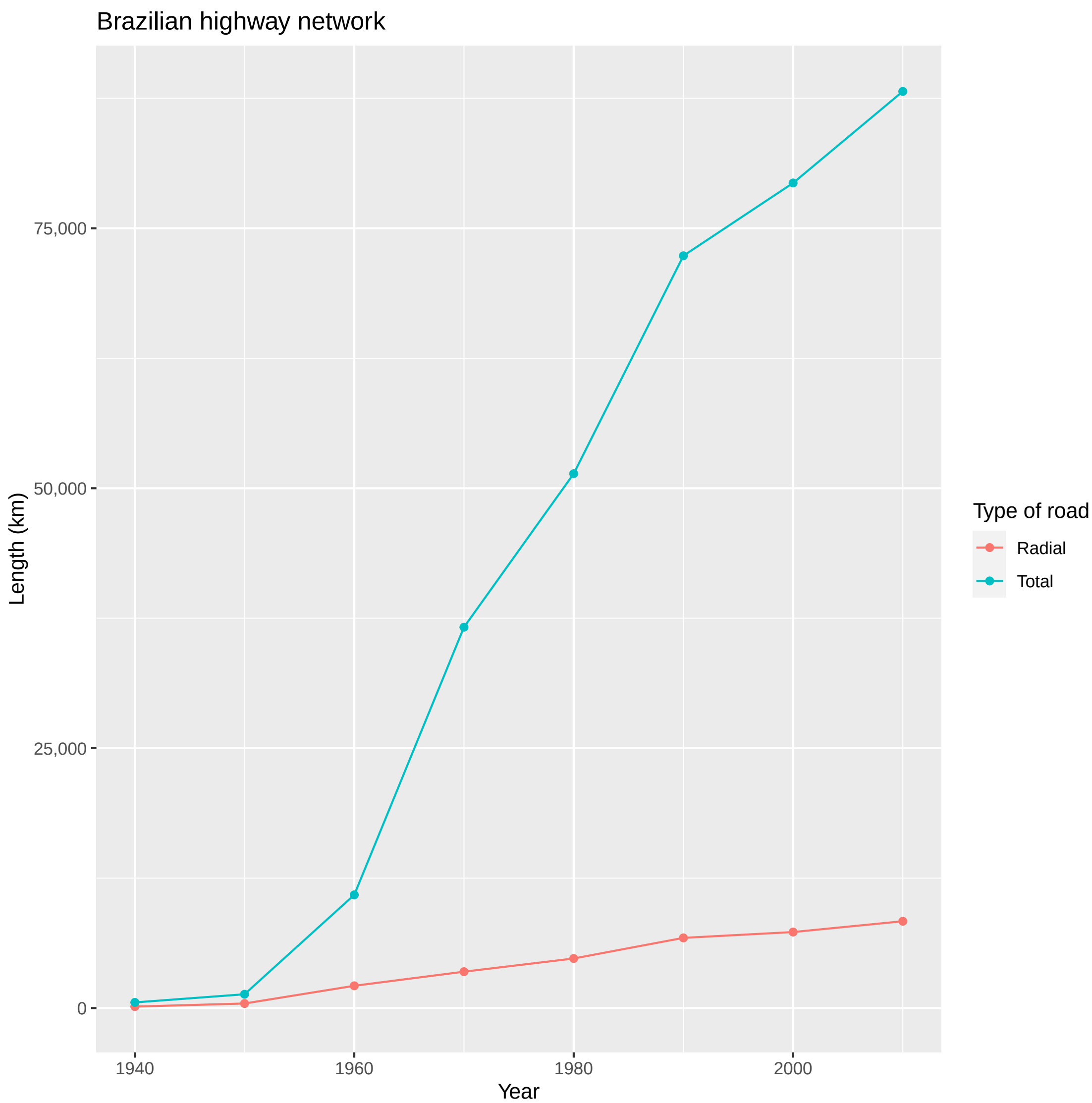
A Appendix Figures and Tables

Appendix Figure 1: Evolution of Brazil's federal highway system, 1940-2010



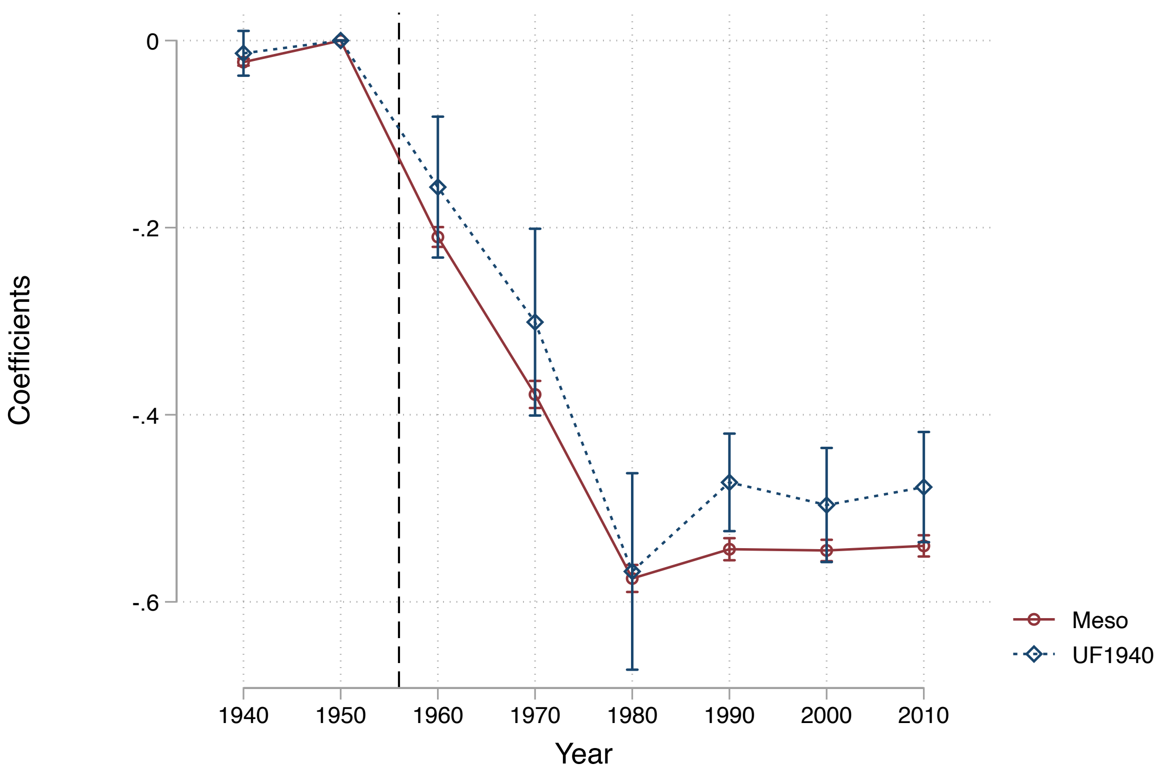
Notes: Figure is a map of the Brazilian road network. The map shows the non-radial highways in blue and radial highways in pink. Consistent state boundaries appear in the background of the map. Source: authors' calculations based on maps obtained from the Brazilian Ministry of Transportation.

Appendix Figure 2: The growth of the road network



Notes: Figure shows the total length of paved highways in Brazil. The figure shows non-radial highways in blue and radial highways in pink. Source: authors' calculations based on maps obtained from the Brazilian Ministry of Transportation.

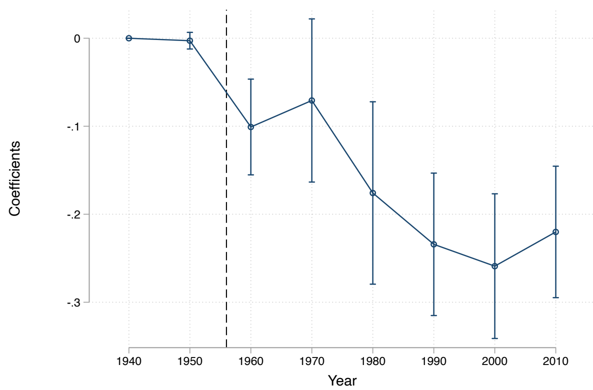
Appendix Figure 3: Time-varying effects of the predicted road network (State vs Meso geo-units)



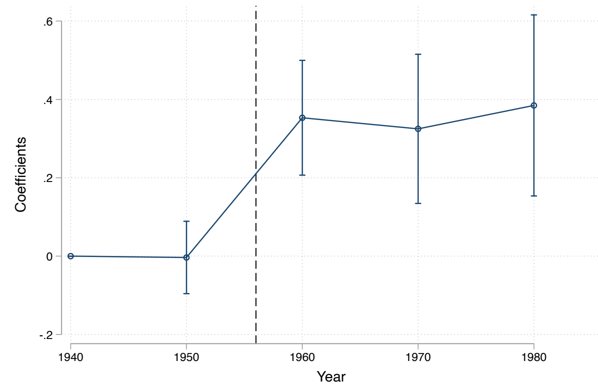
Notes: The graphs plot the estimated α_t coefficients from Equation (2) for state pairs (UF1940) and meso pairs (meso). The coefficient corresponding to 1950, the year immediately before Brasilia, is normalized to zero. For all other years, the coefficient measures the difference in the MST-induced changes in the outcome of interest between that year and the year 1950. Standard errors are clustered at the pair level.

