Rural-Urban Migration at High Urbanization Levels*

Matias Busso, Juan Pablo Chauvin, and Nicolás Herrera L.

Research Department, Inter-American Development Bank

March 20, 2020

Abstract

This study assesses the empirical relevance of the Harris-Todaro (HT) model at high levels of urbanization using data from Brazil. Comparing observed and model-based predictions of the equilibrium urban population and unemployment at the city level, it finds little support 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, more so for urban population than for unemployment predictions. HT equilibrium relationships are more frequently found in cities that are, in relative terms, larger and more specialized in manufacturing, and where the associated rural area is located at relatively shorter distances, is more specialized in agriculture, and has a population demographically more prone to migrate.

Keywords: Harris-Todaro, rural-urban migration, urban unemployment, developing countries.
JEL Codes: J46, J61, O18, R23

1. Introduction

Since its original publication, the Harris-Todaro (1970) framework (henceforth HT) has been one of the key conceptual tools for the study of rural-urban migration and its connection to urban unemployment. At that time, 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 the year 2007 global urbanization crossed the 50% threshold, and in 2018 the world was 55.3% urban (World Bank, 2020). In years to come rural-urban migration will take place in increasingly urbanized low-and-middle-income countries.[1] In this environment, the HT model may lose empirical validity (e.g. because relative scarcity of rural labor renders the rural-urban wage gap negligible), or may require significant extensions to remain useful.

This paper explores the extent to which the HT framework is empirically relevant at high-levels of urbanization, and what extensions of the model may be most adequate in an increasingly urbanized developing world. Our data comes from Brazil, a country that in 1970 had an urbanization rate similar to that of the world today (55.9%), and which went on to reach rates north of 85% in the following five decades (World Bank, 2020). In addition to providing a context that resembles what other developing countries may face in upcoming decades, Brazil has the advantage of having a large number of urban centers, which differ in size, education levels, and the economic structure of their local economies. This allows us to examine at a more granular level the conditions under which the model can have more or less empirical bearing.

In order to take advantage of the rich data available, we bring the analysis to the subnational level. This requires us to select an appropriate unit of observation. Our approach consists in defining, for each of 449 cities, a rural-urban migration "catchment area" as the set of rural municipalities of origin of prior migrants. We then construct, at the city level, both urban labor market variables, calculated with observations from within the urban boundary, and their corresponding rural variables, calculated as weighted averages from rural municipalities in the city's catchment area, where the weights are the municipalities' historical migration shares.

We assess the extent to which the HT equilibrium condition (by which the rural wage equals the expected urban wage) holds in the data, starting with the original framework and subsequently considering extensions. For each version of the model, we characterize the corresponding HT equilibrium condition in terms of the urban population/unemployment that should be observed in a given city if we take the values of the other variables from the data and assume that the equilibrium holds. We then contrast the population and unemployment actually observed with these predicted values to create error measures, which we use to quantitatively assess the empirical performance of

---

[1]While high urbanization levels used to be treated almost as a measure of development, in recent decades it has become more common to see countries achieve high levels of urbanization while remaining at relatively low income levels (Glaeser 2014; Jedwab and Vollrath 2015).

the model.

We look first at the equilibrium condition in the the original HT environment in which the urban wage is the legislated minimum wage. We find that this basic model does not perform well at predicting what we observe in the data particularly in the decades in which the minimum wage is below the average rural wage (which violates one of the key assumptions of the model). We then consider urban wages observed in the market. A challenge is that, as is typical in many developing countries, urban labor markets nest both formal and informal jobs. We show that using just the formal wage or just the informal wage to compute the expected return to migration renders too many localities to appear to be off-equilibrium. When using only the formal wage, the model tends to under-estimate the actual urban population observed in the data and over-estimate the rate of unemployment. When using only the informal wage, the prediction errors reverse directions.

Next, we explore three 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 both in terms of the population and the unemployment equilibrium quantities. Second, we extend the model to accommodate heterogeneous cost of living across localities which, again, improves significantly the model's correspondence with the observed data. Third, the HT model assumes that workers are homogeneous. 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 HT condition better describes the equilibrium for the relatively less educated workers and for the younger population (who are the ones most likely to migrate).

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 distance between the city and its catchment rural area are shorter. This is in line with one of the assumptions of the HT model, namely that moving costs are low. We also find that the errors are smaller for cities better suited to absorb migration flows (those that are larger have a higher share of employment in manufacturing, in line with Henderson et al., 2017), and have in their catchment area more archetypical rural areas (those that are less densely populated and have a higher share of employment in agriculture). Prediction errors are also smaller when the rural population is younger, which the literature has identified as the segment of the population more likely to migrate to cities (Kennan and Walker, 2011), and they tend to get larger over time as urbanization progresses.

Throughout our analysis, model-based predictions are more consistent with the data of urban population than of unemployment. One of the model's key insights, that urban development policies affecting wages lead to rural-urban migration responses that may shape the policies' outcomes, 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 leads to the migration of more than one rural worker, such that

efforts to tackle urban unemployment could *worsen* it – the well-known "Todaro paradox" (Todaro, 1969).[2] We find the evidence to be less conclusive with regards to model-based unemployment predictions. We see this as a reflection of the model's parsimony in its characterization of the urban labor market, and conjecture that, in order to be more informative for unemployment policy, it would need to consider not only the urban informal sector but also the labor force participation margin.

Our paper contributes to the empirical literature on rural-urban migration by providing what we believe to be, to date, the most comprehensive test of the validity of the HT framework in a highly urbanized developing country. This literature has long recognized that rural to 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 elsewhere (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, which are surprisingly few given how influential the model has been on theory and policy, 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 in individual subnational locations, and is replicable with data that is publicly available in many developing countries. Most existing studies either use individual-level analysis, in which a migrant indicator is regressed on characteristics of places of origin and destination (e.g. Fields, 1982; Schultz, 1982), or are based in country-level statistics (e.g. Todaro, 1976; Collier, 1979; Salvatore, 1981). Our approach matches better 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 method is in essence an accounting exercise (closer in spirit to the "growth accounting" method 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 to inform policy. Prior literature has studied how rural-urban migration decisions vary with personal and place characteristics (Fields, 1982; Schultz, 1982; Lucas,

---

[2]Fields (2005) suggests that this emphasis in unemployment may be undue by showing that, 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.

[1985](). Our approach allows us to have 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 to explain rural-urban migration and urban unemployment. At the same time, it helps us gauge the extent to which we can improve the empirical traction of the model in a given time and place by incorporating the informal sector and urban costs of living. This opens the door for a more informed and context-sensitive use of the framework in policy design.

The rest of the paper proceed as follows. Section 2 describes the data and our definition of localities, and 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 extend each of these conditions hold in the data. Section 4 discusses how the predictive power of the HT framework correlates with characteristics of the cities and their rural "catchment areas". Lastly, Section 5 concludes.

2.   Descriptive facts

This section characterizes the context in which our study takes place. It starts by describing how we define cities and their rural migration catchment areas. In turn, this gives us the functional definition of rural-urban migrants that we use in most of the analysis. We then move on to describe key socioeconomic characteristics of these migrants, and their labor market performance relative to non-migrants and to their region of origin. All the variables used in the analysis are constructed from microdata of the Brazilian population census, made publicly available by the Institute of Geography and Statistics (IBGE). The data appendix provides details of the construction of each variable.

