Heterogeneous migration responses to trade policy

Hundanol Kebede * Jianfeng Wu †

March 23, 2026

Abstract

We document that migration and sectoral reallocation in response to trade liberalization in China were predominantly driven by young workers, especially those entering the labor market around 2002. To explain this pronounced age heterogeneity, we develop and structurally estimate a novel migration model in which workers face income uncertainty at destination locations and exhibit constant relative risk aversion (CRRA) utility. Migration decisions respond to both expected income differentials and differences in income risk across locations. We estimate migration elasticities with respect to these income moments and recover age-specific risk-aversion parameters. Older cohorts display significantly higher risk aversion than younger ones, producing migration costs that rise steeply with age and closely account for the observed sharp decline in migration rates across age groups. Counterfactual analysis imposing uniform risk aversion across cohorts yields migration costs that remain essentially flat with respect to age.
Keywords: China, Factor mobility, Internal migration, Trade policy uncertainty
JEL Codes: F16, J61, O15

*Department of Economics, Southern Illinois University Carbondale
†China Center for Economic Studies & School of Economics, Fudan University

1 Introduction

The reallocation of labor across sectors and geographic locations constitutes a central mechanism through which trade liberalization generates aggregate welfare gains. Yet a growing body of evidence documents substantial mobility frictions that slow and attenuate these adjustments, implying that the full effects of trade shocks often materialize only over extended periods (Artuç et al., 2010; Dix-Carneiro, 2014; Fajgelbaum et al., 2018). Despite the centrality of such frictions to quantitative trade models and policy evaluation, surprisingly little is known about their heterogeneity across demographic groups—in particular, how mobility costs and adjustment elasticities vary systematically with worker age.

Young adults, typically aged 15–25, are widely regarded as the most geographically and occupationally mobile segment of the population.[1] This heightened mobility at early career stages raises a natural and policy-relevant question: do younger workers also exhibit markedly greater responsiveness to trade-policy shocks? If so, the speed, magnitude, and distributional consequences of trade-induced structural transformation may depend critically on the age composition of the labor force at the time of reform.

In this paper, we investigate precisely this form of age-based heterogeneity in labor-market responses to trade liberalization. We ask two interrelated questions. First, to what extent do the aggregate sectoral reallocation and migration flows observed in the data reflect the decisions of young workers rather than adjustments by more established cohorts? Second, what underlying mechanisms—economic, institutional, or behavioral—drive systematically larger responses among the youth? We address these questions in the context of China, which underwent one of the most rapid and far-reaching economic transformations in modern history following its 2001 accession to the World Trade Organization. The Chinese experience therefore offers a uniquely powerful setting in which to identify and quantify how age shapes the elasticity of labor supply to trade-induced demand shifts.

We begin by presenting two compelling empirical regularities that highlight the distinctive nature of labor market adjustments in China following its 2001 accession to the World Trade Organization (WTO).

First, a substantial portion of the migration surge triggered by trade liberalization was disproportionately driven by younger cohorts, especially those entering the labor market at precisely age 16. Strikingly, among individuals born in 1986—who turned 16 in 2001—approximately 45% had become migrant workers by 2005. In sharp contrast, only 27% of the slightly older cohort born in 1984 were migrants in the same

[1] See, for example, https://www.brookings.edu/articles/just-before-covid-19-american-migration-hit-a-73-year-low/.

year. This marked differential underscores how trade-induced opportunities primarily shaped the initial labor market decisions of new entrants rather than prompting widespread geographic mobility among more established workers.

Second, the pronounced rise in the share of employment in manufacturing (the secondary sector) similarly reflects the disproportionate absorption of younger cohorts into this expanding sector, rather than substantial sectoral switching by incumbent workers. By 2005, more than 35% of the 1986 birth cohort was employed in the secondary sector, compared with just 25% of the 1984 cohort. These patterns reveal that the structural reallocation of labor toward manufacturing was predominantly facilitated through the career-start choices of youth cohorts entering the workforce amid rapidly expanding export-oriented opportunities.

Taken together, these findings demonstrate that the most significant margins of adjustment—both in terms of rural-to-urban migration and sectoral reallocation following China's WTO accession—were overwhelmingly driven by cohorts born between 1986 and 1990. This cohort-specific responsiveness not only amplified the speed and scale of China's labor market transformation but also illustrates how trade shocks can exert particularly powerful effects on individuals at the pivotal early stages of their working lives, when occupational and locational decisions exhibit greater elasticity. This is not peculiar to China, as we also show that migration response to Vietnam's 2002 expanded access to the U.S market was driven by the youth of roughly similar age to the Chinese case.

Next, we seek to provide an economic mechanism that explains such stark heterogeneous response to trade policy change across different cohorts. To do so, we propose a novel model of migration friction in which workers are uncertain about their income at potential destination locations and have constant relative risk aversion (CRRA) utility function. In this environment, workers' migration is influenced not only by expected income at the destinations relative to origin, but also by the variance of income (income risks) at the potential destinations relative to origin. Our model gives migration flow equation between location pairs as a function of differences in their expected income, differences in their variance of income, differences in amenities, and moving costs. Two parameters govern migration costs in our model: elasticity of migration to difference in expected incomes between destination and origin location (which is standard in the literature and we assume is constant across cohorts) and elasticity of migration to difference in variances of income between destination and origin location (i.e., risk parameter), which we allow to vary across cohorts. This latter parameter is a novel feature of our model.

Migration flow equations derived from models where agents face no income uncertainty (or are risk-neutral) omit differences in variance of income between destination and origin location. If agents face income uncertainty and are not risk-

neutral, this causes serious omitted variable bias in estimation of migration elasticity because the variance of income is a function of the mean. Importantly, this bias cannot be resolved by using IV estimation because any relevant instrument for the mean of income is correlated to the variance of income and thus fails the exclusion restriction. Our approach resolves this problem as it gives estimation equation of migration flow that includes differences in both mean and variance of income between destination and origin. Another identification issue is that unobserved factors, such as amenities or quality of institutions, could be correlated with mean and variances of income at destination and origin locations and influence migration flows. To address this endogeneity issue, we follow the literature and use income of neighboring provinces to construct IVs.

We use the 2005 mini-census data which includes information on birth-place, wages and sectors of employment for workers. This allows us to construct the mean and variances of wages for each province-sector-cohort cells. We aggregate the sectors into two – agriculture and non-agriculture – and construct bilateral migration flow between every province-sector pair for each cohort.

We estimate elasticity of migration flow to the difference in average destination and origin income of 3.5. Our estimates for the elasticity of migration to the difference in income variance between destination and origin ranges from 0.3 for the youngest cohorts (aged 15-19) to -1.47 for the oldest cohorts (aged 60-64), thus decreasing with cohort ages. We infer each cohort's risk preference parameter from the above two elasticities. We find risk parameters ranging from 0.14 for the youngest cohorts to 2.4 for the oldest cohorts. Our estimates consistently suggest that the youngest cohorts (aged 15-19) are far less risk-averse than their older cohorts. The degree of risk-aversion generally increases with age of cohorts, sharply rising after 50s.

Using our parameters, we infer bilateral migration costs between province-sector pairs for each cohort. We find significant migration costs averaging more than 5 times expected current utility of workers.[2] We find that the migration costs are predominantly driven by costs of moving across provinces, which are about 10 times the costs of switching sectors within a province. Most importantly, these migration costs significantly increase with cohort ages. On average, migration costs for the youngest cohorts are one-third of the migration costs facing the oldest cohorts. This monotonic variation in migration costs is primarily driven by their attitude towards income risks. To show this, we calculate counterfactual migration costs if all cohorts had similar risk parameter and find that the resulting migration costs are flat across cohorts.

[2]However, compared to previous related studies, our migration cost estimates are modest. This is perhaps because these past studies generally underestimate the migration elasticity which would cause overestimation of migration costs.

Our results also show that migration costs in general, and costs of switching sectors in particular, are highly asymmetric. For instance, costs of moving from agriculture to non-agriculture are roughly twice the costs of moving in the opposite direction. This is true both for within- and across-province moves.

This paper advances the literature on the relationship between trade liberalization and the sectoral reallocation of labor (Erten and Leight, 2021; McCaig and Pavcnik, 2018).[3] Prominent studies have documented substantial shifts from agriculture to manufacturing following improved export access; for example, Erten and Leight (2021) demonstrate that China's expanded access to the U.S. market after 2001 accelerated the movement of workers out of agriculture and into manufacturing at the local level. The prevailing interpretation of such patterns attributes structural transformation primarily to incumbent workers switching sectors. Indeed, influential discussions of productivity-enhancing structural change often emphasize the relative ease of occupational transitions in manufacturing-led paths, as articulated by McMillan et al. (2017): "It is comparatively easy *to turn a rice farmer into a garment factory worker* ..." without requiring substantial human or physical capital investments. Yet this paper challenges that conventional view by providing robust evidence that, in the Chinese context, sectoral reallocation following WTO accession was overwhelmingly driven by the entry decisions of younger cohorts rather than by sector switching among established workers. We document discontinuous jumps in manufacturing employment shares—and corresponding sharp declines in agricultural employment—precisely for cohorts born in 1986 and in the immediately following years. These cohort-specific shifts are markedly more pronounced in prefectures with greater exposure to trade liberalization than in less-exposed areas. Consequently, the bulk of observed structural change reflects the disproportionate absorption of new labor-market entrants into the expanding manufacturing sector, rather than widespread reallocation among older cohorts. We argue that this pattern arises from systematically higher sectoral switching costs faced by older workers, whose accumulated location-, firm-, or occupation-specific human capital renders transitions substantially more costly than for youth at the outset of their careers.

A parallel strand of research examines the migration responses to trade liberalization (Facchini et al., 2019; Imbert and Papp, 2020; Ngai et al., 2018; Zi, 2022).[4] For instance, Facchini et al. (2019) establish that reductions in trade policy uncertainty following WTO accession triggered significant increases in internal migration toward prefectures most exposed to export opportunities, with inflows concentrated among non-*hukou* migrants. This paper extends that line of inquiry by demonstrating that the migration surge was similarly dominated by the choices of cohorts born

[3] See also McCaig (2011); Kis-Katos and Sparrow (2015).

[4] See also Donovan and Schoellman (2023); Cai et al. (2024); Porzio et al. (2022).

around 1986 onward. Migrant shares exhibit a sharp discontinuous increase for these young cohorts, with the effect amplified in high-exposure destination prefectures. We attribute this age-based heterogeneity to substantially elevated migration costs borne by older workers relative to more mobile young adults.

In terms of theoretical modeling and structural estimation, our work builds closely on frameworks that quantify sectoral and geographic mobility frictions in developing and emerging economies (Tombe and Zhu, 2019; Fajgelbaum et al., 2018; Bryan and Morten, 2019; Morten and Oliveira, 2024; Redding, 2016; Allen and Arkolakis, 2014). These studies typically assume risk-neutral agents who face deterministic income prospects when deciding whether to migrate or switch sectors. A key innovation of the present paper lies in relaxing these assumptions: we incorporate destination income uncertainty and allow for heterogeneous risk preferences across age cohorts. By endogenizing risk attitudes in this way, the model generates cohort-varying mobility costs that align closely with the empirically documented patterns. This richer specification not only rationalizes the pronounced age gradients in migration and sectoral choices but also yields more accurate estimates of underlying frictions, with important implications for the speed, distributional consequences, and welfare effects of trade-induced structural transformation.