Appendix Figure 4: Time-varying effects of the predicted road network (Imputation Estimator)

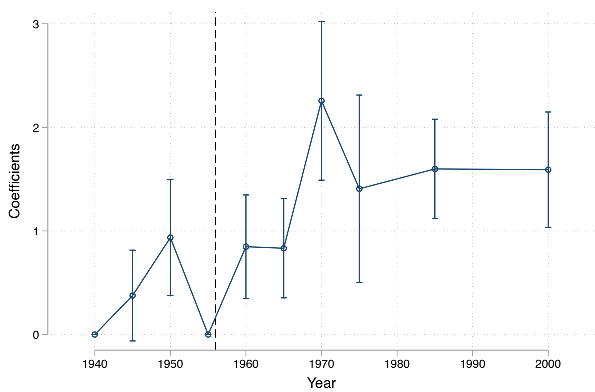
(a) Roads



(b) Migration



(c) Trade



Notes: The graphs plot the estimated α_t coefficients from Equation (2) for (a) travel time on the actual road network, (b) migration, and (c) trade using the did imputation estimator of Borusyak et al. (2021). did imputation needs a discrete treatment and a region is defined to be treated if the reduction of its travel time on the MST road network is above median.

Appendix Table 1: Correlation of MST-induced reduction in travel time with geographic variables

Dep. variable: reduction in travel time	(1)	(2)
Log distance	0.281	0.334
	(0.015)***	(0.013)***
Log distance origin to coast		0.020
		(0.016)
Log distance destination to coast		0.020
		(0.016)
Log distance origin to Brasilia		-0.152
		(0.019)***
Log distance destination to Brasilia		-0.158
		(0.019)***
Log distance origin nearest state capital		-0.082
		(0.023)***
Log distance destination nearest state capital		-0.083
		(0.023)***
N	420	420

Notes: Table shows regression coefficients form a regression of MST-induced reduction in travel time on geographical characteristics of the pair. The unit of analysis is a state-state pair. Regression is unweighted. Standard errors are unclustered as there is only one observation per pair. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Appendix Table 2: Time-varying effects of the predicted road network.

	Travel Time on Roads		Migration		Trade	
	(1) Before/After	(2) Year dummies	(3) Before/After	(4) Year dummies	(5) Before/After	(6) Year dummies
MST-induced Reduction in TT x After 1950	-0.392		0.746		0.879	
	(0.031)***		(0.174)***		(0.455)*	
MST-induced Reduction in TT x 1940		-0.014		-0.090		1.655
		(0.012)		(0.103)		(0.409)***
MST-induced Reduction in TT x 1945						1.094
						(0.342)***
MST-induced Reduction in TT x 1955						
MST-induced Reduction in TT x 1960		-0.157		0.414		1.276
		(0.038)***		(0.093)***		(0.429)***
MST-induced Reduction in TT x 1965						1.586
						(0.438)***
MST-induced Reduction in TT x 1970		-0.301		0.347		1.226
		(0.051)***		(0.144)**		(0.453)***
MST-induced Reduction in TT x 1975						4.900
						(1.386)***
MST-induced Reduction in TT x 1980		-0.568		0.589		
		(0.054)***		(0.166)***		
MST-induced Reduction in TT x 1985						2.327
						(0.521)***
MST-induced Reduction in TT x 1990		-0.472		0.796		
		(0.027)***		(0.188)***		
MST-induced Reduction in TT x 2000		-0.496		0.992		1.500
		(0.031)***		(0.202)***		(0.545)***
N	2940	2940	2939	2939	2913	2913

Notes: An observation is a state-pair-year. **Travel time**: travel time between state o and state d centroids on the actual federal highway network in year t, calculated using a Fast-Marching Algorithm. Data available decennially from 1940-2000. Road data source: Brazilian Ministry of Transportation. **MST-induced Reduction in TT**: change in travel times on the predicted road network (PRN), relative to travel times on an empty map. The PRN is created by using a Minimum Spanning Tree (MST) algorithm aiming to connect Brasilia's centroid to those of all other existing state capitals within each one of the eight pie slices defined by the cardinal and intercardinal directions. **Migration**: Measure of migration is the stock of people born in state o living in destination state d in time t. Data available decennially from 1940-2000. The maximum number of observations is 2940 (21 states*20 states*7 years). Migration data source: pre-1990: digitized from historical yearbooks. 1991-2000: microdata from population census. **Trade**: Measure of trade value is the value of trade from state o to destination d in time t. Data is available annnually from 1942-1949, 1959-1974, 1985, and 1998-1999. The maximum number of observations is 11340 (21 states*20 states*27 years) but some pairs are missing trade data because there was no trade between them or because it was not reported. Furthermore, to ease vizualization we have averaged the trade data over five years. Trade data source: pre-1998: digitized from historical yearbooks. 1998-99: de Vasconcelos and de Oliveira (2006). The omitted year in columns (2), (4), and (6) is 1950. Dependent variables are in logs. After 1950 is an indicator for whether the observation comes from years after 1950. Regressions are weighted by the migration and normalized trade flows so that each individual migration move or good transaction has equal weight. All regressions include state origin-year, state destination-year, and state pair fixed effects. Standard errors clustered at the pair level are reported in paretheses. Migration regression weighted by migration share. Trade regression weighted by trade share. * p<0.1, ** p<0.05, *** p<0.01

Appendix Table 3: Migration elasticities to travel time on roads - three-way clustered standard errors

	Migration			
	(1)	(2)	(3)	(4)
	OLS	IV	PPML	PPML
Log Travel Time on Roads	-0.332	-1.753	-0.560	-2.123
	(0.242)	(0.703)**	(0.283)**	(0.735)***
N	2939	2939	2940	2940
F-stat	1.883	6.219		
First-stage F stat		54.617		
First-stage residuals control			no	yes

Notes: An observation is a state-pair-year. **Log travel time on roads**: travel time between state o and state d centroids on the actual federal highway network in year t, calculated using a Fast-Marching Algorithm. Data available decennially from 1940-2000. Road data source: Brazilian Ministry of Transportation. Log travel time on roads is instrumented by the change in travel times on the predicted road network, relative to travel times on an empty map, interacted with year dummies (omitted year is 1950). The predicted road network is created by using a Minimum Spanning Tree (MST) algorithm aiming to connect Brasilia's centroid to those of all other existing state capitals within each one of the eight pie slices defined by the cardinal and intercardinal directions. **Migration**: Measure of migration is the stock of people born in state o living in destination state d in time t. Data is decennial covering 1940-2000. The maximum number of observations is 2940 (21 states*20 states*7 years). Data source: pre-1990: digitized from historical yearbooks. 1991-2000: microdata from Brazilian Population Census. **Trade**: Measure of trade value is the value of trade from state o to destination d in time t. Data is annual covering 1942-1949, 1959-1974, 1985 and 1998-1999. The maximum number of observations is 11340 (21 states*20 states*27 years) but some pairs are missing trade data because there was no trade between them or it was not reported. Data source: pre-1998: digitized from historical yearbooks. 1998-99: de Vasconcelos and de Oliveira (2006). All regressions include state origin-year, state destination-year, and state pair fixed effects. The dependent variables are in logs in columns (1), (2), (5), and (6), and in levels in columns (3), (4), (7), and (8). The Kleibergen-Paap F statistic (First-stage F stat) for weak identification is reported for IV estimates. Columns (3), (4), (7) and (8) present Poisson Pseudo Maximum Likelihood (PPML) estimates. Columns (4) and (8) add control for the first-stage regression residuals on the entire sample, including those with zero flows. Standard errors clustered at the origin, destination, and year levels reported in paretheses. OLS/IV regressions are weighted by the migration and normalized trade flows so that each individual migration move or good transaction has equal weight. Regressions are unweighted for the PPML regressions. * p<0.1, ** p<0.05, *** p<0.01.

