Rural-Urban Migration and Market Integration*

Dennis Egger[†], Benjamin Faber[‡], Ming Li[§] and Wei Lin[¶]

July 2025

Abstract

We combine a new collection of microdata from China with a natural policy experiment to investigate the extent to which reductions in rural-urban migration barriers affect flows of trade and investments between cities and the countryside. We find that increases in worker eligibility for urban residence registration (Hukou) across origin-destination pairs increase rural-urban exports, imports, capital inflows and outflows, both in terms of bilateral transaction values and the number of unique buyer-seller matches. To quantify the implications at the regional level, we interpret these estimates through the lens of a spatial equilibrium model in which migrants can reduce buyer-seller matching frictions. We find that a 10% increase in a rural county's migration market access on average leads to a 1.5% increase in the county's trade market access and a 2% increase in investment market access. In the context of China's recent Hukou reforms, we find that these knock-on effects on market integration were on average larger among the urban destinations compared to the rural origins, reinforcing incentives for rural-urban migration.

Keywords: Economic development, rural-urban migration, trade integration
JEL Classification: F63, O12, R13

*This research was supported by funding from the Sloan Foundation and the Weiss Foundation. John Kadlick, Jianmeng Lyu, Gerard Martin-Escofet and Wenzhang Zhang provided excellent research assistance. We thank audiences at several seminars and conferences for helpful comments and Zibin Huang, Sam Kortum and Meredith Startz for insightful discussions. Li gratefully acknowledges the National Natural Science Foundation of China (No. 72403216 and 72192804) and Guangdong Basic and Applied Basic Research Foundation (No. 2022B1515120060). Lin gratefully acknowledges the National Natural Science Foundation of China grant No. 72303200, and the Natural Science Foundation of Guangdong grant No. 2024A1515011279. Errors are our own.

†Department of Economics, Oxford University

‡Department of Economics, UC Berkeley, CEPR and NBER

§School of Management and Economics, Chinese University of Hong Kong, Shenzhen

¶School of Management and Economics, Chinese University of Hong Kong, Shenzhen

1 Introduction

Much of the anecdotal evidence from successful cases of rural economic development involve rural-to-urban migration: migrants learn about urban demand, supply, buyer-seller connections and sources of capital to seize opportunities back home. But this potential driver of development does not feature in the bulk of the existing theory and evidence on rural-urban migration (see e.g. Gollin (2014), Lagakos (2020) and Lagakos, Mobarak & Waugh (2023) for recent reviews). Do policies aimed at lowering rural-urban migration barriers lead to additional economic gains for rural origins and urban destinations through better market integration in trade and investment? How large are these gains, and is their incidence concentrated among rural origins or urban destinations – reducing or reinforcing incentives for rural-urban migration?

Answering these questions poses challenges. We rarely have access to a data environment with information on flows of migration, trade and investments at a geographically granular level within countries, in particular among low- and middle-income countries that are in the process of urbanization. Another challenge is identification: many policies that lower migration costs, such as transport investments, could also affect bilateral trade and investments directly. And origin-level (push-) or destination-level (pull-) shocks, that could be used to construct shift-share instrumental variables for bilateral migration, may fail the exclusion restriction in this setting, as regional pairs connected by (past) migration could also have bilateral exposure to shocks regardless of migration. Ideally, we would like to directly trace the effects of a policy affecting rural-urban migration barriers, using a natural experiment that is otherwise unrelated to flows of trade and investments.

To make progress on these challenges, we combine a unique collection of microdata from China with a new empirical strategy. We bring to bear four main datasets in the analysis. First, we build a comprehensive database of the legal provisions contained in the urban residence registration (Hukou) policy reforms in China starting in 1978 until 2020. The Hukou system imposes substantial costs on working and living in a city without local residence eligibility, primarily through restricted access to public services, employment rights and housing market restrictions. Our database contains roughly 25,000 Hukou eligibility changes among 900 of the roughly 1000 urban county-level destinations (including urban districts (shixiaqu) and county-level cities) over the 40-year period. For each of the rule changes, we record the urban destination, the year of implementation and the conditions that apply for worker eligibility. These include geographical restrictions on which origin regions are considered under the rule, demographic conditions on what types of workers are eligible and other conditions such as thresholds for required investments in the destination or employment by the local government. As we describe below, we use this database to measure origin-destination-level exposure to Hukou reforms over time.

Second, we use the universe of firm-to-firm sales transactions with VAT reporting requirements from China's State Taxation Administration (STA) over the period 2014-2018 to measure county-to-county (by sector) trade flows. The database includes roughly 16 billion transaction records between 16 million establishments. Third, we use administrative data from China's State Administration for Market Regulation (SAMR) with information on firm ownership stakes by both other firms and individual investors. We use these data to construct roughly 90 million county-to-county (by sector) investment flows in newly created establishments over the period 2010-2019. Fourth, we use representative individual-level samples from the Chinese population censuses in 2010 and 2015 with information on location, demographics and migration.

We use these data to propose an empirical strategy based on changes in origin-destination flows of trade, investment, migration and Hukou eligibility. In each year of data between 2010-2020, we consider all Hukou reforms that have been passed for a given urban destination since 1978, and quantify the proportion of the 2010 population for each of the roughly 1650 rural origin counties that were eligible to obtain urban Hukou in the destination (fractions from 0-1 for each rural origin-by-urban-destination pair and year). To do so, we focus on eligibility conditions that we can measure with precision in the census microdata, including a worker's origin location in 2010, type of Hukou status held by the worker, age, sex and years of education. We then implement a Poisson pseudo-maximum likelihood (PPML) estimation, including origin-destination(-sector), destination-time(-sector) and origin-time(-sector) fixed effects. The estimation relies on 'long' changes in bilateral flows of trade (2014-18), investment (2014-19) and migration (2005-10 vs. 2011-15) between rural origins and the rest of China on the left-hand side, and increases in bilateral Hukou eligibility on the right-hand side, conditioning on overall changes in origin(-sector)-level and destination(-sector)-level outcomes over time.

The identifying assumption is motivated by a growing literature on the causes and consequences of Chinese Hukou reforms across city regions, documenting that policy changes respond to urban labor demand conditions.[1] A city's planning commission may determine that more female workers between the ages of 20-40 with at least completed middle-school degrees would be beneficial for local industry. In addition, they may decide to implement this new eligibility rule first among rural origins within the same province, before extending it to other regions over time. Such a common type of policy in our database clearly responds to urban labor demand (destination-level shocks), but can lead to rich variation in changes of bilateral Hukou eligibility across rural origins over time –both because the pre-existing composition of workers with a given age, sex and education profile differs across regions and because Hukou reforms include geographical restrictions on eligible origin regions. The origin-time(-sector) fixed effects we include further capture the possibility that cities could be more or less likely to extend Hukou eligibility among rural origins with better recent economic performance. We use the database to investigate several potentially remaining concerns, including placebo tests on bilateral-specific pre-trends, testing for endogenous improvements in bilateral transport links, previously informal firms becoming formal in the wake of Hukou reforms and addressing concerns about correlated shocks across space or within sectors.

Implementing this design, we find that a 10 percentage point increase in a rural origin's Hukou eligibility to an urban destination –which corresponds to the average increase over the sample period 2014-19– increases rural-urban exports and imports by about 1.5% over a 4-year period. The effect on urban-to-rural investment inflows is an average 4.5% increase and rural-to-urban investment outflows increase by roughly 3.5% over a 5-year period. The effect on rural-urban migration flows is a 3% increase over a 5-year period. Bilateral migration stocks increase by about 1.5%.[2] Consistent with the increases in trade and capital flows being driven by better information and reduced matching frictions between rural and urban markets, we find that the number of unique buyer-seller and investor-investment pairs goes up by about 3% on average, and that Hukou reforms have no significant effect, with

[1]A recent example is Tian (2024), who finds that prefectures more exposed to export tariff reductions are more likely to increase Hukou eligibility for migrant workers. Other recent work on Hukou policies include Bosker et al. (2012), Ngai et al. (2019) and Zhang et al. (2019). The existing literature have generally relied on destination-level variation in Hukou reforms over time, either at the province or the prefecture levels.

[2]As we discuss in more detail in Sections 2 and 3, given limitations of the available data on migration flows in China, we focus the analysis on directly estimating the effects of the policy (reductions in rural-urban migration barriers) on flows of trade, investments and migration, instead of pursuing a two-stage least squares IV design. As part of Section 3, we then also quantify, with these caveats in mind, the second-stage point estimates of the split-sample IV implied by our estimates and relate them to existing findings in the literature.

point estimates close to zero, among sellers and buyers with pre-existing bilateral migrant linkages. In terms of heterogeneity, we find that the effect on rural exports is non-linear, declining at higher levels of origin eligibility, and is more pronounced among bilateral connections that are initially less integrated. We find no significant heterogeneity across initially richer/poorer or more/less educated rural origins. In terms of sectors, the effects on rural exports are significant for all broad economic sectors, agriculture, manufacturing and traded services, but strongest for agricultural exports and traded services, such as equipment rentals and repair, warehousing, transportation services and construction contracts. For rural investment inflows, we find that the effects are concentrated in the wholesale and retail sector and the same traded services mentioned above. Reductions in rural-ruban migration barriers therefore seem to have promoted market integration between rural and urban regions by facilitating the formation of new buyer-seller and investment linkages that increase flows of trade and capital in both directions. The findings also suggest some complementarity between investment inflows to rural regions and rural exports to cities, including in trade facilitating activities such as warehousing and wholesale and retail trade sectors.

In the second part of the analysis, we interpret these estimates through the lens of a simple spatial equilibrium model to answer two main questions. The first is to quantify the effect of Chinese Hukou reforms over the last decade 2010-2020 on migration market access and the resulting knock-on effects on trade and investment market access among the roughly 1650 rural origin counties in China. While our empirical design above relies on relative changes across bilateral trading routes, we thus aim to quantify the relative regional implications of reductions in policy barriers to migration for migration, trade and investment market access among rural origin regions. Second, while Hukou reforms affect migration costs unilaterally (rural-to-urban), the evidence suggests that the resulting effects on trade and capital flows are similar in both directions. The second question we then ask in this context is whether the additional gains in market integration through lifting migration restrictions have been stronger among the rural origins or the urban destinations during this period of Hukou reforms.

To answer these questions, we introduce three features of our empirical setting into an otherwise standard quantitative spatial equilibrium model. We allow for bilateral migration costs that include both distance-related costs as well as policy barriers, such as

Hukou restrictions. We introduce mobile capital in addition to labor as inputs to production, and we allow for buyer-seller matching frictions in trade and investment transactions following recent work by Eaton et al. (2023). We let these frictions be a function of bilateral migrant stocks, capturing the idea that migrants can act as a reduction in information and communication costs (e.g. Rauch & Trindade (2002)). The model yields three types of the well-known market access expressions for a given region; one for migration market access that is directly affected by changes in Hukou restrictions; as well as for trade and investment market access that can be indirectly affected by knock-on effects since they depend on both distance-related costs and bilateral matching frictions. A strength of our approach here lies in the fact that we are able to directly trace the policy's effects on flows of migration, trade and investment, so that we do not have to take a stance on the underlying structural parameters to quantify changes in market access.[3] To quantify changes in these expressions, we thus make direct use of the estimation results discussed above, in addition to estimating distance decay elasticities for flows of migration, trade and investment. We then proceed to measure each county's changes in market access for migration, trade and investment between 2010 (based on Hukou reforms passed 1978-2010) and 2020 (based on Hukou reforms passed until 2020).

We find that a 10% increase in migration market access among rural counties has on average led to a 1.5% increase in trade market access and a 2% increase in investment market access –a significant amplification of the traditional gains from migration through knock-on effects on market integration in trade and investment. While the gains in migration market access are naturally larger among the rural origins, we also find that the resulting gains in trade and investment market access are on average roughly 40% larger among the urban destinations over this period. The reason is that Hukou reforms tend to affect multiple rural origin regions for a given urban destination, so that market access gains from reductions in matching frictions tend to be larger among the urban destinations in this setting. Overall, our findings imply that the recent wave of Chinese Hukou reforms have brought significant additional gains to both rural origins and urban destinations beyond the traditional gains from migration, and that those knock-on effects have reinforced the incentives

[3] One would have to know these parameters for welfare analysis, as we derive in the theory. Our main goal here is to document the magnitude –implied by the estimates of our empirical analysis– of the policy's impact on the market access expressions that have been widely used in the literature. The model delivers those expressions parsimoniously, but we do not attempt to fully calibrate it for counterfactual analysis as the setup is deliberately stylized.

for rural-urban migration because they were larger among the urban destinations.

Our paper contributes to a longstanding literature on the role of rural-urban migration in economic development (e.g., Lewis (1954), Harris & Todaro (1970)). Recent work has focused on selection versus productivity differences to rationalize rural-urban wage gaps (e.g., Young (2013), Lagakos et al. (2020), Hamory et al. (2021)), the role of social networks in migration (e.g. Munshi & Rosenzweig (2016), Munshi (2020)), the income gains to rural migrants and their families (e.g., Bryan et al. (2014)), the labor market implications in rural origins or urban destinations (e.g., Kinnan et al. (2018), Akram et al. (2017), Imbert et al. (2022), Imbert et al. (2023), Guo et al. (2025)), and the aggregate gains from removing migration frictions due to sorting and agglomeration (e.g., Bryan & Morten (2019), Tombe & Zhu (2019), Lagakos et al. (2023)). We contribute to this large body of work by investigating a new channel through which rural-urban migration can affect both rural and urban economic development, and the distribution of economic activity in space.

We also build on the existing literature studying whether migration causes trade and investment flows. The bulk of this work have used data on cross-country bilateral migrant stocks in combination with international trade flows or foreign direct investment (FDI) (Javorcik et al. (2011), Cohen et al. (2015), Parsons & Vézina (2018), Burchardi et al. (2019), Bonadio (2023), McCully et al. (2024)). Closest in spirit to our empirical analysis are Burchardi & Hassan (2013) who estimate the effect of pre-existing social connections between East and West German households and regions on economic activity and entrepreneurship in the wake of the fall of the Berlin wall.[4] Relative to existing work in this space, we analyze these questions in the context of rural-urban migration, a ubiquitous phenomenon driving urbanization within developing countries. Empirically, our paper is the first to bring to bear within-country flows of trade, investments and migration to investigate these questions, combined with a natural policy experiment that we unlock with the Hukou policy database.[5]

The remainder proceeds as follows. Section 2 discusses the policy context and datasets

[4]Recent work by McCully et al. (2024) uses US Nielsen data to highlight the effect of migrants' home-biased consumption on imports. In our context, we can trace rural-urban food exports, but also trade and investment in all other sectors outside of groceries, including intermediate goods and services that account for the vast majority of bilateral flows. Our findings suggest that home-biased migrant consumption within China is not a discernible factor in the overall effect on trade flows in this setting.

[5]The theme of buyer-seller matching frictions also relates to a growing literature on firm-to-firm linkages (e.g. Bernard et al. (2019), Miyauchi (2024)), the role of business networks for firm performance (Cai & Szeidl (2018), Cai et al. (2024)), and the role of information frictions across agricultural markets (e.g., Goyal (2010), Allen (2014), Bergquist et al. (2023)).

used in the analysis. Section 3 presents the empirical strategy and estimation results. Section 4 presents the model. Section 5 presents the quantification at the regional level. Section 6 concludes.

2 Policy context and data

This section describes the institutional background of Hukou policy implementation and reform in China. We also describe the database that we construct to measure county-to-county trade flows, investment flows, migration and Hukou eligibility over time.

2.1 Context and Hukou reform database

Policy context

Introduced in 1958, the Hukou system in China was designed to control labor mobility through a mechanism of local residence registration. Each person is registered at birth as a permanent resident within a location and category (typically two main Hukou types: urban or rural). This registration (the Hukou) defines individuals' rights and obligations, such as the right to purchase property as well as eligibility for a wide range of government services, including public housing, education, child care, social insurance, health care and pensions. The Hukou system thus imposes substantial costs on individuals for leaving their Hukou location. While initially applied to almost all inter-regional migration, the system has been reformed gradually over the past decades in a decentralized manner.

Prior to the economic reforms initiated in 1978, Chinese workers were prohibited from seeking employment outside their designated Hukou registration. As there were no provisions for changing Hukou status, this effectively restricted inter-regional labor mobility. Subsequently, a series of incremental reforms progressively reduced the constraints imposed by the Hukou system. Hukou liberalization was slow at first and proceeded in four main waves. The first phase, known as the 'Blue Stamp Hukou', was rolled out from 1984 to 1998 to a very limited set of individuals. Entrepreneurs engaging in substantial investments, white-collar professionals, and farmers displaced due to governmental land acquisitions were for the first time able to apply for a change in their Hukou status and get registered in a different urban location. A second (still minor) set of reforms, occurring between 1997 and 2001, extended the eligibility criteria to migrants who had established permanent residence in a small number of selected (primarily smaller) prefectures, permitting their application for local Hukou status.

