Economic Theory (2022) 74:1–67
https://doi.org/10.1007/s00199-021-01369-2

RESEARCH ARTICLE

Educational choice, rural–urban migration and economic development

Pei-Ju Liao[1] · Ping Wang[2] · Yin-Chi Wang[3] · Chong K. Yip[4]

Received: 23 August 2019 / Accepted: 5 May 2021 / Published online: 14 June 2021

Abstract

We develop an overlapping-generations framework of education-based migration that takes place prior to labor-market participation and explore its role for economic development, urbanization and workforce composition. We show that education-based and work-based migration are substitutes and the equilibrium outcome depends crucially on children's talent distribution, college costs and selectiveness, urban job opportunities, and migration barriers. We establish conflicting partial- and general-equilibrium effects at work for comparative statics, and examine their locational as well as macroeconomic implications for assessing education and migration policies. Applying our model to fit the data from China over 1980–2007, we find that, although education-based migration only amounts to one-fifth of that of work-based migration, it contributes more to per capita output growth than work-based migration owing to its high-skilled nature. Moreover, the abolishment of education-based migration policy and the relaxation of the work-based migration are found to have limited effects on per capita output and urbanization.

We are grateful for comments from Rick Bond, Kaiji Chen, Victor Couture, Suchin Ge, Chang-Tai Hsieh, B. Ravikumar, Ray Riezman, Michael Song and Dennis Yang, two insightful referees, an associate editor, as well as participants at the AREUEA-ASSA Annual Meetings, the Asian Meetings of the Econometric Society, the Asian Bureau of Finance and Economic Research Conference, the Midwest Macro Meetings, the Public Economic Theory Conference, the Society for Advanced Economic Theory Meeting, the Symposium on Growth and Development, the Society for Economic Dynamics Meeting, and the Regional Science Association International Meeting, and seminar participants at Academia Sinica, Chinese University of Hong Kong, National Taiwan University, National Sun Yat-sen University, Washington University in St. Louis and University of Washington-Seattle. Travel support from the Chinese University of Hong Kong, Tsinghua University, and the Center for Dynamic Economics of Washington University are gratefully acknowledged. An earlier draft has been circulated as NBER Working Paper #23939. Liao thanks the research grant from the Ministry of Science and Technology of Taiwan (MOST 103-2410-H-001-016-MY2). Yip acknowledges support received through a GRF grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (No. CUHK 14501618). Needless to say, the usual disclaimer applies.

✉ Yin-Chi Wang
yinwang@mail.ntpu.edu.tw

Extended author information available on the last page of the article

Keywords Educational choice · Rural–urban migration · Urbanization · Skill composition · Development

JEL Classification O15 · O53 · R23 · R28

1 Introduction

During the post-WWII period, many developing countries have experienced rapid structural transformation from traditional agricultural societies to modern economies. Accompanied by industrialization is a continual process of rural to urban migration, with labor force shifting toward more productive sectors in cities. Its importance has led to a renewed interest in studying structural change induced rural–urban migration, decades after the celebrated contribution by Todaro (1969) and Harris and Todaro (1970). This newer literature has focused primarily on work-based (WB) migration, with two noticeable exceptions by Bénabou (1996) and Lucas (2004).[1] This is somewhat surprising: Since an influential workshop on "Education and Migration" held at Liverpool University (UK) organized by the Education Study Group of the Development Studies Association (ESGDSA), many empirical scholars in the areas of education and economics have identified a positive relationship between migration propensities from rural to urban areas and educational attainment. Such empirical evidence may suggest that better urban education induces internal migration or the better educated to migrate to urban. While the former may be called "education-based migration," the latter may be referred to as "migration-induced education." In this paper, we shall fill the knowledge gap by constructing a dynamic general equilibrium model of education-based (EB) migration and then fitting the model to data to examine its macroeconomic consequences for economic development, urbanization and city workforce composition.

Using census-based, internationally compatible dataset put together by Bernard et al. (2018), one may study (i) (total) migration intensities measured by the ratio of migrants to total population at age 15 or above and (ii) ages at peaked migration intensities. Figure 1 shows a key stylized fact: Age at migration peak is *younger* in countries with higher migration intensities—some of those young migrants appeared to be not purely work-based. In a subsample of the above dataset (only 10 countries available, all developing economies), reasons for migration are collected. We find that EB migration in these developing countries accounted for about 13 percent of total migration, comparable to the work-based figure of 16 percent.[2] Thus, the evidence provides an empirical ground on which our paper is designed to understand the individual decision on EB migration in dynamic general equilibrium.

[1] Bénabou (1996) stresses on within municipal relocation for better schooling and the resulting phenomenon of human-capital based locational stratification. Lucas (2004), on the contrary, emphasizes on an important force for migration—namely, to accumulate human capital when working in a city—which may be viewed as work-based migration with educational purposes.

[2] The subsample includes China, Cambodia, Colombia, Egypt, India, Indonesia, Iran, Iraq, Mexico and Thailand. More than 70 percent of migrants are for other reasons such as marriage and relatives.

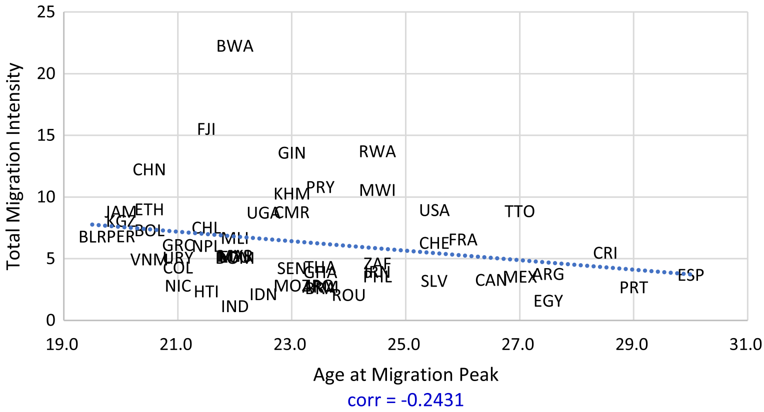

Fig. 1 Age at migration peak and migration intensity

As in the strand of the intergenerational human capital transmission literature, we construct a two-period overlapping generations framework to model rural–urban migration, where altruistic parents make crucial education-migration decisions for their children, allowing for intergenerational human capital accumulation and income mobility. Specifically, rural parents decide whether to send children to urban areas to receive high-quality education. This EB migration would take place prior to the participation in the job market. As stressed by Heckman (1976) and Rosen (1976), schooling not only leads to higher initial human capital at the entry to the job market but also improves the efficacy of on-the-job learning. That is, those sent by parents to take high-quality education in cities are expected to accumulate human capital at higher rates under the learning mechanism elaborated by Lucas (2004). For completeness and fair comparison, we introduce WB migration which does not require parental investment—as a result, we model the WB channel as a lottery draw for simplicity. Finally, to better understand the role of EB migration played in the process of economic development, we incorporate various institutional factors that may affect EB and WB migration differently.

We establish sufficient conditions under which the EB migration motive is positive. We characterize a unique cutoff in children's ability so that those whose ability above it will be sent to urban areas for higher education. The sufficient conditions require that (i) the probability of finding an urban job via education is reasonably high (Assumption 1) and rewarding (Condition NM) and (ii) the positive "intergenerational effect" to dominate the negative "direct consumption effect" at least for some parents whose children are sufficiently talented (Condition I). Basically, the sufficient conditions ensure that the expected net payoff of college education dominates the outside options inclusive of WB migration and rural production.

We further delve into the theory by examining the comparative statics. We refer to the channel via the direct incentives for EB migration as the partial-equilibrium effect. The comparative static outcomes are complicated because of a general-equilibrium effect via changes in employment and wages. Nonetheless, we show that, if (i) the probability for the high-skilled to get a low-skilled job in urban areas is higher than that for the low-skilled to migrate to cities via lottery draw, (ii) the probability for the high-skilled to get a high-skilled job is sufficiently low, (iii) the wage markdown of the high-skilled is sufficiently large, (iv) human capital of the high-skilled is not too high, and (v) the urban-rural total factor productivity (TFP) gap is not too large (Condition W1), then the general-equilibrium effect reinforces the partial-equilibrium effect. In this case, more EB migration occurs when (i) children are more talented, college admission is less selective, and education becomes less costly, (ii) the EB migration cost decreases permanently or the WB migration cost increases permanently, (iii) the chance for children to obtain a high-skilled urban job rises or the chance for children to encounter a low-skilled migration falls. We also establish a sufficient condition (Condition W2) under which the general-equilibrium effects of the aforementioned parameter changes always dampen the partial-equilibrium effects, leading to generally ambiguous comparative-static outcomes. Such potentially opposite effects require us to check the dominance of such effects in the quantitative applications.

To quantify the importance of EB migration for economic development, we calibrate the model to fit data from a large developing economy, China. Based on the international migration data mentioned above, in a group of 55 countries, China is ranked fifth highest in migration intensity, third highest in migration intensity at peak and fourth youngest in age at migration peak. While high migration intensity signifies the study of migration decision and consequences, the young age at migration peak gives greater chances for EB migration. The case of China is interesting also for its various institutional factors that may affect education and migration decisions. Such institutional factors include (i) a household registration system that restricts migration and tightly controls both EB migration and WB migration, (ii) a Guaranteed Job Assignment (GJA) policy, assigning high-skilled jobs to college graduates prior to mid-1990s, and (iii) a rapid rise in education cost since 1990s followed by a rapid college expansion toward the end of the 1990s. Despite its promotion on EB migration, the relative share of EB migration compared to WB has declined over the past three decades as China grew to as the world factory: On average EB migration only amounts to one-fifth of that of WB migration (see Table 1).

We discipline our model to match several key observations from Chinese data during the period from 1980 to 2007 prior to the Great Recession, including: (i) education and work based migration flows, (ii) urban production shares, (iii) high to low skilled employment shares, (iv) urban premium and skill premium, (v) expenditure shares on child rearing, child college education and rural to urban migration, and (vi) Mincerian rate of return of college education. In addition to TFP growth in rural and urban sectors, we are particularly interested in changes in (i) the cost of migration, (ii) the cost and the selectivity of college education, and (iii) the availability of urban low-skilled jobs, as explicitly examined in our model. To properly capture some key policy changes, we separate our sample period into two regimes, 1980–1994 and 1995–2007. Most prominently, the changes include the abolishment of the GJA in 1994, the relaxation of household registration-induced migration barriers since the mid-1990s, as well as the rise in college tuition and the expansion of college admission toward the second half of the 1990s, where all changes occurred around mid-1990s, thus granting the validity of the division into two regimes.

Table 1 Migration by reasons

Year	Population outflow	Job transfer	Job assignment	Work or business	Study or training	Other reasons
Percentage						
1985	100.00	29.57	8.04	3.08	11.26	48.05
2000	100.00	5.32	3.76	33.55	6.84	50.53
Average	100.00	17.44	5.90	18.32	9.05	49.29
Population (thousand persons)						
1985	10,770.00	3184.23	866.43	331.75	1212.81	5174.78
2000	21,580.00	1148.73	810.36	7240.20	1475.87	10,904.83
Annual Growth	4.74%	− 6.57%	− 0.45%	22.82%	1.32%	5.09%

Data source: Migration by reasons (percentage) is obtained from *The 10 Percent Sampling Tabulation on the 1990 Population Census of the People's Republic of China* and *The Tabulation on the 2000 Population Census of the People's Republic of China*. Migration reasons include migration due to job transfer, job assignment, work or business, study and training, to relative and friend, retired or resigned (1985 data only), moved with family, marriage, pull down and move (2000 data only) and other reasons. We categorize migration due to job transfer, job assignment and work or business as work-based migration, and migration due to study or training as education-based migration. Note: There is no available national-wide survey on population outflow (rural–urban migration) in China. Thus, we use changes in urban population as a proxy for population outflow. In the table, migrant population by reasons is computed based on the proxy for population outflow

Upon calibrating the model over these two regimes, we investigate the influences of both types of migration on China's development and urbanization, and decompose the effects of macroeconomic and institutional factors. This is done by conducting counterfactual exercises shutting down each of the two migration channels one-by-one and by comparing the counterfactual outcomes with the benchmark counterparts to obtain the contribution of each migration channel to changes in per capita output and other measures. We find that EB migration accounts for 6.3 percent of changes in per capita output, larger than that of WB migration (4.5 percent). Interestingly, even in regime 2 over the sub-sample period of 1995–2007, we obtain a similar pattern for the comparable contributions of EB and WB migration (8.0 and 5.9 percent, respectively), despite that the share of EB migration is only about 20 percent of the WB share. The intuition of the importance of the EB migration to per capita output is closely related to the rise in skill premium over this sub-sample period, as a result of higher urban TFP and expanded employment of low-skilled workers via WB migration. This finding suggests that without including the quality dimension via the education channel, the picture of rural to urban migration in China could be severely misleading.

We also conduct counterfactual policy experiments on various economic and institutional factors. We begin by verifying that the general-equilibrium effects of key parameter changes discussed in the theory all turn out to dampen the direct partial-equilibrium effects. We then find that the TFP growth and the improvement in human capital together account for about two-thirds of changes in per capita output. Surprisingly, the impacts of the termination of GJA and the relaxation of WB migration on per capita output are found limited. This is a result of the conflicting partial- and general-equilibrium effects. The latter finding on WB migration also reinforces our emphasis on the important role played by the quality dimension of migration. Thus, the general concern with the termination of GJA and the much appreciated temporary permits for migrant workers need not be supported by a general-equilibrium framework that incorporates the quality dimension of migration. Last but not least, as college admission in our calibrated benchmark economy is rationed, we further construct an unrationed counterfactual economy. In this case, we find that there would have been more EB migration than that in the benchmark and, as a result, total per capita output, urbanization rates and high-skilled composition in urban areas are strengthened while the skill premium is lower. Due to a reduction in skill premium, the relative importance played by EB migration in this unrationed counterfactual economy is weakened with its contribution dropped by more than 40%.

These nontrivial and somewhat surprising findings signify our contribution to the literature. They are useful for developing countries to better design internal migration and education policy if industrialization accompanied by skill-enhanced output growth is an important objective.

Related Literature

The older literature on migration is mostly empirical adopting reduced-from approach or theoretical under static or partial equilibrium setting. One exception is Glomm (1992), which developed a dynamic general equilibrium model with persistent urbanization along the equilibrium path; another is Lucas (2004) which rested the analysis in a continuous-time lifecycle framework.

The main migration incentive in Lucas (2004) is that after migration workers can accumulate human capital and have larger life earnings in urban. In our paper, the main migration incentives (by parents) is to enable their children to obtain urban residency and possibly obtain high-skilled jobs. Notably, in Lucas (2004), urban workers are all self-employed Robinson Crusoe's and hence there is no direct interactions among them in the benchmark model without an external effect in learning (to be further discussed in Section 5.3 below). In our paper, urban workers, whether high- or low-skilled, are all directly connected via an aggregate production function. This provides a natural avenue of agglomeration economies.

Our paper adopts a two-location, two-period lived overlapping generations model to study a new, namely, education-based, channel of rural–urban migration in China. It can therefore be compared with the recent, dynamic model based studies on job-related internal migration. Bond et al. (2015) examined the effects of reductions in trade and migration barriers on China's growth and urbanization, focusing on China's accession to the World Trade Organization in 2002, highlighting migration barriers as a main driver for the surplus labor in rural areas and sizable rural–urban migration. Laing et al. (2005) constructed a dynamic search equilibrium model to study the macroeconomic consequences of illegal WB migrants in China (the so-called "blind-flow" or peasant flood) due to the presence of surplus labor and labor search frictions. As rural to urban migration may depend not only on the urban-rural wage gap but also on rural land productivity, Ngai et al. (2019) showed that land policy is a major barrier on industrialization in China. Finally, Garriga et al. (2020) studied the housing-market boom in China as a consequence of its structural transformation and the resulting reallocation of labor from rural to urban areas. They found that the rapid increase in urban housing prices can be attributed to this urbanization processes in conjunction with a significant reduction in the associated migration costs.[3]

Methodologically, our emphasis on the idea of parental motivation is in line with Albornoz et al. (2018) who presented a model of endogenous immigration to study how parents, students and schools interact so as to affect school systems and students' performances in host countries. We also echo Ellickson and Zame (2005) who stress on the valuable implications of a competitive model for location in the presence of heterogeneous locations and costly transportation—in our model, rural and urban differ in many aspects whereas migration is costly.

2 The model

To facilitate the study of the continual process of rural–urban migration covering both EB and WB channels, we construct a dynamic spatial equilibrium model with two-period lived overlapping generations making education and location choices.

In order to have a better understanding of our model setup, we first provide an institutional background to support some essential features to govern our modeling strategy (see the online appendix or our working paper, Liao et al. 2021, for a detailed institutional documentation). We begin with two important institutional features that are commonly observed in developing countries. First, we restrict skill acquisition to urban college education only, as usually seen in many developing countries. Second, we permit admission selectivity to be relaxed over time as a result of improved education systems, though education-related costs are rising over time in response to increased education demands.[4]

Because we shall calibrate our model to fit the case of China, it is also worth highlighting two important, China-specific institutional features related to our model of EB migration: the *hukou* or household registration system and the *zhaosheng* or admission policy of higher education. The *hukou* system maintained a tight control that essentially rationed WB migration through the assignment of the *hukou* certificates. With this *hukou* system, it is better justified to model WB migration by a lottery. On the contrary, the *zhaosheng* policy enables much less regulated EB migration. It allows rural students to obtain the urban *hukou* certificate through college education. Accompanied with the *zhaosheng* policy is the GJA policy prior to 1994 that granted high-skilled jobs to college graduates. Generally, we consider the probabilities for college graduates to obtain either a high-skilled or a low-skilled job, or none and hence to return to rural areas after graduation. This setup enables us to capture the

[3] There are other studies investigating quantitatively or empirically the relationship between migration barriers and rural–urban development in China. These studies usually consider static or partial equilibrium settings with different methodologies and research agenda. For brevity, we are thus abstracting them from our literature review.

[4] For the particular relevance to the case of China, we note that, using the 2015 data from the Chinese Ministry of Education, 2541 out of the 2553 (or 99.53 percent) junior colleges, colleges and universities in China were located in prefectural-level cities or municipalities. Moreover, in China, there was a college education expansion since 1998 and a lift of college tuition control since 1990 that induced sharply rising costs of college education toward late 1990s.

GJA policy in a tractable manner, because under such a policy the latter two non-high-skilled job acquisition probabilities can be simply set to zero.

2.1 The basic setup

There are two geographical regions, rural (R) and urban (U), with only the latter location that can offer higher education required for high-skilled jobs.[5] The initial masses of high- and low-skilled workers in urban areas are exogenously given by (N_H, N_L). We restrict our attention to rural–urban migration, thus for the sake of simplicity, leaving reverse migration from urban to rural areas as exogenous. This is consistent with most of the rural–urban migration research that basically abstracts from reverse migration. Under an overlapping-generations setting, agents live for two periods and study passively in the location chosen by their parents during their first period of life. In the second period, they make decisions for a sequence of events that take place simultaneously. Each agent consumes and gives birth to a single child. Given the talent of the children, parents decide whether to send their children to urban areas to have college education. The residency of urban households are assumed to pass from one generation to another. By focusing on internal migration, we assume away natural birth or international immigration so that the total population is constant over time.

2.1.1 Production

Output is produced using labor inputs in either location, rural or urban.[6] We consider two factor-market distortions by introducing two wedges. One is on the factor price side as a result of unequalized valuation of marginal product, as in the standard misallocation literature, e.g., Hsieh and Klenow (2009). Another is on the factor quantity side related to the deviation from the optimal composition of production inputs, which captures the production technique wedge in Uras and Wang (2017) or the factor-technique mismatch in Wang et al. (2018).

The urban technology (with factor quantity distortion) uses both high-skilled and low-skilled labor and is given by

$$Y_U = AF\left(\tilde{N}_H, N_L\right), \quad \tilde{N}_H = (N_H + \psi)\, h, \tag{1}$$

where $A > 0$ is the urban TFP parameter and h is the level of human capital possessed by high-skilled workers. The outcome of urban education is the acquisition of h,

[5] We assume that every person in the economy is entitled to a basic level of low-skilled education. This basic level of education is sufficient to handle the farming job in rural areas and the low-skilled job in urban areas. However, in order to be a high-skilled worker, one has to upgrade herself with a high-skilled education which is only available in urban areas. We also assume that over-qualification is not a problem so that high-skilled workers can always handle low-skilled jobs and rural farming.

[6] We abstract from physical capital to simplify the dynamics and to sharpen the focus on rural–urban migration. Including physical capital into our model will enhance the importance of EB migration under capital-skill complementarity.

which is assumed to be constant.[7] The introduction of the high-skilled labor wedge, ψ, enables us to capture any possible input-quantity distortion in production, allowing us to fill the gap relating the employment ratio to the relative factor price. Quantitatively, this permits us to use employment shares to back out intergenerational mobility, and to use skill premia to pin down the urban relative TFP as well as the high-skilled labor wedge. Finally, we assume F satisfies all the properties of a neoclassical production function in its arguments, $\tilde{N}_H$ and N_L: $\partial F/\partial m > 0 > \partial^2 F/\partial m^2 \left(m = \tilde{N}_H, N_L\right)$, $\lim_{m\to 0} \partial F/\partial m = \infty$ and $\lim_{m\to\infty} \partial F/\partial m = 0$ (Inada conditions) and F is constant returns in $\left(\tilde{N}_H, N_L\right)$.[8]

Since the classic of Harris and Todaro (1970), it is well documented in the economic development literature that the urban labor market is subject to many institutional distortions. To capture this type of factor market distortion, we introduce a labor market wedge $\tau \in (-1, \infty)$ faced by urban firms when hiring high-skilled workers. Denoting $\tilde{w}_H$ as the effective high-skilled wage received by high-skilled workers and w_L as the low-skilled wage, we obtain the urban wage rates as follows:

$$(1+\tau)\,\tilde{w}_H = \frac{\partial Y_U}{\partial \tilde{N}_H} = AF_{\tilde{N}_H}, \tag{2}$$

$$w_L = \frac{\partial Y_U}{\partial N_L} = AF_L, \tag{3}$$

where $F_{\tilde{N}_H} = \partial F/\partial \tilde{N}_H$ and $F_L = \partial F/\partial N_L$.[9] Then we have

$$\tilde{w}_H = \left(\frac{1}{1+\tau}\frac{F_{\tilde{N}_H}}{F_L}\right) w_L, \tag{4}$$

that is, the values of marginal products of high- and low-skilled labor are not equalized in efficiency unit. When $\tau > 0$, the high-skilled labor would suffer a wage markdown.[10]

Turning to rural production, the constant-return production technology uses only raw (or low-skilled) labor:

$$Y_R = BN_R, \tag{5}$$

[7] We can think of h as an index on labor quality or human capital that results from the total number of years in higher education. Because urban education in our model is measured relative to rural, an education reform improving the quality of rural schools can be translated into a reduction in h.

