Migrants and Firms: Evidence from China *

Clement Imbert   Marlon Seror   Yifan Zhang
Yanos Zylberberg

December 12, 2017

Abstract

This paper estimates the causal impact of rural migrant inflows on urban firms in China between 1998 and 2007. We combine international agricultural commodity price shocks with a gravity model to isolate exogenous variation in rural-to-urban migration. Using a census of above-scale firms covering most of the manufacturing sector, we find that migrant inflows decrease labor costs, and increase employment at destination. As capital does not adjust, the labor supply shift strongly affects the factor mix for the average urban firm. There are wide disparities across firms: employment growth is concentrated on capital-rich, private and exporting firms. Overall, rural-urban migration alleviates labor misallocation across production units of the same sector and fosters manufacturing growth.

JEL codes: D24; J23; J61; O15.

*Imbert: Warwick University, clemimbert@gmail.com; Seror: PSE, marlon.seror@ens.fr; Zhang: CUHK, yifan.zhang@cuhk.edu.hk; Zylberberg: University of Bristol, yanos.zylberberg@bristol.ac.uk. We are grateful to Samuel Bazzi, Loren Brandt, Holger Breinlich, Gharad Bryan, Juan Chen, Giacomo De Giorgi, Maelys De La Rupelle, Taryn Dinkelman, Christian Dustmann, Giovanni Facchini, Greg Fischer, Richard Freeman, Albrecht Glitz, Doug Gollin, Andre Groeger, Flore Gubert, Naijia Guo, Marc Gurgand, Marieke Kleemans, Jessica Leight, Florian Mayneris, David McKenzie, Dilip Mokherjee, Joan Monras, Albert Park, Sandra Poncet, Markus Poschke, Simon Quinn, Mark Rosenzweig, Gabriella Santangelo, Michael Song, Jon Temple, Christine Valente, Thomas Vendryes, Chris Udry and Chris Woodruff for useful discussions and comments. We also thank participants in Bristol, CUHK, Harvard, Manchester, Oxford, Nottingham, St Andrews, St Gallen, Warwick and the Yale Agri-Devo Conference, the China Economic Summer Institute, the CEPR "Labor Markets in Developing Countries" Workshop, NEUDC, the BGSE Summer Forum, the European Economic Association meetings in Geneva for helpful comments. The usual disclaimer applies.

1 Introduction

In the process of structural transformation, i.e., the reallocation of production factors between the traditional and the modern sectors, large numbers of rural workers migrate to urban centers (Lewis, 1954; Kuznets, 1964; Harris and Todaro, 1970). In China, this transformation reached an unprecedented scale and pace. The share of agricultural employment went from 70% to 31% between 1980 and 2014, a shift which spanned more than 100 years in most countries (ADB, 2014; Alvarez-Cuadrado and Poschke, 2011). In 2010, there were 200 million rural migrants in Chinese cities, as many as international migrants worldwide (Chan, 2012b; United Nations, 2015).

At the same time, economic development often implies a transformation of the modern sector itself, with a constant reallocation of resources toward small, young and productive firms. The economic literature has identified the misallocation of factors within sectors and across production units as having non-negligible consequences on aggregate productivity in transition economies (Hsieh and Klenow, 2009). This process has been crucial in sustaining economic growth during the past decades in China, which experienced high output growth with rapid reallocation within the manufacturing sector (Song et al., 2011).

The objective of this paper is to study the effect of labor reallocation from agriculture to manufacturing on the reallocation of factors *within* the manufacturing sector in China between 1998 and 2007. Our analysis combines migration data from three population censuses with an exhaustive panel of medium and large manufacturing establishments and answers the following questions. How do firms respond to large variations in (unskilled) labor supply? Does this response vary across heterogeneous firms? Does the rapid movement of labor from agriculture to manufacturing improve or worsen the allocation of factors within the modern sector?

Providing empirical evidence on the impact of rural-urban migration at destination is challenging because it requires to identify large, systematic and exogenous immigrant flows into cities. Our methodology proceeds in two steps. In the first step, we isolate exogenous variation in agricultural labor returns *at origin* from the interaction of (i) an exogenous time-varying factor (world prices for agricultural commodities) and (ii) time-invariant cropping patterns in each rural prefecture. Interacting price variations and prefecture-specific cropping patterns, we produce a measure of *residual agricultural income*. This measure of return to the traditional sector exhibits large year-to-year variation due to fluctuations in world demand and supply for agricultural products, but also wide cross-sectional differences due to the variety of cropping patterns across prefectures. Fluctuations in the residual agricultural income generate significant variations in outflows from rural areas, as measured in population censuses.[1] An excess value of 10% for the agricultural portfolio—about 1 standard deviation—is associated with a 0.25 p.p. lower outmigration incidence. In the second step, we combine the exogenous changes in outmigration due to international commodity prices with a gravity model, including travel time between origin and destination and population at destination, in order to generate exogenous fluctuations in migrant flows into urban areas.[2] This method yields fluctuations in immigrant inflows which satisfy three important properties: they are orthogonal to factor demand in the urban sector, their generate economically significant changes in migrant inflows, and they exhibit substantial variation across years and destinations.

We next use these migration predictions to instrument actual immigrant inflow and estimate its causal impact on the urban economy. We document four novel empirical findings: (i) rural-urban migration in China strongly shifts labor supply at destination and the wage elasticity with respect to migration is large; (ii) production in the average manufacturing firm becomes more labor intensive; (iii) rural-urban migration benefits private exporting firms, which suffer from hiring constraints; (iv) by alleviating labor market distorsions, rural-urban migration fosters manufacturing growth.

First, we quantify the labor supply shift at destination. We find that the vast majority of migrants are hired by medium and large manufacturing firms: an increase of one percentage point in the ratio of migrants to manufacturing employment at destination is associated with a percentage increase in employment among the "above-scale" firms surveyed by the National Bureau of Statistics. This shift has also implications for labor costs at destination. We find that migrant inflows exert a downward pressure on the average compensation per worker. This effect may be due to a composition effect, if firms were to replace urban workers with rural migrants who are less productive. Using the Urban Household Surveys (2002–2008), we find that "natives" wages decline at the same rate as firms' labor costs, and the implied wage elasticity with respect to migration is about 0.9. Finally, wage employment among urban permanent residents slightly declines, indicating that migrants displace urban "natives".

Second, we look at the effect of migrant inflows on factor use and factor productivity in the average manufacturing firm. The increase in employment induces a large decrease in the capital to labor ratio. Capital does not adjust immediately to

[1] We use retrospective information on migration spells from the 2000 Population Census, the 2005 1% Population Survey (or "Mini-Census") and the 2010 Population Census to reconstruct migration flows.

[2] Our approach is similar in that respect to Boustan et al. (2010).

changes in employment, and there are little signs of such adjustment in the following years. As for returns to factors, we find a drop in the marginal productivity of labor of the same order of magnitude as the decrease in labor wages. This outcome would be inconsistent with migrants inducing additional distortions on labor markets. By contrast, the slow or inexistent capital adjustment underlines possible frictions in access to capital for the average urban firm, and migrants do not seem to relax these constraints (at least in the short-run).

Third, we explore whether these effects are heterogeneous across firms. We show that only private sector firms change their factor use in response to migration, as public sector firms face political pressures not to hire migrants. Among private firms, the response is stronger among firms which have a higher capital to labor ratio as compared to the average firm in their sector, which may be indicative of high hiring constraints (relative to capital constraints). The rise in employment is also driven by exporting firms, which face an infinitely elastic demand for their good, and expand when migrant labor is available.

Fourth, we show that migration flows have non-negligible aggregate consequences at the destination level. The positive labor supply shock shifts the entire firm size distribution to the right. It also reduces the dispersion in marginal returns to labor, which is consistent with the fact that capital-intensive firms absorb part of the migrant inflows. These findings suggest that migration favors manufacturing sector growth by alleviating labor market distortions.

This paper makes significant contributions to several strands of the literature.

The research closely relates to the nascent literature studying the impact of shifts in labor supply on the structure of firms and their relative factor intensities (Peri, 2012; Accetturo et al., 2012; Dustmann and Glitz, 2015; Kerr et al., 2015; Olney, 2013).[3] While many features of our empirical analysis are similar with these papers (especially Dustmann and Glitz (2015)), the context we study—a developing economy with massive disparities in productivity and relative factor intensities— is very different. To the best of our knowledge, our paper is the first microeconomic paper to investigate and provide evidence of the effect of labor supply shocks on firm outcomes in a developing economy.

The empirical investigation sheds light on the importance of granular disparities in productivity and factor allocation across firms in developing economies in general, and in China in particular (Hsieh and Klenow, 2009; Brandt et al., 2013; Hsieh and

[3] Giesing and Laurentsyeva (2015) provide evidence of firm adaptation to a different type of shifts in labor supply, i.e., the emigration of skilled workers in Eastern European countries.

Song, 2015). Few recent contributions have questioned the role of credit market imperfections in generating the wide dispersion in factor use and factor returns across firms (Buera et al., 2011; Midrigan and Xu, 2014; Moll, 2014). As developed theoretically in Song et al. (2011), factor market imperfections may generate large disparities in returns to capital and labor, and structural transformation should imply a constant reallocation of resources across production units. Our contribution is to provide empirical evidence on this reallocation and to show it is influenced by rural to urban migration, a major feature of the process of structural transformation.

A large literature has studied the process of structural transformation, and the aggregate reallocation of resources from the traditional and rural sector to the modern and urban sector.[4] We relate to this literature in three distinct ways.

First, we study China, which experienced a remarkably rapid structural transformation since the 1980s, with a sharp fall in the share of agriculture, a symmetric rise in manufacturing and services and massive migration flows from rural to urban areas. In our empirical exercise, we study migration and manufacturing growth with coherent data sources spanning a significant part of the structural transformation period, from 1995 to 2010. Much attention has been given to the patterns of Chinese growth (Song et al., 2011; Hsieh and Song, 2015). However, while the role of rural-to-urban migration in fueling growth finds a large echo in the policy debate in China,[5] the economic literature has so far shown less interest in the topic.

Second, the finding that migration lowers wages and boosts urban employment relate to "labor push" models, which generally imply that, by releasing labor, agricultural productivity gains may trigger industrialization (Gollin et al., 2002; Alvarez-Cuadrado and Poschke, 2011; Bustos et al., 2015). By contrast, we rely on *negative shocks* to agricultural productivity at origin in order to trigger rural-urban migration. We show that following negative (short-term) fluctuations in crop prices, the gap between rural and urban returns to labor widens which increases migration from rural areas. This implies that, in our context, an improvement of conditions at origin increases the opportunity cost of migrating, rather than fostering migration by relaxing liquidity constraints (Angelucci, 2015; Bazzi, 2016).

Third, we relate to "labor pull" models, in which structural change—and thus migration—is prompted by an increased productivity in manufacturing. In China, there are non-negligible labor market frictions across space, possibly explaining the large productivity gaps between rural and urban areas. Our period of interest coincides with large changes in these productivity gaps, due to a take-off of the urban

[4]See Herrendorf et al. (2013) for a review.

[5]See Meng and Zhang (2010) for a survey.

sector in presence of mobility restrictions.[6] Consistent with this labor pull interpretation, Facchini et al. (2015) and Cheng and Potlogea (2017) show that trade shocks increase demand for labor in the manufacturing sector and stimulate internal migration in China. We take the opposite approach to these papers and study how changes in migration can affect manufacturing growth.

The paper also contributes to the large literature on the effects of immigration on labor markets (Borjas, 2003), and more specifically to studies of internal migration.[7] In China, the evidence is mixed: while De Sousa and Poncet (2011) find that migration has alleviated upward pressures on Chinese wages between 1995 and 2007, Meng and Zhang (2010) find a modestly positive or zero effect of rural migrants on native urban workers' labor market outcomes, and Combes et al. (2015) find a strong positive impact on local wages. In a more structural approach, Ge and Yang (2014) use wage decomposition methods and a simple calibration to show that migration depressed unskilled wages in urban areas by at least 20% throughout the 1990s and 2000s. Our findings indicate a wage elasticity to unskilled labor comparable to those found in the literature on developing economies.

The remainder of the paper is organized as follows. In Section 2, we show how we combine shocks to agricultural productivity at origin and a gravity model to isolate exogenous variation in migration to urban areas. In Section 3, we describe how we estimate the causal effect of migrant flows on firm and workers. We present our main results in section 4. Section 5 briefly concludes.

2 Predicting migration to urban areas

This section presents the data sources and the construction of shocks to agricultural livelihoods. We describe how we generate exogenous rural-to-urban migrant flows based on our variations in labor returns at origin. Our strategy closely follows Boustan et al. (2010).[8]

[6] See Bryan and Morten (2015) for an analysis of the implications of labor market frictions for the spatial distribution of activity in the context of developing economies, and Au and Henderson (2006, 2007); Vendryes (2011); Bosker et al. (2012); Tombe and Zhu (2015) for studies of the specific Chinese context.

[7] Boustan et al. (2010), Feng et al. (2015a), El Badaoui et al. (2014), Imbert and Papp (2016), and Kleemans and Magruder (2017), among others, study the labor market effects of migration in the United States during the Great Depression and more recently, Thailand, India, and Indonesia respectively.

[8] A similar approach is adopted by El Badaoui et al. (2014), Feng et al. (2015a) and Kleemans and Magruder (2017).

2.1 Migration flows

In order to measure migration flows, we use the 2000 and 2010 Censuses, and the 2005 1% Population Survey, also called "2005 Mini-Census." The latter is a representative survey with a similar coverage to that of a census. Rural migrants and all types of non-locally registered individuals are interviewed. The censuses contain information on occupation, industry, income, ethnicity, education level and housing characteristics, and, crucially, migration history.

Unlike most studies, which compute migration flows as a difference between population stocks, we directly observe migration spells.[9] First, we observe the household registration type (agricultural or non-agricultural) and places of registration and residence, which are available down to the prefecture level. This information allows us to construct the migrant stock at destination from any possible origins. Second, migrants are asked the main reason for leaving their place of registration and when they did so.[10] The information on places of origin and residence can be combined with retrospective data on the year that the respondent left her place of registration in order to create a matrix of yearly migration flows between all Chinese prefectures between 1996 and 2010.

While all individuals are interviewed in Population Censuses, not all their migration spells are observed. Using primary sources, we can only reconstitute single migration spells from the registered *hukou* location. When a migrant transits through another city before reaching destination, the *step migration* would not be observed at the time of interview and we would incorrectly infer either the initial departure date or the final arrival date. When a migrant returns to her *hukou* location before the interview, we would miss the entire migration spell as if she had never left (*return migration*). A raw measure of migration flows based on retrospective data (and primary sources) may suffer from measurement error due to both step migration and return migration. We discuss these possible biases and describe our adjustments in Appendix A.[11]

Census data allows us to recover yearly migration flows from each prefecture of origin to each prefecture of destination. The main explanatory variable in our analysis are immigrant inflows into each prefecture for every year. To identify its causal impact on urban outcomes, however, we need to isolate exogenous variation

[9]Because of their quality and degree of detail, the census data collected by the National Bureau of Statistics are widely used in the literature (Combes et al., 2015; Facchini et al., 2015; Meng and Zhang, 2010; Tombe and Zhu, 2015, among others), although few authors have been able to use all three censuses.

[10]The year of emigration is censored above five years prior to the interview.

[11]Return migration is substantial, while step migration is negligible. The results presented in Section 4 are corrected for return migration but remain robust to using non-adjusted flows.

in these migration inflows. For this, we combine exogenous variation in migration outflows with fixed migration patterns between origins and destinations. We describe these steps in the next two sections.

2.2 Shocks to agricultural labor returns

As exogenous source of variation in outmigration, we use shocks to labor returns in agriculture, which come from the interaction of origin-specific cropping patterns and time-varying exogenous price fluctuations.

Potential Agricultural Output We construct the potential output for each crop in each prefecture, by combining a measure of harvested area and a measure of potential yield, both provided by the Food and Agriculture Organization (FAO) and the International Institute for Applied Systems Analysis (IIASA).

First, we extract from the 2000 World Census of Agriculture the geo-coded map of harvested area for each crop (in a 30 arc-second resolution, approximately 10km). We then overlay this map with a map of prefectures, and we construct total harvested area h_{co} for a given crop c and a given prefecture o.[12]

Second, we use a measure of potential yield per hectare as computed in the Global Agro-Ecological Zones (GAEZ) Agricultural Suitability and Potential Yields dataset. The measure is model-based and uses information on crop requirements (e.g., the length of yield formation period and the stage-specific crop water requirements), soil characteristics (i.e., the ability of the soil to retain and supply nutrients) in order to generate a potential yield for a given crop and a given soil under 5 scenarios: rain-fed (high/intermediate/low water input), and irrigated crop (high/intermediate water input). For each crop c and prefecture o, we use the intermediate scenarios and information on the share of rain-fed or irrigated harvest in 2000 to construct potential yield yi_{co}.

The interaction between harvested area and potential yield $h_{co}yi_{co}$ is our measure of potential agricultural output for each crop in each prefecture. By construction, $h_{co}yi_{co}$ is time-invariant and supposed to capture long-term cropping patterns at origin.[13] Figure 2 displays the cross-sectional variation in potential output $h_{co}yi_{co}$ for rice and cotton, and illustrates the large geographic disparity in agricultural portfolios. We provide a more detailed description of the existing variation across prefectures and regions in Appendix B.

[12]We collapse our analysis at the prefecture level to match migration data but agricultural shocks can be constructed at a 30 arc-second resolution all over the country.

[13]Note, however, that the measure is computed in 2000 and may be partly affected by fluctuations in prices. This measurement error may potentially bias our first-stage downward.

Price fluctuations As a measure of exogenous changes in international demand for crops, we use the World Bank Commodities Price Data ("The Pink Sheet").[14]. We consider prices in constant 2010 USD and per kg between 1980 and 2009 for the following commodities: banana, cassava, coffee, cotton, an index of fodder crops, groundnut, maize, millet, potato, pulses, rapeseed, rice, sorghum, soybean, sugar beet, sugar cane, sunflower, tea and wheat.[15] These crops account for the lion's share of China's agricultural production over the period of interest: they represent 90% of total agricultural output in 1998 and 80% in 2007.[16] We also collect producer prices, exports and production as reported by the FAO between 1991 and 2013 for China (and other countries) to check that international price variations translate into producer price variations.

In order to identify shocks in international prices, we use a percentage deviation from long-term trend hp_{ct} by applying a Hodrick-Prescott filter on the logarithm of nominal prices. Fluctuations in hp_{ct} capture short- and medium-run fluctuations in international crop prices. We provide in the Appendix B descriptive statistics about the magnitude of fluctuations across crops. On the whole, fluctuations in prices behave as an Auto-Regressive process. The amplitude of innovation shocks is non-negligible: the market value of rice production decreased by 40% between 1998 and 2001 and increased by 70% between 2007 and 2008. In both instances, the rice-producing regions of China experienced large fluctuations in the value of agricultural production.

