Migrants and Firms: Evidence from China

Clement Imbert Marlon Seror Yifan Zhang
Yanos Zylberberg*

September 3, 2021

Abstract

How does rural-urban migration shape urban production in developing countries? We use longitudinal data on Chinese manufacturing firms between 2000 and 2006, and exploit exogenous variation in rural-urban migration induced by agricultural income shocks for identification. We find that, when immigration increases, manufacturing production becomes more labor-intensive and productivity declines. We investigate the reorganization of production using patent applications and product information. We show that rural-urban migration induces both labor-oriented technological change and the adoption of labor intensive product varieties.

JEL codes: D24; J23; J61; O15.

Firms in developing countries have lower productivity per worker (Hall and Jones, 1999). A number of factors explain this pattern, such as an imperfect access to capital (Banerjee and Duflo, 2014), inputs (Boehm and Oberfield, 2020), technology (Howitt, 2000), international markets (Verhoogen, 2008), or poor management

*Imbert: University of Warwick, BREAD, CEPR and JPAL, c.imbert@warwick.ac.uk; Seror: Université du Québec à Montréal, University of Bristol, DIAL, Institut Convergences Migrations, seror.marlon@uqam.ca; Zhang: Chinese University of Hong Kong, yifan.zhang@cuhk.edu.hk; Zylberberg: University of Bristol, CESifo, the Alan Turing Institute, yanos.zylberberg@bristol.ac.uk.
We thank Samuel Bazzi, Loren Brandt, Holger Breinlich, Gharad Bryan, Juan Chen, Giacomo De Giorgi, Maelys De La Rupelle, Sylvie Démurger, Taryn Dinkelman, Christian Dustmann, Ben Faber, Giovanni Facchini, Greg Fischer, Richard Freeman, Albrecht Glitz, Doug Gollin, Andre Groeger, Flore Gubert, Naijia Guo, Marc Gurgand, Seema Jayachandran, Marieke Kleemans, Jessica Leight, Florian Mayneris, David McKenzie, Alice Mesnard, Dilip Mookherjee, Joan Monras, Melanie Morten, Albert Park, Sandra Poncet, Markus Poschke, Simon Quinn, Mark Rosenzweig, Gabriella Santangelo, Michael Song, Jon Temple, Chris Udry, Gabriel Ulyssea, Thomas Vendryes, Chris Woodruff, Daniel Yi Xu, Dean Yang, and numerous conference and seminar participants for useful comments. We are grateful to Christine Valente and Eric Verhoogen for their insights at the early stage of the project. Seror gratefully acknowledges support from the Arthur Sachs Scholarship Fund. The usual disclaimer applies.

practices (Bloom, Schankerman and Van Reenen, 2013; Atkin et al., 2017). Another potential factor may be the abundance of migrant labor. The process of economic development induces large movements of rural workers from agriculture to manufacturing (Lewis, 1954), which could reduce firms' incentives to adopt productivity-enhancing technologies (Lewis, 2011). Despite its relevance, empirical evidence on the role of rural-urban migration in shaping urban production in developing countries is scarce.

This paper is the first to estimate the causal effect of rural migrant inflows on urban production in the process of structural transformation. We use longitudinal micro data on Chinese manufacturing firms between 2000 and 2006 and a population micro-census to measure rural-urban migration. Our study period coincides with a unique episode of rural-urban migration: more than 45 million rural workers came to Chinese cities in only five years (16% of the urban population in 2000). We instrument migrant flows to Chinese cities using exogenous shocks to agricultural income in rural areas, which trigger rural-urban migration. We first identify the effect of migration on factor cost, factor use, and factor productivity at destination. We then characterize the restructuring of production at destination through the analysis of technological innovation and product choice within production units.