[8] Given our specification of the production technology, the presence of the quantity distortion ψ does not affect any of our analytical findings in the model section. It is only helpful for our quantitative analysis in data matching.

[9] Similar to the quantity distortion ψ, the presence of the factor-price distortion τ does not affect any of our analytical findings in the model section. We only use it for the quantitative analysis in data matching.

[10] It is observed that the high-skilled labor wage of planned economies is usually suppressed. For the case of China, see Maurer-Fazio (1999) for a discussion of this common feature that is generally analyzed in the development literature.

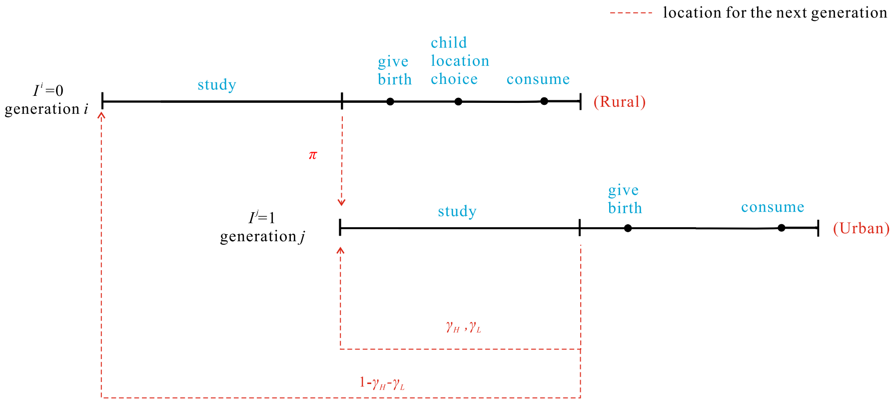

Fig. 2 Timeline of the model

where N_R is the number of "farmers" in the rural area and B is the TFP in the rural area.[11] A competitive labor market implies that the rural wage rate is:

$$w_R = B. \tag{6}$$

2.1.2 Rural households

The economy is populated with all females with each adult woman giving birth to one daughter. The interconnection of dynasties can be fully captured by three adjacent generations, labeled sequentially as (i, j, k). Because rural to urban migration is the focus of our paper, altruistic rural households are the key players in the economy. Rural households of generation i can derive utility from their own consumption (c^i) and their children's consumption (c^j), subject to an altruistic discounting factor $\beta \in (0, 1)$. The generational flow felicity function is common, denoted by $u(\cdot)$ and assumed to be strictly increasing and strictly concave.

For a rural agent, she can relocate to urban if she wins the work-migration lottery draw in her second-period life. However, with only rural education, she can only serve as a low-skilled worker in a city. If she stays in her hometown throughout her adult life, beside consumption, she also chooses, after giving birth to a daughter, whether to acquire urban high-skilled education for her child. Such urban education acquisition is the only way to make her daughter turn into a high-skilled worker. If the mother decides not to acquire urban education for her daughter, the child will repeat the same life span and choices as her rural parent has. To better illustrate the sequence of events within a generation, we plot the timeline of a rural youngster of generation i and her EB-migrated urban child of generation j, in Fig. 2.

[11] We have implicitly assume that rural farming does not require human capital or skill from urban education. So educated high-skilled workers that come back from urban areas do not have higher productivity in rural production.

The representative household's (generation i) objective depends on three consecutive generations (herself i, her child j and her grandchild k) and is given by:

$$\max_{\mathbf{I}^j} \Omega^i \left(\mathbf{I}^j | \mathbf{I}^i = 0, \mathbf{I}^k, x^j \right) = \max_{\mathbf{I}^j} \left[u\left(c^i\right) + \beta \mathbb{E}_X u\left(c^j\right) \right], \tag{7}$$

where β is the altruistic factor on children, and $\mathbf{I}^j$ is an indicator function of migration such that

$$\mathbf{I}^j = \begin{cases} 0 & \text{if generation } j \text{ is not sent to college in an urban area;} \\ 1 & \text{if generation } j \text{ is sent to college in an urban area.} \end{cases}$$

An agent i's discrete choice problem is to decide whether to acquire urban education for her child of talent z^j ($\mathbf{I}^j = 1$ versus $\mathbf{I}^j = 0$), given her rural residency ($\mathbf{I}^i = 0$) and the education choice for her grandchild ($\mathbf{I}^k$). Although her grandchild's education $\mathbf{I}^k$ is chosen by her child, it is indirectly related to the rural parent's decision on the study location of her child. In an independent work on early childhood development, Daruich (2020) emphasized "investing in a child not only improves her skills but also creates a better parent for the next generation". This argument is consistent with our paper: Once an agent sent her child to college with an urban *hukou*, her child is in a superior position to raise the grand child with all urban facilities including college education. As we elaborate below, the talent of the child z^j translates directly into a learning cost variable x^j. Thus, the agent compares $\Omega^i \left(1|0, \mathbf{I}^k, x^j\right)$ to $\Omega^i \left(0|0, \mathbf{I}^k, x^j\right)$ and chooses the highest value between the two.

There are two types of costs in raising children. First, there is a basic requirement for resources, which we assume to be a constant child-rearing cost, denoted by ϕ. Second, there are costs to improve the child's quality, which we can summarize as the education cost. There are two components of the education cost. As high-skilled education is available only in urban areas, there is a constant migration cost for education denoted by σ_e which captures the basic moving expenses.[12] This is the first component of the education cost for children. The second component is the cost of skill acquisition: the learning cost x^j. Since talent matters for learning because people who are more talented study more efficiently, we assume that part of the learning cost depends on the talent of the child. Specifically, x^j is a random variable that depends inversely on both the talents of the child z^j and the college admission selectivity parameter a, and positively on the non-talent related basic learning expenses b:

$$x^j \equiv \chi\left(az^j\right) + b, \tag{8}$$

where $\chi' < 0$, $\chi(0) = \infty$ and $\chi(\infty) = 0$. The college admission selectivity parameter, a, captures the institutional friction of the education system. A decrease in a implies that the urban college education program is more selective in admission so that the learning cost x^j becomes higher for the child with the talent z^j. We note that

12 The migration costs can be interpreted as the costs of obtaining the legal right to stay in cities, transportation costs between hometowns and cities and urban living costs.

$z^j \in \left(z^j_{\min}, \infty\right)$ is drawn from a distribution with cumulative distribution function denoted by $G\left(z^j\right)$, and $z^j_{\min}$ is the minimum support of the talent distribution. For simplicity, we assume that parents perfectly observe children's talent draw and that children's ability in college learning does not affect the human capital measure in production. While imperfect observability requires more complicated expected utility maximization, linking human capital to college learning results in ex post heterogeneity within the high-skilled group and hence complicated aggregate production. Either aspect of generalization would reduce the tractability of the model significantly. While heterogeneities in preferences lead to different locational choices in Castro et al. (2021), heterogeneities in children's innate abilities in *higher education* (e.g., Futagami and Ishiguro 2004; Eckwert and Zilcha 2020) play the key role in different migration decisions in our model. We assume that $z^j_{\min} \leq z^j_0$ which is a cutoff level defined as:

$$w_R - x^j_0 - \sigma_e - \phi = 0,$$

where $x^j_0 \equiv \chi\left(az^j_0\right) + b$. That is, z^j_0 is the talent of the marginally affordable child whose education and rearing costs fully exhaust the income of her rural parent (i.e., $c^i = 0$). As a result, children whose talent z^j that is less than or equal to z^j_0 would not be sent by their parents to acquire urban education. Thus**,** the budget constraint for a rural parent is:

$$c^i + \mathbf{I}^j \cdot \left(x^j + \sigma_e\right) + \phi = w_R. \tag{9}$$

Notably, while there is no income variation within the rural area, allowing children to have different abilities in schooling implies individual parent's expenditure and net income for consumption purposes are all different.

Children who are sent to urban areas become high-skilled after receiving their education. Following the pivotal studies by Todaro (1969) and Harris and Todaro (1970), we assume that they are not guaranteed upon graduation to be high-skilled workers. Specifically, as a college graduate, she may be (i) a high-skilled worker with probability γ_H earning a wage $w_H = \tilde{w}_H h$, (ii) a low-skilled worker with probability γ_L earning a wage w_L, or (iii) unable to find an urban career and forced to return to rural to become a farmer with probability $1 - \gamma_H - \gamma_L$ (reverse migration) earning a rural wage w_R.[13] Children that remain in the rural area do not incur any cost in education or migration for their parents. When these children turn into adults, they either may get recruited via a lottery as low-skilled workers in urban areas with a probability π to earn w_L or work in the rural area to earn w_R. The resulting valuation are equalized in the sense of Todaro (1969) and Harris and Todaro (1970) when taking into account the fact that more low-skilled are drawn in would push down the urban low-skilled wage and thus there must be a value of π consistent with the "net" rural–urban migration

[13] To focus on the endogenous decision of EB migration, we abstract from the decision of reverse migration from urban to rural as the latter requires the explicit modeling of the optimization problem of an urban household. Nonetheless, we conduct robustness checks quantitatively, as reported in Table 8, for various values of this reverse migration probability by varying γ_L.

(i.e., migration inflows to cities net of outflows) given the ratio of low-skilled workers to total rural population.[14]

The expected income earned by a household in generation j in the adulthood is given by:

$$W^j = \mathbf{I}^j \left[\gamma_H w_H + \gamma_L w_L + (1-\gamma_H-\gamma_L)\, w_R\right] \\ + \left(1-\mathbf{I}^j\right)\left[(1-\pi)\, w_R + \pi\,(w_L-\sigma_w)\right], \tag{10}$$

where σ_w is the constant WB migration cost for the low-skilled workers.[15] Then, the children's budget constraint is given by:

$$c^j + \mathbf{I}^k \cdot \left[\mathbf{I}^j\,(1-\gamma_H-\gamma_L) + \left(1-\mathbf{I}^j\right)(1-\pi)\right]\left(x^k+\sigma_e\right) + \phi = W^j, \tag{11}$$

where

$$\mathbf{I}^k = \begin{cases} 0 & \text{if children do not send generation } k \text{ (grandchildren) to college in an urban area,} \\ 1 & \text{if children send generation } k \text{ (grandchildren) to college in an urban area,} \end{cases}$$

and $\left(x^k+\sigma_e\right)$ are the total costs of grandchild going to college in cities. When households of generation i decide $\mathbf{I}^j$, x^k is unknown. We use X to denote the random variable of education cost in their objective function Ω^i.

To compute $\Omega^i\left(1|0,\mathbf{I}^k,x^j\right)$ and $\Omega^i\left(0|0,\mathbf{I}^k,x^j\right)$, we substitute $c^i = w_R - \mathbf{I}^j \cdot \left(x^j+\sigma_e\right) - \phi$ and $c^j = W^j - \mathbf{I}^k \cdot \left[\mathbf{I}^j\,(1-\gamma_H-\gamma_L) + \left(1-\mathbf{I}^j\right)(1-\pi)\right] \cdot \left(x^k+\sigma_e\right) - \phi$ into the value functions Ω^i, where W^j is given by (10):

$$\Omega^i\left(1|0,\mathbf{I}^k,x^j\right) = u\left(w_R - x^j - \sigma_e - \phi\right) \\ + \beta \mathbb{E}_X u \begin{bmatrix} \gamma_H w_H + \gamma_L w_L + (1-\gamma_H-\gamma_L)\, w_R \\ -\mathbf{I}^k(X)\,(1-\gamma_H-\gamma_L)\,(X+\sigma_e) - \phi \end{bmatrix},$$

$$\text{and} \quad \Omega^i\left(0|0,\mathbf{I}^k,x^j\right) = u\,(w_R-\phi) \\ + \beta \mathbb{E}_X u \left[(1-\pi)\, w_R + \pi\,(w_L-\sigma_w) \right. \\ \left. - \mathbf{I}^k(X)\,(1-\pi)\,(X+\sigma_e) - \phi\right].$$

[14] While for the sake of simplicity these probabilities $(\gamma_H, \gamma_L, \pi)$ are exogenous, the scale and shares of migration from these two channels are both endogenous as long as rural households are solving the discrete choice problem to decide on whether to send their children to urban colleges. Thus, this simplifying assumption is viewed innocuous.

[15] The differentiation of migration costs between EB and WB facilitates our understanding on the costs of rural–urban migration. The effects that rural productivity and urban productivity could have on the cost of migration (e.g., via improvement in rural schools, internet access and communication cost, and transport cost) can be captured by these cost parameters. It is also realistic, for instance, college students usually enjoy cheaper housing provided by the universities which migrant workers do not have.

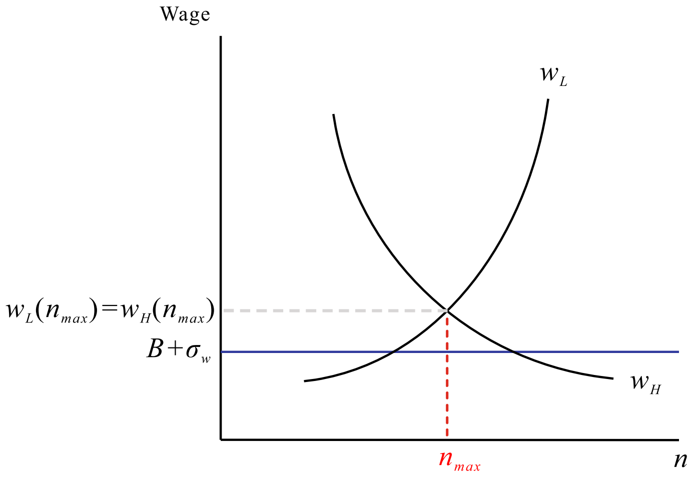

Fig. 3 High- and low-skilled wages and rural wage rate

For comparison, we define $\Delta^i\left(\mathbf{I}^k, x^j\right)$ as the net gain in value for sending children to urban areas to continue their education:

$$\begin{aligned}\Delta^i\left(\mathbf{I}^k, x^j\right) &\equiv \Omega^i\left(1|0, \mathbf{I}^k, x^j\right) - \Omega^i\left(0|0, \mathbf{I}^k, x^j\right)\\ &= u\left(w_R - x^j - \sigma_e - \phi\right) - u\left(w_R - \phi\right)\\ &\quad +\beta\mathbb{E}_X\left\{\begin{array}{c} u\left[\gamma_H w_H + \gamma_L w_L + (1\text{-}\gamma_H\text{-}\gamma_L)\, w_R\text{-}\mathbf{I}^k\,(X)\,(1\text{-}\gamma_H\text{-}\gamma_L)\,(X+\sigma_e)\text{-}\phi\right]\\ \text{-}u[(1\text{-}\pi)\, w_R + \pi\,(w_L\text{-}\sigma_w)\,\text{-}\mathbf{I}^k(X)\,(1-\pi)\,(X+\sigma_e)\,\text{-}\phi]\end{array}\right\}.\end{aligned} \tag{12}$$

Then we have:

$$\mathbf{I}^j = \begin{cases} 0 & \text{if } \Delta^i\left(\mathbf{I}^k, x^j\right) < 0\\ 1 & \text{if } \Delta^i\left(\mathbf{I}^k, x^j\right) > 0.\end{cases}$$

Further, we define $n \equiv (N_H + \psi)h/N_L$ to be the high-skilled to low-skilled labor ratio. Then the high-skilled and low-skilled effective wage in (2) and (3) can be rewritten as:

$$(1+\tau)\, w_H = Af'(n)\, h, \quad w_L = A\left[f(n) - nf'(n)\right], \tag{13}$$

where $Af(n) = AF(n, 1) = Y_U/N_L$. With w_H (w_L) is decreasing (increasing) in n, the high-skilled to low-skilled wage ratio w_H/w_L is decreasing in n with a lower bound at unity. Defining $n_{\max} \geq n$ such that $w_H(n_{\max})/w_L(n_{\max}) = 1$, we impose the following condition:

Condition NM: (Sufficiency for Nondegenerate Migration) $w_H(n_{\max}) = w_L(n_{\max}) > B + \sigma_w$.

If Condition NM holds, then any urban job pays (net of the WB migration cost) better than the rural job. To better understand Condition NM, we plot the high- and low-skilled wages against n in Fig. 3.

Condition NM guarantees that, as long as children can find a job in cities, rural parents will send them to urban areas to attend college. Our next concern is the likelihood of finding a job in the urban area. We impose an assumption on the probabilities of acquiring an urban high-skilled job: The probability of finding an urban high-skilled job via education must be higher than that of finding a low-skilled one through any channels.

Assumption 1 (*Better Job Opportunity for the High-Skilled*) $\gamma_H > \max(\gamma_L, \pi)$.

Assumption 1 states that the probability of securing a high-skilled job after receiving education is higher than the probability of finding a low-skilled job through any channels in the urban area.[16] Thus, Condition NM and Assumption 1 together imply that the expected urban wage income is higher than the rural wage income. Since urbanization and development depend on the composition and relative size of the urban workforce, Condition NM and Assumption 1 simply highlight the fact that urban jobs, especially high-skilled ones, are more attractive than rural jobs to the household. When the talent of the children is sufficiently high, rural parents will then consider sending their children to cities to receive education. As a result, the relative supply of high- to low-skilled workers is expected to rise.

We can easily connect our model to various institutional factors often seen in developing countries. First of all, the relaxation of internal migration restrictions that has raised migrants' chance to get urban jobs is summarized by the probability parameters γ_H, γ_L and π. A relatively higher value of $\gamma_H + \gamma_L$ may be due to better urban job opportunities or as a result of encouraging education policy, both lowering the probability of reverse migration. Next, changes in the education policy that alter the value of the EB migration are given by the admission selectivity parameter a and the basic expenditure parameter b in the learning cost variable x^j. These education parameters provide a short cut for studying the effects of education reforms that affect college admission and tuition. Finally, better transportation system and relaxed migration restrictions can also be captured by the resulting reduction in the moving costs of rural–urban migration given by σ_e and σ_w.

In summary, despite some simplification, the migration setup in our model economy is capable of capturing key factors that affect migration decisions via both EB and WB migration channels—for example, relative TFP in urban and rural areas, urban premium, as well as various education policy and institutional barriers. Nonetheless, we note that a potential endogenous effect not considered here is the rising cost for urban living (including the housing price hike). However, our quantitative results would be "conservative" by shutting down the positive impact of EB migration on the urban cost of living and hence the potentially negative impact on WB migration. Should we include such an effect, the relative importance of EB migration would be even strengthened. The reader should be warned that, however, generalization in either direction would make the model intractable, especially because we must examine decision making by three adjacent generations in which the number of urban (high-skilled and low-skilled) and rural workers are state variables as a result of changing migration flows over time.

[16] Note that Assumption 1 implies $\gamma_H + \gamma_L > \pi$.

2.2 Population dynamics

In this section, we study the population dynamics of rural–urban migration. Recall that adults supply labor to the market and that each one gives birth to only one child, so the entire adult population participates in the labor market. Let $\left(N_H^t, N_L^t\right)$ be the number of high-skilled and low-skilled workers in the urban area, respectively, and N_R^t be the rural labor force, all at time t. Denote $J, K \in \{H, L\}$ as the type of jobs for generation-j and generation-k urban workers respectively. Let δ_{JK} be the transitional probability for an urban generation-k worker born to a generation-j urban worker with job J, working as a type K worker in an urban area. Thus, δ_{JK} captures job mobility across generations in urban areas. We then assume:

Assumption 2 (*Parental Skill Transmission*) $\delta_{JJ} > \delta_{JK}$ for $J \neq K$.

Assumption 2 implies that the child is more likely to be high-skilled (low-skilled) when the parent is high-skilled (low-skilled). Given that the residency of urban households are passed from one generation to another, we have:

$$\sum_K \delta_{JK} = 1. \tag{14}$$

Then, the populations of high-skilled, low-skilled and rural workers evolve according to the following law of motion equations:

$$\begin{aligned}
N_H^{t+1} &= \delta_{HH} N_H^t + \delta_{LH} N_L^t + N_R^t \int \mathbf{I}^j\left(z^j, \mathbf{I}^k\right) \gamma_H dG(z^j), \\
N_L^{t+1} &= \delta_{HL} N_H^t + \delta_{LL} N_L^t + N_R^t \left\{ \int \mathbf{I}^j\left(z^j, \mathbf{I}^k\right) \gamma_L dG(z^j) + \int \left[1 - \mathbf{I}^j\left(z^j, \mathbf{I}^k\right)\right] \pi dG(z^j) \right\}, \\
N_R^{t+1} &= \left(1 - \delta_{HH} - \delta_{HL}\right) N_H^t + \left(1 - \delta_{LH} - \delta_{LL}\right) N_L^t \\
&\quad + N_R^t \left\{ \int \mathbf{I}^j\left(z^j, \mathbf{I}^k\right) \left(1 - \gamma_H - \gamma_L\right) dG(z^j) + \int \left[1 - \mathbf{I}^j\left(z^j, \mathbf{I}^k\right)\right] (1 - \pi) dG(z^j) \right\},
\end{aligned}$$

where the initial urban and rural labor forces are denoted by N_H^0, N_L^0 and N_R^0, respectively. Using (14), we can simplify the above law of motion expressions to:

$$N_H^{t+1} = \delta_{HH} N_H^t + \left(1 - \delta_{LL}\right) N_L^t + N_R^t \int \mathbf{I}^j\left(z^j, \mathbf{I}^k\right) \gamma_H dG(z^j), \tag{15}$$

$$N_L^{t+1} = \left(1 - \delta_{HH}\right) N_H^t + \delta_{LL} N_L^t + N_R^t \left\{ \pi + \int \mathbf{I}^j\left(z^j, \mathbf{I}^k\right) \left(\gamma_L - \pi\right) dG(z^j) \right\}, \tag{16}$$

$$N_R^{t+1} = N_R^t \left\{ (1 - \pi) - \int \mathbf{I}^j\left(z^j, \mathbf{I}^k\right) \left(\gamma_H + \gamma_L - \pi\right) dG(z^j) \right\}. \tag{17}$$

Finally, combining (15) and (16), we have:

$$N_U^{t+1} = N_U^t + \underbrace{\underbrace{N_R^t \left\{1 - \int \mathbf{I}^j \left(z^j, \mathbf{I}^k\right) dG(z^j)\right\} \pi}_{\text{WB}} + \underbrace{N_R^t \int \mathbf{I}^j \left(z^j, \mathbf{I}^k\right) (\gamma_H + \gamma_L)\, dG(z^j)}_{\text{EB}}}_{\text{migrants}},$$

where $N_U^t \equiv N_H^t + N_L^t$ denotes the total urban workforce at time t.