Finally, we combine the fluctuations in world prices with cropping patterns to construct the excess value of crop production for each year in each prefecture o. The *value gap* for the agricultural portfolio is a weighted average of the crop-specific deviations from long-term trend, $\{hp_{ct}\}_c$, weighted by the expected share of agricultural revenue for crop c in prefecture o. These shares are $\{h_{co}yi_{co}\bar{p}_c\}_c$ where $h_{co}yi_{co}$ is potential output in 1990 described above and $\bar{p}_c$ is a snapshot of international crop prices in 1980.

$$p_{ot} = \left(\sum_c h_{co}yi_{co}\bar{p}_c hp_{ct} \right) / \left(\sum_c h_{co}yi_{co}\bar{p}_c \right) \tag{1}$$

The price shocks p_{ot} exhibit some time-varying volatility coming from World demand

[14] The data are freely available online at http://data.worldbank.org/data-catalog/commodity-price-data

[15] We exclude from our analysis one crop, tobacco, for which (i) China has a dominant position and directly influences international prices and (ii) the China National Tobacco Corporation, a state-owned enterprise, has a monopoly.

[16] http://data.stats.gov.cn/english/easyquery.htm?cn=C01

and supply, but there are also large cross-sectional differences. A prefecture is only exposed to variations in the prices of crops that it produces. The wide variety of harvested crops across China guarantees a large cross-sectional variance in prices p_{ot} that will be exploited in our empirical strategy. Appendix Figure A7 displays the spatial dispersion in price shocks p_{ot} in 2001, just after farmers experienced a crisis across China due to a sharp decrease in the price of rice, and in 2004, after recovery.

As the fluctuations in p_{ot} entirely come from fluctuations in World commodity prices, the identification assumption is that demand and supply fluctuations in commodities are orthogonal to Chinese urban labor demand, i.e., prices are driven by supply shocks in other exporting countries, demand fluctuations in importing countries or the world agricultural market integration. Another important requirement is that there is some pass-through from international prices to domestic prices faced by rural farmers. We show in Appendix B the extent to which fluctuations in international prices are transmitted to the average Chinese farmer.

2.3 Predicting rural-urban migration flows

Let M_{odt} denote migrant flows between origin o (rural areas of a prefecture o) and destination d (a "city," i.e., urban areas in a prefecture d) in a given year $t = 1995, \ldots, 2010$, which we construct using retrospective questions from the 2000 and 2010 Censuses and the 2005 Mini-Census.[17] We construct the outmigration rate in year t, m_{ot}, by dividing the sum of migrants who left o (for labor reasons) in year t by the number of adults who still reside in o, which we denote with R_o. Formally, we have:

$$m_{ot} = \frac{\sum_d M_{odt}}{R_o}.$$

We also construct the probability that a migrant from an origin o goes to destination d at time t, which we denote with $p_{odt} = \frac{M_{odt}}{\sum_d M_{odt}}$.

For the sake of exposition, we describe our strategy for a given shock s_{ot} to the rural origin o in year t, which will be a price shock in the baseline and a rainfall shock in some robustness checks.

In order to estimate the causal effect of migrant inflows on urban destinations, we need variations in migration that are unrelated to potential destination outcomes. Our empirical strategy follows Boustan et al. (2010), and interacts two sources of exogenous variation. First, we use origin variations in returns to agriculture as exogenous determinants of migrant outflows in each rural prefecture. Second, we

[17] There is some debate on how well urbanization is captured in Chinese data—see Chan (2007). However, our population of interest, i.e., rural migrant workers, overwhelmingly settles in urban areas.

use a time-invariant gravity model that includes travel time between prefectures and urban population in 1990 to allocate rural migrants to urban destinations. This two-stage method gives a prediction of migrant inflows to urban areas that is exogenous with respect to fluctuations in urban labor demand.

Exogenous variations in migrant outflows We first regress rural migrant outflows on shocks to agricultural income. Formally, we estimate the following equation:

$$m_{ot} = \beta_0 + \beta_s \bar{s}_{ot} + \delta_t + \nu_o + \varepsilon_{o,t}, \tag{2}$$

where o indexes the origin, and t indexes time $t = 1995, \ldots, 2005$. $m_{o,t}$ and s_{ot} denote is the outmigration rate and the agricultural residual income at origin o, respectively. ν_o denotes origin fixed effects and captures any time-invariant characteristics of origins, e.g., barriers to mobility. We use 1990 population at origin as a weight to generate consistent outmigration predictions in the number of migrants.

As a measure of shock s_{ot}, we use the average residual agricultural income in $t-1$ and $t-2$. A migration spell recorded at date $t = 2005$, for instance, corresponds to a migrant worker who moved between October 2004 and October 2005. Hence, given the timing of the growing cycle for most crops in our sample, migration spells in period t are most likely to be impacted by variations in $t-1$ and before—especially if there are lags in the decision to migrate.[18]

Equation 2 yields the predicted migration rate $\widetilde{m_{ot}}$ from origin o in year t:

$$\widetilde{m_{ot}} = \widetilde{\beta_0} + \widetilde{\beta_1} s_{ot} + \widetilde{\nu_o} + \overline{\overline{\widetilde{\delta_t}}}$$

where $\overline{\overline{\widetilde{\delta_t}}}$ is the average of the time effect.[19] We then multiply the migration rate by rural population at origin R_o to compute predicted migration flows from o:

$$\widetilde{M_{ot}} = \widetilde{m_{ot}} \times R_o$$

We present the estimation of equation (2) in Panel A of Table 1. Between 2001 and 2005 out-migration is negatively correlated with price fluctuations. A 10% lower return to agriculture, as measured by the value gap, is associated with a 1.04 p.p. higher migration incidence. Equivalently, an additional standard deviation

[18]Incorporating contemporary price/rainfall shocks in the analysis does not change the results. We also estimate the same specification using forward shocks, i.e., the average residual agricultural income in $t+1$ and $t+2$, to show that shocks are not anticipated.

[19]We remove time variation from our predictions, in order to avoid correlation between our migrant flows and destination trends in outcomes.

in the value gap decreases migration incidence by 0.19 standard deviations. The corresponding plot in Figure 1 shows that the relationship is globally linear. In Appendix B, we test whether shocks are anticipated and find that forward variations in prices do not predict migration outflows. The estimated coefficients on the lags and forwards of our constructed shocks in the joint regression are similar to those in the separate specifications.

Fluctuations in returns to agriculture may have two opposite effects on migration Bazzi (2016). On the one hand, if rural workers compare the return to labor in rural areas versus the return to labor in urban areas, then a low return to agriculture should push them toward urban centers (*substitution* or *opportunity cost* effect). On the other hand, low returns to agriculture may affect household wealth and the ability to finance migration to urban centers (*liquidity* or *wealth* effect). In the Chinese context, where (i) migrants migrate without their family, and (ii) low-skilled jobs in cities are easy to find, the fixed cost of migration may be relatively low.[20] Since wealth accummulation happens over time, it may also be less affected by short-term fluctuations in agricultural prices. The negative relationship between the value of the agricultural portfolio and migration suggests that migration decisions are driven by the opportunity cost of migrating.[21]

Exogenous variations in origin-destination migration flows We next estimate the following equation:

$$p_{od} = f(dist_{od}) + \gamma Pop_d + \mu_o + \varepsilon_{od}, \tag{3}$$

where p_{od} is the share of migrants from prefecture o who went to prefecture d, $dist_{od}$ is the distance between o and d, f is a parametric function of distance and Pop_d is the total urban population of prefecture d in 1990. Equation 3 yields $\widehat{p_{od}}$, the predicted probability for migrants from prefecture o to go to prefecture d based on distance, a fixed and exogenous characteristic of the pair (od), and the attractiveness of d captured by its lagged population. The specifications are weighted by Pop_d.

We report the estimation of equation 3 in Panel B of Table 1. We choose the inverse of distance for function f.[22] As apparent in this table, distance and population

[20]In the period we study, migration restrictions still exist through the *hukou* or registration system, they tend to make long-term settlement of rural migrants into urban areas difficult rather than impede rural to urban migration itself. See Appendix A for more details.

[21]In Appendix B, we show that negative rainfall shocks have independent positive effects on migration outflows, which is also consistent with the substitution effect.

[22]The inverse function gives a better fit than a linear or quadratic specification (see Appendix Figure A3).

at destination are very strong predictors of the destination choice.

Predicted migration flows Finally, we combine predicted migrant outflows (Equation 2) and the probabilities to transit from each origin to each destination (Equation 3), and predict migrant inflows to urban destinations. Formally, we compute:

$$\widetilde{M_{dt}} = \sum_{o \neq d} \widetilde{M_{ot}} \times \widehat{p_{od}}, \tag{4}$$

where $\widetilde{M_{dt}}$ are migrant inflows to destination d in year t, $\widetilde{M_{ot}}$ is the predicted migrant outflow from origin o in year t and p_{od} is the predicted probability that a migrant from o goes to d. In order to avoid that migrant inflows are correlated with destination outcomes, we exclude from $\widetilde{M_{dt}}$ inflows originating from rural areas of prefecture d.

This two-stage process yields synthetic migrant inflows to prefectures of destination that are exogenous with respect to destination outcomes. We provide some intuition about the nature of these exogenous variations in the Appendix Figure A8 (measure $\widetilde{M_{dt}}$ as predicted by price variations). We report these measures cleaned for cross-sectional time-invariant factors in 2001 (left panels) and 2004 (right panels). As shown in the Appendix Figure A8, there is some spatial auto-correlation in these measures arising from the spatial auto-correlation of crop composition across prefectures and the transformation of outflows into inflows involving distance between prefectures. There is some auto-correlation across periods as international prices exhibit persistence in their fluctuations. Fortunately, there are also large cross-sectional and time-varying fluctuations that we can use for our analysis (see the Appendix B for a variance decomposition of the shocks).

In order to test whether our migration predictions are accurate, we regress the actual migrant inflows on the predicted immigrant inflows. Panel C of Table 1 reports the correlation between actual and predicted migration rates. The relationship is strong, positive and significant throughout the sample period. The coefficient in both specifications is close to one. This suggests, as expected and by construction, that our instrument successfully predicts variation in migrant inflows between years for a given prefecture and across prefectures for a given year, even if they do not explain most of the total variation in migration rates. This baseline relationship between actual and exogenous variations in immigration rates will serve as a first stage in our analysis to estimate the impact of migration on the urban economy.

3 Empirical strategy

This section first presents our data sources, then explain how we construct productivity measures, and finally presents our estimation strategy.

3.1 Data

The empirical analysis mostly relies on establishment-level data spanning 1998-2007 from the National Bureau of Statistics (NBS).[23] The NBS implements every year a census of all state-owned manufacturing enterprises and all non-state manufacturing firms with sales exceeding 5 million RMB or about $600,000. While small firms are not included in the census, the firms we consider account for 90% of total gross output in the manufacturing sector. The firms can be matched across years, so that we can use either the total sample of firms, whose size ranges from 150,000 and 300,000 per year, or the balanced sample of 45,000 firms. The NBS census contain information on each firm's location, industry, ownership type, exporting activity, number of employees and a wide range of accounting variables (e.g., output, input, value added, wage bill, fixed assets, financial assets, etc.). We use these data to construct firm level capital, employment and labor costs, which is defined as total wage bill divided by employment. We also construct measures of factor productivity, which we describe below.

There are three main challenges with using the NBS census. First, matching firms over time in the NBS is difficult because of frequent changes in firm identifiers. In order to match "identifier-switchers," we extend the fuzzy algorithm (using name, address or phone number) developed by Brandt et al. (2014) to cover the period 1992–2009. Second, although we will use the terms "firm" and "enterprise" interchangeably in the remainder of the paper, the NBS data cover "legal units" (*faren danwei*). Subsequently, different subsidiaries of the same enterprise may be surveyed separately if they are separate legal entities and if they are financially indedependant. Third, the 5 million RMB threshold that defines whether a non-publically owned firm belongs to the NBS census was not perfectly implemented. In effect, few firms may have entered the database few years after having reached the sales cutoff. Conversely, some private firms continue to participate in the survey even if their annual sales fall short of the threshold. However, the share of firms below the threshold is negligible, as shown in Figure 3, and dropping them does not affect the results.

[23]The following description partly borrows from a detailed discussion in Brandt et al. (2014), and a more detailed discussion is provided in Appendix C.

3.2 Accounting framework

In this section, we develop a simple model of firm production to guide our empirical analysis. As in Hsieh and Klenow (2009), the economy is divided into sectors in which there is monopolistic competition between heterogeneous firms. A final good is produced from the combination of sectoral outputs, and sectoral output itself is a CES aggregate of firm-specific differentiated good (where σ denotes the elasticity of substitution between differentiated final goods).

Each firm i in sector s is producing according to a CES production function:

$$Y_{is} = A_{is}\left[a_s K_{is}^{\rho} + (1-a_s)L_{is}^{\rho}\right]^{\frac{1}{\rho}}, \tag{5}$$

where a_s governing the sectoral capital share, is assumed to be constant within each sector and ρ, governing the elasticity of substitution between factors, is constant across all firms.[24]

Let τ_{is}^{L} denote the firm-specific labor market distortions and τ_{is}^{K} denote firm-specific capital market frictions (both assumed to be constant over time), respectively impacting the marginal cost of labor and capital. Firm i in sector s maximizes the following program,

$$\pi_{is} = P_{is}Y_{is} - (1+\tau_{is}^{L})wL_{is} - (1+\tau_{is}^{K})rK_{is}. \tag{6}$$

Consequently, the factor demand for firm i in sector s can be summarized by the capital-labor ratio:

$$\ln\left(K_{is}/L_{is}\right) = \frac{1}{1-\rho}\ln\left(\frac{a_s}{1-a_s}\right) + \frac{1}{1-\rho}\ln\left(\frac{1+\tau_{is}^{L}}{1+\tau_{is}^{K}}\right) + \frac{1}{1-\rho}\ln\left(w/r\right), \tag{7}$$

which depends on (i) an industry fixed-effect, (ii) firm-specific relative distortions between the factor markets and (iii) the relative prices at the destination level.

The marginal revenue product of capital and labor, and the revenue productivity verify:

$$\begin{cases} \ln\left(MPK_{is}\right) = \ln\left(a_s P_{is} A_{is}^{\rho}\right) + (1-\rho)\ln\left(\dfrac{Y_{is}}{K_{is}}\right) = \ln(r(1+\tau_{is}^{K})) \\ \ln\left(MPL_{is}\right) = \ln\left((1-a_s)P_{is}A_{is}^{\rho}\right) + (1-\rho)\ln\left(\dfrac{Y_{is}}{L_{is}}\right) = \ln(w(1+\tau_{is}^{L})) \\ \ln\left(TFP_{is}\right) = \ln\left[\dfrac{P_{is}Y_{is}}{\left[a_s K_{is}^{\rho} + (1-a_s)L_{is}^{\rho}\right]^{\frac{1}{\rho}}}\right] = \ln(P_{is}A_{is}) \end{cases} \tag{8}$$

[24] The Cobb-Douglas production function corresponds to ρ converging toward 0.

In theory, the labor supply shift generated by the arrival of migrants should shift wages w downward (and employment upward—a *labor supply* effect). In parallel, demand for the final good may also be affected which would shift all prices P_{is} in the same proportion (a *demand* effect). In such framework where factor market distortions are constant, all firms in the same industry should be affected equally. In the empirical analysis, we will interpret deviations from this benchmark as indirect evidence of dynamic distortions in labor and capital markets.

Some quantities governing production at the firm level are not directly observed in the data, and we must estimate or calibrate them. We will proceed as follows. As in Hsieh and Klenow (2009), we use US data at the firm level from the 2015 Annual Survey of Manufactures in order to calibrate the sectoral-specific capital shares a_s. As for the constant elasticity of substitution between capital and labor, we use the recent estimates in Oberfield and Raval (2014), i.e., we set $\rho = 0.7$. The implied residuals τ_{is}^K and τ_{is}^L, capturing the distance to the unconstrained allocation of factors (for a certain firm i), will be estimated as firm fixed-effects using the panel dimension of our firm dataset.

We use this accounting framework to discipline the empirical analysis in three ways. First, we define *fixed* categories of firms depending on their factor use during the baseline period. We create, for instance, a relative measure of frictions between factors: Firms facing relatively high frictions on capital markets are defined as firms with a ratio K_{is}/L_{is} above their sectoral median during the baseline period. Second, we analyze the dynamic adjustment of firm level outcomes, including capital to labor ratio, the marginal revenue products of labor and capital, and total factor productivity. Third, in the spirit of Hsieh and Klenow (2009), we construct aggregated measures of productivity dispersion in each industry*destination.

An important assumption of this framework is that labor is homogeneous, which implies in particular that there is no productivity difference between migrant and resident workers. This assumption is driven by data limitations, since the NBS census does not break down firm employment by skill or by migrant status.[25] It can be relaxed as long as an efficient unit of labor, whether provided by a resident or a migrant, is equally costly to the firm. However, any discrepancy between the productivity of natives and migrants would generate a downward bias when estimating the effect of migrant inflows on returns to capital, and total factor productivity.

[25] The census data presented in Appendix A does in fact suggest that resident workers are more skilled than migrant workers.

3.3 Empirical strategy

We take advantage of the panel structure of the NBS data and implement a 2SLS-FE specification in which we regress the variable of interest y_{idt} for a firm i in year t in urban prefecture d on the migrant inflow to d, which we denote $M_{d,t}$, using predicted migration $\widetilde{M_{d,t}}$ as instrument and including firm fixed effects η_i and time fixed effects ν_t. Our first outcomes of interest are compensation per worker and employment, which measure the impact of the labor supply shock on the labor market equilibrium. Second, we study the capital to labor ratio in order to analyze the adjustment along the other factor. Finally, we estimate the effects on the marginal productivities of labor and capital and total factor productivity, and compare them to the benchmark model developed in the previous section.

$$\begin{cases} M_{dt} = b_0 + b_m \widetilde{M_{dt}} + c_i + n_t + e_{idt} \\ \\ y_{idt} = \beta_0 + \beta_m \widehat{M}_{dt} + \gamma_i + \nu_t + \varepsilon_{idt} \end{cases}, \tag{9}$$

with standard errors clustered at the level of the prefecture of destination.[26]

We will perform a series of robustness checks to provide support to the main identification assumptions. A first cause of concern is the failure of the exclusion restriction assumption, i.e., if the agricultural commodity price shocks have a direct effect on firms. In order to test this, we will exclude industries that process agricultural products. A second concern is that predicted migrant flows, which are constructed using distance and destination population, capture market access, which has independent effect on firm growth. To test this, we estimate Equation (9) controlling for the (log of the) destination population interacted with year fixed effects, to allow destinations to follow different trends depending on their size. A third and more general concern is that our estimation may be capturing different sectoral trends, which could be correlated with migration patterns through the geographical distribution of manufacturing activities or the diffusion of agricultural price shocks. To alleviate this concern, we include industry×year fixed effects into Equation (9), so that our estimates only capture within-industry variation in firm outcomes. Finally, we perform a standard placebo check and test whether future migration shocks have any effect on firm outcomes.