The third wave of Hukou reforms that introduced the first set of major changes started in the early 2000s and coincided with a substantial increase in internal migration. During this period, urban locations, in particular, saw a sharp increase in labor demand in the aftermath of China's accession to the WTO). This led to growing numbers of rural-urban migrants, most of whom were not able to change their Hukou status. This 'floating population' of migrant workers still had restricted rights and little access to urban amenities and government services. In response, governments at various geographic levels (local, prefectural, and provincial) gained more autonomy to adapt and experiment with immigration policies to meet their region's specific labor demand.

The fourth and most recent wave of Hukou reforms, encapsulated in 'The New Urbanization Plan (2014–2020)', initiated an evolution towards more inclusive urban societies in China. This wave coincides with the beginning of our main estimation samples for flows of trade and investment that we describe below. The aim was to increase access to cities for rural workers and foster urbanization patterns and market integration. Many of the initial policies over this period started by geographically targeting rural areas within the same province or prefecture before extending the reforms to a broader set of eligible regions over time. Despite some coordination at the national and provincial levels, implementation at the prefecture and county levels varied widely, leading to significant variation in migration costs across space, time, and types of eligible workers.

Hukou reform database

While existing work has documented and used destination-level panel variation in Hukou reforms, typically at the level of provinces or prefectures (e.g., Bosker et al. (2012), Zhang et al. (2019), Tian (2024)), our objective here is to account for variation in the incidence of these reforms at a more granular level of aggregation.

We make use of the fact that Hukou registration rules need to be publicly available to citizens in order to function. We thus attempt to collect the universe of local government documents containing legal provisions on Hukou residence and migrant rules covering the period between 1978-2020. To do so, we created two separate research teams ('double-blind') who are tasked to implement both manual and code-word searches across paper-based (archives and libraries) and digital sources of local government legal provisions. This yields a universe of 1801 policy documents containing legal provisions related to local Hukou eligibility.[6] Figure 1 depicts the evolution of Hukou reforms across the urban county-level administrative regions in China (see discussion of geographical units just below). The figure shows the number of city regions with any type of Hukou reform by year, depicting noticeable upticks during the 3rd and 4th waves in early 2000 and starting in 2014.

For each reform document, we collect information on the geographic urban destination unit(s) that each of the included rules apply to, the type of reform, the target group (based on workers' origin, Hukou status, demographics, additional requirements like local investments), and the implementation date of the policy. These 1801 policy documents yield 9,620 policy rules with conditions on worker origins, demographics and other requirements that are necessary for obtaining urban Hukou registration at the level of roughly 900 (of a total of about 1000) urban destinations (districts and county-level cities).[7] Roughly 45% of these policies are rules specific to a single county-level urban destination, whereas the remaining 55% apply to multiple urban county units within a prefecture or the province. When thus expanded to the level of county (6-digit) urban destinations, we obtain a total of 25,278 policy rules linked to county-level urban destinations for obtaining local Hukou registration.

To calculate bilateral eligibility for urban Hukou registration between origins and destinations, we compute the share of the population from each origin county that qualifies for Hukou reforms at a given destination, utilizing microdata from the Chinese population census in 2010 (see discussion below). To do this, we need to define a time-consistent geography at which we can measure both policies and outcome data: as we discuss below, we create 2661 time-consistent 6-digit codes corresponding to counties and districts. Not all eligibility criteria are either ex ante immutable or observable in the census data: For instance, investment requirements in the destination are a type of conditionality that we can not assign to workers across rural origins. Other conditions, such as occupation (e.g., working for the local government) after arrival in the destination are not ex ante immutable across rural workers. We therefore focus on five conditions that are observable in the census and

[6] Sources for policy documentation are wide-ranging, incorporating platforms such as Beidafabao ('ChinaLawInfo') (PKlaw, www.pkulaw.com), collections of government-issued documents, official online portals of governmental entities, government gazettes, archives of laws and regulations, as well as documents provided by relevant administrative units. Additionally, we conduct searches using a two-layer keyword approach across multiple platforms to unearth policies pertinent to migration.

[7] There were an additional 13% of policies that are targeted at family reunion –for instance, allowing children or parents of previous migrants to move where the worker is. We are primarily interested in migration of workers and measuring exposure to family reunion at an origin based on the initial stock at the destination is also complicated. For these reasons, we focus on non-family related Hukou relaxations –a total of 9,620 policies.

immutable ex ante: origin at the six-digit county level, type of Hukou (e.g., urban vs. rural, agricultural), age, gender, and educational years to determine whether each individual is eligible for each policy. 5,219 of the 9,620 policies (54%) include such conditions that we can measure with precision in the census microdata, leading to 11,793 rules when expanded to the level of urban county-level units.[8]

For each of these roughly 12 thousand policy rules, we calculate the percentage of the working-age population (ages 15-60) in each of the approximately 1,650 rural counties that satisfy these eligibility criteria. By iteratively analyzing the Hukou policy database from 1978 onwards, for each year between 2010 and 2020, we thus quantify changes in the eligible proportion of the local population for securing Hukou residency from rural origin counties to urban destination districts, based on the population observed in 2010. The resulting database provides the origin-by-destination-by-year proportion of an (initial) origin population that is eligible for local Hukou at the destination in that year, based on all eligibility rules that have been passed at a given time.

Figure 2 shows the upward trajectory in the average urban Hukou eligibility proportion between 2010-2020. Over the period of our main estimation sample 2014-2019, the average rural origin population eligibility share has increased by roughly 10 percentage points for the average rural-urban pair, from about 25% to 35%. In most other country contexts, such residence eligibility shares would be at 100% throughout.

2.2 Data on flows of trade, investment, migration and transportation

We combine the above with several additional datasets. Figure 3 presents a map of the geographical units used in the analysis. Tables 1 and 2 present descriptive statistics of the main datasets used in the analysis. Figures A.1-A.3 present non-parametric estimates of the distance decay in county-to-county flows of investments, trade and migration using this database.

[8]We thus omit Hukou reforms that do not include either geographical or demographic conditions, mostly related to investment targets, home ownership upon arrival or local government/public employment status and occupation after arrival. Such policies by definition do not have observable variation in exposure across rural origins ex ante, and our baseline assumption is that they are thus soaked up by the destination-time(-sector) fixed effects that we include in our estimation in Section 3. To the extent that rules targeting non-immutable characteristics are positively correlated with those we observe in the census, our measures will understate eligibility changes across origins, and our reduced-form estimates will over-estimate the effect of eligibility changes on migration, trade, and investment flows. The quantification in Section 5, however, would still reflect the policy impacts including due to potentically correlated unobserved eligibility changes, as we make direct use of the estimated impacts of measured eligibility changes in the model-based quantification.

Time-consistent spatial units

China's administrative structure is generally organized as a province (2-digit)-prefecture (4-digit)-county/urban district (6-digit) hierarchy, where county/urban district is the third level of administrative divisions. The two types of urban county-level units are county-level cities and urban districts (shixiaqu). Urban districts are regions within prefecture-level cities. Our analysis is mainly conducted at the county/district level based on two reasons: First, counties and districts (we will sometimes use these terms interchangeably in what follows) are the basic level of local government and play an important role in the administration of rural and urban areas. The official division of rural and urban regions within a prefecture is mainly based on county vs. district divisions. Second, counties and districts have a certain degree of administrative autonomy, allowing them to implement policies and manage resources within their jurisdiction. Many policies in China, including the most recent wave of Hukou reforms, are implemented at the district level (as discussed above).

One challenge of conducting county-level analysis is that the county boundaries can change over time due to splits, mergers or boundary modifications. We therefore construct the smallest time-consistent spatial units at the 6-digit level in China, spanning the years from 2000 to 2020. This involves documentation of all alterations in administrative boundaries over the period across all 6-digit administrative counties/districts in China, which we then classify into rural counties, county-level cities and urban districts. This yields roughly 1650 rural counties and 1000 urban districts and county-level cities, as shown in Figure 3, allowing us to track economic activity and migration at a granular geographical scale.

Firm registry data and investment flows

We use the firm registration database collected by the Chinese State Administration for Market Regulation (SAMR), which catalogs the universe of registered firms in China. It contains information on firm location, the year of establishment and exit (if any), the sector of activity, the value of registered capital at establishment, as well as the sources of firm equity at establishment. Importantly, establishments in China with separate tax registrations are considered distinct firms, even if they fall under the same parent company (Chen et al. 2021).[9] So what we mainly refer to as 'firms' in the text are establishments.

[9] For instance, Starbucks establishments, even within the same county-level location, appear as multiple separate entities with separate VAT registrations.

We use the equity ownership records for each newly registered establishment in the data to measure flows of capital (equity) in China over the period 2010-2020. The data allow us to distinguish between a) firm-to-firm investments through shareholding and subsidiaries, and b) individual-to-firm investment activities, where individuals are identified by unique shareholder IDs. The database contains a total of about 5 million firm-to-firm investments and 83 million investments by individuals over the period 2010-2020. We then combine the registered locations and reported investments to calculate bilateral equity investment flows in newly created establishments between each of the 2661 time-consistent 6-digit county pair i and j in each year t by adding up the total investment made by firms in county i to firms in county j. For individual-to-firm investments, we determine their origin locations using the first investment location that is linked to the investor ID in our data.[10]

We are also able to distinguish investment flows by the sector of activity of the new establishments. As shown in Table 2, we can distinguish between 19 main sectors of activity, including intermediate and final goods as well as services.[11]

VAT transaction data and trade flows

The second establishment-level dataset is from the State Taxation Administration (STA) that records all value-added tax (VAT) invoices issued by firms in mainland China from 2014 through 2018. To the best of our knowledge, this is the first study to use these data. It encompasses roughly 16.1 billion transactions among 16 million entities, providing a granular view of commercial exchanges within the Chinese market. It provides information on each transaction, including the location, industry classification and ownership type of buyer and seller firms, and transaction values for all transactions subject to VAT reporting requirements in China.

VAT filing in China is required for all transactions in order to claim input tax credits, and failures to comply are punishable by law. However, it is important to note that there are certain exceptions to transactions that fall outside the purview of the VAT invoice policy. In particular, invoices are not issued for direct sales to consumers and small firms are exempt from VAT payments: firms with average monthly sales below RMB 30,000 (approximately

[10]While we are thus able to trace investment flows in newly created establishments, one limitation of the data to keep in mind is that they do not allow us to also measure changes in ownership stakes of existing firms.

[11]The 19 sectors are accommodation, agriculture, construction, education, electricity, entertainment, finance, health, IT, leasing and business services, manufacturing, mining, real estate, residential services, technical services, transportation, utilities, wholesale and retail, and other.

USD 4,300) in a given year.

As for the investment flows, we are able to distinguish trade flows by the sector of the selling firm. For trade flows, we observe 18 of the 19 sectors included in the investment flows, in addition to breaking up the manufacturing sector into 16 more granular sectoral classifications, leading to a total of 34 sectors of trade flows.[12]

Population census and migration flows

To construct county-to-county bilateral migration flows, our study leverages individual-level microdata from several random samples of the Chinese population censuses of 2010 and 2015. For each census wave, we have access to two independent random samples that we combine in our database, yielding about 7 million observations for 2010 and 3.5 million observations for 2015.[13] These microdata allow us to implement our computation of the bilateral Hukou eligibility database, by determining for each individual included in the 2010 census whether they are eligible for any of the roughly 12 thousand Hukou residence reforms implemented in our database over the period 1978-2020.

While the random census samples provide rich information on the demographic characteristics across different counties (with samples of more than 2500 per spatial unit on average in 2010), these data are many orders of magnitude sparser for measuring county-to-county bilateral migration flows or changes in bilateral migration stocks compared to the administrative data we are able to bring to bear on flows of trade and investments that we describe above, which capture 100% of the administrative datasets. With this data limitation in mind, we use information on the current residence, the previous residence and the time of arrival in the current residence to construct the sum of migration flows over the 5-year period 2005-2010 in the 2010 combined census sample, and the period 2011-2015 for the 2015 combined census sample. The 2010 population census fortunately also includes information on migration anytime before 2005, which we use to obtain estimates of the changes in the stock of bilateral migrants over the period 2005-2010.

[12]The additional classifications within manufacturing are: processed food and agriculture, chemical industry, computing equipment, electronics, sports and entertainment goods, general merchandise, instruments, metal goods, nonmetal mineral goods, printing goods, smelting production, special equipment, textiles, transportation equipment, wood and furniture and other manufacturing.

[13]China conducted a full population census in 2010 and a 'mini-census' covering about 1.5% of the population in 2015. We have access to two independent random samples for research use for each of the two rounds. These data have been used to measure migration flows in the existing literature (e.g., Combes et al. (2015), Tombe & Zhu (2019), Imbert et al. (2022)).

3 Empirical analysis

This section presents the empirical strategy and estimation results.

3.1 Empirical strategy

Using the data that we describe in the previous section, we propose the following specification that we estimate in PPML:

$$Y_{ijst} = exp\left[\alpha_{ijs} + \alpha_{jst} + \alpha_{ist} + \beta_1 \text{Eligibility}_{ijt}\right] \times \epsilon_{ijst} \,, \tag{1}$$

i indexes the roughly 1650 rural origin counties in China, j indexes roughly 2650 county destinations in China, s indexes sectors of production (for trade and investment flows) and t indexes years. Y_{ijst} are bilateral flows of trade and investment between rural origins and their trading pairs in the rest of China. α_{ijs}, α_{jst} and α_{ist} indicate origin-destination-sector, destination-sector-time and origin-sector-time fixed effects respectively. ϵ_{ijst} is an error term that we cluster at the level of destination-year –the level at which the treatments in our data are correlated across units (Abadie et al. 2023). For migration flows, specification (1) does not have a sectoral dimension s.

For trade flows, we measure rural exports to destination markets and rural imports from destination markets across 34 sectors of production, and use the long change in the total value of bilateral flows between the first year of our data in 2014 and the last year in 2018 (using two years of trade data on the left-hand side in (1)). For investment flows, we measure either rural inflows or outflows and use long changes over the period 2014-2019. For migration, we use changes between migration flows from the rural origins during the period 2005-2010 (observed in 2010 population census) and those during the period 2011-2015 (observed in the 2015 census), again with two periods on the left-hand side. To measure changes in bilateral migration stocks, we make use of the fact that the 2010 census included a question on whether the individual had migrated to the current location at any time before 2005 in addition to their origin location. When having migration on the left-hand side, the regressions do not feature a sectoral dimension (s).

Eligibility_{ijt} is the fraction (0-1) of the origin population eligible for Hukou registration in the destination in a given year. To limit potential concerns about past treatments that may continue to affect the control group during our samples, we exclude bilateral flows to urban destinations that implemented Hukou reforms during the 5-year period before the start of our sample in 2014 for trade and investment flows, or before 2010 for migration flows.

Following recent contributions at the intersection of applied econometrics and gravity specifications in international trade and economic geography (e.g. Wooldridge (1999), Silva & Tenreyro (2006), Fally (2015), Sotelo (2019), Roth & Chen (2023)), we estimate specification (1) in PPML. This reflects the reality that many bilateral pairs in our five main connectivity matrices (exports, imports, investment inflows, outflows and migration) report zero bilateral flows in a given year of data, and allows us to estimate the effect of Hukou reforms capturing both intensive and extensive margin changes on the left-hand side.[14]

Identification in (1) is thus based on changes in bilateral flows of rural exports, imports, investment inflows, outflows and migration outflows on the left-hand side, and increases in origin-destination Hukou eligibility on the right-hand side, conditional on overall changes in either origin-level or destination-level economic outcomes over time. The identifying assumption is that urban Hukou reforms respond to urban labor market conditions in the destination, but are not specifically targeted at contemporary idiosyncratic changes on a given bilateral route. Targeting based on past or contemporary economics changes across origins or origin-sector pairs, on the other hand, would be captured by the inclusion of $\alpha_{i(s)t}$.

There are several potential concerns that we use our database to further investigate. We first estimate (1) including α_{ij}, α_{jt} and α_{it}. A concern with this could be that urban labor market demand shocks at the destination may be correlated with rural labor market shocks at the origin due to sectors of production that may be more similar between origin-destination pairs experiencing increases in Hukou eligibility. To assess this possibility, we estimate the effects of bilateral increases in Hukou eligibility after including origin-destination-sector, destination-time-sector and origin-time-sector fixed effects. We consider this our baseline specification, as shown in (1).