Appendix Table 4: Migration and Trade elasticities to travel time on roads: First Stage

	Migration	Trade
	(1) OLS	(2) OLS
MST-induced Reduction in TT x 1940	0.042	
	(0.050)	
MST-induced Reduction in TT x 1942		-0.116
		(0.060)*
MST-induced Reduction in TT x 1943		-0.116
		(0.057)**
MST-induced Reduction in TT x 1944		-0.042
		(0.042)
MST-induced Reduction in TT x 1945		-0.011
		(0.041)
MST-induced Reduction in TT x 1946		-0.015
		(0.036)
MST-induced Reduction in TT x 1947		-0.012
		(0.038)
MST-induced Reduction in TT x 1948		-0.063
		(0.043)
MST-induced Reduction in TT x 1959		0.185
		(0.065)***
MST-induced Reduction in TT x 1960	-0.012	-0.131
	(0.075)	(0.114)
MST-induced Reduction in TT x 1961		-0.132
		(0.109)
MST-induced Reduction in TT x 1962		-0.130
		(0.112)
MST-induced Reduction in TT x 1963		-0.138
		(0.093)
MST-induced Reduction in TT x 1964		-0.119
		(0.109)
MST-induced Reduction in TT x 1965		-0.042
		(0.090)
MST-induced Reduction in TT x 1966		-0.136
		(0.112)
MST-induced Reduction in TT x 1967		0.006
		(0.087)
MST-induced Reduction in TT x 1968		-0.145
		(0.097)
MST-induced Reduction in TT x 1969		0.037
		(0.091)
MST-induced Reduction in TT x 1970	-0.041	0.089
	(0.106)	(0.149)
MST-induced Reduction in TT x 1971		-0.053
		(0.105)
MST-induced Reduction in TT x 1972		-0.098
		(0.103)
MST-induced Reduction in TT x 1973		-0.014
		(0.130)
MST-induced Reduction in TT x 1974		0.078
		(0.187)
MST-induced Reduction in TT x 1980	-0.331	
	(0.114)***	
MST-induced Reduction in TT x 1985		-0.414
		(0.096)***
MST-induced Reduction in TT x 1990	-0.311	
	(0.089)***	
MST-induced Reduction in TT x 1998		-0.384
		(0.081)***
MST-induced Reduction in TT x 1999		-0.394
		(0.082)***
MST-induced Reduction in TT x 2000	-0.294	
	(0.085)***	
N	2939	7451
First-stage F stat	22.086	11.217

Notes: An observation is a state-pair-year. Dependent variable in first-stage regressions is the log of travel time on roads. **Log travel time on roads**: travel time between state *o* and state *d* centroids on the actual federal highway network in year *t*, calculated using a Fast-Marching Algorithm. Data available decennially from 1940-2000. Road data source: Brazilian Ministry of Transportation.**MST-induced Reduction in TT**: change in travel times on the predicted road network, relative to travel times on an empty map. The predicted road network is created by using a Minimum Spanning Tree (MST) algorithm aiming to connect Brasilia's centroid to those of all other existing state capitals within each one of the eight pie slices defined by the cardinal and intercardinal directions. **Migration**: Measure of migration is the stock of people born in state *o* living in destination state *d* in time *t*. Data is decennial covering 1940-2000. The maximum number of observations is 2940 (21 states*20 states*7 years). Data source: pre-1990: digitized from historical yearbooks. 1991-2000: microdata from Brazilian Population Census. **Trade**: Measure of trade value is the value of trade from state *o* to destination *d* in time *t*. Data is annual covering 1942-1949, 1959-1974, 1985 and 1998-1999. The maximum number of observations is 11340 (21 states*20 states*27 years) but some pairs are missing trade data because there was no trade between them or it was not reported. Data source: pre-1998: digitized from historical yearbooks. 1998-99: de Vasconcelos and de Oliveira (2006). All regressions include state origin-year, state destination-year, and state pair fixed effects. Omitted year is 1950. The Kleibergen-Paap F statistic (First-stage F stat) for weak identification is reported for IV estimates. Standard errors clustered at the pair level are reported in paretheses. Regressions are weighted by the migration and normalized trade flows so that each individual migration move (or good transaction) has equal weight. * p<0.1, ** p<0.05, *** p<0.01.

Appendix Table 5: Elasticities to travel time on roads: PPML robustness to higher-order first-stage residual polynomials

	Migration					Trade				
	(1) 2nd	(2) 4th	(3) 6th	(4) 8th	(5) 10th	(6) 2nd	(7) 4th	(8) 6th	(9) 8th	(10) 10th
Log Travel Time on Roads	-2.155	-2.157	-2.098	-2.158	-2.168	-1.908	-1.938	-1.960	-1.585	-1.556
	(0.336)***	(0.336)***	(0.333)***	(0.321)***	(0.316)***	(0.442)***	(0.434)***	(0.440)***	(0.468)***	(0.471)***
N	2940	2940	2940	2940	2940	8960	8960	8960	8960	8960

Notes: Each column shows PPML estimates that add controls for polynomials of first-stage residuals up to the order displayed on the column heading. An observation is a state-pair-year. **Log travel time on roads**: travel time between state o and state d centroids on the actual federal highway network in year t, calculated using a Fast-Marching Algorithm. Data available decennially from 1940-2000. Road data source: Brazilian Ministry of Transportation. Log travel time on roads is instrumented by the change in travel times on the predicted road network, relative to travel times on an empty map, interacted with year dummies (omitted year is 1950). The predicted road network is created by using a Minimum Spanning Tree (MST) algorithm aiming to connect Brasilia's centroid to those of all other existing state capitals within each one of the eight pie slices defined by the cardinal and intercardinal directions. **Migration**: Measure of migration is the stock of people born in state o living in destination state d in time t. Data is decennial covering 1940-2000. The maximum number of observations is 2940 (21 states*20 states*7 years). Data source: pre-1990: digitized from historical yearbooks. 1991-2000: microdata from Brazilian Population Census. **Trade**: Measure of trade value is the value of trade from state o to destination d in time t. Data is annual covering 1942-1949, 1959-1974, 1985 and 1998-1999. The maximum number of observations is 11340 (21 states*20 states*27 years) but some pairs are missing trade data because there was no trade between them or it was not reported. Data source: pre-1998: digitized from historical yearbooks. 1998-99: de Vasconcelos and de Oliveira (2006). All regressions include state origin-year, state destination-year, and state pair fixed effects. Standard errors clustered at the pair level are reported in paretheses. OLS/IV regressions are weighted by the migration and normalized trade flows so that each individual migration move or good transaction has equal weight. Regressions are unweighted for the PPML regressions. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Appendix Table 6: Elasticity of trade flows to distance

Dep. var: log trade flows	(1)	(2)	(3)	(4)
Log distance	-1.477			-0.509
	(0.053)***			(0.207)**
Log traveltime		-1.633	-0.978	-1.101
		(0.054)***	(0.139)***	(0.218)***
Pair FE			✓	
N	7453	7453	7451	7453
F stat	790.839	927.396	49.771	463.823
widstat				

Notes: An observation is a state-pair-year. The dependent variable is log trade flows. Standard errors clustered at the pair level are reported in paretheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Appendix Table 7: Migration and Trade elasticities to travel time on roads - robustness to dropping Goias

	Migration				Trade			
	(1) OLS	(2) IV	(3) PPML	(4) PPML	(5) OLS	(6) IV	(7) PPML	(8) PPML
Log Travel Time on Roads	-0.419	-2.409	-0.676	-2.481	-0.786	-2.307	0.021	-1.866
	(0.119)***	(0.536)***	(0.125)***	(0.336)***	(0.158)***	(0.556)***	(0.222)	(0.451)***
N	2660	2660	2660	2660	6734	6734	8132	8132
F-stat	12.474	20.214			24.735	17.227		
First-stage F stat		18.681				14.036		
First-stage residuals control			no	yes			no	yes

