IDB WORKING PAPER SERIES Nº IDB-WP-1173

Rural-Urban Migration at High Urbanization Levels

Matías Busso
Juan Pablo Chauvin
Nicolás Herrera L.

Inter-American Development Bank
Department of Research and Chief Economist

December 2020

Rural-Urban Migration at High Urbanization Levels

Matías Busso
Juan Pablo Chauvin
Nicolás Herrera L.

Inter-American Development Bank

December 2020

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Busso, Matías.
Rural-urban migration at high urbanization levels / Matías Busso, Juan Pablo Chauvin, Nicolás Herrera L.
p. cm. — (IDB Working Paper Series ; 1173)
Includes bibliographic references.
1. Rural-urban migration-Brazil-Mathematical models. 2. Unemployment-Brazil-Mathematical models. 3. Migrant labor-Brazil-Mathematical models. 4. Urbanization-Brazil-Mathematical models. I. Chauvin, Juan Pablo. II. Herrera L., Nicolás. III. Inter-American Development Bank. Department of Research and Chief Economist. IV. Title. V. Series.
IDB-WP-1173

http://www.iadb.org

Abstract[*]

This study assesses the empirical relevance of the Harris-Todaro model at high levels of urbanization – a feature that characterizes an increasing number of developing countries, which were largely rural when the model was created 50 years ago. Using data from Brazil, the paper compares observed and model-based predictions of the equilibrium urban employment rate of 449 cities and the rural regions that are the historic sources of their migrant populations. Little support is found in the data for the most basic version of the model. However, extensions that incorporate labor informality and housing markets have much better empirical traction. Harris-Todaro equilibrium relationships are relatively stronger among workers with primary but no high school education, and those relationships are more frequently found under certain conditions: when cities are relatively larger; and when associated rural areas are closer to the magnet city and populated to a greater degree by young adults, who are most likely to migrate.

Keywords: Harris-Todaro, Rural-urban migration, Urban unemployment, Developing countries

JEL Codes: J46, J61, O18, R23

[*]Corresponding authors: Busso (mbusso@iadb.org) and Chauvin (juancha@iadb.org).

We thank Adam Storeygard and two anonymous referees for very useful comments and suggestions. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.

1. Introduction

Since its original publication, the Harris-Todaro (1970) framework (henceforth HT) has been one of the key conceptual tools for studying rural-urban migration and its links to urban unemployment. Fifty years ago, the majority of the world's population lived in rural areas, with a global urbanization rate of 36.6%. Over the decades that followed, urbanization proceeded at a rapid pace. In 2007 global urbanization crossed the 50% threshold. Today 55.3% of the world population lives in urban areas (World Bank, 2020), and the proportion is expected to increase to 68% by 2050 (United Nations, 2019). In future decades, rural-urban migration will take place in increasingly urbanized low- and middle-income countries.[1] In such an environment, the standard HT model may lose empirical validity. This could happen because the reality that the model seeks to explain changes (for example, the relative scarcity of rural labor could render the rural-urban wage gap negligible), or simply because the framework is too parsimonious to meaningfully characterize increasingly complex and interconnected economies. In this paper we explore the extent to which the HT framework is empirically relevant as an analytical tool at high levels of urbanization, and we evaluate extensions of the model to determine which of these may be the most appropriate to use in an increasingly urbanized developing world.

Our study focuses on Brazilian cities, which offer an ideal context in which to pursue these research questions. In 1970, the year in which the seminal paper by Harris and Todaro was published, Brazil had an urbanization rate of 55.9%, similar to that of the world as a whole today (World Bank, 2020). Over the following 50 years, the country continued to urbanize at a rapid pace, reaching a rate above 85% by 2018. In this sense, the Brazilian experience may anticipate the experiences that other developing countries are likely to face in coming decades. Moreover, Brazil's many cities differ in terms of their size, the education levels of local residents, and industry structures of local economies – features that allow us to examine at a granular level the conditions under which the model can have more or less empirical bearing.

Our data come primarily from three rounds of the Brazilian population census (1991, 2000, and 2010). We approach our study by defining a rural-urban migration "catchment area," the set of rural municipalities of origin of migrants that have previously arrived in

[1]High urbanization levels used to be treated as a measure of development. In recent decades, however, a trend has emerged in which countries achieve high levels of urbanization even though they continue to have relatively low per capita income levels (Glaeser 2014; Jedwab and Vollrath 2015).

each of the 449 cities in our sample. We then construct two sets of city-level variables: urban labor market variables, calculated with observations from within the urban boundary; and corresponding rural variables, calculated as weighted averages from rural municipalities in the city's catchment area, with the municipalities' historical migration shares as the weights.

We use these to examine the HT equilibrium condition, the point at which the rural wage equals the expected urban wage. We assess the extent to which the equilibrium condition holds, starting with the original framework, and subsequently considering several extensions. For each version of the model, we characterize the corresponding HT equilibrium condition in terms of the urban employment rate and the rural-urban wage ratio, and we use it to create an error measure to quantitatively assess the empirical performance of the model.

For most of the period under study, the minimum wage is below the average rural wage – which violates one of the key assumptions of the model. This gap prevents us from using the legislated minimum wage as the urban wage measure, as in the original model's formulation. Instead we take another approach by considering urban wages observed in the market. A challenge is that urban labor markets nest both formal and informal jobs, with the informal sector playing a more prominent role in developing than in high-income countries. We show that using a single urban sector – either the formal or the informal – renders too many Brazilian localities to appear to be off-equilibrium. In both cases the observed urban employment rate is smaller than what the model would anticipate – and more severely so when we use the informal sector to compute the reference urban wage for rural migrants. This is because, in the model, high employment rates in urban areas result from low rural-urban migration, which in turn results from a small gap between the expected urban wage and the rural wage. Given that the expected urban wage in both the formal and the informal sector is larger than in rural areas, ignoring one of the two sectors leads to lower migration predictions than if both sectors are considered.

Next, we explore two extensions to the original HT framework. First, we expand the model to accommodate two urban sectors, formal and informal, and assume that they are fully segmented. This extension improves the empirical performance of the model. Second, we extend the model to accommodate differences in costs of living across urban and rural areas. This extension further improves the model's correspondence with the observed data. These results are robust to the way we construct our sample and the way we measure several key variables and parameters involved in the model's equilibrium condition.

The HT model assumes that workers are homogeneous; however, in practice, both rural and urban workers are heterogeneous. Thus, the framework's predictive ability may be

weaker if workers systematically sort into migration based on personal characteristics that are associated with their wage and employment probability. Even though we do not formally model worker heterogeneity, we explore the empirical traction of the framework for different sub-populations. We find that the model has similar predictive power in younger (aged 15 to 39) and older (aged 40 and older) cohorts. Across different schooling subgroups, the HT equilibrium better fits the behavior of those who have a primary school-level of education than of those with either lower (less than primary) or higher (completed high school or have a higher degree) levels of education. The model has less empirical traction when we consider workers from each gender separately; the model underestimates the observed urban employment rates of males, and overestimates female employment rates. In both the schooling and the gender categories, subgroups for which the model performs poorly offset each other, improving the fit of the model in the full sample.

Finally, we analyze how the HT prediction error varies with characteristics of the city and of their associated rural areas of influence. We find that the errors are smaller in the cases in which the distances between the city and its rural catchment area are shorter. This is in line with one of the HT model assumptions: that moving costs are low. We also find that the errors are smaller for cities better suited to absorb migration flows – that is, those that are larger and have experienced growth in their service sector in the past; and for those cities whose catchment areas are more archetypically rural in nature, with less densely populated geographies. Prediction errors are also smaller when rural areas has more young people, the segment of the population that is most likely to migrate to cities (Kennan and Walker, 2011); and they tend to get larger over time as urbanization progresses.

One of the model's key insights – that higher urban wages lead to rural-urban migration responses – does appear to have empirical traction in highly urbanized Brazil. That said, much of the policy discussion associated with the HT framework has been concerned specifically with the case in which the creation of one additional job in urban areas leads to the migration of more than one rural worker, such that efforts to tackle urban unemployment could *worsen* it. This is the well-known "Todaro paradox" (Todaro, 1969).[2] Our analysis suggests that the lack of empirical evidence in the literature in support of this prediction (Lall et al., 2006) is a reflection of the model's parsimony in its characterization of the urban labor market, which overlooks both the critical role the urban informal sector, and the labor

[2] Fields (2005) suggests that this emphasis on unemployment may not be warranted: even if formal sector development leads to excess supply in the urban labor market, the net welfare effect is not necessarily negative because the social benefit of the increase in high-paying jobs and the reduction of poverty may outweigh the social costs of unemployment.

force participation decisions of households. We conjecture that expanding the model in these directions could improve our understanding of the relationship between rural-urban migration and urban informality, and help inform whether and when policies promoting formal employment could lead to unanticipated effects on urban informality rates.

Our paper contributes to the empirical literature on rural-urban migration by providing a comprehensive test of the validity of the HT framework in a highly urbanized developing country. The literature has long recognized that rural-urban migration is a fundamental part of the economic development process. Several papers have recently studied the causes and consequences of internal migration in Latin America (e.g., Jiang and O'Neill, 2018, Rodríguez-Vignoli and Rowe, 2018, Bernard et al., 2017), sub-Saharan Africa (e.g., de Brauw et al., 2014), China (e.g., Laing et al., 2005, Combes et al., 2015), India (e.g., Munshi and Rosenzweig, 2016, Hnatkovska and Lahiri, 2015), and Indonesia (e.g., Bryan and Morten, 2019). As urbanization advances in the developing world, where the urban population is expected to be in 2030 twice as large as in 2000 (World Bank, 2013), much of the world's rural-urban migration in the foreseeable future is likely to take place in contexts similar to that of our study. Prior empirical tests of the HT model are surprisingly few given how influential the model has been on theory and policy. Most studies have taken place at either lower levels of urbanization (e.g., Barnum and Sabot, 1977; Collier, 1979; Fields, 1982; Schultz, 1982; Lucas, 1985) or in more developed economies (e.g., García-Ferrer, 1980; Salvatore, 1981; Petrov, 2007).

We follow an approach that allows us to test model-based predictions at the level of individual subnational locations, and one that is replicable with data that are publicly available in many developing countries. Most existing studies either rely on country-level statistics (e.g., Todaro, 1976; Collier, 1979; Salvatore, 1981), or on individual-level analysis, in which a migrant indicator is regressed on characteristics of places of origin and destination (e.g., Fields, 1982; Schultz, 1982). Our approach better matches the geographic level at which, arguably, most of the related policy decisions are made and/or implemented. Methodologically, rather than estimating reduced-form or structural elasticities to contrast them with the signs implied by the HT model, our approach is in essence an accounting exercise – closer in spirit to "growth accounting" in macroeconomics – which we see as complementary to prior approaches.

In addition, we contribute to the understanding of the specific circumstances in which the HT framework may be most relevant for informing policy. Prior studies have analyzed how rural-urban migration decisions vary with personal and place characteristics (Fields, 1982;

Schultz, 1982; Lucas, 1985). Our approach provides a broader sense of where (across place characteristics), when (across time, as the country continues to urbanize), and for whom (across subpopulations) the model is useful in seeking to explain rural-urban migration and urban unemployment. At the same time, by incorporating both the effects of the informal sector and urban costs of living, our approach helps gauge the extent to which refinements can improve the model's empirical traction at a given time and place. This opens the door for a more informed and context-sensitive use of the framework in policy design.

The rest of the paper proceeds as follows. Section 2 describes the data and our definition of cities and their rural catchment areas. It also presents a set of descriptive facts to contextualize the analysis that follows. Section 3 presents the equilibrium conditions of the basic HT model and of a few extensions, and assesses to what extent each of these conditions hold in the data. Section 4 addresses the role of human capital heterogeneity in the empirical traction of the model. Section 5 discusses how the predictive power of the HT framework correlates with characteristics of the cities and their rural "catchment areas." Section 6 concludes.

2. Descriptive Facts

All variables used in the analysis are constructed from microdata of the Brazilian population censuses, which are publicly available from the Institute of Geography and Statistics (IBGE). The Data Appendix provides details of the construction of each variable. Our analysis requires information on the (rural) municipality of origin of migrants that arrived in each city over the previous 10 years; because these data are available only beginning in 1991, we assess the suitability of the HT model for the period from 1991 to 2010. In addition, we rely on the population censuses of 1970 and 1980 to measure other auxiliary variables.

2.1. Defining Cities and Their Rural Catchment Area

To proceed with our analysis, we need to define a unit of observation that both takes advantage of the richness of the Brazilian context and the data available, and captures the level at which the economic forces contemplated in the model are likely to operate. In the original HT framework there are only two geographies, an urban and a rural area. While it is relatively straightforward to bring this to the data at the national level, where all the rural-urban migrants effectively move from one geography to the other, this is not the case

at the city level. The rural-urban migrants that arrive in a given city come at different intensities from different subsets of the country's rural areas. Conversely, potential migrants living in rural locations frequently have more than one possible urban destination, and therefore the measures of (expected) urban wage and employment probabilities relevant for their migration decisions can be more complex.

Our analysis focuses on cities and data-driven definitions of their rural catchment areas.[3] To identify cities, we rely on the boundaries of urban commuting zones (*"arranjos populacionais"*), which are defined by the Brazilian Institute of Geography and Statistics, IBGE (2016), as sets of adjacent municipalities linked by high levels of commuting for work or study in the 2010 census. We also use this to distinguish, in the microdata, between urban observations – individuals living within the boundaries of a commuting zone – and rural observations – individuals living in municipalities that are not part of any of these zones.[4]

We define the rural-urban migration catchment area of a city as the set of rural municipalities from which migrants originated in the past.[5] For the purpose of our analysis, in the case of the rural wage and the rural housing rent we attribute to each city a single measure, calculated as the weighted average of the corresponding variable in the rural municipalities in the catchment area. We use as weights the share of each municipality in the city's historical rural-urban migration.[6] To improve the precision of the measures, corresponding calculations use only observations classified as "urban" or "rural" in both our own definition and that of the census. The Data Appendix provides further details on our computations.

A challenge with this approach is that the number of Brazilian municipalities grew significantly over the period of analysis (from 3,950 in 1970 to 5,565 in 2010). In several cases

[3] A few existing papers have looked at geographies below the national level, but have not explicitly linked cities with the rural areas from which their migrants originate. Barnum and Sabot (1977) group Tanzanian regions into six areas (three urban and three rural), and estimate migration equations using as their unit of observation cells based on area, education category, and time period. Fields (1982) uses departments' borders to divide Colombian territory into six urban and six rural areas, and studies the determinants of migration from and to any of these locations. García-Ferrer (1980) uses 50 Spanish provinces as its unit of observation, without distinguishing between rural and urban areas.