[26]Because the regressor of interest, the migration rate, is itself predicted, correct inference requires to bootstrap the first stage. The standard errors in the second stage are, however, correctly estimated through 2SLS.

4 Results

In this section, we present our findings on the absorption of migrant labor supply by the urban economy. First, we quantify the labor supply shift using employment costs for the firm, and the wage declared by residents (to clear from possible compositional effects). Second, we analyze the effects of the labor supply shock on the factor mix and returns to factors for the average firm. Third, we present results on the heterogeneity of these effects depending on firm characteristics. We also estimate the impact of such heterogeneity on the aggregate distribution of firm size and productivity at destination.

4.1 The labor supply shift

We first analyze the impact of exogenous changes in migrant inflows on labor costs and employment at destination.

In Table 2, we analyze specification 9 on the subsample of firms present from 1998 to 2006. We estimate the effect of migration on compensation per employee (including fringe benefits), employment and capital to labor ratio. For each of these outcomes, we report first OLS estimates from a regression on the actual migration rate, and, in a second column, we report IV estimates, where we use our migration prediction as an instrument for migration (see Table 1 for the first stage). All regressions in Table 2 include firm and year fixed effects and the standard errors are clustered at the prefecture level.

As columns 1 and 2 in Table 2 indicates, the inflow of rural migrants has a strong negative effect on labor costs. A one percentage point increase in the immigration rate is associated with a 0.45% decrease in wages. The IV estimates are negative and larger in magnitude: If migrants are attracted to cities that offer higher wages, OLS estimates should indeed be biased upwards. These estimates suggest that, once cleaned for demand-driven fluctuations, an influx of rural migrants depresses urban labor costs. Following Borjas (2003), we can recover the elasticity of urban wages with respect to migration by multiplying the coefficient by $\frac{1}{(1+m)^2}$, where m is the ratio of migrants to natives. In our context, the migration rate is about 20%, hence $\frac{1}{(1+m)^2} \approx 0.69$. The implied wage elasticities from our estimates is -0.8, which is markedly higher than Borjas's (2003) own estimates of -0.4. It is, however, comparable to other studies on internal migration in developing countries which use a similar strategy (Kleemans and Magruder, 2017). It may that immigrants and native are much more substitutable in the case of internal than international migration. It may also be that the labor market for unskilled labor in urban China in the period we study was relatively unregulated, which made it easier for firms to adjust wages and employment.[27]

In parallel to this decrease in labor costs, firms expand employment and absorb the excess labor force. A one percentage point increase in the migration rate increases employment by about 0.75% to 1.1% (columns 3 and 4 of Table 2). As shown in columns 5 and 6 of Table 2, the increase in employment translates into a strong fall in the capital to labor ratio, which is expected if capital is fixed in the short run. However, capital to labor ratio decreases slightly more than employment increases for the average firm in our sample. Capital may be decreasing due to higher substitutability between capital and labor in the Chinese than in the benchmark US economy. Firms may also sell some assets in order to finance the hiring of migrant workers.

In order to interpret these effects as a pure labor supply shift, we provide in Appendix D a series of robustness checks which isolate the indirect impact on urban firms through the arrival of workers from potential direct effects of agricultural shocks (through demand for non-tradable goods or the provision of intermediate goods). First, we exclude industries that process agricultural products in order to show that our results are not driven by the direct effect of agricultural price shocks on manufacturing units. Seond, we allow firms in larger urban areas to experience different trends. Third, we control flexibly for industry-specific trends in order to show that we are not simply capturing urban dynamics linked to sectoral specialization or market power. Fourth, we change our definition for migrant workers restricting them to extra-provincial flows. None of these changes affect our results.

4.2 Returns to factors

We now study the consequences of the labor supply shift on returns to factors for the average firm. Specifically, we estimate specification 9 using as outcomes the marginal revenue product of labor, the marginal revenue product of capital and total factor productivity in revenue terms (all in logs). The estimates are persented in Table 3. As in Table 2, we report OLS estimates and IV estimates, with origin-driven migration prediction as an instrument. Consistent with the system of equations 8, marginal returns to labor decrease following a positive labor supply shift (see columns 1 and 2 of Table 3). The magnitude of the decline is similar to that of the wage rate, which suggests that the difference between the marginal product and the marginal cost remains stable for the average firm. This finding is consistent with the

[27] Minimum wage regulations only came into force toward the end of our observation period (Mayneris et al., 2014).

theoretical framework assuming constant firm-specific distortions on labor markets (as captured by a constant tax over labor costs).

Surprisingly, the marginal revenue product of capital respond to the labor supply shift (see columns 3 and 4 of Table 3) while it should remain constant over time and orthogonal to migration flows with constant firm-specific distortions on capital markets. The same decrease is observed for Total Factor Productivity (see columns 5 and 6 of Table 3). This suggests that hiring cheaper migrant labor may come at a cost for the firm, unaccounted for in the theoretical framework. A likely explanation is that new workers are less skilled than average, e.g., because they do not know how to operate machinery. The equations 8 may be verified by firms in our sample, conditional on L capturing efficient labor units. Assuming, as we do in practice, that migrants are equally productive than natives, we may generate a negative correlation between migration flows and returns to factors. We provide more evidence on this below.

4.3 Dynamic effects

In order to study the dynamic effect of migrant inflows, we modify equation (9) to include lagged migration rates M_{dT} with $T \in \{t-3, t-2, t-1, t\}$ instrumented by lagged predicted migration rates. Introducing more lags reduces the number of years included in the estimation, and we estimate equation (9) with one lag, then two, then three lags to test whether the results are consistent across samples. The estimating equation writes:

$$\begin{cases} \mathbf{M}_{dt} = \mathbf{b}_0 + \mathbf{b_m}\widetilde{\mathbf{M}_{d\tau}} + \mathbf{c}_i + \mathbf{n}_{s\tau} + \mathbf{e}_{isd\tau} \\ \\ y_{idt} = \beta_0 + \sum_{\tau=t-3}^{t} \beta_m^\tau \widehat{M}_{d\tau} + \gamma + \nu_t + \varepsilon_{idt} \end{cases} \tag{10}$$

where $\mathbf{M}_{dt}$ and $\widetilde{\mathbf{M}_{dt}}$ are vectors composed of lags of M_{dt} and $\widetilde{M_{dt}}$.

Table 4 presents the results from the estimation. As column 1 to 3 show, the negative effect of migration shocks on labor costs of urban firms persists for at least two years, while the capital to ratio slowly returns to the mean. Interestingly, the negative effects on total factor productivity quickly fade away. On the whole, these results suggest that a positive migration shock depresses wages for a few years. The firm readjusts its factor mix after the first year, and the negative effects on total factor productivity on impact are short lived. These findings are consistent with a reorganization of firms to alleviate the capital to labor imbalances.

4.4 Reallocation of resources across firms

The results of this section have provided some stylized facts for the average firm. In the next section, we seek to identify which firms gain from the newly-available resources. Specifically, we study the heterogeneity in the response to migrant inflows by interacting migration shocks with different firm characteristics $\mathbf{X}_i$. We focus on three characteristics: public owned firms, exporting firms, and firms which have a capital to labor ratio above the median of firms in their sector and prefecture. The first category of firms faces frictions on labor market that are positively correlated with the immigrant inflows. The second category of firms face a highly elastic demand for the final good, implying that they have fewer constraints in expanding their production. The latter category represents firms with a relatively high residual τ_{is}^L compared to τ_{is}^K, or equivalently higher constraints on labor than on capital. The estimating equation writes:

$$\begin{cases} M_{dt} = b_0 + b_m \widetilde{M_{dt}} + c_i + n_{st} + e_{isdt} \\ \\ y_{idt} = \beta_0 + \beta_m \widehat{M_{dt}} + \beta_x \widehat{M_{dt}} \times X_i + \gamma_i + \nu_t + \mu_t \times X_i + \varepsilon_{idt} \end{cases} \tag{11}$$

The results shown in table 5 explore the heterogeneity in firm responses to the labor supply shock induced by migrant inflows. Panel A shows that public sector firms experience a slightly (not significantly) lower decrease in labor costs, but do not expand their employment when the migration rate increases. This could be due to restricted access to migrant labor for public sector firms which were at the time massively laying off urban resident workers (Naughton, 2007). Panel B shows that firms with high capital labor ratio experience higher increase in employment, lower decrease in capital to labor ratio and smaller decrease in the marginal revenue productivity of capital and the total factor productivity. These firms may have faced a relative shortage of workers, are are now expanding their workforce with little negative effect on their productivity. Finally, Panel C shows that a sizeable share of the increase in employment is due to firms which export part of their production. Taken together, these findings suggests that private, capital-intensive firms, with good access to international markets are the ones which benefit from an influx of migrants.

4.5 Aggregate effects

Our analysis so far has been limited to firms which were present every year in the sample between 1998 and 2006. In order to study the effect of migration on the manufacturing sector as a whole, we consider all firms covered by the census, and aggregate outcomes at the industry-prefecture level. We estimate the same specification as equation 9, except that the unit of observation is an industry-prefecture instead of the firm.

Table 6 presents the effects of migration on the firm size distribution. As column 2 shows, a one percentage point increase in the migration rate increases the number of firms present in the census for the average industry-prefecture cell by 16 firms (4% on the average of 400).[28] In columns 3 to 6, we consider the number of firms with revenues above 10M RMB, and 20M. We find that migration increases the number of firms by the same proportion at all thresholds. In other terms, migration shifts the entire firm size distribution to the right.

We now turn to the effect of migration on the distribution of merginal returns to factors across firms in a given industry-prefecture. In columns 1 and 2 of Table 7, we present our estimates for the causal effect of migration on the dispersion of marginal revenue product of labor. Panel A, B and C use alternative measures of this dispersion: standard deviation, difference between the 25th and 75th, and between the 10th and 90th percentiles respectively. Across measures, we find that a positive labor supply shocks decreases the dispersion of labor productivity: a one percentage point increase in migration decreases the 10th-90th percentile range by 0.7%. This finding is consistent with the heterogenous effects observed with the firm-based analysis: firms with a relatively high capital to labor ratio (and thus a high marginal revenue product of labor) are more likely to hire thereby reducing their marginal revenue product of labor. By contrast, columns 3 and 6 show that migration slightly increases dispersion of marginal revenue of capital and total factor productivity. These findings suggest that an inflow of migrant alleviates the effect of labor market distortions but may reinforce that of capital market imperfections.

4.6 Composition effects

A major shortcoming of our analysis of firms outcomes is that migration inflows may have two effects, they may change the quantity of labor supplied and the composition of the labor force in urban areas. In order to shed light on these issues, we seven cross-sections of the Urban Household Survey (2002–2008), a representative survey of urban "natives" (see description in Appendix C).

Let y_{jdt} be the labor market outcome of individual j in destination d in year t. We consider the four following outcomes: real monthly wages, the probability of

[28] For this exercise, we limit our analysis to firms just above the entry threshold, and drop the few firms below the threshold.

being wage employed, unemployed and self-employed. We regress y_{jdt} on predicted migration the year before, $\widetilde{M_{dt}}$, and a vector of individual characteristics X_j, including marital status, gender, education level, and age. As before, the effect of M_{dt} on y_{jdt} is estimated through Two-Stage Least Squares (2SLS), using $\widetilde{M_{dt}}$ as an instrument:

$$\begin{cases} M_{dt} = b_0 + b_m\widetilde{M_{dt}} + b_xX_j + c_d + n_t + e_{dt} \\ \\ y_{jdt} = \beta_0 + \beta_m M_{dt} + \delta X_j + \gamma_d + \nu_t + \varepsilon_{jdt} \end{cases}, \tag{12}$$

Since unskilled urban residents are more likely to be competing for jobs with migrant workers, they may experience larger changes in wages and occupation in response to migration inflows. In order to test this, we estimate the same specification interacting the migration shock with a dummy $LowSkill_j$ equal to one if the worker has lower secondary education or less. Table 8 presents the results.

We first consider the impact on natives' real wages. The estimates are very similar to the estimates for labor costs using establishment-level data: a one percentage point increase in the immigration rate is associated with a 1.4% decrease in wages (column 2 in Panel A). As expected, the wage results are stronger for urban residents with at most lower secondary education, who are more substitutable to migrant workers (column 2 in Panel B). These results suggests that the decline in labor costs estimated using firm data did reflect a change in equilibrium wages due to a labor supply shock and not only a composition effect.

We next consider the effect of rural to urban migration on the occupation status of urban residents (column 3 to 8 in Panel A). The OLS estimates are close to zero and mostly insignificant. IV estimates, however show that a one percentage point increase in migration decreases wage employment of urban residents by 0.3 percentage points (the average participation to wage employment is above 90%). Interestingly self-employment seems to increase in the same proportion (by 0.3 percentage points). These results provide suggestive evidence that migrants displace urban residents, pushing them into self-employment.

The estimates presented in Panel B column 4 suggest, that the negative effect on wage employment is driven by high skill workers: the (statistically insignificant) coefficient of the interaction with the low skill indicator is of opposite sign and of similar magnitude as the main coefficient. These findings suggests that less skilled urban residents suffer large wage cuts but do not have the possibility to become self-employed. They also provide modest support to the idea that migration changes the composition of wage workers by pushing out more skilled natives.

5 Conclusion

This paper provides some of the first causal empirical evidence of the impact of rural to urban migration on the allocation of factors within the urban sector. It relies on the unique combination of population censuses and a census of above scale manufacturing firms in China between 1998 in 2007, a period of rapid structural transformation and sustained manufacturing growth. We build predictions of migrant flows into urban areas based on shocks to agricultural incomes in rural origins and distance between prefectures of origin and destination. These predictions are exogenous with respect to urban outcomes, which allows us to tackle the issue of migrants self-selecting into buoyant labour markets and provide causal estimates of the effect of migration on the urban economy.

We find that the average firm experiences a large increase in employment together with a marked decrease in labor costs, which indicates that migration changes the urban labor market equilibrium. The magnitudes suggest that migrants and natives are close substitutes (the wage elasticity with respect to migration is about 0.8) and that labor demand is highly elastic (with a demand elasticity close to 0.7). These labor market effects are independently confirmed by a representative survey of urban workers. As a repsonse to this positive labor supply shock, firms increase employment so that the marginal product of labor decreases proportionally to the wage change. Capital does not adjust, capital and Total Factor Productivity decreases. These effects are temporary and suggest costly factor adjustment in the short-run.

The analysis of the heterogeneous impact of migration confirms the importance of factor market distortions. The increase in employment is concentrated in private sector firms, since public sector firms face constraints in hiring unregistered migrants. Among private sector firms, exporting firms and capital-intensive firms are those who benefit most from a positive labor supply shock, as they face a very elastic demand and relatively lower capital distortions. This heterogeneous impact has aggregate implications. Migration decreases the dispersion in the marginal product of labor, with relatively capital-intensive firms hiring more workers. The entire firm size distribution shifts to the right. These findings show that the movement of labor from the agricultural to the urban sector has important implications on the reallocation of factors within the urban economy and on manufacturing growth.

References

Accetturo, Antonio, Matteo Bugamelli, and Andrea Roberto Lamorgese, "Welcome to the machine: firms' reaction to low-skilled immigration," *Bank of Italy Working Paper*, 2012, *846.*

ADB, "The Declining Share of Agricultural Employment in the People's Republic of China: How Fast?," ADB Economics Working Paper Series 419, Asian Development Bank (ADB) November 2014.

Alvarez-Cuadrado, Francisco and Markus Poschke, "Structural Change Out of Agriculture: Labor Push versus Labor Pull," *American Economic Journal: Macroeconomics*, July 2011, *3* (3), 127–58.

Angelucci, Manuela, "Migration and Financial Constraints: Evidence from Mexico," *The Review of Economics and Statistics*, March 2015, *97* (1), 224–228.

Au, Chun-Chung and J. Vernon Henderson, "How migration restrictions limit agglomeration and productivity in China," *Journal of Development Economics*, 2006, *80*, 350–88.

_ and _ , "A numerical simulation analysis of (Hukou) labour mobility restrictions in China," *Journal of Development Economics*, 2007, *83*, 392–410.

Bazzi, Samuel, "Wealth Heterogeneity and the Income Elasticity of Migration," *American Economic Journal: Applied Economics*, 2016.

Borjas, George J., "The Labor Demand Curve Is Downward Sloping: Reexamining The Impact Of Immigration On The Labor Market," *The Quarterly Journal of Economics*, November 2003, *118* (4), 1335–1374.

Bosker, Maarten, Steven Brakman, Harry Garretsen, and Marc Schramm, "Relaxing Hukou: Increased labor mobility and China's economic geography," *Journal of Urban Economics*, 2012, *72*, 252–66.

Boustan, Leah Platt, Price V. Fishback, and Shawn Kantor, "The Effect of Internal Migration on Local Labor Markets: American Cities during the Great Depression," *Journal of Labor Economics*, October 2010, *28* (4), 719–746.

Brandt, Loren, Johannes Van Biesebroeck, and Yifan Zhang, "Challenges of working with the Chinese NBS firm-level data," *China Economic Review*, 2014, *30* (C), 339–352.

_ , Trevor Tombe, and Xiaodong Zhu, "Factor market distortions across time, space and sectors in China," *Review of Economic Dynamics*, 2013, *16* (1), 39–58.

Bryan, Gharad and Melanie Morten, "Economic Development and the Spatial Allocation of Labor: Evidence from Indonesia," Technical Report 2015.

Buera, Francisco J, Joseph P Kaboski, and Yongseok Shin, "Finance and development: A tale of two sectors," *The American Economic Review*, 2011, *101* (5), 1964–2002.

Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli, "Agricultural productivity and structural transformation: evidence from Brazil.," 2015. Forthcoming *American Economic Review.*

Chan, Kam Wing, "Misconceptions and Complexities in the Study of China's Cities: Definitions, Statistics, and Implications," *Eurasian Geography and Economics*, 2007, *48* (4), 383–412.

_ , "The Chinese hukou system at 50," *Eurasian geography and economics*, 2009, *50* (2), 197–221.

_ , "The household registration system and migrant labor in China: Notes on a debate," *Population and development review*, 2010, *36* (2), 357–364.

_ , "Crossing the 50 Percent Population Rubicon: Can China Urbanize to Prosperity?," *Eurasian Geography and Economics*, 2012, *53* (1), 63–86.

_ , "Migration and development in China: trends, geography and current issues," *Migration and Development*, 2012, *1* (2), 187–205.

_ **and Will Buckingham**, "Is China abolishing the hukou system?," *The China Quarterly*, 2008, *195*, 582–606.

Cheng, Wenya and Andrei Potlogea, "Trade Liberalization and Economic Development: Evidence from China's WTO Accession," 2017. Manuscript.

Combes, Pierre-Philippe, Sylvie Démurger, and Shi Li, "Migration externalities in Chinese cities," *European Economic Review*, 2015, *76* (C), 152–167.

Demurger, Sylvie, Marc Gurgand, Shi Li, and Ximing Yue, "Migrants as second-class workers in urban China? A decomposition analysis," *Journal of Comparative Economics*, December 2009, *37* (4), 610–628.