3 Equilibrium

We begin by characterizing the decision on EB migration and examining the resulting policy implications by presenting some partial-equilibrium comparative static findings without taking into account general-equilibrium effects of migration on market wages. Upon defining the dynamic competitive spatial equilibrium, we then characterize the equilibrium by performing full comparative statics incorporating the general-equilibrium effects. Finally, we describe a counterfactual economy eliminating the possibility of EB migration that will be used for counterfactual analysis in the quantitative exercises.

3.1 Migration decision and partial-equilibrium comparative statics

To have a better understanding of such comparative statics, we separate the effect on the utility difference of the marginal parent under EB migration into two parts according to (12):

$$\Delta^i \left(\mathbf{I}^k, x^j\right) = \underbrace{u\left(w_R - x^j - \sigma_e - \phi\right) - u\left(w_R - \phi\right)}_{\text{direct consumption effect}} + \underbrace{\beta \mathbb{E}_X \left\{u\left(c_U^j\right) - u\left(c_R^j\right)\right\}}_{\text{intergenerational effect}},$$

where c_U^j denotes the consumption of children if they are sent to an urban area and c_R^j is the consumption of children if they are kept in a rural area. The *direct consumption effect* (DCE) is always negative because parents' consumption is lower due to the costs of urban education, whereas the *intergenerational effect* (IE) is ambiguous. Condition NM and Assumption 1 together assure that the intergenerational effect is positive which is necessary for parents to acquire urban education for their children:

Proposition 1 (Positive Motive for Urban Education Acquisition) *Under Assumption 1 and Condition NM, the intergenerational effect of migration is positive.*

Proof All proofs are relegated to the Appendix. □

The intuition of the above proposition is straightforward. If the probability of finding an urban job via education is reasonably high (Assumption 1) and rewarding (Condition NM), then the higher expected urban wage provides an incentive for parents to pay the costs of their children's education via altruism. Otherwise, urban education would not be a good "investment" from the parents' perspective.

We next examine how the net gain in education $\Delta^i\left(\mathbf{I}^k, x^j\right)$ responds to changes in the parameterization, i.e., we examine whether the "marginal" parent (a parent who is indifferent between sending her child to attend college in an urban area or keeping the child in the rural area so that $\Delta^i\left(\mathbf{I}^k, x^j\right) = 0$) will send her child to receive an education. By characterizing $\Delta^i\left(\mathbf{I}^k, x^j\right)$, we obtain the following proposition for the comparative statics of EB migration from a partial-equilibrium perspective:

Proposition 2 (Urban Education Acquisition) *Under Assumption 1 and Condition NM, more parents will be willing to acquire urban education for their children if*

1. *their children are more talented (z^j higher), college admission is less selective (a higher), or education becomes less costly (b lower);*
2. *the chance for their children to obtain an urban job is higher (γ_H or γ_L higher);*
3. *the chance for their children to encounter a low-skilled migration decreases (π lower);*
4. *the EB migration cost decreases permanently (σ_e lower);*
5. *the WB migration cost increases permanently (σ_w higher).*

We have studied the EB migration decision as an outcome of two opposing effects: a negative direct consumption effect on the parents and a positive intergenerational effect on the offsprings. If the latter dominates the former, then EB migration takes place. Proposition 2 indicates that EB migration is more likely to arise when children are more talented, when urban education better facilitates the acquisition of higher-paid urban jobs and is not too costly, or when WB migration becomes less available. From the cost perspective, it also provides a general guidance under which various institutional factors as well as education and migration policies may affect the process of rural–urban migration and economic development.

3.2 Dynamic competitive spatial equilibrium

In equilibrium, all labor markets clear under the factor prices $\{w_H, w_L, w_R\}$ given by (2), (3) and (6):

$$N_M^{dt} = N_M^t, \quad M = H, L, R, \tag{18}$$

where N_M^{dt} denotes labor demand of type-M workers. In addition, the overall population size for each period is constant:

$$N_H^t + N_L^t + N_R^t = N, \tag{19}$$

where N is the constant population size.

We define the competitive equilibrium for our model:

Definition (Dynamic Competitive Spatial Equilibrium) Under Condition NM, Assumptions 1 and 2, a **dynamic competitive spatial equilibrium** (DCSE) of the model consists of migration choices $\left\{\mathbf{I}^j\right\}$ and wage rates $\{w_H, w_L, w_R\}$, such that for each period

(i) (Optimization) given wage rates $\{w_H, w_L, w_R\}$, $\{\mathbf{I}^j\}$ solves (7) subject to (9), (10) and (11);
(ii) (Market clearing) wage rates $\{w_H, w_L, w_R\}$ satisfy (2), (3) and (6), and labor markets clear according to (18); and
(iii) (Population evolution) given the initial population $\{N_H^0, N_L^0, N_R^0\}$ and the distribution of talent $G(z^j)$, the population evolves according to (15)–(17) and is restricted by (19).

To conclude this section, we show that there exists a nondegenerate DCSE under the following condition:

Condition I: (Interiority for EB Migration) $\beta\pi\sigma_w > b + \sigma_e$.

Condition I ensures that the positive intergenerational effect identified in Proposition 1 dominates the negative direct consumption effect at least for some parents whose children are sufficiently talented. Intuitively, when the children's talent distribution $G\left(z^j\right)$ has an unbounded upper support, this condition requires the EB migration costs incurred ($b + \sigma_e$) to be smaller than the expected altruistic discounted WB migration costs ($\beta\pi\sigma_w$). With this additional condition, we can establish:

Theorem 1 (Nondegenerate Dynamic Competitive Spatial Equilibrium) *Under Assumption 1 and Conditions NM and I, a nondegenerate dynamic competitive spatial equilibrium exists in which a positive measure of parents will acquire urban education for their children whose talents are above a unique cutoff.*

3.3 Partial- versus general-equilibrium effects

The results provided by Proposition 2 can be regarded as partial-equilibrium comparative-static analysis, i.e., it shows the responses of incentive to EB migration given the differential wages in the rural and urban regions. The general-equilibrium outcomes require solving out for these wages based on (2) and (3), which in turn demand the equilibrium urban high-low skilled labor ratio n. To differentiate the partial- versus general-equilibrium effects, we first note that, the difference is due to the employment effect of n which in turn affects the wages. According to (15)–(17), any parameters that influence the migration decision of parents will affect the population transition $\left(N_H^t, N_L^t, N_R^t\right)$. As a result, the effects of the parameter changes on the urban education decision outcomes studied in Proposition 2 are all partial and we are going to compute their general-equilibrium effects in this sub-section. Although the EB migration comparative statics shown in Propositions 2 are partial, deriving the general-equilibrium ones by solving out for n does not alter the intuition or properties that they illustrate. As we are going to show below, it is possible that the general-equilibrium effects reinforce the partial-equilibrium ones under plausible conditions.

We next deliver the general-equilibrium version of the comparative statics findings presented in Proposition 2. To begin, we would like to explain the nature of the general-equilibrium effects: Via migration, the supply of a particular type of labor, high- or low-skilled, changes, subsequently resulting in changes in the respective market wages and the incentives for migration. Since the general-equilibrium effects work through the relative labor supply (n) and hence on wages (w_H and w_L), it is therefore convenient to decompose them into two components: a relative labor supply effect and a labor induced wage effect. The decomposition is done as follows: Consider a permanent change in the parameter Q studied in Proposition 2 ($Q = x^j, \sigma_e, \sigma_w, \gamma_H, \gamma_L, \pi$):

$$\frac{d\left[\Delta^i\left(\mathbf{I}^k, x^j\right)\right]}{dQ} = \underbrace{\frac{\partial\left[\Omega^i\left(1|0, \mathbf{I}^k, x^j\right) - \Omega^i\left(0|0, \mathbf{I}^k, x^j\right)\right]}{\partial Q}}_{\textit{partial}\text{-equilibrium effect (Prop 2)}} + \underbrace{\frac{\partial\left[\Omega^i\left(1|0, \mathbf{I}^k, x^j\right) - \Omega^i\left(0|0, \mathbf{I}^k, x^j\right)\right]}{\partial n}\frac{dn}{dQ}}_{\textit{general}\text{-equilibrium effect}}$$

$$= \frac{\partial\left[\Omega^i\left(1|0, \mathbf{I}^k, x^j\right) - \Omega^i\left(0|0, \mathbf{I}^k, x^j\right)\right]}{\partial Q} + \beta\mathbb{E}_X\left\{\underbrace{\Gamma(n)}_{\text{labor induced wage effect}} \times \underbrace{\frac{dn}{dQ}}_{\text{labor supply effect}}\right\}$$

where $\Gamma(n)$ is given by,

$$\Gamma(n) = u^j_{c_U}\gamma_H\frac{dw_H}{dn} + \left(u^j_{c_U}\gamma_L - u^j_{c_R}\pi\right)\frac{dw_L}{dn} \tag{20}$$

and $u^j_{c_S} = u_c\left(c^j_S\right)$, $S = U, R$, denote the location-S marginal utilities facing a marginal parent.

We first examine how the partial-equilibrium comparative statics results of Proposition 2 affect n to get the relative labor supply effect in the following lemma:

Lemma 1 (The Relative Labor Supply Effect) *Under Assumption 1 and Condition NM, the relative supply of high- to low-skilled workers (n) rises if*

1. *children are more talented (z^j higher), college admission is less selective (a higher), or education becomes less costly (b lower);*
2. *the chance for children to obtain a high-skilled urban job is higher (γ_H higher);*
3. *the chance for children to encounter a low-skilled migration decreases (π lower);*
4. *the EB migration cost decreases permanently (σ_e lower);*
5. *the WB migration cost increases permanently (σ_w higher).*

However, the effect of a change in the chance for children to obtain a low-skilled urban job (γ_L) on the relative labor supply (n) is generally ambiguous.

We are now ready to study how wages respond to changes in the relative labor supply, i.e., the labor induced wage effect. The next lemma characterizes the labor induced wage component of the general-equilibrium effect. Specifically, it provides sufficient conditions that help to sign the labor induced wage effect $\Gamma(n)$, which measures the expected wage gain from EB migration in response to changes in the relative labor supply.

Lemma 2 (The Labor Induced Wage Effect) *Let*

$$n_c \equiv \frac{\gamma_H h}{(1+\tau)(\gamma_L - \pi)}$$

and $\Upsilon(\bar{n}) = 0$*, where*

$$\Upsilon(n) \equiv n\gamma_L u_c \left(\frac{hAf'(n)}{1+\tau} - \phi\right) - \left(\pi n_{\max} + \frac{\gamma_H h}{1+\tau}\right) u_c(\tilde{w}_R) \tag{21}$$

where $\tilde{w}_R \equiv w_R - (1-\pi)\left(\chi\left(az^k\right) + b + \sigma_e\right) - \phi$ *and* $w_H(n_{\max})/w_L(n_{\max}) = 1$*. The labor induced wage effect (*$\Gamma(n)$*) can be characterized as follows:*

1. *If* $\gamma_L < \pi$ *or* $n < n_c$*, then* $\Gamma(n) < 0$*;*
2. *If* $n > \max\{\bar{n}, n_c\}$*, then* $\Gamma(n) > 0$.

Following Lemmas 1 and 2 above, we consider two sufficient conditions for signing $\Gamma(n)$:

Condition W1: (sufficient for $\Gamma(n) > 0$) $\psi/N > \max\{\bar{n}, n_c\}$.

Condition W2: (sufficient for $\Gamma(n) < 0$) $\{n_{\max} < n_c\} \cup \{\gamma_L < \pi\}$.
While Condition W1 is a sufficient condition for $\Gamma(n) > 0$, Condition W2 is a sufficient condition for $\Gamma(n) < 0$ (noting that $n_{\min} = \psi/N_U$ and $N_U < N$). To study the role played by these conditions on the determination of the general-equilibrium effects of a permanent change in parameter Q on EB migration, we recall

$$\frac{d\left[\Delta^i\left(\mathbf{I}^k, x^j; Q\right)\right]}{dQ} = \frac{\partial\left[\Delta^i\left(\mathbf{I}^k, x^j; Q\right)\right]}{\partial Q} + \beta\mathbb{E}_X\left\{\Gamma(n) \times \frac{dn}{dQ}\right\}. \tag{22}$$

The partial effects of Q is given by the first term on the RHS of (22), which are characterized in Proposition 2. For given wages, an increase in $\left(\gamma_H, \gamma_L, a, z^j, \sigma_w\right)$ or a decrease in (b, π, σ_e) raises the likelihood to earn a higher urban net wage via EB migration channel and hence raises the relative gain of EB migration. The second term highlights the general-equilibrium consideration under a change in Q. It works through the change in the urban high- to low-skilled labor supply and hence the urban wages. Recall from Lemma 1 that the effect of a change in the chance for children to obtain a low-skilled urban job (γ_L) on the relative labor supply (n) is ambiguous. As a result, the general-equilibrium wage effects of γ_L, namely, $\Gamma(dn/d\gamma_L)$, cannot be signed analytically. With regard to changes in other parameters, we have the following general-equilibrium version of Proposition 2:

Proposition 3 (The General-Equilibrium Comparative Statics) *Under Assumption 1 and Conditions NM and I, a DCSE possesses the following properties:*

1. *When Condition W1 is met, the general-equilibrium wage effects of a change in* $z^j, a, b, \sigma_e, \sigma_w, \gamma_H,$ *or* π *always reinforce the partial-equilibrium effects and more EB migration occurs if*

(a) children are more talented (z^j higher), college admission is less selective (a higher), education becomes less costly (b lower), the EB migration cost decreases permanently (σ_e lower), or the WB migration cost increases permanently (σ_w higher);

(b) the chance for children to obtain a high-skilled urban job rises (γ_H higher);

(c) the chance for children to encounter a low-skilled migration falls (π lower);

2. *When Condition W2 is met, the general-equilibrium wage effects of a change in $z^j, a, b, \sigma_e, \sigma_w, \gamma_H,$ or π always dampen the partial-equilibrium effects, leading to generally ambiguous comparative-static outcomes.*

Notably, Condition W1 is more likely to be satisfied if $\{\bar{n}, n_c\}$ become smaller which requires: (i) the probability for the high-skilled to get a low-skilled job be higher than lottery draw for the low-skilled to migrate to cities ($\gamma_L > \pi$); (ii) the probability for the high-skilled to get a high-skilled job (γ_H) be sufficiently low; (iii) the wage markdown of the high-skilled (τ) be sufficiently large; (iv) human capital of the high-skilled (h) be not too high; and (v) the urban-rural TFP gap (A/B) be not too large. Under this condition, it is guaranteed that the direct positive partial-equilibrium effect of the aforementioned parameter changes on the EB migration is accompanied by a reinforcing increase in the expected wage gain from EB migration, thereby leading to definite comparative statics. On the contrary, Condition W2 is more likely to be satisfied if $\{\bar{n}, n_c\}$ become larger which requires: (i) the probability for the high-skilled to get a high-skilled job (γ_H) be sufficiently higher than that of a low-skilled job (γ_L); (ii) the wage markdown of the high-skilled (τ) be sufficiently small; (iii) human capital of the high-skilled (h) be sufficiently high. In this case, the direct positive partial-equilibrium effect may be overturned by the induced reduction in the expected wage gain via rising relative labor supply (n), thus causing ambiguity in comparative statics.

As shown in Proposition 3, depending on the parameterization, the general-equilibrium effect of a migration-related parameter change can work against the partial-equilibrium effect, thereby leading to ambiguous net effects on EB migration. In this case, we will source to quantitative analysis to conclude plausible outcomes based on a calibrated economy.

3.4 A counterfactual economy with no education-based migration

Before closing the theory, we note that, in the absence of EB migration, we have $\mathbf{I}^i = \mathbf{I}^j = \mathbf{I}^k = 0$ and hence the representative household's expected utility is:

$$u(w_R - \phi) + \beta \mathbb{E}_X u[(1-\pi) w_R + \pi (w_L - \sigma_w) - \phi]. \tag{23}$$

The populations of high-skilled, low-skilled and rural workers evolve according to the following law of motion equations:

$$N_H^{t+1} = \delta_{HH} N_H^t + \delta_{LH} N_L^t,$$
$$N_L^{t+1} = \delta_{HL} N_H^t + \delta_{LL} N_L^t + \pi N_R^t,$$

$$N_R^{t+1} = (1 - \delta_{HH} - \delta_{HL})\, N_H^t + (1 - \delta_{LH} - \delta_{LL})\, N_L^t + (1 - \pi)\, N_R^t. \quad (24)$$

In equilibrium, all labor markets clear under the factor prices $\{w_H, w_L, w_R\}$:

$$N_M^{dt} = N_M^t, \quad M = H, L, R.$$

Finally, the overall population size for each period is constant as before:

$$N_H^t + N_L^t + N_R^t = N.$$

In our counterfactual quantitative analysis when EB migration is absent, these changes will be modified accordingly. If WB migration is further eliminated ($\pi = 0$) in the counterfactual economy, then rural–urban migration ceases completely. Both scenarios will be studied in Sect. 4.2.

4 Quantitative analysis

This section presents a calibrated version of our model to study the contribution of the EB migration to the development of the Chinese economy within the post-reform regime but before the financial tsunami, namely, over the period of 1980–2007. During this period, China has experienced rapid economic growth and urbanization. Real per capita GDP has grown at an annual rate of approximately 6.0 percent, whereas the comparable figure since Deng Xiao-Ping's Southern Trip in 1992 is 7.6 percent. Meanwhile, as shown in Fig. 4, urbanization rates (urban population shares) and urban value-added shares have increased from 19.4 to 44.9 percent and from 66.7 to 87.3 percent, respectively, and the migration flows (proxied by changes in urban population) over rural population have nearly quadrupled, increasing from 0.5 to 1.9 percent.[17] Concurrently, more rural students were attending colleges because of the college expansion in the late 1990s, while empirical studies have pointed out the phenomenon that fewer rural students were admitted to top universities. The above observations motivate us to take China as an interesting example for our quantitative analysis.

Since most of the high-quality universities are located in large metropolises in China, we consider cities as places for higher education to take place.[18] This complements Lucas (2004) who views cities as places for immigrants to accumulate human capital when working. In so doing we explore a potentially important, EB channel of rural to urban migration which has become a unique channel that mitigates migration barriers in China. Students attend colleges by passing the National College Entrance Examination to migrate to cities. This institutional migration channel enables us to examine the role of the EB migration in the development of China and to compare

[17] For urban output shares, urbanization rates and migration outflows, the correlation coefficients range from 0.71 to 0.96.

[18] See the online appendix for the related literature, the detailed discussion on the rural–urban disparities in college admission rates and the inequality in the distribution of educational resources in China.

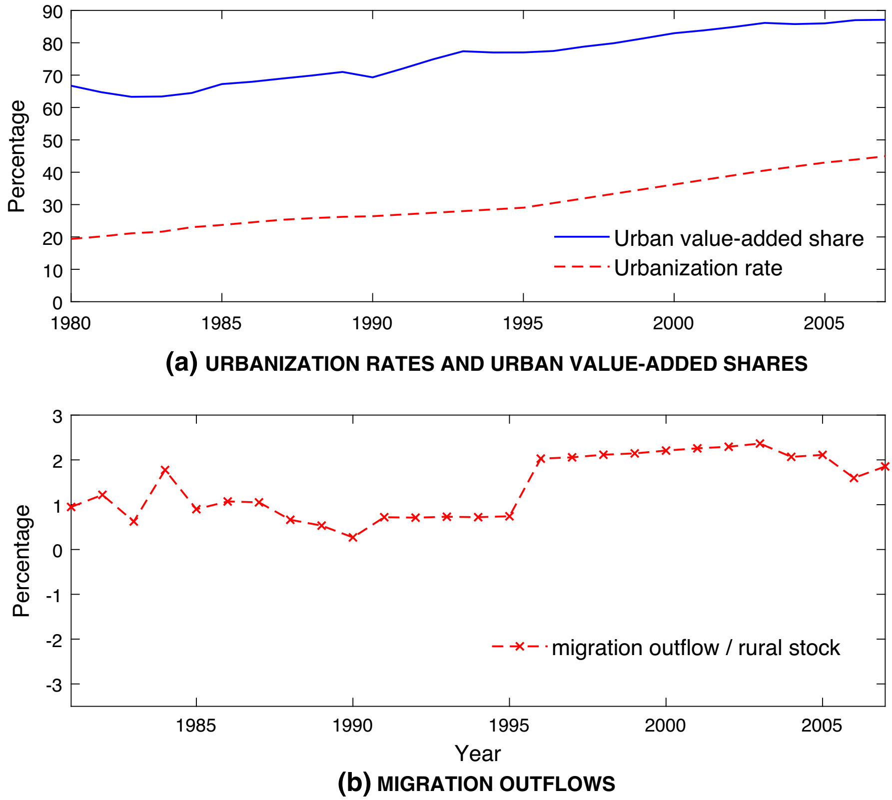

Fig. 4 Urbanization process in China over 1980–2007. Note: Urbanization rate is defined as urban population shares out of total population. Urban value-added share is defined as the total value-added share of the industry sector, the construction sector and the service sector. See the appendix for the data source. Because there is no good data on migration, we use changes in urbanization as a proxy for migration outflow

the importance of this channel to that of WB migration. Moreover, due to the college expansion and the facts that most universities are located in cities and urban high-skilled jobs are much better paid, one would expect that there shall be more youngsters migrating to cities for higher education. However, as shown in Table 1, the number and the annual growth rate of WB (inclusive of job transfer, job assignment, and work or business) migration far outweighed those of EB (including studying or training). Therefore, it is worth to examine factors that shaped rural youngsters' migration patterns and the causes leading to the growing difficulty for rural students to attend top colleges in urban areas.

Because of the major changes in the macroeconomic environment for migration and education starting in the mid-1990s, we break the entire period into two sub-periods: regime 1, spanning from 1980 to 1994, and regime 2, ranging from 1995 to 2007.[19] We first provide the calibration and simulation methodology for the quantitative model.

[19] Analogous to our theoretical model, we consider the whole Chinese economy to be two geographical regions, rural and urban, and dismiss the differentiation of within- and cross-provincial migration.

We then decompose the contribution of EB and WB migration to the development of the Chinese economy. Robustness tests are conducted to reexamine the importance of EB and WB migration. Lastly, the quantitative effects of changes in key parameters are assessed by an analysis of factor decomposition. The investigation of factor decomposition is relegated to Sect. 5.