[4] Information on the type of area (rural or urban) where the migrant lived immediately prior to migrating is not consistently available across census rounds. Our approach instead measures migrants' location of origin consistently throughout the period of analysis.

[5] An alternative would be to define, for each rural area, a set of potential urban destinations for migrants. However, a typical rural municipality is too small to significantly affect by itself quantities and prices in its destinations' urban labor markets, making this approach inappropriate to assess the existence of HT-style equilibrium relationships in the data.

[6] We calculate these migration weights using data from migrants that arrived in the city between six and 10 years before, excluding "recent migrants" (those that arrived five years ago or more recently), who are more likely to be responding to current incentives (such as wage gaps and unemployment) in the census year.

the parents of new municipalities belonged to different commuting zones. This requires us to adjust our geographic units of observation (commuting zones and rural municipalities) to make them consistent across census years. We follow Chauvin (2018) and Kovak (2013) to build a set of time-consistent cities, defined by aggregating the original commuting zones that share the same family tree into a larger commuting zone. This yields 449 cities that are considered urban areas throughout the period of analysis. Similarly, the set of rural municipalities is also constant over time.[7] Appendix A shows the geographic distribution of these locations and discusses the definitions we use in more detail, and Appendix Table B.2 reports descriptive statistics of these cities and their catchment areas.[8]

2.2. Rural-Urban Migration in Brazil

Even though internal mobility in Brazil has historically not been as high as in the United States, it has until recently been larger than in many other developing countries, including China and India (Chauvin et al., 2017). Figure 1 shows the percentage of the working-age population that had migrated in the last five years (blue line), and the percentage of the working-age population living in urban areas that had migrated in the last five years (red line) over the period 1970 to 2010. In 1980, around 18% of the working-age population declared that they had changed municipality of residence at some point in the previous five years. This figure, which had increased from an initial 15% in 1970, dropped steadily over the following three decades until reaching 11% in 2010. Cities were the destination for the lion's share of these population movements. Migrants living in urban areas at the time of the 1970 census accounted for 70% of total migration. This share rose to 78% in 1980 and remained stable at that level in all subsequent census years.

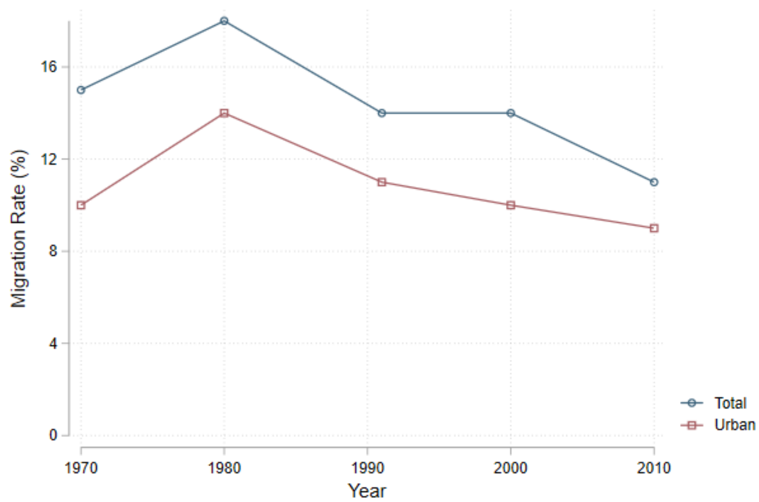
About two-thirds of all rural migrants move to urban destinations, with the vast majority of them going to cities in the top quartile of the population distribution (Appendix Table B.1). The majority of urban people who migrate tend to move to other cities, usually one that is larger than the city of origin. A non-negligible fraction of people living in urban areas move to rural locations. This is consistent with the prevalence of return migration, which

[7] In other words, we do not allow for a reclassification of rural areas into urban areas.

[8] The median commuting zone (i.e., a city in our analysis) is composed of 2 municipalities, and it has an area of 1,319 square kilometers and a population of 95,646. The median urban municipality that is part of our commuting zones has a population density of 81 persons per square kilometer. The median rural catchment area is composed of 44 municipalities, has an area of 1,294 square kilometers and 27,344 people. The median (rural) municipality that is part of our catchment areas has a population density of 24 persons per square kilometer.

could be due to an unsuccessful performance in the destination labor market (Hirvonen and Lilleør, 2015), negative shocks experienced in the city, or improved conditions in the place of origin (Nguyen et al., 2017). Some of these returned migrants re-enter the labor market in their original rural communities as wage workers or as entrepreneurs (Dustmann and Kirchkamp, 2002).

Figure 1: Internal Migration of Working-Age Individuals in Brazil, 1970-2010



Notes: Authors' calculations using census microdata. The blue line shows the percentage of the country's total working-age population who migrated internally in the five years previous to the census year. The red line shows the share of the working-age population who migrate internally in the five years before the census year towards cities. Section 2.1 describes in detail our definition of urban zones.

In spite of the country's high levels of urbanization, an important share of urban immigrants continues to come from rural areas. Table 1 (Panel A) documents that, over the five years preceding each census, 2%-3% of the working-age population moved from rural municipalities to cities. In 2010, this amounted to almost 2.8 million people, representing 9.7% of the rural population in the prior census, and almost 30% of the total flow of internal migrants into cities. The gender split of rural-urban migrants in Brazil is even, in contrast to the pattern of male-dominated migration in other developing-world regions.[9]

Rural-urban migrants tend to be young and relatively less educated, but both age and education levels of migrants have increased over time as primary school became nearly universal and high school enrollment grew at a fast pace throughout the country (Busso et al., 2017). Four out of five rural-urban migrants were younger than 40 at the time of their move.

[9]Even levels of male and female internal migration as those found in Brazil have also been documented in various other Latin American countries as noted by Mazumdar (1987) and Lall et al. (2006), who attribute the pattern to the prevalence of migrant female domestic workers.

Regarding educational achievement, 44% had less than primary schooling in the 1991 census. This share declined to 24% in 2000, and 21% in 2010. By contrast, the shares of migrants with high school education or higher increased over the same time from 13% (1991), to 23% (2000) and to 38% (2010).

The statistics reported in Panel B of Table 1 describe how rural-urban migrants perform in the labor market after they have arrived in the city. Rural migrants tend to have similar or higher rates of employment than natives, and they earn wages that are 2%-5% lower than those earned by non-migrant workers in the same city. [10] Relative to their rural municipality of origin, however, migrants receive a wage premium of around 8%.[11]

As is the case in many developing countries, labor informality is high in Brazil. Ulyssea (2018) reports that in 2003 two-thirds of firms and 35 percent of workers were in the informal sector. These informal workers tend to have lower levels of education and earn lower wages than formal employees (Ulyssea, 2010). We consider workers to be informally employed if they lack a signed booklet (*carteira de trabalho*) that registers the entire worker's history of employment in the formal sector - a standard definition used in the literature (e.g., Dix-Carneiro and Kovak 2019). Census data are collected at the household level, without identifying individuals or their households; thus, workers face no risk of penalties for reporting an informal working status to the census. Rural-urban migrants are more likely than non-migrants to be employed in informal jobs, although the informality gap between migrants and natives has declined over time.[12] Panel C describes the housing conditions of rural-urban migrants. Migrants are more likely than urban local residents to rent (rather than own) houses. The houses they occupy are also, on average, of lower quality than those occupied by non-migrants.[13] Migrants also tend to pay rents that are 15-20 percentage points

[10]Part of these differences observed in aggregate employment rates, formality rates, and wages could, in principle, reflect differences in the characteristics of people that choose to migrate and the local (non-migrant) population. Self-selection into migration has been long and extensively recognized in the literature of cross-border migration (e.g., Borjas (1987), Chiquiar and Hanson (2005)) as well as rural-urban internal migration (see Lagakos (2020) for a recent review). We assume that self-selection explains a similar proportion of the wage gaps across all commuting zones and their respective rural catchment areas.

[11]This wage gap is similar, for instance, to the one reported by Michaelsen and Haisken-DeNew (2015) for the case of Mexico.

[12]The observed wage gap between migrants and non-migrants from their rural municipalities of origin is virtually the same for formal and informal workers in all census years.

[13]To succinctly characterize the variation in housing quality, we estimated a housing quality index using a simple factor model of housing characteristics (wall material, people per room, number of rooms, number of bedrooms, and water connection) predicting an index and then standardizing it so that the prediction is mean zero and standard deviation one for urban areas in 1991. This is the measure used in Table 1, Panel C.

lower than those paid by non-migrant urban residents, and, as expected, they face rental prices that are more than double those of residents in their rural communities of origin.

Table 1: Characteristics and Labor Market Performance of Rural-Urban Migrants

	1991	2000	2010
Panel A: Characteristics of rural-urban migrants			
Working-age rural-urban migrants (in 1000s)	2,293	2,194	2,822
Percent of the national working-age population	2.9%	2.2%	2.4%
Percent of rural population (prior census)	10.6%	8.7%	9.7%
Share in total migration to cities	33.1%	27.6%	28.8%
Percent of females	51.4%	52.4%	50.5%
Age at the time of migrating			
Percent 15 to 39	84.4%	82.8%	81.5%
Percent 40 or older	15.6%	17.1%	18.5%
*Education**			
Percent less than primary	43.7%	23.2%	21.0%
Percent primary but less than high school	43.5%	53.8%	40.9%
Percent high school or higher	12.8%	22.9%	38.1%
Panel B: Labor market performance of rural-urban migrants			
Employment rate	63.2%	57.2%	64.3%
Difference from the urban average (ppts.)	3.6	0.9	0.7
Informality rate	43.6%	48.0%	36.8%
Difference from the urban average (ppts.)	9.18	4.38	-0.56
Wage gap (ratio)			
Relative to non-migrant urban residents	98.4%	95.5%	97.9%
Relative to rural municipality or origin	108.6%	107.6%	107.7%
Panel C: Housing conditions of rural-urban migrants			
Percentage of households that rent	37.8%	N/A	59.5%
Difference from the urban average (ppts.)	16.5	N/A	36.7
Average Housing quality index	-0.42	N/A	-0.35
Difference from the urban average (s.d)	-0.39	N/A	-0.33
Rent gap (ratio)			
Relative to non-migrant urban residents	78.3%	N/A	86.4%
Relative to rural municipality or origin	247.0%	N/A	181.7%

Notes: Authors' calculations of national-level estimates using census microdata. Rural-urban migrants include all individuals moving from rural municipalities to cities in the five years before the census year. Section 2.1 describes in detail our definition of urban and rural areas. We restrict the sample to working-age migrants (at the time of migrating) and working-age stayers. We estimate the household quality index using principal component analysis. The variables used are the quality of wall material, the type of dwelling, the number of people per room, the number of rooms, the number of bedrooms, and access to water connection. We standardized the index using the mean and the standard deviation of the urban non-migrants' index distribution. Housing rents not available in 2000. ***** To capture pre-migration education attainment, these measures are calculated restricting the sample to individuals aged 18 or older at the time of migrating (i.e., the age in which individuals are expected to have finished high school education in Brazil).

3. The HT Equilibrium in the Data

With this background, we explore the extent to which the the equilibrium relationships predicted by the Harris-Todaro model are observed in the data. We start with a version of the model that is very close to the one originally formulated by Harris and Todaro (1970). We then analyze two variants that relax some of the original assumptions while keeping the equilibrium concept of the original model, namely that the expected payoff of moving to the city equals the payoff of staying in the rural areas.

3.1. The Basic Model

The economy has two sectors, urban and rural. They have isomorphic production functions assumed to be $Y_s = \psi_s L_s^{\gamma}$, where ψ_s is a labor productivity shifter, subindex $s = \{u, r\}$ denotes urban or rural sector, and $0 \leq \gamma \leq 1$. Workers' marginal productivity is $\psi_s \gamma L_s^{\gamma-1}$, and labor demand is given by:

$$L_s = \left(\frac{w_s}{\psi_s \gamma}\right)^{\frac{1}{\gamma-1}}. \tag{1}$$

In line with the HT framework we assume that, in the urban sector, the observed wage is above the competitive equilibrium wage. In our formulation this is due to an exogenous friction term τ, such that $w_u = w_u^{comp} + \tau$, where w_u^{comp} is the wage under perfect competition. The wedge between the observed and the competitive wages prevents the urban market from clearing, and leads to excess urban labor supply (i.e., unemployment). In the original HT formulation, the wedge emerged from an institutionally set urban minimum wage ($\underline{w}$), which in our expression corresponds to $\tau = \underline{\mathrm{w}} - w_u^{comp}$. Subsequent extensions have proposed alternative wage-setting mechanisms for the urban sector, such as market wages with labor turnover costs (Stiglitz, 1974; Sato, 2004), costly employee supervision (Calvo and Wellisz, 1978; Zenou, 2011), or wages shaped by employer-union bargaining (Calvo, 1978). Our generic formulation allows us to develop expansions of the model that both remain parsimonious and allow for multiple alternative explanations of what drives the urban wage above equilibrium.

On the labor supply side, homogeneous workers derive utility exclusively from the consumption of a tradable good C. Assuming that the good is priced at one, the consumer maximizes her (expected) wage income, which corresponds to w_r if the worker locates in a rural area, and $m_1 w_u$ if the worker locates in an urban area, where the population consists

of urban-employed (L_u) and non-employed (L_n) individuals, and $m_1 = \frac{L_u}{L_u+L_n}$ is the urban employment rate. Workers will locate and inelastically supply one unit of labor in the area where their expected labor income is higher. Per the original HT assumptions, workers living in rural areas cannot search for jobs in the city and vice versa. For simplicity, we assume away migration costs.