Second, maybe urban Hukou reforms are on average explicitly targeted at particular bilateral connections that have shown greater promise in terms of trade and investment flows –in a way that is also not captured by the origin-sector-time fixed effects–, violating

[14] Weidner & Zylkin (2021) also investigate sensitivity to potential incidental parameter concerns in a three-way fixed effects PPML specification. Our empirical context is reassuring in this respect, as it offers substantially larger N compared to the more common cross-country setting.

the parallel trends assumption across bilateral routes. To investigate this possibility, we make use of the fact that we can measure investment flows not only for the 5-year period we use in the baseline estimation (2014-2019), but also for the previous period between 2010-2014. We thus estimate a placebo falsification test whether changes in bilateral investment flows 2010-14 are systematically related to (future) bilateral changes in Hukou policy reforms 2014-19.

Third, urban destination-level shocks, that are likely related to Hukou reform decisions, could disproportionately affect rural origins that are more likely targeted by the Hukou reforms –for other reasons than lowering rural-urban migration barriers. In our standard models of trade (with structural gravity and ad-valorem trade costs), any such destination-level shock affects all bilateral routes proportionately, and would thus be fully captured by the destination-sector-time fixed effects. But with plausible deviations from that theoretical benchmark –e.g. allowing for additive trade costs– local shocks could propagate differently as a function of bilateral distances. To investigate this possibility, we also estimate the effects of changes in bilateral Hukou eligibility conditional on destination-time-distance fixed effects –effectively conditioning on changes in bilateral outcomes within a given distance of each of the destinations over time. In particular, we interact the α_{jst} of the baseline specification with 20 bilateral distance bands (20 five-percentile bins of bilateral distances) around each individual destination. This specification thus compares changes over time in flows to or from the same destination across origins that are more or less exposed to Hukou reforms, but now within the same distance-radius around the destination.

Fourth, it could be the case that trade and investment flows have on average changed systematically as a function of bilateral distances in China over this period –e.g. transport investments could have led to larger increases among more distant routes–, while at the same time bilateral distances could be on average related to the likelihood of experiencing increases in urban Hukou eligibility. Furthermore, such a secular trend toward more market integration could affect rural-to-rural flows differently over time compared to rural-to-urban flows, giving rise to bias in specification (1). To assess this, we estimate effects after also including bilateral distance-by-time fixed effects that we also allow to differ between rural-to-rural vs. rural-to-urban flows of trade and investments. We use the same 20 five-percentile bins of bilateral distances to create those additional fixed effects.

Finally, there are questions about the interpretation of the effect of Hukou reforms, beyond assessing potential bias in specification (1). Even if changes in Eligibility_{ijt} are as good as randomly assigned across bilateral pairs over time, the interpretation of the effects on trade flows could be complicated by the fact that bilateral eligibility changes can also affect the pre-existing stock of (potentially unregistered) migrants from i residing in j. In particular, Hukou reforms could make it easier for existing unregistered migrants from i operating in j to claim local Hukou, then register their business and start reporting transactions under VAT requirements –even if both purchases from the countryside (rural exports) and sales to the countryside (imports) were pre-existing but unregistered.[15] To investigate this possibility, we can use the STA VAT database to estimate a robustness check after limiting the sample to trade transactions by firms that were already registered (and reporting VAT transactions) at the beginning of the sample in 2014. Another question about interpretation is the possibility that trade costs may react endogenously to increases in Hukou eligibility on a given rural-urban bilateral pair –e.g., more roads or railway connections could be established or roads and railways could be upgraded on those routes to facilitate increased travel activity. To assess this, we make use of the transportation panel database provided in recent work by Davis et al. (2025) to test for endogenous reductions in bilateral trade costs.

3.2 Results

Trade flows

Main results Table 3 presents the estimation results of specification (1) for rural exports and imports. We first present estimates of the effect of bilateral Hukou reforms after including origin-destination, origin-time and destination-time fixed effects. We then present our baseline specification with origin-destination-sector, origin-sector-time and destination-sector-time fixed effects. Estimation samples are identical across the columns (with numbers of observations reflecting multicollinearity with the increasing number of fixed effects included).

In our baseline specification in column 2, we find that a 10 percentage point increase in an origin's Hukou eligibility for an urban destination increases bilateral exports by about 1.5 percent over a 4-year period. In column 5, we find roughly symmetric effects for rural

[15] Having said that, local governments tend to be welcoming of local business creation in case applicants, including migrants, have the investment equity to found a new establishment. Business activities based in fixed-point locations are also predominantly registered by the state in the Chinese compliance context (e.g., State Council (2015)).

imports. In columns 3 and 6, we then assess the robustness to including destination-sector-distance-time fixed effects and distance-sector-time-rural/urban fixed effects as discussed in the empirical strategy above. Despite the inclusion of a large number of additional fixed effects, we find very similar point estimates and significance. This provides some reassurance that differential exposure to destination-level shocks or overall trends of growth in trade flows as a function of bilateral distances within rural-rural or rural-urban pairs are unlikely driving the point estimates.

Channels and robustness We start in Table 4 by presenting the effects of Hukou eligibility on the number of unique bilateral buyer-seller matches in the trade flow data. We also report the effects on the number of sellers and buyers separately. In support of a mechanism based on better information and reduced frictions to buyer-selling matching across markets, we find that a 10 percentage point increase in the origin's Hukou eligibility increases the number of unique buyer-seller pairs for rural export flows and import flows by about 3% using the same baseline estimation specification respectively. We find that the number of buyers per seller increases for both exports and imports, in line with the motivating evidence for buyer-seller matching frictions in Eaton et al. (2023). Increased rural-urban migration leads to the formation of new trading relationships in both directions.

To further assess the role of better information and matching, we also restrict the estimation sample to buyers or sellers with pre-existing migrant linkages –either urban establishments operated by previous migrants from the rural origin that already existed at the beginning of the sample in 2014 or owners with existing businesses in 2014 in rural markets who had migrated from the city. In the first two columns of Table 5 we find no significant effects of Hukou reforms on trade flows between establishments that already had bilateral migrant linkages, in support of an information-based mechanism. These results are also relevant when considering a demand-based explanation, that migrants create rural exports by bringing home-biased tastes to the city. One would expect similar (or larger) positive effects on rural exports among this subset of connected establishments if they were providing household consumption goods from the rural origins in the city. We return to this theme when documenting the sectoral effects in the heterogeneity analysis below.

We also assess the possibility that the effect on trade could in part be driven by knock-on effects of the Hukou reforms on bilateral transportation improvements –reductions in trade costs that could be driven by increased demand for bilateral transportation. To do so, we make use of the comprehensive transportation panel database provided by Davis et al. (2025) that tracks the existence and characteristics of transport links for Chinese roads and highways, controlled-access motorways (expressways), high-speed railways and traditional railways over the period 1993-2020. We use information on the speed limits for each segment of these four transport networks in a given year of data in order to compute least-cost travel times between the centroids of all rural counties to all other county centroid destinations (matrices of roughly 1650×2650) for each of the 4 networks in the years 2014 and 2019.[16] Columns 3 and 4 in Table 5 report the results from a fixed-effects specification with the average or minimum travel times (in hours), computed across the four transport networks, between rural origins and their destinations on the left-hand side. We find close to zero and statistically insignificant point estimates, suggesting that endogenous bilateral upgrades in transportation infrastructure are unlikely a mechanism driving the observed effects on trade flows in our setting.

Finally, we assess the concern that the increase in rural exports and imports to and from the city may be in part driven by previously informal merchants who become more likely to register their establishments in the wake of Hukou reforms in the city and start reporting trade flows under the VAT requirements as a result. To do so, we re-estimate our baseline specification from Table 3 for county-to-county exports and imports after restricting the estimation sample in urban markets to firms that were already formal and reporting VAT transactions at the beginning of the sample in 2014. In Columns 5 and 6 of Table 5 we find similar point estimates and levels of statistical significance as in our baseline results (and cannot reject that they are equal). This provides some reassurance that transactions between previously informal merchants are unlikely to drive the positive effects on rural-urban trade integration. Instead, much of the increased flows appear to be driven by existing urban establishments that form new matches with rural sellers and buyers.

Heterogeneity Table 6 reports additional results about the heterogeneity of the effect on rural exports across types of origin regions, pairs and sectors of production. In column 1, we test for non-linearity in the effect of bilateral Hukou eligibility shares and find that the positive effect on average increases up to about 50% of initial local population eligibility before decreasing at higher shares. In column 2, we find that the effect on rural exports

[16] We compute travel times between centroids and the nearest segment of the networks by assigning the lowest road travel speed (40 km/h) to straight lines connecting the centroid. Results remain unchanged (close to zero effects and not statistically significant) if we instead "snap" the centroids directly onto the network.

is significantly stronger as a function of percentiles of bilateral distances between origin-destination pairs. This suggests that migrants may reduce information and communication frictions more strongly among bilateral pairs that were initially less integrated.

Conversely, we do not find significant heterogeneity as a function of percentiles of rural origins' initial output per capita (measured in 2014) or average years of education (measured in 2010). Across sectors (columns 2-5), we find significant positive effects on rural exports for agriculture (1 sector), manufacturing (16 sectors) and services (16 sectors).[17] The effects are most pronounced for agricultural exports and traded services. Traded services are IT services, equipment leasing and other business services, temporary accommodation, transportation and warehousing and construction contracts. Positive effects on exports in the VAT data imply that urban buyers purchased from rural establishments in those sectors.

Investment flows

Main results Table 7 presents the estimation results for rural investment inflows and outflows, following the same sequence of specifications as in the trade flow Table 3. In column 2, we find that a 10 percentage point increase in an origin's Hukou eligibility to an urban destination increases urban-rural investment inflows by about 4.4% over a 5-year period. In column 5, we find roughly symmetric effects for rural-urban investment outflows of about 3.4%. To interpret the investment outflow results, it is helpful to clarify that counties in China classified as rural almost always include urban parts (county capitals and smaller-tier municipalities, such as township centers, within the countryside). Columns 3 and 6 confirm the positive point estimates after including the saturated set of fixed effects discussed above, but we start to lack power in the case of rural-urban investment flows.

Channels and robustness Table 8 uses the fact that we are able to observe investment flows before 2014 to present results from a placebo falsification exercise. In columns 1 and 3, we present the baseline results from Table 7 on rural investment outflows and inflows 2014-19. In columns 2 and 4, we keep the identical estimation samples and right-hand side of the estimation, but replace investment outflows or inflows with the years 2010 and 2014. Reassuringly, we find no evidence suggesting that bilateral routes affected by Hukou reforms during the period 2014-2019 were already subject to positive (or negative) trends in the previous period 2010-2014, for either rural-urban investments or urban-rural invest-

[17] We do not separately analyze mining due to few observations.

ments. The point estimates are close to zero, while the standard errors are very similar to the baseline estimation period. We view this as corroborating evidence that urban Hukou reforms are unlikely targeted at specific bilateral routes that have experienced contemporaneous bilateral-(and sector-)specific shocks, in line with the existing finding that Hukou reforms respond to urban labor market developments.

To assess the role of improved information and matching frictions, Table 9 presents the effects of Hukou eligibility on the number of unique bilateral investor-investment matches in the investment data. We also report the effects on the number of investors and investments separately. Similar to the buyer-seller effects underlying trade flows, we find that number of unique matches goes up by about 3% for both inflows and outflows due to a 10 percentage point increase in rural-urban Hukou eligibility. And we find suggestive evidence that the number of investors per investment increases, even though the effects on numbers of investors and investment projects are not statistically significantly different.

Heterogeneity Table 10 presents results on the heterogeneity of the effect on rural investment inflows across types of origin regions, pairs and sectors. In columns 1 and 2, we find no significant non-linearity or heterogeneity as a function of percentiles of bilateral distances, origin output per capita or average years of education. Across sectors, we find positive but imprecisely estimated point estimates for newly created agricultural firms or manufacturing firms. We find significant positive effects on equity inflows in the wholesale and retail sectors. We also find significant positive effects in the same traded services categories that respond in terms of exporting in the previous Table 6. These results suggest some complementarity between investment inflows to rural regions and rural exports to cities, including in trade facilitating activities such as warehousing and wholesale and retail trade sectors.

Migration flows

While the database we were able to build combining Hukou reforms with trade and investment flows over the period 2014-2019 above are based on the universe of existing administrative records, the available data environment is, unfortunately, much less complete for studying migration flows or changes in bilateral migration stocks, as the Chinese data on county-to-county migration are only available for tiny random subsamples of the population. In addition, the way that migration flows are recorded in the data maps less clearly to estimating the effects of changes in bilateral Hukou eligibility, as we discuss below.[18] With these limitations in mind, we use the best available information to implement the empirical strategy in specification (1).

Table 11 presents the estimation results for rural migration flows, including the same baseline and the more saturated fixed effects specifications from the previous Tables 3 and 7 (but without a sectoral dimension). We compare migration flows reported by individuals in the 2010 census over the period 2005-10 to migration flows over the period 2011-15 in the 2015 census, and project them on changes in bilateral Hukou eligibility over the period 2010-2015. The first (cautionary) element to note is the drastic reduction in the estimation samples compared to the previous analysis on trade and investment flows (with the number of observations in Column 1 being less than 0.3% compared to the same specification in Column 1 of Table Tables 3). In our baseline specification in column 1 (with origin-destination, origin-time and destination-time fixed effects), we find that a 10 percentage point increase in an origin's Hukou eligibility leads to a roughly 3% increase in bilateral migration flows. In column 2, we use the same saturation as for trade and investment flows, including destination-distance-time fixed effects and distance-time-rural/urban fixed effects. The point estimate sightly decreases from 2.9 to 2.2% and remains statistically significant.

In columns 4 and 5, we make use of the fact that the 2010 census also included a question on whether the individual had migrated to the current location at any time before 2005 (and their origin location). We use this to compute the change in the bilateral migration stock between 2005-2010 (adding the 5-year migration after 2005 in 2010 to the pre-existing stock before 2005). Here, we find that the bilateral stocks increase by about 1.5%, when averaging between the more and less saturated fixed effects specifications that are both statistically significant.

Two-stage least squares interpretation Given the data limitations for migration flows in our empirical context, we focus the empirical analysis on documenting the effects of policy changes in migration restrictions directly on flows of trade, investments and migration, instead of a two-stage-least-squares design with effects on migration as the first stage –which would have to be a split-sample IV strategy in our context, as the 2020 Chinese population

[18] Having said this, there are several existing studies documenting a positive and significant effect of Hukou reforms on (in) migration at the level of destination regions (e.g., An et al. (2024), Fan (2019), Sieg et al. (2023), Tian (2024)).

census has not been made available for researchers, regardless of concerns about tiny random samples in migration flows. Having said this, we can compare the implied 2nd-stage point estimate from such a split-sample IV analysis to the existing literature on the effects of cross-country migration on trade and investment flows. Using the baseline estimates for effects on trade (1.5%), investment (4%) and bilateral migration stocks (1.5%), the implied effect of a 1% increase in bilateral migrant stocks would be roughly 1% on trade flows and 2.5% on investment flows.

These effect sizes are broadly in line with some of the most recent empirical evidence using cross-country flow data. Parsons & Vézina (2018) use an IV strategy based on the quasi-random allocation of Vietnamese migrants across US states after the fall of Saigon to estimate that a 1% increase in a State's stock of Vietnamese migrants led to a 1.4% increase in exports to Vietnam.[19] For the effect of migrants on investment flows, recent influential work by Burchardi et al. (2019) use an IV strategy based on the interaction of past out- and in-migration waves between Europe and US states to estimate that doubling the population with foreign ancestry increase in the likelihood of at least one foreign direct investment (FDI) project on the ground from the origin country by about 4%. While the comparison of this binary outcome across countries to percent changes in the value of county-to-county investment flows in our setting is somewhat heroic, it does not seem implausible given the average value of FDI projects in the US and the fact that our PPML estimates include extensive-margin effects.

Benchmarking to distance elasticities

Table 12 benchmarks the estimated effects of Hukou eligibility by estimating the distance elasticity of the respective bilateral county-to-county(-by-sector) flows that we use in the estimation above, using the identical samples. To this end, we keep the origin-time(-sector) and destination-time(-sector) fixed effects, but exclude the origin-destination fixed effects since bilateral distances are time-invariant. Migration flows have the highest estimated distance decay elasticity of about -2.1, followed by trade flows (-1.3 for exports and imports). Investment flows have slightly lower estimated distance decay elasticities, with roughly -1 for outflows and -0.8 for inflows.

We can use these estimates to benchmark the estimated effects on trade, investment and migration flows. The average effect of a 10 pp increase in Hukou eligibility on rural exports

[19]Based on the IV specification with controls in Column 6 of Table 2 in Parsons & Vézina (2018).

and imports would be equivalent to a roughly 1.2% reduction in bilateral distances (1.5%/-1.3). The estimated average effect on investment inflows and outflows would be equivalent to a roughly 4.4% reduction in bilateral distances (4%/-0.9). And the effect on migration flows would be equivalent to a roughly 1.4% reduction in distances (3%/-2.1). In addition to benchmarking our estimates, we will use the bilateral distance decay elasticities in Table 12 to calibrate the market access expressions that we derive from the model below.