2.1.   Defining cities and their rural catchment area

In order 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 a subset of the country's rural areas. Conversely, potential migrants living in rural locations frequently have more than one possible urban destination, and the measures of (expected) urban wage and unemployment relevant for their migration decision are not obvious.

Our analysis focuses on cities and data-driven definitions of their rural catchment areas.[3] To

---

[3]A few existing papers have looked at geographies below the national level, but have not explicitly linked cities

4

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 enhance the precision of our measures, we use in the corresponding calculations only observations that are classified as "urban" or "rural" both in our definition and in the census definition. 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. This requires us to adjust our geographic units of observation (commuting zones and rural municipalities) to make them consistent across census years. We adopt the time-consistent municipalities defined by Chauvin (2018), and employ the same procedure to define time-consistent commuting zones, yielding 449 cities. The rural catchment area in the median city includes 44 rural municipalities. Appendix Figure A.1 shows the geographic distribution of these locations, and Appendix Table A.1 reports descriptive statistics of these cities and their catchment areas.

---

with the rural areas from which their migrants originate. Barnum and Sabot (1977) groups Tanzanian regions in 6 areas (3 urban and 3 rural), and estimate migration equations using as their unit of observation area–education category–time period cells. Fields (1982) uses departments' borders to divide the Colombian territory in 6 urban and 6 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]The type of area (rural or urban) that the migrant was living in prior to migrating is not consistently available across census rounds. Our approach allows us to measure 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]In 1991, 2000 and 2010 we observe the municipality of origin of migrants up to ten years prior to the census. We calculate these migration weights using data from migrants that arrived in the city between 10 and 6 years before, excluding "recent migrants" (those that arrived 5 years or less ago), who are more like to be responding to current incentives (such as wage-gaps and unemployment) in the census year.

2.2. Rural-urban migration

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 how migration evolved over five rounds of the population census. In 1980, around 18% of the working-age population declared that they had changed municipality of residence at some point in the prior 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. The lion's share of these population movements were directed towards cities. In 1970, migrants that were living in urban areas at the time of the census accounted for 70% of total migration. This share went up to 78% in 1980, and remained stable at those levels in all subsequent census years.

Figure 1: Internal migration of working-age individuals in Brazil, 1970-2010
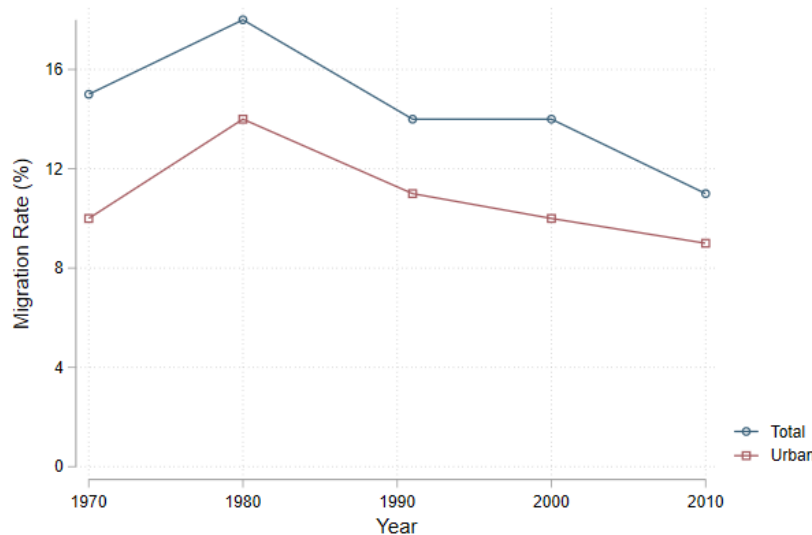
Notes: National-level estimates constructed from individual census microdata. We define migrants as working-age individuals who lived in a different municipality five years before the census. Urban migrants are those residing in cities (per the definition described in section 2.1) at the time of the census.

Table A.2 reports the origins and destinations of internal migrants in Brazil. 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. The majority of the urban population that migrate also tend to migrate to other cities usually larger than the city of origin. There is also a non-negligible fraction of the people living in urban areas that move to rural locations. This is consistent with the prevalence of return migration which could be due to an unsuccessful performance in the destination labor market (Hirvonen and Lilleør, 2015), to 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).

In spite of the country's high levels of urbanization, an important fraction of urban immigrants continue 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.[7] 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. In contrast with other developing countries, the gender split of rural-urban migrants is even.[8]

Rural-urban migrants tend to be young and relatively less educated, but as high-school enrollment grew at a fast pace and primary school became nearly universal (Busso et al., 2017), both age and education levels of migrants have increased over time. Four out of five rural-urban migrants were younger that 40 at the time of migrating. Regarding educational achievement, 44% had less than primary schooling in the 1991 census. The share went down to 24% in 2000 and 21% in 2010. In contrast, the shares with high-school education or higher for the same years, increased from 13%, to 23% and to 38%.

The statistics reported in Panel B of Table 1 describe how rural-urban migrants perform in the labor market after they have arrived to the city. Rural migrants tend to have similar or higher rates of employment than natives. They also tend to be employed in more informal jobs than the incumbent population; although the informality gap between migrants and natives has declined over time. These differences in the quality of jobs also manifest itself in wages: on average, migrants earn that are 2%-5% lower than those earned by non-migrant workers working in the same sector and in the same city of residence. This gap is virtually the same for formal and informal workers in all census years. Relative to their rural municipality of origin, migrants receive a wage premium of around 8%.[9] With this background we turn now to the central question of this paper, which concerns the empirical relevance of the Harris-Todaro model of rural-urban migration in the highly urbanized Brazilian context.

---

[7]Note that starting in this table, and for the remaining of the paper, we adopt the rural-urban definition described in section 2.1. We are unable to measure rural-urban migration using this definition prior to 1991 because of data limitations.

[8]This difference relative to the rest of the developing world is found in various 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.

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

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 24 or younger	52.8%	50.0%	46.5%
Percent 25 to 39	31.7%	32.8%	35.0%
Percent 40 or older	15.6%	17.1%	18.5%
Education*			
Percent less than primary	43.68%	23.22%	20.97%
Percent primary but less than high school	43.54%	53.84%	40.90%
Percent high school or higher	12.8%	22.9%	38.1%
Panel B: Labor market performance of rural-urban migrants			
Non-employment rate	36.8%	42.8%	35.7%
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			
All urban workers	98.4%	95.5%	97.9%
Formal urban workers	98.5%	95.5%	97.7%
Informal urban workers	98.5%	96.1%	98.2%
Wage gap (ratio) relative to rural municipality or origin			
All workers	108.6%	107.6%	107.7%
Formal urban workers	111.9%	110.7%	109.7%
Informal urban workers	104.1%	104.3%	104.1%

Notes: Own calculations using census microdata. We restrict the sample to working-age migrants (at the time of migrating) and working-age stayers. Rural-urban migrants include all individuals moving from rural municipalities to municipalities located in cities, per the definition described in section 2.1.

* 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

In this section 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 originally formulated by Harris and Todaro (1970). We then analyze two variants that release 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

We start with a simplified version of the original formulation of the HT framework. 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 because an institutionally set urban minimum wage ($\underline{W}$), which in our expression corresponds to $\tau = \underline{W} - W_U^{comp}$. 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 their (expected) wage income, which corresponds to W_R if the worker locates in a rural area, and $\frac{1}{1+\lambda} W_U$ if the worker locates in an urban area, where $\lambda = \frac{U}{L_U}$ is the incidence of unemployment (unemployment as a share of urban employment), and $\frac{1}{1+\lambda} = \frac{L_U}{L_U+U}$ is the probability of being employed. 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 viceversa. For simplicity, we assume away migration costs.