In equilibrium, the expected urban wage is equalized to the rural wage, $m_1 w_u = w_r$, which can be expressed in terms of the relationship between the urban employment rate and the rural-urban wage gap:

$$m_1 = w_1$$

where $w_1 = \frac{w_r}{w_u}$ is the ratio between the nominal rural and urban wages. This suggests that we can use the ratio between the left-hand side and the right-hand side of equation 2 as a measure of the empirical "prediction error" of the HT equilibrium in a given city:

$$\varepsilon_1^{HT} = \frac{m_1}{w_1}. \tag{2}$$

This captures how distant the data are from the theoretical HT equilibrium. The numerator of equation 2 is the employment rate observed in the data. It is useful to think of the denominator – the rural-urban wage ratio – as representing the model's "predicted" urban employment rate. This is the case because, if the model fully reflected reality and the local economy was at equilibrium, we would be able to predict the employment rate of an urban area based on this wage ratio. This allows us to interpret the error term ε_1^{HT} in terms of "excess urban employment rate" relative to the equilibrium condition. In the model, migratory responses arbitrage away rural-urban expected wage differences, making $\varepsilon_1^{HT} = 1$. Values larger than one imply that an urban area has a higher employment rate than the model would predict, and values smaller than one imply a lower than expected urban employment rate.

The error ε_1^{HT} can reflect either misspecification in the model,[14] or the fact that the city is temporarily off equilibrium due to contemporary shocks or slow adjustments to prior shocks. Our approach does not allow us to empirically differentiate between these two explanations.

[14] The literature has pointed out several potentially key determinants of rural-urban migration and urban unemployment that the standard model does not consider. Rural-urban migrants may be responding to incentives other than a potentially higher expected income, including public services unavailable in rural areas (Brueckner and Lall, 2015; Lall et al., 2009), the strength of migrants' social networks in their potential destinations (Giulietti et al., 2018), the risk of losing informal insurance networks in their places of origin (Munshi and Rosenzweig, 2016), and conflict-related displacement (Henderson et al., 2017; Calderón-Mejía and Ibáñez, 2016).

However, we make progress in understanding the role of misspecification by empirically assessing extensions of the model that account for potentially relevant variables that are not part of the basic framework.

Rural-Urban Wage Gaps, Informality, and Unemployment in Brazilian Cities

To obtain city-level measures of equation 2, we use census microdata to calculate constituent variables for each city and its correspondent rural migration catchment area (see Appendix A and the data appendix for details). Table 2 reports the averages (taken across cities) of the main variables used to measure the HT error in the basic model and extensions. In addition to the rural wage, the table contains four alternative measures of the urban wage: the official minimum wage at the time of the census, wages of formal and informal workers measured separately, and a weighted average of the formal and informal wages.

On the basis of these summary statistics alone, it is clear that the official minimum wage is unlikely to be the driver of rural-urban migration in Brazil. In both 1991 and 2000, the average rural wage was in fact larger than the minimum wage. This gap closed after the minimum wage increased by 73% in real terms over the 2000s, but even by 2010, it was only 12% larger than the rural wage. Using the minimum wage as the urban wage in our empirical analysis would violate a key assumption of the original formulation of the HT model, namely, that the institutionally set urban wage is higher than the competitive urban wage, which in turn is higher than the rural wage.

By contrast, the rural wage has been consistently smaller than the average market urban wage. Moreover, the rural-urban wage ratio has remained, in real terms, strikingly constant over time between 56% and 58%. This gap is larger relative to the formal urban wage: in the average city, workers living in its rural catchment area earn just around half the wage of formal urban workers, but only around 30% less than informal urban workers. Rural areas also have significantly lower costs of living. The average rural rent represented 35% of the average urban rent in 1991, and 48% in 2010, the two census years for which this information is available. This large gap partly reflects the substantive differences in development levels, productivity, and costs of living between the richer, more urbanized regions of Brazil (the South and Southwest) and the poorer, less urbanized regions of the country (North and Northeast). To account for these spatial differences in the empirical analysis that follows, we adjust our price measures (wages and rents) by macroregion and census-year fixed effects (see the Data Appendix for further details).

Table 2: Variables Used in the HT Prediction Errors

	1991	2000	2010
Average rural wage (w_r)	275.3	357.8	453.5
Minimum wage ($\underline{w}$)	186.0	294.8	510.0
Average urban wage ($\bar{w}_u = {}^{(w_f L_f + w_i L_i)}/_{(L_f + L_i + L_n)}$)	495.6	640.3	782.0
Formal workers (w_f)	572.0	780.3	916.5
Informal workers (w_i)	401.5	505.7	624.7
Rural / minimum wage ratio (${}^{w_r}/_{\underline{w}}$)	148.0%	121.4%	88.9%
Rural / urban wage ratio (${}^{w_r}/_{\bar{w}_u}$)	55.5%	55.9%	58.0%
Rural / formal urban wage ratio (${}^{w_r}/_{w_f}$)	48.1%	45.9%	49.5%
Rural / informal urban wage ratio (${}^{w_r}/_{w_i}$)	68.6%	70.8%	72.6%
Average urban housing rent (r_u)	177.1	N/A	263.5
Average rural housing rent (r_r)	61.9	N/A	125.1
Urban employment rate (m)	59.6%	56.3%	62.7%

Notes: Authors' calculations using census microdata. The table reports the average of labor and housing market aggregates across cities. Rural variables refer to the average across city's rural catchment areas. The Data Appendix explains in detail the computation of urban and rural aggregates as well as the definitions of each variable. All monetary values are expressed in 2010 Reais.

The original HT model assumed away non-participation, such that urban dwellers can be either employed or unemployed. Incorporating labor force participation decisions into the HT framework is beyond the scope of this paper. At the same time, in taking the model to the data we cannot simply exclude non-participants from the analysis, given that they quantitatively matter for the measurement of key variables such as the total urban population and the expected urban wage (through the probability of being employed).[15]

[15] The fact that the HT framework assumes away the non-participant population poses two important challenges for empirical work. First, the employment rate – used to calculate the expected urban wage – corresponds in this case to ${}^{Employed}/_{Employed+Unemployed}$. In reality, the share of the urban population that chooses not to participate in the labor force is counted neither as employed or unemployed, even though they reside in the city. This implies that the actual jobs/population ratio is smaller than suggested by the original framework, in turn implying that, from the perspective of potential rural migrants, the likelihood of getting a job and the expected urban wage are also smaller. Second, labor force participation – and therefore measured unemployment – can change in response to shifts in urban labor demand *even in the absence rural-urban migration.* Moreover, the extent to which labor demand shifts lead to rural-urban migration or higher labor force participation of urban local residents may vary significantly across regions of the country. In Brazil, changes in labor force participation across census years have had a higher variance in cities than in rural areas (see Appendix Table B.2). Our choice of defining employment rates considering the entire urban population (${}^{Employed}/_{Employed+Unemployed+Nonparticipant}$) addresses both of these issues, and allows us

Our approach is to use the non-employed urban population (unemployed plus non-participant) as the empirical counterpart of the HT unemployment variable. A limitation of this solution is that we may overestimate unemployment, in the sense that the risk of *involuntary non-employment* may not be as large in the eyes of potential migrants as our measure suggests. If labor supply is upward sloping among urban dwellers, such mismeasurement would be negatively correlated with the urban-rural wage gap.[16] We take note of this issue in the analysis of our empirical results, and consider how it may affect their interpretation.[17]

The average urban employment rate is reported at the bottom of Table 2. Non-employed workers represent a large share of the urban population in Brazil. In 1991, there were as many as 67 non-employed for each 100 employed workers in the average Brazilian city. Employment rates dropped, on average, between 1991 and 2000, before rising again in 2010.

Bringing the Model's Predictions to the Data

We turn now to the question of the empirical traction of the HT equilibrium, quantifying the extent to which the equilibrium condition from the most basic version of the HT model holds in the data.

During most of the period of study, the minimum wage lies below the rural wage, making it a poor measure of the urban wage in the context of explaining rural-urban migration. Thus, we use alternative measures of the urban wage observed in local markets. An important challenge with this approach is that formal and informal labor markets coexist in Brazilian cities, as is the case in most developing countries. The minimum wage and other regulations that effectively increase income are enforced in the formal sector but not in the informal sector. If aspiring to a well-paid formal job is the main driver of rural migration to cities, the formal wage should matter the most for measuring the HT equilibrium. By contrast, if finding informal jobs is relatively easier, and if the urban informal wage still represents

to abstract from explicitly modeling the labor force participation decision of urban dwellers.

[16] Holding the rural wage constant, a higher urban wage would lead to higher participation, and thus the strictly unemployed (i.e., people who are willing to work but are unable to find a job) would be a larger share of the non-employed.

[17] Alternatively, we could use the standard definition of unemployment (without non-participants) and still include all the non-employed in the urban population totals. However, because the urban unemployment is close to zero in multiple cities, the model-based urban employment rates measures are unrealistically high. Other studies (e.g., Fields, 1982; García-Ferrer, 1980) have also found unemployment to be a poor measure of the attractiveness of local labor markets for potential migrants.

an improvement relative to the rural wage, then the informal market may be the relevant urban outside option for potential rural migrants, and, thus, the most appropriate metric to capture HT equilibrium relationships in the data.

We explore the relative merits of these two views in the first two panels of Table 3. Here we report calculations of the errors in equation 2 using, alternatively, the formal labor market ($w_u = w_f$ and $m_1 = \frac{L_f}{L_u + L_n}$) in Panel A, and the informal labor market ($w_u = w_i$ and $m_1 = \frac{L_i}{L_u + L_n}$) in panel B.[18] The table reports, for each version of the error and each census year, the average across cities of the errors and two statistics characterizing the error's distribution: the percentage of the cities whose error falls within half a standard deviation around one (the "perfect fit" benchmark), and the percentage that falls within one standard deviation around one, respectively. To facilitate comparisons we use the same standard deviation for all cases (corresponding to the estimates of ε_2^{HT} in 1991).

When we use either the formal or the informal sector to compute the reference expected urban wage, the observed urban employment rate tends to be smaller than the one predicted by the HT equilibrium. This is because, when we assume away the opportunities for income-generating employment in one of the sectors, the expected urban wage appears closer to the rural wage than if we consider both sectors. Accordingly, the model predicts that a relatively smaller migratory response is needed to restore equilibrium, resulting in a relatively high predicted urban employment rate. In practice, however, rural-urban migrants that fail to obtain a job in one of the urban sectors do not necessarily face unemployment; and the prospect of working in the other sector – with wages frequently larger than in rural areas – raises the incentives to migrate to the city.

While both single-sector models fit the data poorly, deviations from the predicted equilibrium are less pronounced in the model featuring the formal sector than in the model featuring the informal sector. In the formal-sector model (Table 3, Panel A) the average errors are relatively close to one (ranging from 0.68 to 0.87), and a larger fraction of the errors lie in the vicinity of one (between 20% and 26% lie within half a standard deviation, and between 40% and 51% lie within one standard deviation). By contrast, in the informal-sector model (Table 3, Panel B) the average errors are smaller (between 0.41 and 0.43), and a very small share of the errors fall close to one.

In these versions of the model the size and direction of the errors may be particularly

[18] Fields (1975) was one of the first to introduce the informal sector –which he dubbed the "murky sector"– to the HT framework; he used a similar assumption, constraining rural-urban migrants to finding only informal jobs in the city.

Table 3: Empirical Deviations from the Harris-Todaro equilibrium

	1991	2000	2010
Panel A: Basic model using the urban formal sector			
Average prediction error ($\varepsilon^{HT}_{1.f}$)	0.77	0.68	0.87
Percent of cities within $1 \pm \frac{1}{2}\sigma_{\varepsilon^{HT}_{2}.91}$	25.7%	20.0%	24.5%
Percent of cities within $1 \pm \sigma_{\varepsilon^{HT}_{2}.91}$	51.4%	40.8%	49.7%
Panel B: Basic model using the urban informal sector			
Average prediction error ($\varepsilon^{HT}_{1.i}$)	0.41	0.44	0.43
Percent of cities within $1 \pm \frac{1}{2}\sigma_{\varepsilon^{HT}_{2}.91}$	2.3%	3.6%	5.6%
Percent of cities within $1 \pm \sigma_{\varepsilon^{HT}_{2}.91}$	1.1%	1.4%	1.2%
Panel C: Model with two urban sectors			
Average prediction error (ε^{HT}_{2})	1.18	1.12	1.30
Percent of cities within $1 \pm \frac{1}{2}\sigma_{\varepsilon^{HT}_{2}.91}$	43.1%	40.8%	31.4%
Percent of cities within $1 \pm \sigma_{\varepsilon^{HT}_{2}.91}$	72.5%	69.9%	57.4%
Panel D: Model with housing market			
Average prediction error (ε^{HT}_{3})	0.99	N/A	1.06
Percent of cities within $1 \pm \frac{1}{2}\sigma_{\varepsilon^{HT}_{2}.91}$	51.03%	N/A	44.44%
Percent of cities within $1 \pm \sigma_{\varepsilon^{HT}_{2}.91}$	87.24%	N/A	77.78%

Notes: Authors' calculations using census microdata. The errors in Panels A and B are computed using equation 2. Errors in Panels C and D are calculated using equations 3 and 4, respectively, assuming$\alpha = 1/3$ in the latter. We report the average and the percentage of cities whose errors fall within half/one standard deviation of one (the "perfect fit" benchmark) across cities and separately for each census year. To make columns and rows comparable we use, in all cases, the standard deviation of the error of the model with two urban sectors in 1991, $\varepsilon^{HT}_{2,91}$. We residualize wages and rents on the interaction of macroregion and year fixed effects before calculating the prediction errors. We trim errors at 1%. Appendix Table B.4 shows very similar results obtained without trimming. The Data Appendix further describes the precise computation of HT prediction errors.

sensitive to the limitations of our measurement approach. Stepping outside of the basic framework to consider the labor force participation margin – under the assumption of an upward-sloping urban labor supply curve – note that a lower reference wage would in principle lead to fewer labor market participants, yielding a higher employment rate for any given employment level. Since we assume away non-participants, we do not capture this effect, potentially introducing a negative bias into the measurement of the employment rate and the HT error. That bias may be more pronounced in the informal-sector model given that the informal wage is smaller than the formal wage.

3.2. Model Extensions

This section moves beyond the basic model to explores two extensions. The first simultaneously introduces a formal and an informal sector in the urban labor market, and the second introduces a housing market, which allows us to account for the effects of the increased costs of living in urban areas.

Segmented Urban Labor markets

In developing-country cities, unemployment tends to be lower than in the developed world. This is because, at low income and saving levels, people oftentimes cannot afford to survive without an income stream. Facing scarcity of formal jobs, these workers frequently engage in informal economic activities. The existence of an urban informal sector has been considered in the Harris-Todaro literature since its early years (e.g., Fields, 1975; Mazumdar, 1976). In some variations of the model, the urban informal sector takes the place of unemployment, such that workers in cities can be either formally employed or informally employed (e.g., Brueckner and Zenou, 1999). By contrast, our approach is to incorporate the informal sector into the model, but to retain the possibility that urban dwellers may be unemployed as in Gupta (1993).