4 Model

The model builds on the workhorse structure of recent quantitative spatial models (Allen & Arkolakis 2014, Redding 2016). Motivated by the empirical setting, we extend this setup in three dimensions to derive the market access expressions that we use in the quantification of the region-level implications in Section 5. Those are allowing for bilateral migration costs that can be a function of regional distances as well as government policies, such as Hukou regulations, allowing for mobile capital in addition to labor as inputs to production, and incorporating buyer-seller matching frictions following recent contributions by Eaton et al. (2023) and Eaton & Kortum (2024) that we allow to be functions of bilateral migrant stocks. We proceed with a streamlined exposition here, while the Appendix provides additional details and derivations.

Environment

There are N regions that are connected across ij regional pairs through flows of trade, migration and capital investments. Regions are populated by workers, capitalists, producers of goods and retailers. After making a migration decision, workers supply labor inelastically in their region. Capitalists are immobile, but they can invest in other regions. Producers use local labor and capital to produce differentiated varieties of goods, which are shipped across regions to be purchased by retailers. Retailers purchase the goods and sell them at-cost to workers in their respective regions.

Trade flows in goods, migration flows and investment flows are subject to iceberg-type bilateral costs, τ_{ij}, τ_{ij}^{m} and τ_{ij}^{k} respectively. In addition, transactions between capitalists and producers, and between producers and retailers take place through buyer-seller matching. The matching rates of these interactions between two regions, denoted by λ_{ij}^{k} and λ_{ij} respectively, measure the fraction of matches given a number of buyers and sellers in the destination market.[20] Capitalists can match with multiple producers and producers can match with multiple retailers, but for simplicity retailers are assumed to match with a single variety. We now describe the behavior of each agent in more detail.

Workers

Workers inelastically supply a unit of labor in their region of residence i. They consume varieties of goods from different origins j that are offered by the number of active retailers (stores $s \in \mathcal{S}_i$) in their region, such that they solve[21]

$$\max_{\{c_{ji}(s)\}} \left(\frac{1}{\mathcal{S}_i} \int_{s \in S_i} c_{ji}(s)^{1-\frac{1}{\sigma}} ds \right)^{\left(1-\frac{1}{\sigma}\right)^{-1}}$$
$$\text{s.t.} \int_{s \in \mathcal{S}_i} p_{ji}(s) c_{ji}(s) ds = w_i,$$

where w_i is the wage income of workers in i. Demand is: $c_{ji}(s) = \left(\frac{p_{ji}(s)}{P_i}\right)^{-\sigma} \frac{w_i}{P_i}$ for $s \in \mathcal{S}_i$, where $P_i = \left(\frac{1}{\mathcal{S}_i} \int_{s \in \mathcal{S}_i} p_{ji}(s)^{1-\sigma} ds\right)^{\frac{1}{1-\sigma}}$.

Migration

A mass of $L_{j,0}$ workers are born in each location j. Each worker ψ differs in their taste $A_{ji}(\psi)$ for amenities between the bilateral pair of their birthplace j and a destination i, where $A_{ji}(\psi)$ is an i.i.d. draw across workers from a Frechet distribution with shape parameter ϕ. Workers pay the migration cost $\tau^m_{ji} \in (0, 1)$ for each unit consumed in i implying indirect utility is: $V_{ji}(\psi) = \max_i A_{ji}(\psi) \frac{w_i}{P_i}(1 - \tau^m_{ji})$. The number of workers migrating from j to i is then:

$$L_{ji} = \frac{\left(\frac{w_i}{P_i}(1 - \tau^m_{ji})\right)^{\phi}}{\sum_{l=1}^{N} \left(\frac{w_l}{P_l}(1 - \tau^m_{jl})\right)^{\phi}} L_{j,0},$$

where the fraction on the right determines the share of workers migrating from j to i: $L_{ji} = \pi^m_{ji} L_{j,0}$.

[20] Both trade costs (τ_{ij}) and matching frictions (λ_{ij}) are are assumed to be symmetric (e.g. $\tau_{ij} = \tau_{ji}$) and costless for interactions within the same region ($\tau_{ii} = \tau^k_{ii} = 1$ and $\tau^m_{ii} = 0$ below).

[21] As discussed in Eaton & Kortum (2024), setting up preferences in this way shuts down feedback from changes in the mass of active retailers on the local price index through love of variety.

Capitalists

Each region i has a distribution of capitalists indexed by l. The returns to capital in region i are denoted R_i.[22] There are idiosyncratic monitoring costs $1/z_i^k(l)$ per unit of capital invested in i, with $z_i^k(l)$ drawn i.i.d. from a Pareto distribution with shape parameter θ, and an iceberg cost τ_{ji}^k of investing from region j in region i. Capitalists have deep pockets, meaning their supply of capital to producers is perfectly elastic conditional on their willingness to invest, i.e., given $\frac{R_i z_i^k(l)}{\tau_{ij}^k} \geq \bar{r}$, where $\bar{r}$ is the variable cost of investing. The mass of capitalists from j who can invest capital in i is then

$$\mu_{ji}^k(R_i) = \left(\bar{r}\tau_{ji}^k\right)^{-\theta} R_i^\theta.$$

The number of capitalists in j who offer a rate less than or equal to R and match with an intermediate producer in i is distributed Poisson with rate: $\rho_{ji}^k(R) = \lambda_{ji}\left(\bar{r}\tau_{ji}^k\right)^{-\theta} R_i^\theta$. The total number of capitalists in j who can offer a rate less than or equal to R and match with an intermediates producer from anywhere is distributed Poisson with rate: $\xi_j^k(R) = \sum_{i=1}^N \lambda_{ji}\left(\bar{r}\tau_{ji}^k\right)^{-\theta} R_i^\theta$; and the number of capitalists from anywhere who offer below R and match with an intermediates producer in i is distributed Poisson with rate: $\delta_i^k(R) = \sum_{j=1}^N \lambda_{ji}\left(\bar{r}\tau_{ji}^k\right)^{-\theta} R_i^\theta$. This setup implies that the share of capital that capitalists who can finance at below rate R send from j to i is given by:

$$\Delta_{ji} = \frac{\lambda_{ji}^k(\tau_{ji}^k)^{-\theta} R_i^\theta}{\sum_{l=1}^N \lambda_{jl}^k(\tau_{jl}^k)^{-\theta} R_l^\theta}.$$

And the share of capital that producers in i receive from capitalists in j is: $\pi_{ji}^k = \frac{\lambda_{ji}^k(\tau_{ji}^k)^{-\theta}}{\sum_{l=1}^N \lambda_{li}^k(\tau_{li}^k)^{-\theta}}$.

Producers and retailers

A producer ω in location i has the production function

$$Y_i(\omega) = z_i(\omega) Z_i \left(\frac{L_i(\omega)}{1-\beta}\right)^{1-\beta}\left(\frac{K_i(\omega)}{\beta}\right)^{\beta},$$

where Z_i is an exogenous parameter that captures regional differences in productivities and $z_i(\omega) \sim \text{Pareto}(1,\gamma)$. The measure of firms with $z(\omega) \leq z : \mu(z)$ is then $\mu(z) = z^{-\gamma}$. Cobb-

[22] If multiple capitalists match with the same producer, we assume they evenly divide the financing at the common interest rate.

Douglas production implies the following cost given wages w_i and interest rate $R_i : c_i(\omega) = \frac{w_i^{1-\beta} R_i^{\beta}}{Z_i z_i(\omega)}$. Since producers and retailers sell at-cost, the price of variety ω, i sold in location j is

$$p_{ij}(\omega) = \frac{\tau_{ij} w_i^{1-\beta} R_i^{\beta}}{Z_i z_i(\omega)},$$

given the iceberg trade cost τ_{ij}. The measure of producers that can ship from i to a retailer in j at a cost below c is: $\mu_{ij}(c) = \left(\frac{\tau_{ij} w_i^{1-\beta} R_i^{\beta}}{Z_i}\right)^{-\gamma} c^{\gamma}$. The number of matches between producers in i and retailers in j for a cost at or below c is distributed Poisson with the rate parameter: $\rho_{ij}^{p}(c) = \int_0^c \lambda_{ij} d\mu_{ij}(c') = \lambda_{ij}\mu_{ij}(c)$. The total number of matches between retailers in j and producers from anywhere who can sell at or below c is: $\rho_j^p(c) = \sum_{i=1}^{N} \rho_{ij}^p(c) = \sum_{i=1}^{N} \lambda_{ij}\mu_{ij}(c)$. Thus, the share of matches that retailers in j have with firms who can ship at cost at or below c from i is

$$\pi_{ij} = \frac{\lambda_{ij}\left(\frac{\tau_{ij} w_i^{1-\beta} R_i^{\beta}}{Z_i}\right)^{-\gamma}}{\sum_{k=1}^{N} \lambda_{kj}\left(\frac{\tau_{kj} w_k^{1-\beta} R_k^{\beta}}{Z_k}\right)^{-\gamma}}.$$

Equilibrium

Given location fundamentals $\{Z_i, L_{i,0}\}$, elasticities $\{\sigma, \beta, \phi, \theta, \gamma\}$, sets of bilateral iceberg trade, capital market, and migration costs $\left\{\tau_{ij}, \tau_{ij}^{k}, \tau_{ij}^{m}\right\}$, and bilateral matching frictions between capitalists, producers and retailers $\left\{\lambda_{ij}, \lambda_{ij}^{k}\right\}$, an equilibrium is defined by vectors of wages and capital prices $\{w_i, R_i\}$, populations $\{L_i\}$ and total expenditures $\{E_i\}$ such that trade is balanced and goods markets clear:

$$E_i = \sum_{j=1}^{N} \frac{\lambda_{ij}\left(\frac{\tau_{ij} w_i^{1-\beta} R_i^{\beta}}{Z_i}\right)^{-\gamma}}{\sum_{k=1}^{N} \lambda_{kj}\left(\frac{\tau_{kj} w_k^{1-\beta} R_k^{\beta}}{Z_k}\right)^{-\gamma}} E_j,$$

capital markets clear and are balanced across regions

$$R_i K_i = \sum_{j=1}^{N} \frac{\lambda_{ji}^{k} (\tau_{ji}^{k})^{-\theta} R_i^{\theta}}{\sum_{l=1}^{N} \lambda_{jl}^{k} (\tau_{jl}^{k})^{-\theta} R_l^{\theta}} R_j K_j,$$

agents migrate to their chosen regions according to the expression above: $L_{ji} = \pi_{ji}^{m} L_{j,0}$. And firms maximize profits such that $R_i K_i = \beta E_i$ and $w_i L_i = (1-\beta) E_i$. The regional price index, up to a common scaling factor, is given by:

$$P_i = \left(\sum_{j=1}^{N} \lambda_{ji} \left(\frac{\tau_{ji} w_j^{1-\beta} R_j^{\beta}}{Z_j} \right)^{-\gamma} \right)^{-\frac{1}{\gamma}}.$$

Market access expressions

To derive the trade and investment market access expressions, we start by rewriting regional total expenditures, E_i above, in terms of local production (factor prices and productivity) and trade market access (TMA) to expenditures across space:

$$E_i = \left(\frac{R_i^{\beta} w_i^{1-\beta}}{Z_i} \right)^{-\gamma} \underbrace{\sum_{j=1}^{N} \frac{\lambda_{ij} \tau_{ij}^{-\gamma}}{\sum_{k=1}^{N} \lambda_{kj} \tau_{kj}^{-\gamma} \left(\frac{w_k^{1-\beta} R_k^{\beta}}{Z_k} \right)^{-\gamma}} E_j}_{\text{TMA}_i}. \tag{2}$$

In equilibrium, TMA affects regional labor earnings and capital returns, up to a common scaling factor, as follows:

$$\begin{aligned} R_i &= Z_i^{\frac{\gamma}{1+\gamma}} TMA_i^{\frac{1}{1+\gamma}} \left(\frac{L_i}{K_i} \right)^{\frac{\gamma(1-\beta)}{1+\gamma}} K_i^{\frac{-1}{1+\gamma}} \\ w_i &= Z_i^{\frac{\gamma}{1+\gamma}} TMA_i^{\frac{1}{1+\gamma}} \left(\frac{K_i}{L_i} \right)^{\frac{\gamma\beta}{1+\gamma}} L_i^{\frac{-1}{1+\gamma}}. \end{aligned}$$

For capital market access (KMA), we can write the supply of local capital, K_i, as a function of local interest rates and the region's access to capitalists across space:

$$K_i = R_i^{\theta-1} \underbrace{\sum_{j=1}^{N} \frac{\lambda_{ji}^k (\tau_{ji}^k)^{-\theta}}{\sum_{l=1}^{N} \lambda_{jl}^k (\tau_{jl}^k)^{-\theta} R_l} \beta E_j}_{\text{KMA}_i}. \tag{3}$$

Equating regional capital demand and supply in equilibrium, regional returns to capital can similarly be expressed as a function of KMA: $R_i = \left(\frac{\beta E_i}{\text{KMA}_i} \right)^{\frac{1}{\theta}}$.

Counterfactuals

Following previous work on the link between migrants and trade (e.g. Rauch & Trindade (2002)), we let the rates of buyer-seller matching across regions for trade in goods and capital, λ_{ij} and λ_{ij}^k respectively, be increasing functions of the bilateral stock of migrants from i residing in j. A policy shock that affects the matrix of bilateral migration costs, τ_{ij}^m, can then have knock-on effects on regional integration in trade and capital markets, through TMA_i and KMA_i, that we set out to quantify in the next Section.

Before doing so, we derive below how the canonical gains-from-trade expression in Arkolakis et al. (2012), extended to a setting with matching frictions by Eaton et al. (2023), extends also to this environment. Using 'hat' notation for proportional changes between two equilibria ($\hat{y} = \frac{y_1}{y_0}$), we can relate regional changes in real wages, $\hat{W}_i = \hat{w}_i/\hat{P}_i$, to changes in regional market access that we quantify in the next Section as follows:

$$\hat{W}_i = \hat{Z}_i^{\Theta}\, T\hat{M}A_i^{\Lambda}\, K\hat{M}A_i^{\Omega}\, \hat{\pi}_{ii}^{\Xi}\, L\hat{M}A_i^{\Gamma},$$

where $\pi_{ii} = \left(\frac{R_i^{\beta} w_i^{1-\beta}}{Z_i P_i}\right)^{-\gamma}$ is the region's own trade share, LMA is a region i's access to workers across space, $\sum_{j=1}^{N} \frac{(1-\tau_{ji}^{m})^{\phi}}{\sum_{l=1}^{N}\left(\frac{w_l}{P_l}(1-\tau_{jl}^{m})\right)^{\phi}} L_{j,0}$, and the exponents Θ, Λ, Ω, Ξ and Γ are functions of the model's key elasticities $\{\sigma, \beta, \phi, \theta, \gamma\}$, as we derive in the Appendix. Without capital, as we take $\beta \to 0$, this expression collapses to the familiar: $\lim_{\beta\to 0} \hat{W}_i = \hat{Z}_i \hat{\pi}_{ii}^{-\frac{1}{\gamma}}$.

5 Quantification of regional knock-on effects

In the final part, we interpret the estimates from Section 3 through the lens of the model to answer two additional questions. The first is to quantify the overall effect of Chinese Hukou reforms over the last decade 2010-2020 on migration market access among rural regions –defined as access of local workers to real wages in other regions: $MMA_i = \sum_{j\neq i}^{N}(1-\tau_{ij}^{m})^{\phi}\frac{w_j}{P_j}$–, and the resulting gains in trade and investment market access due to reductions in matching frictions.[23] While our empirical design above relies on relative changes across bilateral trading routes, we thus aim to quantify the overall local effects of migration cost reductions on migration and trade market access among rural origin regions.

While Hukou reforms affect migration costs unilaterally (rural-to-urban), our estimates suggest that the resulting effects on trade flows are similar in both directions. In this context, our second question is whether these additional gains from market integration due to reductions in migration barriers have been stronger among the rural origins or the urban destinations during this period of significant Hukou policy reforms in China. The answer

[23] From the model above, this definition of MMA corresponds to the numerator of a region i's migration shares to regions j: $\pi_{ij}^{m} = \frac{\left(\frac{w_j}{P_j}(1-\tau_{ij}^{m})\right)^{\phi}}{\sum_{l=1}^{N}\left(\frac{w_l}{P_l}(1-\tau_{il}^{m})\right)^{\phi}}$.

to this question will determine whether the additional gains tend to reduce or reinforce incentives for rural-to-urban migration in China over this period.

To answer these questions, we set out to quantify to what extent a rural region's changes in MMA due to Hukou reforms affect TMA and KMA across regions in China. In a first period, workers make migration decisions based on changes in bilateral migration costs, as described in the model above, until a new equilibrium is reached. In a second period, those realized migration flows can then lead to reductions in bilateral matching frictions for trade in goods and capital flows, while workers remain put during that second period. A strength of our approach here lies in the fact that we are able to directly trace the policy's effects on flows of migration, trade and investment, so that we do not have to take a stance on the underlying structural parameters to quantify changes in market access.