The rest of the paper is organized as follows. Section 2 describes the data and presents empirical facts. Section 3 presents our novel model of migration friction and estimation of the structural parameters. Section 4 discusses the estimated migration costs in comparison with self-reported migration costs from survey data. Section 5 concludes the paper.

2 Data and motivating facts

2.1 Data

Our main dataset comes from rounds of Chinese national census data. For our analysis of trends on migration flow and sectoral allocation across cohorts, we combine successive rounds of Chinese national census data. More specifically, we use the 2000 and 2010 rounds of national census and the 2005 and 2015 mini-censuses. We use these data sets to construct prefecture-level share of workers employed in the primary, secondary, tertiary and manufacturing sectors for each of the cohorts born since 1935. We also use the data to construct the share of workers of each cohort who are migrant in the last five years before the census date. Our measure of migrants includes those who move within prefecture and between prefectures. Past studies usually focus on between prefecture migrations which underestimate the true extent of migration because they exclude some common migration patterns such as

rural-urban migrations within a prefecture. We combine these data with a measure of exposure to China's accession to the U.S market at prefecture-level. We obtain data on this measure and a host of prefecture-level control variables from Facchini et al. (2019).

For our structural estimation of migration flow equation, we rely on the 2005 mini-census which includes information on employment, income, migration history and other basic characteristics of workers. We construct migration flows between provinces and sectoral changes within province for each cohorts (age groups) We choose province, instead of lower level of administrative units such as prefecture, because the sample size of the 2005 census does not allow us to construct reliable measure of bilateral migration at cohort and sector level.

We complement our census-based analysis with the 2007 China Household Income Project (CHIP) survey dataset. This survey includes migrants' self-reported migration costs, which we use to compare and validate our model-based estimates of migration costs at a cohort level.

2.2 Motivating facts

We start by presenting important summary statistics on trends in manufacturing employment and migration for youth and adult workers across census rounds in Table 1. The table shows that only 8% of adult (26 or older) workers in 2000 census are migrants, compared to 24% of young (age 16-25) workers who are migrants in the same year. The share of migrant adult workers increased to 24% while the share of youth workers who are migrants increased to 52% in 2010.

The table shows similar trend in manufacturing employment rates of adults and youth. In 2000 census, 12% of adult workers and 20% of youth worked in manufacturing. These numbers increased to 17% and 31% for the adults and youth, respectively, in 2010 census. Overall, the table shows that not only are the youth becoming increasingly more mobile compared to the adults but also that they are increasingly absorbed into the manufacturing sector. Below we explore how these trends are driven by trade policy by exploiting geographic variation in exposure to trade policy change.

2.2.1 Methodology

To explore heterogeneous sectoral reallocation and migration responses to trade policy across cohorts, we utilize an identification strategy that combines regression discontinuity design (RDD) and generalized DID design.

The RDD component comes from discontinuous exposure to trade liberalization across cohorts presented in figures 1 & 2. Figure 1 shows that employment share

in the secondary sector discontinuously jumps for cohorts born in 1986 (who turn 16 by 2002) compared to older cohorts. Similarly, Figure 2 shows that share of migrant workers discontinuously jumps for the cohorts born in 1986 relative to the older cohorts. The discontinuous jumps in these two figures suggest that there is an underlying difference in mobility frictions across ages that causes discontinuous variation in exposure to trade policy change across birth-years, and we exploit this for our identification.[5]

We combine the above discontinuous variation in exposure to trade policy across cohorts with exogenous spatial variation in exposure to trade liberalization across prefectures. This spatial variation is attributed to differences in the basket of goods exported by prefectures and variation in change in tariff uncertainty across different baskets of goods (see below).

Our strategy can be viewed as a generalized version of Bargain and Jonassen (2024), who use age discontinuity in eligibility to social assistance to study its effect on employment and benefit take up.[6] Our framework is generalized because in addition to age discontinuity in exposure to trade policy change, individuals in different prefectures are exposed to trade shock to a different degree depending on the industrial composition of their prefectures and variation in the pre-liberalization tariff uncertainty across industries. Intuitively, our identification strategy can be understood as comparing the discontinuous changes shown in Figures 1 & 2 across prefectures with various degree of exposure to trade policy. In the prefectures less exposed to the policy change, the discontinuous jumps in these figures are muted; in highly exposed prefectures the jump is significantly pronounced. We use these variations to estimate heterogeneous response to trade liberalization across cohorts.

Our main estimation equation is written as follows:

$$y_{pc} = \beta_0 + \beta_1\Big(D_c \times \text{NTRgap}_p\Big) + \beta_2\Big(D_c \times I_c \times \text{NTRgap}_p\Big) + \gamma_p + \gamma_c + \varepsilon_{pc} \qquad (1)$$

where y_{pc} denotes outcome variables (mainly, share of employment in manufacturing and migrant share of employment), p indexes prefectures, and c is cohorts (birth year). D_c is a dummy variable which equals 1 for cohorts born in 1986 or later and zero otherwise. Cohorts born in 1986 turn 16 years of age (legal age of employment) in 2001 when China became a member of WTO. I_c is birth year of cohorts *minus* 1986. This variable is a re-centered birth year around 1986 and in interaction with D_c captures discontinuous exposure to trade liberalization for cohorts born in 1986, relative to other cohorts.

[5] In appendix figure A.1 we show similar figures for the migrant and manufacturing share of workers by age. While the Vietnamese case not as striking as the Chinese one, it shows similar pattern.

[6] See also Lemieux and Milligan (2008).

NTRgap_p is a measure of exposure to trade liberalization across prefectures. Prior to 2001, the U.S congress votes annually to renew China's Normal Trade Relation (NTR) status in the U.S market. If China's NTR status is approved in congress, Chinese goods will be subject to low NTR tariffs (MFN tariff rates reserved for WTO members). If congress votes against China's NTR status, however, Chinese exports to the U.S. market would be subjected to very high non-NTR tariff rates (also known as column 2 tariff rates) usually imposed on nonmarket economies. The non-NTR tariff rates were determined under Smoot-Hawley Tariff Act of 1930 (Pierce and Schott, 2016). In October 2000, the U.S. congress voted to grant Permanent Normal Trade Relation (PNTR) status to China which would become effective upon China's accession to WTO in 2001. This reduced significant uncertainty faced by Chinese firms in their access to the U.S market. Moreover, because the NTR-gap, difference between the non-NTR and NTR tariff rates, varied massively across industries, the decrease in uncertainty varied across industries. We exploit this significant heterogeneity in the level of change in tariff uncertainty across industries together with geographic variation in the concentration of industries across China to construct prefecture-level exposure to trade liberalization. The resulting measure is given as $\text{NTRgap}_p = \sum_i \alpha_{pi,1999} \times \text{NTRgap}_i$, where $\alpha_{ip,1999}$ is share of prefecture p's export of good i and $\text{NTRgap}_i = \text{NonNTR tarrif}_i - \text{NTR tariff}_i$ is the difference between non-NTR tariff rate and NTR tariff rate for product i. Erten and Leight (2021) use similar methodology to study the effect of trade liberalization on structural transformation in China while Facchini et al. (2019) use it to study migration.

Alternatively, we estimate our regression using variation across census rounds for each cohort and prefecture to obtain cohort-by-prefecture-by-census round variation. We can then run the following regression.

$$y_{pct} = \beta_0 + \beta_1\Big(D_c \times \text{NTRgap}_p\Big) + \beta_2\Big(D_c \times I_c \times \text{NTRgap}_p\Big) + \gamma_p + \gamma_c + \gamma_t + \varepsilon_{pct} \tag{2}$$

The main difference between regression equations 1 and 2 is that in the former, we are implicitly assuming that cohorts born in the same year have equal share of employment in a given sector (say secondary sector) across census years, up to a constant factor. If significant fraction of these cohorts change employment sector over census years, this assumption is wrong. Equation 2 does not impose this assumption but any difference in average employment share of a cohort in a given sector is subsumed in census-year fixed effects.

2.2.2 Results

Sectoral reallocation of workers: We now present our estimation results for sectoral reallocation in table 2. The outcome variables are shares of employment in manufacturing, secondary, primary and tertiary sectors. All regressions include prefecture and cohort fixed effects and the post-1986 dummy (D) interacted with the following control variables obtained from Facchini et al. (2019): export license share, foreign tariff, import tariff, production subsidy, and NTRrate.

Column 1 shows that manufacturing share of employment increased by 17.3 percentage points for cohorts born in 1986 or later relative to older cohorts in a prefecture with 100 percentage point decrease NTRgap. Column 2 shows that share of employment in the secondary sector increased by 12.3 percentage points while Column 3 shows that the share of employment in the primary sector decreased by 13.4 for cohorts born in 1986 or later relative to older cohorts. Column 4 shows that the employment share in tertiary sector did not change significantly for cohorts born in 1986 or later following trade liberalization.

To benchmark the magnitude of these estimated effects, we compare a prefecture at the 10th percentile against a prefecture at the 90th percentile. The latter experienced a decrease in NTRgap by 46.5 percentage points more than the former. Thus multiplying the estimated coefficients in table 2 by 0.465 gives us change in employment shares for cohorts born in 1986 or later (relative to older cohorts) in a prefecture at 90th percentile of exposure to trade liberalization (relative to a prefecture at the 10th percentile). The resulting numbers are 8ppt increase in manufacturing share of employment, 5.7ppt increase in share of employment in the secondary sector, and 6.2ppt decrease in share of employment in the primary sector.

In equation 1 and table 2, our estimation is based on data from multiple censuses aggregated at prefecture-cohort level. An alternative to this approach is to use variations at prefecture-cohort-census round level as written in equation 2. The results are reported in Table 3. The estimated effects closely resemble their counterparts in table 2 implying that the share of a cohort employment in a given sector remains relatively stable across census years. That is, almost all of the variation in sectoral share of employment is across cohorts born in different years rather than within the same cohort and across census years. This result lends strong support to the main thesis of this paper that structural change in response to trade liberalization is strongly driven by difference in sectoral share of employment across cohorts who joined the labor market after the trade liberalization relative to cohorts who joined the labor market earlier.

Migration: We now turn to age heterogeneity in inter-prefecture migration in response to trade liberalization. Figure 2 show share of workers in a prefecture

who are migrants from other prefectures for each of cohorts born between 1935 and 2000. The figure presents migrant share weighted by number of workers in each cohort-prefecture cell. The figure clearly shows that there is a discontinuous jump in the share of migrant workers starting at cohorts born in 1985 and 1986. The share of workers who are migrants increased from approximately 27% for cohorts born in 1984 to nearly 45% for cohorts born in 1986.

We present the estimation results using across prefecture variation in exposure to and across cohort variation in responsiveness to trade liberalization in Table 4. Column 1 is based on data across census years aggregated to cohort-prefecture level (equation 1) whereas Column 2 conducts the estimation without aggregation across census years but instead includes census-year fixed effects (equation 2). The results across both columns show that migrant share of workers increases by 13 percentage points for cohorts born after 1986, relative to older cohorts, following one unit (100 percentage point) change in NTRgap. This result shows that most of the migration in response to decrease in tariff uncertainty is driven by the youth who just joined the labor market.

3 Model

Why are younger cohorts more responsive to trade policy? In this section we present a partial equilibrium model of migration. We abstract from general equilibrium in which wages, prices and rent are endogenously determined to focus on heterogeneous migration friction across cohorts.