Providing empirical evidence on the causal impact of labor inflows on manufacturing production requires large and exogenous migrant flows to cities. Our methodology proceeds in two steps. In the first step, we isolate exogenous variation in agricultural income by combining innovations in world prices for agricultural commodities with variation in cropping patterns across prefectures of origin.[1] In the second step, we combine these predictors of rural emigration (the "shift") with historical migration patterns between prefectures (the "share"). The resulting shift-share instrument strongly predicts migrant inflows to cities and exhibits substantial variation across prefectures of destination.

We use the shift-share design to estimate the impact of rural-urban migration flows on manufacturing firms. We find that migration exerts a downward pressure on labor cost.[2] After an influx of migrants, manufacturing production becomes much more labor-intensive, as capital does not adjust to changes in employment, and value added per worker sharply decreases. The effects are large in economic terms: a ten

[1]Prefectures are the second administrative division in China, below the province. There were about 330 prefectures in 2000.

[2]This effect may reflect a shift in labor supply, but also a compositional effect with immigrants receiving lower wages than natives. We quantify the bias in the wage effect, as induced by heterogeneity in labor efficiency between migrants and urban residents, in Appendix D.3. We further show in Appendix E.3 that low-skilled natives do experience a decrease in returns to labor following a migration shock. By contrast, we find no such effects for natives with tertiary education.

percentage point increase in the migrant share at destination causes a 1.5% decrease in compensation per employee and a 4.3% decrease in the capital-to-labor ratio.

Although our main results focus on firms that are observed every year between 2000 and 2006, the composition of the manufacturing sector is in constant evolution during the period, and many firms enter and exit the sample. When we consider all firms present in the sample at any point in time, the shift towards labor intensive production and the decline in productivity following a migration shock are even more pronounced. This suggests that following a migration shock, labor-intensive, low-productivity firms are less likely to die, and more likely to grow.

Changes in input mix may reflect changes in the nature of manufacturing goods, or changes in technology to produce a similar output (as in Beaudry and Green, 2003). We look for evidence of adjustments along both margins. To better characterize the impact of labor inflows on the production process, we first exploit the textual descriptions of products as reported by manufacturing establishments. More specifically, we associate a unique product code (HS-6) to each product, and we characterize the direction of a change in output mix by looking at factor use within product classes.[3] We find that rural-urban migration tilts production toward products that are labor-intensive and with low human capital intensity. Using a match between manufacturing firms and patent applications (He et al., 2018), we also find evidence of a marked decrease in patenting, concentrated in categories linked to fundamental innovation and new production methods, and in patent classes that are capital- and skill-intensive. These effects are modest but not negligible: a ten percentage point increase in the migration rate increases the probability to change products by 2% and decreases the probability to submit a patent by 10%. Finally, we explore whether changes in technology and input mix are systematically related to endogenous product choice. We find that firms that experience an increase in labor supply through migration change their input mix, irrespective of a possible change in product. By contrast, the decline in patenting only occurs among firms that *also* adjust their product in response to migration. These results suggest that there is directed technological change for a given output mix (as in Beaudry and Green, 2003), but that product choice is also a margin of adjustment: innovative firms, which would have pushed the technological frontier through capital-augmenting innovation, now prefer to shift along the frontier and adopt more labor-intensive product varieties.

Our findings are robust to numerous sensitivity checks, e.g., excluding industries that process agricultural goods, controlling for industry fixed effects, or accounting

[3]The HS classification (Harmonized Commodity Description and Coding Systems) is an international standard adopted by the United Nations for the classification of traded goods; its finest classification is at the 6-digit level (HS-6) and distinguishes more than 5,000 product categories.

for a demand shift for the final good. The identification assumption underlying our shift-share design is that "shifts", i.e., agricultural income shocks driven by cropping patterns and price innovations, are numerous and as good as random (Borusyak, Hull and Jaravel, 2018).[4] We provide support for this assumption by transforming our baseline specification at destination into a specification looking at the effect of shifts on transformed outcomes across origins (Borusyak, Hull and Jaravel, 2018). We use this specification to test that pre-trends in outcomes are not correlated with future migration shocks. We also test that our results are not driven by the lagged effects of past migration waves (Jaeger, Ruist and Stuhler, 2018). Finally, we use the recent contribution of Adão, Kolesár and Morales (2019) to discuss the correct inference for our shift-share design.