In equilibrium, the expected urban wage is equalized to the rural wage, namely:

$$W_R = \frac{1}{1+\lambda} W_U. \tag{2}$$

This condition can be rearranged, and the difference between the right and left hand sides of the equality can be used to create measures of how distant the data is from the predicted equilibrium condition. This "error" could be expressed in terms of any of the variables included in equation 2. We are interested in assessing deviations from the equilibrium in terms of the model's two key endogenous variables: migration and urban unemployment. The migratory response will determine how urban population varies with the rural-urban wage gap. Urban population is implicitly defined in equation 2 as:

$$L_U + U = U \frac{W_U}{W_U - W_R} \tag{3}$$

which makes clear that, taking the model at face value, one could predict the urban population of a location based on its rural wage, urban wage, and unemployment, as long as the location is at HT equilibrium. This suggests that we can use equation 3 to compute population predictions for each city, and treat the ratio of the observed and the predicted values $(L_U+U/\widehat{L_U+U})$ as a measure of the empirical "prediction error" of the HT equilibrium in a given location.

$$\varepsilon_{1,L_u}^{HT} = \left(1 + \frac{L_U}{U}\right)\left(1 - \frac{W_R}{W_U}\right) - 1. \tag{4}$$

The minuend term in expression 4 corresponds to the ratio between the observed and the predicted urban population. Subtracting one allows us to interpret the error in terms of excess urban population. When the ratio is larger than one the error becomes positive, implying that the observed urban population is larger than the HT model would predict based on the measured rural and urban wages and urban unemployment. When the ratio is smaller than one the error becomes negative, indicating that the observed population falls below the prediction.

Alternatively, we can express the equilibrium condition 2 in terms of excess urban unemployment –measured as a fraction of the unemployment predicted by the model– by defining the HT prediction error as:

$$\varepsilon_{1,U}^{HT} = \left(\frac{U}{L_U}\right)\left(\frac{W_U}{W_R} - 1\right)^{-1} - 1. \tag{5}$$

The structure of this error is similar to the one in equation 4, with a term related to employment and another to wages. In equation 5 the minuend term corresponds to the ratio between the observed and the predicted unemployment.

We interpret ε_{1,L_u}^{HT} and $\varepsilon_{1,U}^{HT}$ as measuring how far off a city's population and unemployment are, respectively, relative to where the HT model would predict them to be at equilibrium. This

implies that the error can reflect either mis-specification in the model,[10] 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 components. However, we make progress in understanding the role of mis-specification by empirically assessing extensions of the model that account potentially relevant variables that are not part of the basic framework.

Rural-urban wage gaps, informality, and unemployment in Brazilian cities

In order to obtain city-level measures of equations 4 and 5, we use census microdata to calculate their constituent variables for each city and its correspondent rural migration catchment area (see the data appendix for details). Table 2 reports the averages (taken across cities) of the main variables used to measure the HT equilibrium in the basic model and extensions. In addition to rural wage it 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.[11]

On the basis of this summary statistics alone, the minimum wage appears unlikely to be an important driver or 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, and even then it was only 7% smaller than the rural wage in 2010.

In contrast, the rural wage has been consistently smaller than the market urban wage. Moreover, the rural-urban wage ratio has remained strikingly constant over time at around 60%. 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 25% less than informal urban workers. Rural areas also have lower costs of living. The average rural rent represented 75% of the average urban rent in 1991, and 85% in 2010, the two census years for which this information is available.

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

---

[10]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).

[11]As is standard in the literature on Brazilian labor markets, we define formal workers as those who report having a signed work card, which grants them access to the benefits and protection of the local labor laws. Workers without a signed work card are considered informal.

measurement of key variables such as the total urban population and the expected urban wage (through the probability of being employed).

Table 2: Variables used to compute HT prediction errors

	1991	2000	2010
Average rural wage (W_R)	285.0	371.2	475.2
Minimum wage $(\underline{W})$	186.0	294.8	510.0
Average urban wage $(\delta_F W_F + \delta_I W_I)$	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 $(W_R/\underline{W})$	153.3%	125.9%	93.2%
Rural / urban wage ratio $(W_R/\bar{w}_U)$	57.5%	58.0%	60.8%
Rural / formal urban wage ratio $(W_R/\underline{W})$	49.8%	47.6%	51.8%
Rural / informal urban wage ratio (W_R/W_I)	71.0%	73.4%	76.1%
Average urban housing rent (R_U)	199.6	N/A	253.8
Average rural housing rent (R_R)	150.4	N/A	214.1
Urban non-employed / employed ratio (λ)	69.4%	80.5%	62.4%
Urban informality rate (δ_I)	44.8%	50.1%	43.3%

Notes: The table reports the average of the variables taken across cities. All monetary values are expressed in 2010 Reais. We describe the precise computation of each variable in the data appendix.

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 over estimate unemployment, in the sense that the risk of *involuntary non-employment* may not be as large as our measure suggests in the eyes of potential migrants. If labor supply is upward slopping among urban dwellers, such mismeasurement would be negatively correlated with the urban-rural wage gap.[12] We take note of this issue in the analysis of our empirical results, and consider how it may affect their interpretation.[13]

The averages of the non-employed/employed ratios and the informality rates in urban areas are reported at the bottom of Table 2. Non-employed and informal workers represent a large share of the urban population in Brazil. In 2010, there were as many as 62 non-employed for each 100

---

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

[13]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 in multiple cities urban unemployment is close to zero, the urban population predictions are unrealistically low. 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.

employed workers in the average Brazilian city, and among those who did have a job, 43% worked in the informal sector. Both non-employment and informality had increased significantly between 1991 and 2000, before coming back down 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.

We start with the original Harris-Todaro assumption in which this wage is institutionally set, and estimate the prediction errors using the official minimum wage ($W_U = \underline{W}$). Panel A in Table 3 reports averages of these city-level errors, as well as the fraction of cities for which the predicted error is within a 10% neighborhood around zero. By these measures, the model performs dismally at predicting what we observe in the real world, particularly in 1991 and 2000. This is not surprising given that, as discussed before, the average rural wage was larger than the minimum wage in those years, violating one of the key assumptions of the model. The gaps were so large that, for the model's equilibrium to hold in the case of many cities, negative values of urban populations and unemployment would have been needed. Observed populations are, of course, positive, yielding errors that are between 144% and 316% the size of the negative benchmark.

Since the minimum wage appears to be non-binding and to have little bearing in rural-urban migration decisions in Brazil during this period, we turn to considering market-determined urban wages. An important challenge with this approach is that, as in most other developing countries, formal and informal labor markets coexists in Brazilian cities. The minimum wage and other regulations that effectively increase income are enforced in the formal sector but not in the informal. 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. In contrast, if finding informal jobs is relatively easier and the urban informal wage still represents an improvement relative to the rural wage, 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 next two panels of Table 3. Here we report calculations of the errors in equations 4 and 5 using, alternatively, the formal wage ($W_U = W_F$) in Panel B, and the informal wage ($W_U = W_I$) in panel C.