4.1 Calibration and Simulation

When bringing a two-period overlapping generations model of rural–urban migration to the data, one always faces the problem: How to compromise the long model period and the timing of the decision in the model to data? One method to overcome this problem is to extend the two-period model into a multi-period model for the quantitative analysis.[20] Although the modified quantitative model corresponds its model period to the data period, it is a different model in nature, and one needs to make sure that all the theoretical results are still valid. In our paper, we directly carry the theory to data by considering that each cohort of the rural parents make migration decisions immediately upon turning adults and giving birth of their children. This one-to-one mapping between decision timing and cohorts' generation allows us to assume that every year features repeated cohorts with stationary distribution. In this way, the model period (which we set to be twenty-five years for our two-period overlapping generations model of rural–urban migration) and the annual data become consistent with each other.[21]

To proceed with the quantitative analysis, we first perform a two-regime calibration to match the regime average data to pin down the regime-common and regime-specific parameters. The former category includes deep parameters in preferences, technologies and the talent distribution, whereas the latter category consists of parameters that describe the specific environment of the regime, such as urban and rural TFP, job finding rates, migration and education costs. The decision rules solved from the two-stage calibration can thus be interpreted as the "mean-year" agent's decision rules in each regime. Given these regime parameters, we then solve the annual decisions from the EB migration flow data. We calibrate the urban TFPs and the distortionary wedges annually by matching the skill premium and the urban premium data.

Below we describe how we conduct the two-regime calibration and the calibration for the annual TFPs and distortions. The reader is reminded that the return to education is measured in *urban-to-rural differences* throughout the paper. Calibration details and data sources are relegated to the online appendix.

4.1.1 Two-regime calibration

Total population is normalized to one in every period. Urban (rural) population is the share of urban (rural) to total population and is computed using the data on populations

[20] For an example of this approach, see Song et al. (2011).

[21] The limitation of this approach is that one will not be able to analyze the age composition of workers in the quantitative analysis. As aging related issues (such as pension and population dividends) are not our focus, once the workers evolution equations are taken care of and the model implied population stock ratios are matched to the data, this approach shall yield similar results to that of a quantitative multi-period model.

by rural and urban residence. We term workers with educational attainment of college and above (below) as high (low)-skilled. Then, using the data on urban employment by educational attainment and urban population, we compute the stocks of high- and low-skilled workers.

The utility function is assumed to take the standard CRRA form:

$$u\left(c\right)=\frac{c^{1-\varepsilon}-1}{1-\varepsilon},\quad \varepsilon>1,$$

where ε is the inverse of the elasticity of intertemporal substitution (EIS). In the literature, the Pareto distribution is commonly associated with wealth and income, which are believed to be closely related to one's talent.[22] Therefore, we assume that children's talents z^j follow a Pareto distribution, with the CDF given by:

$$G\left(z^j\right)=1-\left(\frac{z_{\min}}{z^j}\right)^{\theta},\quad z^j\geq z_{\min},$$

where $z_{\min}$ and θ are the location and shape parameters of the Pareto distribution, respectively. The learning cost x^j is inversely related to z^j and is assumed to take the form of:

$$x^j=\frac{1}{az^j}+b.$$

With this setup, the higher the college admission selectivity parameter a is, the less selective the college admission is, and the lower cost born by rural parents to acquire urban education for their children. The urban production Y_U takes the following form:

$$Y_U=AF\left(\tilde{N}_H,N_L\right)=A\left[\alpha\tilde{N}_H^{\rho}+(1-\alpha)\,N_L^{\rho}\right]^{1/\rho},\ \alpha\in(0,1)\,,\ \rho<1$$

where $\tilde{N}_H=(N_H+\psi)\,h$, $1/\left(1-\rho\right)$ is the elasticity of substitution in production for high-skilled and low-skilled inputs, and α is the distribution parameter (which yields the high-skilled labor income share when $\rho=0$). Below, we first describe the preset common parameters and then the preset regime-specific parameters, followed by the methods of identifying the remaining parameters.

China is well known for its high saving rates and low annual time preference rates. We thus set the annual time preference at 1 percent, which is close to Song et al. (2011). The parental altruistic factor for children β is hence equal to 0.7798. The inverse of the EIS parameter ε is set at 1.5, which is common in the literature. There is no nationwide survey of child-rearing costs for rural China. We follow the estimate in the literature to set ϕ such that the percentage of the child-rearing cost to rural household income $\tilde{\phi}$ is 17.4 percent in both regimes.[23] For the Pareto distribution parameters, we normalize $z_{\min}$ to one as typically set in the literature.[24] Since talents are unobservable but are

22 For example, Feenberg and Poterba (1993) assumed that the U.S. income follows a Pareto distribution. Their estimated Pareto shape parameter for the U.S. over the 1980–1990 period is 1.92.

23 See the online appendix for the details.

24 See, e.g., Ghironi and Melitz (2005), Bernard et al. (2003), and Eaton and Kortum (2002).

found to be correlated with income levels, we set θ to 2.5 using rural household net income data from the Chinese Household Income Project (CHIP). Our value is close to the estimate for the United States. The last preset common parameter is the elasticity of substitution between high-skilled related inputs and low-skilled labor, $1/(1-\rho)$. We proxy it by the estimates on the elasticity of substitution between high- and low-skilled workers. As pointed out by previous studies, the estimated values for Asian economies are usually larger, mostly falling between 2 and 7, than those for developed countries, ranging from 1 to 3. We thus choose $1/(1-\rho)$ to be 3 so that ρ equals 0.6667.

Denote $\tilde{\sigma}_e$ ($\tilde{\sigma}_w$) as the EB (WB) migration cost as a percentage of rural household income. Considering WB migration cost as urban living costs and the required costs for moving to urban areas, we compute $\tilde{\sigma}_w$ from CHIP with the periods over 1980–2002 and obtain a value of 55.54 percent and 30.79 percent of rural household income for regimes 1 and 2, respectively. EB migration cost includes the costs of food and dormitory for a college student. Assuming that a student stays in college for four years and adjusting for model periods, we obtain the EB migration cost $\tilde{\sigma}_e$ to be 0.1021 for regime 2. The data on EB migration cost prior to 1996 is not available, so we compute $\tilde{\sigma}_e$ for regime 1 by assuming that $\tilde{\sigma}_e$ and $\tilde{\sigma}_w$ grow at the same rate between the two regimes and obtain $\tilde{\sigma}_e = 0.1841$ for regime 1.[25]

The main spirits of China's education reforms are captured by the endogenous threshold in talents, controlling admission selectivity parameter (a) and the cost of college education (b). We will address how to pin down the threshold talent and a using model equations later. Similar to the cases for calibrating EB and WB migration costs, we denote $\tilde{b}$ as the college education cost as a percentage of rural household income. College education was almost free of charge before 1990. Thus, we set $\tilde{b}$ in regime 1 to only include stationary, materials and textbooks while $\tilde{b}$ in regime 2 further includes tuition costs. Using Urban Household Survey (UHS) 2007 and 2008, $\tilde{b}$ equals 0.48 percent and 5.28 percent of rural household income in regimes 1 and 2, respectively.[26]

We note that by calibrating σ_e, σ_w and b in ratios of rural incomes, urban and rural TFPs or any productivity effects of education via the human capital measure h would have quantitative impacts on migration and college education costs. Thus, the endogenous effects of income factors on migration decisions via changes in migration and college education costs are quantitatively accounted for in our policy experiments.

Under the linear rural production technology, the scaling factor B is equal to the rural wage rate. Being interested in the relative economic positions of rural and urban China and understanding how regional technological disparities shape individuals' migration decisions, we normalize rural per capita income in 2007 to one. Then we compute the rural per capita income over 1980–2007. The averages of B are 0.3685 and 0.7177 for regimes 1 and 2, respectively. It is notable that such normalization of rural per capita income together with $z_{\min} = 1$ imply that only parents with relatively talented children can afford to send their children to college. This is because rural

25 We relegate the details of computing the WB and EB migration costs to items 11 and 12 in the online appendix.

26 One model period is twenty-five years. Therefore, the cost of college education here is not directly comparable to annual data.

Table 2 Education-based migration flow and the probability of work-based migration

	Education-based migration flow	Prob. of work-based migration
Regime 1	0.00058946	0.003554486
Regime 2	0.00114381	0.008281515

Source: Authors' calculation using the average of 1985 and 2000 migration reasons in Table 1

parents have to maintain their own consumption and pay the child-rearing cost in addition to costs of college education and EB migration.

We now turn to the rates at which college graduates (only originally from rural China) find jobs and the migration probability for rural workers. All the job finding probabilities are those faced by each cohort. Denote $\gamma \equiv \gamma_H + \gamma_L$ as the college graduates' job finding probability, or the urban employment rate for college graduates. During the years of the GJA policy (1951–1994), a college graduate was assigned a stable job (either in the government or in state-owned enterprises), usually in an urban work unit. In contrast, after the termination of the GJA policy, jobs for college graduates were no longer guaranteed. In line with the GJA policy, we set $\gamma = \gamma_H = 1$ and $\gamma_L = 0$ in regime 1, meaning that college graduates from rural China are fully employed as high-skilled workers.[27] In regime 2, $\gamma < 1$. As the data on the employment rate of college graduates from rural China is not available, we use urban employment rates from CHIP in 1995, 2002 and 2007 to proxy for the employment rate of college graduates from rural areas. The average value, 0.9209, is set to be the employment rate in city districts for college graduates in regime 2. Note that γ_L is the job mismatching rate for college graduates, which we do not have information about. We set γ_L to 0.05, and γ_H is thus solved as 0.8709 in regime 2.[28] For the probability capturing the rate of WB net migration flows π, as there is no nationwide survey on rural–urban migration in China during the periods under examination, we use changes in urban population as a proxy for rural–urban migration flows. We compute π as a ratio of migration flows due to work-related reasons to rural population.[29] As reported in Table 2, the average migration probabilities for rural workers π in regime 1 and regime 2 are 0.0036 and 0.0083, respectively.[30]

The next one is the human capital possessed by high-skilled workers relative to low-skilled workers, h. We first compute the average years of schooling for high- and

[27] The main spirit of the GJA policy was that the government provided jobs to college graduates. However, in practice, there existed "mismatch" problems in the job assignment system. Hence, starting in 1983, the government allowed the hiring units and college graduates to meet prior to the job assignment, which has essentially eliminated the mismatch problem (cf. Qi 2015). It is therefore appropriate to assume $\gamma_H = 1$ in regime 1.

[28] As shown in Sect. 4.3, our quantitative results are not sensitive to the choice of γ_L.

[29] See the online appendix for the details. Although Longitudinal Survey on Rural Urban Migration in China provides migration information, it only starts in 2008 and does not cover the period that we examine in this paper.

[30] We notice that migrants with different household registration status would have different urban benefits. However, our focus is the overall contribution of WB migration compared to that of EB. Our calibration is thus employment-based, rather than household registration-based. Nonetheless, considering workers' household registration status would not change our main findings.

low-skilled workers and take Mincerian coefficients from the literature for the two regimes.[31] Following the Mincerian method, we then compute the regime-specific h and obtain 1.3529 and 1.5928 for regimes 1 and 2, respectively. The last preset regime-specific parameters are intergenerational mobility. Assuming that the residences of urban households are passed from one generation to another and allowing upward mobility, we have $\delta_{HH} = 1$, $\delta_{HL} = 0$ and $\delta_{LH} + \delta_{LL} = 1$ in both regimes.[32] The probabilities of remaining low-skilled workers across generations (δ_{LL}) in the two regimes are calibrated to match the N_H/N_L ratios using (15)-(17) and the *EB migration* flows data (computed in the same way as that for migration flows due to employment as reported in Table 2). Thus, δ_{LL} is calibrated as 0.9996 and 0.9883 in regimes 1 and 2, respectively.[33] The fall in δ_{LL} shows that intergenerational mobility in China has improved over 1995–2007.

The regime-specific price distortions τ faced by urban firms when hiring high-skilled workers, the urban TFPs in the two regimes, the CES production high-skilled labor share α and the high-skilled labor wedge ψ are calibrated to match the regime average skill premiums (w_H/w_L), urban premiums (w_L/w_R) and urban production shares (Y_U/Y). The targets of urban production shares contain more information in addition to employment and wage measures. Thus they can serve to calibrate both α and ψ. The calibrated α and ψ are equal to 0.8461 and 0.0618, the regime-specific distortions τ are 7.1103 and 5.4763, and the urban TFPs in the two regimes are equal to 5.3877 and 11.0573, respectively. Our results show that urban TFP is growing faster relative to rural TFP: The implied annual urban TFP growth rate is 5.47 percent. In addition, the price distortion τ faced by urban firms in regime 2 is reduced by more than 22 percent compared to that in regime 1, indicating that the market price distortions due to the planned economy have been greatly alleviated.

Denote $\hat{z}$ to be the threshold in children's talent such that when a child is equipped with the talent $\hat{z}$, her parent is indifferent between sending her to college in urban areas or keeping her in the rural hometown. When a child is talented such that $z^j \equiv 1/\left[a\left(x^j - b\right)\right] \geq \hat{z}$, her parent will definitely send her to college ($\Delta^i(\mathbf{I}^k, x^j) \geq 0$). The endogenous threshold $\hat{z}$ therefore dichotomizes the "destiny" of rural children. Specifically, define $\tilde{N}_E^t$ as the EB migration flow at time t. $\tilde{N}_E^t$ can be written as:

$$\tilde{N}_E^t = N_R^t \int \mathbf{I}^j\left(z^j, \mathbf{I}^k\right) dG(z^j) = N_R^t \left(\frac{z_{\min}}{\hat{z}}\right)^{\theta}. \tag{25}$$

Therefore, $\hat{z} = z_{\min}(N_R^t/\tilde{N}_E^t)^{1/\theta}$ and $\hat{z}$ can be obtained using the EB migration flows data. The computed average $\hat{z}$ for regimes 1 and 2 are equal to 17.7632 and 13.1391, respectively. The decrease in $\hat{z}$ captures the college expansion in China: More rural students are going to colleges. With $\hat{z}$, we can solve the last parameter a by the indifference boundary equation (12). The calibrated a are 1.1489 and 0.4701 for regimes

31 See the online appendix for the details.

32 The average years of schooling in China for people aged 15 and over have increased from 4.86 years in 1980 to 7.51 years in 2010, showing an overall pattern of upward mobility in education.

33 We have indeed matched the N_H/N_L data series, taking into account reverse migration of public employees when computing urban employment rate.

1 and 2. The decrease in a reflects the fact of the draining in rural talents so that college admission is becoming more selective to rural students. This is consistent with the data that it becomes more difficult for rural students to attend top universities in China (e.g. see Yang 2006). Tables 3 and 4 report the calibration results. Based on the above parameters, our next step is to calibrate the annual urban TFP and annual price distortions for 1981–2007 and to perform a simulation to serve as our benchmark model.

4.1.2 Calibration of the annual urban TFP and distortions

To calibrate the annual urban TFP and τ, we first need the annual N_R, N_H and N_L based on the model. Following the same method in the two-regime calibration, we compute the annual *EB migration* flows. Together with the data on N_R, N_H and N_L in 1980 and the calibrated parameters (including γ_H, γ_L, π, δ_{HH}, δ_{HL}, δ_{LH}, δ_{LL} and θ), we solve the threshold $\hat{z}$ of 1980 based on (25). The 1980–1981 WB migration flows are also solved according to the equation: $\tilde{N}_W^t = \pi N_R^{t-1}\left[1 - \left(z_{\min}/\hat{z}\right)^{\theta}\right]$, where $\tilde{N}_W^t$ is WB migration flows. Furthermore, from the evolution of workers equations (15)–(17), we compute the model implied N_R, N_H and N_L for 1981. We then repeat this procedure to obtain annual series for $\hat{z}$, N_R, N_H, and N_L. Assuming that the annual growth rate of human capital is constant over 1980–2007, we compute the annual series of h so that the average human capital in regimes 1 and 2 are exactly equal to those in the two-regime calibration. Finally, with the time series data on rural per capita income, the annual urban TFP and price distortions τ are solved to match the urban premium (w_L/w_R) and skill premium (w_H/w_L).

Figure 5 plots the calibrated urban TFP and rural TFP for 1981–2007. It can be observed that the urban TFP grows relatively faster than the rural TFP after 1985, corresponding to China's economic reform, the privatization of state-owned enterprises and the deregulation of price controls. As reported in Table 5, the relative urban-rural TFP growth rate over our sample period is approximately 0.39 percent per year. Figure 6 provides a comparison between the model and the data on urban production (per capita) and total output per capita.[34] We define the urbanization rates in the model as the shares of urban workers. Figure 7 compares the model to the data on urbanization rates and the stocks of urban high- and low-skilled workers. Our model shows a lower urbanization rate and a smaller stock of low-skilled workers than the data do, with the discrepancies between the model and the data widening over time. The gaps can be explained by the migration flows inputted when we calibrate the model. Because our model only considers two channels of migration, the data on migrants who migrated for non-educational and non-employment reasons (accounting for approximately 50 percent of total migration) are thus excluded in the calibration. However, these migrants could migrate due to other reasons but became low-skilled workers later. As a result, our model underestimates the stock of low-skilled workers and the urbanization rate. Because the model generates fewer workers in urban areas, especially fewer low-skilled workers, the urban production (per capita) in the model is slightly higher than

[34] Based on Eqs. (1) and (5), we use the N_H, N_L and N_R data to calculate urban production (per capita) and total output. See the online appendix for details.

Table 3 Parameters taken from data

	Parameter values		Data source or assumption
	Regime 1	Regime 2	
Regime-common parameter			
δ_{HH}	1	1	Urban residences pass from one generation to another
β	0.7798	0.7798	Annual discount factor=1%; Song et al. (2011)
ε	1.5	1.5	Common setting in the literature
z_{min}	1	1	Eaton and Kortum (2002)
$\tilde{\phi}$	17.4%	17.4%	Zhu and Zhang (1996)
θ	2.5	2.5	CHIPS 1995 and 2002
ρ	0.6667	0.6667	Autor et al. (1998)
Regime-specific parameter			
γ	1	0.9209	Regime 1: due to GJA policy, $\gamma_H = 1$ and $\gamma_L = 0$ Regime 2: urban employment rate, CHIPS 1995, 2002 and 2007
γ_H	1	0.8709	Regime 1: due to GJA policy Regime 2: see Table 4
γ_L	0	0.05	Regime 1: due to GJA policy Regime 2: see Table 4
π	0.0036	0.0083	The probability of work-based migration reported in Table 2
B	0.3685	0.7177	Using rural income in the *China Statistical Yearbook* 2011; Rural income of 2007 is normalized to one
h	1.3529	1.5928	the *China Labor Statistical Yearbook* 2002 and 2009 and the *China Statistical Yearbook* 1998; see the online appendix
$\tilde{\sigma}_e$	18.41%	10.21%	He and Dong (2007); see the online appendix
$\tilde{\sigma_w}$	55.54%	30.79%	CHIPS 2002; see the online appendix
$\tilde{b}$	0.48%	5.28%	UHS 2007 and 2008; see the online appendix

Table 4 Calibrated parameters

	Parameter values		Target	Model result		Data	
	Regime 1	Regime 2	Moment	Regime 1	Regime 2	Regime 1	Regime 2
Regime-common parameter							
α	0.8461	0.8461	Y_U/Y	0.6922	0.8294	0.6922	0.8294
ψ	0.0618	0.0618	Y_U/Y	0.6922	0.8294	0.6922	0.8294
Regime-specific parameter: jointly calibrated							
γ_L	–	0.05	N_H/N_L	0.0424	0.1466	0.0424	0.1466
δ_{LL}	0.9996	0.9883	w_L/w_R	1.7781	2.0076	1.7781	2.0076
A	5.3877	11.0573	w_H/w_L	1.2296	1.6576	1.2296	1.6576
τ	7.1103	5.4763					
$\hat{z}$	17.7632	13.1391	edu. migration flow	0.059%	0.114%	0.059%	0.114%
Regime-specific parameter: solved by model equation							
γ_H	–	0.8709	solved by $\gamma_H = \gamma - \gamma_L$				
a	1.1489	0.4701	solved by equation (12)				

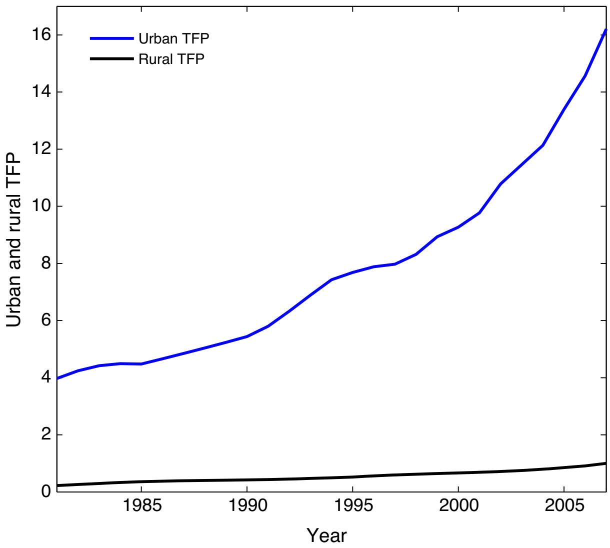

Fig. 5 Calibrated urban and rural TFP during 1981–2007

Table 5 Model implications

	Regime 1	Regime 2	Explanation
$1 - G(\hat{z})$	0.075%	0.160%	Average education-based migration proportion
A/B	14.6188	15.4071	Average urban-rural TFP ratio
ψ_{cost}	0.6459	0.4380	Unit cost reduced by ψ
A_g		5.47%	Average annual growth rate of A from 1981 to 2007
$(A/B)_g$		0.39%	Average annual growth rate of A/B from 1981 to 2007

that observed in the data. Additionally, as there are more rural workers in the model and rural technology is less productive, total output per capita in the model is slightly lower than that observed in the data.

This calibrated economy serves as our benchmark model for the decomposition of migration channels in Sect. 4.2 and for all experiments in Sect. 5. Table 6 summarizes the annual averages of important macroeconomic variables in the benchmark model for regimes 1 and 2 as well as for the entire sample period. As expected, total output per capita in regime 2 is more than doubled that in regime 1, the urban production share increases about 18 percent (from 0.6585 to 0.7754), and the urban employment share increases about 33 percent (from 0.2174 to 0.2883). The increases in urban production and urban employment shares imply that urban production becomes more important in regime 2. Furthermore, our model shows that the high-skilled employment shares in urban areas are more than quadruple in regime 2, while the skill premium still increases. These trends are all consistent with the experience of China's development.