To quantify this scenario, we begin by estimating changes in rural regions' access to real wages across other regions due to the Hukou reforms:

$$MMA_{it} \approx \sum_{j \neq i}^{N} (1-\tau_{ijt}^{m})^{\phi} \frac{w_{j,2010}}{P_{j,2010}}, \text{where}$$
$$\left(1-\tau_{ijt}^{m}\right)^{\phi} = exp\left[-2.1 \, logDist_{ij} + .29 \text{ Eligibility}_{ijt}\right].$$

We estimate this expression for every rural county in China in both 2010 and 2020, keeping destinations' real wages fixed at the beginning in 2010.[24] The point estimates to calibrate bilateral migration frictions in the second row of the expression above ($-2.1 \, logDist_{ij}$ and $.3$ Eligibility$_{ijt}$) are taken from column 5 of Table 12 and column 1 of Table 11, respectively. We thus obtain measures of rural migration access to employment opportunities in 2010 (based on bilateral distances and Hukou reforms passed 1978-2010) and 2020 (based on distances and Hukou reforms passed 1978-2020).

To quantify the knock-on effects on regional integration in trade and investment flows, we then take first-order approximations of the TMA and KMA expressions in (2) and (3) above and fix the destination-level economic masses (E_j) to their initial levels of GDP in 2010 across counties.[25] For TMA, we quantify

[24]We use data on counties' real GDP per capita from provincial statistical yearbooks in 2010 for this purpose.

[25]Following, e.g., Donaldson & Hornbeck (2016) who show that results based on the first order are very close to the exact expression (including multilateral resistance terms).

$$TMA_{it} \approx \sum_{j \neq i}^{N} \lambda_{ijt} \tau_{ij}^{-\gamma} E_{j,2010}, \text{where}$$
$$\tau_{ij}^{-\gamma} = exp\left[-1.3\ logDist_{ij}\right] \text{ and } \lambda_{ijt} = exp\left[.15 \text{ Eligibility}_{ijt}\right],$$

the point estimates to calibrate bilateral trade costs and buyer-selling matching frictions in the second row ($-1.3\ logDist_{ij}$ and .15 Eligibility_{ijt}) are taken from the average of columns 1 and 2 of Table 12 and the average of columns 2 and 5 of Table 3, respectively. To measure changes in investment market access, the analogous expressions are:[26]

$$KMA_{it} \approx \sum_{j \neq i}^{N} \lambda_{ijt}^{k} \left(\tau_{ij}^{k}\right)^{-\theta} E_{j,2010}, \text{where}$$
$$\left(\tau_{ij}^{k}\right)^{-\theta} = exp\left[-0.9\ logDist_{ij}\right] \text{ and } \lambda_{ijt}^{k} = exp\left[.4 \text{ Eligibility}_{ijt}\right],$$

the point estimates to calibrate bilateral investment costs and investor-investment matching frictions in the second row ($-0.9\ logDist_{ij}$ and .4 Eligibility_{ijt}) are taken from the average of columns 3 and 4 of Table 12 and the average of columns 2 and 5 of Table 7, respectively.

For each of the three market access expressions above, we then proceed to quantify log changes, e.g. $T\hat{M}A_i \approx logTMA_i^{2020} - logTMA_i^{2010}$. With these county-level measures at hand, we proceed to the quantification. Figures 4-6 and Table 13 present the estimation results. The average rural origin has experienced a roughly 3% increase in overall access to real wages in other regions (MMA) due to Hukou reforms passed between 2010-2020; and a 5.5% increase in urban migration market access (limiting destinations to urban regions in both periods). Figure 4 plots those changes as a function of percentiles in 2010 GDP per capita across all counties (rural and urban) in China. Since reductions in migration barriers due to urban Hukou reforms are unilateral in nature, it is unsurprising that initially poorer (mostly rural) regions have seen significantly larger increases in migration market access compared to richer (mostly urban) counties in China.

Table 13 proceeds to quantify the average increase in trade and investment market access due to Hukou-induced changes in migration market access within the sample of roughly 1650 rural origin counties in China. We find that a 10% increase in migration market access on average leads to a 1.6% increase in trade market access and a 2% increase in investment market access. When including province fixed effects, these effects remain statistically significant, but slightly decrease in magnitude (1.4% and 1.6% respectively).

[26]Note that the β in front of E_j in the KMA expression drops out when computing log changes.

These findings suggest a significant amplification of the traditional gains from migration through knock-on effects of migration linkages on flows of trade and investment.

Turning to the second question, while the gains in migration market access are concentrated among the rural origins over this period, we find that the resulting gains in TMA and KMA were on average larger among the urban destinations compared to the rural regions. Figures 5 and 6 plot county-level changes in log TMA and KMA due to Hukou reforms as a function of initial percentiles of county-level GDP per capita in 2010, again including all regions of China (rural and urban).

In contrast to Figure 4, we find that gains in both trade and investment market access were on average larger among the initially richer (mostly urban) counties compared to the poorer rural regions. In line with these graphs, we find that the gains in TMA were on average about 45% larger among urban counties compared to rural counties over this time period, and the gains in KMA were on average about 38% larger. Through the lens of the model, this is not a general result in the context of lowering migration costs, but is related to the policy's empirical context in China: urban destinations tend to implement Hukou reforms that affect a large number of rural origins all at once –thus on average gaining access to a much larger number of markets through reductions in matching frictions compared to rural origins the other way around.

Because the knock-on gains in TMA and KMA were on average larger among the urban destinations, real wages in the model would experience upward pressure in these locations, as can be seen from the market access expressions in the previous section. The direct effect of reducing migration barriers through Hukou reforms on rural-urban migration has thus been reinforced by asymmetric gains in trade and investment access that favored the urban destinations.

Overall, our findings imply that the recent wave of Chinese Hukou reforms have (i) brought significant additional gains to rural origins and urban destinations due to increased market integration in trade and investments, and (ii) those gains reinforced the incentives for rural-urban migration because they were larger among the urban destinations.

6 Conclusion

We investigate the extent to which reductions in rural-urban migration barriers give rise to additional economic gains for rural origins and urban destinations from knock-on effects on better market integration through flows of trade and investments. To do so, we bring to bear a unique collection of microdata from China and combine this with a natural experiment that allows us to trace the effects of government policies aimed at reducing migration barriers on migration, trade, and investment flows. We find support for the hypothesis that rural-urban migration increases market integration through flows of trade and investments in both directions.

We then interpret these estimates through a spatial equilibrium model that features several of the well-known market access expressions. This allows us to quantify the implications of China's Hukou reforms over the past decade on migration market access among rural counties and the resulting knock-on effects on trade and investment market access at the regional level. We find evidence in support of a significant amplification of the traditional gains from migration through better rural-urban market integration in trade and investments.

Our findings point to a new source of gains from rural-urban migration within countries, with implications for origins, destinations and the distribution of economic activity across space. We hope that these findings can complement a rich and longstanding literature on rural-urban migration to inform ongoing policy debates on the role of migration and urbanization for economic development and spatial inequality.

References

Abadie, A., Athey, S., Imbens, G. W. & Wooldridge, J. M. (2023), 'When should you adjust standard errors for clustering?', *The Quarterly Journal of Economics* **138**(1), 1–35.

Akram, A. A., Chowdhury, S. & Mobarak, A. M. (2017), Effects of emigration on rural labor markets, Technical report, National Bureau of Economic Research.

Allen, T. (2014), 'Information frictions in trade', *Econometrica* **82**(6), 2041–2083.

Allen, T. & Arkolakis, C. (2014), 'Trade and the topography of the spatial economy', *The Quarterly Journal of Economics* **129**(3), 1085–1140.

An, L., Qin, Y., Wu, J. & You, W. (2024), 'The local labor market effect of relaxing internal migration restrictions: evidence from china', *Journal of Labor Economics* **42**(1), 161–200.

Arkolakis, C., Costinot, A. & Rodríguez-Clare, A. (2012), 'New trade models, same old gains?', *American Economic Review* **102**(1), 94–130.

Bergquist, L. F., McIntosh, C. & Startz, M. (2023), 'Search cost, intermediation, and trade: Experimental evidence from ugandan agricultural markets'.

Bernard, A. B., Moxnes, A. & Saito, Y. U. (2019), 'Production networks, geography, and firm performance', *Journal of Political Economy* **127**(2), 639–688.

Bonadio, B. (2023), 'Migrants, trade and market access', *Review of Economics and Statistics* pp. 1–45.

Bosker, M., Brakman, S., Garretsen, H. & Schramm, M. (2012), 'Relaxing hukou: Increased labor mobility and chinaâs economic geography', *Journal of Urban Economics* **72**(2-3), 252–266.

Bryan, G., Chowdhury, S. & Mobarak, A. M. (2014), 'Underinvestment in a profitable technology: The case of seasonal migration in bangladesh', *Econometrica* **82**(5), 1671–1748.

Bryan, G. & Morten, M. (2019), 'The aggregate productivity effects of internal migration: Evidence from indonesia', *Journal of Political Economy* **127**(5), 2229–2268.

Burchardi, K. B., Chaney, T. & Hassan, T. A. (2019), 'Migrants, ancestors, and foreign investments', *The Review of Economic Studies* **86**(4), 1448–1486.

Burchardi, K. B. & Hassan, T. A. (2013), 'The economic impact of social ties: Evidence from german reunification', *The Quarterly Journal of Economics* **128**(3), 1219–1271.

Cai, J., Lin, W. & Szeidl, A. (2024), Firm-to-firm referrals, Technical report, National Bureau of Economic Research.

Cai, J. & Szeidl, A. (2018), 'Interfirm relationships and business performance', *The Quarterly Journal of Economics* **133**(3), 1229–1282.

Chen, Q., Chen, Z., Liu, Z., Serrato, J. C. S. & Xu, D. (2021), Regulating conglomerates in china: Evidence from an energy conservation program, Technical report, National Bureau of Economic Research.

Cohen, L., Gurun, U. & Malloy, C. (2015), 'Resident networks and firm value', *The Journal of Finance* .

Combes, P.-P., Démurger, S. & Li, S. (2015), 'Migration externalities in chinese cities', *European Economic Review* **76**, 152–167.

Davis, S. J., Qian, M. & Zeng, W. (2025), A comprehensive gis database for china's surface transport network with implications for transport and socioeconomics research, Technical report, National Bureau of Economic Research.

Donaldson, D. & Hornbeck, R. (2016), 'Railroads and american economic growth: A âmarket accessâ approach', *The Quarterly Journal of Economics* **131**(2), 799–858.

Eaton, J. & Kortum, S. (2024), 'Technology in the global economy', *Annual Review of Economics* **16**.

Eaton, J., Kortum, S. S. & Kramarz, F. (2023), 'Firm-to-firm trade: Imports, exports, and the labor market', *Econometrica* .

Fally, T. (2015), 'Structural gravity and fixed effects', *Journal of international economics* **97**(1), 76–85.

Fan, J. (2019), 'Internal geography, labor mobility, and the distributional impacts of trade', *American Economic Journal: Macroeconomics* **11**(3), 252–288.

Gollin, D. (2014), 'The lewis model: A 60-year retrospective', *Journal of Economic Perspectives* **28**(3), 71–88.

Goyal, A. (2010), 'Information, direct access to farmers, and rural market performance in central india', *American Economic Journal: Applied Economics* **2**(3), 22–45.

Guo, N., Zhang, L., Zhang, R. & Zou, B. (2025), 'Migration restrictions and the migrant-native wage gap: The roles of wage setting and sorting', *Available at SSRN* .

Hamory, J., Kleemans, M., Li, N. Y. & Miguel, E. (2021), 'Reevaluating agricultural productivity gaps with longitudinal microdata', *Journal of the European Economic Association* **19**(3), 1522–1555.

Harris, J. R. & Todaro, M. P. (1970), 'Migration, unemployment and development: a two-sector analysis', *The American economic review* **60**(1), 126–142.

Imbert, C., Seror, M., Zhang, Y. & Zylberberg, Y. (2022), 'Migrants and firms: Evidence from china', *American Economic Review* **112**(6), 1885–1914.

Imbert, C., Ulyssea, G. et al. (2023), Rural migrants and urban informality: Evidence from brazil, Technical report, CEPR Discussion Papers.

Javorcik, B. S., Özden, Ç., Spatareanu, M. & Neagu, C. (2011), 'Migrant networks and foreign direct investment', *Journal of development economics* **94**(2), 231–241.

Kinnan, C., Wang, S.-Y. & Wang, Y. (2018), 'Access to migration for rural households', *American Economic Journal: Applied Economics* **10**(4), 79–119.

Lagakos, D. (2020), 'Urban-rural gaps in the developing world: Does internal migration offer opportunities?', *Journal of Economic perspectives* **34**(3), 174–192.

Lagakos, D., Marshall, S., Mobarak, A. M., Vernot, C. & Waugh, M. E. (2020), 'Migration costs and observational returns to migration in the developing world', *Journal of Monetary Economics* **113**, 138–154.

Lagakos, D., Mobarak, A. M. & Waugh, M. E. (2023), 'The welfare effects of encouraging rural–urban migration', *Econometrica* **91**(3), 803–837.

Lewis, W. A. (1954), 'Economic development with unlimited supplies of labour'.

McCully, B., Jaccard, T. & Albert, C. (2024), 'Immigrants, imports, and welfare: Evidence from household purchase data', *Working paper, Rockwool Foundation Berlin (RF Berlin)* .

Miyauchi, Y. (2024), 'Matching and agglomeration: Theory and evidence from japanese firm-to-firm trade', *Econometrica* **92**(6), 1869–1905.

Munshi, K. (2020), 'Social networks and migration', *Annual Review of Economics* **12**, 503–524.

Munshi, K. & Rosenzweig, M. (2016), 'Networks and misallocation: Insurance, migration, and the rural-urban wage gap', *American Economic Review* **106**(01), 46–98.

Ngai, L. R., Pissarides, C. A. & Wang, J. (2019), 'Chinaâs mobility barriers and employment allocations', *Journal of the European Economic Association* **17**(5), 1617–1653.

Parsons, C. & Vézina, P.-L. (2018), 'Migrant networks and trade: The vietnamese boat people as a natural experiment', *The Economic Journal* **128**(612), F210–F234.

Rauch, J. E. & Trindade, V. (2002), 'Ethnic chinese networks in international trade', *Review of Economics and Statistics* **84**(1), 116–130.

Redding, S. J. (2016), 'Goods trade, factor mobility and welfare', *Journal of International Economics* **101**, 148–167.

Roth, J. & Chen, J. (2023), 'Logs with zeros? some problems and solutions', *Quarterly Journal of Economics* .

Sieg, H., Yoon, C. & Zhang, J. (2023), 'The impact of local fiscal and migration policies on human capital accumulation and inequality in china', *International Economic Review* **64**(1), 57–93.

Silva, J. S. & Tenreyro, S. (2006), 'The log of gravity', *The Review of Economics and statistics* **88**(4), 641–658.

Sotelo, S. (2019), 'Practical aspects of implementing the multinomial pml estimator', *Ann Arbor: University of Michigan, mimeo* .

State Council (2015), Several policy measures on vigorously promoting mass entrepreneurship and innovation, Guo fa [2015] no. 32, State Council, Beijing. Accessed 22 July 2025. **URL:** *http://www.gov.cn/zhengce/content/2015-06/16/content$_9$950.htm*

Tian, Y. (2024), 'International trade liberalization and domestic institutional reform: Effects of wto accession on chinese internal migration policy', *Review of Economics and Statistics* pp. 794–813.

Tombe, T. & Zhu, X. (2019), 'Trade, migration, and productivity: A quantitative analysis of china', *American Economic Review* **109**(5), 1843–1872.

Weidner, M. & Zylkin, T. (2021), 'Bias and consistency in three-way gravity models', *Journal of International Economics* **132**, 103513.

Wooldridge, J. M. (1999), 'Quasi-likelihood methods for count data', *Handbook of applied econometrics volume 2: Microeconomics* pp. 321–368.