Our analysis relies on a significant methodological contribution: we process textual product descriptions using a Natural Language Processing (NLP) algorithm to characterize product choice within manufacturing firms in China. A text-based approach to product classification has been used by Hoberg and Phillips (2016) to capture fine-grained product differentiation and study how firms distinguish themselves from close competitors in the United States. Our approach differs from theirs in that our objective is to allocate product descriptions into existing, standard product categories rather than identifying new, and more precise, product clusters. Observing product and technology adoption at the firm level is rare in developing countries. The few exceptions, reviewed in Verhoogen (2020), are contributions looking at specific sub-sectors (e.g., Atkin et al., 2017), or the papers documenting the expansion in product scope when firms gain access to new imported inputs (Goldberg et al., 2010; Bas and Paunov, 2019). Our classification can be useful for research using data on Chinese manufacturing firms, e.g., to study the effect of trade on product choice and technology. The method can also be applied to other contexts in which product or industry information comes as a free-text description.

Our paper contributes to four different strands of literature. First, we use product-level information and patent data to estimate the effect of labor supply shocks on factor use, product choice, and technological adoption at the establishment level. This approach relates to the growing literature that estimates the impact of immigration on factor use at destination (Lewis, 2011; Peri, 2012; Accetturo, Bugamelli and Lamorgese, 2012; Olney, 2013; Dustmann and Glitz, 2015; Kerr, Kerr and Lincoln, 2015; Mitaritonna, Orefice and Peri, 2017). In contrast with a literature

[4] Another recent contribution discusses identification in shift-share designs when "shares" are as good as random, numerous, and dispersed (Goldsmith-Pinkham, Sorkin and Swift, forthcoming). In our shift-share design, shares are previous settlement patterns and may reflect the expectations of migrants about the future evolution of urban production.

that focuses on international migration to developed countries, we study rural-urban migration in a developing country, a context which is less studied but equally important: in 2010, there were as many internal migrants in China alone as international migrants worldwide (205 million). A related literature looks specifically at the positive contribution of the immigration of scientists (e.g., Moser, Voena and Waldinger, 2014) or inventors (e.g., Akcigit, Baslandze and Stantcheva, 2016) on innovation in the United States. In our context, internal migrants are mostly low-skilled, so that they are substitutes for capital and capital-enhancing technological innovation.[5] Our result that firms adopt more labor intensive technology following a migrant inflow is closer to Lewis (2011), who studies the inflow of unskilled migrants to the United States and its impact on the (non-)adoption of automation machinery.

Our focus on the absorption of rural migrants in the urban sector of a fast-growing economy echoes a second, older literature that looks at cities of the developing world (Harris and Todaro, 1970; Fields, 1975). This literature emphasizes the role of labor market imperfections, with rural migrants transiting through unemployment or informal employment upon arrival. By contrast, our findings suggest that migrants swiftly find their way into formal manufacturing firms. We document employment responses to labor supply shocks that are compatible with a relatively flexible labor market, although labor market frictions are likely pervasive in urban China. Such labor market imperfections may be related to job search frictions (Abebe et al., 2016; Alfonsi et al., 2017), informality (Meghir, Narita and Robin, 2015; Ulyssea, 2018) or institutional constraints, e.g., minimum wages (Mayneris, Poncet and Zhang, 2018; Hau, Huang and Wang, 2018). Another source of labor market imperfections is mobility rigidity, leading to large productivity gaps across space and sectors in developing countries (Gollin, Lagakos and Waugh, 2014; Bryan and Morten, 2019), and in China (Brandt, Tombe and Zhu, 2013; Tombe and Zhu, 2019).