We expand the model to accommodate two urban sectors, formal and informal, and we assume that they are fully segmented. In the spirit of the original HT formulation, workers randomly find jobs in one sector or the other, depending on each sector's share in total employment. The formal, informal, and rural sectors have the same generic production function as in the basic model, and their labor demand follows equation 1 with $s = \{f, i, r\}$. The introduction of the informal sector requires us to redefine the expected urban wage as $E[w_u] = \frac{w_f L_f + w_i L_i}{L_f + L_i + L_n}$, where w_f and w_i are the urban formal and informal wages, and L_f

and L_i are urban formal and informal employment, respectively. The equilibrium condition $E[w_u] = w_r$ can be rewritten as $m_2 = w_2$, where $m_2 = \frac{L_f + L_i}{L_f + L_i + L_n}$ is the urban employment rate, and the rural-urban wage ratio $w_2 = \frac{w_r}{E_e[w_u]}$ is calculated using the expected wage of the *employed* urban workers, $E_e[w_u] = \frac{w_f L_f + w_i L_i}{L_f + L_i}$. The corresponding HT prediction error is:

$$\varepsilon_2^{HT} = \frac{m_2}{w_2}. \tag{3}$$

When formal and informal wages are considered simultaneously the model's empirical traction improves significantly. Panel C of Table 3 reports the summary statistics of these errors. They tend to be larger than one (with averages ranging between 1.12 and 1.30), indicating that the model predicts a smaller employment rate (i.e., larger rural-urban migration) than we actually observe. Still, relative to the single-sector models, a larger share of the errors are close to one; between 31% and 43% are within half a standard deviation, and between 57% and 73% are within one standard deviation.

Introducing Housing Markets

Costs of living in the city can, in principle, act as a deterrent of rural-urban migration, effectively reducing the urban real wage. Brueckner and Zenou (1999) and Brueckner and Kim (2001) formally incorporate urban land into a model in the Harris-Todaro tradition, and they show that rural-urban migration can raise the price of urban land, making the city more expensive to live in and deterring further migration. This effect is at the core of the standard within-city (Alonso 1964; Muth 1969) and across-cities (Rosen 1979; Roback 1982) spatial equilibrium models in urban economics.

To keep the model tractable we focus on the housing market, and assume homogeneous housing. This contrasts with other models in the HT literature, which consider multiple land prices within the city (Brueckner and Zenou, 1999). With this choice, we are able to succinctly capture the demand effects that migration has in the urban housing markets, at the expense of abstracting from questions related to the location of rural-urban migrants within the city.[19]

Production and labor demand continue to have three sectors as before. On the labor

[19]The within-city location may in turn be related to urban informality and unemployment. For example, Posada and Moreno-Monroy (2017) show that rural-urban migration increases with cities' decentralization, and argue that this effect is driven by lower costs of housing in the outskirts. They also find that this effect is correlated with a larger decentralization of informal jobs.

supply side, we now assume that homogeneous workers derive utility from the consumption of a tradable good C, which we treat as the numeraire, and housing, which is rented at r_a for areas $a = \{u, r\}$. They choose location a to solve the optimization problem:

$$\max_a \left\{C_a^{1-\alpha} H_a^{\alpha}\right\} \ s.t. \ E\left[w_a\right] = C + r_a H$$

where $E\left[w_u\right]$ is defined as before, and $E\left[w_r\right] = w_r$. It follows that housing demand at their location of choice will be given by:

$$H_a = \alpha \frac{E\left[w_a\right]}{r_a}.$$

Total housing demand in each area is given by $H_a L_a$ where L_a is the number of workers locating in area a. The indirect utility function can thus be expressed as

$$V_a = \alpha_1 E\left[w_a\right] r_a^{-\alpha}$$

with $\alpha_1 := \alpha^{\alpha}\left(1-\alpha\right)^{1-\alpha}$. Workers choose to locate, and inelastically supply one unit of labor, in the area in which their *expected utility* is higher. The equilibrium condition is therefore:

$$\alpha_1 E\left[w_u\right] r_u^{-\alpha} = \alpha_1 w_r r_r^{-\alpha}$$

The relevant wage now is the (expected) *real wage*, that is, the expected wage adjusted by the housing rents. The equilibrium condition can be expressed as the equality of the urban employment rate and the *real* rural-urban wage gap, $m_2 = w_2 r^{-\alpha}$, where $r = \frac{r_r}{r_u}$ is the rural-urban housing rents ratio. This corresponds to an HT prediction error that accounts for differences in costs of living across urban and rural areas:

$$\varepsilon_3^{HT} = \frac{m_2}{w_2 r^{-\alpha}}. \tag{4}$$

Panel D of Table 3 evaluates the errors in equation 4 for the years in which the Brazilian census provides data on housing rents (1991, and 2010), assuming an exogenous share of income spend in housing of $\alpha = 1/3$.[20]

Including the urban cost of living in the model significantly improves its performance in the data. The average prediction errors are now noticeably closer to one (0.99 in 1991

[20]In Appendix Table B.3 we assess the equilibrium condition using formal (w_f) and informal (w_i) wages in the context of a model that considers housing markets.

and 1.06 in 2010). Furthermore, a larger fraction of cities have errors that are close to one, with 87% within one standard deviation in 1991, and 78% in 2010. Relative to the model with two urban sectors but no housing market (error ε_2^{HT}), introducing differences in costs of living between urban and rural areas to the model yields relatively higher urban employment rates (i.e., relatively lower migration) predictions, which are closer to the employment rates observed in the data.[21]

Taking Stock

The analysis above makes it clear that, while the basic framework with a single urban sector fits the Brazilian data poorly, the empirical performance of the model improves significantly when one extends the model to accommodate two segmented urban labor markets and living costs differentials between urban and rural areas.

This can be seen more clearly in Figure 2, which shows the distribution of the prediction errors in the range of zero to two for alternative models, pooling observations from all census years. In the two versions of the model that feature a single urban sector, the equilibrium conditions predict urban employment rates that are higher than the ones observed in the data. This implies that, when the economic opportunities offered by one of the urban sectors are assumed away, the model predicts less rural-urban migration than actually occurs. The gap is more pronounced in the informal-sector model, which neglects the urban sector that, in practice, offers the higher expected wage (i.e., the formal sector). The best-performing reference urban wage is a weighted average of wages in the formal and informal sectors. As the figure shows, the distribution of the resulting error term has a noticeably larger share of its mass around one, the "perfect fit" benchmark. Moreover, if we also account for differences in costs of living across urban and rural areas in the "housing markets" model, the tightness of the distribution around one increases.

[21] It is the combination of housing markets and two urban sectors that improves the performance of the model in the data. Introducing housing markets into a model with one urban sector (formal or informal) does not improve it. See Appendix Table B.3.

Figure 2: Distributions of the HT Prediction Error under Alternative Models

Notes: Authors' calculations using census microdata. The figure shows the kernel density estimates of the pooled HT prediction errors' distributions at the city level (Panels B, C, D, and E of table 3). The figure shows errors trimmed at 1% and truncated distributions at 2. They are very similar except for a longer right tail caused by 11 cities. The Data Appendix further describes the precise computation of HT prediction errors.

The relatively poorer empirical performance of the models with a single urban sector sheds light on why there is relatively scarce evidence supporting the Harris-Todaro predictions related to urban unemployment, and in particular the "Todaro Paradox" (Todaro, 1969). Under the assumption that the only alternatives available to workers that locate in an urban area are to either be employed (in the formal sector) or to be unemployed, the paradox posits that efforts to reduce unemployment through urban jobs promotion may backfire, attracting a disproportionate number of rural-urban migrants and pushing up unemployment rates. Papers that have empirically tested predictions inspired in the HT framework have found empirical support for the premise that rural migrants are driven by the size of the rural-urban wage gap. However, finding empirical evidence of the Todaro Paradox has been more elusive (Lall et al., 2006).[22] Our analysis implies that this may be, at least partially, due

[22]In the studies that are consistent with these predictions, the evidence is mostly indirect or applies only to certain sub-populations. For instance, Todaro (1976), finds indirect empirical support for the Todaro Paradox by estimating rural-urban migration elasticities in 14 developing countries, and arguing that, in most cases, the elasticities fall in the range in which the theory predicts the paradox would hold. Barnum and Sabot (1977) and Schultz (1982) also find some support the hypothesis that migrants respond to favorable employment rates in the destination. However, the finding in Schultz (1982) holds only for males with at least secondary education, and not for workers with lower schooling levels.

to the fact that the urban informal sector offers economic opportunities greater than those in the rural areas, and analyses that treat informal work as equivalent to unemployment overlook this important migration pull factor. The results also suggests a promising avenue to further develop the framework and improve its empirical performance: improving the characterization of the urban labor markets in the model by explicitly incorporating the labor force participation margin (i.e., by adding "non-participant" to the set of potential urban states, and linking the participation decision of workers to the observable economic opportunities in the formal and informal sector).

This line of reasoning also points to the possible existence of an "informality paradox." Recent research has highlighted that high levels of informality can be associated with lower human capital accumulation (Bobba et al., 2017, 2020) and lower overall productivity (Busso et al., 2012; Ulyssea, 2018), pointing to potential welfare gains from policies promoting formalization. Our analysis suggests that – depending on how economic opportunities in the formal sector interact with the informal sector and with labor force participation decisions – policies that foster formal labor demand in cities could result in *higher* urban informality rates if they prompt a disproportionate migratory response. We see the empirical investigation of this hypothesis as a promising avenue for future research.

3.3. Robustness

We perform four robustness checks to the results presented in Table 3, and show that our conclusions are robust to a number of decisions made in the the empirical analysis. Table 4 reports the results. The first three columns show the errors associated with the model with segmented labor markets (equation 3), while the last two columns evaluate the errors associated with the model that incorporates housing markets (equation 4).

First, in the model with housing markets the equilibrium conditions were computed assuming that households spent a third of their income on housing ($\alpha = 1/3$), a share that was assumed to be exogenous and fixed across localities. This is a common assumption in the literature, and is based on the share of housing in income in the United States (Chauvin et al., 2017; Glaeser and Gottlieb, 2009). This value is also very close to the actual average share of household income spent on housing (28.8% in 1991 and 29.7% in 2010) according to the Brazilian data. Panel A evaluates the sensitivity of the results to calculating the errors in equation 4 with the observed average $\alpha_i t$ in each locality i for each census year $t = 1991, 2010$.

Second, in assessing the different models' equilibrium conditions, we used the average wages in cities and their respective rural catchment areas. However, De la Roca and Puga (2017) have shown, using data from Spain, that there are returns to city experience. Potential migrants may use information that correspond to recent migrants (rather than the average wage in the commuting zone) in forming their expectation of their wage returns to migration. Recent-migrant wages may be smaller than the average wage if the returns to living in cities increase over time in Brazil as it does in Spain. To consider this possibility, we compute the equilibrium errors using the average formal and informal wages of migrants who arrived in each city in the five previous years of each census date. Panel B shows the results.

Third, we assess the robustness of the results to our definitions of cities and their respective rural catchment areas. Figure A.1 shows that some cities are located in areas of the country that are, in general, more sparsely populated, such as the Amazon rainforest. Panel C explores how sensitive the results are to excluding cities located in areas with low population density. Relatedly, our definition of rural catchment areas relies on weights that measure the contribution of different rural municipalities to the pool of migrants in each city. In our main results, shown in Table 3, these weights were allowed to vary over time. Panel D in Table 4 recomputes the results using fixed weights, which are calculated using the migration patterns of 1980.

In all four robustness checks, we find that the average prediction error and the percent of cities within 1/2 and 1 standard deviation $\sigma_{\varepsilon^{HT}_{2,91}}$ are largely unchanged, with only marginal drops or improvements of the models' fit to the data. The largest difference comes from the errors that use the urban wages of the most recent migrants in Panel B, which yields an average prediction error that is 10 percent larger than in the benchmark results.

Table 4: Robustness Tests

	Model with two urban sectors			Model with Housing Market	
	1991	2000	2010	1991	2010
Panel A: Share of housing expenditure by city-year					
Average prediction error	1.18	1.12	1.30	1.01	1.12
Percent of cities within $1 \pm \frac{1}{2}\sigma_{\varepsilon_2^{HT},91}$	43.1%	40.8%	31.4%	50.5%	42.2%
Percent of cities within $1 \pm \sigma_{\varepsilon_2^{HT},91}$	72.5%	69.9%	57.4%	87.0%	75.6%
Panel B: Using recent migrants to urban areas					
Average prediction error	1.41	1.28	1.52	1.11	1.16
Percent of cities within $1 \pm \frac{1}{2}\sigma_{\varepsilon_2^{HT},91}$	29.1%	38.1%	23.8%	53.2%	48.3%
Percent of cities within $1 \pm \sigma_{\varepsilon_2^{HT},91}$	51.1%	64.6%	45.3%	84.6%	76.0%
Panel C: Cities above the 10th percentile of density distribution					
Average prediction error	1.19	1.13	1.32	1.00	1.07
Percent of cities within $1 \pm \frac{1}{2}\sigma_{\varepsilon_2^{HT},91}$	42.6%	38.7%	29.7%	52.0%	44.4%
Percent of cities within $1 \pm \sigma_{\varepsilon_2^{HT},91}$	70.4%	68.1%	54.0%	87.2%	76.7%
Panel D: Migration weighting matrix fixed in 1980					
Average prediction error	1.18	1.11	1.27	0.99	1.03
Percent of cities within $1 \pm \frac{1}{2}\sigma_{\varepsilon_2^{HT},91}$	43.1%	40.1%	31.4%	50.9%	48.6%
Percent of cities within $1 \pm \sigma_{\varepsilon_2^{HT},91}$	72.5%	71.1%	56.7%	87.0%	80.3%

Notes: Authors' calculations using census microdata. We report the average and the percentage of cities whose errors fall within half/one standard deviation of one (the "perfect fit" benchmark) across cities and separately for each census year. To make columns and rows comparable we use, in all cases, the standard deviation in 1991 of the error of the model with two urban sectors but no housing market, $\varepsilon_{2,91}^{HT}$. We calculate the errors in columns 1 to 3 using equation 3, while the last two columns evaluate the errors using equation 4. Errors in Panel A use the α (average share of income spent on housing) observed in each city and year. Errors in Panel B use the average urban wage of migrants that arrived in the city during the five years previous to the census. For Panel C, we exclude the cities below the 10th percentile of each year's density distribution. In Panel D, we compute all census years' errors using rural-urban migration' shares to each city between 1981-1986. We residualized wages and rents on the interaction of macroregion and year fixed effects before calculating the prediction errors. We trim errors at 1%. The Data Appendix further describes the precise computation of HT prediction errors.