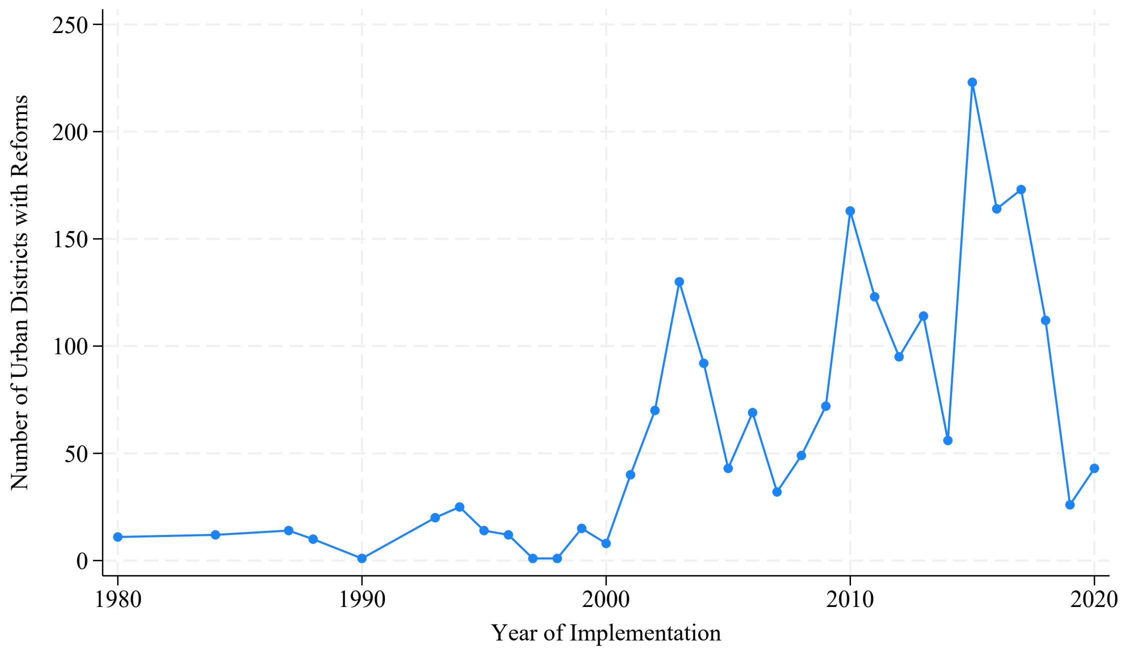
Young, A. (2013), 'Inequality, the urban-rural gap, and migration', *The Quarterly Journal of Economics* **128**(4), 1727–1785.

Zhang, J., Wang, R. & Lu, C. (2019), 'A quantitative analysis of hukou reform in chinese cities: 2000–2016', *Growth and Change* **50**(1), 201–221.

7 Figures and tables

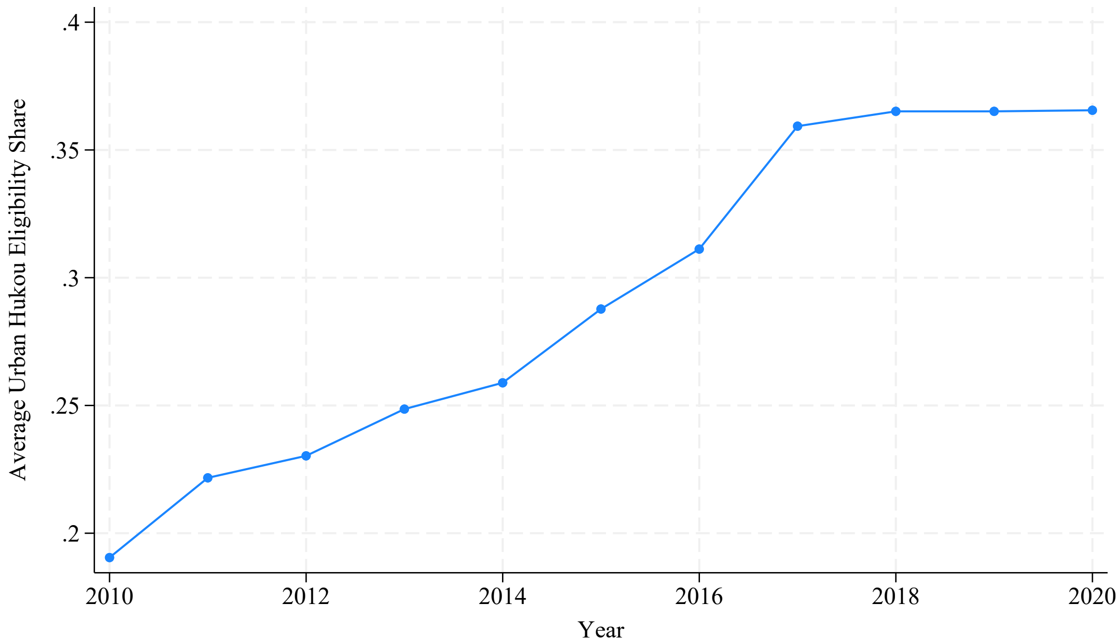
Figures

Figure 1: Hukou reforms



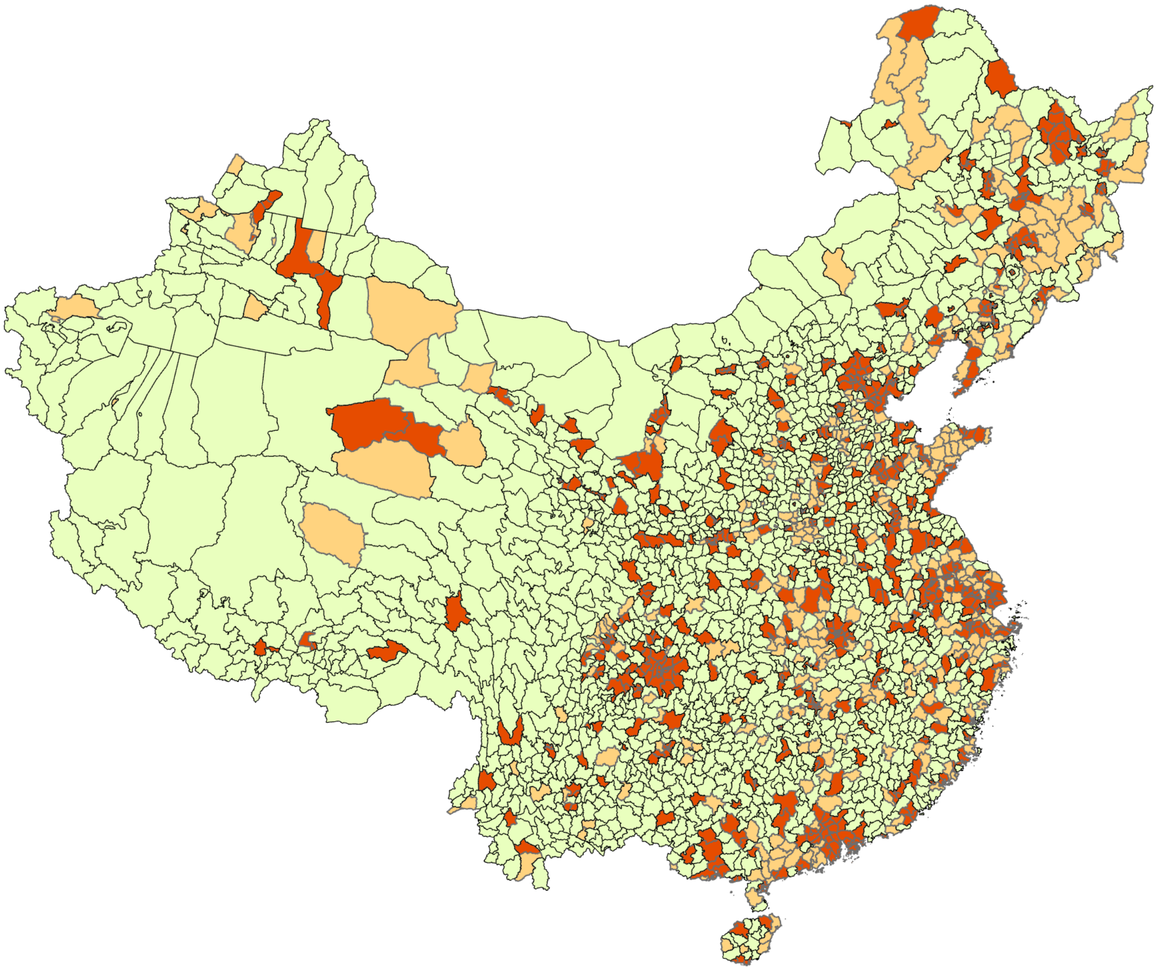
Notes: Figure plots the number of urban county units with any Hukou reform in a given year over time. See Section 2 for further discussion.

Figure 2: Hukou eligibility



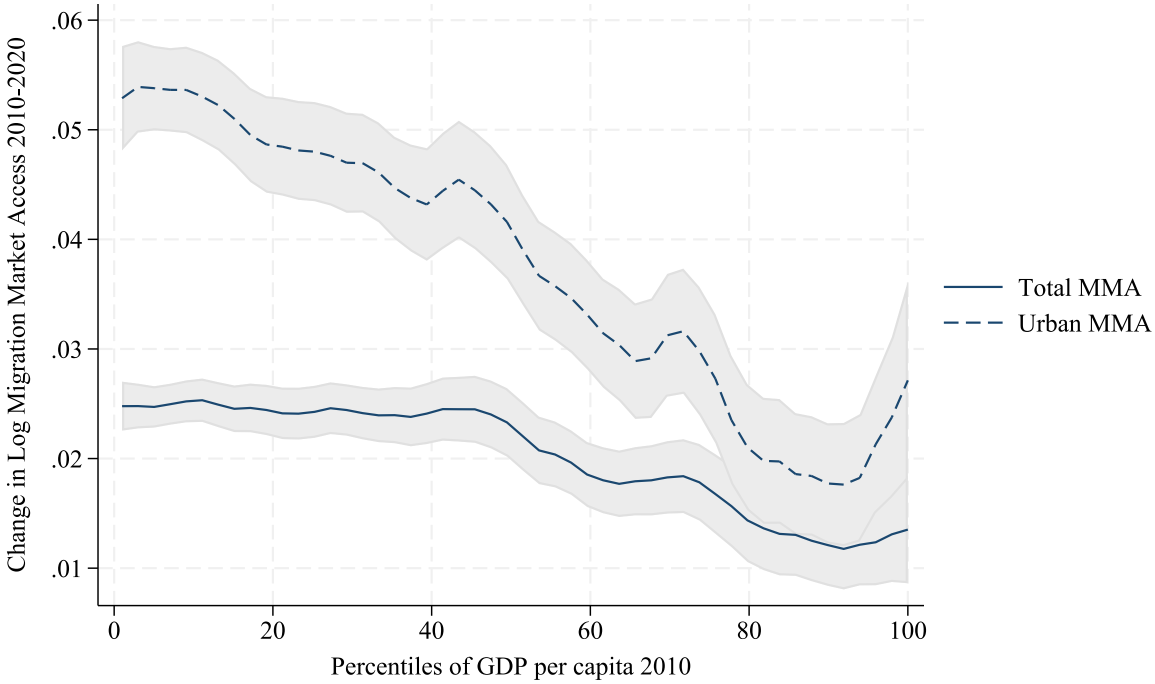
Notes: Figure plots the share of rural origins' working-age populations, fixed in 2010, who are eligible for urban Hukou registration across urban destinations, averaged across bilateral pairs over time. See Section 2 for further discussion.

Figure 3: Geography



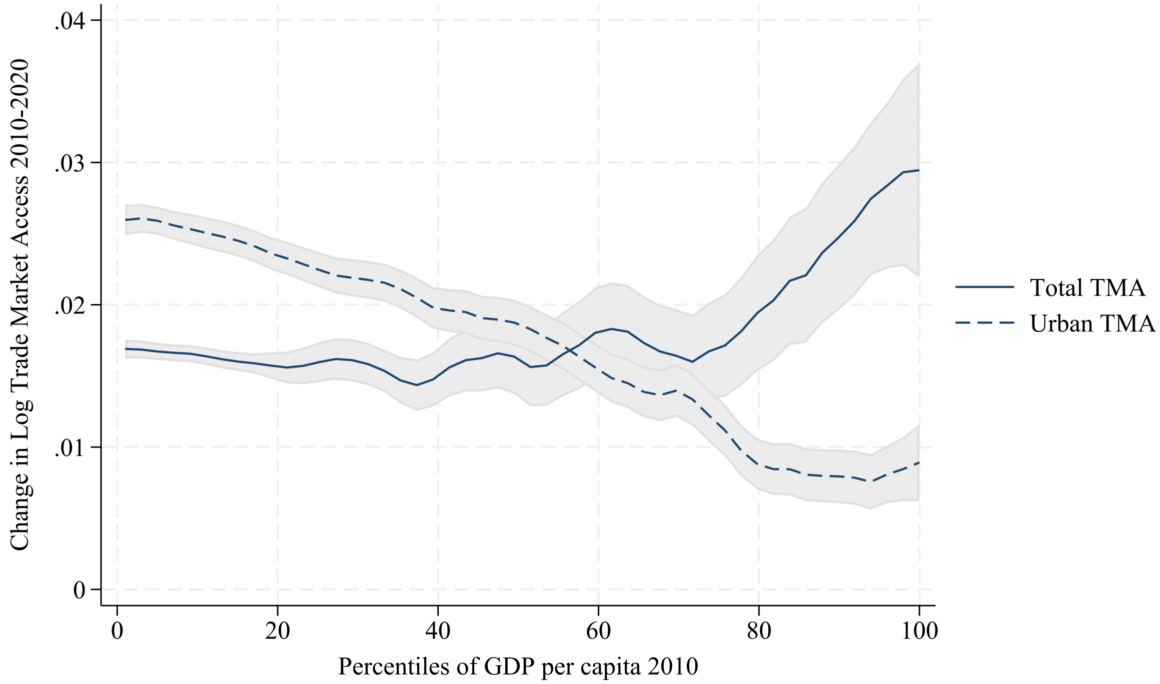
Notes: Figure shows 2661 time-consistent county-level units in mainland China covered in our database. Units in red are urban districts. Units in orange are county-level cities. Units in green are rural counties. See Section 2 for further discussion.

Figure 4: Changes in migration market access due to Hukou reforms 2010-2020



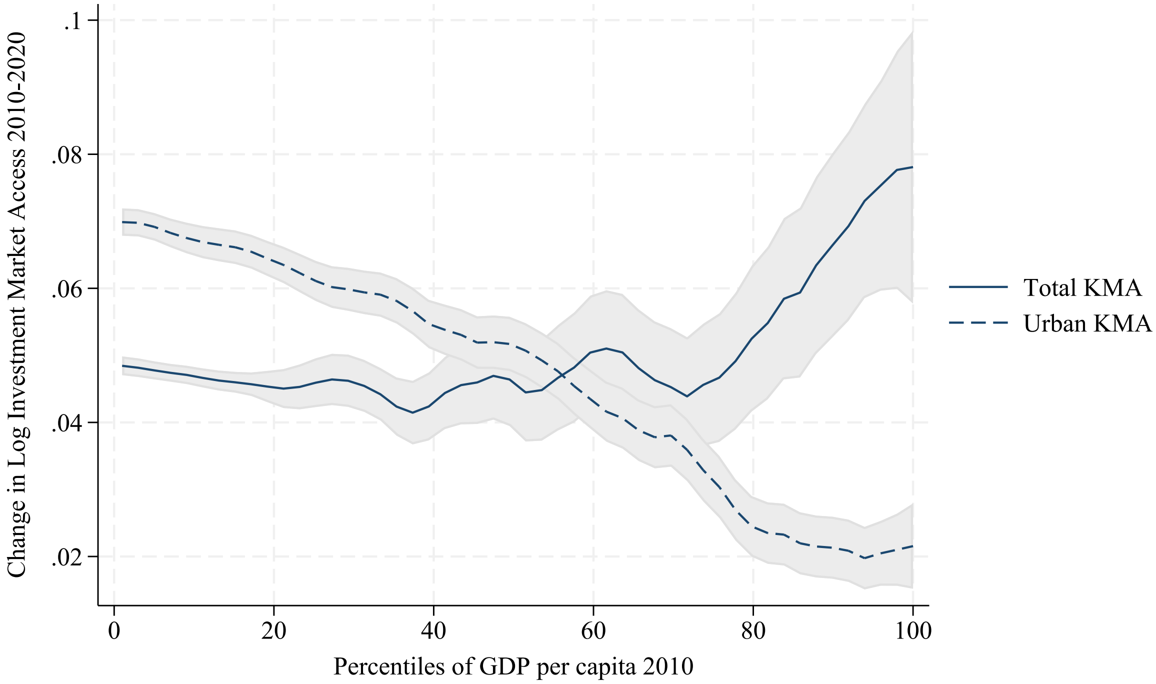
Notes: Figure plots estimates from local polynomial regression of estimated changes in migration market access (MMA) over 2010-2020 on initial percentiles of GDP per capita in 2010 for all (rural and urban) county units. Shaded area indicates 95% confidence intervals. See Section 5 for further discussion.

Figure 5: Changes in trade market access due to Hukou reforms 2010-2020



Notes: Figure plots estimates from local polynomial regression of estimated changes in trade market access (TMA) over 2010-2020 on initial percentiles of GDP per capita in 2010 for all (rural and urban) county units. Shaded area indicates 95% confidence intervals. See Section 5 for further discussion.