When we use the formal sector to compute the reference urban wage (Table 3, Panel B), the observed population tends to exceed the one predicted by the HT equilibrium. On average, it is 23% higher in 1991, 20% higher in 2000, and 31% higher in 2010. In spite of this average error, a significant number of cities were not far from the HT equilibrium prediction. In 1991, 19% of cities had an error smaller than 10% in absolute value. The fraction increased to 22% in 2000, before it went back down to reach 17% in 2010.

13

Table 3: Empirical deviations from the Harris-Todaro equilibrium

	1991	2000	2010
Panel A: Using the minimum wage ($W_U = \underline{W}$)			
Urban population prediction error (ε_{1,L_u}^{HT})	-3.16	-2.17	-1.44
Percent of cities within 0.1	0.2%	0.0%	7.6%
Urban unemployment prediction error ($\varepsilon_{1,U}^{HT}$)	-2.43	-1.70	-0.91
Percent of cities within 0.1	0.2%	0.4%	10.9%
Panel B: Using the urban formal wage ($W_U = W_F$)			
Urban population prediction error (ε_{1,L_u}^{HT})	0.23	0.20	0.31
Percent of cities within 0.1	18.5%	21.6%	16.5%
Urban unemployment prediction error ($\varepsilon_{1,U}^{HT}$)	-0.24	-0.16	-0.28
Percent of cities within 0.1	8.7%	12.0%	8.9%
Panel C: Using the urban informal wage ($W_U = W_I$)			
Urban population prediction error (ε_{1,L_u}^{HT})	-0.36	-0.43	-0.36
Percent of cities within 0.1	10.7%	9.1%	11.6%
Urban unemployment prediction error ($\varepsilon_{1,U}^{HT}$)	0.67	0.83	0.24
Percent of cities within 0.1	6.2%	5.3%	8.9%
Panel D: Model with two urban sectors			
Urban population prediction error (ε_{2,L_U}^{HT})	0.02	-0.05	0.08
Percent of cities within 0.1	20.49%	22.27%	19.38%
Urban unemployment prediction error ($\varepsilon_{2,U}^{HT}$)	0.42	0.40	0.18
Percent of cities within 0.1	12.5%	12.9%	9.6%
Panel E: Model with housing market			
Urban population prediction error (ε_{3,L_U}^{HT})	-0.05	N/A	-0.002
Percent of cities within 0.1	25.61%	N/A	23.83%
Urban unemployment prediction error ($\varepsilon_{3,U}^{HT}$)	0.34	N/A	0.32
Percent of cities within 0.1	18.71%	N/A	15.81%

Notes: The errors in Panels A through D are computed using equations 4 and 5 with alternative wage measures. Errors in Panel E are calculated using equations 9 and 10, and assuming $\alpha = 0.33$. We describe the precise computation of each variable in the data appendix.

Using the formal wage as the reference implies a large urban-rural nominal wage gap, which translates into strong incentives for rural workers to migrate to cities. If we only considered this partial effect, we might expect negative errors to be more frequent than positive – that is, that observed populations tend to fall below the predictions. Yet we find that urban population tends to

be larger than the model predicts. This is because, while in the model migration leads to increases in urban unemployment that act as strong disincentives to further migration, in practice the "urban unemployment" term does not raise by as much as the model requires to achieve equilibrium.[14]

This pattern is apparent in the estimates of the unemployment prediction error, $\varepsilon_{1,U}^{HT}$. In 1991, unemployment was on average 24% smaller than the model's equilibrium would have required given wages and population. In 2000, when the population error performed relatively better, the unemployment error was also closer to zero, and in 2010, when the model's population prediction deteriorated, so did its unemployment prediction. This may be, at least partially, the result of assuming away the urban informal sector in this specification. In practice, rural-urban migrants that fail to obtain a formal-sector job in the city do not necessarily face unemployment, and the perspective of working in the informal sector –with wages frequently larger than in rural areas– may be more of an incentive than a deterrent for many potential migrants.

What happens if, instead, we take the urban informal wage as the reference? Errors computed under this assumption are reported in Panel C of Table 3.[15] In this case, on average, the observed population tends to be smaller that the model's predictions, by 36% in 1991 and 2010, and by 43% in 2000. In the best-performance year (2010), 12 out of 100 cities had an error smaller than 10% in absolute value. As in Panel B, the unemployment prediction errors tend to go in the other direction: while the urban population is on average smaller than predicted, the unemployed population tends to be larger than the prediction.

In this case, the size and direction of the errors may be particularly sensitive to the limitations of our measurement approach. As equation 3 shows, if we observe a small urban-rural nominal wage gap along with sizable urban unemployment, a large urban population is needed for the HT equilibrium condition to be satisfied. The population "prediction" is more sensitive to unemployment the smaller the wage gap is. At the same time, stepping outside of the model to consider the urban labor force participation margin under the assumption of upward-sloping labor supply, note that a lower reference wage implies that non-participants become a relatively larger fraction of the non-employed. Therefore, when the informal wage is used as the reference urban wage, our choice to include non-participants in the unemployment measure may introduce a positive measurement bias in the unemployment error, and a negative bias in the population error.

Having explored how the HT framework performs empirically if we treat either the formal wage or the informal wage as the reference for urban areas, we turn to considering both urban sectors at the same time.

---

[14]This is the case even though to measure U in equation 4 we use non-employment, which may over-estimate the disincentives to migrate as previously discussed.

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

3.2. Model extensions

This section moves beyond the basic model to explores two extensions. The first introduces simultaneously 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 urban costs of living that migration can bring about.

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 locating in cities can be either formally employed or informally employed (e.g. Brueckner and Zenou, 1999). In contrast, our approach is to incorporate the informal sector into the model but keep 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 assume that they are fully segmented. In the spirit of the original HT formulation, workers randomly get 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:

$$E\left[W_U\right] = \frac{\delta_F}{1+\lambda}W_F + \frac{\delta_I}{1+\lambda}W_I \tag{6}$$

where W_F is the average urban formal wage, W_I the average urban informal wage, the incidence of unemployment continues to be $\lambda = \frac{U}{L_U}$ but now with $L_U = L_F + L_I$, and we define $\delta_F = \frac{L_F}{L_U}$ and $\delta_I = \frac{L_I}{L_U}$ as the urban formality and informality rates, respectively.[16]

The location decision rule remains the same as before (i.e. choose the location with the higher expected labor income), implying that the HT equilibrium is given by $W_R = \frac{\delta_F}{1+\lambda}W_F + \frac{\delta_I}{1+\lambda}W_I$. Under these assumptions, the HT prediction errors in terms of urban population and unemployment, ε_{2,L_u}^{HT} and $\varepsilon_{2,U}^{HT}$, are defined as in equations 4 and 5, respectively, replacing W_U with $\frac{\delta_F}{1+\lambda}W_F + \frac{\delta_I}{1+\lambda}W_I$.

When the formal and informal wages are considered simultaneously the model's empirical traction improves significantly, particularly regarding the urban population prediction errors. Panel D

---

[16]Note that the formality and informality rates are defined as a share of total employment, and are not the same as the probability of being employed in sector $S = \{F, I\}$, defined as $\frac{\delta_S}{1+\lambda} = \frac{L_S}{L_U + U}$.

of Table 3 reports the average of errors ε_{2,L_u}^{HT} and $\varepsilon_{2,U}^{HT}$ across cities. The cross-city average of the population error is close to zero in all three years. The observed population was 2% larger than the model-based prediction in 1991, and 8% larger in 2010. In the year 2000, it was 5% *smaller* than the prediction.[17] The fraction of cities with an error smaller than 10% in absolute value was 19% in the worst-performing census year (2010) and 22% in the best (2000).