We hypothesize that there are potentially three reasons why younger cohorts (who just joined the labor market) are more responsive to trade liberalization and drive the sectoral reallocation. The first is that migration barriers are likely to be higher for the older workers with families. In particular, the *hukou* system imposes higher migration barrier on adults with families because their family member may not access public services such as education and healthcare in the destination place. Younger cohorts with no family are unlikely to be deterred by this. Due to these potentially higher migration barriers, older workers demand larger expected income gain at destination location relative to their origin location to be induced to migrate, compared to younger cohorts.

Second, moving costs (such as transport costs and costs of finding accommodation) could be higher for older individuals who are likely accompanied by families.

The third potential factor is that migration involves significant income risk as migrants may not know their income at the destination location with certainty. Instead, their migration decision is likely based on knowledge of distribution of income. In such environment, their choices are influenced not only by the expected

income but also the variance of income at destination locations and the migrant's risk preference. Given an expected income and other factors, risk-averse individuals would like to avoid areas where the variance of income is high. On the contrary, risk-loving individuals would like to take the gamble of high income variance locations. One can imagine that older individuals with established carrier at their origin or with household responsibility are less willing to take large income risks compared to a 16 year old who just joins the labor market. As a result, younger cohorts are likely to be less risk-averse to migrate in response to a policy change. Our focus in this paper is to quantify the role of this third factor.

Suppose that an individual earns a random wage w_j at location j. A fraction β of this income is spent on consumption of goods with prices P_j and the remainder $1-\beta$ fraction is spent on housing with rent R_j. Real income of an individual who lives at location j can be written as $V_j = w_j/(P_j^{\beta} R_j^{1-\beta})$. Note that V_j is random because individuals do not certainly know the wage income they receive at destination location.

We assume that migrants have the following constant relative risk-aversion (CRRA) utility function

$$U_j = \frac{V_j^{1-\rho}}{1-\rho}, \quad \text{for} \quad \rho \neq 1 \tag{3}$$

where U_j is the indirect utility derived from real income at destination location and is a random variable. ρ is the *relative risk-aversion coefficient*. We allow this parameter to vary across cohorts (age-groups) without any restriction. If $\rho = 0$, the utility function becomes linear: $U_j = V_j$, indicating risk neutrality. If $\rho > 1$, the individual is highly risk-averse. If $\rho < 1$, the individual is less risk-averse.

Suppose V_j has a log-normal distribution $\ln(V_j) \sim N(\mu_j, \sigma_j^2)$. The expected utility function is given by:

$$E[U_j] = \frac{1}{1-\rho} \exp\left((1-\rho)\mu_j + \frac{1}{2}(1-\rho)^2\sigma_j^2\right) \tag{4}$$

Note that expected utility decreases with the variance of income distribution for risk-averse individuals. We find it convenient to work with the following certainty-equivalent representation of the expected utility:

$$\tilde{V}_j = \exp\left(\mu_j + \frac{1}{2}(1-\rho)\sigma_j^2\right) \tag{5}$$

where $\tilde{V}_j$ is the certainty-equivalent income for the random income V_j. For later use, it is useful to add location $\in \{i, j\}$, sector $\in \{k, l\}$ and cohort (or age) a indexes to

the expected utility. Thus, the certainty-equivalent income from working in location j and sector l for a person with cohort (or age) a is written as $\tilde{V}_j^{a,l}$. For estimation of the model parameters, we use bilateral migration data for each cohort between sector-province pairs, i.e., (i,k) to (j,l). We define sectors broadly as agricultural and non-agricultural sectors following **?**.

Migration decision: Given $\tilde{V}$, the individual makes migration decisions that maximizes its expected utility net of migration costs and idiosyncratic location preferences. Suppose $C_{ij}^{a,kl}$ denotes utility cost of migrating from (i,k) to (j,l) for a person with age a, which includes costs of moving as well as the inconveniences related to switching sector or living faraway and losing social networks at native location. This is unobserved quantity and our goal is to estimate it. Workers have idiosyncratic province preference ε_j with Fréchet distribution $F_\varepsilon(x) = e^{-x^{-\kappa}}$ where κ is a measure of dispersion of the distribution and governs migration elasticity.

An individual in location-sector (i,k) chooses location-sector (j,l) which maximizes its expected utility net of utility cost of migration and idiosyncratic location preference, given the individual's risk preference:[7]

$$\max_{j,l} \quad \left\{ \frac{\tilde{V}_j^{a,l} \varepsilon_j}{C_{ij}^{a,kl}} \right\} \tag{6}$$

Given the Fréchet distribution of ε_j, the probability of migration from (i,k) to (j,l) for a person with age a, $m_{ij}^{a,kl}$ is given by

$$m_{ij}^{a,kl} = \frac{\left(\tilde{V}_j^{a,l} / C_{ij}^{a,kl}\right)^\kappa}{\sum_{l'} \sum_{j'} \left(\tilde{V}_{j'}^{a,l'} / C_{ij'}^{a,kl'}\right)^\kappa} \quad \text{for} \quad (i,k) \neq (j,l) \tag{7}$$

The law of large numbers implies that $m_{ij}^{a,kl}$ is also the proportion of workers of age a migrating from (i,k) to (j,l). The number of workers of age a in destination j and sector l is given by $L_j^{a,l} = \sum_k \sum_i m_{ij}^{a,kl} L_i^{a,k}$.

The proportion of workers of age a who migrate from (i,k) to (j,l), $m_{ij}^{a,kl}$, relative to the proportion who stay at (i,k), $m_{ii}^{a,kk}$ is given by

$$\frac{m_{ij}^{a,kl}}{m_{ii}^{a,kk}} = \left(\frac{\tilde{V}_j^{a,l}}{\tilde{V}_i^{a,k} C_{ij}^{a,kl}} \right)^\kappa, \quad \text{for} \quad (i,k) \neq (j,l) \tag{8}$$

[7] In extension of our model, we incorporate workers preferences for amenity differences across locations. While this has no significant effect on estimation of our model parameters, the incorporation of amenities in calculation of migration costs inflates the the estimated migration costs significantly.

where $C_{ii}^{a,kk} = 1$ (i.e., stayers face no migration costs) has been used. Using equation 5 in 8 and taking logs gives

$$\ln\left(\frac{m_{ij}^{a,kl}}{m_{ii}^{a,kk}}\right) = \kappa\left(\mu_j^{a,l} - \mu_i^{a,k}\right) + 0.5\kappa(1-\rho^a)\left(\sigma_j^{2a,l} - \sigma_i^{2a,k}\right) - \kappa \ln C_{ij}^{a,kl} \tag{9}$$

This equation has an intuitive interpretation. The first term implies that migration rate increases with average income at destination relative to average income at origin where κ measures elasticity of migration rates to average income differences between destination and origin location-sector cells. The second term implies that higher variance of income at destination relative to variance of income at origin location-sector cells causes more migration rate of individuals with $0 < \rho^a < 1$ but less migration rate of individuals with $\rho^a > 1$ (risk-averse individuals). The last term shows that migration rate decreases with bilateral migration costs.

3.1 Identification and estimation

To take the model to data, we assume moving costs between (i,k) and (j,l) for each age group a take the following form:

$$C_{ij}^{a,kl} = d_{ij}^{\theta}\tilde{\gamma}_i^k\tilde{\gamma}_j\tilde{\gamma}^a\tilde{\epsilon}_{ij}^{a,kl}, \quad \text{for} \quad (i,k) \neq (j,l) \tag{10}$$

where d_{ij} is bilateral distance and θ is the distance elasticity, $\tilde{\gamma}_j$ is destination j specific migration frictions that are constant across age and origin, $\tilde{\gamma}_i^k$ is origin i and sector k specific migration frictions that are constant across age and destinations included to capture asymmetric migration frictions between sectors and location, $\tilde{\gamma}^a$ is age specific migration frictions that are constant across origin-destination pairs, and $\tilde{\epsilon}_{ij}^{a,kl}$ is the residual which we assume is random.

Using equation 10 in equation 9, we have our main estimation equation:

$$\begin{aligned}\ln\left(\frac{m_{ij}^{a,kl}}{m_{ii}^{a,kk}}\right) = \kappa\left(\mu_j^{a,l} - \mu_i^{a,k}\right) + \alpha^a\left(\sigma_j^{2a,l} - \sigma_i^{2a,k}\right) \\ + \gamma_{ij} + \gamma_i^k + \gamma_i^a + \epsilon_{ij}^{a,kl}, \quad \text{for} \quad (i,k) \neq (j,l)\end{aligned} \tag{11}$$

where $\alpha^a \equiv 0.5\kappa(1-\rho^a)$, γ_{ij} are origin-destination pair fixed effects, γ_i^k are origin-sector fixed effects and captures origin-sector specific unobserved factors common across all age groups, γ_i^a is origin-cohort fixed effects capturing origin- and cohort-specific migration frictions and $\epsilon_{ij}^{a,kl}$ is a random error term.

The above regression equations suggest that there is potentially serious omitted variable bias when estimating migration elasticities based on models of migration that ignore migrants' risk preference. This is because these estimations omit the variance

of income, which is correlated with the mean of income. To see this, note that variance of income $V = E[X^2] - E[X]^2$ is a function of its mean $E[X]$. Importantly, this bias cannot be addressed by the use of instrumental variables because any relevant instrument Z ($cov(E[X], Z) \neq 0$) will fail the exclusion restriction ($cov(V, Z) \neq 0$) because V is a function of $E[X]$.[8]

Identification issues: A potential identification issue in estimation of 11 arises if unobserved factors correlated to destination income ($\mu_j^{a,l} - \mu_i^{a,k}$ influence migration flows. A good example is institutional quality and amenities at destination locations relative to origin location, which influence not only relative economic performances of locations but also influence individuals' migration decisions because people generally like to live in locations with better institutions and amenities. Similar endogeneity concern applies to the the second moment term in equation 11 in the sense that locations with better institutions may have lower income uncertainty.

To address these endogeneity concerns, we use instrumental variables (IV) estimation strategy. We follow the literature and use population weighted average and variance of income of neighboring provinces to construct our IV. More specifically, we instrument $\left(\mu_j^{a,l} - \mu_i^{a,k}\right)$ by $\left(\sum_{j'} w_{j'} \mu_{j'}^{a,l} - \sum_{i'} w_{i'} \mu_{i'}^{a,k}\right)$ and $\left(\sigma_j^{2a,l} - \sigma_i^{2a,k}\right)$ by $\left(\sum_{j'} w_{j'} \sigma_{j'}^{2\,a,l} - \sum_{i'} w_{i'} \sigma_{i'}^{2\,a,k}\right)$ where i' denotes the set of neighboring provinces to province i, j' denotes the set of neighboring provinces to province j and w denotes the population weights. A province's income is likely correlated to its neighbor's income as provinces with richer neighbors tend to be rich and provinces in the neighborhood of poorer provinces tend to be poorer. However, income of neighbors is likely exogenous to a given province's income and migration shocks.

Equation 11 can be estimated to obtain the key structural parameters κ and $\rho_a, \forall a$, given data on distribution of income in each age-sector-location cells. We use the 2005 mini-census data which includes information on workers employment, migration, and wage income information to construct the mean and variance of wage in each age-sector-province cells. We use the same data to construct the dependent variable (proportion of migrants relative to stayers). We use provinces as geographic locations. Using lower level geographic units such as prefectures poses the problem of too much sparseness in the bilateral migration data, given that our data is constructed at location-sector-age cell level. Using provinces as geographic units overcomes this problem. As a robustness exercise, we also estimate the above regression replacing the bilateral fixed effects γ_{ij} by bilateral distance to estimate the distance elasticity of migration flows $\kappa\theta$.