Notes: Sample excludes pairs involving Goias, the state which houses Brasilia. An observation is a state-pair-year. **Log travel time on roads**: travel time between state o and state d centroids on the actual federal highway network in year t, calculated using a Fast-Marching Algorithm. Data available decennially from 1940-2000. Road data source: Brazilian Ministry of Transportation. Log travel time on roads is instrumented by the change in travel times on the predicted road network, relative to travel times on an empty map, interacted with year dummies (omitted year is 1950). The predicted road network is created by using a Minimum Spanning Tree (MST) algorithm aiming to connect Brasilia's centroid to those of all other existing state capitals within each one of the eight pie slices defined by the cardinal and intercardinal directions. **Migration**: Measure of migration is the stock of people born in state o living in destination state d in time t. Data is decennial covering 1940-2000. The maximum number of observations is 2940 (21 states*20 states*7 years). Data source: pre-1990: digitized from historical yearbooks. 1991-2000: microdata from Brazilian Population Census. **Trade**: Measure of trade value is the value of trade from state o to destination d in time t. Data is annual covering 1942-1949, 1959-1974, 1985 and 1998-1999. The maximum number of observations is 11340 (21 states*20 states*27 years) but some pairs are missing trade data because there was no trade between them or it was not reported. Data source: pre-1998: digitized from historical yearbooks. 1998-99: de Vasconcelos and de Oliveira (2006). All regressions include state origin-year, state destination-year, and state pair fixed effects. The dependent variables are in logs in columns (1), (2), (5), and (6), and in levels in columns (3), (4), (7), and (8). The Kleibergen-Paap F statistic (First-stage F stat) for weak identification is reported for IV estimates. Columns (3), (4), (7) and (8) present Poisson Pseudo Maximum Likelihood (PPML) estimates. Columns (4) and (8) add control for the first-stage regression residuals on the entire sample, including those with zero flows. Standard errors clustered at the pair level are reported in paretheses. OLS/IV regressions are weighted by the migration and normalized trade flows so that each individual migration move or good transaction has equal weight. Regressions are unweighted for the PPML regressions. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Appendix Table 8: Estimating trade elasticity

	(1) OLS	(2) IV	(3) First stage
Change log wage	0.890	-70.793	
	(1.524)	(544.661)	
Bartik instrument			0.383
			(2.136)
Implied theta	-1.124	0.014	
	(1.925)	(0.109)	
N			21
F stat	0.341	0.015	
First stage F stat		0.016	0.032

Notes: Unit of analysis is a state and each observation is the 1999-1985 difference in estimated origin fixed effect. Instruments for changes in wages are state-level Bartik shocks. The Kleibergen-Paap F statistic (First-stage F stat) for weak identification is reported for IV estimates. Standard errors are unclustered. Regressions unweighted. * p<0.1, ** p<0.05, *** p<0.01.

Appendix Table 9: Comparing response across geography and stock/flow

	Working men			Entire population			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep var: log migration flow	Meso flow	UF flow	UF stock	Meso flow	UF flow	UF stock	Hist. UF stock
Panel a: Distance							
Log distance	-1.352	-1.296	-1.505	-1.564	-1.389	-1.538	-1.947
	0.014***	0.074***	0.086***	0.014***	0.076***	0.086***	0.096***
Panel b: Traveltime on road							
Log traveltime road	-1.602	-1.636	-1.936	-1.863	-1.756	-1.986	-2.108
	0.019***	0.098***	0.111***	0.019***	0.101***	0.110***	0.099***
Panel c: Traveltime on mst							
Log traveltime mst	-1.694	-1.683	-1.965	-1.978	-1.806	-2.005	-2.525
	0.018***	0.105***	0.122***	0.019***	0.109***	0.122***	0.146***
N	43640	1670	1680	51511	1677	2100	2099
r2	0.600	0.822	0.829	0.610	0.821	0.830	0.798
origFE	✓	✓	✓	✓	✓	✓	✓
destFE	✓	✓	✓	✓	✓	✓	✓
histsample							✓
microsample	✓	✓	✓	✓	✓	✓	

Notes: Table shows regression results from estimating gravity equations on log migration. The dependent variable is either meso-meso migration over the last five years (meso flow), state-state migration over the last five years (UF flow), or migration outside state of birth (UF stock). Datasource is the Brazilian census microdata between 1980 and 2010 for columns (1)-(6). The sample for column (7) is the historical data for 1940-1980 digitized from historical yearbooks. Working men are men between 20-65 with non-zero income. All is the entire Brazilian population. Standard errors clustered at the orig-dest pair. Regression is unweighted. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Appendix Table 10: Estimating distance coefficient (meso level)

	Poisson		Linear	
	(1) b/se	(2) b/se	(3) b/se	(4) b/se
Log distance	-1.705	-1.698	-1.596	-1.597
	(0.035)***	(0.035)***	(0.014)***	(0.014)***
Estimated change in mig costs	170.367		-148.573	
	(71.609)**		(179.058)	
N	55896	55896	37939	37939

Notes: An observation is a meso-pair year in the range 1980-2000. The estimated change in mig costs are the results implied from the state-level gravity regressions. In columns (2) and (4) the value of the estimated change in mig (trade) costs is constrained to be 1 and is not reported. Standard errors are clustered at the pair level. Regressions are unweighted. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Appendix Table 11: Estimating distance coefficient (state-level)

	Migration		Trade	
	(1)	(2)	(3)	(4)
Log distance	-1.230	-1.277	-0.996	-1.187
	(0.160)***	(0.124)***	(0.088)***	(0.056)***
Estimated change in mig costs	0.793			
	(0.509)			
Estimated change in trade costs			0.194	
			(0.257)	
N	1260	1260	1260	1260

Notes: An observation is a state-pair year in the range 1980-2000. The estimated change in mig (trade) costs are the results implied from the state-level gravity regressions. In columns (2) and (4) the value of the estimated change in mig (trade) costs is constrained to be 1 and is not reported. Standard errors are clustered at the pair level. Regressions are unweighted. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Appendix Table 12: First stage: migration and housing elasticity

	Mig elasticity		Housing elasticity
	(1)	(2)	(3)
	Cost	Nocost	
Bartik shock	1.801	1.801	-0.347
	(0.461)***	(0.461)***	(1.042)
Predicted change in log pop, IV	0.097	0.097	0.282
	(0.036)***	(0.036)***	(0.079)***
N	137	137	137
First-stage F stat	16.103	16.103	6.395

Notes: Unit of analysis is a meso-region and each observation is the 2010-1980 difference. The dependent variable in columns (1) and (2) is the 2010-1980 change in (log) indirect utility of the meso-region. The indirect utilities for years 2010 and 1980 are obtained as the set of (meso) destination-year fixed effects after estimating a gravity equation from meso-to-meso bilateral migration flows, accounting for bilateral migration costs (as a linear function of distance and travel times in logs) and (meso) origin-year fixed effects. The dependent variable in columns (3) and (4) are also the 2010-1980 change in indirect utilities from meso-to-meso flows, but assuming zero bilateral migration costs. The dependent variable in columns (5) and (6) is the change in (log) rental prices. Rents are calculated from census micro-data in each year as the meso-specific average after netting out housing characteristics such as number of rooms and bedrooms, electricity access, walls, roof, and floor quality, among others. Instruments for changes in wages and housing expenditure are meso-level Bartik shocks and a model-based measure of changes in labor as a function of Bartik shocks of all meso-regions weighted by migration costs. The Kleibergen-Paap F statistic (First-stage F stat) for weak identification is reported for IV estimates. Standard errors clustered at the state level. Regressions unweighted. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Appendix Table 13: Imputed prices: migration and housing elasticity

	Baseline		Imputed prices		Raw rents		F.S. baseline	F.S. prices		F.S. raw. rents
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7)	(8)	(9)	(10)
Change in wages, adjusted for residualized rents	0.029	4.515	1.239	5.841						
	(0.295)	(0.763)***	(0.500)**	(1.136)***						
Change log imputed prices			0.197	7.983						
			(1.496)	(3.925)*						
Change in wages, adjusted for rents					1.553	5.653				
					(0.465)***	(0.815)***				
Bartik shock							1.801	1.471	0.001	1.416
							(0.461)***	(0.421)***	(0.033)	(0.431)***
Predicted change in log pop, IV							0.097	0.090	0.001	0.115
							(0.036)***	(0.035)**	(0.002)	(0.031)***
Change log imputed prices IV								-0.381	0.359	
								(0.160)**	(0.019)***	
N	137	137	137	137	137	137	137	137	137	137
F stat	0.009	35.000	3.286	13.329	11.132	48.112				
SW F		16.103		8.423		21.687	16.103	8.423	8.423	21.687