The unemployment prediction error also improves relative to the estimation that uses the informal wage (Panel C). Looking back at equation 3 this is to be expected, given that the urban-rural wage gap is larger when considering the average rather than the informal wage. However the measured unemployment continues to be larger than the model-based prediction in all specifications, going from 42% larger in 1991 to 18% larger in 2010, which suggests that the measurement issues discussed in section 3.1 may still be at play. Relative to the estimation that uses the formal wage (Panel B), the improvements are ambiguous. The average performance (in absolute value) is worse in two of the three years, but a larger share of cities have an error within $\pm 10\%$.

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 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.[18]

Production and labor demand continues to have three sectors as before. On the labor 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 $A = \{U, R\}$. They

---

[17]With average errors near zero, changes in sign across different draws are not that surprising, particularly if our measure has a random component that is symmetrically distributed and centered around zero.

[18]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 is driven by lower costs of housing in the outskirts. They also find that this is correlated with a larger decentralization of informal jobs.

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 the expected wage is defined as in equation 6. It follow that housing demand at their location of choice will be given by:

$$H_A = \alpha \frac{E\left[W_A\right]}{R_A} \tag{7}$$

where $E\left[W_R\right] = W_R$. Total housing demand in each area will be 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

$$V_A = \alpha_1 E\left[W_A\right] R_A^{-\alpha} \tag{8}$$

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 where their *expected utility* is higher. The equilibrium condition is therefore:

$$\alpha_1 W_R R_R^{-\alpha} = \alpha_1 \left(\frac{\delta_F}{1+\lambda} W_F + \frac{\delta_I}{1+\lambda} W_I \right) R_U^{-\alpha}$$

which corresponds to the following prediction errors, expressed respectively in terms of urban population and unemployment:

$$\varepsilon_{3,L_u}^{HT} = \left(1 + \frac{L_U}{U}\right) \left(1 - \frac{W_R}{\delta_F W_F + \delta_I W_I} \left(\frac{R_U}{R_R}\right)^\alpha \right) - 1 \tag{9}$$

$$\varepsilon_{3,U}^{HT} = \left(\frac{U}{L_U}\right) \left(\frac{\delta W_F + \delta_I W_I}{W_R} \left(\frac{R_R}{R_U}\right)^\alpha - 1 \right)^{-1} - 1 \tag{10}$$

The structure of the errors remain the same as in equations 4 and 5. The first term of the minuend is unchanged, but in the term related to prices now incorporates, in addition to the rural/urban wage ratios, the urban/rural housing rents ratio weighted by the housing share in consumption α. Panel E of Table 3 evaluates the errors in equations 9 and 10 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 $\alpha = 0.33$.

Including urban costs of living in the model significantly improves it's performance in the data. The average population prediction error is 5% below zero in 1991, and only 0.2% below in 2010. In both years, about a quarter of all cities have errors smaller than 10% in absolute value. Regarding the unemployment prediction error also performs better than in the model without a housing market. In this case, the average improves in 1991 and deteriorates in 2010, but in both years the fraction of cities with errors within $\pm 10\%$ increase, reaching 19% and 16%, respectively.

Taking stock

The analysis above makes it clear that even the most basic formulation of the HT model finds some support in the Brazilian data, as long as we use some market-determined urban wage in the computation. Moreover, when we expand the basic framework beyond the single urban sector assumption the empirical performance of the model improves significantly.

This can be seen more clearly in Figure 2, which shows the distribution of the prediction errors in the range -2 to 2 for alternative models, pooling observations from all census years.[19] When we treat the formal wage as the reference urban wage, the model systematically over-predicts urban population and under-predicts unemployment. If instead we use the informal wage, the errors go in the opposite direction and have a higher variance. The best-performing reference urban wage is a linear combination of wages in the formal and informal sector. Using this measure, the center of the errors distribution (particularly in the case of the population error) moves much closer to zero. And if we also account for differences in urban costs of living as proxied by housing rents, the tightness of the distribution around zero meaningfully increases.

Even in the best-performing extension, the model's equilibrium condition is noticeably more useful to predict urban population than to predict unemployment. This may be partly explained by our choice of counting non-participants as part of our unemployment measure, as previously discussed. However, while the existing empirical papers that have tested predictions inspired in the HT framework have found empirical support for the migration/population predictions more often than not, this has not been the case for the unemployment predictions and the related "Todaro Paradox" (Lall et al., 2006).[20] This suggests that at least part of the unemployment prediction errors that we observe may indeed reflect specification issues. In particular, by assuming away the participation margin, the HT framework may be too parsimonious in its treatment of the urban labor markets to be able to predict the response of unemployment to shifts in the other variables of the model.

---

[19]We don't include the model that uses the minimum wage in the figure since we have previously shown that it has almost no traction in the data.

[20]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, but in the latter this only holds for males with at least secondary education, and not for workers with lower schooling levels.

Figure 2: Distributions of the HT prediction error under alternative models

Urban Population

Unemployment
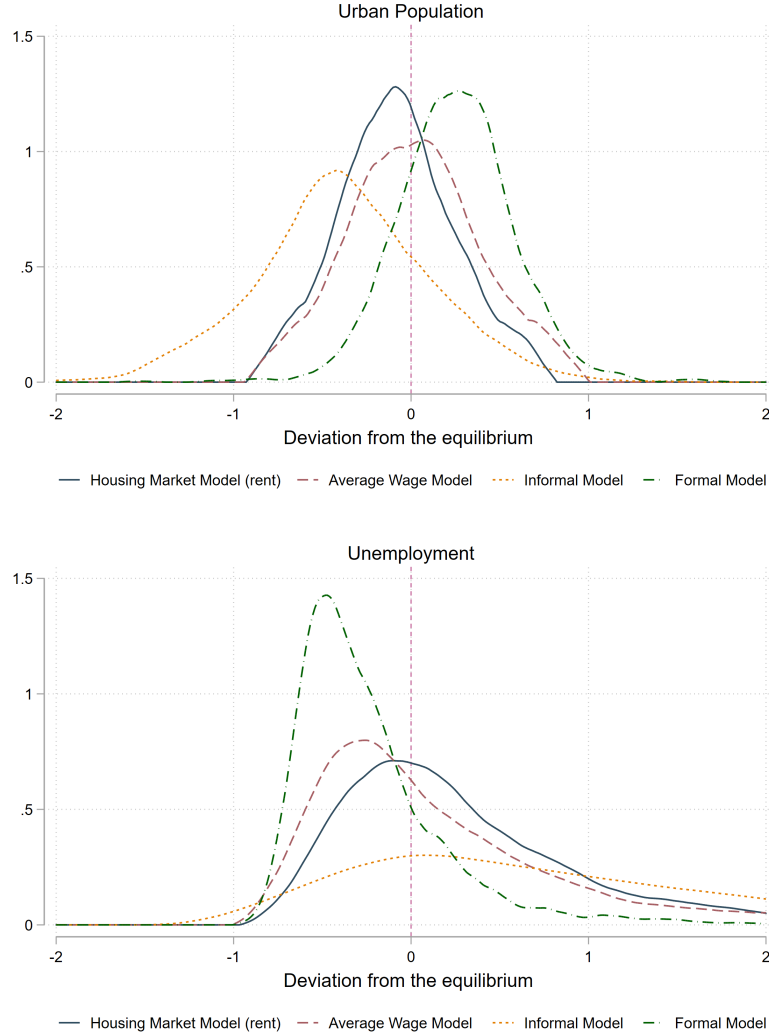
Notes: Kernel density estimates of the HT prediction errors' distributions at the city level (Panel B, C, D and E of 3). Pooled data from all census years with the required data. Average wage model corresponds to the HT prediction error using both formal and informal sectors. We restrict the plot to the interval -2,2 where most of the errors' distribution lie.