Dustmann, Christian and Albrecht Glitz, "How Do Industries and Firms Respond to Changes in Local Labor Supply?," *Journal of Labor Economics*, 2015, *33* (3), 711–750.

Ebenstein, Avraham and Yaohui Zhao, "Tracking rural-to-urban migration in China: Lessons from the 2005 inter-census population survey," *Population Studies*, 2015, *69* (3), 337–353.

El Badaoui, Eliane, Eric Strobl, and Frank Walsh, "The Impact of Internal Migration on Local Labour Markets in Thailand," EconomiX Working Papers 2014-12, University of Paris West - Nanterre la Défense, EconomiX 2014.

Facchini, Giovanni, Maggie Y. Liu, Anna Maria Mayda, and Minghai Zhou, "The impact of China's WTO accession on internal migration," December 2015.

Fan, Cindy C., *China on the Move*, Routledge, 2008.

Feng, Shuaizhang, Michael Oppenheimer, and Wolfram Schlenker, "Weather Anomalies, Crop Yields, and Migration in the US Corn Belt," March 2015.

_ , Yingyao Hu, and Robert Moffit, "Long Run Trends in Unemployment and Labor Force Participation in China," August 2015.

Ge, Suqin and Dennis Tao Yang, "Changes In China's Wage Structure," *Journal of the European Economic Association*, 04 2014, *12* (2), 300–336.

Giesing, Yvonne and Nadzeya Laurentsyeva, "Brain Drain and Firm Productivity: Evidence from the Sequential Opening of EU Labour Markets," 2015. Unpublished.

Gollin, Douglas, Stephen Parente, and Richard Rogerson, "The Role of Agriculture in Development," *American Economic Review*, May 2002, *92* (2), 160–164.

Harris, John R. and Michael P. Todaro, "Migration, Unemployment and Development: A Two-Sector Analysis," *The American Economic Review*, 1970, *60* (1), pp. 126–142.

Herrendorf, Berthold, Christopher Herrington, and Akos Valentinyi, "Sectoral Technology and Structural Transformation," Technical Report 9386, C.E.P.R. Discussion Papers March 2013.

Hodrick, R. and E. Prescott, "Post-war U.S. business cycles: An empirical investigation," *Journal of Money, Credit and Banking*, 1997, *29* (1), 1–16.

Hsieh, Chang-Tai and Peter J. Klenow, "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, November 2009, *124* (4), 1403–1448.

_ and Zheng Song, "Grasp the Large, Let Go of the Small: The Transformation of the State Sector in China," NBER Working Papers 21006, National Bureau of Economic Research, Inc March 2015.

Imbert, Clement and John Papp, "Short-term Migration, Rural Workfare Programs and Urban Labor Markets - Evidence from India," The Warwick Economics Research Paper Series (TWERPS) 1116, University of Warwick, Department of Economics 2016.

Kerr, Sari Pekkala, William R. Kerr, and William F. Lincoln, "Skilled Immigration and the Employment Structures of US Firms," *Journal of Labor Economics*, 2015, *33* (S1), S147 – S186.

Kleemans, Marieke and Jeremy Magruder, "Labor Market Changes in Response to Immigration: Evidence from Internal Migration Driven by Weather Shocks in Indonesia," *Economic Journal*, 2017. Manuscript.

Kuznets, Simon, *Agriculture in Economic Development*, McGraw-Hill Book Company, 1964.

Lewis, Arthur, "Economic Development with Unlimited Supplies of Labour," *The Manchester School*, 1954, *22* (2), 139–191.

Mayneris, Florian, Sandra Poncet, and Tao Zhang, "The cleansing effect of minimum wages. Minimum wages, firm dynamics and aggregate productivity in China," Core Discussion Papers 2014/44 October 2014.

Meng, Xin and Dandan Zhang, "Labour Market Impact of Large Scale Internal Migration on Chinese Urban 'Native' Workers," IZA Discussion Papers 5288 October 2010.

Midrigan, Virgiliu and Daniel Yi Xu, "Finance and misallocation: Evidence from plant-level data," *The American Economic Review*, 2014, *104* (2), 422–458.

Moll, Benjamin, "Productivity losses from financial frictions: can self-financing undo capital misallocation?," *The American Economic Review*, 2014, *104* (10), 3186–3221.

Naughton, Barry, *The Chinese Economy: TTransition and Growth*, MIT Press, 2007.

Oberfield, Ezra and Devesh Raval, "Micro data and macro technology," Technical Report, National Bureau of Economic Research 2014.

Olney, William W, "Immigration and firm expansion," *Journal of Regional Science*, 2013, *53* (1), 142–157.

Park, Albert, *China Urbanizes: Consequences, Strategies, and Policies*, The World Bank, 2008.

Peri, Giovanni, "The Effect Of Immigration On Productivity: Evidence From U.S. States," *The Review of Economics and Statistics*, February 2012, *94* (1), 348–358.

Perkins, D.H. and T.G. Rawski, "Forecasting China's economic growth," in L. Brandt and T.G. Rawski, eds., *China's great economic transformation*, Cambridge University Press, 2008.

Song, Yang, "What should economists know about the current Chinese hukou system?," *China Economic Review*, 2014, *29*, 200–212.

Song, Zheng, Kjetil Storesletten, and Fabrizio Zilibotti, "Growing Like China," *American Economic Review*, 2011, *101* (1), 196–233.

Sousa, Jose De and Sandra Poncet, "How are wages set in Beijing?," *Regional Science and Urban Economics*, 2011, *41* (1), 9–19.

Tombe, Trevor and Xiaodong Zhu, "Trade, Migration and Productivity: A Quantitative Analysis of China," June 2015.

United Nations, "Trends in International Migrant Stock: The 2015 revision.," Technical Report, United Nations, Department of Economics and Social Affairs 2015.

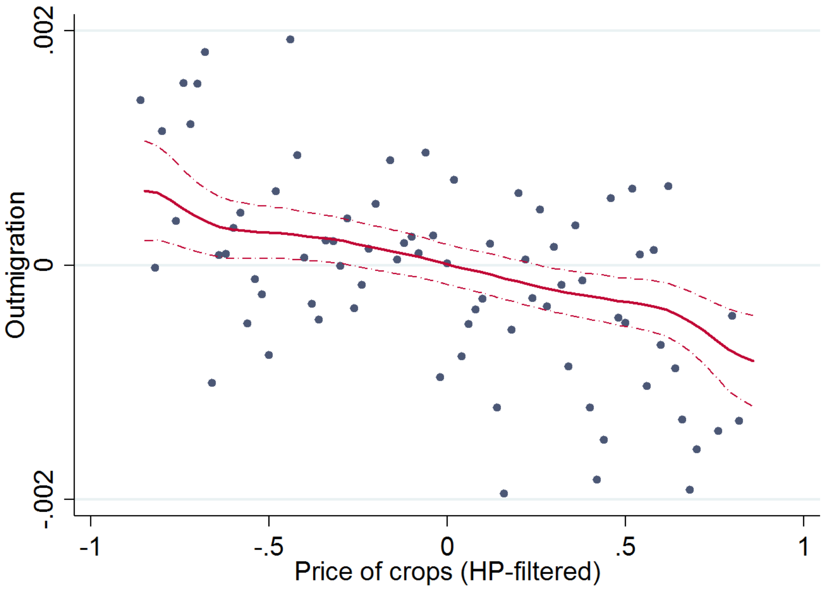
Vendryes, Thomas, "Migration constraints and development: *Hukou* and capital accumulation in China," *China Economic Review*, 2011, *722*, 669–92.

Wang, Fei-Ling, *Organizing through division and exclusion: China's hukou system*, Stanford University Press, 2005.

Zhang, Li and Li Tao, "Barriers to the acquisition of urban hukou in Chinese cities," *Environment and Planning A*, 2012, *44* (12), 2883–2900.

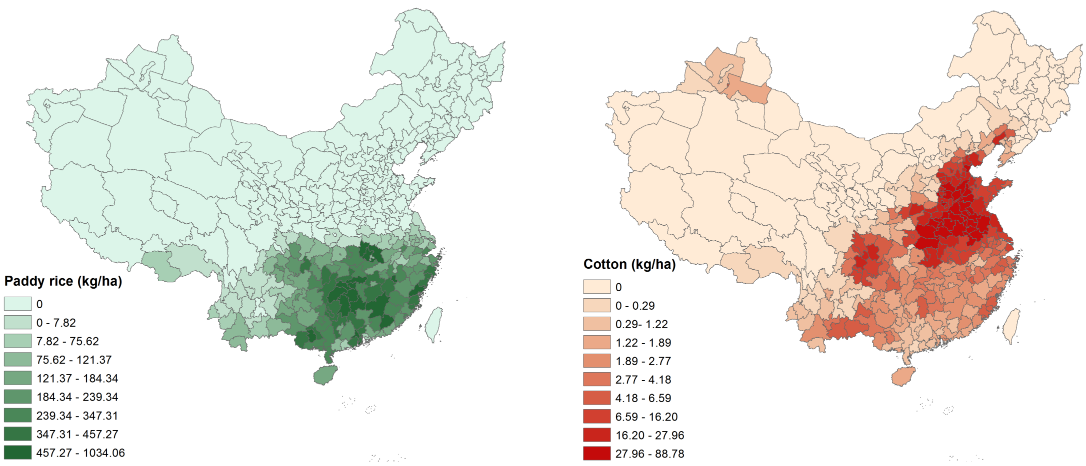
A Figures and tables

Figure 1. Value of agricultural portfolio at origin and outmigration rates.



Notes: This Figure illustrates the relationship between the standardized value of the prefecture-specific agricultural portfolio as predicted by international prices (x-axis) and outmigration (y-axis). We consider the residuals of all measures once cleaned by prefecture and year Fixed-Effects. For the sake of exposure, we group prefecture×year observations, create 100 bins of observations with similar price shock and represent the average outmigration rate within a bin. The lines are locally weighted regressions on all observations.

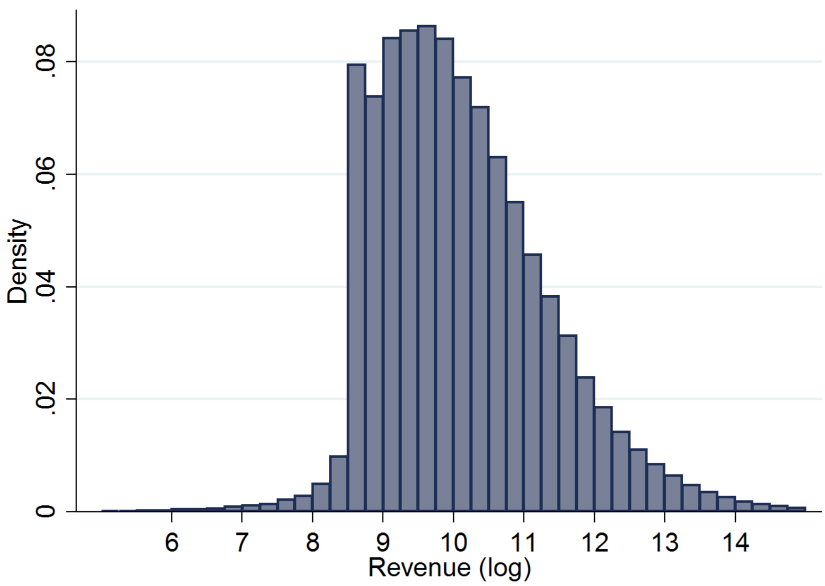
Figure 2. Potential output in China for rice and cotton (2000).



(a) Paddy rice. (b) Cotton.

Notes: These two maps represent the potential output constructed with 2000 harvested areas and potential yield (GAEZ model) in 2000 for 2 common crops in China, i.e. paddy rice (left panel), and cotton (right panel).

Figure 3. Distribution of revenue across firms (NBS, 1997–2008).



Sources: Establishment-level data from the National Bureau of Statistics (NBS), 1997–2008.

Table 1. Migration Predictions

Panel A: Predicting outmigration	
	Outmigration rate
Price Shock (standardized)	-0.104***
	(0.028)
Observations	1,690
Year Fixed-Effects	Yes
Origin Fixed-Effects	Yes
Panel B: Gravity equation	
	Share of migrants
Inverse of distance	8.449***
	(0.065)
Population at destination (millions)	3.824***
	(0.046)
Observations	116,623
Origin Fixed-Effects	Yes
Panel C: Predicting immigration	
	Immigration Rate
Predicted Migration Rate	0.981***
	(0.265)
Observations	1,690
Year Fixed-Effects	Yes
Destination Fixed-Effects	Yes

Notes: Standard errors are clustered at the prefecture level and are reported between parentheses. ***: $p<0.01$, **: $p<0.05$, *: $p<0.1$. In Panel A the sample is all prefectures every year and the outcome variable is the number of rural out-migrants to urban areas divided by the number of rural residents. In Panel B the sample is all province pairs and the outcome variable is the probability of going from each rural origin to each urban destination. In Panel C the sample is all provinces for each year and the outcome variable is the number of rural immigrants from other provinces divided by the number of urban residents.

Table 2. Impact of migration inflows on urban firms – average effect of the labor supply shift.

	Labor cost		Employment		Capital to labor	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Migration	-0.448*	-1.276***	0.721***	1.095***	-0.383	-1.582***
	(0.232)	(0.368)	(0.230)	(0.411)	(0.238)	(0.491)
Observations	353,133	353,133	354,453	354,453	353,538	353,538
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors are clustered at the prefecture level and are reported between parentheses. ***: p<0.01, **: p<0.05, *: p<0.1. The sample is composed of the 44,981 firms present every year in the NBS firm census between 1998 and 2006. All specifications include firm and year fixed effects. In the IV estimation, the instrument is the migration rate predicted using price shocks at origin, distance between origin and destination, and destination population. *Migration* is the immigration rate, i.e., the migration flow over population at baseline. *Labor cost* is the (logarithm of the) compensation per worker including social security. *Employment* is the (logarithm of the) number of workers within the firm. *Capital to labor* is the (logarithm of the) ratio of employment to fixed assets (evaluated at their current prices).

Table 3. Impact of migration inflows on urban firms – marginal product of factors.

	Return to labor		Return to capital		TFP	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Migration	-0.305***	-1.010***	-0.099***	-0.342***	-0.296***	-0.840***
	(0.104)	(0.217)	(0.038)	(0.117)	(0.0881)	(0.243)
Observations	305,055	305,055	304,689	304,689	304,689	304,689
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors are clustered at the prefecture level and are reported between parentheses. ***: p<0.01, **: p<0.05, *: p<0.1. The sample is composed of the 44,981 firms present every year in the NBS firm census between 1998 and 2006. All specifications include firm and year fixed effects. In the IV estimation, the instrument is the migration rate predicted using price shocks at origin, distance between origin and destination, and destination population. *Migration* is the immigration rate, i.e., the migration flow over population at baseline. *Return to labor* is the (logarithm of the) marginal revenue product of labor as defined in Section 3. *Return to capital* is the (logarithm of the) marginal revenue product of capital as defined in Section 3. *TFP* is the (logarithm of the) total factor productivity in revenue terms as defined in Section 3.

Table 4. Impact of migration inflows on urban firms – dynamic adjustment.

	Labor cost		Capital to labor		TFP	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Migration	-0.296**	-1.517***	-0.188*	-1.694***	-0.330***	-0.915***
	(0.117)	(0.293)	(0.111)	(0.412)	(0.080)	(0.274)
Migration L1	-0.153**	-0.925***	-0.028	-0.467	-0.171***	-0.225
	(0.077)	(0.267)	(0.088)	(0.341)	(0.062)	(0.229)
Migration L2	0.038	-0.942***	0.163	-0.600*	-0.208***	-0.246
	(0.107)	(0.345)	(0.111)	(0.354)	(0.068)	(0.239)
Observations	266,922	266,922	267,203	267,203	230,990	230,990
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors are clustered at the prefecture level and are reported between parentheses. ***: p<0.01, **: p<0.05, *: p<0.1. The sample is composed of the 44,981 firms present every year in the NBS firm census between 1998 and 2006. All specifications include firm and year fixed effects. In the IV estimation, the instrument is the migration rate predicted using price shocks at origin, distance between origin and destination, and destination population. *Migration* is the immigration rate, i.e., the migration flow over population at baseline. *Labor cost* is the (logarithm of the) compensation per worker including social security. *Capital to labor* is the (logarithm of the) ratio of employment to fixed assets (evaluated at their current prices). *TFP* is the (logarithm of the) total factor productivity in revenue terms as defined in Section 3

Table 5. Impact of migration inflows on urban firms – heterogeneous effect of the labor supply shift.

	Labor cost		Employment		Capital to labor	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Migration	-0.327	-1.315***	0.583***	0.187	-0.332*	-1.428***
	(0.211)	(0.450)	(0.180)	(0.427)	(0.199)	(0.496)
Migration × Public	0.016	0.531	-0.185	-1.148***	0.094	1.043*
	(0.141)	(0.407)	(0.112)	(0.402)	(0.156)	(0.573)
Migration × Export	0.056	-0.065	-0.120**	0.337**	0.083	-0.672***
	(0.072)	(0.197)	(0.058)	(0.169)	(0.063)	(0.244)
Migration × High K/L	-0.162	-0.091	0.221	1.173***	-0.086	-0.137
	(0.169)	(0.414)	(0.188)	(0.433)	(0.197)	(0.600)
Observations	353,133	353,133	354,453	354,453	353,538	353,538
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors are clustered at the prefecture level and are reported between parentheses. ***: p<0.01, **: p<0.05, *: p<0.1. The sample is composed of the 44,981 firms present every year in the NBS firm census between 1998 and 2006. All specifications include firm and year fixed effects. In the IV estimation, the instrument is the migration rate predicted using price shocks at origin, distance between origin and destination, and destination population. *Migration* is the immigration rate, i.e., the migration flow over population at baseline. *Labor cost* is the (logarithm of the) compensation per worker including social security. *Employment* is the (logarithm of the) number of workers within the firm. *Capital to labor* is the (logarithm of the) ratio of employment to fixed assets (evaluated at their current prices).

Table 6. Impact of migration inflows on urban firms – firm size distribution.

	Firms $\geq$ 5M		Firms $\geq$ 10M		Firms $\geq$ 20M	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Migration	762.2***	1,649***	655.3***	1,203***	454.4***	729.7***
	(147.7)	(272.7)	(134.4)	(207.0)	(97.35)	(136.8)
Observations	17,940	17,940	17,940	17,940	17,940	17,940
Mean in Sample	401	401	304	304	200	200
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors are clustered at the prefecture level and are reported between parentheses. ***: p<0.01, **: p<0.05, *: p<0.1. The aggregation is performed using all firms present at any time in the NBS firm census between 1998 and 2006. The unit of observation is an industry × prefecture × year. In columns 1 and 2 the outcome is the number of firms in this industry. In columns 3 and 4, it is the number of firms with sales above 5M RMB. In columns 5 and 6, it is the number of firms with sales above 20M RMB. In the IV estimation, the instrument is the migration rate predicted using price shocks at origin, distance between origin and destination, and destination population. All specifications include prefecture and year fixed effects, and they are weighted by the number of firms in the industry × prefecture in 1998.