Notes: Unit of analysis is a meso-region and each observation is the 2010-1980 difference. The dependent variable in columns (1) and (2) is the 2010-1980 change in (log) indirect utility of the meso-region. The indirect utilities for years 2010 and 1980 are obtained as the set of (meso) destination-year fixed effects after estimating a gravity equation from meso-to-meso bilateral migration flows, accounting for bilateral migration costs (as a linear function of distance and travel times in logs) and (meso) origin-year fixed effects. The dependent variable in columns (3) and (4) are also the 2010-1980 change in indirect utilities from meso-to-meso flows, but assuming zero bilateral migration costs. The dependent variable in columns (5) and (6) is the change in (log) rental prices. Rents are calculated from census micro-data in each year as the meso-specific average after netting out housing characteristics such as number of rooms and bedrooms, electricity access, walls, roof, and floor quality, among others. Instruments for changes in wages and housing expenditure are meso-level Bartik shocks and a model-based measure of changes in labor as a function of Bartik shocks of all meso-regions weighted by migration costs. The Kleibergen-Paap F statistic (First-stage F stat) for weak identification is reported for IV estimates. Standard errors clustered at the state level. Regressions unweighted. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Appendix Table 14: Robustness of welfare to elasticity and distance elasticity: equalizing coefficients

	Trade dist. elast.	Trade elast.	Mig dist. elast	Mig elast.	Change trade tau	Change mig tau	Change welfare	Share welfare trade
Baseline	0.47	4.00	0.39	4.51	0.10	0.08	1.03	0.76
Set to trade coeff.	0.47	4.00	0.47	4.00	0.10	0.10	1.03	0.71
Set to mig. coeff.	0.39	4.51	0.39	4.51	0.08	0.08	1.02	0.72

Notes: Table shows the relative change. All values are relative to a baseline value of 1. The numbers are computed by simulating the structural model under different elasticities. Table equalizes the migration and trade distance coefficients (and migration and trade elasticities) either to the estimated migration values or the estimated trade values.

Appendix Table 15: Robustness of welfare to elasticity and distance elasticity

	Half			Baseline			Double		
	0.5x	1x	2x	0.5x	1x	2x	0.5x	1x	2x
Trade elasticity/distance elasticity									
Elasticity of trade costs to roads	0.24	0.24	0.24	0.47	0.47	0.47	0.94	0.94	0.94
Trade elasticity	2.00	4.00	8.00	2.00	4.00	8.00	2.00	4.00	8.00
Change in welfare	1.02	1.02	1.02	1.03	1.03	1.03	1.05	1.06	1.08
Implied reduction in trade costs	0.05	0.05	0.05	0.10	0.10	0.09	0.19	0.18	0.18
Share of welfare gain due to trade	0.57	0.59	0.61	0.74	0.76	0.79	0.86	0.88	0.91
Migration elasticity/distance elasticity									
Elasticity of mig costs to roads	0.19	0.19	0.19	0.39	0.39	0.39	0.78	0.78	0.78
Migration elasticity	2.26	4.51	9.03	2.26	4.51	9.03	2.26	4.51	9.03
Change in welfare	1.02	1.02	1.02	1.03	1.03	1.03	1.04	1.04	1.05
Implied reduction in trade costs	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
Share of welfare gain due to trade	0.88	0.87	0.86	0.77	0.76	0.71	0.60	0.55	0.42

Notes: Table shows the relative change. All values are relative to a baseline value of 1. The numbers are computed by simulating the structural model under different elasticities. The first two rows of each panel report the value of the elasticity considered. The results reported in Table 4 correspond to column (5) i.e., Baseline elasticty of trade (migration) costs to roads and 1x the value of the trade (migration) elasticity.

B GIS and Road Data

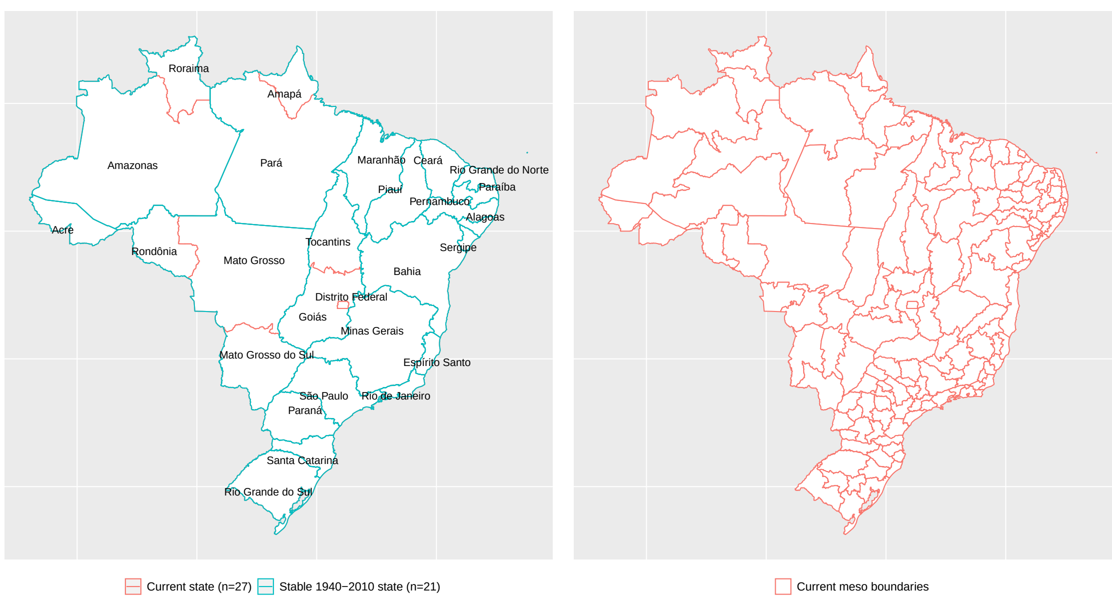
B.1 Geographical units

For all GIS data, we use the SIRGAS 2000 (EPSG:4674) coordinate system. To compute distances, we project the shape file using the Brazil SIRGAS 2000 Polyconic (EPSG:5880) projection.

The analysis in the paper occurs at either the state level or at the meso region level. State boundaries change over time: to generate consistent boundaries from 1940 we construct a set of 21 consistent states.[32] The second unit of analysis are meso regions, of which there are 137 in Brazil.

Appendix Figure 5: Geographical units

(a) States (b) Meso regions



Notes: Figure is a map of Brazil. Map (a) shows the current state boundaries (n=27) and how the consistent state boundaries are constructed (n=21). Map (b) shows the mesoregion boundaries (n=137).

We create a crosswalk between the 2010 meso boundaries and the municipalities in each year between 1970 and 2010. This crosswalk is based on

[32]There were 21 states in Brazil in 1940 and there are 27 states in Brazil now. Those 21 states still exist and 6 states, including the Federal District (Brasilia), split from those states. Three states split from and returned to the 1940 states between 1940 and 2010. The old Federal District became the State of Guanabara which merged into the State of Rio de Janeiro (one of the 1940 states).

the official division of the Brazilian territory (*Divisão Territorial Brasileira*) produced by IBGE. The URL to download these files (as of August 2021) is: https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-te 23701-divisao-territorial-brasileira.html?edicao=23704&t=downloads. We cross-referenced our crosswalk with the crosswalk created by Ehrl (2017), available from download (as of August 2021) at: https://sites.google.com/site/philippehrl/research.

B.2 Road data

B.2.1 1960-2010

The main data source for roads between 1960 and 2010 is the Brazilian Ministry of Transportation. We downloaded the shapefiles for each year between 1960 and 2010 from https://www.gov.br/infraestrutura/pt-br/assuntos/dados-de-transportes/bit/bitmodosmapas#maprodo on 3/21/2021. The shapefiles are appended together to generate the cumulative road network for each year. We call this data source the MOT database.

The MOT database contains a variable that indicates the road surface status (leg_multim). This variable describes whether the road is planned, paved, or unpaved. Our analysis sample is all roads that are either paved or unpaved (drop leg_multim contains "Planejada"). The paved sample is roads that are paved (leg_multim contains the strings "Duplicada" (double-paved road) or "Pavimentada" (paved road)).[33]

B.2.2 1940-1950

The only paved road in the 1940s was the road connecting Rio de Janeiro to Sao Paulo (Rodrigues, 2008). We identified paved roads built in the 1950s using (Planos Nacionais de Viacao) which contains a map showing paved roads in 1956. We selected the 1960 paved roads that matched the road from Rio de Janeiro to Sao Paulo and the paved roads in the 1956 map to construct the 1940 and 1950 road data.