4. The Role of Human Capital Heterogeneity

The standard Harris-Todaro model assumes homogeneous workers, and the extensions of the model discussed above maintain this assumption. However, it is informative to explore how human capital heterogeneity may matter for the empirical traction of the framework. Fully specifying a model that adds heterogeneous human capital is beyond the scope of this study. Instead, we approach this issue empirically, replicating the analysis from Table 3 for subsamples of the population.

To this end we divide the working-age population in subgroups on the basis of their level of age, schooling, and gender. We then use the microdata to re-calculate labor market measures (i.e., wages and employment in each sector) using only the observations belonging to the subgroup. We recompute the errors from the two segmented urban labor markets (equation 3) and the housing market (equation 4) models with these variables. For example, the prediction error of the housing market model for the subpopulation with primary schooling is computed as $\varepsilon_{3,p}^{HT} = \frac{m_{2,p}}{w_{r,p}/E_e[w_{u,p}]r^{-\alpha}}$, where the subindex p indicates that the variable is calculated with individuals who report having completed primary but not high school, such that $m_{2,p} = \frac{L_{f,p}+L_{i,p}}{L_{f,p}+L_{i,p}+L_{n,p}}$, and $E_e[w_{u,p}] = \frac{w_{f,p}L_{f,p}+w_{i,p}L_{i,p}}{L_{f,p}+L_{i,p}}$. Appendix Table B.5 pools observations across census years and reports the cross-city averages of the labor market variables used to compute the errors, Figure 3 plots the distributions of these errors, and Appendix Table B.6 reports summary statistics for these distributions.

We first look at differences across groups of different ages. The distribution of the errors (Figure 3, top left) looks similar for the errors calculated with data from individuals aged 15 to 39 at the time of migrating, and with data of those aged 40 and older. The mean is precisely one for the younger cohort, and barely different (1.02) for the older cohort. In both cases around 80% of cites have an error within one standard deviation from the perfect fit benchmark (Appendix Table B.6). In the version of the model that accounts for geographic differences in the cost of living, the model fits well with the data for both subgroups, even though each faces different sets of incentives. As shown in Appendix Table B.5, the expected urban wage is on average closer to the rural wage for the cohort aged 15 to 39 than for the cohort aged 40 and older. Consistent with the model, the younger cohorts also have, on average, higher urban employment rates.

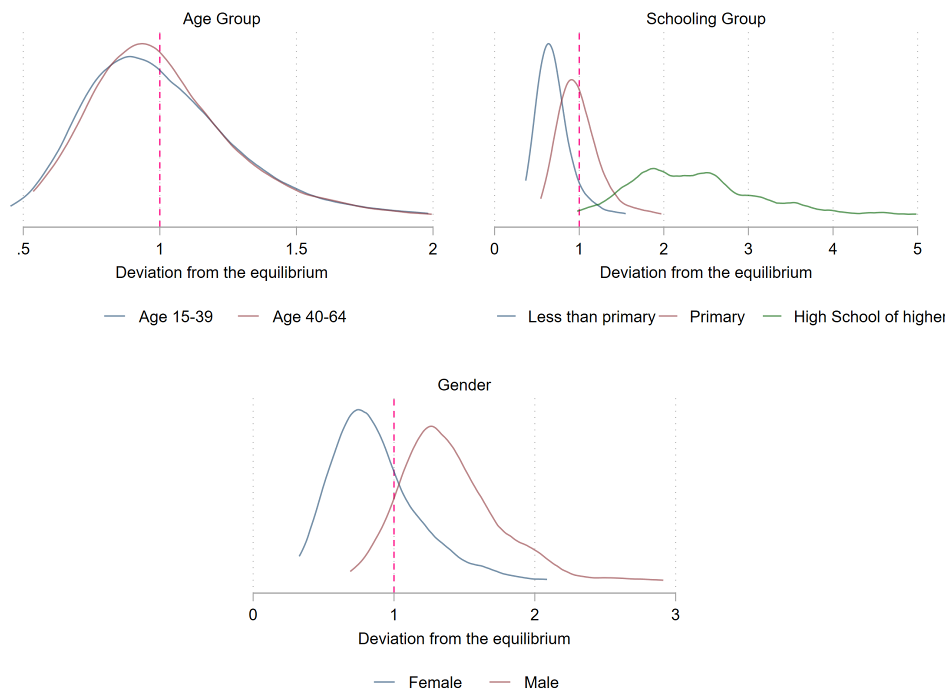
By contrast, the errors vary significantly by subgroup when we split the sample by schooling (Figure 3, top right). Relative to the model-driven predictions, the observed employment rate tends to be smaller than one for the sample of workers with less than primary education,

indicating rural-urban migration flows that are higher than would have been predicted. The framework anticipates relatively lower migration in this case because the rural-urban wage gap is relatively smaller than in other schooling subgroups, with the rural wage being, on average, 71% of the urban wage, and 82% of the urban informal wage (Appendix Table B.5). Yet the observed employment rates are relatively smaller than for other subgroups, implying larger migration rates. A possible explanation is that some migrants with low levels of schooling move to the city to acquire additional education, expecting a higher wage in the future.

Among the three schooling subgroups, the model performs significantly better in the sample of workers with primary but not high school education. In fact, when we restrict our analysis to this subgroup, the housing market model fits the data even better than in the full sample, with an average error of 0.99, and 87% of the cities falling within one standard deviation of one (Appendix Table B.5). By contrast, the errors calculated with the sample of individuals with high school or higher education fall mostly far away from what the model would predict. For this subgroup, the model drastically underestimates the urban employment rate, and for only 3.8% of cities the error falls within one standard deviation from one. This contrasts with the findings of Schultz (1982), who observes migration elasticities consistent with the HT framework only for individuals with at least high school education in Venezuela in the 1960s.

We also consider the empirical performance of the Harris-Todaro model across genders. Figure 3 (bottom) reports errors calculated using data for either males only or females only, showing that the distributions of these errors are centered at either side of the perfect fit benchmark. As shown in Appendix Table B.5, while the gap between the rural wage and the formal urban wage is very similar for both genders, the informal urban wage is significantly larger than the rural wage for men than for women. This leads the model to predict relatively higher migration and relatively lower employment rates for the male sample. However, the observed male employment rate is noticeably above the prediction, and the female employment rate is below the prediction. This likely reflects larger labor force participation rates among males, highlighting how abstracting away from the labor force participation decision of workers limits the empirical traction of the model. These patterns also resonate with prior work showing that migration across Brazilian regions and the relative labor market outcomes of males and females are consistent with a model in which couples move together and labor demand favors male employment (Chauvin, 2018), and, more generally, with the observation that migration and labor force participation decisions of members of the same family units are interrelated (Lall et al., 2006; Gemici, 2011).

Figure 3: HT Prediction Errors Calculated for Population Subgroups



Notes: The figure shows the kernel density estimates of the HT prediction errors' distribution of the two sectors with housing markets model at the city level by population subgroups (Panels A, B, C of Table B.6). We use wages and employment of observations belonging to each subgroup and the city's average rent to calculate the errors. Pooled sample from 1991 and 2010 census years (rents are not available in 2000). We residualize wages and rents on the interaction of macroregion and year fixed effects before calculating the prediction errors. We trim errors at 1%. The Data Appendix further describes the computation of HT prediction errors.

More generally, we note that, when looking at individual subgroups, the housing markets model is not necessarily the one that best fits the data. Appendix Table B.5 compares the summary statistics of the errors of the two urban sectors model without housing markets (ε_2^{HT}) with those of the model with housing markets (ε_3^{HT}) for each subgroup. In all cases, the introduction of housing rents leads the framework to predict lower rural-urban migration and higher employment rates, shifting the distribution of the errors to the left. For the subgroups for which the two urban sectors model predicts more migration than observed (i.e., where errors ε_2^{HT} are centered above one), this makes the housing rents model fit the data better. Conversely, in subgroups where the two urban sectors model without housing markets underpredicts migration (i.e., where the average ε_2^{HT} is below one), introducing housing rents *worsens* the model's fit.

5. HT Equilibrium and Location Characteristics

To understand which location characteristics are associated with the size of the HT prediction errors ε_{it}^{HT} in locations i during census years t, our empirical specification looks at the absolute value of the prediction errors for the model with housing markets, estimating the following regression:

$$|\varepsilon_{3,it}^{HT} - 1| = \beta_0 + \delta C_{i,t-10} + \mu_{year} + \mu_{region} + \varepsilon_{i,jt}$$

where $C_{i,t-10}$ is a vector of variables for city i observed in the previous census year that capture a set of characteristics from the urban zone, its rural catchment area, and the potential pool of migrants, μ_t are census-year fixed effects, and η_{region} are fixed effects for the five macroregions of the country. The dependent variable recenters the errors around zero and uses their absolute values to be able to measure the effects of the covariates on the overall model's fit regardless of whether the main errors lie above or below the one. We estimate this regression pooling the data from the three census rounds together. To assess if certain correlates have asymmetric effects on moving a city closer/farther from the equilibrium, we also present models in which we restrict the sample to observations with prediction errors larger than one ($(\varepsilon_{it} - 1)^+$) and to observations with prediction errors smaller than one ($(\varepsilon_{it} - 1)^-$), respectively. Table 5 reports the results.

The HT model assumes that the urban area and its associated rural area are connected so that people are free to move across space (thus linking both labor markets), and that there are zero migration costs. To assess the role of this assumption, we look at the average distance between the urban area and its catchment area by computing a weighted average of the distance between the centroids of the city[23] and of each rural municipality that contributed migrants in the past (with the weights being the share of migrants coming from each municipality). A negative effect of distance on migration has consistently been found in multiple studies (e.g., Barnum and Sabot, 1977; Fields, 1982; Schultz, 1982). We find that, consistent with the assumptions of the model, the absolute value of the HT prediction errors tend to be larger for those urban areas whose rural areas tend to be farther away (where migration costs are likely higher). Reassuringly, this effect is driven by the correlation of errors that are larger than one. That is, the effect stems from cities for which the model predicts an urban employment rate that is smaller than the one observed; in other words,

[23]Given that our definition of cities includes all municipalities that are part of the same commuting zone, we use the centroid of the most populated municipality in the city.

the model predicts more migration than occurs in reality.

Secondly, we find that the prediction errors are smaller for areas that are more archetypically rural. Rural areas are usually defined by statistical agencies as those regions that have a low population density and are located outside cities. Our catchment areas are a collection of low-density (possibly dispersed) localities that could include small towns that themselves attract migrants. Potential migrants in more dispersed and in more densely populated rural areas might form a different wage expectation of migration than the one implied by the model. We therefore include in our regression a measure of the population density of the rural catchment area, and find that the HT prediction errors are larger for cities for which this area is more densely populated. We also include measures of the size of the rural catchment areas: the prediction errors are larger in cities whose catchment areas are composed by more localities and those that have more population. The positive correlation is driven by errors that are higher than one: for these less archetypically rural areas, the HT model overpredicts migration.

Third, for the model equilibrium to hold, there should be a large urban area that can attract enough potential migrants by providing higher expected utility. We find that the HT prediction error is indeed smaller for larger cities. We next analyze if, once expected urban-rural real wage differentials are taken into account, push-pull factors can explain which cities would be in an HT equilibrium. $C_{i,t-10}$ includes the 10-year lagged change in the share of workers employed in manufacturing and services in the urban areas and 10-year lagged change in the share of workers employed in agriculture in their associated rural areas. We find that cities with an expanding service sector are closer to the HT equilibrium.[24] Similarly, cities whose associated rural areas have an expanding agricultural sector seem to have larger errors (although the coefficient is not statistically significant at standard levels).

Fourth, the original HT model assumed homogeneous labor. In the data, however, workers' heterogeneity can also affect the fit of the model. The literature has consistently found that young adults are more likely to migrate than older or very young individuals. This has typically been associated with the view that migration is an investment, and since the number of working years are finite, it has a higher present value for younger workers (Kennan and Walker, 2011). We find that, consistent with this empirical regularity, rural areas with a higher share of the population that is 15-39 years old have a smaller prediction error. Moreover, the errors tend to be larger where these rural populations have different educa-

[24] In 2010, in the average city, the service sector employed 7.1 times more workers than the manufacturing sector.

tion levels than their associated urban centers – arguably, making it more difficult for these potential migrants to find jobs.

Table 5: Correlates of the HT Prediction Errors with Segmented Labor Markets and Housing Markets

	$\lvert\varepsilon_3^{HT}-1\rvert$	$(\varepsilon_3^{HT}-1)^-$	$(\varepsilon_3^{HT}-1)^+$
Distance urban-rural areas	0.039***	-0.006	0.048**
	(0.011)	(0.011)	(0.016)
Rural: Population density	0.007***	-0.004***	0.011***
	(0.001)	(0.001)	(0.003)
Rural: Municipalities in the catchment area (in 100s)	0.049***	0.029***	0.053***
	(0.007)	(0.007)	(0.006)
Rural: Log(Population)	0.044**	0.006	0.120***
	(0.017)	(0.015)	(0.035)
Urban: Log(Population)	-0.045***	0.015	-0.056***
	(0.008)	(0.008)	(0.012)
Change in urban manufacture emp. share (t-1)	0.045	0.191	0.166
	(0.136)	(0.107)	(0.168)
Change in urban service sector emp. share (t-1)	-0.234	0.265*	-0.275
	(0.138)	(0.117)	(0.238)
Change in rural agriculture sector emp. share (t-1)	0.079	-0.046	0.084
	(0.059)	(0.061)	(0.104)
Rural: Share people aged 15-39	-0.014***	-0.004	-0.019*
	(0.004)	(0.004)	(0.008)
Difference in HS shares (urban-rural)	0.034	0.403**	0.361
	(0.166)	(0.134)	(0.272)
Year 2010	0.067*	-0.019	0.069
	(0.027)	(0.025)	(0.042)
Adjusted R^2	0.202	0.214	0.310
Observations	858	442	416

Notes: Regressions estimated at the city level using pooled data from 1991 and 2010 census years (rents are not available in 2000). The outcome in the first column is the absolute value of the $\varepsilon_3^{HT}-1$ (equation 4). The second and third column use samples restricted to positive and negative values of the $\varepsilon_3^{HT}-1$, respectively. Rural variables refer to the city's rural catchment area. All regressions include macroregion fixed effects. We describe the precise computation and definition of each variable in the Data Appendix. Robust standard errors in parenthesis. * p $<$0.1 , ** p $<$ 0.05, and *** p $<$ 0.01.