To better understand the channel of EB migration, the reader is reminded that the comparative statics depend crucially on the partial-equilibrium versus the general-

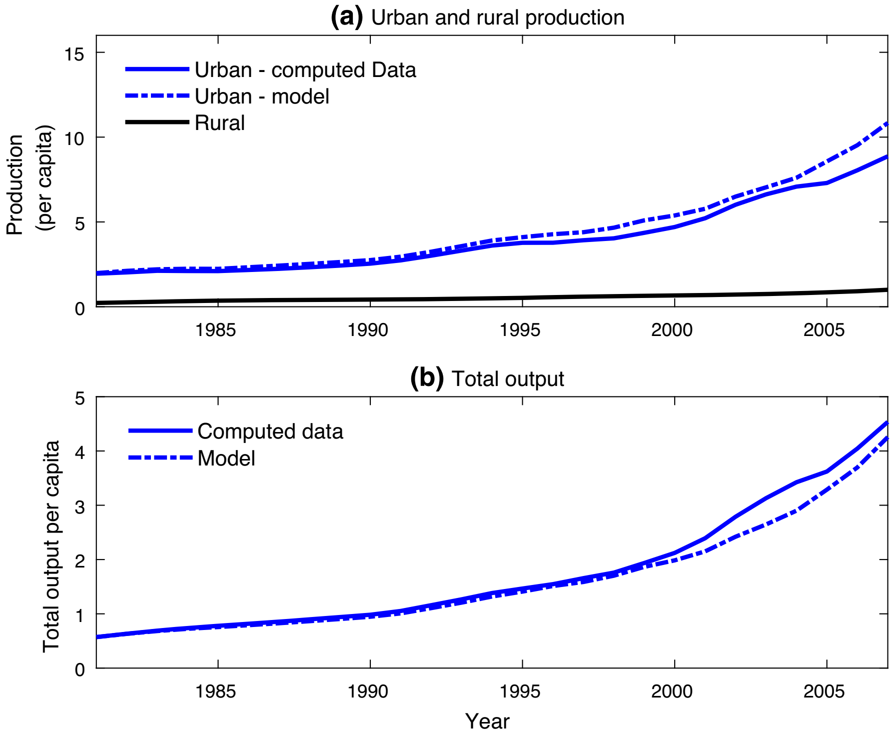

Fig. 6 Benchmark model—production

equilibrium effects. We thus compute in the benchmark calibrated economy several values. First, in Regime 1, $\gamma_L = 0 < \pi$, so by Proposition 3, the general-equilibrium effects of all parameter changes listed there always dampen the direct partial-equilibrium effects. Second, in Regime 2, $\gamma_L = 0.05 > 0.0083 = \pi$. Moreover, we have $n = 0.5426$, $n_{\max} = 2.4714$ and $n_c = 5.1345$. Thus, Condition W2 is met, the expected value of Γ is negative ($= -0.5352$), and by Proposition 3, the general-equilibrium effects again dampen the direct partial-equilibrium effects. The conflicting partial- and general-equilibrium effects are thereby expected regardless of the regimes, which are crucial for some of our quantitative results.

4.2 Decomposition of migration channels

To identify the contribution of each migration channel and to study the total effects of migration on China's development process, we eliminate the migration channels sequentially. The effect of the channel under study is thus the difference between the counterfactual model with the channel being excluded and the benchmark model.

Figure 8 plots urban production (per capita), total output per capita, and urbanization rates under the decomposition. The benchmark model and the three scenarios are plotted for comparison: (i) WB migration is eliminated; (ii) EB migration is eliminated;

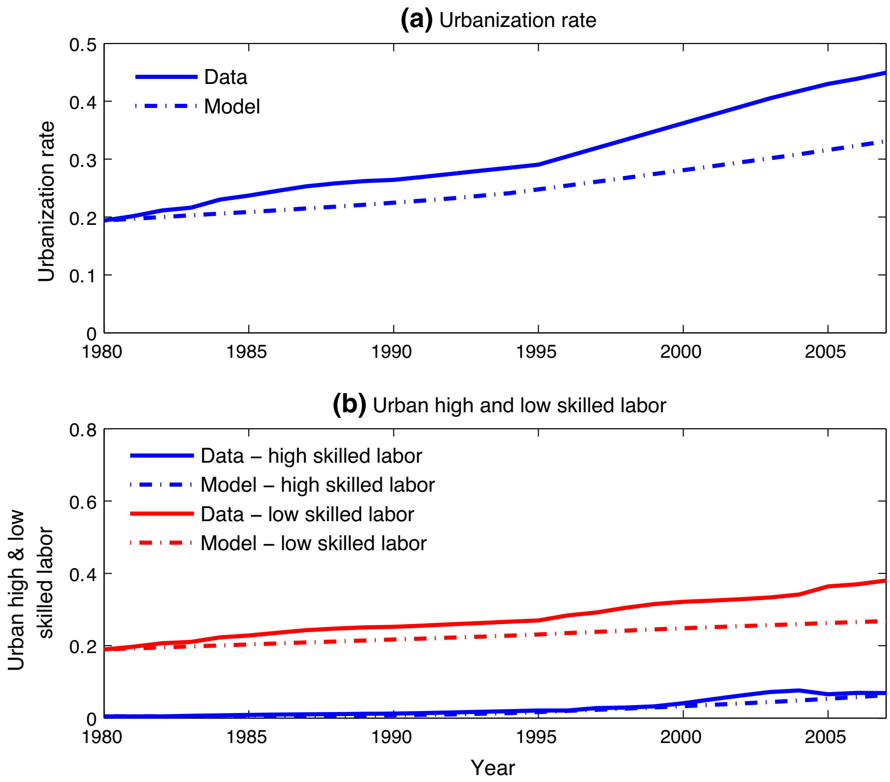

Fig. 7 Benchmark model—urbanization rate and labor share

and (iii) both migration channels are eliminated. In the first scenario, when the WB migration is eliminated, the only "new" source of low-skilled workers coming from countryside is unlucky college graduates. As a consequence, there are much fewer productive low-skilled workers in cities, resulting in a larger high-low skilled labor ratio and a higher urban production (per capita). Furthermore, as the migration volume via the WB migration is large, the urbanization rate in this scenario is much lower than the benchmark case. In the second scenario in which EB migration is eliminated, once again, as the volume of EB migration that is eliminated is not large, the urbanization rate in this case is very close to that in the benchmark model. This shows that the majority of rural–urban migration is WB. With fewer productive high-skilled workers in the cities, urban production (per capita) is now slightly lower than that in the benchmark case.

To identify the magnitude of the contribution of migration types to major macroeconomic variables, Table 7 reports the percentage change relative to the benchmark model for the above three scenarios. Given the large volume of WB migration, the conventional wisdom is that the effects of WB migration on output levels should far outweigh the effects of EB migration. However, our results in Table 7 show that the contribution of EB migration cannot be overlooked: EB migration and WB migration

Table 6 Benchmark model

Period	Total output per capita Y/N	Urban production Y_U/Y	Urban employment $(N_H + N_L)/N$	High-skilled employment share $N_H/(N_H + N_L)$	Skill premium w_H/w_L
Whole: 1981–2007	1.6206	0.7148	0.2516	0.0784	1.4571
Regime 1: 1981–1994	0.8811	0.6585	0.2174	0.0327	1.2575
Regime 2: 1995–2007	2.4169	0.7754	0.2883	0.1277	1.6720

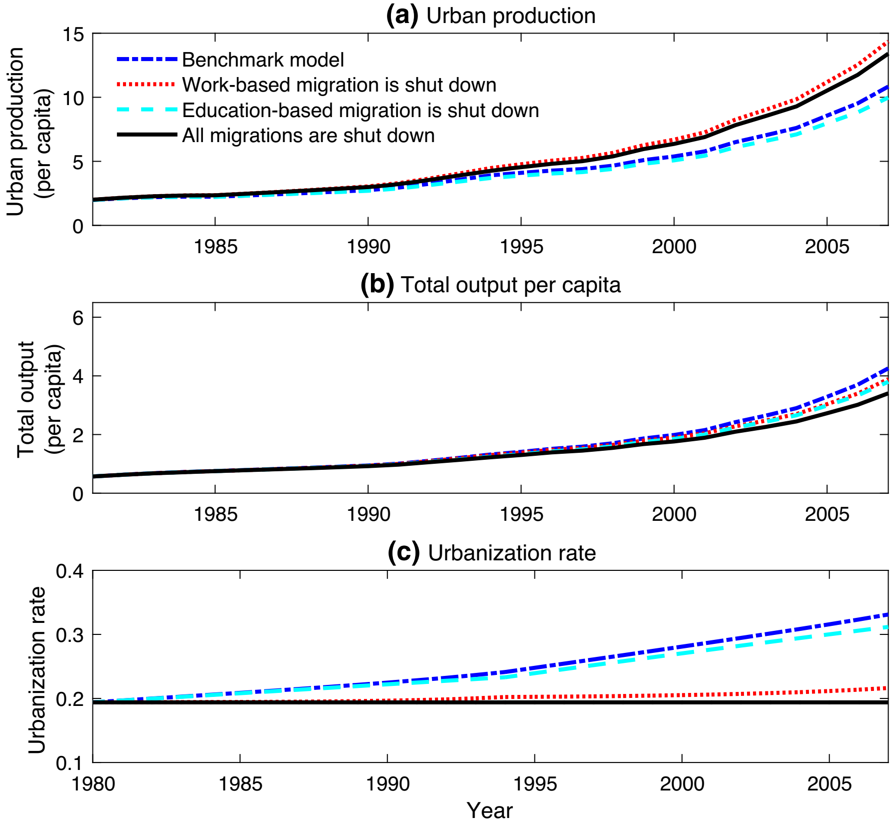

Fig. 8 Decomposition for migration channels

explain 6.3 percent and 4.5 percent of total output per capita in the benchmark model over the entire sample period, respectively. This could be due to the fact that, compared with WB migrants, EB migrants are workers with higher productivity. Therefore, the contribution of EB migrants to per capita output is amplified. We also find that EB migration contributes to roughly one-third of the high-skilled employment share in the benchmark model and thereby lowers the skill premium, while WB migration reduces the high-skilled employment share and boosts the skill premium. Furthermore, the result suggests that EB migration is more important in regime 2 than in regime 1: EB migration in regime 2 explains 8.0 percent of total output per capita in the benchmark model, while it only explains 2.0 percent of total output per capita in regime 1. There are several conflicting forces influencing the effects of EB migration: A higher skill premium, a higher human capital level and a lower EB migration cost in regime 2 attract more migration through the EB migration channel, whereas the higher tuition cost and the termination of the GJA policy depress EB migration. Our quantitative results show that the three positive effects dominate the two negative effects. Therefore, the effects of EB migration on total output per capita and urban employment share in regime 2 are larger than those in regime 1.

As shown in Table 7, our results also show rich interactions between EB and WB migration on the skill premium, high-skilled employment share, total output per capita, urban production and urban employment shares. It is intuitive that the interactive effect is strongest on the high-skilled employment share (accounting for 11 percent of its change over the entire sample period), because WB migration leads to a higher skill premium, attracting more EB migration. For the other variables, several conflicting forces are involved in the resulting interactive effect. First, if WB migration is not allowed, rural residents can still move to urban areas via the EB migration channel. Second, high-skilled workers (mainly from EB migration) and low-skilled workers (mainly from WB migration) are substitutes in production. Third, when there is a larger stock of low-skilled workers in the cities, the skill premium is boosted up. The higher skill premium thus encourages more parents to send their children to cities to attend college. Fourth, there exists upward intergenerational mobility. The last two forces are positive, while the first two are negative. The results show that the skill premium is the dominant effect; thereby, a minor but positive interaction between EB and WB migration is observed.

4.3 Robustness tests

In this subsection, we perform a number of robustness tests, including 5 percent variations in (i) EB or WB migration costs to rural household income ratios ($\tilde{\sigma}_e$ and $\tilde{\sigma}_w$), (ii) annual human capital throughout the years examined (h), (iii) the entire urban TFP series, (iv) the high-skilled labor wedge parameter in urban production (ψ), (v) the child-rearing cost to rural household income ratio ($\tilde{\phi}$), and (vi) the tuition cost to rural household income ratio ($\tilde{b}$). We also recalibrate the model with the probability for a college graduate to join a low-skilled career (γ_L) in the second regime, either to raise from the benchmark value of 5 percent to 10 percent or to fall to zero. The results for robustness tests are reported in Table 8.

Overall, the results suggest that our main findings on the importance of EB migration and its implications for total output per capita, various urban shares and skill premium are all robust. In particular, while the EB migration plays a noticeable role in economic development, more in regime 2 than regime 1, it contributes more significantly to improvements in skill composition of urban employment. Quantitatively, skill composition and skill premium are more sensitive to migration costs to rural household income ratios, which suggests the importance of the *hukou* policy to skill

Table 7 Decomposition of migration channels

Unit: Percentage change

Period	Total output per capita Y/N	Urban production Y_U/Y	Urban employment $(N_H + N_L)/N$	High-skilled employment share $N_H/(N_H + N_L)$	Skill premium w_H/w_L
Education-based migration					
Whole: 1981–2007	6.3%	1.9%	2.8%	30.8%	− 3.1%
Regime 1: 1981–1994	2.0%	1.0%	1.1%	30.6%	− 1.2%
Regime 2: 1995–2007	8.0%	2.8%	4.2%	30.8%	− 4.7%
Work-based migration					
Whole: 1981–2007	4.5%	3.3%	19.9%	− 21.7%	7.2%
Regime 1: 1981–1994	0.8%	1.7%	9.7%	− 11.5%	3.5%
Regime 2: 1995–2007	5.9%	4.8%	28.1%	− 24.5%	10.2%
Interactive migration					
Whole: 1981–2007	0.1%	0.4%	0.2%	11.0%	0.1%
Regime 1: 1981–1994	0.0%	0.0%	0.0%	4.4%	0.1%
Regime 2: 1995–2007	0.2%	0.7%	0.4%	12.8%	0.2%
Non-migration factors					
Whole: 1981–2007	89.1%	94.4%	77.1%	79.9%	95.8%
Regime 1: 1981–1994	97.3%	97.3%	89.2%	76.5%	97.6%
Regime 2: 1995–2007	85.8%	91.8%	67.2%	80.8%	94.3%

Note: Numbers reported in the table are the percentage changes relative to the benchmark model. For example, total output per capita is 1.6206 for the whole period in the benchmark model and 1.5178 in the scenario with the channel of work-based migration only. Therefore, the channel of education-based migration explains 6.3% of total output per capita in the benchmark model

Table 8 Robustness tests for decomposition of migration channels

Unit: Percentage change

Period	Total output per capita Y/N	Urban production Y_U/Y	Urban employment $(N_H+N_L)/N$	High-skilled employment share $N_H/(N_H+N_L)$	Skill premium w_H/w_L	Total output per capita Y/N	Urban production Y_U/Y	Urban employment $(N_H+N_L)/N$	High-skilled employment share $N_H/(N_H+N_L)$	Skill premium w_H/w_L
	$\tilde{\sigma_e}$ increases by 5 % in both regimes					$\tilde{\sigma_e}$ decreases by 5 % in both regimes				
Education-based migration										
Whole: 1981–2007	5.6%	1.7%	2.5%	27.8%	− 2.7%	7.1%	2.2%	3.2%	33.6%	− 3.5%
Regime 1: 1981–1994	1.7%	0.8%	0.9%	26.6%	− 1.0%	2.4%	1.1%	1.3%	34.4%	− 1.5%
Regime 2: 1995–2007	7.2%	2.4%	3.8%	28.2%	− 4.1%	8.9%	3.1%	4.7%	33.3%	− 5.2%
Work-based migration										
Whole: 1981–2007	4.5%	3.4%	20.0%	− 21.3%	7.2%	4.4%	3.3%	19.8%	− 22.0%	7.2%
Regime 1: 1981–1994	0.8%	1.8%	9.8%	− 11.3%	3.5%	0.8%	1.7%	9.7%	− 11.6%	3.5%
Regime 2: 1995–2007	6.0%	4.9%	28.3%	− 24.0%	10.2%	5.8%	4.7%	27.9%	− 24.9%	10.2%
Interactive migration										
Whole: 1981–2007	0.1%	0.3%	0.2%	10.2%	0.1%	0.1%	0.4%	0.3%	11.8%	0.2%
Regime 1: 1981–1994	0.0%	0.0%	0.0%	3.8%	0.1%	0.0%	0.0%	0.0%	4.9%	0.1%
Regime 2: 1995–2007	0.2%	0.6%	0.4%	11.9%	0.2%	0.2%	0.8%	0.5%	13.7%	0.2%
Non-migration factors										
Whole: 1981–2007	89.7%	94.6%	77.4%	83.3%	95.4%	88.4%	94.1%	76.8%	76.7%	96.2%
Regime 1: 1981–1994	97.6%	97.4%	89.4%	80.9%	97.4%	96.9%	97.1%	89.0%	72.3%	97.9%
Regime 2: 1995–2007	86.6%	92.0%	67.6%	83.9%	93.8%	85.0%	91.5%	66.9%	77.9%	94.8%

Table 8 continued

Unit: Percentage change										
Period	Total output per capita Y/N	Urban production Y_U/Y	Urban employment $(N_H+N_L)/N$	High-skilled employment share $N_H/(N_H+N_L)$	Skill premium w_H/w_L	Total output per capita Y/N	Urban production Y_U/Y	Urban employment $(N_H+N_L)/N$	High-skilled employment share $N_H/(N_H+N_L)$	Skill premium w_H/w_L
	$\tilde{\sigma_w}$ increases by 5 % in both regimes					$\tilde{\sigma_w}$ decreases by 5 % in both regimes				
Education-based migration										
Whole: 1981–2007	6.3%	1.9%	2.8%	30.8%	− 3.1%	6.3%	1.9%	2.8%	30.7%	− 3.1%
Regime 1: 1981–1994	2.0%	1.0%	1.1%	30.6%	− 1.3%	2.0%	1.0%	1.1%	30.6%	− 1.2%
Regime 2: 1995–2007	8.0%	2.8%	4.2%	30.8%	− 4.7%	8.0%	2.8%	4.2%	30.8%	− 4.7%
Work-based migration										
Whole: 1981–2007	4.5%	3.3%	19.9%	− 21.6%	7.2%	4.5%	3.3%	19.9%	− 21.7%	7.2%
Regime 1: 1981–1994	0.8%	1.7%	9.7%	− 11.4%	3.5%	0.8%	1.7%	9.7%	− 11.5%	3.5%
Regime 2: 1995–2007	5.9%	4.8%	28.1%	− 24.5%	10.2%	5.9%	4.8%	28.1%	− 24.5%	10.2%
Interactive migration										
Whole: 1981–2007	0.1%	0.4%	0.2%	11.0%	0.1%	0.1%	0.4%	0.2%	11.0%	0.1%
Regime 1: 1981–1994	0.0%	0.0%	0.0%	4.3%	0.1%	0.0%	0.0%	0.0%	4.4%	0.1%
Regime 2: 1995–2007	0.2%	0.7%	0.4%	12.8%	0.2%	0.2%	0.7%	0.4%	12.8%	0.2%
Non-migration factors										
Whole: 1981–2007	89.1%	94.4%	77.1%	79.9%	95.8%	89.1%	94.4%	77.1%	79.9%	95.8%
Regime 1: 1981–1994	97.3%	97.3%	89.2%	76.5%	97.6%	97.3%	97.3%	89.2%	76.5%	97.6%
Regime 2: 1995–2007	85.8%	91.8%	67.2%	80.8%	94.3%	85.8%	91.8%	67.3%	80.9%	94.3%

Table 8 continued

Unit: Percentage change

Period	Total output per capita Y/N	Urban production Y_U/Y	Urban employment $(N_H + N_L)/N$	High-skilled employment share $N_H/(N_H + N_L)$	Skill premium w_H/w_L	Total output per capita Y/N	Urban production Y_U/Y	Urban employment $(N_H + N_L)/N$	High-skilled employment share $N_H/(N_H + N_L)$	Skill premium w_H/w_L
	ϕ increases by 5%					ϕ decreases by 5%				
Education-based migration										
Whole: 1981–2007	6.1%	1.8%	2.7%	30.0%	− 3.0%	6.6%	2.0%	2.9%	31.5%	− 3.2%
Regime 1: 1981–1994	1.9%	0.9%	1.0%	29.7%	− 1.2%	2.1%	1.0%	1.1%	31.4%	− 1.3%
Regime 2: 1995–2007	7.8%	2.7%	4.1%	30.0%	− 4.5%	8.3%	2.9%	4.4%	31.5%	− 4.8%
Work-based migration										
Whole: 1981–2007	4.5%	3.3%	19.9%	− 21.5%	7.2%	4.5%	3.3%	19.8%	− 21.8%	7.2%
Regime 1: 1981–1994	0.8%	1.7%	9.7%	− 11.4%	3.5%	0.8%	1.7%	9.7%	− 11.5%	3.5%
Regime 2: 1995–2007	5.9%	4.8%	28.2%	− 24.3%	10.2%	5.9%	4.8%	28.0%	− 24.6%	10.2%
Interactive migration										
Whole: 1981–2007	0.1%	0.4%	0.2%	10.8%	0.1%	0.1%	0.4%	0.2%	11.2%	0.1%
Regime 1: 1981–1994	0.0%	0.0%	0.0%	4.2%	0.1%	0.0%	0.0%	0.0%	4.5%	0.1%
Regime 2: 1995–2007	0.2%	0.7%	0.4%	12.6%	0.2%	0.2%	0.7%	0.4%	13.1%	0.2%
Non-migration factors										
Whole: 1981–2007	89.3%	94.4%	77.2%	80.8%	95.7%	88.9%	94.3%	77.0%	79.0%	95.9%
Regime 1: 1981–1994	97.3%	97.3%	89.2%	77.5%	97.6%	97.2%	97.2%	89.2%	75.6%	97.7%
Regime 2: 1995–2007	86.1%	91.8%	67.4%	81.7%	94.1%	85.6%	91.7%	67.1%	80.0%	94.4%