[8]This correlation is zero for a normal distribution. However, for any other distribution this correlation is different from zero. For a lognormal distribution the correlation is positive. Income distribution is usually strongly positively skewed and it is best approximated by lognormal.

3.2 Results

3.2.1 Migration elasticity and risk parameters

We present our main estimation results in Table 5. Column 1 of the table presents results for our main estimation equation 11 while column 2 replaces the origin-destination pair fixed effects by bilateral distance. Both columns present IV estimation results.

The first row in Table 5 present the estimated value of κ, the elasticity of migration to mean income gap between destination and origin provinces. In our preferred specification that includes province-pair fixed effects in column 1, the value of this parameter is 3.5. Replacing the province-pair fixed effects by bilateral distance in column 2 increases this parameter to 3.9. The value of κ we estimate is slightly larger than those reported in previous studies including (Tombe and Zhu, 2019; Bryan and Morten, 2019). Bryan and Morten (2019) obtain $\kappa = 2.7$ for the U.S and 3.2 for Indonesia while Tombe and Zhu (2019) find values in the range of 1.2-1.6 for China. However, as the discussion above suggests ignoring risk preference would generally cause underestimation of κ if migrants are risk averse, on average.

Table 5 also reports the elasticity of migration to difference in income variance between destination and origin provinces ($\alpha^a \equiv 0.5\kappa(1-\rho^a)$ for each age cohorts. We use the 60-64 cohort as base group for our estimation. For this group, the elasticity of migration to income variance is -1.47. For those aged 45-49 this elasticity is -0.96 (i.e., -1.47+0.506). For the youth aged 15-19, this elasticity is positive (0.3) implying that this group would like to migrate to provinces with higher income variance. Overall, the elasticity of migration flow to income variance tends to monotonically decrease across ages.

The elasticity of migration flow to differences in mean and variance of income between destination and origin locations govern migrants' risk preference. The risk preference parameter for each cohort can be recovered as $\rho^a = 1 - (2 * \alpha^a)/\kappa$. Table 6 reports the values of these parameters. The two columns in Table 6 report risk preference parameters corresponding to each column in Table 5. Across both columns, we observe that the estimated risk parameters are the lowest for the youngest cohorts and the highest for the oldest cohorts. In other words, the oldest cohorts are the most risk-averse while the youngest are the least risk-averse. In fact, the risk preference parameter for cohorts aged 15-19 reflects that this group are risk-loving. Overall, the degree of risk-aversion tends to increase monotonically with age. These results are robust across all our sensitivity analysis.

The estimated risk preference parameters suggest that migrants are risk averse, on average. As a result, models of migration that ignore migrants risk preferences would underestimate the elasticity of migration to income difference, which biases

the estimated migration costs.

3.2.2 Migration costs

Using our estimated values of κ and ρ^a, we calculate migration cost between province and sector pairs $(i,k) \neq (j,l)$ for each cohort from equation 14 as:

$$C_{ij}^{a,kl} = \left(\frac{m_{ii}^{a,kk}}{m_{ij}^{a,kl}}\right)^{1/\kappa} e^{(\mu_j^{a,l}-\mu_i^{a,k})+0.5(1-\rho^a)(\sigma_j^{2a,l}-\sigma_i^{2a,k})} \tag{12}$$

We normalize this measure by $e^{\mu_j^{a,l}+0.5(1-\rho^a)\sigma_j^{2a,l}}$ so that the migration costs are measured in terms of destination expected utility.

The results are reported in Table 7. Column 1 presents the average migration costs between different province-sector pairs $i \neq j, k \neq l$. The estimated average migration costs range from 2.16 for the youngest cohorts aged 15-19 to 27.7 for the oldest cohorts aged 60-64. Overall, the migration costs tend to increase with ages of cohorts. The large average migration costs are partly driven by high migration costs (low migration flows) between some province-sector pairs. The median migration costs are smaller than the average costs as shown in Table 9 but show the same pattern across cohorts as the average migration costs.

Column 2-5 present the average migration costs between different provinces for different sectoral paths. Column 2 shows migration costs from agriculture (A) to non-agriculture (NA) across provinces while column 3 reports migration in the opposite direction. Column 4 shows migration within agricultural sector between different provinces while column 5 shows migration within non-agricultural sector across provinces. There are a number results worth emphasizing. First, bilateral migration costs across provinces between agriculture and non-agriculture are not symmetric. Across provinces, migration from agriculture to non-agriculture is more costly compared to the costs of migration in the opposite direction. Second, inter-provincial migration within the agricultural sector involve the highest migration costs; in particular they are higher than inter-provincial migration costs in the non-agricultural sector. This is consistent with the Chinese land tenure system acting as a barrier to farmers' migration across provinces (see, for instance, Adamopoulos et al. (2024)). Third, inter-provincial migration costs generally increase with age of cohorts regardless of the origin and destination sectors.

Columns 6-7 report within province migration costs between different sectors (i.e., costs of switching sectors within the same province). These costs are about one-tenth of inter-provincial migration costs. Switching from agriculture to non-agriculture involves slightly higher costs than the move in the opposite direction, particularly for the older cohorts. Similar to inter-provincial migration costs, costs of switching

sectors within a province significantly rise with cohort ages.

Next, we explore how the estimated migration costs vary with distance. Table 8 reports the elasticity of migration costs to distance of 1.35 when migration costs are estimated allowing risk parameters to vary across cohorts. This elasticity is 0.92 for migration costs estimated by imposing risk neutrality on all cohorts (see below).

3.2.3 The role of risk parameter

How much of the increase in migration costs across cohorts is driven by variation in risk preferences? To answer this, we evaluate how the variation in migration costs change if all cohorts are assumed have $\rho^a = 0 \quad \forall a$. The resulting migration cost is $\left(\frac{m_{ii}^{a,kk}}{m_{ij}^{a,kl}}\right)^{1/\kappa} e^{(\mu_j^{a,l}-\mu_i^{a,k})+0.5(\sigma_j^{2a,l}-\sigma_i^{2a,k})}$. Comparing this against the migration cost in equation 17, which allows for heterogeneity in risk preference, enables us to infer how much of the variation in migration costs across cohorts are driven by variation in their attitude towards income risk at destination locations.

The results are reported in Figure 3. Clearly, the migration costs corresponding to constant risk parameter are flat across cohorts, unlike the migration costs with variable risk preference. This is true across most migration paths within or between provinces originating from agriculture. Figure 4 shows the heterogeneous role of risk preference in monotonicity of migration costs for different migration paths. More specifically, migration costs for migrations originating from non-agricultural sector are more sensitive to risk preference, compared to migration costs involving agriculture as the origin sector. This is shown in the left-hand column. All panels (a)-(c) show that migration costs increase with age of cohorts when risk preference is allowed to vary across cohorts but the increase is significantly subdued when risk parameter is kept constant across cohorts.

The right column (panels (d)-(f)) show an even stronger case for the role of risk preference for migrants originating from non-agricultural sector i.e., the monotonicity of migration costs with age of cohorts is strongly driven by the variation in the degree of risk-aversion. This is evident from the fact that when risk parameter is allowed to vary across cohorts, migration costs are monotonic with cohort ages while when a constant risk-preference parameter is imposed across all cohorts, the migration costs became essentially flat across cohorts.

3.3 Robustness exercises

3.3.1 The role of amenities

Our baseline model ignores differences in amenities across provinces. In this section we incorporate amenities in the model and in our estimation of migration costs.

To do so very easily, we modify the distribution of workers' idiosyncratic location preference ε_j as $F_\varepsilon(x) = e^{-A_j x^{-\kappa}}$ where A_j is the location parameter and is a measure of province amenities. This modifies our migration probability from (i,k) to (j,l) for a person with age a as:

$$m_{ij}^{a,kl} = \frac{\left(\tilde{V}_j^{a,l} A_j / C_{ij}^{a,kl}\right)^\kappa}{\sum_{l'}\sum_{j'}\left(\tilde{V}_{j'}^{a,l'} A_{j'} / C_{ij'}^{a,kl'}\right)^\kappa} \quad \text{for} \quad (i,k) \neq (j,l) \tag{13}$$

and the proportion of workers of age a who migrate from (i,k) to (j,l), $m_{ij}^{a,kl}$, relative to the proportion who stay at (i,k), $m_{ii}^{a,kk}$ is given by

$$\frac{m_{ij}^{a,kl}}{m_{ii}^{a,kk}} = \left(\frac{\tilde{V}_j^{a,l} A_j}{\tilde{V}_i^{a,k} A_i C_{ij}^{a,kl}}\right)^\kappa, \quad \text{for} \quad (i,k) \neq (j,l) \tag{14}$$

where $C_{ii}^{a,kk} = 1$ (i.e., stayers face no migration costs) has been used. Using equation 4 in 14 and taking logs gives

$$\ln\left(\frac{m_{ij}^{a,kl}}{m_{ii}^{a,kk}}\right) = \kappa\left(\mu_j^{a,l} - \mu_i^{a,k}\right) + 0.5\kappa(1-\rho^a)\left(\sigma_j^{2a,l} - \sigma_i^{2a,k}\right) + \kappa(A_j - A_i) - \kappa \ln C_{ij}^{a,kl} \tag{15}$$

Compared to our baseline model, this equation includes an additional term $\kappa(A_j - A_i)$ that captures that workers migrate to locations with better amenities, conditional on differences in income levels and income risks. Given our assumption about the structure of moving costs in equation 10, our estimation equation is written as:

$$\begin{aligned}\ln\left(\frac{m_{ij}^{a,kl}}{m_{ii}^{a,kk}}\right) &= \kappa\left(\mu_j^{a,l} - \mu_i^{a,k}\right) + \alpha^a\left(\sigma_j^{2a,l} - \sigma_i^{2a,k}\right) \\ &\quad + \theta \ln d_{ij} + \kappa(A_j - \gamma_j) - \kappa(A_i + \gamma_i) + \kappa\gamma^a + \epsilon_{ij}^{a,kl}, \quad \text{for} \quad (i,k) \neq (j,l)\end{aligned} \tag{16}$$

Note that γ_j is the cost of migrating *to* j that does not vary across origins and γ_i is the cost of migration *out of* i regardless of destination. The larger γ_j, the lower migration flow to j from any origin province. Similarly, the higher γ_i the lower migration flow from origin i. A major problem is that our estimation cannot separately identify these parameters from the location amenities. Conditional on income variables, there is more migration flow to destination j if j has attractive amenities or moving cost to j is lower. Similarly, there is less migration out of i if either i has better amenities or moving out of i is more costly.

One can argue that $\tilde{A}_j \equiv A_j - \gamma_j$ is a better measure of j's amenities as it captures

j's attractiveness net of the cost of moving there. By the same logic $\tilde{A}_i \equiv A_i + \gamma_i$ can be interpreted as amenities given up at i inclusive of the moving costs out of i. With this broader definition of amenities, it is less consequential if we cannot separately identify the γ terms from the A terms as long as we obtain a valid estimate for the combined terms.

We estimate the above regression imposing the constraint that $\tilde{A}_j = -\tilde{A}_i$. The corresponding migration cost is

$$\tilde{C}_{ij}^{a,kl} = \left(\frac{m_{ii}^{a,kk}}{m_{ij}^{a,kl}}\right)^{1/\kappa} e^{(\mu_j^{a,l} - \mu_i^{a,k}) + 0.5(1-\rho^a)(\sigma_j^{2a,l} - \sigma_i^{2a,k}) + (\tilde{A}_j - \tilde{A}_i)} \tag{17}$$

Note that the migration costs $\tilde{C}_{ij}^{a,kl}$ are different from those in our main analysis because the current estimates incorporate differences in amenities net of location specific moving costs.