3.3. 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 sub-samples of the population.

Figure 3 shows the results of these computations. In each case included in the figure we restrict

the microdata to the correspondent subpopulation and re-compute wages, population, unemployment, housing rents, and ultimately the errors in equations 9 and 10. These correspond to the model that includes the housing market (Panel E of Table 3), which is the best-performer in the data. As before, we report the average error across cities and the fraction of cities with errors smaller than 10% in absolute value, this time pooling observations of both census years for which there is rents data available. As a reference, we include previously-reported results for the whole sample.

Figure 3: HT prediction errors calculated for population subgroups
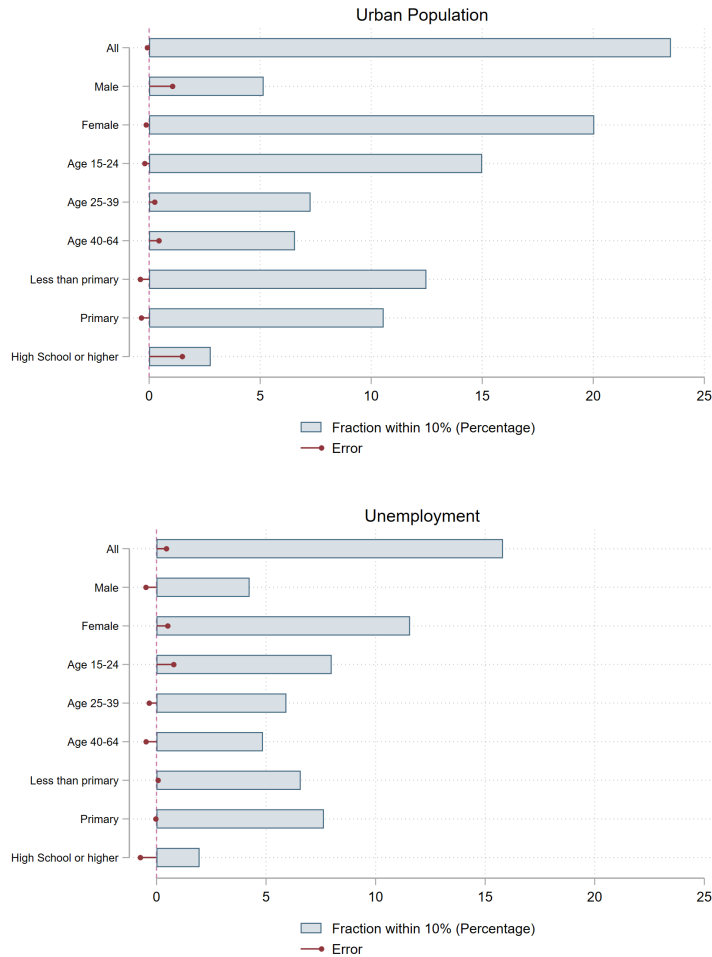
In all cases, restricting our HT equilibrium measures to sub-samples of the working-age population *worsens* the empirical performance of the model relative to the full-sample specification. Within a given demographic or human capital attribute, errors of different subgroups tend to be of different sign, so that when the groups are considered together the resulting errors are closer to

zero.

Take for example gender. Figure 3 reports errors calculated using data of either only males or only females. Similarly to what we found when we considered the formal vis-à-vis the informal wage in sub-section 3.1, using the model we under-estimate the urban population for the empirically larger urban wage (that of males), and over-estimate it for the smaller. As before, population and unemployment prediction errors go in opposite directions. Observed unemployment is smaller than expected in the male sample, and larger than expected in the female sample, which may well reflect the relatively larger labor force participation rates of the former. These patterns also resonate with prior work showing that migration and labor market outcomes of males and females across Brazilian localities are consistent with a joint optimization of couples in a context where 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, and assuming this away may affect the empirical traction from the HT framework (Lall et al., 2006; Gemici, 2011).

We obtain similar findings when we split the sample by age and schooling groups. Relative to the model-driven predictions, the observed urban population is larger among the relatively more educated (at least high school) and older (25 to 64) groups, and smaller among the others. Population and unemployment errors also tend to be of different signs. The only slight deviation from the pattern are the errors corresponding to the group with primary but no high-school education, for whom both population and the unemployment are below the prediction, though the latter is very close to zero. Overall, the errors tend to be smaller among the groups with lower educational attainment. This contrasts with the findings of Schultz (1982), who observes migration elasticities consistent with the HT framework only for individuals of with at least high-school education in Venezuela in the 1960s. These differences are consistent with the sustained growth of primary education enrollment in the region (Busso et al., 2017), which has made relatively scarcer the previously abundant uneducated labor force. As the rural-urban wage gap shrinks for this sub-population, their migration decisions may become more influenced by other considerations such as risk of unemployment.

Still, the most salient result in Figure 3 remains the fact that the model-based predictions perform better in the full sample than in any of the subsamples. Even when the errors average is closer to zero –as is the case of the unemployment error among the lower-schooling subgroups– they still have a larger variance, which translates into a smaller fraction of cities having errors within 10%. We interpret this as reflecting the paucity of the framework. Even if HT captured the key forces behind the rural-urban migration process, its parsimonious representation of the labor market would make it likely to mis-specify the channels affecting specific subpopulations. When considering the working-age population as a whole, specification errors may offset each other, improving the empirical traction of the model.[21]

---

[21]This is similar to the logic described in Kolesár et al. (2015) in the context of multiple invalid instrumental

4. HT equilibrium and location characteristics

This section aims to shed some light on the contexts in which the HT model equilibrium is most likely to be observed. To understand which location characteristics are associated with HT prediction errors that are closer to zero, we take the absolute value of the prediction error for urban population ε_{t,L_U}^{HT} and for urban unemployment $\varepsilon_{t,U}^{HT}$ in each time period $t = 1991, 2000, 2010$. For simplicity we focus on the two models that perform best in the data: the average wage model and the housing market model. Our empirical specification can be written as:

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

where C_{jt-10} is a vector of variables for city j observed in the previous census year that capture a set of characteristics from the urban zone and its catchment area, μ_{year} are census year fixed-effects that allow us to evaluate how the model's performance varies over time. We estimate this regression pooling the data from the three census rounds together, and include as controls indicator variables for five macro-regions (μ_{region}).

Table 4 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). In other words, the cost of migration is low. To approximate this, 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[22] and of each rural municipality that contributed migrants in the past (with the weights being the share of migrants coming from each municipality). The negative effect of distance on migration has consistently been found in multiples studies (e.g. Barnum and Sabot, 1977; Fields, 1982; Schultz, 1982). Consistently with the assumptions of the model, we find that the HT prediction errors tend to be larger for those urban areas whose rural areas tend to be further away (where migration costs are likely higher).