The cross-city correlates of the HT prediction errors suggest that our study may be relevant beyond Brazil. Some of the key characteristics associated with a better model fit – young rural populations, large cities with fast-growing service jobs in the formal and informal sectors, and low-density agricultural regions located not far from cities – are present across the rapidly urbanizing developing world.

We also report the coefficient on a dummy for the year 2010, to evaluate how the model fit changes relative to 1991. The fit of the model was better 30 years ago than in the last available census, indicating that as the country's urbanization increased, the fit of the HT model deteriorated.

Lastly, we evaluate how the ability of city characteristics to predict the size of the HT errors differs in simpler versions of the model. Appendix Table B.7 compares how $C_{i,t-10}$ correlates in the model with segmented labor markets but without housing markets (using $|\varepsilon_2^{HT} - 1|$ as dependent variable) and in the basic model computed with only the formal or the informal urban sector (using $|\varepsilon_1^{HT} - 1|$ as dependent variable). The correlation between $C_{i,t-10}$ and $|\varepsilon_2^{HT} - 1|$ is fairly similar to the correlation with $|\varepsilon_3^{HT} - 1|$, confirming that the model with segmented labor markets and the one that incorporates the housing markets are similar. However, the correlations between $C_{i,t-10}$ and $|\varepsilon_1^{HT} - 1|$ change; especially when only the informal wage is used, reflecting the bad fit of the basic version of the model.

6. Conclusions

Our findings show that the key insights of the HT framework can remain empirically relevant even at high levels of urbanization, so long as features of the original formulation are adapted to include some of the extensions proposed in the literature. The observed average urban wage appears to factor in migrants' decisions. It includes the wage in the urban informal sector, which seems to be seen more as a source of economic opportunities than as a potential risk akin to unemployment. High urban costs of living also play a role, counteracting the pull effects of high city wages.

The insights of the HT model apply to a greater degree to some localities than to others. The empirical traction of the framework did decrease on average as Brazil continued to urbanize; nevertheless, in nearly four out of five of the locations considered, urban employment rates in 2010 remained within one standard deviation of the level that would have been expected in an HT equilibrium. These were locations that closely resembled the environment that Harris and Todaro likely had in mind: one defined by cities large enough to attract rural workers, and by nearby rural areas with traditional agricultural economies, and dispersed populations that have a sizeable proportion of young adults. The predictive power of the model also differs across some population subgroups, in particular across genders and across schooling attainment levels. However, errors on either side of the perfect fit benchmark tend to offset each other, improving the model´s fit in the full sample of working-age individuals.

In terms of policy, our findings support the broader HT proposition that urban policymakers need to factor in the general equilibrium effects brought about by rural migration-driven population growth. These considerations are particularly relevant for medium and large cities with an archetypically rural catchment area nearby.

The prevalence of urban unemployment was a key motivation that led to the framework's creation half a century ago. Meanwhile, decades of policy discussions have debated the possible existence of a "Todaro paradox." Against this backdrop, we find that the basic version of the HT framework in which we consider only the urban formal sector has limited ability to predict employment rates in Brazilian cities. This is largely because it overlooks the critical role played by the urban informal sector, which offers income-generating opportunities to those unable to obtain work in the formal sector, and at wage levels that are still higher than those in rural areas.

Moving forward, it is possible to enrich the classic HT framework to further expand its empirical relevance in highly urbanized environments. Our research suggests that one can likely increase the empirical bearing of the framework by incorporating the labor force participation decisions of individuals and families, and their responses to the conditions of both the formal and informal sectors in cities. Furthermore, in the absence of a binding minimum wage, alternative micro-foundations for the urban-rural wage gap appear to be warranted. These could leverage some of the lessons developed by the urban literature in recent decades. For example, agglomeration economies can explain higher productivity levels and labor demand in cities. However, in the absence of wage rigidities, alternative mechanisms are needed to rationalize why rural-urban migration does not proceed further, to the point where the (real) urban and rural wages are equalized and urban unemployment is driven to zero. Potential explanations include non-wage determinants of migration ("amenities," broadly defined); other frictions, such as migration and search costs; tied migration among family members; and skills mismatch in the supply and demand of labor. Moreover, it may prove useful to nest rural-urban models in broader spatial equilibrium frameworks that also account for urban-urban flows. Further research is needed to understand how rural-urban migration is likely to proceed as the developing world continues to urbanize, and to better inform future policy.

References

Alonso, W., 1964. Location and Land Use. Harvard University Press, Cambridge, Mass, Cambridge, MA.

Barnum, H., Sabot, R., 1977. Education, Enployment Probabilities, and Rural-Urban Migration in Tanzania. Oxford Bulletin of Economics and Statistics 39 (May).

Bernard, A., Rowe, F., Bell, M., Ueffing, P., Charles-Edwards, E., 2017. Comparing internal migration across the countries of Latin America: A multidimensional approach. PLoS ONE 12 (3), 1–25.

Bobba, M., Flabbi, L., Levy, S., 2017. Labor market search, informality and schooling investments. IZA Discussion Paper No. 11170 (12091).

Bobba, M., Flabbi, L., Levy, S., Tejada, M., 2020. Labor market search, informality, and on-the-job human capital accumulation. Journal of Econometrics.

Borjas, G. J., 1987. Self-Selection and the Earnings of Immigrants. American Economic Review 77 (4), 531–553.

Brueckner, J. K., Kim, H. A., 2001. Land markets in the Harris-Todaro model: A new factor equilibrating rural-urban migration. Journal of Regional Science 41 (3), 507–520.

Brueckner, J. K., Lall, S. V., 2015. Cities in Developing Countries. Fueled by Rural-Urban Migration, Lacking in Tenure Security, and Short of Affordable Housing. In: Handbook of Regional and Urban Economics, 1st Edition. Vol. 5. Elsevier B.V., pp. 1399–1455.

Brueckner, J. K., Zenou, Y., 1999. Harris-Todaro models with a land market. Regional Science and Urban Economics 29 (3), 317–339.

Bryan, G., Morten, M., 2019. The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia. Journal of Political Economy 127 (5), 000–000.

Busso, M., Cristia, J., Hincapié, D., Messina, J., Ripani, L. (Eds.), 2017. Learning Better: Public Policy for Skills Development. Inter-American Development Bank, Washington D.C.

Busso, M., Fazio, M. V., Levy, S., 2012. (In)formal and (un)productive: The productivity costs of excessive informality in Mexico. IDB Working Paper NO. IDB-WP-341.

Calderón-Mejía, V., Ibáñez, A. M., 2016. Labour market effects of migration-related supply shocks: Evidence from internal refugees in Colombia. Journal of Economic Geography 16 (3), 695–713.

Calvo, G. A. ., 1978. Urban Unemployment and Wage Determination in LDC'S : Trade Unions in the Harris- Todaro Model. International Economic Review 19 (1), 65–81.

Calvo, G. A., Wellisz, S., 1978. Supervision, loss of control and the optimum size of the firm.

Journal of Political Economy 86 (5), 943–952.

Chauvin, J. P., 2018. Gender-Segmented Labor Markets and the Effects of Local Demand Shocks. Working Paper.

Chauvin, J. P., Glaeser, E., Ma, Y., Tobio, K., 2017. What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States. Journal of Urban Economics 98, 17–49.

Chiquiar, D., Hanson, G. H., 2005. International migration, self?selection, and the distribution of wages: Evidence from mexico and the united states. Journal of Political Economy 113 (2), 239–281.

Collier, P., 1979. Migration and Unemployment: A Dynamic General Equilibrium Analysis Applied to Tanzania. Oxford Economic Papers 31 (2), 205–236.

Combes, P. P., Démurger, S., Li, S., 2015. Migration externalities in Chinese cities. European Economic Review 76, 152–167.

de Brauw, A., Mueller, V., Lee, H. L., 2014. The role of rural-urban migration in the structural transformation of Sub-Saharan Africa. World Development 63, 33–42.

De la Roca, J., Puga, D., 2017. Learning by Working in Big Cities. Review of Economic Studies 84 (1), 106–142.

Dix-Carneiro, R., Kovak, B. K., 2019. Margins of labor market adjustment to trade. Journal of International Economics 117, 125 – 142.

Dustmann, C., Kirchkamp, O., 2002. The optimal migration duration and activity choice after re-migration. Journal of Development Economics 67 (2), 351–372.

Fields, G. S., 1975. Rural-Urban Migration, Urban Unemployment and Underemployment, and Job-search Activity in LDCs. Journal of Development Economics 2, 165–187.

Fields, G. S., 1982. Place-to-Place Migration in Colombia. Economic Development and Cultural Change 30 (3), 539–558.

Fields, G. S., 2005. A welfare economic analysis of labor market policies in the Harris-Todaro model. Journal of Development Economics 76 (1), 127–146.

García-Ferrer, A., 1980. Interactions between internal migration, employment growth, and regional income differences in Spain. Journal of Development Economics 7 (2), 211–229.

Gemici, A., 2011. Family Migration and LaborMarket Outcomes. Unpublished manuscript.

Giulietti, C., Wahba, J., Zenou, Y., 2018. Strong versus weak ties in migration. European Economic Review 104, 111–137.

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

Glaeser, E. L., Gottlieb, J. D., dec 2009. The Wealth of Cities: Agglomeration Economies

and Spatial Equilibrium in the United States. Journal of Economic Literature 47 (4), 983–1028.

Gupta, M. R., 1993. Rural-urban migation, informal sector and development policies A theoretical analysis. Journal of Development Economics 41 (1), 137–151.

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

Henderson, J. V., Storeygard, A., Deichmann, U., 2017. Has climate change driven urbanization in Africa? Journal of Development Economics 124, 60–82.

Hirvonen, K., Lilleør, H. B., 2015. Going Back Home: Internal Return Migration in Rural Tanzania. World Development 70, 186–202.

Hnatkovska, V., Lahiri, A., 2015. Rural and urban migrants in India: 1983-2008. World Bank Economic Review 29, S257–S270.

IBGE, 2016. Arranjos Populacionais e Concentrações Urbanas do Brasil, 2nd Edition. IBGE, Rio de Janeiro.

Jedwab, R., Vollrath, D., 2015. Urbanization without Growth in Historical Perspective. Explorations in Economic History.

Jiang, L., O'Neill, B. C., 2018. Determinants of Urban Growth during Demographic and Mobility Transitions: Evidence from India, Mexico, and the US. Population and Development Review 44 (2), 363–389.

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

Kovak, B. K., 2013. Regional effects of trade reform: What is the correct measure of liberalization? American Economic Review 103 (5), 1960–1976.

Lagakos, D., August 2020. Urban-rural gaps in the developing world: Does internal migration offer opportunities? Journal of Economic Perspectives 34 (3), 174–92.

Laing, D., Park, C., Wang, P., 2005. A Modified Harris-Todaro Model of Rural-Urban Migration For China. In: Critical Issues in China's Growth and Development. pp. 245–264.

Lall, S. V., Selod, H., Shalizi, Z., Dc, W., Bank, W., Paper, W., 2006. Rural-Urban Migration in Developing Countries: A Survey of Theoretical Predictions and Empirical Findings. World Bank Policy Research Paper 33915, 1–63.

Lall, S. V., Timmins, C., Yu, S., 2009. Connecting lagging and leading regions: The role of labor mobility. Brookings-Wharton Papers on Urban Affairs 2009, 151–174.

Lucas, R. E., 1985. Migration Amongst the Batswana. The Economic Journal 95 (378), 538–382.

Mazumdar, D., 1976. The urban informal sector. World Development 4 (8), 655–679.

Mazumdar, D., 1987. Rural- Urban Migration in Developing Countries. Handbook of Regional and Urban Economics II, 1097–1128.

Michaelsen, M. M., Haisken-DeNew, J. P., 2015. Migration magnet: the role of work experience in rural-urban wage differentials. IZA Journal of Migration 4 (1).

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

Muth, R., 1969. Cities and Housing. University of Chicago Press., Chicago and London.

Nguyen, L. D., Grote, U., Sharma, R., 2017. Staying in the cities or returning home? An analysis of the rural-urban migration behavior in Vietnam. IZA Journal of Development and Migration 7 (1).

Petrov, A. N., 2007. Revising the Harris-Todaro framework to model labour migration from the Canadian northern frontier. Journal of Population Research 24 (2), 185–206.

Posada, H. M., Moreno-Monroy, A. I., 2017. Informality, city structure and rural-urban migration in Latin America. Annals of Regional Science 59 (2), 345–369.

Roback, J., 1982. Wages, rents, and the quality of life. The Journal of Political Economy 90 (6), 1257–1278.

Rodríguez-Vignoli, J., Rowe, F., 2018. How is internal migration reshaping metropolitan populations in Latin America? A new method and new evidence. Population Studies 72 (2), 253–273.

Rosen, S., 1979. Wage-Based Indexes of Urban Quality of Life. In: Mieszkowski, P., Stras-Zheim, M. (Eds.), Current Issues in Urban Economics. Johns Hopkins University Press, Baltimore and London, pp. 74–104.

Salvatore, D., 1981. A theoretical and empirical evaluation and extension of the Todaro migration model. Regional Science and Urban Economics 11 (4), 499–508.

Sato, Y., 2004. Migration , Frictional Unemployment, and Welfare-Improving Labor Policies. Journal of Regional Science 44 (4), 773–793.

Schultz, T. P., 1982. Lifetime Migration within Educational Strata in Venezuela: Estimates of a Logistic Model. Economic Development and Cultural Change 30 (3), 559–593.

Stiglitz, J. E., 1974. Alternative Theories of Wage Determination and Unemployment in LDC's: The Labor Turnover Model. The Quarterly Journal of Economics 88 (2), 194–227.

Todaro, M. P., 1969. A Model of Labor Migration and Urban Unemployment in Less Developed Countries.