Table 7. Impact of migration inflows on urban firms – productivity dispersion.

	Dispersion (MRPL)		Dispersion (MRPK)		Dispersion (TFPR)	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Standard Deviation						
Migration	0.022	-0.290***	0.042**	0.161*	0.049	0.181
	(0.020)	(0.107)	(0.020)	(0.095)	(0.040)	(0.189)
Observations	16,018	16,018	16,007	16,007	16,007	16,007
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Log difference 25th-75th percentile						
Migration	0.056	-0.223	0.087**	0.241	0.130	0.503*
	(0.038)	(0.147)	(0.041)	(0.154)	(0.087)	(0.267)
Observations	17,268	17,268	17,259	17,259	17,259	17,259
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes
Panel C: Log difference 10th-90th percentile						
Migration	0.143**	-0.737***	0.244***	0.600**	0.313**	0.430
	(0.067)	(0.269)	(0.074)	(0.254)	(0.132)	(0.487)
Observations	17,940	17,940	17,259	17,259	17,259	17,259
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors are clustered at the prefecture level and are reported between parentheses. ***: p<0.01, **: p<0.05, *: p<0.1. The aggregation is performed using all firms present at any time in the NBS firm census between 1998 and 2006. The unit of observation is an industry × prefecture × year. The outcomes are different measures of the dispersion of MRPL, MRPK and TFPR across firms. In panel A the outcomes are standard deviations, in panel B they are differences between the log of the 75th and 25th percentiles, in panel C they are differences between the log of the 90th and 10th percentile. MRPL, MRPK and TFPR are defined in the data section. All specifications include industry*prefecture and year fixed effects. In the IV estimation, the instrument is the migration rate predicted using price shocks at origin, distance between origin and destination, and destination population. All estimates are weighted by the number of firms in the industry*prefecture in 1998. Standard errors are clustered at the prefecture level.

Table 8. Impact of migration inflows on urban residents – the labor supply shift.

Panel A: Main	Real Monthly Wage		Employee		Unemployed		Self-employed	
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Migration Rate	-0.234*	-1.479*	-0.007	-0.351**	-0.028**	0.041	0.036	0.309**
	(0.122)	(0.800)	(0.038)	(0.177)	(0.014)	(0.065)	(0.034)	(0.146)
Observations	244,020	244,020	271,600	271,600	271,600	271,600	271,600	271,600
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: Unskilled	Real Monthly Wage		Employee		Unemployed		Self-employed	
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Migration Rate	-0.181	-1.344	-0.017	-0.405**	-0.034**	0.054	0.051	0.350**
	(0.117)	(0.826)	(0.036)	(0.177)	(0.013)	(0.066)	(0.033)	(0.145)
MR $\times$ Unskilled	-0.260**	-1.191***	0.042	0.391	0.024	-0.030	-0.067	-0.361
	(0.118)	(0.458)	(0.051)	(0.249)	(0.018)	(0.053)	(0.049)	(0.238)
Observations	244,020	244,020	271,600	271,600	271,600	271,600	271,600	271,600
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors are clustered at the prefecture level and are reported between parentheses. ***: p<0.01, **: p<0.05, *: p<0.1. All specifications include year and destination fixed effects. Unskilled workers are defined as workers with no education, primary education or lower secondary education. *Employee* is a dummy for receiving a wage, while *Self-employed* is a dummy equal to one for individuals that are self-employed or an employer.

ONLINE APPENDIX

A Migration flows: construction and description

In this section, we provide some elements of context about migration in China, focusing on the *hukou* system, its implementation over time and across provinces. We then describe the construction of migration flows from retrospective questions, and the adjustment accounting for return migration. Finally, we discuss few descriptive statistics.

A.1 Elements of context

An important feature of China's society is the division of the population according to its household registration or *hukou* status. Chinese citizens are classified along two dimensions: their *hukou* type (*hukou xingzhi*)—agricultural (*nongye*) or non-agricultural (*fei nongye*)—and *hukou* location (*hukou suozaidi*). Both characteristics, recorded in the household registration booklet, depend on the household one was born into and may not correspond to the actual occupation and location.

Since the inception of the reforms in the late 1970s, rules regarding migration within China have been relaxed. Labor mobility remains subject to legal requirements—e.g., being lawfully employed at destination—but the large flows of internal migrants that have characterized China's recent development illustrate the fact that barriers are low in practice. Migrants however seldom gain local registration status and therefore do not enjoy the same rights as the locally registered population. This is likely to impede mobility but more importantly it reduces migrant workers' bargaining power and means that migrants are locked in a position of "second-class workers" (Demurger et al., 2009). Whereas an agricultural *hukou* grants access to land, non-agricultural *hukou* holders enjoy public services at their place of registration. Given the predominance of rural-urban migration—see below,—we focus on the challenges faced by agricultural *hukou* holders settling in urban areas to briefly describe the *hukou* system.

The type and place of registration have far-reaching consequences. Access to welfare benefits and public services (e.g., enrollment in local schools, access to health care, urban pension plans and subsidized housing) is conditional on being officially recorded as a local urban dweller. Subsequently, migrants face a high cost of living in cities and are supposed to return to their places of registration for basic services such as education and health care or charged higher fees (Song, 2014). Labor outcomes are also affected as local governments may issue regulations restricting access to job opportunities or rely on informal guidelines to employers to favor local permanent residents. As it became possible for state-owned enterprises (SOEs) to lay off "permanent workers" in the 1990s, regulations were introduced to bar them from employing migrant labor instead (Demurger et al., 2009).

Despite the rigidity of the *hukou* system and the persistently low rate of *hukou* conversion, reforms have progressively been introduced during the structural transformation of China. Since the 1980s, China has experienced a gradual devolution of power from the central to local governments in terms of *hukou* policy and management (Chan, 2010). As a consequence, rules and implementation may vary substantially across places and over time. Provincial governments typically set general guidelines and more specific rules are then determined by prefectures (Song, 2014), which in practice hold the most power over *hukou* policy (Wang, 2005). Two major reforms were introduced in recent years. First, the distinction between agricultural and non-agricultural *hukou* has been abolished within local jurisdictions in about one third of Chinese provinces (Chan, 2012a). Albeit an important evolution, this reform does not affect the majority of rural-urban migrants who come from other prefectures or even provinces—see Song (2014) and below. Second, *hukou* conversion rules have been gradually loosened. The main channels to change one's *hukou* from agricultural to non-agricultural (*nongzhuanfei*) used to include recruitment by an SOE, receiving college education or joining the army (Chan, 2009). These conditions have been relaxed since 2000 (Chan and Buckingham, 2008), in particular in small cities and towns (Zhang and Tao, 2012), which however attract fewer migrants. In larger cities, however, conditions for eligibility are tough and annual quotas low, so that *hukou* conversion reforms primarily benefit the richest and highly educated (Song, 2014).

A.2 Data sources and construction of migration flows

Data description In order to measure migration flows, we use the 2000 Population Census, the 2005 1% Population Survey, also called "2005 mini-census" and the 2010 Population Census.

After the beginning of the reforms and loosening of restrictions on mobility, there was a growing disconnect between census data focusing on *hukou* location and the rising "floating population" (*liudong renkou*) of non-locally registered citizens. The 2000 Population Census was the first census to acknowledge this gap and record migrants' place of residence—provided they had been living there for more than 6 months (Ebenstein and Zhao, 2015). In addition to the place of residence (at the prefecture level in our data), *hukou* location (province level) and *hukou* type, the

2000 and 2010 Population Censuses contain retrospective information on the place of residence five years before the survey (at the province level), the place of residence before moving to the current destination (at the prefecture level), the year of arrival at destination (if within the past 5 years) and the reason for departure.

The 2005 1% Population Survey constitutes a 1.3% (!) sample of the population selected from 600,000 primary census enumeration districts thanks to a three-stage cluster sampling (Ebenstein and Zhao, 2015). All Chinese counties are covered. The sampling weights provided by the National Bureau of Statistics (NBS) account for the underlying proportional probability sampling scheme based on the 2004 population registry of the Public Security Bureau, which records people at their places of registration.

A few caveats are in order. First, the sampling frame contained only information on population by registration. High-immigration areas could thus be under-sampled. Comparing the flows for 2000 according to the 2000 Census and 1% Survey, we indeed find a small discrepancy that we attribute to coverage issues. Second, the 2005 1% Population Survey offers a set of variables similar to standard censuses but some discrepancies are worth bearing in mind: (i) Both data sources provide prefecture-level information on the place of residence but it is defined as "current residence" in 2005 and thus also captures migrants who have been established at destination for less than 6 months. (ii) The 2000 and 2010 Censuses contain prefecture-level information on the place of residence prior to departure, while the 1% Survey records *hukou* location at the prefecture level. These two places are one and the same if there is no step migration, i.e., if rural dwellers move directly to their final destination. Along the same lines, the 1% Survey records the timing of *departure* from a migrant's place of registration rather than of *arrival* at destination, so that figures need not exactly coincide—unless, again, there is no step migration. Third, although we do not need this assumption for the empirics as our analyses are carried out at the prefecture level, it is important to bear in mind that the data do not allow us to determine whether a migrant is residing in a rural or urban area. The census and mini-census data do not record the place of residence at high enough resolution to unambiguously infer whether the destination is urban or rural. Nevertheless, it is clear from the literature that rural-to-rural migration represents a small share of outmigration from rural areas, not least because most of it is explained by marriages, which usually give right to local registration (Fan, 2008; Chan, 2012b).[29] Fourth, we cannot account for migrants who changed their *hukou* location or type. This

[29]Only 4.7% of agricultural *hukou* inter-prefectural migrants in the 2005 mini-census reported having left their place of registration to live with their spouses after marriage.

assumption is quite innocuous given that *hukou* conversion is marginal.

Migration flow construction The retrospective data on migration spells contained in the 2000 and 2010 Censuses, and the 2005 1% Population Survey allows us to construct yearly migration flows over the period 1996-2010. Importantly, these flows are directly measured rather than computed as a difference of stocks as is common in the migration literature.

We construct annual migration flows between each prefecture of origin and destination by combining information on the current place of residence (the destination), the place of registration (the origin) and the year in which the migrant left the origin. One advantage of working with Censuses is that it covers–or is representative of—the whole population: all individuals, irrespectively of their *hukou* status, are interviewed in 2000, 2005 and 2010. However, not all migration spells are observed in the Census data. We describe below (i) which migration spells are directly observed and which migration spells are omitted, and (ii) how we can infer some of these unobserved spells and adjust the raw migration flows.

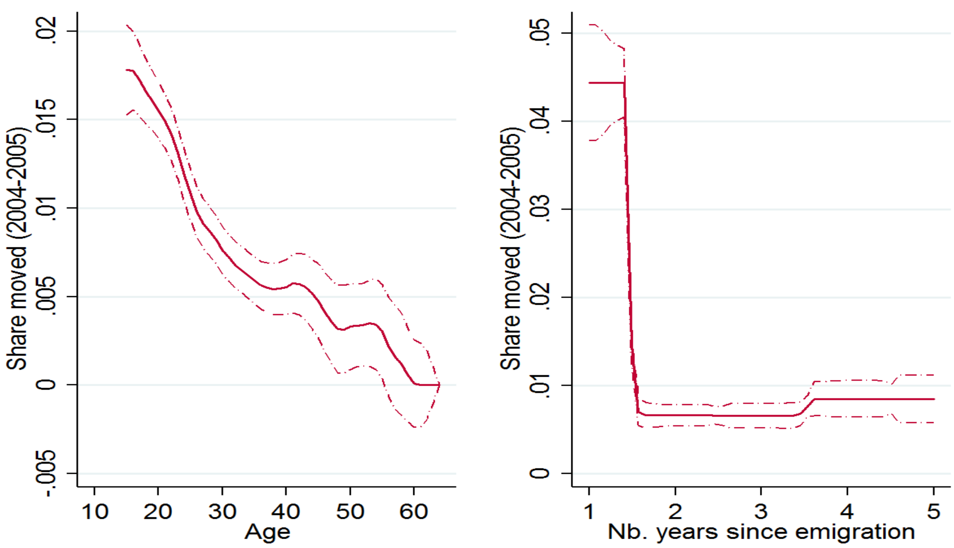
Not all migration spells are observed in the three censuses. We only observe single migration spells, i.e., migration spells in which the interviewed individual is at destination during the survey, and whose origin coincides with the *hukou* location. For these individuals, the origin is deduced from their *hukou* location, and the date of their unique relocation is available. All other types of migration histories during the five years preceding the survey are not easily reconstructed.

For instance, if one individual were to leave her *hukou* location to city A in 2002 and then transit to city B in 2005, we would only record the last relocation. In such *step migration* cases, we would correctly attribute arrival dates at destination for the last spell but we would incorrectly attribute the departure time from origin in the Population Censuses. In the 2005 1% Population Survey, we would incorrectly attribute arrival dates at destination for the last spell but we would correctly specify the departure time from origin. In both data sets, we would miss arrival in city A. If, instead, one individual were to leave her *hukou* location to city A in 2002 and then return to her *hukou* location in 2005, we would miss her entire migration history. In such *return migration* cases, we would incorrectly omit outmigration flows from rural areas, and immigrant inflows to urban settlements.

A fraction of the *step migration* and *return migration* spells can, however, be observed in the restricted sample of individuals interviewed in the 2005 1% Population Survey, and the 2000 and 2010 Censuses. Indeed, the 2005 1% Population Survey includes the locations in which the individuals were living one and five years before the survey (at the province level) while the 2000 and 2010 Censuses only include a question about the residence five years prior to the survey. We can estimate how many migrants report different destinations, which would be a proxy for step migration, and we can observe total return migration between 1995 and 2000, 2000 and 2005, 2004 and 2005, and 2005 and 2010.

We first study the importance of step migration. Among all migrants who lived in their province of registration in 2000 and were living in another province in 2005, we compute the fraction that lived in yet another province in 2004. As Appendix Figure A1 shows, only a minority of migrants have changed provinces of destination in the last year. Step migration is not only small, but concentrated in the very first year after the first migration spell. In other words, step migration induces errors in arrival and departure dates that are also quite small. As adjusting for step migration would require strong assumptions about the intermediate destination, which is not observed in the data, we do not correct the raw flows for step migration.

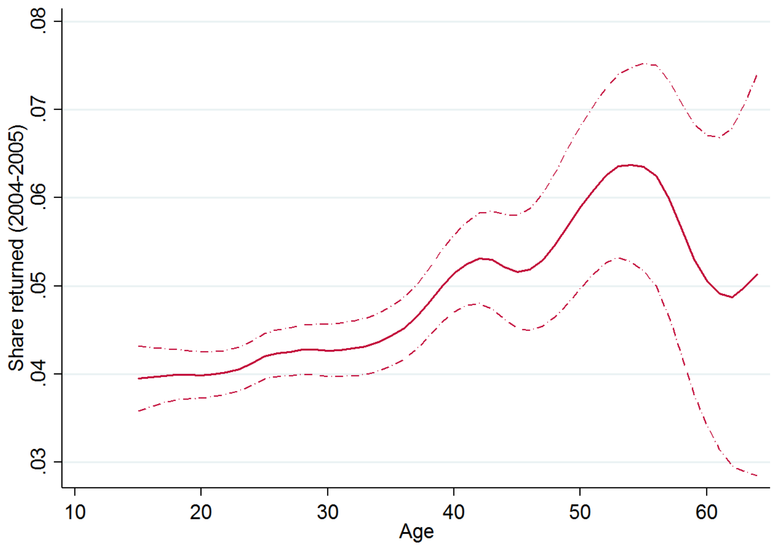
Figure A1. Share of step migrants as a function of age and time since departure.



Sources: 2005 Mini-Census.

We then consider the extent of return migration. Among all migrants from rural areas who lived in their province of registration in 2000 and who lived in another province in 2004, we compute the fraction that had returned to their province of registration by 2005. As Appendix Figure A2 shows, this share is not negligible: in a given year, between 4 and 6% of rural migrants who have left their province of registration in the last six years go back to their *hukou* location. Return migration is hence an important phenomenon, which leads us to underestimate true migration flows and the effect of shocks on out-migration. Because of the retrospective nature of the data, past flows, for instance in 2000 for an individual interviewed in 2005, are mechanically underestimated. In contrast with step migration, however, it is possible—under reasonable assumptions—to adjust migration flows and account for return migration. We provide below a description of these adjustments.

Figure A2. Share of return migrants by age.

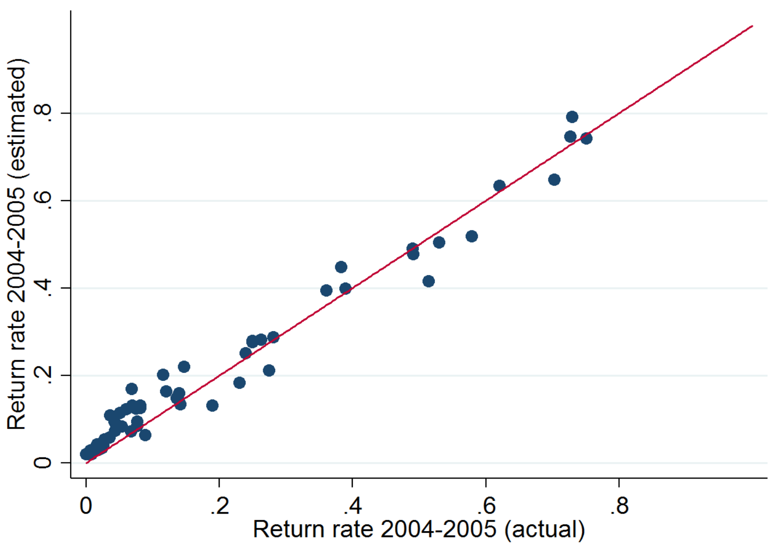


Sources: 2005 Mini-Census.

Adjusting for return migration requires to observe the destination and duration-specific yearly rate of return. Indeed, there is a wide disparity across destinations in return rates, and, as in any survival analysis with censoring, there are large compositional adjustments along the duration of the migration spell. Specifically, the probability for a migrant to return home sharply decreases with the length of the migration spell, mostly reflecting heterogeneity across migrants in their propensity to return. To capture these differences across destinations and length of the migration spell, we make the following assumptions. (i) Each migrant is characterized by a constant Poisson rate f of returning. (ii) We suppose that there is a fixed and destination-specific distribution of migrant types $H(.)$ upon arrival. We also allow the distributions to differ across *hukou* type, i.e., rural or urban. (iii) In order to fit the observed return rates as a function of migration duration, we further assume that $h(f) = \lambda_p^2 f e^{-\lambda_p f}$ where λ_p, the province-specific exponential parameter, will be calibrated using actual return rates within five years, for individuals at destination five years before the survey.