Table 8 continued

Unit: Percentage change										
Period	Total output per capita Y/N	Urban production Y_U/Y	Urban employment $(N_H+N_L)/N$	High-skilled employment share $N_H/(N_H+N_L)$	Skill premium w_H/w_L	Total output per capita Y/N	Urban production Y_U/Y	Urban employment $(N_H+N_L)/N$	High-skilled employment share $N_H/(N_H+N_L)$	Skill premium w_H/w_L
	Reduce $\tilde{b}$ in regime 2 by 20%					Increase the urban TFP by 5%				
Education-based migration										
Whole: 1981–2007	6.6%	2.0%	2.9%	31.4%	− 3.2%	7.3%	2.1%	3.2%	34.0%	− 3.6%
Regime 1: 1981–1994	2.0%	1.0%	1.1%	30.6%	− 1.2%	2.5%	1.1%	1.3%	35.1%	− 1.5%
Regime 2: 1995–2007	8.4%	2.9%	4.4%	31.6%	− 4.8%	9.1%	3.0%	4.8%	33.7%	− 5.3%
Work-based migration										
Whole: 1981–2007	4.4%	3.3%	19.8%	− 21.9%	7.2%	4.5%	3.1%	19.7%	− 22.2%	7.2%
Regime 1: 1981–1994	0.8%	1.7%	9.7%	− 11.5%	3.5%	0.8%	1.7%	9.7%	− 11.8%	3.5%
Regime 2: 1995–2007	5.9%	4.7%	28.0%	− 24.8%	10.2%	6.0%	4.5%	27.9%	− 25.1%	10.2%
Interactive migration										
Whole: 1981–2007	0.2%	0.4%	0.2%	11.4%	0.1%	0.2%	0.4%	0.3%	12.0%	0.2%
Regime 1: 1981–1994	0.0%	0.0%	0.0%	4.4%	0.1%	0.0%	0.1%	0.0%	5.2%	0.1%
Regime 2: 1995–2007	0.2%	0.7%	0.4%	13.3%	0.2%	0.3%	0.8%	0.5%	13.9%	0.2%
Non-migration factors										
Whole: 1981–2007	88.8%	94.3%	77.0%	79.2%	95.9%	88.0%	94.3%	76.7%	76.2%	96.2%
Regime 1: 1981–1994	97.3%	97.3%	89.2%	76.5%	97.6%	96.7%	97.1%	89.0%	71.5%	97.9%
Regime 2: 1995–2007	85.5%	91.7%	67.1%	79.9%	94.4%	84.6%	91.7%	66.9%	77.5%	94.9%

Table 8 continued

Unit: Percentage change

Period	Total output per capita Y/N	Urban production Y_U/Y	Urban employment $(N_H+N_L)/N$	High-skilled employment share $N_H/(N_H+N_L)$	Skill premium w_H/w_L	Total output per capita Y/N	Urban production Y_U/Y	Urban employment $(N_H+N_L)/N$	High-skilled employment share $N_H/(N_H+N_L)$	Skill premium w_H/w_L
	Increase the entire series of h by 5%					Decrease the entire series of h by 5%				
Education-based migration										
Whole: 1981–2007	7.3%	2.1%	3.2%	33.8%	−3.6%	5.5%	1.7%	2.4%	27.6%	−2.7%
Regime 1: 1981–1994	2.5%	1.1%	1.3%	34.8%	−1.5%	1.6%	0.8%	0.9%	26.2%	−1.0%
Regime 2: 1995–2007	9.1%	3.0%	4.8%	33.5%	−5.3%	7.0%	2.5%	3.7%	28.0%	−4.1%
Work-based migration										
Whole: 1981–2007	4.4%	3.1%	19.7%	−22.1%	7.2%	4.5%	3.5%	20.0%	−21.2%	7.2%
Regime 1: 1981–1994	0.8%	1.7%	9.7%	−11.8%	3.5%	0.8%	1.8%	9.8%	−11.1%	3.5%
Regime 2: 1995–2007	5.8%	4.5%	27.9%	−25.0%	10.2%	6.0%	5.1%	28.3%	−23.9%	10.1%
Interactive migration										
Whole: 1981–2007	0.2%	0.4%	0.3%	11.9%	0.2%	0.1%	0.3%	0.2%	10.0%	0.1%
Regime 1: 1981–1994	0.0%	0.0%	0.0%	5.1%	0.1%	0.0%	0.0%	0.0%	3.6%	0.1%
Regime 2: 1995–2007	0.3%	0.8%	0.5%	13.8%	0.2%	0.2%	0.6%	0.4%	11.8%	0.2%
Non-migration factors										
Whole: 1981–2007	88.1%	94.3%	76.8%	76.4%	96.2%	89.9%	94.5%	77.4%	83.5%	95.4%
Regime 1: 1981–1994	96.8%	97.2%	89.0%	71.9%	97.9%	97.7%	97.4%	89.4%	81.4%	97.4%
Regime 2: 1995–2007	84.8%	91.7%	66.9%	77.7%	94.8%	86.8%	91.8%	67.6%	84.1%	93.8%

Table 8 continued

Unit: Percentage change

Period	Total output per capita Y/N	Urban production Y_U/Y	Urban employment $(N_H+N_L)/N$	High-skilled employment share $N_H/(N_H+N_L)$	Skill premium w_H/w_L	Total output per capita Y/N	Urban production Y_U/Y	Urban employment $(N_H+N_L)/N$	High-skilled employment share $N_H/(N_H+N_L)$	Skill premium w_H/w_L
	ψ increases by 5%					ψ decreases by 5%				
Education-based migration										
Whole: 1981–2007	6.1%	1.8%	2.8%	30.5%	−3.0%	6.5%	2.0%	2.8%	31.0%	−3.3%
Regime 1: 1981–1994	1.9%	0.9%	1.1%	30.2%	−1.2%	2.1%	1.0%	1.1%	30.9%	−1.3%
Regime 2: 1995–2007	7.8%	2.6%	4.2%	30.6%	−4.5%	8.3%	2.9%	4.3%	31.0%	−4.9%
Work-based migration										
Whole: 1981–2007	4.4%	3.2%	19.9%	−21.7%	7.2%	4.5%	3.4%	19.9%	−21.6%	7.2%
Regime 1: 1981–1994	0.8%	1.7%	9.7%	−11.5%	3.5%	0.8%	1.8%	9.7%	−11.4%	3.5%
Regime 2: 1995–2007	5.8%	4.7%	28.1%	−24.5%	10.2%	6.0%	4.9%	28.1%	−24.5%	10.1%
Interactive migration										
Whole: 1981–2007	0.1%	0.4%	0.2%	11.0%	0.1%	0.1%	0.4%	0.2%	11.0%	0.1%
Regime 1: 1981–1994	0.0%	0.0%	0.0%	4.4%	0.1%	0.0%	0.0%	0.0%	4.4%	0.1%
Regime 2: 1995–2007	0.2%	0.7%	0.4%	12.8%	0.2%	0.2%	0.7%	0.4%	12.9%	0.2%
Non-migration factors										
Whole: 1981–2007	89.3%	94.6%	77.1%	80.2%	95.6%	88.8%	94.1%	77.1%	79.6%	96.0%
Regime 1: 1981–1994	97.3%	97.4%	89.2%	76.9%	97.6%	97.2%	97.1%	89.2%	76.1%	97.7%
Regime 2: 1995–2007	86.1%	92.1%	67.3%	81.1%	94.0%	85.5%	91.4%	67.2%	80.6%	94.5%

Table 8 continued

Unit: Percentage change

Period	Total output per capita Y/N	Urban production Y_U/Y	Urban employment $(N_H+N_L)/N$	High-skilled employment share $N_H/(N_H+N_L)$	Skill premium w_H/w_L	Total output per capita Y/N	Urban production Y_U/Y	Urban employment $(N_H+N_L)/N$	High-skilled employment share $N_H/(N_H+N_L)$	Skill premium w_H/w_L
	Double regime 2 γ_L from 0.05 to 0.1					Set regime 2 γ_L to 0				
Education-based migration										
Whole: 1981–2007	6.2%	1.9%	2.8%	30.1%	− 3.1%	6.5%	1.9%	2.8%	31.4%	− 3.2%
Regime 1: 1981–1994	2.0%	1.0%	1.1%	30.6%	− 1.2%	2.0%	1.0%	1.1%	30.6%	− 1.2%
Regime 2: 1995–2007	7.9%	2.7%	4.2%	30.0%	− 4.5%	8.2%	2.8%	4.2%	31.6%	− 4.8%
Work-based migration										
Whole: 1981–2007	4.5%	3.3%	19.9%	− 21.4%	7.2%	4.4%	3.3%	19.9%	− 21.9%	7.2%
Regime 1: 1981–1994	0.8%	1.7%	9.7%	− 11.5%	3.5%	0.8%	1.7%	9.7%	− 11.5%	3.5%
Regime 2: 1995–2007	6.0%	4.8%	28.1%	− 24.2%	10.1%	5.9%	4.8%	28.1%	− 24.8%	10.2%
Interactive migration										
Whole: 1981–2007	0.1%	0.4%	0.2%	10.7%	0.1%	0.1%	0.4%	0.2%	11.3%	0.1%
Regime 1: 1981–1994	0.0%	0.0%	0.0%	4.4%	0.1%	0.0%	0.0%	0.0%	4.4%	0.1%
Regime 2: 1995–2007	0.2%	0.7%	0.4%	12.5%	0.2%	0.2%	0.7%	0.4%	13.2%	0.2%
Non-migration factors										
Whole: 1981–2007	89.1%	94.4%	77.1%	80.6%	95.7%	89.0%	94.4%	77.1%	79.2%	95.8%
Regime 1: 1981–1994	97.3%	97.3%	89.2%	76.5%	97.6%	97.3%	97.3%	89.2%	76.5%	97.6%
Regime 2: 1995–2007	86.0%	91.8%	67.3%	81.7%	94.2%	85.7%	91.7%	67.2%	80.0%	94.4%

Note: As mentioned in the main text, we recalibrate γ_H, δ_L, a, τ and urban TFP when performing the robustness tests on γ_L. The corresponding parameters values in regime 2 are:
(i) Experiment with $\gamma_L = 0.1$: $\gamma_H = 0.8709$, $\delta_L = 0.9881$, $a = 0.4768$, and the average of τ and urban TFP in regime 2 are 5.139 and 10.6446, respectively.
(ii) Experiment with $\gamma_L = 0$: $\gamma_H = 0.9210$, $\delta_L = 0.9886$, $a = 0.4639$, and the average of τ and urban TFP in regime 2 are 5.1388 and 10.6441, respectively

measures—a channel largely ignored in the literature. Not surprisingly, macroeconomic performance measured by total output per capita is more sensitive to changes in human capital or urban TFP. Interestingly, the robustness check on the high-skilled labor wedge parameter in urban production also shows the important role of skill premium: As an increase in such wedge tends to lower skill premium, it reduces the contribution of EB migration to per capita output. Finally, we find that the recalibration with a sizable change in the probability for a college graduate to join a low-skilled career essentially leads to quantitatively identical results to our benchmark.

5 Factor decomposition and policy experiments

Based on the benchmark calibration, we are now ready to examine important factors that influence the development and urbanization of China that has implemented large-scaled institutional reforms on education, market intervention and migration regulation since the 1990s. What are the effects of these policies? How did these policy reforms and other underlaying factors shape China's subsequent macroeconomic performance in comparison to its development in the earlier decades? Aiming at answering the above questions, we provide a counterfactual based decomposition analysis. We also perform counterfactual policy experiments on education and labor market policies to study how these policy tools can be adopted to enhance the development of an economy. Finally, we introduce human capital externalities into the model to examine the extent to which the free-rider problems affect the EB migration.

5.1 Factor decomposition

We conduct an eleven-factor decomposition, investigating the separate contribution of (i) the abolishment of the GJA policy (lower γ_H), (ii) better WB job opportunities (higher π), (iii) an increase in the EB migration cost (higher σ_e), (iv) an increase in the WB migration cost (higher σ_w), (v) increases in both urban and rural TFP, (vi) an improvement in human capital (higher h), (vii) an increase in child-rearing cost (higher ϕ), (viii) less market price distortion (lower τ), (ix) better intergenerational mobility (lower δ_{LL}), (x) rising admission selectivity (lower a), and (xi) an increase in college tuition (higher b). Each counterfactual experiment is conducted by setting the corresponding parameter in regime 2 back to the level of regime 1, while others remain unchanged. Then, we compute the percentage change in each of the counterfactual outcome from the benchmark model (regime 2 in Table 6).

The results of factor decomposition are provided in Table 9 and are summarized below. First of all, the TFP growth, the improvement in human capital and the better intergenerational mobility contribute the most to the increases in total output per capita, accounting for 52.9 percent, 10.8 percent and 12.3 percent, respectively, whereas the rising admission selectivity greatly damps total output per capita (depressed by 24.8 percent). Second, the better intergenerational mobility, the improvement in human capital and the TFP growth also matter for the increase in urban production share, accounting for 3.2 percent, 3.0 percent and 1.8 percent of the increase, respectively.

Table 9 Factor decomposition

Unit: Percentage change

Factors	Total output per capita Y/N	Urban production Y_U/Y	Urban employment $(N_H + N_L)/N$	High-skilled employment share $N_H/(N_H + N_L)$	Skill premium w_H/w_L
Abolishment of the GJA (lower γ_H)	− 0.9%	− 0.2%	− 0.4%	− 2.9%	0.4%
Better work-based job opportunities (higher π)	1.5%	1.2%	8.2%	− 7.3%	2.9%
An increase in the education-based migration cost (higher σ_e)	− 0.3%	− 0.1%	− 0.1%	− 0.8%	0.1%
An increase in the work-based migration cost (higher σ_w)	0.0%	0.0%	0.0%	0.0%	− 0.0%
Increases in urban and rural TFP	52.9%	1.8%	1.0%	5.5%	− 0.8%
An improvement in human capital (higher h)	10.8%	3.0%	0.3%	2.1%	9.8%
An increase in the child-rearing cost (higher ϕ)	− 1.1%	− 0.3%	− 0.5%	− 3.2%	0.5%
Lower market distortion (lower τ)	1.2%	0.3%	0.6%	3.5%	21.4%
Better intergenerational mobility (lower δ_{LL})	12.3%	3.2%	− 0.0%	49.3%	− 9.9%
Rising admission selectivity (lower a)	− 24.8%	− 4.9%	− 12.4%	− 64.2%	8.5%
An increase in college tuition (higher b)	− 2.0%	− 0.5%	− 1.0%	− 6.1%	0.9%

However, the effect is offset by the rising admission selectivity (− 4.9%). Third, urban employment share rises due to better WB job opportunities, accounting for 8.2 percent of the increase, but is depressed by the rising admission selectivity (− 12.4%). Fourth, intergenerational mobility and TFP growth are both important in increasing the high-skilled employment share (accounting for 49.3 percent and 5.5 percent respectively), whereas the high-skilled employment share is decreased by the rising admission selectivity, the higher college tuition and better WB job opportunities (− 64.2%, − 6.1% and − 7.3%, respectively). Finally, among all the factors, the lower labor-market price distortion is found to be the most important factor that leads to the increase in skill premium. Other factors contributing to the increase in skill premium include the improvement in the quality of human capital and the rising admission selectivity, whereas the improvement in intergenerational mobility drags the skill premium down.

Compared with other factors, we find that the rising college admission selectivity plays a crucial but negative role in China's development during 1994–2007. Admissions are becoming more selective to rural students. This could be due to the fact that high-skilled parents tend to move to cities, resulting in a brain drain from rural to urban areas. Since it is more difficult for rural students to attend top universities, rural parents have lower incentives to send their children for higher education in urban areas (fewer EB migration). This provides a possible explanation to the imbalanced migrations between the high-skilled and the low-unskilled.

To simplify the analysis, we have abstracted from fertility choice by assuming one child per household. This assumption is innocuous because it is consistent with the spirit of China's one-child policy, and the one-child policy had been implemented throughout the time period we examine. However, we are aware of the fact that the one-child policy was not strictly imposed in rural China. To carefully contemplate this issue, the effect of changes in family size on migrations can be regarded as changes in the probability of WB migration, π. When the implementation of the one-child policy is looser, fertility is higher, and family sizes become larger. This implies that the WB migration is more competitive, and the probability of migrating via working becomes lower. Then EB migration will be more attractive. Therefore, considering changes in family size will enhance the role of EB migration in this paper.

5.2 Policy experiments

We consider two groups of scenarios with policy implications. The first group relates to education policies, discussing the scenario of no quantity rationing on rural students in college admission, and the scenarios with subsidies on tuition and EB migration cost. The second group explores regulations on the labor market by studying the scenarios with the GJA policy repealed, the regulations on the WB migration relaxed and the labor-market price distortion mitigated. The results are summarized in Table 10.

Table 10 Policy experiments

Unit: Percentage change

Period	Total output per capita Y/N	Urban production Y_U/Y	Urban employment $(N_H + N_L)/N$	High-skilled employment share $N_H/(N_H + N_L)$	Skill premium w_H/w_L
(a) Experiments on education policies					
No quantity rationing in college admission in both regimes					
Whole: 1981–2007	3.7%	1.2%	1.8%	19.4%	− 1.8%
Regime 1: 1981–1994	1.7%	0.8%	0.9%	24.9%	− 1.0%
Regime 2: 1995–2007	4.6%	1.5%	2.5%	17.9%	− 2.5%
Tuition subsidies					
A 20% tuition subsidy in regime 2					
Regime 2: 1995–2007	0.40%	0.10%	0.20%	− 3.25%	− 0.17%
A 50% tuition subsidy in regime 2					
Regime 2: 1995–2007	1.04%	0.26%	0.52%	− 10.95%	− 0.45%
Subsidies on education migration cost in regime 2					
Education migration cost = 80% of work-based migration cost					
Regime 2: 1995–2007	− 2.96%	− 0.76%	− 1.45%	− 8.98%	1.34%
Education migration cost = 20% of work-based migration cost					
Regime 2: 1995–2007	1.67%	0.41%	0.83%	5.01%	− 0.72%

Table 10 continued

Unit: Percentage change

Period	Total output per capita Y/N	Urban production Y_U/Y	Urban employment $(N_H + N_L)/N$	High-skilled employment share $N_H/(N_H + N_L)$	Skill premium w_H/w_L
(b) Experiments on labor market policies					
No GJA in regime 1					
Whole: 1981–2007	− 1.2%	− 0.4%	− 0.5%	− 7.0%	0.7%
Regime 1: 1981–1994	− 0.7%	− 0.3%	− 0.3%	− 10.0%	0.4%
Regime 2: 1995–2007	− 1.4%	− 0.5%	− 0.7%	− 6.1%	0.9%
Better job opportunities in regime 1: $\pi_1 = \pi_2$					
Whole: 1981–2007	2.8%	2.4%	14.4%	− 6.6%	4.2%
Regime 1: 1981–1994	0.9%	2.2%	12.5%	− 11.1%	4.1%
Regime 2: 1995–2007	3.6%	2.7%	16.0%	− 5.4%	4.2%
A 20% subsidy on work-based migration cost in both regimes					
Whole: 1981–2007	0.0%	0.0%	0.0%	4.7%	0.0%
Regime 1: 1981–1994	0.0%	0.0%	0.0%	0.0%	0.0%
Regime 2: 1995–2007	0.0%	0.0%	0.0%	6.0%	0.0%
A reduction of market distortion in regime 1 to the lower level of regime 2					
Whole: 1981–2007	13.7%	4.5%	7.1%	− 11.3%	46.3%
Regime 1: 1981–1994	9.0%	4.0%	5.1%	− 5.5%	113.3%
Regime 2: 1995–2007	15.5%	4.9%	8.7%	− 12.9%	− 8.0%

5.2.1 Education policies

As pointed out in the recent literature, for example Gou (2006), college admission quotas are not evenly distributed across regions in China: More developed regions are allocated with higher quotas. Similar to the arguments used to show the equivalence between quota (or quantitative restriction) and tariff (or ad valorem tax), our college admission selectivity parameter a can be used to capture quantity rationing despite it only enters the budget constraint to affect the cost of education. Given the presence of rationing in data, our benchmark model is by construction calibrated to a rationed outcome. One may thus inquire what happens if there were no quantity rationing. To do this counterfactual analysis, we take the data of urban and rural admission rates from Gou (2006) to compute the relative admission rates of rural students to their urban counterparts. This measure can be interpreted as the relative admission opportunity for rural students. As the data is not available for the entire periods of 1980–2007, we perform the second-degree polynomial curve fitting to obtain the computed data for the unavailable years.[35] Our results show that the average relative admission rates in regime 1 and regime 2 are 0.5538 and 0.8521, respectively. The increasing relative admission rate indicates that, indeed, rural students were under more strict rationing in college admission compared to their urban counterparts, but the situation has been improved markedly in regime 2. We then use the series of relative admission rate to back out the "unrationed" or "equal admission opportunity" education-based migration flow, and the associated thresholds in talents. Once we obtain the unrationed thresholds in talents, we can recalibrate the college admission selectivity a under the counterfactual scenario without quantity rationing for rural students by the indifference boundary equation (12). We simulate the model based on the "unrationed" college admission selectivity parameter a and the results are reported in Panel (a) of Table 10. We find that, in the unrationed scenario, the college admission selectivity parameter a becomes 1.4734 and 0.5036 in regime 1 and regime 2, which are 28.2 percent and 7.1 percent higher than the values in the benchmark. That is, the quantity rationing was much less severe in regime 2 when the college expansion policy was institutionalized. Not surprisingly, we find that, in an unrationed counterfactual economy, there is more EB migration than that in the benchmark. As a result, total per capita output, urbanization rates and high-skilled composition in urban areas are strengthened, while the skill premium is lower.

In our benchmark economy, the EB migration cost (σ_e) only amounts to 33.14 percent of the WB migration cost (σ_w). The relatively lower EB migration cost implies the existence of education subsidies in data. Therefore, our benchmark model represents a subsidized outcome. To discuss the subsidies on EB migration cost, we consider two variations on subsidies: (i) EB migration cost is 80 percent of the WB migration cost, and (ii) EB migration cost is 20 percent of the WB migration cost. Compared with the benchmark, the first scenario indicates the subsidy on EB migration cost is reduced so EB migration becomes more costly, coming to roughly 2.4 times of the benchmark

[35] Only the data for 1989, 1990, and 1996–2005 are available. It is noted that China has experienced a rapid expansion in higher education since 1998–1999. We therefore break 1980–2007 into two periods when performing the curve fitting for the relative opportunity for rural students, with the first period spanning from 1980 to 1999, and the second period from 2000 to 2007.

EB migration cost. In contrast, the second scenario considers a subsidy expansion, so that EB migration cost equals only 60 percent of the benchmark EB migration cost. As shown in Panel (a) of Table 10, institutionalizing a larger subsidy on EB migration cost than the benchmark economy strengthens the contribution of EB migration to skill composition and enhances total output per capita. However, the skill premium is lower with a bigger EB migration subsidy.

We also consider an alternative education subsidy, investigating the effect of a 20 percent or 50 percent reduction in tuition cost, b. Notably, one may view a decrease in b as to relax the credit constraint that limits rural parents' ability to send their children to urban colleges. The results in Panel (a) of Table 10 indicate that a subsidy to tuition (or relaxation of the credit constraint) tends to raise the contribution of EB migration to total output per capita but weakens its contribution to skill composition. This is because that such a subsidy makes parental decisions on sending children to urban colleges less dependent on children's talent.