Todaro, M. P., 1976. Urban job expansion, induced migration and rising unemployment. A formulation and simplified empirical test for LDCs. Journal of Development Economics 3 (3), 211–225.

Ulyssea, G., 2010. Regulation of entry, labor market institutions and the informal sector. Journal of Development Economics 91 (1), 87–99.

Ulyssea, G., August 2018. Firms, informality, and development: Theory and evidence from brazil. American Economic Review 108 (8), 2015–47.

United Nations, P. D., 2019. World Urbanization Prospects: The 2018 Revision. Vol. 12. United Nations, New York.

World Bank, 2013. Planning, Connecting, and Financing Cities - Now: Priorities for City Leaders. Washington D.C.

World Bank, 2020. World Development Indicators. URL: databank.worldbank.org/source/world-development-indicators, accessed on 02/03/2020.

Zenou, Y., 2011. Rural-urban migration and unemployment: Theory and policy implications. Journal of Regional Science 51 (1), 65–82.

Appendix

A. Definition of Cities and Their Rural Catchment Areas

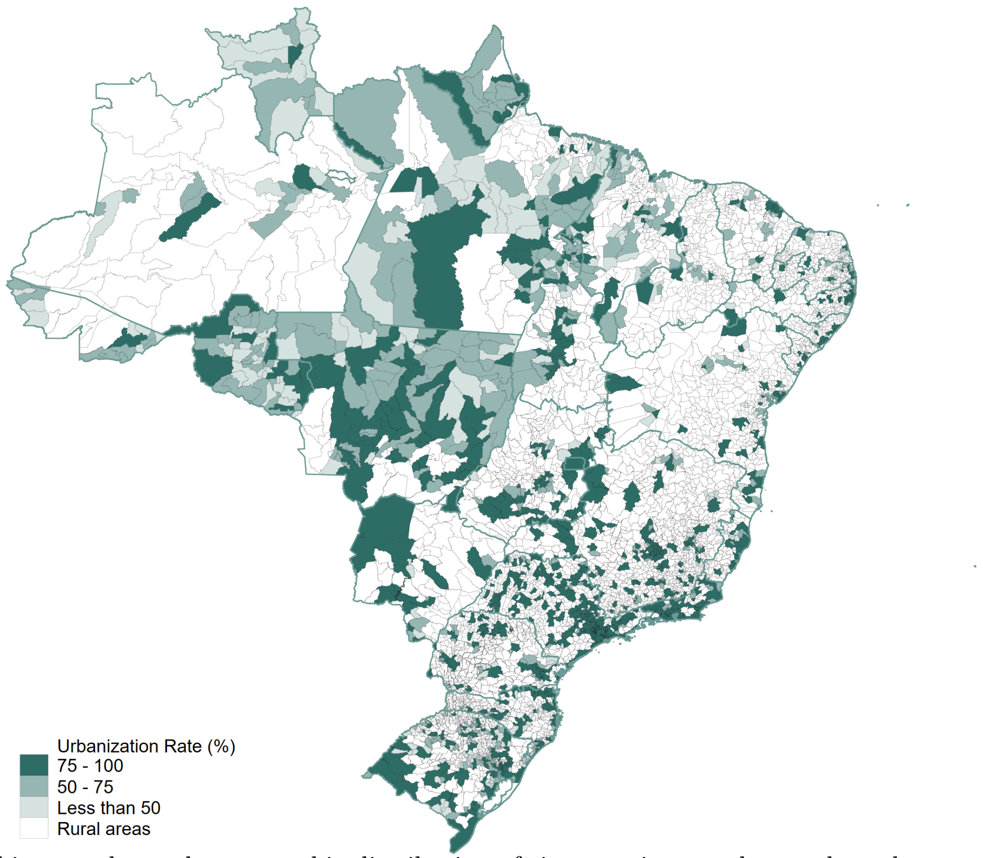
This appendix provides further details on our functional definition of urban and rural areas. We start by defining time-consistent municipalities by grouping together municipalities that share common ancestors (i.e., parent municipalities, or parents of parent municipalities) over the period 1970-2010, as in Chauvin (2018) and Kovak (2013). From the 5,565 municipalities that had been created by 2010, we obtain 3,803 time-consistent municipalities.

Next, we classify these time-consistent municipalities into cities and rural areas. To define cities, we use the "arrranjos populacionais" (commuting zones) definition provided by the Brazilian Institute of Geography and Statistics, IBGE (2016), which consists of sets of adjacent municipalities linked by intensive commuting for work or study in the 2010 census. We join commuting zones that share time-consistent municipalities. This yields 449 "cities" compressing 953 municipalities. The map in Figure A.1 displays the location of these cities, along with their urbanization rates. To calculate urban quantities and prices, we use only the observations from the microdata that come from *urban areas within the commuting zones*. All time-consistent municipalities that are not included in an urban commuting zone are treated as rural areas.

For each city in the sample we define a rural migration "catchment area" as the set of municipalities of origin of past rural-urban migrants – those that migrated between six and 10 years before.[25] We calculate the rural price measures – i.e., rural wages and rural housing rents – that correspond to each city in two steps. First, we compute wage and housing rent averages for each time-consistent rural municipality, using *only the observations identified as residing in rural areas* in the census microdata. Then, for each city, we calculate the weighted average of the rural prices computed in the prior step, using the share of each catchment-area municipality in the city's past rural-urban migration as weights.

[25] The 2000 census does not allow us to identify the municipality of previous residence for individuals who migrated before 1995. In this case, we instead use the share of migrants between 1991-1986 (from the 1991 census).

Figure A.1: Commuting Zones in Brazil, Time-Consistent Borders for 1970-2010

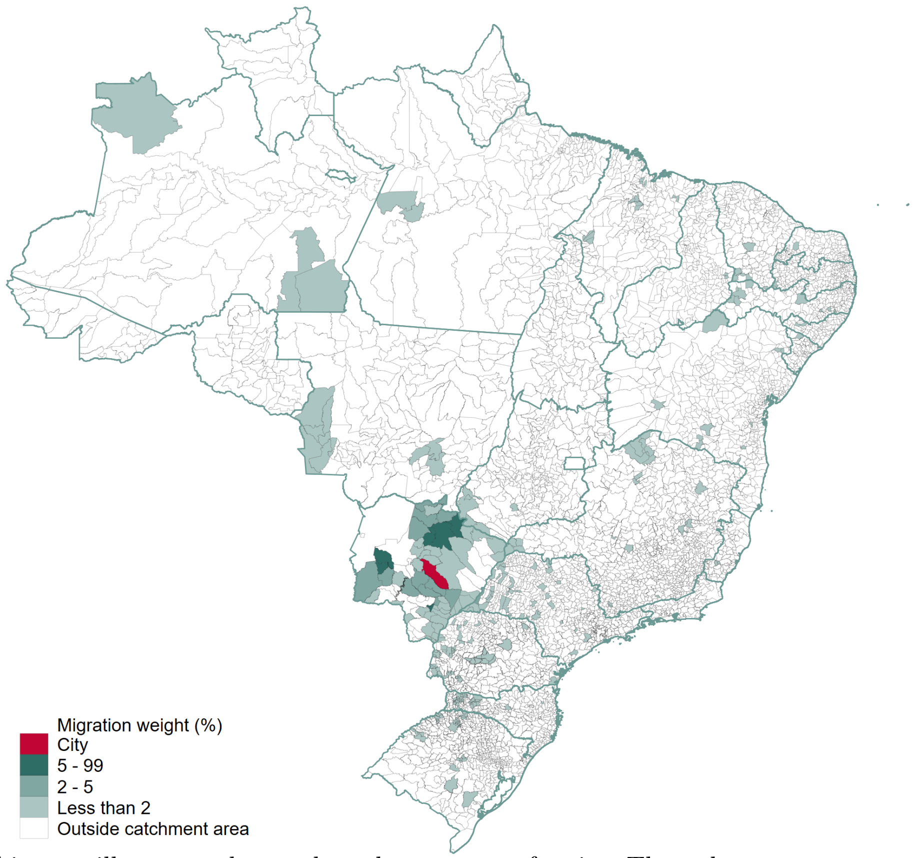


Notes: This map shows the geographic distribution of time-consistent urban and rural areas as described in Section 2.1. White zones represent the rural areas while green zones represent municipalities in an urban area. The urbanization rate represents the percentage of the population living in urban zones according to the Census of 2010 for each urban area, which is the population that we use to calculate urban prices and quantities.

The map in Figure A.2 illustrates this approach with the example of the city of Campo Grande, shown in red. Campo Grande is the capital and largest city the state of Mato Grosso do Sul. Between 2000 and 2005, rural migrants moved to the city from 169 different (time-consistent) rural municipalities, shown in green. For the year 2010, the urban wages and housing rents attributed to Campo Grande correspond to the average of these variables calculated over the individuals residing in urban zones of the municipalities that are part of the city. Employment and non-employment counts are also based on these observations. The rural prices attributed to the city, in turn, are weighted averages of the average wages and housing rents of the 169 rural municipalities in its catchment area, which are themselves calculated for individuals residing in rural zones in these municipalities. The weights employed in the calculation are given by the shares of the rural municipalities in the city's rural-urban

migration between 2000 and 2005. As the map illustrates, most of these migrants originate from a handful of rural municipalities.

Figure A.2: Campo Grande's Rural Catchment Area in 2010



Notes: This map illustrates the rural catchment area of a city. The red area represents the city of Campo Grande; green zones represent rural municipalities of origin of migrants that arrived in the city between 2000 and 2005. Darker regions represent a larger share of migrants from the total. White zones represent areas outside the catchment area.

B. Additional Tables

Table B.1: Internal Migrants' Origins and Destinations

Origins	Destinations				
	Rural	Urban			
		Q1	Q2	Q3	Q4
Panel A: 1991					
Rural	31.7%	2.3%	5.1%	9.0%	51.9%
Urban, Q1	28.1%	11.2%	10.8%	6.4%	43.4%
Urban, Q2	25.0%	2.7%	6.6%	9.6%	56.2%
Urban, Q3	21.4%	2.8%	4.7%	13.4%	57.7%
Urban, Q4	16.4%	1.6%	4.2%	5.8%	72.1%
Panel B: 2000					
Rural	32.7%	2.6%	5.6%	8.9%	50.2%
Urban, Q1	26.0%	13.1%	7.2%	10.3%	43.4%
Urban, Q2	27.1%	3.2%	9.4%	8.2%	52.1%
Urban, Q3	21.8%	1.7%	5.3%	15.7%	55.4%
Urban, Q4	16.7%	1.9%	3.3%	6.4%	71.6%
Panel C: 2010					
Rural	32.1%	2.4%	6.0%	10.4%	49.1%
Urban, Q1	28.4%	11.3%	5.9%	7.9%	46.4%
Urban, Q2	24.9%	3.2%	9.8%	9.6%	52.5%
Urban, Q3	21.1%	2.7%	5.7%	17.7%	52.8%
Urban, Q4	17.6%	1.8%	3.8%	7.0%	69.9%

Notes: Authors' calculations using census microdata. The table reports the origins and destinations of internal migrants in Brazil. Shares represent the proportion of emigrants from a geographical location (column) that arrived at a rural or urban destination (row) during the five years before the period of reference. We divided urban areas into four groups based on the quartile of cities' population distribution by year. Section 2.1 describes in detail our definition of urban and rural areas.

Table B.2: Summary Statistics of Cities and Their Rural Migration Catchment Areas

	Mean	Median	S.D.	Min.	Max.
Panel A: Urban					
Population (1000s)	234.7	76.5	960.8	2.3	19568.7
Share of HS educated (%)	23.4	22.6	11.6	1.6	57.1
Agriculture employment share (%)	10.5	7.3	11.3	0.0	67.8
Manufacturing employment share (%)	7.0	5.0	7.4	0.0	54.6
Services employment share (%)	27.2	33.5	17.3	0.0	58.1
Share of population aged 15-39 (%)	41.1	41.4	3.0	31.5	52.6
Labor force participation (%)	58.0	58.4	8.2	28.5	83.4
Change in labor force participation 1980-2010 (ppts.)	6.8	7.2	7.7	-25.8	33.4
Panel B: Rural					
Population (1000s)	10.2	0.3	16.1	0.1	148.0
Share of HS educated (%)	10.2	9.6	4.7	1.4	23.9
Agriculture employment share (%)	26.7	25.9	7.3	8.8	54.0
Manufacturing employment share (%)	6.1	5.3	3.2	0.5	19.7
Services employment share (%)	26.5	26.9	5.4	8.1	39.6
Share of population aged 15-39 (%)	40.5	40.5	3.1	25.6	61.9
Labor force participation (%)	52.4	54.1	8.2	30.9	74.9
Change in labor force participation 1980-2010 (ppts.)	4.8	4.5	4.0	-9.6	17.9
Distance to rural destinations (100 km)	2.4	1.6	2.3	0.2	19.3
Number of municipalities in catchment area	80.1	44.0	135.8	2.0	2171.0
Distance between rural destinations (100 km)	4.2	3.4	2.7	0.3	22.6
Population density (km^2)	12.6	0.4	20.9	0.1	100.2

Notes: Authors' calculations using census microdata. The table reports summary statistics at city level pooling data from 1991, 2000 and 2010 census. Rural variables refer to the average across city's rural catchment areas. We describe the precise computation and definition of each variable in Appendix A

Table B.3: Empirical Deviations from the Harris-Todaro Equilibrium using Formal and Informal Wages and Housing Markets

	1991	2010
Panel A: Model using the urban formal sector and housing markets		
Average prediction error	0.65	0.71
Percent of cities within $1 \pm \frac{1}{2}\sigma_{\varepsilon_2^{HT},91}$	16.9%	21.0%
Percent of cities within $1 \pm \sigma_{\varepsilon_2^{HT},91}$	40.0%	47.3%
Panel B: Model using the urban informal sector and housing markets		
Average prediction error	0.34	0.35
Percent of cities within $1 \pm \frac{1}{2}\sigma_{\varepsilon_2^{HT},91}$	0.4%	0.7%
Percent of cities within $1 \pm \sigma_{\varepsilon_2^{HT},91}$	1.6%	1.1%

Notes: Authors' calculations using census microdata. We report the average and the percentage of cities whose errors fall within half/one standard deviation $\varepsilon_{2,91}^{HT}$ of one ("perfect fit") across cities and separately for each census year. Errors in columns 1 and 2 are calculated using equation 4 replacing $w_2 = w_r/w_f$ for panel A, $w_2 = w_r/w_i$ for Panel B and assuming $\alpha = 1/3$. We trim errors at 1%. We residualize wages and rents on the interaction of macroregion and year fixed effects before calculating the prediction errors. The Data Appendix describes further the precise computation of HT prediction errors.