Under these assumptions, the evolution of the pool of migrants with duration can easily be computed. In the cross-section (i.e., with all cohorts and not only newly-arrived migrants), the distribution of migrant types is $h_c(f) = \lambda_p e^{-\lambda_p f}$ such that the average return rate in the cross-section is $1/\lambda_p$. The targeted moment, i.e., the return rate over 5 years, is $1/(\lambda_p + 5)$. We calibrate the *hukou*- and province-specific exponential parameter λ_p to match the actual return rate, and we perform this calibration *in each survey* such that we flexibly allow for long-term fluctuations in these province-specific distributions. The correction term that we apply to migration spells is more complicated to compute, and depends upon the return rate over k years for newly-arrived migrants, $1 - \lambda_p^2/(\lambda_p + k)^2$. We carry out this exercise for the 2000 and 2010 Censuses and 2005 Survey data.[30]

Figure A3. Over-identification test for the return migration correction.



Sources: 2005 Mini-Census.

One concern with this methodology is that we may not precisely capture the duration-dependence in return rates, and thus over- or under-estimate return rates for individuals arriving immediately before the interview. Using the 2005 1% Population Survey, we can provide an over-identification test by computing the return probability between 2004 and 2005 for recently-arrived migrants (i.e., between 2000 and 2004), and compare it with the empirical moment. As shown in Figure A3, the

[30]We also adjust for coverage issues in the 1% Population Survey due to its sampling design in order to obtain consistent figures across data sets. We calculate the prefecture-level and *hukou*–specific adjustment term that matches the flows (corrected for return migration) in 2000, the only year for which the Census and the survey overlap.

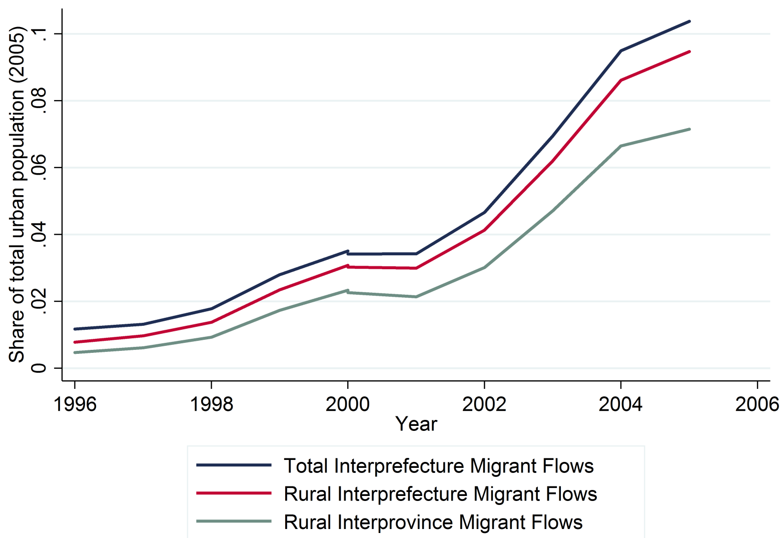
prediction—computed with the cross-sectional return rate over 5 years—matches quite well the actual annual return rate for recently-arrived migrants.

A.3 Description

In this section, we provide some descriptive statistics about migration flows and the selection of migrants.

Migration patterns over time and across regions Figure A4 illustrates the rise in inter-prefectural migrant flows between 1996 and 2005 as a share of the population registered in urban areas in 2005. The rising trend and the magnitude of migration flows is striking: In 2005, the annual inflow of migrants from other prefectures was around 10% of the destination population, as against less than 2% in 1996. Two interesting facts pertain to the composition of the incoming migrants. First, about 80% of the yearly migrant inflows consist of agricultural *hukou* holders ("rural" migrants), the remainder being accounted for by urban dwellers originating from other prefectures. Second, about 80% of inter-prefectural rural-urban migrations recorded over the period 1996-2005 involved the crossing of a provincial border.

Figure A4. Evolution of migration rates between 1996 and 2005.



Sources: 2000 Census, and 2005 Mini-Census.

As far as variation in migration outflows across space is concerned, Table A1 compares outmigration (top panel) and immigration (middle panel) rates in 2000 and 2005 across China's six broad regions. The table also distinguishes between the types of administrative border crossed by migrants: inter-prefectural, inter-provincial and inter-regional. Two obvious patterns emerge from Table A1: There is significant variation in terms of both emigration and immigration rates across regions but no region is left aside from the migration phenomenon.

Table A1. Descriptive statistics of migration flows by region.

	North	North-east	East	South Central	North-west	West
Emigration Rate (%), 2000:						
Within prov., out of pref.	0.15	0.28	0.32	0.42	0.31	0.20
Within region, out of prov.	0.25	0.16	0.64	0.82	0.14	0.15
Out of region	0.06	0.18	0.42	0.38	1.31	0.25
Emigration Rate (%), 2005:						
Within prov., out of pref.	0.40	0.67	0.97	1.05	0.65	0.49
Within region, out of prov.	0.51	0.69	1.34	2.04	0.26	0.27
Out of region	0.30	0.64	0.81	1.37	3.56	1.05
Immigration Rate (%), 2000:						
Within prov., out of pref.	0.37	0.32	0.99	1.47	1.37	0.65
Within region, out of prov.	0.61	0.19	1.97	2.89	0.64	0.49
Out of region	1.65	0.37	1.55	2.26	0.38	1.75
Immigration Rate (%), 2005:						
Within prov., out of pref.	0.97	0.77	2.97	3.67	2.92	1.54
Within region, out of prov.	1.25	0.80	4.09	7.17	1.15	0.85
Out of region	4.11	0.73	6.71	4.98	0.90	2.42
Destination Concentration:						
HHI, 2000	0.42	0.30	0.22	0.20	0.22	0.27
HHI, 2005	0.35	0.35	0.21	0.18	0.21	0.36

Notes: Migration flows are corrected for return migration and adjusted for coverage issues in the 2005 1% Population Survey. The top panel (Emigration Rates) displays yearly migration rates in 2000 and 2005 by region of origin. Rates are expressed as a share of the total rural population in the region in 2000. The middle panel (Immigration Rates) displays yearly migration rates in 2000 and 2005 by region of destination. Rates are expressed as a share of the total urban population in the region in 2000. The bottom panel (Destination Concentration) provides standardized Herfindahl-Hirschmann Indices (HHI) for destination concentration. Prefecture-level HHIs are averaged by region. The index ranges between 0 and 1; an index of 1 indicates that all migrants from a prefecture of origin move to a single prefecture of destination; 0 indicates perfect dispersion.

The bottom panel (Destination Concentration) of Table A1 provides further insights on the spatial patterns of migration. The panel displays prefecture-level Herfindahl-Hirschmann Indices (HHIs) of destination concentration that we then average by region. The HHIs are standardized to facilitate interpretation: An index of 1 means that all migrants from a prefecture of origin move to a single prefecture of destination; a value of 0 indicates perfect dispersion. As we can see from the table, regions differ markedly in terms of destination concentration. Nonetheless, the HHIs are everywhere under 0.42, which suggests that migrants from one region do not all flock to a single destination.

Table A2 displays the shares of rural-to-urban migrants, defined as as agricultural *hukou* holders who crossed a prefecture boundary and belong to working-age cohorts (15-64), in the total urban population of prefectures.

Table A2. Descriptive statistics from the 2005 mini-census.

Migrants share	Count	Share of urban population
Rural migrants	122,756	0.19
...from another province	94,326	0.15
Reason for moving	Count	Share of migrants
Work or business	100,670	82.01
Follow relatives	6,474	5.27
Marriage	5,783	4.71
Support from relatives/friends	4,461	3.63
Education and training	1,367	1.11
Expropriation and relocation	603	0.49
Job transfer	522	0.43
Mission	498	0.41
Recruitment	158	0.13
Deposit *hukou* demand	142	0.12
Other	1,956	1.59

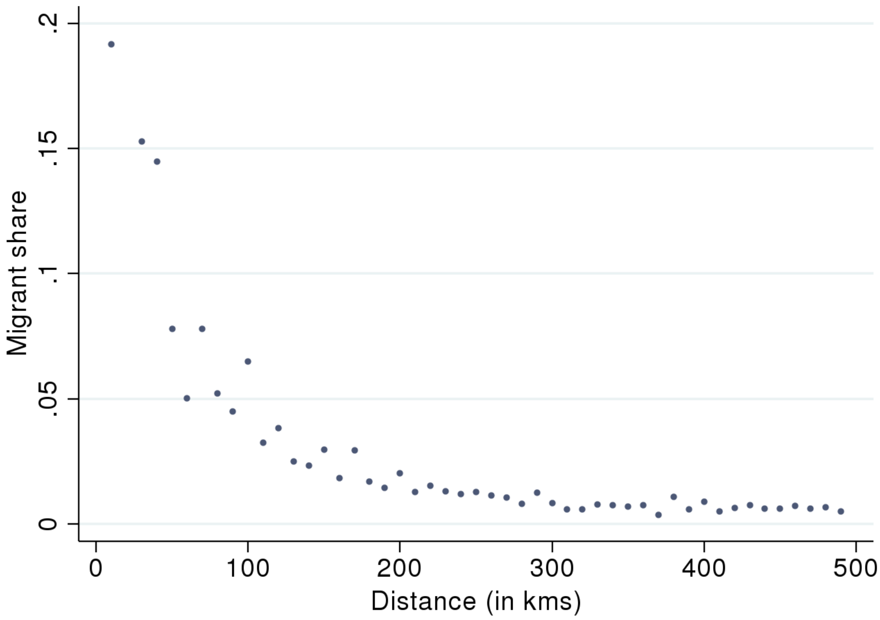
Notes: "Rural migrants" are defined as inter-prefectural migrants with an agricultural *hukou* aged 15-64. "Total resident urban population" refers to the population in the prefecture that is either locally registered and holds a non-agricultural *hukou* or resides in the prefecture but holds an agricultural *hukou* from another prefecture. The sample is restricted to inter-prefectural rural migrants.

The upper panel of Table A2 distinguishes between inter-prefectural migrants and those who left their provinces of origin. We see that inter-prefectural migrants represented 19% of a prefecture's total number of urban residents on average in 2005, while inter-provincial migrants accounted for 15% of it, which confirms that a majority (77%) of inter-prefectural migrations imply the crossing of a provincial boundary.

The high share of inter-provincial migrations among inter-prefectural moves already sheds light on the distance traveled by internal migrants in China. Figure A5 and Table A3 offer additional evidence on the impact of distance on migrants' destination choices. We see that there is a strong and significant inverse relationship between the share of migrants from origin o to destination d (among all migrants from o) and distance between o and d.

Selection of migrants We now provide some descriptive statistics on the profiles of internal migrants in China—in terms of education, demographics and labor market situation. In order to understand the effects of our shocks on outmigration and the

Figure A5. Origin-destination migration predictions—the role of distance.



Notes: Migration flows constructed with the 2000 Census and 2005 Mini-Census.

Table A3. Distance and migration flows between origins and destinations (2000-2005).

Migration flows (share)	(1)	(2)	(3)
Distance $d_{o,d}$ (1,000 km)	-0.0116***	-0.0449***	
	(0.000539)	(0.00286)	
Squared Distance $d^2_{o,d}$		1.04e-08***	
		(8.50e-10)	
Inverse Distance $1/d_{o,d}$			9.424***
			(0.757)
Destination population, 1990 $Pop_{d,1990}$	0.943***	0.956***	0.949***
	(0.0557)	(0.0552)	(0.0546)
Observations	116,622	116,622	116,622
R-squared	0.206	0.231	0.255
Origin FE	Yes	Yes	Yes

Robust standard errors are reported between parentheses. *** p<0.01, ** p<0.05, * p<0.1. The unit of observation is an origin×a destination×a year. Migration flows (share) are the number of migrants going from origin o to destination d normalized by the total number of migrants from origin o. For the sake of exposition, we normalize distance $d_{o,d}$ and destination population $Pop_{d,1990}$ by $1,000$.

impact of rural-to-urban migrants on the urban labor market and firms, it is useful to know the motives behind migration spells and describe the profile of rural migrant workers relative to both non-migrant rural dwellers and "native" urbanites.

The bottom panel of Table A2 presents the reasons put forward by inter-prefectural agricultural *hukou* migrants for leaving their places of registration. A vast majority (82%) moved away in order to seek work ("Work or business"), mostly as labourers, while all other rationales attracted much smaller shares.[31]

Migrants are not a representative sample of the Chinese population with an agricultural *hukou*. As can be seen from Table A4, migrants tend to be younger, more educated and more often single than the non-migrant rural population. They are also more likely to be self-employed or employees and to work in the private sector. Their total monthly income is more than twice larger. As expected, a majority (78%) of rural dwellers work in agriculture, as against 5% among migrants.

Urban workers differ significantly depending on their *hukou* status. As is usual with internal migration, we consider in the main specifications that migrants and locally registered urban *hukou* holders, or "natives," are highly substitutable. However, Chinese rural-to-urban migrants differ in a number of respects, which reduces their ability to compete with urbanites for the same jobs.

Table A4 provides summary statistics on key characteristics of inter-prefectural migrants and compares them with the locally registered urban population. Migrants and natives are significantly different on most accounts, the former being on average younger (and thus less experienced), less educated, more likely to be illiterate, and more often single, and employed without a labor contract. Important facts for the analysis are that rural-to-urban migrants are over-represented in privately owned enterprises and in manufacturing and construction industries: 91% of them are employed in the private sector as against 42% of locally registered non-agricultural *hukou* holders; and the share of rural-to-urban migrants working in manufacturing and construction is 51% and 9%, as against 20% and 4% for urban natives, respectively. Migrants also stand out as earning significantly less: Migrants' monthly income is 17% lower than urban natives'; the difference increases to about 40% when one takes into account the fact that migrants are attracted to prefectures where they can expect higher wages.[32]

Migrant selection into destinations based on economic prospects in those destination poses a serious threat to the identification of the effect of labor inflows on urban areas. Our empirical strategy relies on exogenous variation in agricultural prices at migrants' places of origin, which drive emigration decisions.

[31]The only other reasons that display shares in excess of 1% are "Education and training," "Other," "Live with/Seek refuge from relatives or friends," which Fan (2008) based on metadata from the Population Census Office dubs "Migration to seek the support of relatives or friends," "Following relatives," which should be understood as "Family members following the job transfer of cadres and workers", and "Marriage".

[32]Results available upon request.

Table A4. Migrant selection (2005 mini-census).

	Rural-urban migrants	Local *urban hukou*	Non-migrant *rural hukou*
Age	30.22	38.54	37.43
Female	0.49	0.49	0.51
Married	0.64	0.76	0.75
Education:			
Literate	0.97	0.99	0.91
Primary education	0.20	0.08	0.34
Lower secondary	0.60	0.33	0.47
Higher secondary	0.14	0.33	0.09
Tertiary education	0.02	0.24	0.01
Unemployed	0.02	0.09	0.01
Self-employed/Firm owner	0.20	0.16	0.14
Employee	0.77	0.81	0.11
Employee w/o labour contract	0.48	0.29	0.12
Public sector	0.11	0.72	0.21
Private sector	0.89	0.28	0.79
Total monthly income (RMB)	961.84	1157.07	408.64
Industry:			
Agriculture	0.05	0.06	0.78
Mining	0.01	0.03	0.01
Manufacturing	0.51	0.20	0.08
Utilities	0.00	0.03	0.00
Construction	0.09	0.04	0.03
Transportation	0.03	0.08	0.02
Information transfer, etc.	0.00	0.01	0.00
Wholesale and retail trade	0.15	0.14	0.04
Accommodation and catering	0.06	0.04	0.01
Finance	0.00	0.03	0.00
Real estate	0.01	0.01	0.00
Leasing and commercial services	0.01	0.02	0.00
Scientific research	0.00	0.01	0.00
Public facilities	0.00	0.01	0.00
Resident services	0.05	0.03	0.01
Education	0.00	0.10	0.00
Health care	0.00	0.04	0.00
Culture and entertainment	0.01	0.01	0.00
Public administration	0.00	0.11	0.01
International organisations	0.00	0.00	0.00
Observations	122,756	509,817	1,176,791

Notes: All variables except Age and Income are dummy-coded. Only the income of individuals who reported having a job is considered. The sample is restricted to individuals aged 15-64.

B Shocks to rural livelihoods

Our identification strategy relies on exogenous variation in agricultural livelihoods at migrants' places of origin. The empirical results presented in the paper use international prices, weighted by fixed prefecture-specific cropping patterns, to predict outflows of migrants from rural areas. The methodology is detailed in Section 3.

In this Appendix, we first illustrate our source of cross-sectional variation, i.e., the disparity in cropping patterns across Chinese prefectures. We then analyze our time-varying shocks, and we show that international prices vary substantially from one year to the next, as well as across crops, and that they translate into large fluctuations in domestic returns to agriculture. Finally, we generate similar shocks to rural livelihoods based on the interaction of rainfall and crop-specific growing cycles.

B.1 Crop suitability and use across Chinese prefectures

In order to assign crop-specific international price shocks to prefectures, we weight prices by the expected share of each crop in the prefecture's agricultural revenue. For this, we rely on potential yields and harvested areas in 2000. Yields are defined under different scenarios—low, intermediate and high inputs; rain-fed or irrigated. Harvested areas distinguish between rain-fed and irrigated land.

The top and middle panels of Table A5 show the variation in potential yields and harvested areas by crop and region. We focus on the four most important crops—rice, wheat, maize and soy—and on the high-input scenarios, which are better tailored to China and therefore the ones we use to compute the weights. It is obvious from the table that, as expected, some crops are concentrated in particular regions. This is especially true of rice, which is absent from the colder and drier northern regions. Nonetheless, it is also apparent that there is substantial regional variation in terms of crops and that no crop is cultivated in a single region, nor a region characterized by a single crop. A large part of the cross-sectional variation that we exploit does not come from regional differences, but from more local and granular disparities across prefectures.

An illustration of these regional differences is also provided in Figure 2 of the paper.

B.2 International price variations and domestic prices

Besides cross-sectional variation, our shocks to rural livelihoods rely on international commodity prices for temporal variation.

Table A5. Descriptive statistics of price shocks, potential yields and harvested areas by region.

	North	North-east	East	South Central	North-west	West
Potential Yields:						
Rice, rain-fed	0.000	0.000	4010.1	3656.1	2807.6	0.000
Rice, irrigated	3011.6	2669.2	342.5	798.8	3327.1	3363.1
Wheat, rain-fed	1441.7	1368.7	1643.6	1724.6	2326.2	3363.3
Wheat, irrigated	1477.0	1210.2	1538.1	1965.6	2480.1	3390.6
Maize, rain-fed	3149.2	2298.8	1243.6	1430.0	3651.0	3641.6
Maize, irrigated	2852.7	2654.1	1491.4	1900.6	3635.2	4386.6
Soy, rain-fed	1299.2	1080.8	223.0	304.7	1579.7	1533.8
Soy, irrigated	1004.2	1255.8	368.6	463.9	1587.9	1804.3
Harvested Areas:						
Rice, rain-fed	0.000	0.001	0.026	0.041	0.023	0.000
Rice, irrigated	0.119	0.432	0.935	0.715	0.474	0.083
Wheat, rain-fed	0.066	0.016	0.173	0.139	0.141	0.081
Wheat, irrigated	0.706	0.038	0.696	0.789	0.257	0.332
Maize, rain-fed	0.126	0.375	0.208	0.180	0.287	0.094
Maize, irrigated	0.428	0.215	0.317	0.281	0.062	0.160
Soy, rain-fed	0.045	0.094	0.113	0.061	0.086	0.035
Soy, irrigated	0.071	0.028	0.064	0.038	0.015	0.025
Price Shocks:						
Between variation	0.037	0.024	0.036	0.039	0.040	0.047
Within variation	0.008	0.019	0.006	0.011	0.002	0.038

Notes: This table displays between- and within-region standard deviation in the prefecture-level price shock, and between-region variation in potential yield and harvested area for the main crops under irrigated and rain-fed agriculture. Between variation is measured in 2000. Potential yield is expressed in kg/ha and corresponds to the high-input scenario. Harvested area (ha) refers to the normalized agricultural land devoted to a crop.