The results are presented in Table A.4 and Figure 4. The results in Table A.4 are very similar to our main results. First, most of overall migration costs are driven by between-province migration costs: the costs of switching sectors within-province are about one-tenth of the migration costs between provinces. Second, migration costs are significantly higher for migrants originating from the agricultural sector. Figure 4 shows the role of risk aversion in heterogeneous migration costs across ages for different migration paths. The Figures show that monotonicity of migration costs across ages are driven by older migrants' stronger risk-aversion than their younger counterparts. Counterfactual migration costs calculated assuming a constant risk-aversion parameter for all cohorts are flat across age, particularly for migrants originating from the non-agricultural sector. All these results reinforce our main results.

3.3.2 Alternative specification

As a robustness exercise, we estimate the following variation of our migration equation:

$$\ln\left(\frac{m_{ij}^{a,kl}}{m_{ii}^{a,kk}}\right) = \kappa\mu_j^{a,l} + \alpha^a\sigma_j^{2a,l} + \gamma_{ij} + \gamma_i^k + \gamma_i^a + \epsilon_{ij}^{a,kl}, \quad \text{for} \quad (i,k) \neq (j,l) \tag{18}$$

The only difference between the current regression and our baseline regression is that in the baseline regression, the independent variables are differences in the mean and variance of income between destination and origin locations while in the current regression the regressors are mean and variance of income at destination location. Given our origin-sector fixed effects in the regressions, these two regressions are closely related. We instrument $\mu_j^{a,l}$ by $\sum_{j'} w_{j'}\mu_{j'}^{a,l}$ and $\sigma_j^{2a,l}$ by $\sum_{j'} w_{j'}\sigma_{j'}^{2\,a,l}$ where j'

denotes the set of neighboring provinces to province j and w denotes the population weights.

We present the results in Table A.1. Column 1 presents results with origin-destination pair fixed effects while Column 2 replaces the fixed effect with log distance between origin and destination provinces. The results are similar across both columns. Moreover, the results in this table are very similar to their counterparts in our main analysis, showing that our conclusions are not sensitive to changes in our regression specifications. The first row in Table A.1 reports the elasticity of migration flow with respect to average income at destination. The reported elasticities ranging from 3.5-3.9 are very similar to their counterparts in Table 5. Moreover, the elasticity of migration flows with respect to variance of income at destination reported in Table A.1 are similar to their counterparts in the main analysis. In particular, these elasticities significantly decrease with age of cohorts.

Table A.2 reports the resulting risk preference parameters across age cohorts. Similar to the results in the main analysis, the new estimates show that the youngest cohorts have the lowest risk parameter while the oldest cohorts have the highest risk parameter. Moreover, the youngest cohorts (aged 15-19) have a risk parameter below one. All the other age cohorts have risk parameter larger than one, suggesting that they are risk averse.

We report the migration costs corresponding to the new parameters in Table A.3. These migration costs have similar features as those in the main analysis. First, they are lowest for the youngest cohorts and highest for the oldest. They generally increase with age of the cohorts. Second, migration costs are higher for migrants originating from the agricultural sector compared to those originating from non-agriculture. Third, inter-province migration costs about ten times larger than costs of switching sectors within a province. Overall, the results in this table reinforce our results in the main analysis.[9]

4 Migration costs from survey reports

In this section we compare our estimated migration costs with self-reported migration costs by migrants in 2007 China Household Income Project (CHIP) survey dataset. The survey asks households their current earing as well as their counterfactual income if they were still at their origin location. From this, we calculate the migration costs as the ratio of their current income to their counterfactual income at their origin location for each age group.

The results are presented in figure 5. These self-reported migration costs are

[9]The similarities with the main result are partly driven by the origin province-sector and origin province-cohort fixed effects included in both regressions.

generally lower than the migration costs estimated from our model. However, it is important to note that the self-reported migration costs would underestimate the true migration costs because these costs are estimated based on income of those who actually migrated. One would expect that costs of those who did not migrate would have been higher. Our model-based migration costs estimates account for such selection issue. Hence, this perhaps explains why they are higher than the self-reported costs.

Though the self-reported migration costs are relatively lower, they vary systematically across migrant ages in a pattern that mimics our model-based migration costs. The self-reported costs are lowest for the youngest age group (15-19) and sharply increase across age cohorts until age group 35-39. The migration costs for the 35-39 age group are 70% higher than for those aged 15-19. The self-reported migration costs tends to slightly decrease beyond the 40-44 age groups while our model based migration costs persistently increase with ages. We attribute this divergence in the trend between self-reported and model based estimates to the selection issue we highlighted above. That is, the selection issue likely worsens with age. Overall, with cautious interpretation, these self-reported migration costs reinforce our model-based estimates.

5 Conclusion

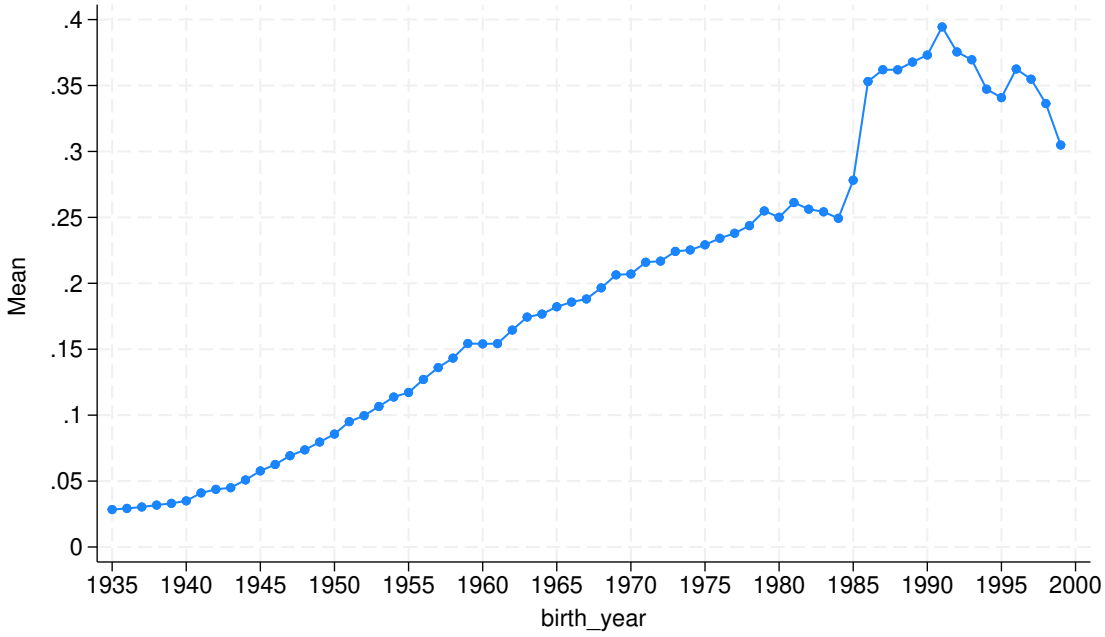
We study how migration and sectoral allocation responses to trade policy change varies across workers of different age groups. We find that most of migration and sectoral adjustments in response to trade policy are driven by younger cohorts, who join the booming sectors at disproportionately larger fraction. We rationalize this by significant variation in mobility friction across workers of different ages driven by their attitude towards income risks at destination location-sector relative to their origin location-sector. Younger cohorts, particularly those who just joined the labor markets, tend to be significantly less risk averse than their older counterparts and thus are less deterred by risks of income loss, compared to their older counterparts.

Our results have important policy implications. Whether structural change is driven by workers switching sectors or by new workers (new entrants to the labor market) choosing to join the booming sector in disproportionately large fraction has important policy implications. Though structural transformation (particularly from agriculture to modern sectors) is key for economic development, the obstacles to such structural transformation is poorly understood. Is structural change in areas such as Sub-Saharan Africa held back due to (i) significant mobility frictions to workers across sectors or locations, or (ii) is the bottleneck sluggish development of the modern sectors that absorbs more workers? Identifying which of these two

is key problem directs policy either towards reducing workers mobility barriers or towards promoting policies that help faster expansion of the modern sectors. The Chinese experience studied in this paper strongly suggests that even state-imposed toughest mobility barriers - such as the *hukou* system - may only slightly delay but not hinder sectoral and geographic reallocation. The reason is that mobility frictions are generally lower for younger workers who just joined the labor market compared their older counterparts. And these younger cohorts are enough to drive significant structural change. That is, a country that manages to create a vibrant manufacturing sector to absorb large number of new workers would eventually achieve relatively fast structural transformation despite barriers to mobility of workers.