5.2.2 Labor-market policies

As the GJA policy had been in force in China from the 1950s to the mid 1990s, we wonder how the economy would have performed if China had not implemented the GJA policy throughout its history. Here we perform a scenario, supposing that there were no GJA policy for the time horizon under study by setting the value of γ_H in regime 1 to that of regime 2. That is, jobs are no longer guaranteed for college graduates in regime 1. There are two opposite effects of this policy. Without guaranteed high-skilled jobs, college education becomes less rewarding, resulting in fewer EB migration. However, the skill premium increases because of the decreasing supply of high-skilled workers, which makes college education more rewarding. Our quantitative result, as shown in Panel (b) of Table 10, suggests that the former effect dominates. Therefore, without the GJA policy throughout the history, urban employment would decrease by 0.5 percent, the share of high-skilled employment would decrease by approximately 7 percent, the skill premium would increase by 0.7 percent and the total output per capita would decline by 1.2 percent. We thus conclude that the impact of no GJA on China's development is relatively small. Notably, the small impact of the GJA policy is due to the ambiguous general-equilibrium effect of γ_L on EB migration discussed in Sect. 3.3 and the conflicting partial- and general-equilibrium effects of γ_H.

The second experiment explores the effect of a more relaxed regulation on the WB migration since 1980. Because of the household registration reforms, the regulations on WB migration have been gradually relaxed. We are curious what China would look like if the government had maintained looser regulations for migrant workers. We thus conduct an experiment by increasing the value of π in regime 1 to that of regime 2. The result in Panel (b) of Table 10 suggests that, with a more relaxed regulation on WB migration, there would be more WB migrants, resulting in a larger share of urban employment and an increase in both urban production share and total output per capita. However, the relaxation leads to a lower share of high-skilled employment; thereby a higher skill premium. Although EB migration in Regime 1 is lower, the conflicting general-equilibrium effect resulting from a higher skill premium reduces the negative impact. This explains why the contribution of EB migration in Regime

1 had not decreased by as much as the whole period. Compared with the GJA policy, the regulation on WB migration has a larger impact on China's urbanization and development.

In addition to the above WB migration lottery, we also investigate two labor policy experiments: (i) a 20 percent subsidy on WB migration cost (σ_w) in both regimes and (ii) a reduction of market price distortion (τ) in regime 1 to the lower level of regime 2. The results in Panel (b) of Table 10 suggest that the effect of a subsidy on WB migration cost is expected to be small as they must work through EB migration choice indirectly. A reduction of market price distortion is found to raise the contribution of EB migration to total output per capita but weaken its contribution to skill composition, similar to those of an education subsidy.

Before closing the discussion on policy experiments, we briefly discuss some interesting but omitted factors and their expected effects on our main findings. First, as Table 7 suggests, the EB migration decision depends negatively on changes in WB migration (via π). Should we allow for two-way interactions (i.e. endogenous EB and WB migration), one would expect that EB migration may have greater contribution by lowering π, and may lead to a larger share of high-skilled employment. Second, it is also plausible that higher EB migration may enhance urban productivity, thus reinforcing the incentives for EB migration. Similarly, should we consider learning by doing and follow Rosen (1976) and Heckman (1976) allowing better-educated to have faster learning on the job, the incentive for EB migration would be even stronger. In either case, our figure about the contribution of EB migration may again be viewed as on the conservative side. Third, another factor affecting migration is the urban benefit that can be regarded as an increase in the expected benefit of migration (see Liao et al. 2020). As a result of household registration regulations, high-skilled workers generally enjoy more urban benefits than lower-skilled workers. Given that the substitution effect dominates, we expect that the overall rural–urban migration would be higher but the increase is biased toward EB migration. On the contrary, land entitlement of rural households may also affect rural–urban migration in an opposite direction: Ngai et al. (2019) regarded the land policy as a barrier to China's industrialization, and Liao et al. (2020) found such a policy can slow down the progress of WB migration. In our model, we can treat land entitlement as an increase in the opportunity cost of migration. Thus, it is expected to provide opposite outcomes to urban benefits. Fourth, a related issue is the housing market performance as examined in Garriga et al. (2020). In our model, the rise in housing prices can be reinterpreted as a rising migration cost to rural households. Because low-skilled workers are expected to be relieved from subsidized policy such as public housing, we can regard housing booms as to raise the relative cost of EB migration. This is yet another case which yields opposite outcomes to changes in urban benefits. Finally, rural–urban migration is often taken to be closely related to the structural transformation of industrialization. As discussed in Garriga et al. (2020), during the process of structural transformation and on-going rural–urban migration, the relative price of agricultural goods rises. This is qualitatively equivalent to an increase in the (self-employed) rural wage in our model. As a result of lower incentives for migration, the overall migration is lower and indeed we can show that EB migration would fall relatively more. Some of those factors could enhance the EB migration but others could favor the WB migration.

5.3 Human capital externality

In a now-classic paper by Lucas (1988), human capital externality is incorporated into an education-based endogenous growth model. An interesting implication is: Although such within-the-generation positive externality enhances production, the presence of the free-rider problem reduces individual incentive to undertake education, thus resulting in under-investment in human capital. In an economy with regulations on population mobility, such as China, how important is the effect of such externality on an economy? To introduce human capital externality into our framework, the urban production function is modified as:

$$Y_U = A\left\{\alpha\left[\left(N_H+\psi\right)h^{1-\xi}H^{\xi}\right]^{\rho}+(1-\alpha)\,N_L^{\rho}\right\}^{\frac{1}{\rho}},\ \alpha\in(0,1)\,,\ \rho<1$$

where H is the aggregate stock of human capital in the economy and $H = N_H h$ in equilibrium; ξ represents the degree of human capital externality. Other model setup remains unchanged.

Empirically, however, there are many issues regarding the identification of pure education-related human capital externality, typically using Mincerian approach. By instrumenting with compulsory education, Acemoglu and Angrist (2000) found that within the Mincerian framework human capital externality is marginal, about 1 percent using the U.S. data. To avoid the problems associated with the Mincerian approach, Ciccone and Peri (2006) proposed a more rigorous method without requiring estimates of individual return to human capital to which many problems are related. They find no evidence of significant human capital externality in American Cities. In line with their findings, we thus choose to conduct policy experiments with the degree of human capital externality at modest levels of 1 percent and 4 percent, respectively. The results are summarized in Table 11. Our results reconfirm Lucas (1988) that the presence of the free-rider problem reduces the incentives for EB migration. Nonetheless, EB and WB migrations both play comparable roles in income advancing while their quantitative consequences for urbanization and wage premium remain valid.

In an extended model by Lucas (2004), a learning technology with external human capital spillovers from the leaders to the followers is introduced, through which self-employed urban workers are connected. New migrants in Lucas (2004) with human capital lower than the leaders would not work but rather invest all their time in accumulating human capital (a corner solution due to linear production technology). Upon catching up, they behave the same as those leaders. Thus, new migrants incur a delay to earn income while new migration after the first wave also incurs a delay. The migration and production delays are consistent with our negative impact on EB migration resulting from the free-rider problem. However, in Lucas (2004), as the leaders' human capital rises over time, migrating to cities to take advantage of the learning externality becomes increasingly attractive, which adds to a positive migration incentive that we do not have. Nonetheless, should such positive incentive effect be considered, the contribution of EB migration would be even greater.

Table 11 Human capital externality: decomposition of migration channels

Unit: Percentage change					
Period	Total output per capita Y/N	Urban production Y_U/Y	Urban employment $(N_H + N_L)/N$	High-skilled employment share $N_H/(N_H + N_L)$	Skill premium w_H/w_L
The scenario with $\xi = 0.01$					
Education-based migration					
Whole: 1981–2007	5.8%	1.8%	2.5%	27.7%	− 2.5%
Regime 1: 1981–1994	1.7%	0.9%	0.9%	26.0%	− 0.8%
Regime 2: 1995–2007	7.3%	2.6%	3.8%	28.2%	− 3.8%
Work-based migration					
Whole: 1981–2007	4.6%	3.5%	20.0%	− 21.2%	7.3%
Regime 1: 1981–1994	0.8%	1.8%	9.8%	− 11.1%	3.6%
Regime 2: 1995–2007	6.1%	5.0%	28.3%	− 23.9%	10.2%
Interactive migration					
Whole: 1981–2007	0.1%	0.4%	0.2%	10.1%	0.1%
Regime 1: 1981–1994	0.0%	0.0%	0.0%	3.6%	0.1%
Regime 2: 1995–2007	0.2%	0.7%	0.4%	11.8%	0.2%
Non-migration factors					
Whole: 1981–2007	89.5%	94.4%	77.4%	83.4%	95.1%
Regime 1: 1981–1994	97.5%	97.3%	89.4%	81.6%	97.2%
Regime 2: 1995–2007	86.4%	91.7%	67.6%	83.9%	93.4%

Table 11 continued

Unit: Percentage change					
Period	Total output per capita Y/N	Urban production Y_U/Y	Urban employment $(N_H+N_L)/N$	High-skilled employment share $N_H/(N_H+N_L)$	Skill premium w_H/w_L
The scenario with $\xi = 0.04$					
Education-based migration					
Whole: 1981–2007	4.0%	1.2%	1.5%	18.4%	− 1.1%
Regime 1: 1981–1994	0.9%	0.5%	0.3%	12.2%	0.0%
Regime 2: 1995–2007	5.2%	1.9%	2.5%	19.9%	− 1.9%
Work-based migration					
Whole: 1981–2007	4.8%	4.1%	20.3%	− 19.6%	7.5%
Regime 1: 1981–1994	0.8%	2.1%	9.8%	− 10.3%	3.6%
Regime 2: 1995–2007	6.5%	5.8%	28.9%	− 22.0%	10.4%
Interactive migration					
Whole: 1981–2007	0.1%	0.3%	0.1%	7.0%	0.1%
Regime 1: 1981–1994	− 0.1%	0.0%	0.0%	1.3%	0.0%
Regime 2: 1995–2007	0.1%	0.5%	0.2%	8.5%	0.2%
Non-migration factors					
Whole: 1981–2007	91.1%	94.5%	78.1%	94.2%	93.5%
Regime 1: 1981–1994	98.4%	97.5%	89.9%	96.8%	96.5%
Regime 2: 1995–2007	88.2%	91.8%	68.5%	93.5%	91.3%

6 Concluding remarks

Economic development is usually associated with a process of structural transformation and urbanization. Rural to urban migration triggers the process. In this paper, we have constructed a dynamic spatial equilibrium model with a focus on a largely unexplored migration channel: EB migration. We have then conducted quantitative analysis, taking China as an example of special interest to examine the causes and consequences of EB and WB rural–urban migration in its development process. We have performed various counterfactual based decomposition analysis and policy experiments.

The main takeaway of our quantitative analysis is that migration played an important role in the development process of China: Rural–urban migration accounted for nearly 11 percent of per capita output changes throughout the 1981–2007 period. Particularly, we find that the contribution of EB migration is even larger than that of WB migration. Because of the considerable impact of EB migration, ignoring the education channel would severely under-estimate the effects of migration, particularly the skill-enhanced process of migration. This strong skill enhancing effect of education is consistent with the celebrated contribution by Heckman (1976) and Rosen (1976).

We would, however, like to acknowledge some major limitations of our study. The first, and most importantly, is to recognize that our quantitative analysis is calibration-based. As such, we have tried to fit limited data moments, using averages, growth rates or some key ratios over the entire or each of the sub-sample periods, but leaving other data variations aside. Thus, while our findings are viewed valid for investigating macroeconomic consequences, they should not be taken to micro-level issues typically addressed in the micro development and labor economics literature. Moreover, to accommodate theoretical analysis, we have to maintained tractability, which limits the generality of the model. A list of various omitted factors have been discussed at the end of Section 5.2, with some enhancing the contribution of EB migration but others dampening it. It is possible that some of these factors might be incorporated in a pure numerical oriented paper to quantify their precise impacts on the role of EB migration. Another limitation is that our analysis is exclusively positive. Thus, normative analysis such as welfare evaluation of various policies is left behind. To conduct welfare analysis is actually not straightforward. In addition to various distortionary factors and intergenerational spillovers, migration itself also causes spatial externality. To properly account for all such complicated welfare effects is beyond the scope of the current study.

Along these lines, it would be interesting to extend our framework to study various migration issues in developing countries. For instance, it has been recognized that rural–urban migration can affect the housing market (for example, Garriga et al. 2020). One may include more formally housing costs as part of the migration decision for this purpose. Another possible extension is to allow urban low-skilled workers to accumulate human capital in cities, as in Lucas (2004). This will further enhance the importance of the EB migration channel. One could also examine different underlying channels of the WB migration, in particular, the early sample stage of the WB migration channel into state-owned enterprises and the later stage into both state-owned enterprises and private sector jobs. Moreover, the investment-oriented channel via the blue-stamp scheme for setting up private businesses as well as investments in properties and factories is worth exploring.[36] Furthermore, one may consider an alternation search-theoretic framework to study information spillovers via job networks in the process of urbanization. We leave these interesting topics with nontrivial extensions for future research.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00199-021-01369-2.

Appendix: Mathematical proofs

Proof of Proposition 1

Denote c_U^j as the consumption of children if they are sent to an urban area and c_R^j as the consumption of children if they are kept in a rural area. From (10) and (11) we have:

$$c_U^j = \gamma_H w_H + \gamma_L w_L + (1 - \gamma_H - \gamma_L) w_R - \mathbf{I}^k (X) (1 - \gamma_H - \gamma_L) (X + \sigma_e) - \phi, \tag{26}$$

$$c_R^j = (1 - \pi) w_R + \pi (w_L - \sigma_w) - \mathbf{I}^k (X) (X + \sigma_e) - \phi. \tag{27}$$

By subtracting (27) from (26) and rearranging terms, under Condition NM, we have:

$$\begin{aligned} c_U^j - c_R^j &= \gamma_H w_H + \gamma_L w_L + (\pi - \gamma_H - \gamma_L) w_R + \mathbf{I}^k (X) \\ &\quad \times (\gamma_H + \gamma_L) (X + \sigma_e) - \pi (w_L - \sigma_w) \\ &= \gamma_H w_H + \gamma_L w_L - \pi w_L + (\pi - \gamma_H - \gamma_L) w_R \\ &\quad + \mathbf{I}^k (X) (\gamma_H + \gamma_L) (X + \sigma_e) + \pi \sigma_w \\ &> (\gamma_H + \gamma_L - \pi) [w_L (n_{\max}) - w_R] + \mathbf{I}^k (X) (\gamma_H + \gamma_L) (X + \sigma_e) + \pi \sigma_w \\ &> 0. \end{aligned}$$

Because $u(\cdot)$ is strictly increasing and strictly concave, we have:

$$u\left(c_U^j\right) > u\left(c_R^j\right).$$

Thus, Assumption 1 and Condition NM together guarantee that $\mathbb{E}_X \left\{ u\left(c_U^j\right) - u\left(c_R^j\right) \right\} > 0$ for all $x^k \in (b, x_{\max}^k]$, where $x_{\max}^k \equiv \chi\left(a z_{\min}^k\right) + b$. □

[36] Due to economic development, several state governments introduced the blue-stamp urban *hukou* in the early 1990s to attract professional workers and investors. The blue-stamp *hukou* required an urban infrastructure construction fee for any newcomer in order to obtain a temporary urban *hukou*. A detailed discussion is provided in the online appendix.

Proof of Proposition 2

For notation convenience, we denote $u^j_{cS} = u_c\left(c^j_S\right)$, $S = U, R$ as the location-S marginal utilities. Recall the arguments of $\Delta^i\left(\mathbf{I}^k, x^j\right)$, we have

$$\Delta^i\left(\mathbf{I}^k, x^j\right) = u\left(w_R - x^j - \sigma_e - \phi\right) - u\left(w_R - \phi\right) + \beta\mathbb{E}_X\left\{u\left(c^j_U\right) - u\left(c^j_R\right)\right\},$$

$$c^j_U = \gamma_H w_H + \gamma_L w_L + (1 - \gamma_H - \gamma_L)\, w_R - \mathbf{I}^k(X)(1 - \gamma_H - \gamma_L)(X + \sigma_e) - \phi,$$

$$c^j_R = (1 - \pi)\, w_R + \pi\,(w_L - \sigma_w) - (1 - \pi)\,\mathbf{I}^k(X)(X + \sigma_e) - \phi.$$

We compute:

$$\frac{\partial \Delta^i\left(\mathbf{I}^k, x^j\right)}{\partial x^j} = -u^i_{cR} < 0$$

$$\frac{\partial \Delta^i\left(\mathbf{I}^k, x^j\right)}{\partial \gamma_H} = \beta\mathbb{E}_X\left\{u^j_{cU}\left[(w_H - w_R) + \mathbf{I}^k(X)(X + \sigma_e)\right]\right\} > 0$$

$$\frac{\partial \Delta^i\left(\mathbf{I}^k, x^j\right)}{\partial \gamma_L} = \beta\mathbb{E}_X\left\{u^j_{cU}\left[(w_L - w_R) + \mathbf{I}^k(X)(X + \sigma_e)\right]\right\} > 0$$

$$\frac{\partial \Delta^i\left(\mathbf{I}^k, x^j\right)}{\partial \pi} = \beta\mathbb{E}_X u^j_{cR}\left[w_R - (w_L - \sigma_w) - \mathbf{I}^k(X)(X + \sigma_e)\right] < 0$$

Since x^j is decreasing in a and z^j, but increasing in b, the results follow.

To show the comparative statics of the cost parameters, it is straightforward to show the effect of σ_w on EB migration:

$$\frac{\partial \Delta^i\left(\mathbf{I}^k, x^j\right)}{\partial \sigma_w} = \pi\beta\mathbb{E}_X\left(u^j_{cR}\right) > 0.$$

For the EB migration cost σ_e, we have

$$\frac{\partial \Delta^i\left(\mathbf{I}^k, x^j\right)}{\partial \sigma_e} = -u^i_{cR} + \beta\mathbb{E}_X\left\{\left[-(1 - \gamma_H - \gamma_L)\, u^j_{cU} + (1 - \pi)\, u^j_{cR}\right]\mathbf{I}^k(X)\right\}.$$

Next, we note that under Condition NM, $c^j_U > c^j_R$ and $u^j_{cU} < u^j_{cR}$. Define $\Lambda \equiv \left[-(1 - \gamma_H - \gamma_L)\, u^j_{cU} + (1 - \pi)\, u^j_{cR}\right] \cdot \mathbf{I}^k(X)$, then we have $\beta\mathbb{E}_X\Lambda < u^i_{cU}$ iff $\partial\Delta^i\left(\mathbf{I}^k, x^j\right)/\partial\sigma_e < 0$, i.e.,

$$\partial\Delta^i\left(\mathbf{I}^k, x^j\right)/\partial\sigma_e < 0 \Leftrightarrow \beta\mathbb{E}_X\left\{\left[-(1 - \gamma_H - \gamma_L)\, u^j_{cU} + (1 - \pi)\, u^j_{cR}\right]\mathbf{I}^k(X)\right\} < u^i_{cR}.$$

Recall that (9) and (11),

$$c_R^i = w_R - \mathbf{I}^j(X)(X + \sigma_e) - \phi,$$
$$c_R^j = (1-\pi) w_R + \pi (w_L - \sigma_w) - (1-\pi) \mathbf{I}^k(X)(X + \sigma_e) - \phi,$$

we then compute:

$$\begin{aligned} c_R^i - c_R^j &= \left(w_R - \mathbf{I}^j(X)(X+\sigma_e) - \phi\right) \\ &\quad - \left[(1-\pi) w_R + \pi (w_L - \sigma_w) - (1-\pi)\mathbf{I}^k(X)(X+\sigma_e) - \phi\right] \\ &= -\pi (w_L - \sigma_w - w_R) - \mathbf{I}^j(X)(X+\sigma_e) + (1-\pi)\mathbf{I}^k(X)(X+\sigma_e) \\ &\leq -\pi \left[(w_L - \sigma_w - w_R) + \mathbf{I}^k(X)(X+\sigma_e)\right] < 0 \\ &\Rightarrow u_{c_R}^i > u_{c_R}^j \end{aligned}$$

owing to the fact that generation j has a higher expected income and lower migration cost compared to generation i due to their possibility of WB migration. Putting these results together, we get:

$$\begin{aligned} \frac{\partial \Delta^i\left(\mathbf{I}^k, x^j\right)}{\partial \sigma_e} &= \beta \mathbb{E}_X \left\{\left[(1-\pi) u_{c_R}^j - (1-\gamma_H-\gamma_L) u_{c_U}^j\right] \mathbf{I}^k(X)\right\} - u_{c_R}^i \\ &\leq \beta \mathbb{E}_X \left\{\left[(1-\pi) u_{c_R}^j - (1-\gamma_H-\gamma_L) u_{c_U}^j - \frac{u_{c_R}^i}{\beta}\right] \mathbf{I}^k(X)\right\} \\ &< \beta \mathbb{E}_X \left\{\left[(1-\pi) u_{c_R}^j - (1-\gamma_H-\gamma_L) u_{c_U}^j - \frac{u_{c_R}^j}{\beta}\right] \mathbf{I}^k(X)\right\} \\ &< 0. \end{aligned}$$

The first weak inequality comes from the fact that $\mathbf{I}^k(X)$ is a binary choice of $\{0, 1\}$, where the second strict inequality combines the following facts that $\beta \in (0, 1)$ and $u_{c_R}^i > u_{c_R}^j$.

□

Proof of Theorem 1

Recall the net gain in education $\Delta^i\left(\mathbf{I}^k, x^j\right)$:

$$\Delta^i\left(\mathbf{I}^k, x^j\right) = \underbrace{u\left(w_R - x^j - \sigma_e - \phi\right) - u(w_R - \phi)}_{\text{direct consumption effect (DCE)}} + \underbrace{\beta \mathbb{E}_X \left\{u\left(c_U^j\right) - u\left(c_R^j\right)\right\}}_{\text{intergenerational effect (IE)}}.$$

Recall the definition that $z_0^j > 0$ such that $x_0^j \equiv \chi\left(az_0^j\right) + b$ and

$$w_R - x_0^j - \sigma_e - \phi = 0.$$

In this case, $c^i = 0$ when $\mathbf{I}^j = 1$ so that it is not a sustainable equilibrium. As a result, we have

$$\lim_{z^j \to \underline{z}^j} \Delta^i\left(\mathbf{I}^k, x_0^j\right) < 0$$

so that it is not worth sending children to urban to get educated given the low talent level and hence high cost. Otherwise, $\Delta^i\left(\mathbf{I}^k, x_0^j\right) > 0$ for all levels of talent.