[33] The variable ds_superfi also indicates whether a road is paved, unpaved, or planned. This variable appears to designate the federal status of the road and not the status of the road itself. For example, an existing unpaved road may be a planned paved road, or an existing paved state highway may be a planned paved federal highway. We use the leg_multim variable instead to be consistent with the actual road that exists.

B.2.3 Verification against OSM

To check for the completeness of the road network we verify it against Open Street Maps.[34] We downloaded the entire OSM road network for Brazil from http://download.geofabrik.de/south-america/brazil-latest.osm.pbf on 7/22/2021. From this database, we keep road segments that have a highway classification as motorway, trunk, primary, or secondary. The OSM data contains both federal and state roads. To match the MOT data, which is only federal highways, we create a crosslisted dataset which contains all road segments that include the prefix for a federal highway in Brazil (a "BR-").[35] The road surface (i.e., paved or unpaved) is not consistently marked in the OSM database. To account for small differences in the two shape files we consider a 1km buffer when intersecting.

Share of MOT roads in OSM

- 1.5% of the paved 2010 MOT roads are not in the 2021 OSM database
- 5.3% of the paved 2010 MOT roads are not in the crosslisted 2021 OSM database

This is a fairly high match rate and the small degree of mismatch when we restrict the OSM to the crosslisted roads is likely due to the imperfect inclusion of highway codes.

Share of OSM roads in MOT

- 12.4% of the crosslisted 2021 OSM roads are not in the 2010 MOT analysis data
- 12.1% of the crosslisted 2021 OSM roads are not in the 2010 MOT data (including planned roads)

This number is slightly high. One reason that OSM roads could be missing from the 2010 database is that they were constructed after 2010 and not necessarily planned in 2010. To check if this is the explanation, we use an additional dataset available from the Brazilian Ministry of Transportation (at the same website listed above) that has the 2019 highway system. 3.8%

[34] We cannot systematically use Google maps for the verification exercise because the data are not downloadable.

[35] Some roads are coincidental state and federal highways (for example, the BR-010 highway runs from Belém to Brasília and contains crosslisted state and federal segments including GO-118 (in the state of Goias), DF-345 (in the Distrito Federal), and TO-387 (in the state of Tocantins)).

of the crosslisted 2021 OSM roads are not in the 2019 MOT analysis data The remaining mismatch is primarily due to roads built between 2019 and 2021: if we expand the 2019 MOT to include planned roads, only 2.9% of the crosslisted 2021 OSM roads are not in the 2019 MOT data.

Overall, the mismatch rate is around 5% between the MOT and OSM database. Given this number, we feel confident that our road data matches the existing highway system in Brazil.

B.3 Construction of MST network

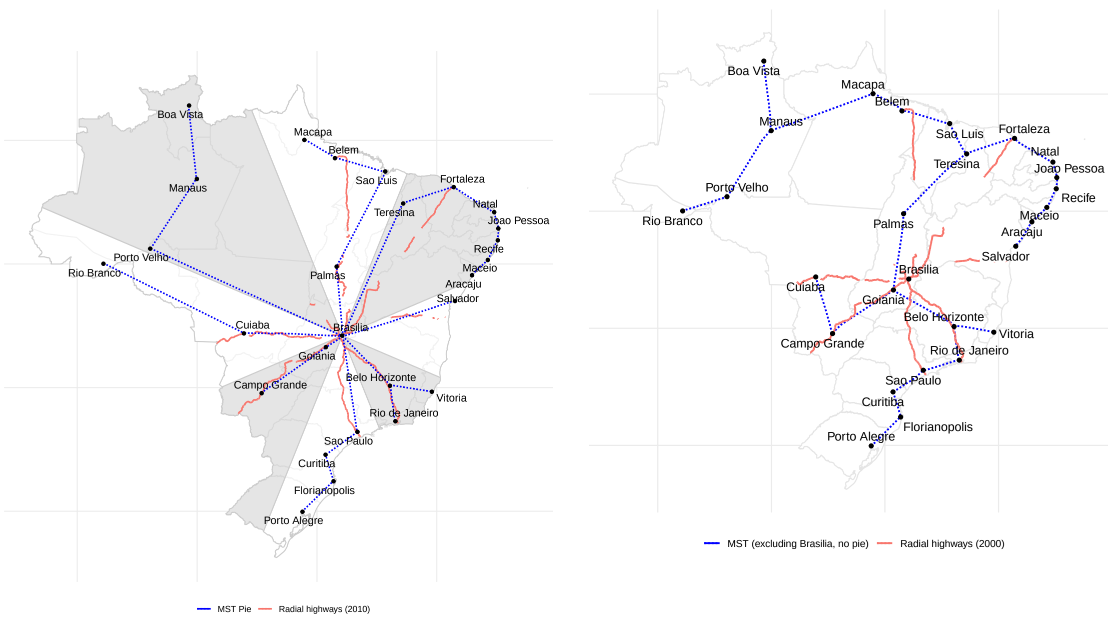
We use ArcGIS to compute the minimum spanning tree (MST) network. We use latitude-longitude coordinates of the geo-units' centroids to create point features representing the location of Brasilia and the 26 state capitals. Next, we divide the country into 8 exogenous equal-sized slices, centered around 0N. We then use the *Optimal Region Connection* tool in ArcGIS Pro to connect the cities within the pie slice. Figure 6 illustrates this process.

To construct the counterfactual Rio network, we eliminate the Brasilia city point and then connect the remaining 26 cities using the *Optimal Region Connection* tool in ArcGIS Pro.

Appendix Figure 6: Construction of MST networks

(a) MST pie network

(b) MST network without Brasilia



Notes: Figure is a map of the Brazilian road network indicating Brasilia and the 26 state capitals. The map shows radial highways out of Brasilia and the straight-line instrument for roads. Panel (a) shows the eight equal-sized pie slices used to construct the MST instrument centered around Brasilia. Panel (b) shows the alternative MST network. Consistent state boundaries appear in the background of the map. Source: authors' calculations based on maps obtained from the Brazilian Ministry of Transportation.

B.4 Construction of road travel times

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 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. We compute the fast marching method in R using the algorithm **fastmaRching.R**, available on CRAN at `https://cran.r-project.org/package=fastmaRching`. To run the algorithm we assign a speed of 10 to non-paved roads and a speed of 100 to paved roads.

B.4.1 Validation of road travel time measures

To validate the road travel time measure, we calculated travel time between pairs of 398 populated areas in Brazil using our travel time measure on the 2010 road network and travel time scraped from Google Maps. Google Maps requires a location to be relatively close to a road network to compute travel times. The thus use a list of 398 populated places and run our fast marching algorithm and compare it to Google maps travel times. Appendix Table 16 shows that the two measures of travel time are highly correlated, even after controlling for the log distance between the two locations.

C Datasources

C.1 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, 1954-1975, and 1985. 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. The yearbook

Appendix Table 16: Validation of travel time measure with google travel times

	All cities		Exclude close to capital	
	(1)	(2)	(3)	(4)
Dep var: log google time				
Log road travel time	1.121	0.183	1.144	0.178
	0.001***	0.002***	0.001***	0.002***
Log dist km		0.793		0.800
		0.001***		0.002***
origFE	✓	✓	✓	✓
destFE	✓	✓	✓	✓
N	133605	133605	74265	74265
r2	0.958	0.989	0.956	0.989

Notes: Table shows regression results from estimating a gravity equation where the dependent variable is the log travel time between two cities computed by Google. Log road travel time is computed on the 2010 road network. An observation is a city-city pair. The all cities sample is travel time between 398 populated cities. Excluding close to capital excludes any origin or destination within 100 km of a state capital. Standard errors are not clustered as there is only one observation per pair. Regression is unweighted.

data reported imports and exports. However, not all states are reported. We differentiate between missing data and zero flows as follows. If we observe a reported import or reported export from a location we assume that location reports all its data and so treat any missing data as true zeros. If we do not observe any reported imports or exports we treat the data as missing. Appendix Table 17 summarizes the pair-level coverage of the trade data. We drop any year where more than 75% of the pair-level data is missing. Years that are dropped (1954-1958 and 1975) are indicated in the final column of the table.