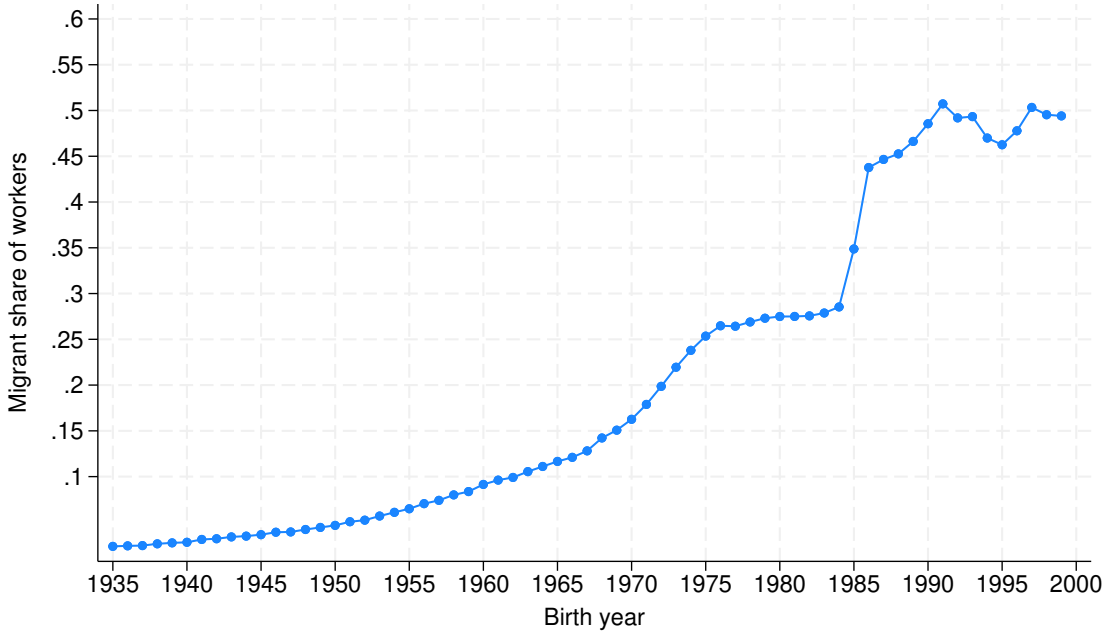
Figures

Figure 1: Share of employment in the secondary sector across cohorts



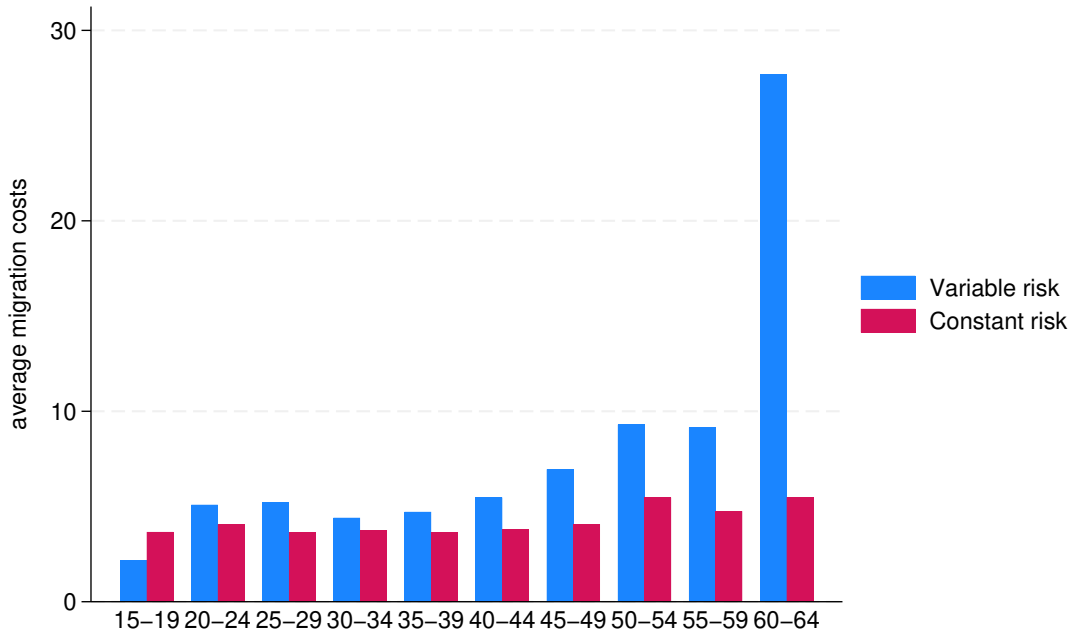
Notes: This figure plots the average share of employment in the manufacturing sector in Prefectures across cohorts.

Figure 2: Share of migrant workers across cohorts (weighted by number of workers)



Notes: This figure plots the average share of migrants workers in Prefectures across cohorts. The number of workers in each prefecture is used as weights in calculation of the national average.

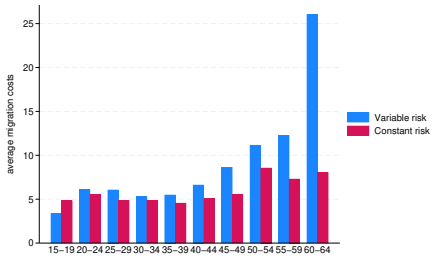
Figure 3: Migration costs when risk parameter varies vs when it is constant across cohorts ($\rho^a = 0, \forall a$)



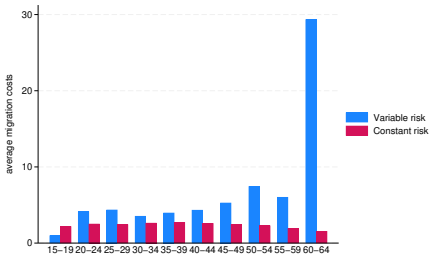
Notes: This figure shows variation of migration costs with cohort age. Migration costs are calculated under two scenarios:(i) using empirically estimated risk parameters varies across cohorts and (ii) assuming a uniform risk parameter across cohorts ($\rho^a = 0, \forall a$).

Figure 4: Migration costs across different paths when risk parameter varies across cohorts vs when it is assumed to be constant ($\rho^a = 0, \forall a$)

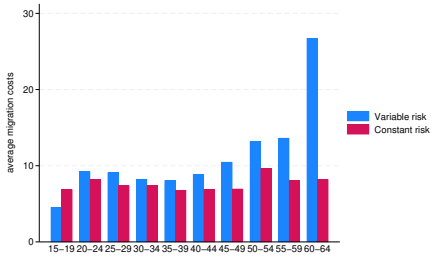
(a) Inter-province Ag → Na



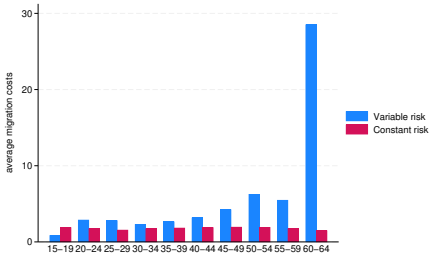
(d) Inter-province Na → Ag



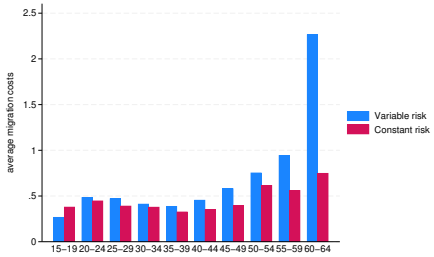
(b) Inter-province Ag → Ag



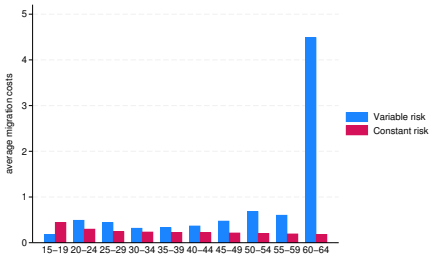
(e) Inter-province Na → Na



(c) Intra-province Ag → Na

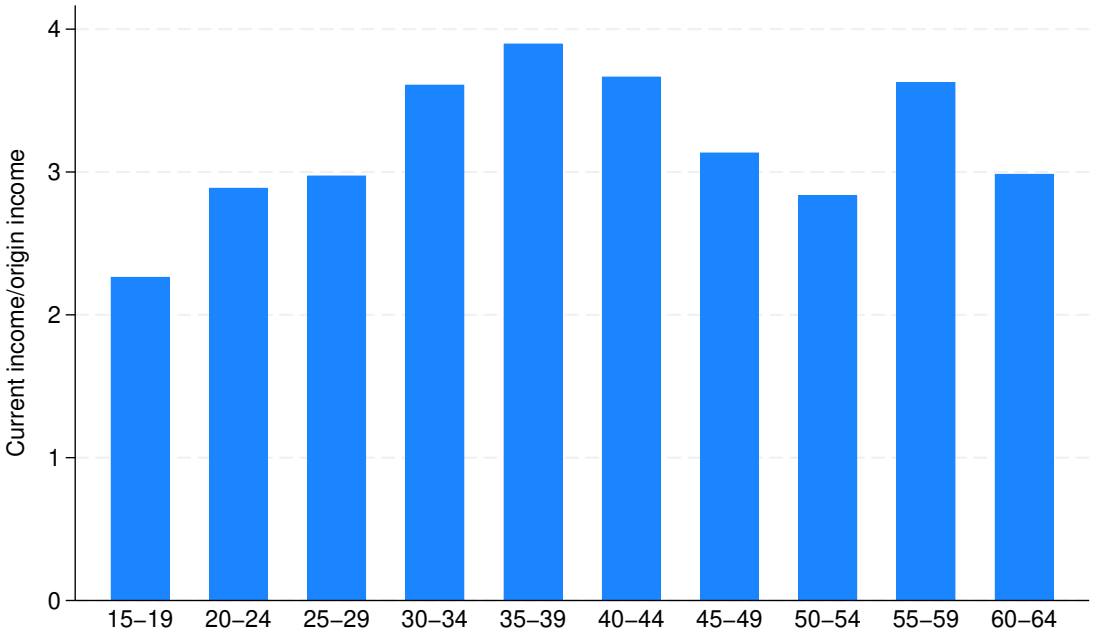


(f) Inta-province Na → Ag



Notes: The figures present estimated migration costs across cohorts for different migration paths under two scenarios: (i) under observed risk parameter of each cohort and (ii) assuming all cohorts have the same risk preference parameter ($\rho^a = 0, \forall a$). Panels (a)-(c) present migrations originating from agriculture. Panels (d)-(f) present migrations originating from non-agriculture. Panels (c) and (f) present the costs of switching sectors within a province. Migration costs are measured relative to destination utility.

Figure 5: Estimation of migration costs from survey (CHIP 2007)



Notes: This figure presents estimation of migration costs from CHIP 2007 data. Migration cost is measured as workers current income relative to the income that the worker would have earned if he/she remained at origin location. The graph presents the average of this measure across workers within age groups.

Tables

Table 1: Employment and migration statistics across the 2000, 2005, and 2010 censuses

		2000	2005	2010
	Adults (Age >=26)			
1	Number of adult workers (mil.)	523	562	640
2	Number of adult migrant workers (mil.)	44	49	157
3	Number of adult manufacturing workers (mil.)	61	69	110
	Youth (Age 16-25)			
4	Number of young workers (mil.)	118	117	150
5	Number of young migrant workers (mil.)	28	32	78
6	Number of young manufacturing workers (mil.)	24	28	46
	Shares of adult workers			
	who are migrant (2/1)	8%	9%	24%
	who are in manufacturing (3/1)	12%	12%	17%
	Shares of young workers			
	who are migrant (5/4)	24%	27%	52%
	who are in manufacturing (6/4)	20%	24%	31%

Notes: This table presents employment statistics based on the 2000, 2005, and 2010 censuses. Migrants in each census include those who arrived in the last five years of the census date. Migration numbers include both within and between prefecture movers.

Table 2: The effect of trade liberalization on sectoral employment shares

	(1) Manufacturing	(2) Secondary	(3) Primary	(4) Tertiary
D*NTRgap	0.173***	0.123***	-0.134**	0.011
	(0.037)	(0.040)	(0.052)	(0.024)
D*I*NTRgap	0.003	0.005	-0.011**	0.006*
	(0.003)	(0.004)	(0.004)	(0.003)
N	19291	19291	19291	19291
R^2	0.884	0.900	0.929	0.865

Notes: Standard errors are clustered at Prefecture level. All regressions include prefecture and cohort fixed effects and the following control variables interacted with D: export license share, foreign tariff, import tariff, production subsidy, and NTRrate.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: The effect of trade liberalization on sectoral employment shares (no aggregation across census years)

	(1) Manufacturing	(2) Secondary	(3) Primary	(4) Tertiary
D*NTRgap	0.165***	0.103***	-0.094**	-0.009
	(0.034)	(0.035)	(0.047)	(0.023)
D*I*NTRgap	0.004	0.007*	-0.013***	0.006**
	(0.003)	(0.004)	(0.004)	(0.003)
N	66110	66110	66110	66110
R^2	0.810	0.809	0.839	0.711

Notes: Standard errors are clustered at Prefecture level. All regressions include prefecture, census year and cohort fixed effects, and the following control variables interacted with D: export license share, foreign tariff, import tariff, production subsidy, and NTRrate.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: The effect of trade liberalization on migrant share of workers

	(1)	(2)
	Migrant share	Migrant share
D*NTRgap	0.142***	0.139***
	(0.052)	(0.035)
D*I*NTRgap	0.003	0.005
	(0.005)	(0.004)
N	18354	53962
R^2	0.801	0.668

Notes: Standard errors are clustered at Prefecture level. Column 1 is based on data from multiple census years aggregated by cohort-prefecture level. Column 2 is is based on cohort-prefecture-census year variation. Column 1 includes cohort and prefecture fixed effects. Column 2 includes cohort, prefecture and census-year fixed effects. All regressions include the following control variables interacted with D: export license share, foreign tariff, import tariff, production subsidy, and NTRrate.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Income and income risk elasticity of migration by age groups