Our study also contributes to the literature on structural transformation, which describes the shift of production factors from the traditional sector to the modern sector in developing economies (Lewis, 1954; Herrendorf, Herrington and Valentinyi, 2015). The finding that migration boosts urban employment relates to "labor push"

[5]The non-adjustment of capital to labor supply may reflect a high substitutability between capital and low-skilled labor, or the existence of credit constraints. We calibrate a CES production function at the sectoral level using estimates for the United States (Oberfield and Raval, 2014) and show that, when accounting for the complementarity/substitutability between factors, the marginal product of labor falls sharply, the marginal product of capital rises faintly, and total factor productivity slightly decreases. This finding would be consistent with some degree of credit market imperfections. We also look at treatment heterogeneity across firms with different characteristics at baseline (e.g., factor returns or ownership structure), possibly facing different access to capital (Song, Storesletten and Zilibotti, 2011; Midrigan and Xu, 2014). We do not find large treatment heterogeneity along these baseline firm characteristics.

models which argue that, by releasing labor, agricultural productivity gains may trigger industrialization (Alvarez-Cuadrado and Poschke, 2011; Gollin, Parente and Rogerson, 2002; Bustos, Caprettini and Ponticelli, 2016).[6] In order to identify migration inflows that are exogenous to labor demand at destination, our paper takes the opposite approach to "labor pull" models, in which rural migrants are attracted by increased labor productivity in manufacturing (Facchini et al., 2015). Closely related to our paper, Bustos et al. (2018) find that regions of Brazil that benefited from genetically-engineered soy specialized in low-productivity, low-innovation manufacturing, and argue that the effect is driven by the inflow of unskilled labor released by agriculture. Our contribution is to identify the effect of rural migrant labor supply on urban production independently from factors such as consumer demand (Santangelo, 2016) and capital availability (Marden, 2015; Bustos, Caprettini and Ponticelli, 2016), and to document changes in products and technology at the firm level.

Our empirical analysis finally relates to the literature that estimates the effect of immigration on labor markets (Card and DiNardo, 2000; Card, 2001; Borjas, 2003), and more specifically to studies of internal migration (e.g., Boustan, Fishback and Kantor, 2010; El Badaoui, Strobl and Walsh, 2017; Imbert and Papp, 2019; Kleemans and Magruder, 2018). Since internal migrants are usually closer substitutes to resident workers than international migrants to natives, the literature on internal migration tends to find larger negative effects on wages at destination. In China, the evidence is mixed: De Sousa and Poncet (2011); Ge and Yang (2014) find a negative effect, Meng and Zhang (2010) no effect and Combes, Démurger and Li (2015) a positive effect. We find moderate negative effects: the immediate negative effect of the migrant labor supply shock is partly compensated by the shift of firms towards more labor-intensive production and technology, which increases labor demand.

The remainder of the paper is organized as follows. Section I presents data sources and the empirical strategy. Section II describes the effect of immigration on factor cost and factor use. Section III characterizes the reorganization of production through product choice and technological innovation. Section IV briefly concludes.

[6] Our results depart from the traditional "labor push" interpretation in that migration from rural areas is triggered by a *negative* shock to agricultural productivity (as in Gröger and Zylberberg, 2016; Feng, Hu and Moffitt, 2017; Minale, 2018). Worse economic conditions at origin lower the opportunity cost of migrating, an effect which dominates an (opposite) effect operating through tighter liquidity constraints (Angelucci, 2015; Bazzi, 2017).

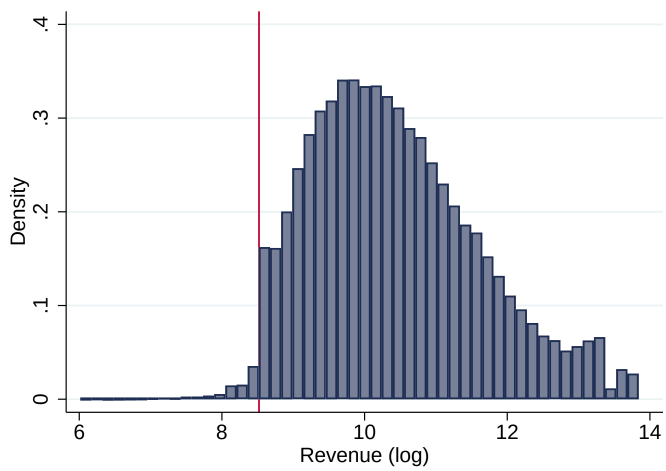
I Data and empirical strategy

This section describes the data on production and migrants, explains how we identify exogenous variation in migrant inflows, and presents the empirical specification.