For 1998 and 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. The data are collected by double accounting, converting reported imports and reported exports. If we observe a reported import or reported export from a location we assume that location reports all its data and so treat any missing data as true zeros. If we do not observe any reported imports or exports we treat the data as missing. Appendix Table 17 summarizes the pair-level coverage of the trade data. We drop any year where more than 75% of the pair-level data is missing. Years that are dropped (1954-1958 and 1975) are indicated in the final column of the table.

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.

C.2 Census database, 1970-2010

We construct a regional database of migration, wages and roads at the meso level between 1970–2010. Summary statistics for the regional database are presented in Appendix Table 18). The primary datasource is the individual data files from the Brazilian Census, 1970–2010, collected by the Brazilian Institute of Geography and Statistics (IBGE).[36]

[36]For the purposes of sampling, the national territory is divided in "setores censitarios" (census blocks). Within each sector, a fraction of the households is randomly selected and the questionnaires are administered. The fraction of households sampled within a sector has varied across census years. In 1970 and 1980,

Appendix Table 17: Classification of trade data by year

	(1) Positive	(2) Zero	(3) Missing	(4) Dropped from analysis
1942	0.52	0.48	0.00	0.00
1943	0.65	0.35	0.00	0.00
1944	0.58	0.28	0.14	0.00
1945	0.61	0.20	0.19	0.00
1946	0.60	0.25	0.14	0.00
1947	0.62	0.19	0.19	0.00
1948	0.53	0.28	0.19	0.00
1949	0.59	0.27	0.14	0.00
1954	0.14	0.00	0.86	1.00
1955	0.19	0.00	0.81	1.00
1956	0.05	0.00	0.95	1.00
1957	0.05	0.00	0.95	1.00
1958	0.19	0.00	0.81	1.00
1959	0.54	0.08	0.38	0.00
1960	0.68	0.08	0.24	0.00
1961	0.77	0.09	0.14	0.00
1962	0.75	0.10	0.14	0.00
1963	0.76	0.10	0.14	0.00
1964	0.73	0.13	0.14	0.00
1965	0.67	0.09	0.24	0.00
1966	0.75	0.06	0.19	0.00
1967	0.69	0.07	0.24	0.00
1968	0.68	0.08	0.24	0.00
1969	0.66	0.06	0.29	0.00
1970	0.52	0.05	0.43	0.00
1971	0.52	0.05	0.43	0.00
1972	0.52	0.10	0.38	0.00
1973	0.49	0.04	0.48	0.00
1974	0.35	0.08	0.57	0.00
1975	0.11	0.03	0.86	1.00
1985	0.99	0.01	0.00	0.00
1998	0.97	0.03	0.00	0.00
1999	0.99	0.01	0.00	0.00
Total	0.56	0.11	0.33	0.18
N	13860	13860	13860	14553

Notes: Classification of trade data into missing, zero, or positive flow. 420 total pairs (21 origin x 20 destination states as own-trade is never measured). We drop years from the analysis if 75% or more of the pair-level data is missing for that year.

Appendix Table 18: Census summary stats

	(1) 1970	(2) 1980	(3) 1991	(4) 2010
Live in Brasilia	0.006	0.011	0.011	0.015
Live in Rio de Janeiro	0.085	0.088	0.074	0.068
Migrated between UF	.	0.054	0.043	0.032
Migrated between meso	.	0.095	0.071	0.053
Report previous meso	0.000	0.986	0.995	0.992
Report previous UF	0.000	0.986	0.995	0.992
Mean wage	2.979	6.124	5.184	8.403
Mean rent	219.033	295.956	302.135	364.283
Share renting	0.176	0.224	0.158	0.186
Avg years school	2.781	3.812	5.438	8.623
N	4733022	5896078	3540516	4471780

Notes: Summary stats computed on sample of men with non-zero earnings between 20-65 years old. All nominal variables converted to 2010 Brazilian reals. Migration measured based on location five years prior. Summary stats weighted by IBGE-provided individual weights. Data source: Brazilian census microdata/

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. [37]

C.2.1 Employment and wages

Wage data are sourced from the census. The census asks both the average earnings per month in the main occupation,[38] as well as the usual hours worked. We use earnings from main occupation and the hours worked to

25% of households were drawn from the population. In 1991 and 2000, the fractions varied according to municipality size. For municipalities with up to 15,000 inhabitants, 20% of their population was sampled; the fraction was 10% for municipalities with more than 15,000 people. In 2010, there were five fractions. In municipalities with up to 2,500 inhabitants, 50% were sampled; in municipalities with more than 2,500 but less than 8,000 inhabitants, the sampling fraction was 33%; in municipalities with more than 8,000 and less than 20,000 inhabitants, 20% were sampled; in municipalities with more than 20,000 and less than 500,000 people, 10% were sampled; and in municipalties with mora than 500,000 inhabitants, the fraction was 5%.

[37] 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.

[38] The exception is 1970, where only total earnings, rather than earnings in the main occupation, is asked.

construct an equivalent hourly wage rate. Assuming a standard 2000 hour work year, the annual wage of 6.4 BRL in 2010 would be equivalent to annual income of $5565. The per capita GDP figures for Brazil was $5600 in 2010 (World Development Indicators) and so the wage is the correct magnitude.

C.2.2 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. To get consistent geographic boundaries over time we aggregate municipalities to match the 2010 meso and UF boundaries. Migration data is first reported in the 1980 census; we see that the interstate migration rate is between 5.3% in 1980 to 3.2% in 2010. The inter-meso migration rate is naturally higher, at 9.5% in 1980 and 5.3% in 2010.

C.2.3 Rental prices

For rental rates we use census data on the rents paid for housing. The mean rental rate in 2010 is 321 BRL a month, equivalent to 50 hours of work at the mean wage. 18.5% of the population report paying rents for their housing in 2010. While this may seem low, the equivalent number for US houses in 2005 is 24%.

Appendix Table 19 tabulates the observable characteristics for housing quality collected in each year. Appendix Table 20 shows the hedonic regressions for rent on observables by year. The residuals from these regressions are used to construct the residualized rent for each meso region.

D Theoretical derivations

D.1 Exact hat derivation

This appendix describes the procedure to compute the exact-hat changes in labor and prices:

D.1.1 Labor

As given by Equation 5, labor in location d is given by:

Appendix Table 19: Summary statistics of hedonic variables

	(1) 1970	(2) 1980	(3) 1991	(4) 2010
Number of rooms	4.855	5.274	5.640	5.922
Number of bedrooms	2.190	2.216	2.150	2.216
Sanitation	0.122	0.288	0.346	0.577
Public water	0.311	0.560	0.696	0.836
Has electricity	0.460	0.709	0.867	0.992
Rural	0.453	0.294	0.234	0.126
House made from good materials	0.260			
House not apartment		0.934	0.922	0.996
Walls made from durable materials		0.701	0.793	0.978
Roof made from durable materials		0.914	0.945	
Floor made from durable materials		0.830		
Has toilet			0.783	0.954
N	4733048	5782473	3540519	4471780

Notes: An observation an individual from our sample of men with non-zero earnings aged between 20 and 65 years. We note that the sample includes individuals from all kinds of housing units, whether or not the unit is a rental. Table presents average values of available housing variables for each census year. Averages are weighted by IBGE-provided individual weights. Data source: Brazilian census microdata.