	Model-1	Model-2
Diff mean income	3.536***	3.889***
	(0.262)	(0.354)
Diff var income	-2.416***	-2.788***
	(0.182)	(0.254)
Age 15-19 * Diff var income	3.943***	4.905***
	(0.333)	(0.448)
Age 20-24 * Diff var income	1.894***	2.067***
	(0.086)	(0.107)
Age 25-29 * Diff var income	1.912***	2.101***
	(0.099)	(0.129)
Age 30-34 * Diff var income	2.161***	2.360***
	(0.100)	(0.131)
Age 35-39 * Diff var income	2.106***	2.362***
	(0.105)	(0.135)
Age 40-44 * Diff var income	2.002***	2.194***
	(0.103)	(0.132)
Age 45-49 * Diff var income	1.754***	1.971***
	(0.106)	(0.137)
Age 50-54 * Diff var income	1.356***	1.525***
	(0.093)	(0.120)
Age 55-59 * Diff var income	1.374***	1.590***
	(0.107)	(0.142)
Distance		-1.236***
		(0.012)
Origin-sector FE	Yes	Yes
Origin-destination FE	Yes	No
Age group FE	Yes	Yes
N	37766	37766
R-square	-0.87	-0.26
First-stage F-stat	24.39	15.65

Notes: This table presents estimation results for migration flow equation. Diff mean income is the difference in average income between destination province/sector and origin province/sector. Diff var income represents differences in variance of income between destination province/sector and origin province/sector. Column 1 includes origin-destination fixed effects while column 2 replaces these fixed effects by log of bilateral distance. 60-64 age group are base group.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Implied risk preference parameter by age group ρ^a

Age group	ρ^a (Model-1)	ρ^a (Model-2)
15-19	0.14	-0.09
20-24	1.30	1.37
25-29	1.28	1.35
30-34	1.14	1.22
35-39	1.18	1.22
40-44	1.23	1.31
45-49	1.37	1.42
50-54	1.60	1.65
55-59	1.59	1.62
60-64	2.37	2.43

Notes: This table presents risk preference parameters corresponding to the migration flow equation presented in the previous table.

Table 7: Average migration costs by age group and migration path

	Different Provinces					Same Province	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Overall	$A \to$ NA	NA $\to A$	$A \to A$	NA $\to$ NA	$A \to$ NA	NA $\to A$
15-19	2.16	3.34	0.98	4.52	0.85	0.27	0.18
20-24	5.09	6.08	4.11	9.21	2.88	0.49	0.49
25-29	5.21	6.06	4.36	9.07	2.80	0.47	0.44
30-34	4.41	5.28	3.53	8.23	2.31	0.41	0.32
35-39	4.72	5.48	3.96	8.05	2.64	0.39	0.33
40-44	5.47	6.62	4.33	8.86	3.17	0.46	0.37
45-49	6.95	8.63	5.27	10.47	4.24	0.58	0.48
50-54	9.30	11.16	7.44	13.16	6.26	0.75	0.68
55-59	9.15	12.28	6.02	13.62	5.48	0.95	0.60
60-64	27.69	26.03	29.35	26.72	28.52	2.26	4.49
N	16548	8274	8274	7562	9182	277	277

Notes: The table presents average migration costs under different conditions. Column 1 reports average migration costs when origin and destination provinces are different and origin and destination sectors are different. Columns 2-5 reports average migration costs when origin and destination provinces are different between and within agriculture (A) and non-agriculture (NA) sectors. Column 6-7 reports average migration costs between A and NA sectors. In all cases, migration costs are calculated as a ratio to destination (current) utility of workers.

Table 8: Migration costs and distance

	(1) Migration costs	(2) Migration costs
Log Distance	1.371***	0.911***
	(0.095)	(0.049)
Origin Province-Sector FEs	Y	Y
Destination Province-Sector FEs	Y	Y
Cohort FEs	Y	Y
N	16548	18300
R2	0.423	0.521

Notes: In Column 1, the migration costs are calculated allowing the risk parameter to vary across cohorts. In Column 2, the migration costs are calculated assuming constant risk parameter $\rho^a = 1$ across all cohorts.

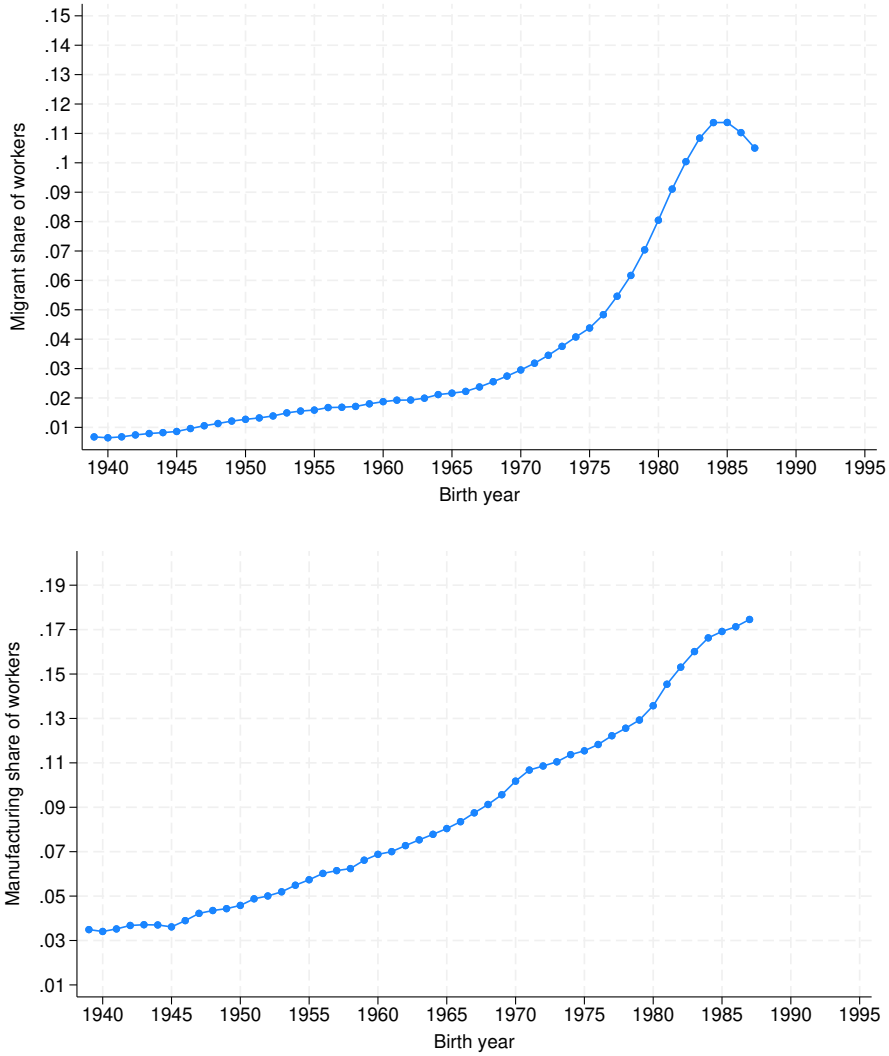
Table 9: Median migration costs by age group and migration path

	Different Provinces					Same Province	
	(1) Overall	(2) $A \to$ NA	(3) NA $\to A$	(4) $A \to A$	(5) NA $\to$ NA	(6) $A \to$ NA	(7) NA $\to A$
15-19	1.13	2.35	0.88	3.18	0.71	0.18	0.13
20-24	4.06	3.20	4.13	8.98	3.61	0.49	0.46
25-29	4.28	3.12	4.51	8.81	3.23	0.49	0.43
30-34	3.65	2.62	3.73	7.84	2.84	0.38	0.31
35-39	3.87	3.19	3.98	7.68	3.39	0.40	0.33
40-44	4.46	5.04	4.41	9.14	3.81	0.44	0.37
45-49	5.71	7.54	5.45	9.64	4.67	0.52	0.48
50-54	7.80	10.38	7.53	13.13	7.01	0.74	0.69
55-59	6.71	12.42	6.26	14.17	5.83	0.91	0.61
60-64	18.26	19.26	18.26	19.26	17.22	1.56	2.62
N	16548	8274	8274	7562	9182	277	277

Notes: The table presents median migration costs under different conditions. Column 1 reports median migration costs when origin and destination provinces are different and origin and destination sectors are different. Columns 2-5 reports median migration costs when origin and destination provinces are different between and within agriculture (A) and non-agriculture (NA) sectors. Column 6-7 reports median migration costs between A and NA sectors. In all cases, migration costs are calculated as a ratio to destination (current) utility of workers.

Appendix A Robustness exercises

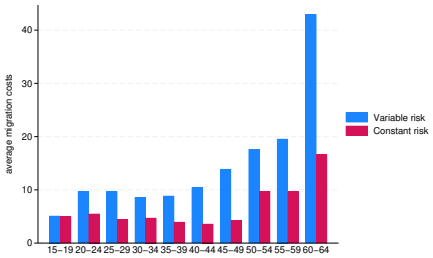
Figure A.1: Migrants and manufacturing share of workers across ages in Vietnam



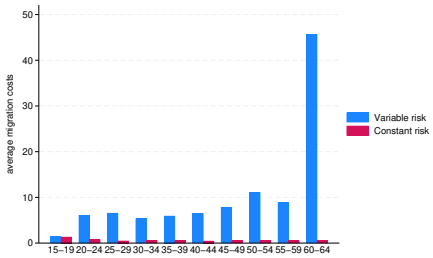
Notes: This figure presents the share of workers who are migrants and share of manufacturing workers for by birth year. Vietnam had expanded access to the U.S market in 2002. Thus individuals born during 1980s or later are the most exposed to the trade policy shock.

Figure A.2: Migration costs (incorporating amenities) across different paths when all agents are assumed to be risk-neutral vs when risk parameter varies across agents