Table B.4: Empirical Deviations from the Harris-Todaro Equilibrium (No trimming)

	1991	2000	2010
Panel A: Basic model using the urban formal sector			
Average prediction error ($\varepsilon_{1,f}^{HT}$)	0.78	0.70	0.91
Percent of cities within $1 \pm \sigma_{\varepsilon_2^{HT},91}$	58.8%	44.8%	56.3%
Percent of cities within $1 \pm \frac{1}{2}\sigma_{\varepsilon_2^{HT},91}$	29.5%	22.9%	30.1%
Panel B: Basic model using the urban informal sector			
Average prediction error ($\varepsilon_{1,i}^{HT}$)	0.41	0.46	0.44
Percent of cities within $1 \pm \sigma_{\varepsilon_2^{HT},91}$	1.3%	1.8%	1.8%
Percent of cities within $1 \pm \frac{1}{2}\sigma_{\varepsilon_2^{HT},91}$	4.5%	7.6%	10.2%
Panel C: Model with two urban sectors			
Average prediction error (ε_2^{HT})	1.19	1.16	1.34
Percent of cities within $1 \pm \sigma_{\varepsilon_2^{HT},91}$	77.6%	75.7%	60.2%
Percent of cities within $1 \pm \frac{1}{2}\sigma_{\varepsilon_2^{HT},91}$	49.9%	45.2%	35.5%
Panel D: Model with housing market			
Average prediction error (ε_3^{HT})	1.00	N/A	1.07
Percent of cities within $1 \pm \sigma_{\varepsilon_2^{HT},91}$	89.66%	N/A	81.67%
Percent of cities within $1 \pm \frac{1}{2}\sigma_{\varepsilon_2^{HT},91}$	57.98%	N/A	52.94%

Notes: Authors' calculations using census microdata. The errors in Panels A and B are computed using equation 2. Errors in panels C and D are calculated using equations 3 and 4, respectively, assuming$\alpha = 1/3$ in the latter. We report the average and the percentage of cities whose errors falls within half/one standard deviation of $\varepsilon_{2,91}^{HT}$ across cities and separately for each census year. We residualize wages and rents on the interaction of macroregion and year fixed effects before calculating the prediction errors. Unlike Table 3.1, we do not remove extreme values to compute average prediction errors. The Data Appendix further describes the precise computation of HT prediction errors.

Table B.5: Variables Used to compute HT Prediction Errors by Subgroups

	Rural-urban wage ratio $\left({}^{w_r}\!/_{\bar{w}_u}\right)$	Rural-formal wage ratio $\left({}^{w_r}\!/_{w_f}\right)$	Rural-informal wage ratio $\left({}^{w_r}\!/_{w_i}\right)$	Urban employment rate $\left(\frac{L_f+L_i}{L_f+L_i+L_n}\right)$
Panel A: Age Group				
Age 15-39	61.6%	52.4%	77.5%	60.6%
Age 40-64	51.2%	42.1%	65.4%	57.8%
Panel B: Schooling Group				
Less than primary	71.4%	60.1%	81.6%	53.8%
Primary	66.8%	58.9%	75.2%	62.0%
High school of higher	56.0%	53.1%	63.2%	75.9%
Panel C: Gender				
Males	52.6%	45.8%	61.6%	75.6%
Females	54.5%	43.9%	76.1%	44.4%

Notes: Authors' calculations using census microdata. The table reports the average of the variables taken across cities by population subgroups and pooled across census years. Rural variables refer to the city's rural catchment area. We describe the precise computation and definition of each variable in Appendix A.

Table B.6: Empirical Deviations from the Harris-Todaro Equilibrium by Subgroups

	Model with two urban sectors (ε_2^{HT})	Model with housing market (ε_3^{HT})
Panel A: Age Group		
Average prediction error for age 15-39	1.18	1.00
Percent of cities within $1 \pm \frac{1}{2}\sigma_{\varepsilon_2^{HT},91}$	38.1%	44.8%
Percent of cities within $1 \pm \sigma_{\varepsilon_2^{HT},91}$	66.6%	79.4%
Average prediction error for age 40-64	1.20	1.02
Percent of cities within $1 \pm \frac{1}{2}\sigma_{\varepsilon_2^{HT},91}$	35.7%	49.9%
Percent of cities within $1 \pm \sigma_{\varepsilon_2^{HT},91}$	64.9%	80.3%
Panel B: Schooling Group		
Average prediction error for less than primary	0.81	0.69
Percent of cities within $1 \pm \frac{1}{2}\sigma_{\varepsilon_2^{HT},91}$	25.6%	14.2%
Percent of cities within $1 \pm \sigma_{\varepsilon_2^{HT},91}$	62.3%	41.4%
Average prediction error for primary	1.16	0.99
Percent of cities within $1 \pm \frac{1}{2}\sigma_{\varepsilon_2^{HT},91}$	45.8%	51.7%
Percent of cities within $1 \pm \sigma_{\varepsilon_2^{HT},91}$	73.6%	86.9%
Average prediction error for more than high school	3.13	2.70
Percent of cities within $1 \pm \frac{1}{2}\sigma_{\varepsilon_2^{HT},91}$	0.7%	1.9%
Percent of cities within $1 \pm \sigma_{\varepsilon_2^{HT},91}$	1.9%	3.8%
Panel C: Gender		
Average prediction error for males	1.64	1.39
Percent of cities within $1 \pm \frac{1}{2}\sigma_{\varepsilon_2^{HT},91}$	15.2%	22.4%
Percent of cities within $1 \pm \sigma_{\varepsilon_2^{HT},91}$	29.5%	47.0%
Average prediction error for females	1.02	0.87
Percent of cities within $1 \pm \frac{1}{2}\sigma_{\varepsilon_2^{HT},91}$	32.7%	29.8%
Percent of cities within $1 \pm \sigma_{\varepsilon_2^{HT},91}$	62.3%	60.8%

Notes: Authors' calculations using census microdata. We report the average and the percentage of cities whose errors fall within half/one standard deviation $\varepsilon_{2,91}^{HT}$ of one ("perfect fit") across cities by population subgroup. Errors in columns 1 and 2 are calculated using equations 3 and 4, respectively, and assuming $\alpha = 1/3$ in the latter. We use wages and employment of observations belonging to each subgroup and the city's average rent to calculate the error. We residualized wages and rents on the interaction of macroregion and year fixed effects before calculating the prediction errors. We trim errors at 1%. The Data Appendix further describes the precise computation of HT prediction errors.

Table B.7: Correlates of the HT Prediction Errors

	$\lvert\varepsilon_{1,i}^{HT}-1\rvert$	$\lvert\varepsilon_{1,f}^{HT}-1\rvert$	$\lvert\varepsilon_{2}^{HT}-1\rvert$
Distance urban-rural areas	-0.047***	0.006	0.127***
	(0.006)	(0.011)	(0.015)
Rural: Population density	0.001	0.001	0.010***
	(0.001)	(0.001)	(0.001)
Rural: Municipalities in the catchment area (in 100s)	-0.028***	0.060***	0.125***
	(0.006)	(0.016)	(0.018)
Rural: Log(Population)	-0.053***	0.027	0.058**
	(0.010)	(0.017)	(0.021)
Urban: Log(Population)	0.005	-0.061***	-0.056***
	(0.005)	(0.010)	(0.013)
Change in urban manufacture emp. share (t-1)	0.123*	-0.277	0.518***
	(0.061)	(0.145)	(0.150)
Change in urban service sector emp. share (t-1)	-0.399***	-0.062	0.200
	(0.090)	(0.117)	(0.172)
Change in rural agriculture sector emp. share (t-1)	0.124***	-0.079	-0.104
	(0.031)	(0.055)	(0.067)
Rural share people aged 15-39	0.012***	-0.008*	-0.043***
	(0.003)	(0.004)	(0.005)
Difference in HS shares (urban-rural)	-0.294**	-1.086***	1.633***
	(0.091)	(0.168)	(0.221)
Year 2000	-0.012	0.085***	0.033
	(0.011)	(0.017)	(0.022)
Year 2010	0.014	0.129***	0.075*
	(0.017)	(0.027)	(0.036)
Adjusted R^2	0.393	0.332	0.436
Observations	1298	1297	1298

Notes: Regressions estimated at the city level using pooled data from 1991, 2000 and 2010 census years. The outcomes refer to the absolute value of the formal, informal, and two urban sectors' prediction errors minus one. Rural variables refer to city's rural catchment areas. All regressions include macroregion fixed effects. We describe the precise computation of each variable in the Data Appendix. Robust standard errors in parenthesis. * p <0.1 , ** p < 0.05, and *** p < 0.01.

C. Data Appendix

Table B.8: Definitions

Definitions	**Description / comments**
Working age population	Individuals between 15 and 64 years old in the period of reference.
Formally employed	Individual that worked over the period of reference with a signed work card , or was an employer.
Informally employed	Individual that worked over the period of reference without a signed work card, or was self-employed.
Employed	Individual either formally or informally employed.
Non-employed	Working age individual declared as non employed.
Migrant	Individual that declares that its time of residence in their current municipality is less or equal to 5 years.
High skill	Individuals that completed at least high-school-equivalent education (2do grau, colegial o medio 2do ciclo).
Wage	Monthly labor income in main occupation in the reference period.
Rent	Montly value of housing rent.
Industry of employment	Four major industries based on CNAE - Domiciliar definition
Catchment area*	Set of rural municipalities with a positive rate of emigration to a given city. The rate of emigration uses individuals who migrated 5 to 10 years before the reference period.
Arranjos populacionales (City)	Grouping of two or more municipalities where there is a strong population integration due to commuting to work or study, or due to contiguity between the spots main urbanized areas (IBGE 2016). We use a time-consistent definition joining the arranjos that share a common municipality for the period 1970-2010 .
Urban	Individuals identified as living in the city's urban area during the census year.
Rural	Individuals identified as living in the rural area of municipalities not considered as cities during the census year.

Notes: Housing rents are not available for the 2000 census.
***** The 2000 census does not allow us to identify the municipality of the previous residence for individuals who migrated before 1995. We instead use the share of migrants between 1991-1986 (from the 1991 census) as weights for this year.

Table B.9: Variables Used in HT Error Computations

Variable	**Samples**	**Description / comments**
Average rural wage	1991, 2000, 2010	Weighted average log-wage of invididuals identified as living in rural areas in the catchment area of a city.
Minimum wage	1991, 2000, 2010	National minimum wage published by the Ministry of Labor and Employment in the reference period, in 2010 reais.
Average urban wage*	1991, 2000, 2010	Average log-wage of individuals identified as living in urban areas of a city.
Average formal urban wage	1991, 2000, 2010	Average log-wage of individuals identified as formally employed living in urban areas of a city.
Average informal urban wage	1991, 2000, 2010	Average log-wage of individuals identified as informally employed living in urban areas of a city.
Rural/minimum wage ratio	1991, 2000, 2010	Ratio between the average rural wage and minimum wage.
Rural/urban wage ratio	1991, 2000, 2010	Ratio between the average rural wage in the catchment area of a city and average the urban wage.
Rural/formal urban wage ratio	1991, 2000, 2010	Ratio between the average rural wage in the catchment area of a city and average urban formal wage.
Rural/informal urban wage ratio	1991, 2000, 2010	Ratio between the average rural wage in the catchment area of a city e and average urban informal wage.
Urban housing rent	1991, 2010	Average log-rent of households identified as living in urban areas of a city.
Rural housing rent	1991, 2010	Weighted average log-rent of households identified as living in rural areas in the catchment area of a city.
Urban employment rate	1991, 2000, 2010	Share of employees from the working age population of a city.

Notes: The Data Appendix provides a detailed explanation on the computation of urban and rural variables.
***** It is computed separately for formal and informal workers within a city. Finally, we compute a weighted average of both average values using the share of formality/informality in a given city as weights.

(Table B.9 continued)

Variable	Samples	Description / comments
Average prediction error (formal/informal)	1991, 2000, 2010	Average across cities of the value of the error defined in equation 2. For each city, the error is computed using as inputs the average rural wage in the city's catchment area, the city's average formal/informal wage and the city's formal/informal employment rate.
Average prediction error (two urban sectors)	1991, 2000, 2010	Average across cities of the value of the error defined in equation 3. For each city, the error is computed using as inputs the average rural wage in the city's catchment area, the city's weighted average between formal and informal salaries and the city's employment rate.
Average prediction error (Housing market)	1991, 2010	Average across cities of the value of the error defined in equation 4. For each city, the error is computed using as inputs the average rural wage in the city's catchment area, the city's weighted average between formal and informal salaries, the city's employment rate and, the ratio between rural rents in city's catchment area and the city's average rents. The last ratio is raised to the power 1/3
Percent of cities within $1 \pm \frac{1}{2}\sigma_{\varepsilon_2^{HT},91}$ and $1 \pm \sigma_{\varepsilon_2^{HT},91}$	1991, 2000, 2010	Share of errors lying within half/one standard deviation of ε_2^{HT} calculated for the full working-age sample in the year 1991.

Table B.10: City and Catchment Areas Characteristics

Variable	**Samples**	**Description / comments**
Distance to rural destinations	1991 , 2000 , 2010	Weighted average distance between the centroid of the most populated municipality within an urban area and the centroid of each rural municipalities in the catchment area.
Population	1991 , 2000 , 2010	Working-age population of a municipality.
Population density	1991 , 2000 , 2010	Ratio between the working-age population in a municipality and its area in square kilometers.
Number of municipalities in the catchment area	1991 , 2000 , 2010	Number of rural municipalities with positive emigration to a city in a reference period.
Agriculture/Manufacture/Services employment share	1991 , 2000 , 2010	Share of employed individuals working in manufacturing/agriculture in the reference period based on CNAE - Domiciliar definition. We use the lagged change in industry employment's shares as covariates for tables 5 and B.7.
Share of HS educated	1991 , 2000 , 2010	Share of individuals with high school or degree or higher in the working-age population a of a municipality with education information.
Share of population aged 15-39	1991 , 2000 , 2010	Share of individuals aged 15-39 from the working-age population.

Notes: The Data Appendix provides a detailed explanation on the computation of urban and rural variables.