I.1 Firms and migrants

Measure of production in cities Our main data source is a census of Chinese manufacturing firms conducted by the National Bureau of Statistics (NBS).[7] The NBS implements a yearly census of all state-owned manufacturing enterprises and all non-state manufacturing firms with sales exceeding RMB 5 million (approx. $600,000). Although the sample does not include smaller firms, it accounts for 90% of total manufacturing output. Firms can be matched across years: our main analysis focuses on the balanced panel of 31,886 firms present every year between 2000 and 2006. The NBS collects information on location, industry, ownership type, exporting activity, number of employees, and a wide range of accounting variables (sales, inputs, value added, wage bill, fixed assets, financial assets, etc.). We divide total compensation (including housing and pension benefits) by employment to compute the compensation rate and we construct real capital as in Brandt, Van Biesebroeck and Zhang (2014).

Figure 1. Distribution of revenue across firms (NBS, 2000–2006).



Sources: Firm-level data from the National Bureau of Statistics (NBS), 2000–2006. The revenue threshold for appearing in the NBS Census of above-scale firms is RMB 5,000,000, corresponding to $ln(5,000) \approx 8.52$ along the logarithmic scale (of revenues expressed in thousands of RMB).

[7]The following discussion partly borrows from Brandt, Van Biesebroeck and Zhang (2014), and a detailed description of data construction choices is provided in Appendix A.1.

There are three potential issues with the NBS data. First, matching firms over time is difficult because of frequent changes in identifiers. We apply the fuzzy algorithm from Brandt, Van Biesebroeck and Zhang (2014) to detect "identifier-switchers", using firm name, address, and phone number etc. Second, although we use the term "firm" in this paper, the NBS data cover "legal units" (*faren danwei*), which roughly correspond to the definition of "establishments" in the United States.[8] Third, the RMB 5 million threshold may not be strictly implemented: private firms may enter the database a few years after having reached the sales cut-off or continue to participate in the census even if their annual sales fall below the threshold. We cannot measure delayed entry into the sample, but there are very few surveyed firms below the threshold, as Figure 1 shows.

Our baseline outcomes include compensation per worker, employment, capital-to-labor ratio, and value added per worker. Table 1 provides descriptive statistics of our key outcomes in 2000 and 2006 for the balanced sample and for the whole sample of firms. The study period is one of fast manufacturing growth: employment in the balanced sample increases by 30%, capital per worker by 29%, and value added per worker by 76%. Labor costs per worker also increase by about 10% per year, much faster than inflation, which is about 2% per year over the period. There is also rapid growth in the number of manufacturing establishments: sample size is multiplied by about six between 2000 and 2006, so that firms of the balanced sample represent 40% of manufacturing establishments in 2000, but only 7% in 2006. Firms in the balanced sample are larger, and more capital-intensive at baseline, and they grow faster than the average firm between 2000 and 2006. While they have higher productivity per worker and pay higher wages than the average firm at baseline, productivity and wages grow faster in the average firm.

The NBS does not provide a precise, systematic classification of products. Instead, it collects textual descriptions of up to three main products without standardizing them. We develop a Natural Language Processing algorithm to match product information with a unique HS 6-digit code, through the systematic comparison of the textual description provided by manufacturing firms with descriptions of each HS 6-digit code. In a first step, we collect descriptions in Mandarin of the standardized HS 6-digit product classification. We clean all textual descriptions by applying a tokenizer ("jieba") which groups characters into words, and by deleting stop words.