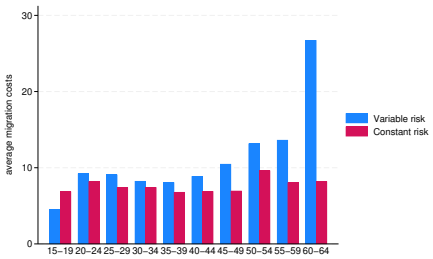
(a) Inter-province Ag → Na



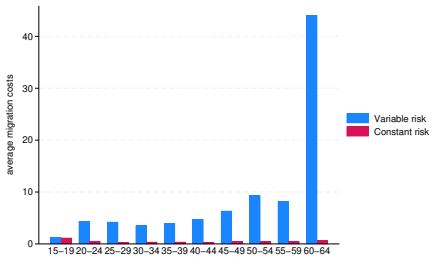
(d) Inter-province Na → Ag



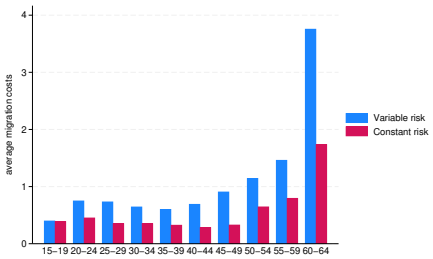
(b) Inter-province Ag → Ag



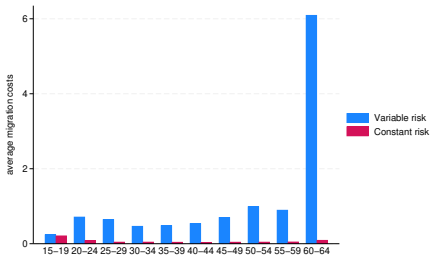
(e) Inter-province Na → Na



(c) Intra-province Ag → Na

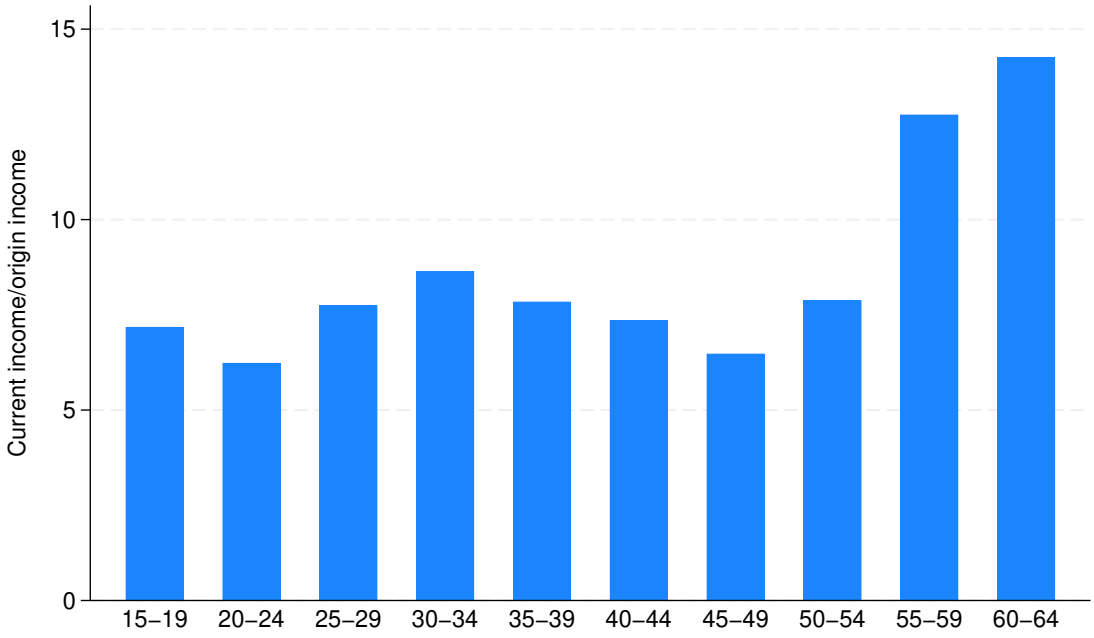


(f) Inta-province Na → Ag



Notes: The figures present estimated migration costs across cohorts for different migration paths under two scenarios: (i) under observed risk parameter of each cohort and (ii) assuming all cohorts have the same risk preference parameter ($\rho^a = 0, \forall a$). Panels (a)-(c) present migrations originating from agriculture. Panels (d)-(f) present migrations originating from non-agriculture. Panels (c) and (f) present the costs of switching sectors within a province. Migration costs are measured relative to destination utility.

Figure A.3: Estimation of migration costs from survey (CHIP 2002)



Notes: This figure presents estimation of migration costs from CHIP 2002 data. Migration cost is measured as workers current income relative to the income that the worker would have earned if he/she remained at origin location. The graph presents the average of this measure across workers within age groups.

Table A.1: Income and income risk elasticity of migration by age groups

	Model-1	Model-2
Dest income mean	3.530***	4.579***
	(0.285)	(0.578)
Dest income var	-1.470***	-2.199***
	(0.140)	(0.296)
Age 15-19 * Dest income var	1.773***	3.771***
	(0.257)	(0.504)
Age 20-24 * Dest income var	1.030***	1.158***
	(0.077)	(0.094)
Age 25-29 * Dest income var	0.870***	1.079***
	(0.069)	(0.093)
Age 30-34 * Dest income var	1.295***	1.498***
	(0.078)	(0.106)
Age 35-39 * Dest income var	1.098***	1.447***
	(0.079)	(0.107)
Age 40-44 * Dest income var	0.893***	1.098***
	(0.075)	(0.097)
Age 45-49 * Dest income var	0.506***	0.768***
	(0.071)	(0.099)
Age 50-54 * Dest income var	0.557***	0.755***
	(0.072)	(0.104)
Age 55-59 * Dest income var	0.481***	0.752***
	(0.072)	(0.111)
Distance		-1.246***
Origin-sector FE	Yes	Yes
Origin-destination FE	Yes	No
Age group FE	Yes	Yes
N	36208	36208
R-square	-0.08	0.15
First-stage F-stat	27.19	8.39

Notes: This table presents estimation results for migration flow equation. Dest income mean is the average income at destination province/sector. Dest income var represents variance of income at destination province/sector. Column 1 includes origin-destination fixed effects while column 2 replaces these fixed effects by log of bilateral distance. 60-64 age group are base group..
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.2: Implied risk preference parameter by age group ρ^a

Age group	ρ^a (Model-1)	ρ^a (Model-2)
15-19	0.83	0.31
20-24	1.25	1.45
25-29	1.34	1.49
30-34	1.10	1.31
35-39	1.21	1.33
40-44	1.33	1.48
45-49	1.55	1.63
50-54	1.52	1.63
55-59	1.56	1.63
60-64	1.83	1.96

Notes: This table presents risk preference parameters corresponding to the migration flow equation presented in the previous table.

Table A.3: Average migration costs by age group and migration path: Robustness check

	Different Provinces					Same Province	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Overall	$A \to$ NA	NA $\to A$	$A \to A$	NA $\to$ NA	$A \to$ NA	NA $\to A$
15-19	3.07	4.30	1.85	5.82	1.60	0.33	0.37
20-24	4.88	5.95	3.82	9.03	2.68	0.47	0.46
25-29	5.64	6.36	4.91	9.54	3.16	0.49	0.50
30-34	4.19	5.13	3.24	8.00	2.12	0.40	0.29
35-39	5.01	5.74	4.29	8.41	2.86	0.40	0.36
40-44	6.40	7.46	5.34	9.93	3.90	0.51	0.45
45-49	9.18	10.72	7.64	12.94	6.13	0.71	0.69
50-54	8.40	10.45	6.34	12.30	5.33	0.70	0.58
55-59	8.86	12.01	5.72	13.33	5.20	0.92	0.57
60-64	11.26	15.07	7.46	15.43	7.18	1.31	1.01
N	16548	8274	8274	7562	9182	277	277

Notes: The table presents average migration costs under different conditions. Column 1 reports average migration costs when origin and destination provinces are different and origin and destination sectors are different. Columns 2-5 reports average migration costs when origin and destination provinces are different between and within agriculture (A) and non-agriculture (NA) sectors. Column 6-7 reports average migration costs between A and NA sectors. In all cases, migration costs are calculated as a ratio to destination (current) utility of workers.

Table A.4: Average migration costs by age group and migration path (incorporating amenities)

	Different Provinces					Same Province	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Overall	$A \to$ NA	NA $\to A$	$A \to A$	NA $\to$ NA	$A \to$ NA	NA $\to A$
15-19	1.24	1.97	0.51	2.67	0.44	0.21	0.10
20-24	3.21	3.69	2.72	5.54	1.93	0.39	0.39
25-29	3.28	3.67	2.89	5.46	1.89	0.38	0.36
30-34	2.77	3.17	2.36	4.95	1.55	0.33	0.27
35-39	2.87	3.26	2.48	4.77	1.68	0.31	0.26
40-44	3.48	4.07	2.89	5.45	2.13	0.37	0.31
45-49	4.26	5.19	3.34	6.33	2.70	0.46	0.38
50-54	5.76	6.69	4.83	7.99	4.08	0.59	0.55
55-59	5.55	7.25	3.84	8.04	3.51	0.71	0.48
60-64	19.47	16.50	22.45	16.94	21.86	1.82	4.16
N	16548	8274	8274	7562	9182	277	277

Notes: The table presents average migration costs under different conditions. Column 1 reports average migration costs when origin and destination provinces are different and origin and destination sectors are different. Columns 2-5 reports average migration costs when origin and destination provinces are different between and within agriculture (A) and non-agriculture (NA) sectors. Column 6-7 reports average migration costs between A and NA sectors. In all cases, migration costs are calculated as a ratio to destination (current) utility of workers.

References

Adamopoulos, T., Brandt, L., Chen, C., Restuccia, D., and Wei, X. (2024). Land security and mobility frictions*. *The Quarterly Journal of Economics*, 139(3):1941–1987.

Allen, T. and Arkolakis, C. (2014). Trade and the topography of the spatial economy *. *The Quarterly Journal of Economics*, 129(3):1085–1140.

Artuç, E., Chaudhuri, S., and McLaren, J. (2010). Trade shocks and labor adjustment: A structural empirical approach. *American Economic Review*, 100(3):1008–45.

Bargain, O. B. and Jonassen, A. B. (2024). New evidence on welfare's disincentive for the youth using administrative panel data. *The Review of Economics and Statistics*, 106(3):655–670.

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

Cai, S., Shi, X., and Xu, Z. (2024). Migration networks, export shocks, and human capital acquisition: Evidence from china. *Journal of Comparative Economics*, 52(2):568–589.

Dix-Carneiro, R. (2014). Trade liberalization and labor market dynamics. *Econometrica*, 82(3):825–885.

Donovan, K. and Schoellman, T. (2023). The role of labor market frictions in structural transformation*. *Oxford Development Studies*, 51(4):362–374.

Erten, B. and Leight, J. (2021). Exporting out of agriculture: The impact of wto accession on structural transformation in china. *The Review of Economics and Statistics*, 103(2):364–380.

Facchini, G., Liu, M. Y., Mayda, A. M., and Zhou, M. (2019). China's "great migration": The impact of the reduction in trade policy uncertainty. *Journal of International Economics*, 120:126–144.

Fajgelbaum, P. D., Morales, E., Suárez Serrato, J. C., and Zidar, O. (2018). State taxes and spatial misallocation. *The Review of Economic Studies*, 86(1):333–376.

Imbert, C. and Papp, J. (2020). Costs and benefits of rural-urban migration: Evidence from india. *Journal of Development Economics*, 146:102473.

Kis-Katos, K. and Sparrow, R. (2015). Poverty, labor markets and trade liberalization in indonesia. *Journal of Development Economics*, 117:94–106.

Lemieux, T. and Milligan, K. (2008). Incentive effects of social assistance: A regression discontinuity approach. *Journal of Econometrics*, 142(2):807–828. The regression discontinuity design: Theory and applications.

McCaig, B. (2011). Exporting out of poverty: Provincial poverty in vietnam and u.s. market access. *Journal of International Economics*, 85(1):102–113.

McCaig, B. and Pavcnik, N. (2018). Export markets and labor allocation in a low-income country. *American Economic Review*, 108(7):1899–1941.

McMillan, M. S., Rodrik, D., and Sepúlveda, C., editors (2017). *Structural change, fundamentals, and growth: A framework and case studies*. Number 978-0-89629-214-7 in IFPRI books. International Food Policy Research Institute (IFPRI).

Morten, M. and Oliveira, J. (2024). The effects of roads on trade and migration: Evidence from a planned capital city. *American Economic Journal: Applied Economics*, 16(2):389–421.

Ngai, L. R., Pissarides, C. A., and Wang, J. (2018). China's mobility barriers and employment allocations. *Journal of the European Economic Association*, 17(5):1617–1653.

Pierce, J. R. and Schott, P. K. (2016). The Surprisingly Swift Decline of US Manufacturing Employment. *American Economic Review*, 106(7):1632–62.

Porzio, T., Rossi, F., and Santangelo, G. (2022). The human side of structural transformation. *American Economic Review*, 112(8):2774–2814.

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

Tombe, T. and Zhu, X. (2019). Trade, migration, and productivity: A quantitative analysis of china. *American Economic Review*, 109(5):1843–72.

Zi, Y. (2022). Trade liberalization and the great labor reallocation. Working Paper .