We next examine $\lim_{z^j \to \infty} \Delta^i\left(\mathbf{I}^k, x^j\right) = \Delta^i\left(\mathbf{I}^k, b\right)$ and note that $\Delta^i\left(\mathbf{I}^k, x^j\right)$ diminishes as z^j increases (x^j decreases). For the DCE, we have

$$\lim_{z^j \to \infty} DCE = u\left(w_R - b - \sigma_e - \phi\right) - u\left(w_R - \phi\right) < 0.$$

For the IE, we first compare the arguments of the utility terms and get

$$\begin{aligned} c_U^j - c_R^j &= \gamma_H\left(w_H - \phi\right) + \gamma_L\left(w_L - \phi\right) + \left(1 - \gamma_H - \gamma_L\right) \\ &\quad \times \left[w_R - \mathbf{I}^k(X)\left(X + \sigma_e\right) - \phi\right] \\ &\quad -\pi\left(w_L - \sigma_w - \phi\right) - (1 - \pi)\left[w_R - \mathbf{I}^k(X)\left(X + \sigma_e\right) - \phi\right] \\ &= \gamma_H w_H + \gamma_L w_L - \pi\left(w_L - \sigma_w\right) - \left(\gamma_H + \gamma_L - \pi\right)\phi \\ &\quad - \left(\gamma_H + \gamma_L - \pi\right)\left(w_R - \phi\right) + \left(\gamma_H + \gamma_L - \pi\right)\mathbf{I}^k(X)\left(X + \sigma_e\right). \end{aligned}$$

So we conclude that the IE effect is at its minimum level (or the consumption difference is the smallest) when $\mathbf{I}^k(X) = 0\ \forall X$:

$$\Delta^i\left(\mathbf{I}^k, x_0^j\right) \geq \Delta^i\left(\mathbf{I}^k = 0, b\right).$$

We assume this to be the case for a conservative analysis:

$$\begin{aligned} &\lim_{z^j \to \infty} \Delta^i\left(\mathbf{I}^k = 0, b\right) \\ &\quad = u\left(w_R - b - \sigma_e - \phi\right) - u\left(w_R - \phi\right) \\ &\qquad + \beta\left[\begin{array}{c} u\left[\gamma_H\left(w_H - \phi\right) + \gamma_L\left(w_L - \phi\right) + \left(1 - \gamma_H - \gamma_L\right)\left(w_R - \phi\right)\right] \\ -u\left[\pi\left(w_L - \sigma_w - \phi\right) + (1 - \pi)\left(w_R - \phi\right)\right] \end{array}\right]. \end{aligned} \tag{28}$$

Combining the arguments of the DCE terms of (28), we have

$$\left(w_R - b - \sigma_e - \phi\right) - \left(w_R - \phi\right) = -\left(b + \sigma_e\right). \tag{29}$$

Combining the arguments of the IE terms of (28), we have

$$\begin{aligned} &\left[\gamma_H\left(w_H-\phi\right)+\gamma_L\left(w_L-\phi\right)+\left(1-\gamma_H-\gamma_L\right)\left(w_R-\phi\right)\right] \\ &\quad-\left[\pi\left(w_L-\sigma_w-\phi\right)+\left(1-\pi\right)\left(w_R-\phi\right)\right] \\ &=\gamma_H w_H+\gamma_L w_L-\pi w_L-\left(\gamma_H+\gamma_L-\pi\right)w_R+\pi\sigma_w. \end{aligned} \tag{30}$$

Putting (29) and (30) together we have

$$\begin{aligned} &-\left(b+\sigma_e\right)+\beta\left[\gamma_H w_H+\gamma_L w_L-\pi w_L-\left(\gamma_H+\gamma_L-\pi\right)w_R+\pi\sigma_w\right] \\ &\quad>-\left(b+\sigma_e\right)+\beta\left[\gamma_H w_R+\gamma_L w_R-\pi w_R-\left(\gamma_H+\gamma_L-\pi\right)w_R+\pi\sigma_w\right] \\ &\quad=-\left(b+\sigma_e\right)+\beta\pi\sigma_w. \end{aligned}$$

So we have

$$\lim_{z^j\to\infty} 1^i\left(\mathbf{I}^k,b\right)\geq\lim_{z^j\to\infty}\Delta^i\left(\mathbf{I}^k=0,b\right)>0\ \ if\ \ \beta\pi\sigma_w>b+\sigma_e.$$

Since $\Delta^i\left(\mathbf{I}^k,x^j\right)$ is decreasing in x^j (increasing in z^j), a sufficient condition for the existence of a nondegenerate dynamic competitive spatial equilibrium is $\beta\pi\sigma_w>b+\sigma_e$ so that $\Delta^i\left(\mathbf{I}^k,x^j\right)=0$ for some $z^j\in\left(z^j_{\min},\infty\right)$. □

Proof of Lemma 1

From (8) and Proposition 2, we can see that all five cases in Lemma 1 result in a rise in migration because

$$\begin{aligned} &\frac{\partial\Delta^i\left(\mathbf{I}^k,x^j\right)}{\partial x^j}<0,\ \frac{\partial\Delta^i\left(\mathbf{I}^k,x^j\right)}{\partial\gamma_H}>0,\ \frac{\partial\Delta^i\left(\mathbf{I}^k,x^j\right)}{\partial\pi}<0, \\ &\frac{d\Delta^i\left(\mathbf{I}^k,x^j\right)}{d\sigma_e}<0,\ \frac{d\Delta^i\left(\mathbf{I}^k,x^j\right)}{d\sigma_w}>0,\ \frac{\partial\Delta^i\left(\mathbf{I}^k,x^j\right)}{\partial\gamma_L}>0. \end{aligned} \tag{31}$$

Under Assumption 1, migration leads to more high-skilled workers than low-skilled ones, and hence the relative supply n rises as long as the non-homothetic parameter ψ is not too large. Specifically, it can be shown that

$$\begin{aligned} \left|\frac{\partial N_H^{t+1}}{\partial Q}\right| &=\gamma_H N_R^t\left|\frac{\partial}{\partial Q}\left[\int\mathbf{I}^j\left(z^j,\mathbf{I}^k\right)dG(z^j)\right]\right|>\left|\frac{\partial N_L^{t+1}}{\partial Q}\right| \\ &=\left(\gamma_L-\pi\right)N_R^t\left|\frac{\partial}{\partial Q}\left[\int\mathbf{I}^j\left(z^j,\mathbf{I}^k\right)dG(z^j)\right]\right|, \end{aligned}$$

for $Q=x^j,\sigma_e,\sigma_w,\gamma_H,\gamma_L$ and π. In addition, there is an additional direct effect of these job acquisition probabilities (γ_H,γ_L,π) on n given migration (i.e., for a given

$\Delta^i\left(\mathbf{I}^k, x^j\right)$). Writing out the comparative statics, we have:

$$\frac{\partial N_H^{t+1}}{\partial \gamma_H} = N_R^t \left\{ \gamma_H \frac{\partial}{\partial \gamma_H} \left[\int \mathbf{I}^j \left(z^j, \mathbf{I}^k\right) dG(z^j) \right] + \underbrace{\int \mathbf{I}^j \left(z^j, \mathbf{I}^k\right) dG(z^j)}_{\text{direct}} \right\} > 0, \tag{32}$$

$$\frac{\partial N_L^{t+1}}{\partial \gamma_H} = N_R^t \left(\gamma_L - \pi\right) \frac{\partial}{\partial \gamma_H} \left[\int \mathbf{I}^j \left(z^j, \mathbf{I}^k\right) dG(z^j) \right] > 0, \tag{33}$$

$$\frac{\partial N_H^{t+1}}{\partial \gamma_L} = N_R^t \gamma_H \frac{\partial}{\partial \gamma_L} \left[\int \mathbf{I}^j \left(z^j, \mathbf{I}^k\right) dG(z^j) \right] > 0, \tag{34}$$

$$\frac{\partial N_L^{t+1}}{\partial \gamma_L} = N_R^t \left\{ \left(\gamma_L - \pi\right) \frac{\partial}{\partial \gamma_L} \left[\int \mathbf{I}^j \left(z^j, \mathbf{I}^k\right) dG(z^j) \right] + \underbrace{\int \mathbf{I}^j \left(z^j, \mathbf{I}^k\right) dG(z^j)}_{\text{direct}} \right\}, \tag{35}$$

$$\frac{\partial N_H^{t+1}}{\partial \pi} = N_R^t \gamma_H \frac{\partial}{\partial \pi} \left[\int \mathbf{I}^j \left(z^j, \mathbf{I}^k\right) dG(z^j) \right] < 0, \tag{36}$$

$$\begin{aligned}\frac{\partial N_L^{t+1}}{\partial \pi} = N_R^t \Bigg\{ &\left(\gamma_L - \pi\right) \frac{\partial}{\partial \pi} \left[\int \mathbf{I}^j \left(z^j, \mathbf{I}^k\right) dG(z^j) \right] \\ &+ \underbrace{\left[1 - \int \mathbf{I}^j \left(z^j, \mathbf{I}^k\right) dG(z^j) \right]}_{\text{direct}} \Bigg\}.\end{aligned} \tag{37}$$

From (32), the direct effect of γ_H on high-skilled labor supply is positive so that n must rise. For an increase in the probability of getting a low-skilled job in urban from WB migration (π) based on (37), the positive direct effect expands the low-skilled labor force in urban areas so that n must fall. Finally, for an increase in the probability of getting a low-skilled job in urban from EB migration (γ_L), (35) shows that the positive direct effect expands the low-skilled labor force in urban areas. As a result, the EB migration effect and the direct job finding effect work in opposite directions so that the net outcomes on n is ambiguous for this case. □

Proof of Lemma 2

We apply (13) and compute $\Gamma(n)$ as follows:

$$\begin{aligned}\Gamma &\equiv u_{c_U}^j \gamma_H \frac{dw_H}{dn} + \left(u_{c_U}^j \gamma_L - u_{c_R}^j \pi\right) \frac{dw_L}{dn} \\ &= -Af''(n) \left[\left(u_{c_U}^j \gamma_L - u_{c_R}^j \pi\right) n - u_{c_U}^j \frac{\gamma_H h}{1+\tau} \right]\end{aligned}$$

$$< -Af''(n)\,u^j_{c_U}\left[(\gamma_L-\pi)\,n-\frac{\gamma_H h}{1+\tau}\right]$$
$$\text{or, sign}\,(\Gamma)=\text{sign}\left[\left(u^j_{c_U}\gamma_L-u^j_{c_R}\pi\right)n-u^j_{c_U}\frac{\gamma_H h}{1+\tau}\right]$$

where the inequality follows from the fact that $u^j_{c_U} < u^j_{c_R}$. As a result, we have:

$$(\gamma_L-\pi)\,n-\frac{\gamma_H h}{1+\tau}<0\Leftrightarrow(\gamma_L-\pi)\,(n-n_c)<0\Rightarrow\Gamma<0. \tag{38}$$

Next, we recall that

$$\frac{w_H}{w_L}=\frac{\tilde{w}_H h}{w_L}=\frac{1}{1+\tau}\frac{hf'(n)}{f(n)-nf'(n)}>1. \tag{39}$$

and let $n_{\max}$ denote the upper bound for n and is determined by $w_H\,(n_{\max})=w_L\,(n_{\max})$, or, $f'\,(n_{\max})=\frac{(1+\tau)\,f(n_{\max})}{[h+(1+\tau)n_{\max}]}$. To a urban worker, we get that c^j_U is maximized, or $u^j_{c_U}$ is minimized, at the highest urban net income, i.e.,

$$\max c^j_U=w_H-\phi=\frac{hAf'(n)}{1+\tau}-\phi.$$

To a rural worker, we get that c^j_R is minimized, or $u^j_{c_R}$ is minimized, at the lowest rural net income, i.e.,

$$\begin{aligned}\min c^j_R&=\min\left[(1-\pi)\,w_R+\pi\,(w_L-\sigma_w)-(1-\pi)\,\mathbf{I}^k\,(X)\,(X+\sigma_e)-\phi\right]\\&=w_R-(1-\pi)\left(x^k+\sigma_e\right)-\phi\\&=w_R-(1-\pi)\left(\chi\left(az^k\right)+b+\sigma_e\right)-\phi\equiv\tilde{w}_R\\&\therefore\tilde{w}_R=\min c^j_R\le c^j_R\Rightarrow u_c\,(\tilde{w}_R)\ge u^j_{c_R}.\end{aligned}$$

Thus, recalling the location-S marginal utility notation that $u^j_{c_S}=u_c\left(c^j_S\right)$ where $S=R,U$, we have:

$$\begin{aligned}&\left(u^j_{c_U}\gamma_L-u^j_{c_R}\pi\right)n-u^j_{c_U}\frac{\gamma_H h}{1+\tau}\\&>\left(u^j_{c_U}\gamma_L-u^j_{c_R}\pi\right)n-u^j_{c_R}\frac{\gamma_H h}{1+\tau}\\&\ge n\gamma_L u_c\left(\frac{hAf'(n)}{1+\tau}-\phi\right)-\left(\pi n+\frac{\gamma_H h}{1+\tau}\right)u^j_{c_R}\\&\ge n\gamma_L u_c\left(\frac{hAf'(n)}{1+\tau}-\phi\right)-\left(\pi n+\frac{\gamma_H h}{1+\tau}\right)u_c\,(\tilde{w}_R)\end{aligned}$$

$$\geq \underbrace{n\gamma_L u_c\left(\frac{hAf'(n)}{1+\tau}-\phi\right)}_{\uparrow \text{ in } n} - \left(\pi n_{\max} + \frac{\gamma_H h}{1+\tau}\right) u_c\left(\tilde{w}_R\right) \equiv \Upsilon(n).$$

The first inequality follows from $u_{cU}^j < u_{cR}^j$. The second weak inequality follows from the fact that $u_{cU}^j \geq u_c\left(\max c_U^j\right) = u_c\left(\frac{hAf'(n)}{1+\tau}-\phi\right)$. The third weak inequality follows from the fact that $u_{cR}^j \leq u_c\left(\min c_R^j\right) = u_c\left(\tilde{w}_R\right)$. Finally, the last weak inequality is straightforward because $n \leq n_{\max}$.

Next, notice that $nu_c\left(\frac{hAf'(n)}{1+\tau}-\phi\right)$ and hence $\Upsilon(n)$ is increasing in n . Also, Let $\bar{n}$ solves $\Upsilon(n) = 0$. Then we obtain:

$$n > \bar{n} \Leftrightarrow \Upsilon(n) > 0 \Rightarrow \Gamma > 0. \tag{40}$$

Inequalities given in (38) and (40) together yield:

$$n < \min\{\bar{n}, n_c\} \Rightarrow \Gamma < 0 \text{ and } n > \max\{\bar{n}, n_c\} \Rightarrow \Gamma > 0,$$

which completes the proof. □

Proof of Proposition 3

From Proposition 2 and Lemma 1, except for γ_L, we know that the partial-equilibrium effects on EB migration work in the same direction as the relative labor supply component of the general-equilibrium effect for a change in Q, i.e.,

$$\text{sign}\left(\frac{\partial\left[\Delta^i\left(\mathbf{I}^k, x^j; Q\right)\right]}{\partial Q}\right) = \text{sign}\left(\frac{dn}{dQ}\right), \quad Q = z^j, a, b, \sigma_e, \sigma_w, \gamma_H, \pi.$$

From (22), under Condition W1 that gives $\Gamma > 0$, the general-equilibrium effect of Q on EB migration reinforces the partial-equilibrium effect. On the contrary, if Condition W2 is imposed so that $\Gamma < 0$, then the overall effect of a change in Q on EB migration is ambiguous because the partial- and the general-equilibrium effects work in opposite direction. □

References

Acemoglu, D., Angrist, J.: How large are human-capital externalities? Evidence from compulsory schooling laws. NBER Macroecon. Annu. **15**, 9–59 (2000)

Albornoz, F., Cabrales, A., Hauk, E.: Immigration and the school system. Econ. Theory **65**, 855–890 (2018)

Autor, D.H., Katz, L.F., Krueger, A.B.: Computing inequality: Have computers changed the labor market? Q. J. Econ. **113**(4), 1169–1213 (1998)

Bénabou, R.: Equity and efficiency in human capital investment: the local connection. Rev. Econ. Stud. **63**(2), 237–264 (1996)

Bernard, A., Bell, M., Cooper, J.: Internal Migration and Education: A Cross-National Comparison. Background Report to the 2019 Global Education Monitoring Report. UNESCO, Paris (2018)
Bernard, A.B., Eaton, J., Jensen, J.B., Kortum, S.: Plants and productivity in international trade. Am. Econ. Rev. **93**(4), 1268–1290 (2003)
Bond, E., Riezman, R., Wang, P.: Urbanization and economic development: a tale of two barriers. Working Paper (2015)
Castro, S.B.S.D., Correia-da-Silva, J., Gaspar, J.M.: Economic geography meets hotelling: the home-sweet-home effect. Econ. Theory (2021). https://doi.org/10.1007/s00199-020-01331-8
Ciccone, A., Peri, G.: Identifying human-capital externalities: theory with applications. Rev. Econ. Stud. **73**(2), 381–412 (2006)
Daruich, D.: The macroeconomic consequences of early childhood development policies. Working Paper (2020)
Eaton, J., Kortum, S.: Technology, geography and trade. Econometrica **70**, 1741–1779 (2002)
Eckwert, B., Zilcha, I.: The role of colleges within the higher education sector. Econ. Theory **69**, 315–336 (2020)
Ellickson, B., Zame, W.R.: A competitive model of economic geography. Econ. Theory **25**, 89–103 (2005)
Feenberg, D.R., Poterba, J.M.: Income inequality and the incomes of very high-income taxpayers: evidence from tax returns? In: Poterba, J. (ed.) Tax Policy and the Economy, 7. NBER, MIT Press, Cambridge (1993)
Futagami, K., Ishiguro, S.: Signal-extracting education in an overlapping generations model. Econ. Theory **24**, 129–146 (2004)
Garriga, C., Hedlund, A., Tang, Y., Wang, P.: Rural-urban migration, structural transformation, and housing markets in China. NBER Working Paper No. 23819 (2020)
Ghironi, F., Melitz, M.: International trade and macroeconomic dynamics with heterogeneous firms. Q. J. Econ. **120**, 865–915 (2005)
Glomm, G.: A model of growth and migration. Can. J. Econ. **25**(4), 901–922 (1992)
Gou, R.: Examining equality in higher education from the perspective of rural-urban access to higher education. Res. Educ. Dev. **5**, 29–31 (2006). ((**in Chinese**))
Harris, J.R., Todaro, M.P.: Migration, unemployment and development: a two-sector analysis. Am. Econ. Rev. **60**(1), 126–142 (1970)
He, Z., Dong, N.: A study of family consumption of higher education in Zhejiang Province based on the construction of a harmonious society: a comparative analysis with other seven provinces and municipalities. China High. Educ. Res. **5**, 23–26 (2007)
Heckman, J.J.: A life-cycle model of earnings, learning, and consumption. J. Polit. Econ. **84**(4, Part 2, Part 2), S11–S44 (1976)
Hsieh, C.-T., Klenow, P.J.: Misallocation and manufacturing TFP in China and India. Q. J. Econ. **124**(4), 1403–1448 (2009)
Laing, D., Park, C., Wang, P.: A modified Harris–Todaro model of rural–urban migration for China. In: Kwan, F., Yu, E. (eds.) Critical Issues in China's Growth and Development, pp. 245–264. Ashgate, London (2005)
Liao, P.-J., Wang, P., Wang, Y.-C., Yip, C.K.: To stay or to migrate? When Becker meets Harris-Todaro. NBER Working paper No. 27767 (2020)
Liao, P.-J., Wang, P., Wang, Y.-C., Yip, C.K.: Educational choice, rural-urban migration and economic development. Working Paper (2021). https://bit.ly/3qi96ZD
Lucas, R.E.: On the mechanics of economic development. J. Monet. Econ. **22**(1), 3–42 (1988)
Lucas, R.E.: Life earnings and rural-urban migration. J. Polit. Econ. **112**(S1), S29–S59 (2004)
Maurer-Fazio, M.: Earnings and education in China's transition to a market economy Survey evidence from 1989 and 1992. China Econ. Rev. **10**(1), 17–40 (1999)
Ngai, L.R., Pissarides, C.A., Wang, J.: China's mobility barriers and employment allocations. J. Eur. Econ. Assoc. **17**(5), 1617–1653 (2019)
Qi, W.N.: The state logic of college graduates' employment system in Mainland China: employment as a political issue. In: National Academy for Educational Research (eds.) New Visions for Higher Education: Comparison & prospects, 63–96. Angle, Taipei. https://books.google.com.tw/books?id=Eh-ADwAAQBAJ (2015) (**in Chinese**)
Rosen, S.: A theory of life earnings. J. Polit. Econ. **84**(4, Part 2), S45–S67 (1976)
Song, Z., Storesletten, K., Zilibotti, F.: Growing like China. Am. Econ. Rev. **101**(1), 196–233 (2011)

Todaro, M.P.: A model of labor migration and urban unemployment in less developed countries. Am. Econ. Rev. **59**, 138–148 (1969)

Uras, B., Wang, P.: Production flexibility, misallocation and total factor productivity. NBER Working Paper No. 23970 (2017)

Wang, P., Wong, T.-N., Yip, C.K.: Mismatch and assimilation. NBER Working Paper No. 24960 (2018)

Yang, D.P.: Access to higher education: widening social class disparities. Tsinghua J. Educ. **27**(1), 19–25 (2006). ((**in Chinese**))

Zhu, C., Zhang, Y.: Cost and benefit of rural children in the Xianyang province of China. Popul. Econ. **5**, 13–22 (1996). ((**in Chinese**))

Authors and Affiliations

Pei-Ju Liao[1] · Ping Wang[2] · Yin-Chi Wang[3] · Chong K. Yip[4]

✉ Yin-Chi Wang
yinwang@mail.ntpu.edu.tw

Pei-Ju Liao
pjliao@ntu.edu.tw

Ping Wang
pingwang@wustl.edu

Chong K. Yip
chongkeeyip@cuhk.edu.hk

1 Department of Economics, National Taiwan University, Taipei, Taiwan

2 Department of Economics, Washington University in St. Louis & Federal Reserve Bank of St. Louis and NBER, St. Louis, Cambridge, USA

3 Department of Economics, National Taipei University, New Taipei City, Taiwan

4 Department of Economics, Chinese University of Hong Kong, Hong Kong SAR, China