Secondly, we find that the prediction errors are smaller for more archetypical rural areas. Rural areas are usually defined by statistical agencies as those regions that have a low population density and are located outside cities. In our case, we consider the rural area associated to each city as all those rural localities from which a city received migrants in the recent past. Our catchment areas are a collection of smaller (possibly dispersed) localities that could include a small town which itself attracts 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 a measure of the population density of the catchment rural area and find that the HT prediction errors are larger for cities with a more dense populated catchment rural area.

---

variables.

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

We also include the log of the number of rural localities that are part of the catchment areas: the prediction errors are larger in cities whose catchment areas are composed by more localities.

Table 4: Correlates of the HT prediction errors

	Two urban sectors		+ Housing market	
	Pop. $(\varepsilon_{2,L_U}^{HT})$ (1)	Unemp. $(\varepsilon_{2,U}^{HT})$ (2)	Pop. $(\varepsilon_{3,L_U}^{HT})$ (3)	Unemp. $(\varepsilon_{3,U}^{HT})$ (4)
Panel A: Rural catchment area characteristics				
Distance	0.055***	0.002	0.036***	0.048
	(0.010)	(0.037)	(0.011)	(0.048)
Population density	0.005***	0.012*	0.001	0.002
	(0.001)	(0.007)	(0.001)	(0.008)
Municipalities in the catchment area (log)	0.029**	-0.006	-0.007	-0.071
	(0.012)	(0.048)	(0.014)	(0.064)
Log Population	-0.014	-0.168*	-0.019	-0.245**
	(0.021)	(0.093)	(0.023)	(0.114)
Share of HS educated	-0.012**	0.001	0.000	0.014
	(0.005)	(0.016)	(0.005)	(0.018)
Share of population aged 15-39	-0.007**	-0.047***	-0.008*	-0.036*
	(0.003)	(0.015)	(0.004)	(0.021)
Agriculture employment share	-0.001	-0.019***	-0.002	-0.020***
	(0.001)	(0.005)	(0.001)	(0.006)
Panel B: Urban area characteristics				
Log Population	-0.043***	-0.124***	-0.018	-0.042
	(0.011)	(0.043)	(0.012)	(0.051)
Share of HS educated	0.003*	-0.030***	-0.003	-0.037***
	(0.002)	(0.006)	(0.002)	(0.007)
Manufacturing employment share	0.006***	-0.012***	0.002*	-0.019***
	(0.001)	(0.003)	(0.001)	(0.003)
Panel C: Time trend				
Year 2000	-0.021	0.238***		
	(0.017)	(0.071)		
Year 2010	0.039	0.426***	0.031	0.341**
	(0.028)	(0.119)	(0.030)	(0.144)
Adjusted R^2	0.104	0.142	0.0302	0.137
Observations	1,347	1,347	898	898

Notes: Regressions estimated at the city level using pooled data from all census years with available information. 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 *** $p < 0.01$.

Third, for the HT model to work there should be a large urban area that can attract enough potential migrants by providing higher expected utility. Henderson et al. (2017) have documented that city-level characteristics such as manufacturing share affect the extent to which urban areas can absorb inflows of rural migrants, although not explicitly connecting this to the HT framework.

We find that the HT prediction error is indeed smaller for larger cities that have a more educated labor force and a larger share of employment in the manufacturing sectors.

Fourth, a feature of the HT model is that workers in the rural area work in agriculture at a fairly low wage. We find that cities whose catchment rural areas have a higher share of their employment in agriculture have a smaller prediction error. Similarly, the literature has consistently found that young adults are more likely to migrate that 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). Consistently with this empirical regularity, we find that rural areas with a lower share of the population that are 15-39 years old have a smaller prediction error.

Finally, we report the census year fixed effects. They show that the fit of the model was better 30 years ago than in the last available census. This is consistent with our conjecture that as the country urbanization increased, the fit of the HT model worsen.

5. Conclusions

Our findings show that, even at high levels of urbanization, the key insights of the HT framework can remain empirically relevant as long as features of the original formulation are adapted to include some of the extensions proposed in the literature. As the population becomes more educated, the minimum wage becomes less binding and thus less relevant for the formation of expectations among potential rural-urban migrants. The observed average urban wage does appear to factor in the migrants' decisions. This includes the wage in the urban informal sector, which seems to be seen more as a source of (second-best) economic opportunities than as a potential risk akin to unemployment. Urban costs of living also play a role, counter-acting the pull effects of high city wages.

The insights of the HT model apply more to some localities than others. Although the empirical traction of the framework did decrease on average as Brazil continued to urbanize we find that, in 1/4 of the locations considered, the population remains within 10% of what would be expected if they were at a HT equilibrium. These were locations that resembled the environment than Harris and Todaro likely had in mind. One in which, on average, the rural areas were dispersed, had low population density, and where cities were large enough to attract potential workers and had rural agricultural economies with sizeable young-adults populations located at relatively short distances.

In terms of policy, our findings support the 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. However, even though the prevalence of urban unemployment was a

key motivation for the authors and the possible existence of a "Todaro paradox" has been present in policy discussions for decades, we find the framework to be relatively less useful to think about urban unemployment reduction policies. The paucity of the model, which is one of its most attractive features, also limits its ability to capture some of the key forces shaping unemployment.

Moving forward, the classic HT framework can be enriched to further expand its empirical relevance in highly urbanized environments. Our research suggests that one of the most consequential extensions would be to incorporate labor force participation decisions of individuals and families in contexts where informal jobs are widely available. Furthermore, in the absence of a binding minimum wage, alternative micro-foundations for the urban-rural wage gap appear to be in order. These could leverage some of the lessons developed by the urban literature in recent decades. Agglomeration economies, for example, 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.

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.

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.

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.

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.

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.

Henderson, J. V., Wang, H. G., 2007. Urbanization and city growth: The role of institutions. Regional Science and Urban Economics 37, 283–313.

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.

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.

Kolesár, M., Chetty, R., Friedman, J., Glaeser, E., Imbens, G. W., 2015. Identification and Inference With Many Invalid Instruments. Journal of Business & Economic Statistics 33 (4), 474–484.

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.

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. Figures

Figure A.1: Commuting zones in Brazil, time-consistent borders for 1970-2010
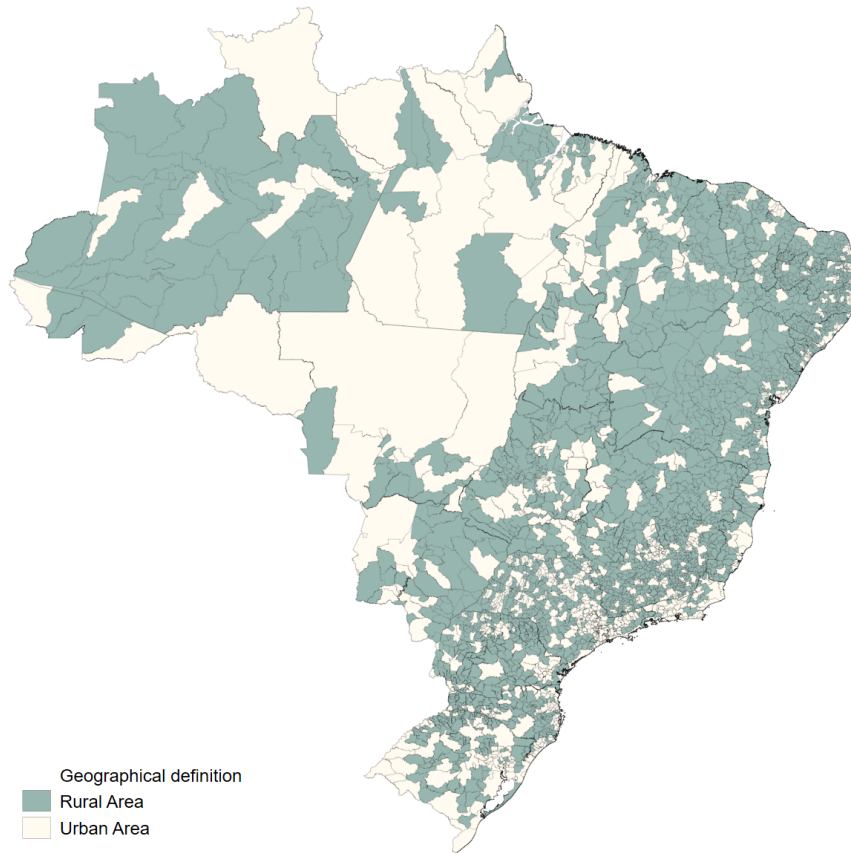
Notes: Geographic distribution of time-consistent urban and rural areas as described in section 2.1. The Amazonian states of Rondônia and Roraima contain only one time-consistent commuting zone each, and represent less than 1% of the population of the country in 2010.