One important assumption behind our empirical strategy is that China's agricultural prices are not insulated from world market fluctuations. Table A6 confirms that international price variations do translate into price fluctuations in the Chinese domestic market. The first column provides the correlation between Chinese domestic prices for different crops in different years and international prices. We find that a 10% increase in the latter yields a 4% hike in the former, which constitutes a substantial pass-through from the international to the domestic market. The second column looks at the logarithm of output as the dependent variable and explains it by international and domestic prices. We can see that both prices are positively associated to crop production over the period of interest. While output and local prices are both determined by local demand and supply, international prices better explain the variation in local output that local prices. One explanation could be that local demand and local supply have opposite effects on the comovement of output and prices, while international price shocks should be perceived as a pure demand shock from the viewpoint of Chinese producers.

Table A6. Correlation between crop international prices and local Chinese prices/production.

VARIABLES	Prices	Output
Price (International)	.402***	.201**
	(.0861)	(0.062)
Price (China)		.0824*
		(.0432)
Observations	210	210
R-squared	.579	.337
Trend	Yes	Yes
Crop FEs	Yes	Yes

Robust standard errors are reported between parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The unit of observation is a crop × year. The two regressions include a time trend and crop fixed effects, and are weighted by the average crop production (in tons) over the period 1991-2010. All variables are in logs. Standard errors are clustered at the crop level.

Another important assumption is that there are relevant short-term fluctuations in international crop prices. Figure A6 plots the evolution of international prices for a selection of crops.[33] We interpret these short-term fluctuations as random shocks on the international market due to fluctuations in World supply and demand for each crop. As can be seen from Figure A6, the HP-filtered price series resemble AR(1) processes with jumps. There are large swings followed by a gradual return to the mean. Importantly, all crops display such large fluctuations over time, and the fluctuations may not coincide across crops.

B.3 Shocks over time and across regions

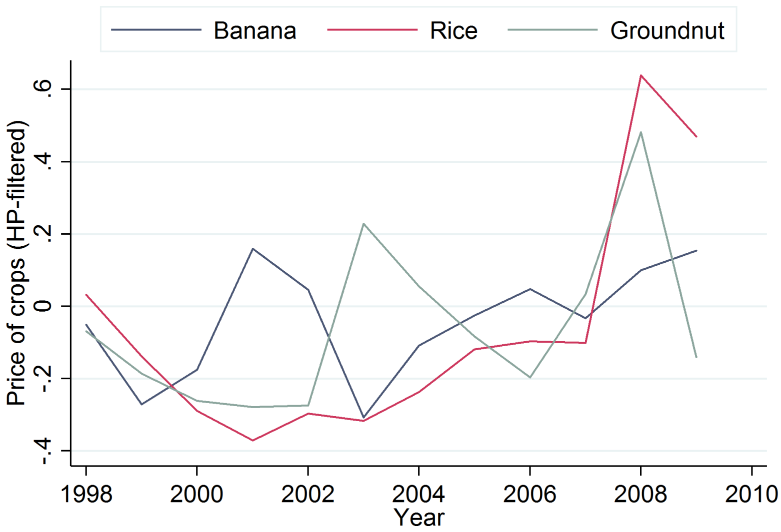
The price shocks, i.e., the HP-filtered international prices weighted by potential output at the prefecture level, exhibit variation both across space and over time.

The bottom panel of Table A5 provides between- and within-region variation in the price shock for China's six major regions. Between variation is measured in 2000. Reassuringly for our identification strategy, we can see that all regions display fluctuations in the price shocks, both across prefectures and over time. No region stands out as being particularly subject to such shocks or immune to them either.

Evidence from Table A5 can be illustrated by some maps. Figure A7 displays the price shocks in 2001 (left panel) and 2004 (right panel). Despite some spatial

[33] As in the main empirical specification, we have filtered out the long-term trend component Hodrick and Prescott (1997).

Figure A6. Price deviations from trends on International Commodity Markets 1998-2010.



Note: These series represent the Hodrick-Prescott residual applied to the logarithm of international commodity prices for three commodities: banana, rice and groundnut. For instance, the price of rice can be interpreted as being 35% below its long-term value in 2001.

correlation due to the underlying cropping patterns, we see that there is substantial variation across prefectures (delimited by dark lines) and that the picture also changes noticeably from one year to the other.

These cross-sectional and time variations carry over from the price shocks to the predicted flows of immigrants. Figure A8 represents immigration rates at the prefecture level in 2001 (left panel) and 2004 (right panel), as predicted by agricultural price shocks in migrants' prefectures of origin. Here again, we see that predicted immigration varies widely both across prefectures and over time.

B.4 An additional source of variation: rainfall shocks

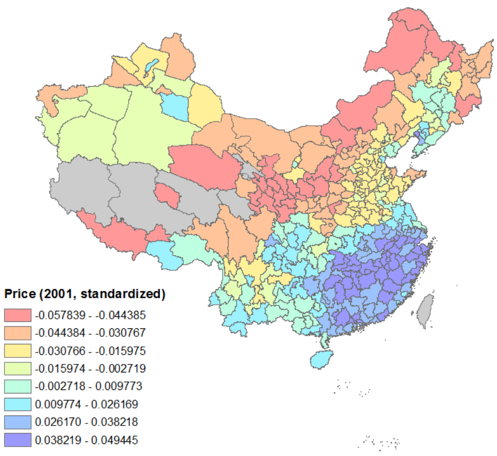
As a robustness check, we construct a second type of shocks to agricultural income based on rainfall deficit during the growing period of each crop.

The monthly precipitation measure (0.5 degree latitude $\times$ 0.5 degree longitude precision) covers the period 1901-2011 and mostly relies on the Global Historical Climatology Network.[34] Once collapsed at the prefecture level, this provides us with a measure ra_{omt} of rainfall for prefecture o in month m and year t.

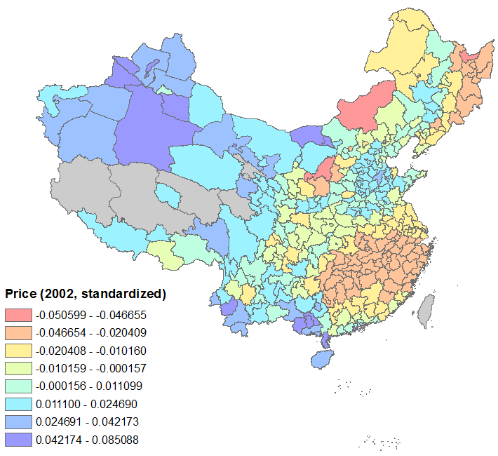
We refine this rainfall measure to account for the growing cycle of each crop, i.e.,

[34] UDel_AirT_Precip data was provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.esrl.noaa.gov/psd/.

Figure A7. Shocks to rural livelihoods across Chinese prefectures in 2001 and 2004.



(a) 2001



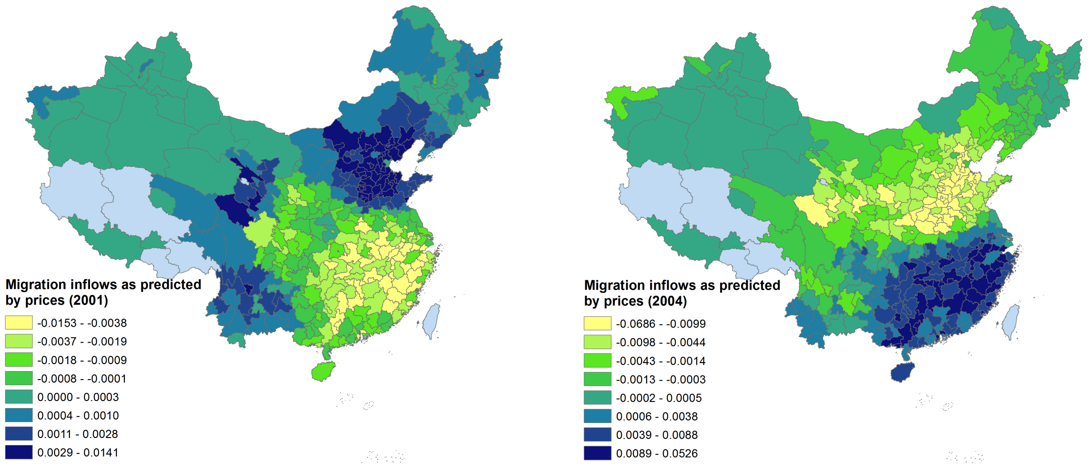
(b) 2004

Notes: These two maps represent the standardized *price shock* p_{ot} in 2001 (left panel), and 2002 (right panel). Note that, in 2001, the price of rice decreases which generates a very negative shock across China concentrated in rice-producing prefectures.

(i) the harvest season and (ii) the crop-specific rainfall requirements. For a given year, there are several sources of variation across Chinese prefectures in actual yields due to rainfall. First, different locations receive different levels of rainfall. Second, exposure to rainfall depends on the growing cycle of the different harvested crops (winter, spring or summer/fall crops). In addition, some crops are resistant to large water deficits while others immediately perish with low rainfall. The large cross-sectional variation in each year may come from (i) a direct effect of local rainfall, (ii) an indirect effect coming from the interaction with the crop-specific growing cycle and the variety of crops grown across China.

We rely on the measure ra_{omt} of rainfall for prefecture o in month m and year t and we construct for each crop a measure wr_c of the minimum crop-specific water requirement during the growing season M_c as predicted by the yield response to

Figure A8. Measure $\widehat{M_{dt}}$ of immigrant inflows to cities as predicted by prices in 2001 and 2004.



(a) 2001 (b) 2004

Notes: These two maps represent the quantities $\widehat{M_{d,2001}}$ and $\widehat{M_{d,2004}}$, where $\widehat{M_{dt}}$ is the measure of immigrant inflows as predicted by price variations and the weighting distance matrix between origins and destinations.

water.[35] We then generate

$$r_{ot} = \left(\sum_c \left(\frac{\max\{\sum_{m \in M_c} wr_c - ra_{omt}, 0\}}{wr_c} \right)^{\alpha} h_{co} yi_{co} \bar{p}_c \right) / \left(\sum_c h_{co} yi_{co} \bar{p}_c \right). \quad (13)$$

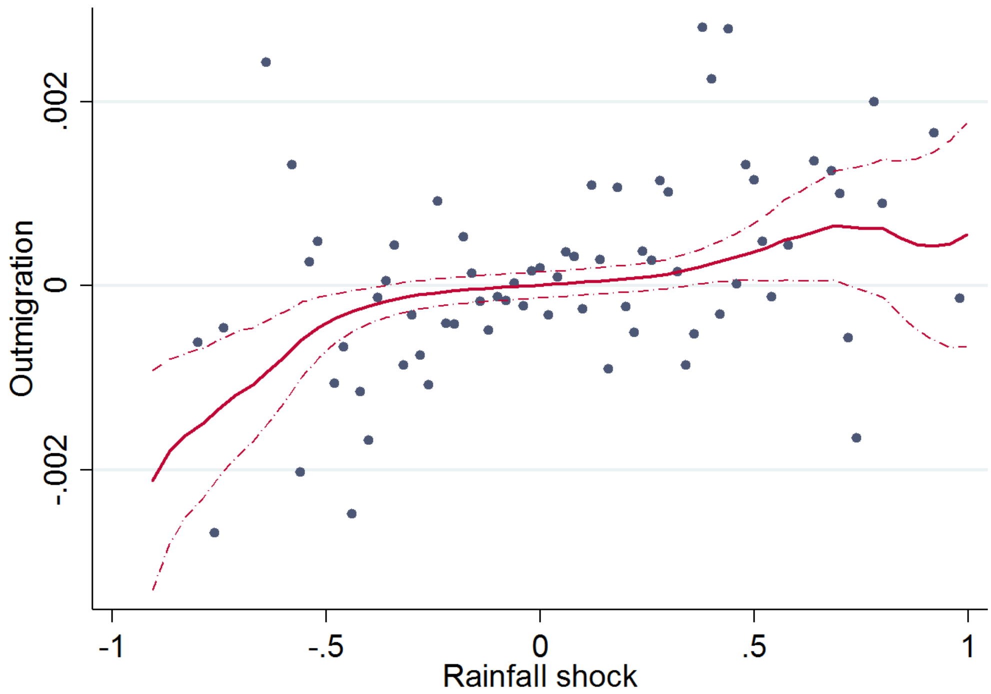
This measure has a very intuitive interpretation. The quantity $\max\{\sum_{m \in M_c} wr_c - ra_{omt}, 0\}$ is the deficit between actual rainfall and the minimum crop water requirement wr_c during the growing season. We then penalize this deficit with a factor α capturing potential non-linearities in the impact of rainfall deficit. In our baseline specification, this penalization parameter α is set equal to 3.[36] A high ratio $\max\{\sum_{m \in M_c} wr_c - ra_{omt}, 0\}/wr_c$ would be associated with a bad harvest for the corresponding crop. We then weight these ratios by potential output for each crop in each prefecture.

There is large year-to-year variation in rainfall deficits. Also, for a given year, because of differences in cropping patterns across prefectures, the spatial autocorrelation of rainfall shocks is much lower than the correlation of rainfall itself. While the exogeneity of rainfall shocks is not questionable, we still need to assume that urban labor demand is not directly affected by rainfall in order to use it as instrument for rural to urban migration.[37]

[35] http://www.fao.org/nr/water/cropinfo.html
[36] The results are robust to more conservative values for α, e.g., $\alpha = 1$ or $\alpha = 2$.
[37] In the robustness checks, we show that our results hold when controlling for local rainfall

The relationship between rainfall shocks and rural emigration is illustrated in Figure A9. Remember that a high value of the rainfall shock is associated with severe water deficit during the growing seasons of locally grown crops. We see from Figure A9 that low rainfall in a given year pushes rural dwellers out of their prefectures of origin.

Figure A9. Rainfall deficits relative to water requirements at origin and outmigration rates.



Notes: This Figure illustrates the relationship between the standardized rainfall deficit relative to water requirements for the origin-specific agricultural portfolio (x-axis) and outmigration (y-axis). We consider the residuals of all measures once cleaned by prefecture and year fixed effects. For the sake of exposure, we group prefecture×year observations, create 100 bins of observations with similar rainfall shock and represent the average outmigration rate within a bin. The lines are locally weighted regressions on all observations.

shocks.

C Data description

C.1 Firm data

We present here in greater detail our firm survey data. We first summarize the main characteristics of the data, along with some descriptive statistics, and then discuss some possible issues and how we tackle them.[38]

Description The firm data that we rely on for the better part of the empirics come from the National Bureau of Statistics (NBS). The NBS implements every year a census of all state-owned manufacturing enterprises and all non-state manufacturing firms with sales exceeding 5 million RMB, or about $600,000 over that period.[39] This threshold gives the data their common name of "above-scale" manufacturing firm surveys (*"xian'e"* or *"guimo yishang" gongye qiye diaocha*), despite the fact that the data constitute a census of SOEs irrespective of their size.

The data that we use cover the period 1998-2007. Data for 1992, 1993, 1995, 1997 and 2008-2009 are also available but sometimes offer a markedly different set of variables and cannot easily be used to create a panel of firms—see below. Our data cover the manufacturing sector—Chinese Industrial Classification (CIC) codes 1311-4392.

Although we shall use the terms "firm" and "enterprise" interchangeably in the remainder of the paper, the NBS data cover "legal units" (*faren danwei*). This implies that different subsidiaries of the same enterprise may be surveyed, provided they meet a number of criteria, including having their own names, being able to sign contracts, possessing and using assets independently, assuming their liabilities and being financially independent. While this definition of units of observation may be unfamiliar to readers accustomed to U.S. or European data, the concept of "legal units" almost perfectly overlaps with that of establishments in practice: In 2007, almost 97% of the units in our data corresponded to single plants.

For that reason, we restrict ourselves to the balanced panel of firms over the period in most of our analysis, and we only study entrants and exiters separately. In contrast with firm data in some developed countries, matching firms over time in the NBS is difficult because of frequent changes in unique firm identifiers. In order to match "identifier-switchers," we use the fuzzy algorithm developed by Brandt et al. (2014), which uses slowly-changing firm characteristics such as its name, address

[38]Please refer to Brandt et al. (2014) for an exhaustive treatment. This section partly summarizes the challenges that they highlight.

[39]The average exchange rate over the period of interest was 8.26 RMB to the USD, so 5 million RMB represents about $605,000. Note that the threshold was raised to 20 million RMB in 2011.

or phone number. While total sample size ranges between 150,000 and 300,000 per year, we end up with 70,000 firms when we limit the sample to the balanced panel.

The "above-scale" firm data contain a wealth of information on large manufacturing firms. Besides each firm's location, industry, ownership type, exporting activity and number of employees, they offer a wide range of accounting variables (e.g., output, input, value added, wage bill, fixed assets, financial assets, etc.). These are the data we rely on to construct the firm-level measures of factor choices and costs, as well as the measures of productivity that we constructed for the empirics.

Table A7 displays key descriptive statistics across public, domestic private and foreign private firm ownership over the period 1998-2007.[40] Public enterprises, a broad category that encompasses state-owned and collective enterprises, have a larger capital stock, spend more on their wage bills and have more employees than domestic private firms. Conversely, the latter report significantly higher sales revenues and perform better in terms of value added. Table A7 yields a very different image of state-owned and collective enterprises when compared to the foreign private sector: Real capital stock, sales revenues, value added and the total wage bill are all higher in foreign-owned firms; only the total number of employees is higher in the public sector.

Table A7. Descriptive statistics from the NBS firm-level data.

	Public sector	Domestic private sector	Foreign private sector
Real capital stock	37,539.7	20,346.0	47,592.4
Sales revenue	63,149.1	71,267.7	167,520.8
Value added	18,470.8	17,106.1	40,216.0
Total wage bill	3,695.91	2,938.08	6,613.63
Total number of employees	340.20	216.93	318.76

All variables except "Total number of employees" are in RMB 1,000. The table displays averages over the period 1998-2007.

In terms of time patterns, private firms still accounted for a relatively small share of total real capital stock, value added, sales revenues, wage bill and employment in 1998 but represented over 80% of the total under all five indicators by 2007. The evolution in terms of employment is particularly striking: Whereas only 32% of total employment could be attributed to private firms in the NBS firm sample in 1998, they accounted for 89% of it in 2007.