Figure 6: Changes in investment market access due to Hukou reforms 2010-2020



Notes: Figure plots estimates from local polynomial regression of estimated changes in investment market access (KMA) over 2010-2020 on initial percentiles of GDP per capita in 2010 for all (rural and urban) county units. Shaded area indicates 95% confidence intervals. See Section 5 for further discussion.

Tables

Table 1: Summary statistics of county-to-county-by-sector flows

	(1)	(2)
Panel A Flows of Trade 2014 & 18		
	Rural Import	Rural Export
Av. No. County-to-county-by-sector Pairs	4,239,706	3,879,593
No. Buyer-seller Matches (in a Given Year)	39,038,030	35,082,105
Average Values in a Pair (10,000 RMB)	1,069.5	1,216.8
SD	(53,377.7)	(56,041.6)
Panel B Flows of Investment 2014 & 19		
	Rural Investment Inflows	Rural Investment Outflows
Av. No. County-to-county-by-sector Pairs	1,667,550	1,302,714
No. Investor-Investee Matches (in a Given Year)	3,565,381	2,789,801
Average Values in a Pair (10,000 RMB)	6,516.9	5,008.3
SD	(187,640.4)	(208,486.6)
Panel C Flows of Investment 2010 & 14		
	Rural Investment Inflows	Rural Investment Outflows
Av. No. County-to-county-by-sector Pairs	1,101,401	810,720
No. Investor-Investee Matches (in a Given Year)	2,363,461	1,763,765
Average Values in a Pair (10,000 RMB)	10,161.2	8,265.8
SD	(240,096.9)	(274,914.1)

Notes: Table presents summary statistics for rural imports, rural exports, rural investment inflows, and rural investment outflows to or from all Chinese counties in the years 2014 and 2018 (trade flows) or 2019 (investment flows). First row under each panel counts the number of county pairs by sector with positive flows, then displays the average across the two sample years. Second row counts the total number of all unique buyer-seller (or investor-investment) matches across county pairs and sectors in a year (with at least one transaction in that year), then displays the average of that sum across the two sample years. Third row displays the mean transaction value across county-county-sector pairs and sample years. See Section 2.2 for further discussion.

Table 2: Sectoral shares of county-to-county flows

Sectoral Share Distribution	Rural Import 2014 & 18		Rural Export 2014 & 18		Rural Inflow 2014 & 19		Rural Outflow 2014 & 19	
	(1) Value	(2) #Pairs	(3) Value	(4) #Pairs	(5) Value	(6) #Pairs	(7) Value	(8) #Pairs
Accommodation and catering	0.025	0.067	0.025	0.063	0.041	0.049	0.041	0.048
Agriculture, forestry, animal husbandry and fishery	0.015	0.017	0.016	0.018	0.391	0.157	0.367	0.115
Construction	0.041	0.022	0.036	0.021	0.025	0.049	0.028	0.051
Education	0.001	0.001	0.001	0.001	0.002	0.005	0.002	0.005
Electricity, heat, gas, and water production and supply	0.023	0.009	0.026	0.010	0.002	0.002	0.002	0.002
Entertainment, sports and culture	0.005	0.007	0.004	0.006	0.007	0.011	0.007	0.012
Finance	0.011	0.017	0.010	0.013	0.004	0.004	0.005	0.005
Health and social work	0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.002
Information transmission, software and IT services	0.017	0.033	0.016	0.024	0.016	0.029	0.017	0.035
Leasing and business services	0.051	0.071	0.053	0.067	0.053	0.101	0.060	0.110
Manufacturing	0.265	0.204	0.290	0.251	0.059	0.067	0.061	0.065
Mining	0.019	0.002	0.022	0.003	0.001	0.001	0.001	0.001
Real estate	0.007	0.011	0.008	0.010	0.010	0.017	0.010	0.017
Residential services, repairs and other services	0.018	0.025	0.018	0.023	0.026	0.030	0.025	0.029
Technical services and scientific research	0.037	0.059	0.030	0.051	0.019	0.040	0.021	0.047
Transportation, warehousing and postal services	0.029	0.033	0.032	0.038	0.014	0.020	0.014	0.020
Utilities, water conservancy and environmental services	0.002	0.002	0.002	0.002	0.002	0.004	0.002	0.004
Wholesale and retail trade	0.405	0.395	0.394	0.380	0.272	0.328	0.276	0.341
Other	0.027	0.025	0.017	0.018	0.056	0.084	0.060	0.090

Notes: Table presents sectoral shares of unique firm-to-firm pairs and transaction values across 19 sectors for rural exports, imports, investment inflows and outflows to or from all Chinese counties. For trade flows (imports and exports), the shares are averaged across the two sample years 2014 and 2018. For investment flows (inflows and outflows), they are averaged for the sample years 2014 and 2019. See Section 2.2 for further discussion.

Table 3: Migration barriers and rural exports and imports

VARIABLES	(1) Exports 2014-2018	(2) Exports 2014-2018	(3) Exports 2014-2018	(4) Imports 2014-2018	(5) Imports 2014-2018	(6) Imports 2014-2018
Eligibility Share	0.184***	0.152***	0.122**	0.195***	0.145***	0.140***
	(0.050)	(0.052)	(0.051)	(0.051)	(0.053)	(0.047)
Observations	81,648,857	9,221,476	8,292,460	85,153,850	9,977,502	9,051,472
Orig-Dest FE	✓	.	.	✓	.	.
Orig-Year FE	✓	.	.	✓	.	.
Dest-Year FE	✓	.	.	✓	.	.
Orig-Dest-Sec FE	.	✓	✓	.	✓	✓
Orig-Sec-Year FE	.	✓	✓	.	✓	✓
Dest-Sec-Year FE	.	✓	✓	.	✓	✓
Dest-Sec-Year-Dist FE	.	.	✓	.	.	✓
RU-Year-Dist FE	.	.	✓	.	.	✓

Notes: Table reports the PPML estimates from specification (1) for bilateral flows from or to rural counties. Unit of observation is a county-to-county-by-sector pair in a given year of data. Trade flows are reported for 34 sectors. We estimate the long difference between 2014 and 2018. Standard errors are clustered at the level of destination-by-year. * 10%, ** 5%, *** 1% significance levels. See Section 3 for further discussion.

Table 4: Migration barriers and numbers of buyers, sellers and pairs

VARIABLES	(1) Exports N Pairs 2014-2018	(2) Exports N Sellers 2014-2018	(3) Exports N Buyers 2014-2018	(4) Imports N Pairs 2014-2018	(5) Imports N Sellers 2014-2018	(6) Imports N Buyers 2014-2018
Eligibility Share	0.274***	0.100***	0.232***	0.275***	0.064***	0.241***
	(0.064)	(0.022)	(0.050)	(0.061)	(0.022)	(0.049)
Observations	9,221,476	9,221,476	9,221,476	9,977,502	9,977,502	9,977,502
Orig-Dest-Sec FE	✓	✓	✓	✓	✓	✓
Orig-Sec-Year FE	✓	✓	✓	✓	✓	✓
Dest-Sec-Year FE	✓	✓	✓	✓	✓	✓

Notes: Table reports the PPML estimates from specification (1) for bilateral flows from or to rural counties. Unit of observation is a county-to-county-by-sector pair in a given year of data. Trade flows are reported for 34 sectors. We estimate the long difference between 2014 and 2018. Standard errors are clustered at the level of destination-by-year. * 10%, ** 5%, *** 1% significance levels. See Section 3 for further discussion.

Table 5: Channels and robustness

VARIABLES	(1) Existing link Exports 2014-2018	(2) Existing link Imports 2014-2018	(3) Travel time Average 2014-2019	(4) Travel time Minimum 2014-2019	(5) Formal only Exports 2014-2018	(6) Formal Only Imports 2014-2018
Eligibility Share	-0.045	0.053	0.060	0.033	0.215***	0.238***
	(0.212)	(0.154)	(0.130)	(0.294)	(0.050)	(0.051)
Observations	158,832	173,892	8,044,706	8,044,706	8,712,838	9,233,808
Orig-Dest FE	.	.	✓	✓	.	.
Orig-Year FE	.	.	✓	✓	.	.
Dest-Year FE	.	.	✓	✓	.	.
Orig-Dest-Sec FE	✓	✓	.	.	✓	✓
Orig-Sec-Year FE	✓	✓	.	.	✓	✓
Dest-Sec-Year FE	✓	✓	.	.	✓	✓

Notes: Table reports estimates from specification (1) for bilateral flows from or to rural counties. Columns 1 and 2 report effects from PPML estimation on rural exports and imports among firms with pre-existing migration linkages on a given bilateral route. Columns 3 and 4 report effects from a fixed-effects specification with average or minimum travel times (in hours) between bilateral pairs on the left-hand side, computed across four transport networks: high-speed rail, traditional rail, roads and highways and controlled-access motorways in 2014 and 2019. Columns 5 and 6 report effects from PPML estimation on rural exports or imports among formal firms that already reported their transactions in the 2014 VAT data. Standard errors are clustered at the level of destination-by-year. * 10%, ** 5%, *** 1% significance levels. See Section 3 for further discussion.

Table 6: Heterogeneity across regions, pairs and sectors - Rural exports

VARIABLES	(1) Exports All 2014-2018	(2) Exports All 2014-2018	(3) Exports Agri 2014-2018	(4) Exports Manuf 2014-2018	(5) Exports All Serv 2014-2018	(6) Exports Traded Serv 2014-2018
Eligibility Share	0.894***	0.221	0.661***	0.111***	0.156**	0.551***
	(0.126)	(0.154)	(0.135)	(0.039)	(0.067)	(0.091)
Elig x Pctl Distance		0.005***				
		(0.001)				
Elig x Education		-0.013				
		(0.015)				
Elig x Pctl Output/Pop		-0.001				
		(0.001)				
Eligibility Share^2	-0.757***					
	(0.118)					
Observations	9,221,476	8,037,614	315,084	3,968,386	4,608,800	2,023,166
Orig-Dest-Sec FE	✓	✓	✓	✓	✓	✓
Orig-Sec-Year FE	✓	✓	✓	✓	✓	✓
Dest-Sec-Year FE	✓	✓	✓	✓	✓	✓

Notes: Table reports the PPML estimates from specification (1) with additional interaction terms as indicated for bilateral flows from or to rural counties. Unit of observation is a county-to-county-by-sector pair in a given year of data. We estimate the long difference between 2014 and 2018. Standard errors are clustered at the level of destination-by-year. * 10%, ** 5%, *** 1% significance levels. See Section 3 for further discussion.

Table 7: Migration barriers and rural investment inflows and outflows

VARIABLES	(1) Inflows 2014-2019	(2) Inflows 2014-2019	(3) Inflows 2014-2019	(4) Outflows 2014-2019	(5) Outflows 2014-2019	(6) Outflows 2014-2019
Eligibility Share	0.272**	0.437***	0.553***	0.274**	0.344**	0.206
	(0.110)	(0.140)	(0.187)	(0.127)	(0.167)	(0.211)
Observations	41,134,202	4,804,286	4,130,456	35,981,763	3,826,712	3,045,392
Orig-Dest FE	✓	.	.	✓	.	.
Orig-Year FE	✓	.	.	✓	.	.
Dest-Year FE	✓	.	.	✓	.	.
Orig-Dest-Sec FE	.	✓	✓	.	✓	✓
Orig-Sec-Year FE	.	✓	✓	.	✓	✓
Dest-Sec-Year FE	.	✓	✓	.	✓	✓
Dest-Sec-Year-Dist FE	.	.	✓	.	.	✓
RU-Year-Dist FE	.	.	✓	.	.	✓

Notes: Table reports the PPML estimates from specification (1) for bilateral flows from or to rural counties. Unit of observation is a county-to-county-by-sector pair in a given year of data. Investment flows are reported for 19 sectors. We estimate the long difference between 2014 and 2019. Standard errors are clustered at the level of destination-by-year. * 10%, ** 5%, *** 1% significance levels. See Section 3 for further discussion.

Table 8: Placebo test on pre-trends

VARIABLES	(1) Inflows 2014-2019	(2) Inflows 2010-2014	(3) Outflows 2014-2019	(4) Outflows 2010-2014
Eligibility Share (2014-2019)	0.437***	0.062	0.344**	0.057
	(0.140)	(0.165)	(0.167)	(0.176)
Observations	4,804,286	3,125,814	3,826,712	2,399,398
Orig-Dest-Sec FE	✓	✓	✓	✓
Orig-Sec-Year FE	✓	✓	✓	✓
Dest-Sec-Year FE	✓	✓	✓	✓

Notes: Table reports the PPML estimates from specification (1) for bilateral flows from or to rural counties. Unit of observation is a county-to-county-by-sector pair in a given year of data. Investment flows are reported for 19 sectors. In Columns 1 and 3, we estimate the long difference between 2014 and 2019. In Columns 2 and 4 we report the long difference between 2010 and 2014. Standard errors are clustered at the level of destination-by-year. * 10%, ** 5%, *** 1% significance levels. See Section 3 for further discussion.

Table 9: Migration barriers and numbers of investments, investors and pairs

VARIABLES	(1) Inflows N Pairs 2014-2019	(2) Inflows N Invested 2014-2019	(3) Inflows N Investor 2014-2019	(4) Outflows N Pairs 2014-2019	(5) Outflows N Invested 2014-2019	(6) Outflows N Investor 2014-2019
Eligibility Share	0.317***	0.240***	0.313***	0.341***	0.257***	0.339***
	(0.018)	(0.017)	(0.017)	(0.020)	(0.019)	(0.018)
Observations	4,804,286	4,804,286	4,804,286	3,826,712	3,826,712	3,826,712
Orig-Dest-Sec FE	✓	✓	✓	✓	✓	✓
Orig-Sec-Year FE	✓	✓	✓	✓	✓	✓
Dest-Sec-Year FE	✓	✓	✓	✓	✓	✓

Notes: Table reports the PPML estimates from specification (1) for bilateral flows from or to rural counties. Unit of observation is a county-to-county-by-sector pair in a given year of data. Investment flows are reported for 19 sectors. We estimate the long difference between 2014 and 2019. Standard errors are clustered at the level of destination-by-year. * 10%, ** 5%, *** 1% significance levels. See Section 3 for further discussion.

Table 10: Heterogeneity across regions, pairs and sectors - Rural investment inflows

VARIABLES	(1) Inflows All 2014-2019	(2) Inflows All 2014-2019	(3) Inflows Agri 2014-2019	(4) Inflows Manuf 2014-2019	(5) Inflows Wholesale 2014-2019	(6) Inflows All Serv 2014-2019	(7) Inflows Traded Serv 2014-2019
Eligibility Share	0.415	1.861***	0.268	0.394	0.710***	0.639***	0.602***
	(0.378)	(0.662)	(0.406)	(0.408)	(0.196)	(0.138)	(0.193)
Elig x Pctl Distance		-0.002					
		(0.002)					
Elig x Education		-0.118					
		(0.075)					
Elig x Pctl Output/Pop		-0.002					
		(0.002)					
Eligibility Share^2	0.179						
	(0.347)						
Observations	4,804,286	4,205,918	740,920	377,292	1,141,268	3,254,878	1,392,214
Orig-Dest-Sec FE	✓	✓	✓	✓	✓	✓	✓
Orig-Sec-Year FE	✓	✓	✓	✓	✓	✓	✓
Dest-Sec-Year FE	✓	✓	✓	✓	✓	✓	✓

Notes: Table reports the PPML estimates from specification (1) with additional interaction terms as indicated for bilateral flows from or to rural counties. Unit of observation is a county-to-county-by-sector pair in a given year of data. Investment flows are reported for 19 sectors. We estimate the long difference between 2014 and 2019. Standard errors are clustered at the level of destination-by-year. * 10%, ** 5%, *** 1% significance levels. See Section 3 for further discussion.

Table 11: Migration barriers and migration

VARIABLES	(1) Migration 2005-10 & 2011-15	(2) Migration 2005-10 & 2011-15	(3) Migrant Stock 2005-2010	(4) Migrant Stock 2005-2010
Eligibility Share	0.285***	0.221**	0.123**	0.173***
	(0.083)	(0.099)	(0.058)	(0.056)
Observations	226,186	201,818	256,070	229,332
Orig-Dest FE	✓	✓	✓	✓
Orig-Census FE	✓	✓	.	.
Dest-Census FE	✓	.	.	.
Dest-Dist-Census FE	.	✓	.	.
RU-Dist-Census FE	.	✓	.	.
Orig-Year FE	.	.	✓	✓
Dest-Year FE	.	.	✓	.
Dest-Dist-Year FE	.	.	.	✓
RU-Dist-Year FE	.	.	.	✓

Notes: Table reports the PPML estimates from specification (1) for the sum of migration flows from rural counties during the period 2005-2010 and 2011-2015 in columns 1 and 2. Unit of observation is a county-to-county pair in a given year of data. Columns 3 and 4 use the bilateral stock of migrants reported in 2005 and 2010. Standard errors are clustered at the level of destination-by-year. * 10%, ** 5%, *** 1% significance levels. See Section 3 for further discussion.