Appendix Table 20: Hedonic Regressions

Dep. var: log rent	(1) 1970	(2) 1980	(3) 1991	(4) 2010
Number of rooms	0.154	0.178	0.162	0.159
	0.016***	0.007***	0.003***	0.004***
Number of bedrooms	-0.013	-0.003	-0.001	0.024
	0.006**	0.007	0.007	0.003***
Sanitation	0.339	0.280	0.243	0.262
	0.119***	0.015***	0.026***	0.011***
Public water	0.316	0.271	0.245	0.041
	0.043***	0.016***	0.016***	0.025*
Has electricity	0.351	0.429	0.390	0.373
	0.023***	0.021***	0.025***	0.030***
Rural	-0.271	-0.325	-0.469	-0.314
	0.040***	0.030***	0.042***	0.022***
House made from good materials	-0.183			
	0.024***			
House not apartment		-0.424	-0.369	0.000
		0.087***	0.032***	.
Walls made from durable materials		0.302	0.302	0.356
		0.029***	0.021***	0.017***
Roof made from durable materials		0.047	0.051	
		0.012***	0.020**	
Floor made from durable materials		0.287		
		0.020***		
Has toilet			0.244	0.291
			0.018***	0.044***
N	808468	1257066	498763	730640

Notes: Table reports regression results from running hedonic regressions. The dependent variable is log rent. Not all variables are available in each survey year. Standard errors are clustered at the mesoregion. Regression is unweighted. Data source: Brazilian census microdata.

$$
\begin{aligned}
L_{dt} &= \sum_o N_{odt} \\
&= \sum_o V_{dt}^{\epsilon} \kappa_{odt}^{-\epsilon} \Phi_{ot}^{-1} L_{ot}
\end{aligned}
$$

We can derive an expression in changes by first dividing and multiplying through by lagged terms, and then rearranging:

$$
\begin{aligned}
L_{dt} &= \sum_o V_{dt}^{\epsilon} \frac{V_{dt-1}^{\epsilon}}{V_{dt-1}^{\epsilon}} \kappa_{odt}^{-\epsilon} \frac{\kappa_{odt-1}^{-\epsilon}}{\kappa_{odt-1}^{-\epsilon}} \Phi_{ot}^{-1} \frac{\Phi_{ot-1}^{-1}}{\Phi_{ot-1}^{-1}} L_{ot} \frac{L_{ot-1}}{L_{ot-1}} \\
&= \sum_o \widehat{V}_{dt}^{\epsilon} \widehat{\kappa}_{odt}^{-\epsilon} \widehat{\Phi}_{ot}^{-1} \widehat{L}_{ot} V_{dt-1}^{\epsilon} \kappa_{odt-1}^{-\epsilon} \Phi_{ot-1}^{-1} L_{ot-1} \\
&= \sum_o \widehat{V}_{dt}^{\epsilon} \widehat{\kappa}_{odt}^{-\epsilon} \widehat{\Phi}_{ot}^{-1} \widehat{L}_{ot} V_{dt-1}^{\epsilon} N_{odt-1} \\
\frac{L_{dt}}{L_{dt-1}} &= \sum_o \left(\frac{N_{odt-1}}{L_{dt-1}} \right) \widehat{V}_{dt}^{\epsilon} \widehat{\kappa}_{odt}^{-\epsilon} \widehat{\Phi}_{ot}^{-1} \widehat{L}_{ot} \\
\widehat{L}_{dt} &= \sum_o \pi_{odt-1}^{d} \widehat{V}_{dt}^{\epsilon} \widehat{\kappa}_{odt}^{-\epsilon} \widehat{\Phi}_{ot}^{-1} \widehat{L}_{ot}
\end{aligned}
$$

Where π_{odt-1}^{d} is the destination share of labor from location o, $\frac{N_{odt-1}}{L_{dt-1}}$. We can proceed in a similar way to derive $\widehat{\Phi}_{ot}^{-1}$:

$$
\begin{aligned}
\Phi_{ot} &= \sum_d V_{dt}^{\epsilon} \tau_{odt}^{-\epsilon} \\
&= \sum_d \widehat{V}_{dt}^{\epsilon} \widehat{\tau}_{odt}^{-\epsilon} V_{dt-1}^{\epsilon} \tau_{odt-1}^{-\epsilon} \\
\frac{\Phi_{ot}}{\Phi_{ot-1}} &= \sum_d \left(\frac{V_{dt-1}^{\epsilon} \tau_{odt-1}^{-\epsilon}}{\Phi_{ot}} \right) \widehat{V}_{dt}^{\epsilon} \widehat{\tau}_{odt}^{-\epsilon} \\
\widehat{\Phi}_{ot} &= \sum_d \pi_{odt-1} \widehat{V}_{dt}^{\epsilon} \widehat{\tau}_{odt}^{-\epsilon}
\end{aligned}
$$

Putting the two together yields:

$$
\begin{aligned}
\widehat{L}_{dt} &= \sum_o \pi^d_{odt-1} \widehat{V}^{\epsilon}_{dt} \widehat{\kappa}^{-\epsilon}_{odt} \left(\sum_d \pi_{odt-1} \widehat{V}^{\epsilon}_{dt} \widehat{\tau}^{-\epsilon}_{odt} \right)^{-1} \widehat{L}_{ot} \\
&= \underbrace{\widehat{V}^{\epsilon}_{dt}}_{\text{Direct effect}} \sum_o \underbrace{\pi^d_{odt-1} \widehat{\kappa}^{-\epsilon}_{odt} \left(\sum_d \pi_{odt-1} \widehat{V}^{\epsilon}_{dt} \widehat{\tau}^{-\epsilon}_{odt} \right)^{-1} \widehat{L}_{ot}}_{\text{Origin labor market access effect}}
\end{aligned}
$$

Since we focus on the post-Brasilia period, we assume that there are no changes in migration costs i.e., $\widehat{\kappa}_{odt} = 1$. We also abstract from exogenous population growth at the origin $\widehat{L}_{ot} = 1$. That leaves the final instrument that labor change in destination d is a composition of the direct increase in the benefits of being in location d, given by $\widehat{V}^{\epsilon}_{dt}$, and then a market access term that shows that when migrants have easy access to other destinations with a large direct gain, they will be less likely to migrate to d.

$$
\widehat{L}_{dt} = \widehat{V}^{\epsilon}_{dt} \sum_o \pi^d_{odt-1} \left(\sum_d \pi_{odt-1} \widehat{V}^{\epsilon}_{dt} \right)^{-1}
$$

D.1.2 Price Index

A similar procedure shows that the change in the price index for location d is given by:

$$
\begin{aligned}
P_{dt} &\propto \left(\sum_{o'} A_{o't} w_{o't}^{-\theta} \tau_{o'dt}^{-\theta}\right)^{\frac{-1}{\theta}} \\
P_{dt} &= \left(\sum_{o'} A_{o't-1} w_{o't-1}^{-\theta} \tau_{o'dt-1}^{-\theta} \hat{A}_{o't} \hat{w}_{o't}^{-\theta} \hat{\tau}_{o'dt}^{-\theta}\right)^{\frac{-1}{\theta}} \\
\widehat{P_{dt}} &= \frac{1}{P_{dt-1}} \left(\sum_{o'} A_{o't-1} w_{o't-1}^{-\theta} \tau_{o'dt-1}^{-\theta} \hat{A}_{o't} \hat{w}_{o't}^{-\theta} \hat{\tau}_{o'dt}^{-\theta}\right)^{\frac{-1}{\theta}} \\
\widehat{P_{dt}} &= \left(\sum_{o'} \frac{A_{o't-1} w_{o't-1}^{-\theta} \tau_{o'dt-1}^{-\theta}}{\sum_{o'} A_{o't} w_{o't}^{-\theta} \tau_{o'dt}^{-\theta}} \hat{A}_{o't} \hat{w}_{o't}^{-\theta} \hat{\tau}_{o'dt}^{-\theta}\right)^{\frac{-1}{\theta}} \\
\widehat{P_{dt}} &= \left(\sum_{o'} \pi_{odt-1} \hat{A}_{o't} \hat{w}_{o't}^{-\theta} \hat{\tau}_{o'dt}^{-\theta}\right)^{\frac{-1}{\theta}}
\end{aligned}
$$

D.2 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 Artuç et al. (2010),[39] at a given time t, the (log) utility flow that worker i living in o and moving to d enjoys is:

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

[39] Caliendo and Parro (2015) adopts this methodology to study the dynamic effects in the US from increased trade with China.

[40] where $\tilde{B}_{dt} = \log B_{dt}$, $\tilde{w}_{dt} = \log w_{dt} - \alpha \log P_{dt} - (1-\alpha) \log r_{dt}$ 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:

$$M_{odt} = \frac{\exp(\tilde{B}_{dt} + \tilde{w}_{dt} + \beta E_t V_{d,t+1} - \tilde{\kappa}_{odt})}{\sum_{s \in N} \exp\left(\tilde{B}_{st} + \tilde{w}_{st} + \beta E_t V_{s,t+1} - \tilde{\kappa}_{ost}\right)} \times L_{ot}. \tag{18}$$

The estimating equation from this model is exactly the same as our main gravity 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 $\hat{\delta}^{\kappa}_{dt}$:

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

[40] Artuç et al. (2010) assumes that a worker starting in location o at time t enjoys the real wage at the origin, $\tilde{w}_{ot}$ 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, $\tilde{w}_{dt}$.