Possible issues There are a number of issues with using the NBS firm data. We now discuss these issues and explain how we take them into account.

[40]Ownership type is defined based on official registration (*qiye dengji zhuce leixing*). Out of 23 exhaustive categories, Table A7 uses three categories: (i) state-owned, hybrid or collective, (ii) domestic private, and (iii) foreign private firms, including those from Hong Kong, Macau, and Taiwan.

First, the 5 million RMB threshold that defines whether a non-publicly owned firm belongs to the NBS census was not perfectly implemented. It is indeed impossible to know the exact level of sales before implementing the survey and some firms only entered the database several years after having reached the sales cut-off.[41] We can however show that this is unlikely to be a serious issue since the threshold is quite sharp, as can be seen from Figure 3. As firms that are below the threshold represent but a small share of the total sample, dropping them does not affect the results.

Second, the truncation due to sample restrictions on private and collective firms potentially introduces a selection bias. While the NBS data offer a *census* of SOEs, the sample tends to over-represent more productive private firms that report high sales given their number of employees. This concern about representativeness should however be alleviated by the fact that our firms account for 90% of total gross output in the manufacturing sector and 70% of the industrial workforce.

Third, although each "legal unit" in each year contains a unique identifier, changes were introduced over time so that identifiers need not be consistent from one year to the next. Matching firms over time in the NBS data is therefore a challenge. In order to match "identifier-switchers," we extend the fuzzy algorithm (using firm identifier, firm name, name of the legal representative, address, phone number, founding year and main products) developed by Brandt et al. (2014). While total sample size ranges between 150,000 and 300,000 per year, we end up with 45,000 firms when we limit the sample to the fully balanced panel between 1998 and 2007.

Fourth, firms may have an incentive to under-report the number of workers as it serves as the basis for taxation by the local labor department. This should be of particular concern with migrants, who represent a large share of the workforce and may be easier to under-report. Along the same lines, workers hired through a "labor dispatching" (*laodong paiqian*) company are not included in the employment variable. This implies that migrant workers might be under-counted in the firm data. This is why we provide additional evidence on the effect of immigration on urban labor markets thanks to the Urban Household Survey data described below. Wage bill may also be slightly under-estimated as some components of worker compensation are not recorded in all years, e.g., pension contributions and housing subsidies, which are reported only since 2003 and 2004, respectively but accounted for only 3.5% of total worker compensation in 2007.

Fifth, some variables are not documented in the same way as in standard firm-

[41] Conversely, about 5% of private and collectively-owned firms, which are subject to the threshold, continue to participate in the survey even if their annual sales fall short of the threshold.

level datasets. In particular, fixed assets are reported in each data wave by summing nominal values for different years. We use the procedure developed in Brandt et al. (2014): (i) We calculate the nominal rate of growth in the capital stock (using a 2-digit industry by province average between 1993 and 1998) to compute nominal capital stock in the firm's start-up year. (ii) Real capital in the start-up year is obtained thanks to the chain-linked investment deflator developed by Perkins and Rawski (2008) (based on separate price indices for equipment-machinery and buildings-structures, and weighted by fixed investment shares provided by the NBS). (iii) We move forward to the first year in the database, assuming a rate of depreciation of 9% per year and using annual deflators. (iv) Once a firm enters the database, we can use the nominal figures provided in the data to compute the change in nominal capital stock in a given year, and deflate it. If the firm's past investments and depreciation are not available in the data, we use information on the age of the firm and estimates of the average growth rate of nominal capital stock at the 2–digit industry level between 1993 and the firm's year of entry in the database.

C.2 UHS data

Description In order to study labor market outcomes from the workers' point of view, we use the national Urban Household Survey (UHS) collected by the National Bureau of Statistics. The UHS is a survey of Urban China, with a consistent questionnaire since 1986 but considered representative from 2002 onward, and our description will correspond to this latter period. The survey is based on a three-stage stratified random sampling. Its design is similar to that of the Current Population Survey in the United States (Ge and Yang, 2014; Feng et al., 2015b) and includes 18 provinces and 207 prefectures.[42] The data we use for our analysis are annual cross-sections, with a sample size that ranges from 68,376 in 2002 to 94,428 individuals in 2008. Our analysis will be restricted to the locally registered urban population.[43]

The UHS is a very rich dataset with detailed information on individual employment, income—including monthly wages, bonuses, allowances, housing and medical subsidies, overtime, and other income from the work unit—and household-level characteristics—see Feng et al. (2015b) for a comprehensive description of the survey. Our measure of real wages relies on monthly wages divided by a prefecture- and

[42]The provinces are Beijing, Shanxi, Liaoning, Heilongjiang, Shanghai, Jiangsu, Zhejiang, Anhui, Jiangxi, Shandong, Henan, Hubei, Guangdong, Chongqing, Sichuan, Yunnan, Shaanxi and Gansu.

[43]While all households living in urban areas are eligible, sampling still ignores urban dwellers living in townships and in the suburban districts of Beijing, Chongqing, Shanghai, and Tianjin (Park, 2008). Rural-urban migrants, who are more likely to live in peripheral areas of cities, are therefore under-represented.

year-specific consumer price index, which we computed using the detailed household-level consumption data available in the UHS. We also construct three employment outcomes: wage employment, unemployment and self-employment (which also includes firm owners).[44] Appendix Table A8 provides some descriptive statistics of key variables over the period 2002-2008.

Empirical strategy Let y_{jdt} be the labor market outcome of worker j in urban destination d in year t. We regress y_{jdt} on predicted migration that year, $\widetilde{M_{dt}}$, and a vector of individual characteristics $\mathbf{X}_j$. The vector $\mathbf{X}_j$ includes dummy variables for individual j's marital status, gender, education level (primary, lower secondary, upper secondary and tertiary), and age. We also include year/occupation dummies in order to control for workers' skills, and sectoral specific fluctuations in labor costs.[45] In order to control for labor market conditions at destination and aggregate fluctuations in labor market outcomes, we also include destination fixed effects and year fixed effects. The effect of M_{dt} on y_{jdt} is estimated through Two-Stage Least Squares (2SLS), using $\widetilde{M_{dt}}$ as an instrument:

$$\begin{cases} M_{dt} = b_0 + b_m \widetilde{M_{dt}} + b_x \mathbf{X}_j + c_d + n_t + e_{dt} \\ \\ y_{jdt} = \beta_0 + \beta_m M_{dt} + \delta \mathbf{X}_j + \gamma_d + \nu_t + \varepsilon_{jdt} \end{cases}, \tag{14}$$

and standard errors are clustered at the level of the prefecture of destination. Since unskilled urban residents are more likely to be competing for jobs with migrant workers, they may experience larger changes in wages and occupation as a response to migration inflows. In order to test this, we estimate the same specification interacting the migration shock with a dummy $LowSkill_j$ equal to 1 if the worker has lower secondary education or less, and 0 otherwise.

[44]Working hours in the month preceding the survey were also recorded in UHS 2002-2006. However, as pointed out by Ge and Yang (2014), they vary within a very narrow range, which means that the UHS measure might understate actual variations in working hours. For this reason, we do not use hours of work in our analysis.

[45]UHS occupation categories are "Head of organization," "Professional skill worker," "Staff," "Commercial and service worker," "Agriculture," "Production operator," "Soldier" and "Other occupations." Since occupation itself may be an outcome of immigration, we check that our results are robust to excluding it from the vector of controls.

Table A8. Descriptive statistics from the UHS data (2002-2008).

	Mean	St. Dev.
Age	43.17	11.00
Female	0.50	0.50
Married	0.88	0.33
Born in prefecture of residence	0.61	0.49
Education:		
Primary education	0.05	0.21
Lower secondary	0.27	0.45
Higher secondary	0.25	0.43
Tertiary education	0.42	0.49
Unemployed	0.02	0.14
Self-employed/Firm owner	0.05	0.23
Employee	0.71	0.45
Public sector	0.63	0.48
Private sector	0.37	0.48
Total monthly income (RMB)	1537.52	1416.81
Monthly wage income (RMB)	1353.36	1264.84
Monthly transfer income (RMB)	56.71	287.76
Industry:		
Agriculture	0.01	0.10
Mining	0.02	0.14
Manufacturing	0.22	0.42
Utilities	0.03	0.18
Construction	0.03	0.17
Transportation	0.06	0.24
Information transfer, etc.	0.04	0.18
Wholesale and retail trade	0.12	0.33
Accommodation and catering	0.03	0.16
Finance	0.02	0.15
Real estate	0.04	0.19
Leasing and commercial services	0.02	0.15
Scientific research	0.03	0.18
Public facilities	0.01	0.11
Resident services	0.10	0.30
Education	0.06	0.23
Health care	0.03	0.18
Culture and entertainment	0.01	0.11
Public administration	0.10	0.30
Obs.		
2002	54,564	
2003	62,194	
2004	65,806	
2005	77,976	
2006	70,853	
2007	75,539	
2008	76,874	

All variables except Age and Income are dummy-coded. The table displays averages over the period 2002-2008. The sample is restricted to locally registered urban *hukou* holders aged 15-64.

D Robustness checks and sensitivity analysis

In Appendix D, we provide some checks of the robustness of our results. We focus first on the solidity of the effect of agricultural shocks on rural outmigration and then proceed to relaxing the identification assumptions that underpin the use of 2SLS.

D.1 Shocks to rural livelihoods

First, we investigate whether rural outmigration reacts in a similar and consistent manner to another type of agricultural shock. We compare the effect of prices with a rainfall deficit index based on precipitation during the growing period of the crops that are cultivated locally. Rainfall shocks are constructed as described in Appendix B.

The results presented in Table A9 show that rainfall shocks are also strong predictors of rural outmigration. As expected, a more severe rainfall *deficit* reduces farmers' expected output and leads to more outmigration. This effect is consistent with that of price shocks and thus reinforces our interpretation of fluctuations in international agricultural commodity prices as shocks to peasants' livelihoods. Table A9 further shows that prices and rainfall constitute two independent sources of variation in rural outmigration, as can be seen from their independent effects in column 3.

Table A9. Comparison of actual and predicted immigration rate in urban areas (Census, 2000-2005).

	Outmigration rate		
	(1)	(2)	(3)
Rainfall Shock (standardized)	0.0488**		0.0617***
	(0.019)		(0.020)
Price Shock (standardized)		-0.104***	-0.116***
		(0.019)	(0.018)
Observations	1,690	1,690	1,690
R-squared	0.861	0.864	0.867
Year FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes

Notes: Standard errors are clustered at the prefecture level and are reported between parentheses. ***: p<0.01, **: p<0.05, *: p<0.1. An observation is a prefecture *times* year. The outcome variable is the number of rural out-migrants to urban areas divided by the number of rural residents.

Price fluctuations and rainfall deficits could not be used as instruments for migration flows if they were foreseeable and farmers had the time to protect their revenues, thus potentially reintroducing endogeneity in migration flows and jeopardizing identification. The construction of our shock variables is designed to alleviate this concern. We nevertheless check that rural dwellers indeed do not anticipate adverse changes in their revenues by emigrating before the realization of a price shock or rainfall deficit. Table A10 shows that contrary to contemporaneous shocks, average residual agricultural income in $t+1$ and $t+2$ has no impact on outmigration. The coefficients are small and not statistically significant.

Table A10. Predicting Outmigration – Forward Shocks

	Outmigration rate (1)
Price Shock (standardized)	.003 (.006)
Rainfall Shock (standardized)	.007 (.005)
Observations	1,690
Year Fixed-Effects	Yes
Origin Fixed-Effects	Yes

Notes: Standard errors are clustered at the prefecture level and are reported between parentheses. ***: p<0.01, **: p<0.05, *: p<0.1. The sample is all prefectures every year and the outcome variable is the number of rural out-migrants to urban areas in year t divided by the number of rural residents. "Price Shock" ("Rainfall Shock") is the average of forward shocks in $t+1$ and $t+2$.

D.2 Main specification

The results on the effect of immigration on firms in urban areas can be interpreted as causal only if the instrument satisfies the exclusion restriction, i.e., if agricultural revenue shocks have no impact on firm outcomes other than through the influx of rural workers that they trigger.

Tables A11 and A12 control for channels through which price shocks might have a direct effect on firms and thus imperil the causal interpretation of the effects reported in Tables 2 and 3, respectively. Panel A reproduces the baseline results of Tables 2 and 3. Panels B to E present the results of four robustness checks.

A first cause of concern is that urban firms may use agricultural commodities as inputs, or more generally be directly affected by agricultural prices in migrants' places of origin. We test this potential confound in two ways. In Panel B, we focus on migrants who crossed a provincial border instead of inter-prefecture migrants as in Section 4. Provinces are the second level of government, immediately below the central government, and thus constitute much larger geographical entities and distinct markets. We see that results remain virtually unchanged when we focus on such far-flung migrants and thus exclude those originating from areas that are likely to supply destination firms or workers in agricultural products. Next, as shown in Panel C, the IV estimates remain large and significant when we explicitly exclude industries that use agricultural commodities in their production processes, which suggests that our results are not driven by the direct effect of agricultural price shocks on manufacturing units.

Another concern is that our predicted migrant flows, which are constructed using distance and destination population, might capture market access, which may imply different firm dynamics. We test this by controlling for the log of the destination population interacted with a time trend to allow larger destinations to evolve differently. Our estimates do not change when we allow for such differential trends (Panel D).

Our estimation may also be capturing different shocks to industrial sectors, which may be correlated with migration through the geographical distribution of manufacturing activities or the diffusion of agricultural price shocks. However, all results go through when we control flexibly for industry-specific year fixed effects (Panel E), which suggests that we are not simply capturing urban dynamics linked to sectoral specialization or market power.

Finally, in Panel F we perform a standard placebo check and test whether future migration shocks have any effect on firm outcomes. We find that future shocks have no effect on current firm outcomes, which further alleviates concerns that our estimates are driven by trends unrelated to migration.

Table A11. Impact of migration inflows on urban firms – robustness checks (1/2).

	Labor cost		Employment		Capital to labor	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Main Results						
Migration	-0.448*	-1.276***	0.721***	1.095***	-0.383	-1.582***
	(0.232)	(0.368)	(0.230)	(0.411)	(0.238)	(0.491)
Obs.	353,133	353,133	354,453	354,453	353,538	353,538
Panel B: Extra-Provincial Migration						
Migration	-0.587**	-1.681***	0.830***	1.467***	-0.458*	-1.957***
	(0.246)	(0.484)	(0.262)	(0.507)	(0.270)	(0.529)
Obs.	353,133	353,133	354,453	354,453	353,538	353,538
Panel C: Excluding Processing Industries						
Migration	-0.461*	-1.235***	0.744***	1.202***	-0.380	-1.577***
	(0.239)	(0.376)	(0.239)	(0.432)	(0.248)	(0.519)
Obs.	317,401	317,401	318,488	318,488	317,694	317,694
Panel D: Controlling for log(Population) $\times$ Year						
Migration	-0.449*	-1.271***	0.720***	1.092***	-0.379	-1.591***
	(0.228)	(0.356)	(0.226)	(0.389)	(0.234)	(0.487)
Obs.	353,133	353,133	354,453	354,453	353,538	353,538
Panel E: Industry $\times$ Year Fixed Effects						
Migration	0.436*	-1.297***	0.694***	1.008**	-0.363	-1.605***
	(0.235)	(0.384)	(0.233)	(0.415)	(0.241)	(0.500)
Obs.	352,795	352,795	354,112	354,112	353,197	353,197
Panel F: Forward Shocks						
Migration	-0.142	0.589	0.101	-0.328	-0.023	0.538
in $t+1$	(0.095)	(0.393)	(0.090)	(0.368)	(0.089)	(0.362)
Obs.	308,414	308,414	309,665	309,665	308,829	308,829

Notes: Standard errors are clustered at the prefecture level and are reported between parentheses. ***: p<0.01, **: p<0.05, *: p<0.1. The sample is composed of the 44,981 firms present every year in the NBS firm census between 1998 and 2006. All specifications include firm and year fixed effects. In the IV estimation, the instrument is the migration rate predicted using price shocks at origin, distance between origin and destination, and destination population. *Migration* is the immigration rate, i.e., the migration flow over population in 2000. *Labor cost* is the logarithm of the compensation per worker including social security. *Employment* is the logarithm of the number of workers within the firm. *Capital to labor* is the logarithm of the ratio of employment to fixed assets (evaluated at their current prices).

Table A12. Impact of migration inflows on urban firms – robustness checks (2/2).

	Return to labor		Return to capital		TFP	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Main Results						
Migration	-0.31***	-1.01***	-0.10***	-0.34***	-0.30***	-0.84***
	(0.10)	(0.22)	(0.04)	(0.12)	(0.09)	(0.24)
Obs.	305,055	305,055	304,689	304,689	304,689	304,689
Panel B: Extra-Provincial Migration						
Migration	-0.40***	-1.20***	-0.16***	-0.38***	-0.43***	-0.90***
	(0.11)	(0.27)	(0.04)	(0.13)	(0.09)	(0.28)
Obs.	305,055	305,055	304,689	304,689	304,689	304,689
Panel C: Excluding Processing Industries						
Migration	-0.30***	-1.00***	-0.12***	-0.40***	-0.32***	-0.90***
	(0.11)	(0.22)	(0.04)	(0.11)	(0.09)	(0.24)
Obs.	275,250	275,250	274,923	274,923	274,923	274,923
Panel D: Controlling for log(Population) $\times$ Year						
Migration	-0.31***	-1.02***	-0.10***	-0.34***	-0.30***	-0.84***
	(0.10)	(0.22)	(0.04)	(0.12)	(0.09)	(0.24)
Obs.	305,055	305,055	304,689	304,689	304,689	304,689
Panel E: Industry $\times$ Year Fixed Effects						
Migration	-0.26**	-0.91***	-0.08**	-0.29**	-0.24***	-0.64**
	(0.11)	(0.21)	(0.04)	(0.12)	(0.08)	(0.25)
Obs.	305,055	305,055	304,689	304,689	304,689	304,689
Panel F: Forward Shocks						
Migration	-0.024	0.336*	-0.009	0.004	-0.045	0.016
in $t+1$	(0.048)	(0.203)	(0.024)	(0.052)	(0.054)	(0.101)
Obs.	265,820	265,820	265,505	265,505	265,505	265,505

Notes: Standard errors are clustered at the prefecture level and are reported between parentheses. ***: p<0.01, **: p<0.05, *: p<0.1. The sample is composed of the 44,981 firms present every year in the NBS firm census between 1998 and 2006. All specifications include firm and year fixed effects. In the IV estimation, the instrument is the migration rate predicted using price shocks at origin, distance between origin and destination, and destination population. *Migration* is the immigration rate, i.e., the migration flow over population in 2000. *Return to labor* is the logarithm of the marginal revenue product of labor as defined in Section 3. *Return to capital* is the logarithm of the marginal revenue product of capital as defined in Section 3. *TFP* is the logarithm of the total factor productivity in revenue terms as defined in Section 3.