Table 12: Benchmarking effects to distance elasticities

VARIABLES	(1) Exports 2014 & 2018	(2) Imports 2014 & 2018	(3) Inflows 2014 & 2019	(4) Outflows 2014 & 2019	(5) Migration 2005-10 & 2011-15
Log Distance	-1.288***	-1.300***	-0.758***	-0.990***	-2.104***
	(0.009)	(0.010)	(0.013)	(0.017)	(0.007)
Observations	181,304,014	191,995,489	119,321,909	117,850,582	7,934,477
Orig-Dest-Sec FE	.	.	.	.	.
Orig-Sec-Year FE	✓	✓	✓	✓	.
Dest-Sec-Year FE	✓	✓	✓	✓	.
Orig-Dest FE	.	.	.	.	.
Orig-Year FE	.	.	.	.	✓
Dest-Year FE	.	.	.	.	✓

Notes: Table reports the PPML estimates from specification (1) for bilateral flows from or to rural counties, but not including origin-destination(-sector) fixed effects so that we can estimate the bilateral distance elasticites for different types of flows. Origin-time-sector and destination-time-sector fixed effects are included (for migration only origin-time and destination-time). Standard errors are clustered at the level of origin-destination pairs in this specification. * 10%, ** 5%, *** 1% significance levels. See Section 3 for further discussion.

Table 13: Effect of migration market access on trade and investment market access

VARIABLES	(1) Rural Counties Δ Log TMA 2010-2020	(2) Rural Counties Δ Log TMA 2010-2020	(3) Rural Counties Δ Log KMA 2010-2020	(4) Rural Counties Δ Log KMA 2010-2020
Δ Log MMA	0.157***	0.141***	0.201***	0.163***
	(0.016)	(0.013)	(0.031)	(0.020)
Observations	1,663	1,663	1,663	1,663
R-squared	0.631	0.812	0.453	0.777
Province FE	.	✓	.	✓

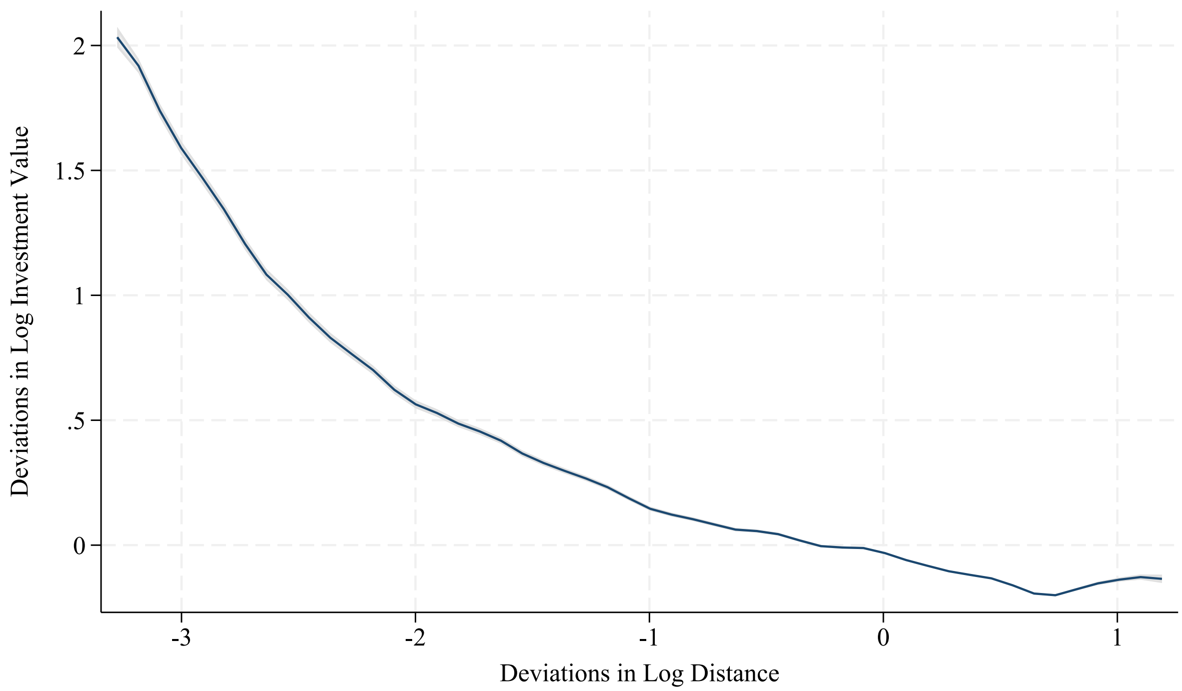
Notes: Table presents regression results of estimated changes in log trade (TMA) and investment market access (KMA) on estimated changes in log migration market access (MMA) due to Hukou reforms passed 2010-2020. The estimation sample is based on roughly 1650 rural counties. Unit of observation is a county. Standard errors are clustered at the level of provinces. * 10%, ** 5%, *** 1% significance levels. See Section 5 for further discussion.

Appendix - For Online Publication: Rural-Urban Migration and Market Integration

Appendix 1 presents additional figures and tables referenced in the main text. Appendix 2 provides additional model derivations.

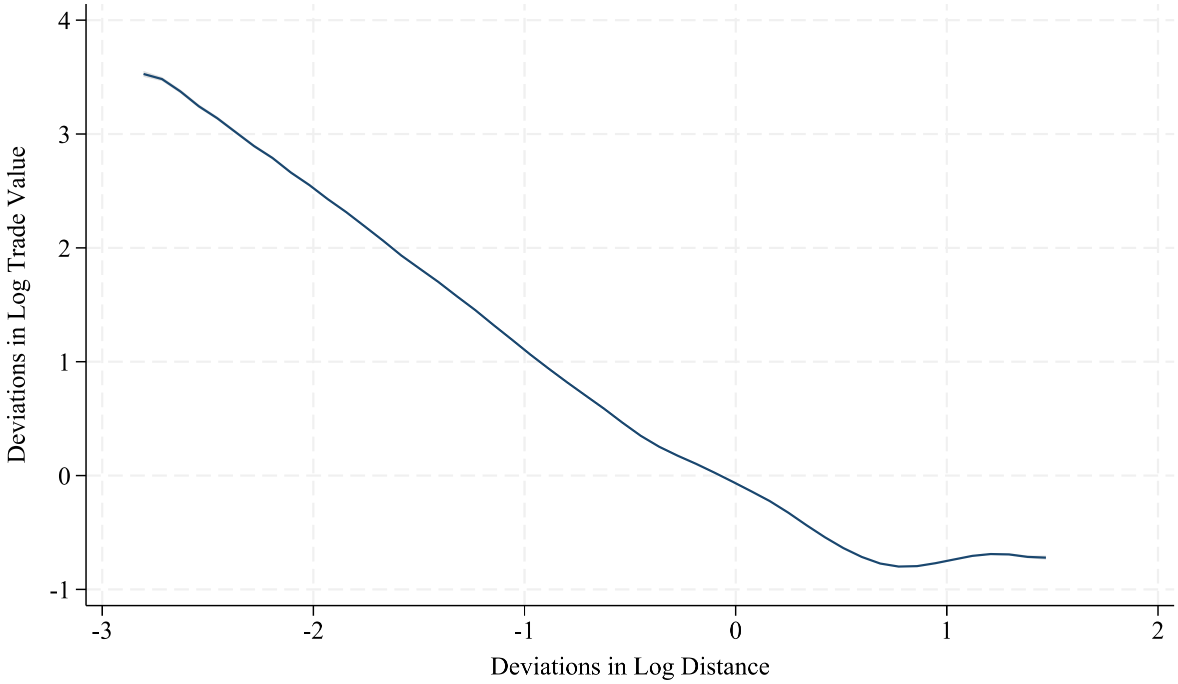
Appendix 1: Additional figures and tables

Figure A.1: Distance decay in investment flows



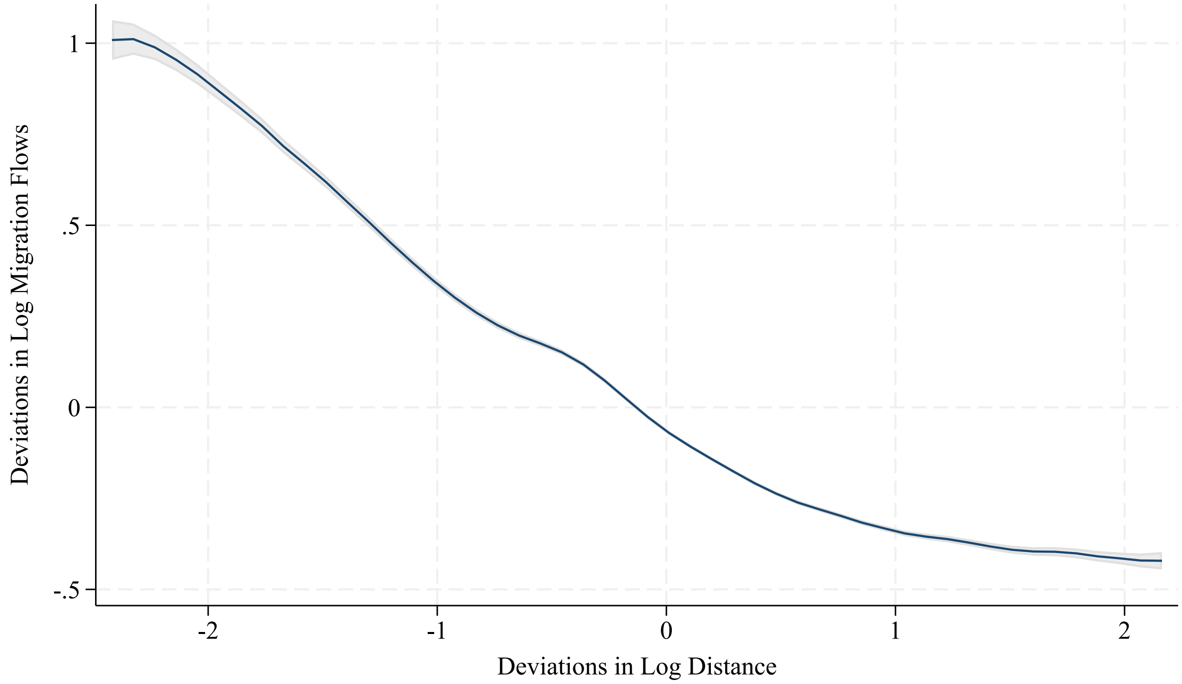
Notes: Figure plots estimates from local polynomial regression of deviations (residuals) in the logarithm of investment flows on deviations in the logarithm of county-to-county distances. Deviations are log residuals after conditioning on origin-sector-time and destination-sector-time fixed effects for both axes. The sample includes all bilateral flows in 2014 and 2019. The shaded area indicates 95% confidence intervals. See Section 2 for further discussion.

Figure A.2: Distance decay in trade flows



Notes: Figure plots estimates from local polynomial regression of deviations (residuals) in the logarithm of trade flows on deviations in the logarithm of county-to-county distances. Deviations are log residuals after conditioning on origin-sector-time and destination-sector-time fixed effects for both axes. The sample includes all bilateral flows in 2014 and 2018. The shaded area indicates 95% confidence intervals. See Section 2 for further discussion.

Figure A.3: Distance decay in migration flows



Notes: Figure plots estimates from local polynomial regression of deviations (residuals) in the logarithm of migration flows on deviations in the logarithm of county-to-county distances. Deviations are log residuals after conditioning on origin-time and destination-time fixed effects for both axes. The sample includes bilateral flows over the period 2005-2010 and 2011-2015. The shaded area indicates 95% confidence intervals. See Section 2 for further discussion.

Appendix 2: Model appendix

From potential to active firms and capitalists

Producers

First, let's think about the number of active producers located in a given market i. Consider a producer in i who can ship to j with cost c in the home market. The number of retailers who buy the producer's variety is distributed Poisson with parameter

$$\eta_{ij}(\tau_{ij}c) = \underbrace{\lambda_{ij}^{p}}_{\text{Likelihood of matching}} \underbrace{1}_{\text{Unit measure of potential retailers}} \underbrace{\left(1 - G_j^{(1)}(\tau_{ij}c)\right)}_{\text{Likelihood of a sale}}.$$

Here, $G_j^{(1)}(\tau_{ij}c)$ is the distribution of the minimum price in j which is

$$
\begin{aligned}
G_j^{(1)}(\tau_{ij}c) =& 1 - \exp(-\rho_j^p(\tau_{ij}c)) \\
=& 1 - \exp\left(-\sum_{k=1}^{N}\lambda_{kj}\left(\frac{\tau_{kj}w_k^{1-\beta}R_k^{\beta}}{Z_k}\right)^{-\gamma}(\tau_{ij}c)^{\gamma}\right) \\
=& 1 - \exp\left(-P_j^{-\gamma}\tau_{ij}^{\gamma}c^{\gamma}\right).
\end{aligned}
$$

Hence, for a producer with cost c, the number of buyers from anywhere is distributed Poisson with rate:

$$\eta_i(c) = \sum_{j=1}^{N}\lambda_{ij}\exp\left(-P_j^{-\gamma}\tau_{ij}^{\gamma}c^{\gamma}\right).$$

Thus, the measure of active intermediates in i is:

$$F_i = \int_0^{\infty}\left(1 - \exp\left(-\eta_i(c)\right)\right)d\mu_{ii}(c).$$

Retailers

Now, how many varieties are sold in i? Consider the producer in j who can ship to i with cost c in the home market. The number of producers from j with production cost c in their home market that a retailer in i matches with is distributed Poisson with rate

$$\zeta_{ji}(\tau_{ji}c) = \underbrace{\lambda_{ji}^p}_{\text{Likelihood of matching}} \underbrace{d\mu_{jj}(c)}_{\text{Measure of potential producers with } c}.$$

Hence, the number of matches between a retailer in i and a producer from j with any production cost in its home market is distributed Poisson with rate

$$\zeta_{ji} = \int_0^{\infty}\lambda_{ji}^p d\mu_{jj}(c) = \lambda_{ji}^p.$$

Likewise, the number of matches between a retailer in i and producers from anywhere is

$$\zeta_i = \sum_{j=1}^{N}\zeta_{ji},$$

and the number of active retailers is:

$$\mathcal{S}_i = 1 - \exp(-\zeta_i).$$

Capitalists

The number of capitalists in i who finance firms in j at or below rate R should be distributed Poisson with parameter

$$\gamma_{ij}(R) = \underbrace{\lambda_{ij}^k}_{\text{Likelihood of matching}} \underbrace{F_j}_{\text{Measure of active producers}} \underbrace{1}_{\text{Likelihood of a sale}}.$$

Hence, the number of active capitalists in i is distributed Poisson with parameter

$$\gamma_i = \sum_{j=1}^{N} \lambda_{ij}^{k} F_j \exp\left(-\sum_{l=1}^{N} \lambda_{lj}^{k} F_j\right).$$

Thus, the measure of active capitalists is

$$\mathcal{K}_i = \int_0^\infty \left(1 - \exp\left(-\gamma_i\right)\right) d\mu_{ii}^{k}(R) = 1 - \exp\left(-\gamma_i\right).$$

Relating hat-changes in real wages to market access expressions

Using expressions for the numerator and denominator in $W_i = \frac{w_i}{P_i}$ that we lay out in Section 4, we can solve for hat changes in real wages as a function of market access changes as follows:

$$\hat{W}_i = \hat{Z}_i^{\frac{1+\psi}{1+\psi(1-\beta)+\gamma\beta}} T\hat{M}A_i^{\frac{1}{\gamma}\frac{\psi\beta}{1+\beta\gamma+\psi(1-\beta)}} K\hat{M}A_i^{\frac{1+\gamma}{\gamma(\theta-1)}\frac{\psi\beta}{1+\beta\gamma+\psi(1-\beta)}} \hat{\pi}_{ii}^{-\frac{1}{\gamma}\frac{1+\psi(1-\beta)}{1+\beta\gamma+\psi(1-\beta)}} L\hat{M}A_i^{\frac{-\beta}{1+\beta\gamma+\psi(1-\beta)}},$$

where $\psi = \frac{\gamma(\theta-1)}{\gamma+\theta}$, $\pi_{ii} = \left(\frac{R_i^{\beta} w_i^{1-\beta}}{Z_i P_i}\right)^{-\gamma}$ is the region's own trade share and LMA is a region i's access to workers across space: $\sum_{j=1}^{N} \frac{(1-\tau_{ji}^{m})^{\phi}}{\sum_{l=1}^{N}\left(\frac{w_l}{P_l}(1-\tau_{jl}^{m})\right)^{\phi}} L_{j,0}$.