B. Tables

Table A.1: Summary statistics of cities and their rural migration catchment areas

	Mean	Median	S.D.	Min.	Max.
Panel A: Urban					
Population (1000s)	224.30	72.16	924.55	1.97	17831.65
Share of HS educated	13.67	12.01	8.78	0.15	41.98
Agriculture employment share	17.36	14.00	13.02	0.41	80.21
Manufacturing employment share	9.08	6.91	7.39	0.43	64.73
Services employment share	32.37	33.72	9.62	5.10	55.26
Share of population aged 15-39	41.53	42.05	2.99	30.32	49.31
Panel B: Rural					
Population (1000s)	25.15	23.63	10.19	5.68	131.07
Share of HS educated	4.84	4.33	3.02	0.28	14.24
Agriculture employment share	33.15	32.81	7.89	12.87	64.82
Manufacturing employment share	5.04	4.54	2.49	1.17	16.60
Services employment share	21.84	21.98	5.59	6.39	35.92
Share of population aged 15-39	39.41	40.07	2.43	31.12	43.59
Distance to rural destinations (100 km)	2.38	1.58	2.28	0.17	19.31
Number of municipalities in catchment area	80.28	44.00	136.89	2.00	2171.00
Distance between rural destinations (100 km)	4.17	3.43	2.72	0.26	22.62
Population density (km^2)	8.92	8.02	5.25	1.10	30.26

Notes: The table reports summary statistics for urban and rural areas, as described in section 2.1. Rural values represent the weighted average over the catchment area of a city. Pooled data from all census years with available information. We describe the precise computation of each variable in the data appendix.

Table A.2: Internal migrants' origins and destinations

Origins	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%

The header "Destinations" spans the Rural and Urban columns.

Notes: The table reports origins and destinations of internal migrants in Brazil. We describe the geographic areas used in section 2.1. Shares represent the ratio between the number of immigrants to a rural/urban destination over the total amount of emigrants from a rural/urban origin during the 5 years before the period of reference. We compute the population quantiles for cities only.

C. Data Appendix

Table A.3: 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*	Monthly 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 populacional	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	Individual living inside an arranjo populacional and identified as living in an urban area in the census of the reference period.
Rural	Individual living outside an arranjo populacional and identified as living in a rural area in the census of the reference period.

* Housing rents are not available for the 2000 census. We discarded the upper 1% extreme values to compute the average values of this variable.
** The 2000 census does not allow to identify the municipality of the previous residence for individuals who migrated before 1995. We use instead, the share of migrants between 1991-1986 (from the 1991 census) as weights for this year.

Table A.4: Variables used in HT error computations

Variable	Samples	Description / comments
Average rural wage*	1991, 2000, 2010	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	Average across cities of the ratio between the average rural wage and minimum wage.
Rural-urban wage ratio	1991, 2000, 2010	Average across cities of the ratio between the average rural wage and average urban wage.
Rural-formal urban wage ratio	1991, 2000, 2010	Average across cities of the ratio between the average rural wage and average urban formal wage.
Rural-informal urban wage ratio	1991, 2000, 2010	Average across cities of the ratio between the average rural wage 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	Average log-rent of households identified as living in rural areas in the catchment area of a city.
Urban non-employed / employed	1991, 2000, 2010	Average across cities of the ratio between the number of the ratio between non-employed and employed individuals.
Informality rate	1991, 2000, 2010	Share of informally employed in a city.

* For rural areas, we compute the average value for a location; subsequently, we compute the weighted average of the rural areas within the catchment area of a city. We used the share of emigrants from each rural municipality to an urban area as weights. Migration shares represent the total number of individuals who migrated from a rural area to an urban area between 10 and 5 years previous to the census year over the total number of immigrants of the urban area.

** 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.

Variable	Samples	Description / comments
Urban population prediction error (minimum wage /formal/informal/two urban sectors)***	1991, 2000, 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 ratio between employed and non-employed working age-urban population, and the ratio between the average rural wage and the average minimum/urban formal/urban informal/urban wage.
Urban unemployment prediction error (minimum wage /formal/informal/two urban sectors)***	1991, 2000, 2010	Average across cities of the value of the error defined in equation 5. For each city, the error is computed using as inputs the ratio between non-employed and employed working age-urban population and the ratio between the average minimum/urban formal/urban informal/urban wage and the average rural wage.
Urban population prediction error with housing market***	1991, 2010	Average across cities of the value of the error defined in equation 9. For each city, the error is computed using as inputs the ratio between employed and non-employed working age-urban population, the ratio between the average rural wage and the average urban wage, and the ratio between rural and urban housing rent to the power of 0.33.
Urban unemployment prediction error with housing market***	1991, 2010	Average across cities of the value of the error defined in equation 10. For each city, the error is computed using as inputs the ratio between non-employed and employed working age-urban population, the ratio between the average urban wage and the average rural wage, and the ratio between rural and urban housing rent to the power of 0.33.
Percent of cities within 0.1		Share of errors lying in the interval -0.1 to 0.1 for cities in a reference period.

*** To avoid extreme values, errors are winzorized at 1% and 99%.

Table A.5: City and catchment areas characteristics

Variable	Samples	Description / comments
Distance*		Average distance between the centroid of the most populated municipality within an urban area and the centroid of each rural municipality in the catchment area.
Population density*		Ratio between the working-age population in a municipality and its area in square kilometers.
Rural areas in the catchment area		Number of rural municipalities with positive emigration to a city in a reference period.
Population*		Working-age population of a municipality.
Share of HS educated*		Share of high skilled individuals in the working-age population a of a municipality with education information.
Share of population aged 15-39*		Share of individuals aged 15-39 from the working-age population.
Agriculture/Manufacture employment share*		Share of employed individuals working in manufacturing/agriculture in the reference period.

* For rural areas, we compute the average value for a location; subsequently, we compute the weighted average of the rural areas within the catchment area of a city. We used the share of emigrants from each rural municipality to an urban area as weights. Migration shares represent the total number of individuals who migrated from a rural area to an urban area between 10 and 5 years previous to the census year over the total number of immigrants of the urban area.