[8]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. The share of single-plant firms is above 90% over our period of interest (Brandt, Van Biesebroeck and Zhang, 2014).

Table 1. Summary statistics of key firm-level outcomes.

	Balanced sample of firms		Unbalanced sample of firms	
	2000	2006–2000	2000	2006–2000
Labor cost	2.282	0.572	2.170	0.615
Employment	4.901	0.261	4.406	0.040
K/L ratio	3.802	0.254	3.701	0.069
Y/L ratio	3.680	0.565	3.570	0.638
Immigration rate	-	0.329	-	0.316
Observations	31,886	31,886	79,980	454,781

Sources: NBS firm-level data (2000, 2006). This table presents the baseline value in 2000 and growth between 2000–2006 for the key outcome variables. The first and second columns report statistics based on the balanced sample of firms; the third and fourth columns report statistics based on all firms present in the NBS data. *Labor cost* is the (log) compensation per worker including social security and housing benefits. *Employment* is the (log) number of workers. *K/L ratio* is the (log) ratio of fixed assets to employment. *Y/L ratio* is the (log) ratio of value added to employment. *Immigration rate* is the ratio of immigrants between 2000 and 2005 to the number of urban residents at destination in 2000.

This step transforms a list of characters into a sequence of identified, relevant words. In a second step, we use the powerful neural net developed by Google ("word2vec") in order to represent every contiguous sequence of words in a vector space. The neural net needs to be trained, and we rely on the word embeddings provided by Li et al. (2018) and trained on the Wikipedia corpus. This representation allows us to compute the distance between any sequences of words, by using their projection onto the vector space. We then compute the average similarity score between all contiguous sequences of words within the product description and word sequences within HS 6-digit descriptions. The output of this procedure is a classification of products with (i) the most likely HS 6-digit code—with the highest similarity score—and (ii) a similarity score to account for possible measurement error.[9] Linking products to HS 6-digit codes allows for a precise characterization of firm production, e.g., by using production patterns within a product class. We classify the products associated at baseline with firms whose capital intensity is higher than the median as "high capital-to-labor ratio" and those associated with firms with a share of workers with high school education higher than the median as "high education" products.[10]

Finally, to measure technological innovation, we use the bridge constructed by He et al. (2018) to match the NBS firm data with all patents submitted to the State Intellectual Property Office (SIPO). The data cover three main categories of

[9] We provide a detailed description of our text-based classification in Appendix A.2.

[10] We also use HS-6 digit codes to identify possible linkages across firms through input-output accounts or technological closeness, as in Bloom, Schankerman and Van Reenen (2013).

patents: design (external appearance of the final product), innovation (fundamental innovations in methods) and utility (changes in processing, shape or structure of products). We also use the patent code and categorize patents by the characteristics of firms that submitted patents with the same code at baseline. Specifically, we define as "high capital-to-labor ratio" all patent codes associated in 2000 with firms that had a capital-to-labor ratio above median (measured in 2000), and as "high education" all patent codes associated in 2000 with firms that had above median share of employees with at least high school education (measured in 2004).

Migration flows To measure migration flows, we use the representative 2005 1% Population Survey (hereafter, "2005 Mini-Census"), collected by the National Bureau of Statistics.[11] The sampling frame of the 2005 Mini-Census covers the entire population at their current place of residence, regardless of whether they hold local household registration (*hukou*), i.e., including migrants. The census collects information on occupation, industry, income, ethnicity, education level, and housing characteristics; it also provides us with key information regarding migration history. First, we observe the household registration type (agricultural or non-agricultural), place of registration, and place of residence at the prefecture level.[12] Second, migrants are asked the main reason for leaving their place of registration, which year they left, and their place of residence one and five years before the interview.[13] We combine these two pieces of information to create a matrix of rural-urban migration flows between all Chinese prefectures every year between 2000 and 2005. We only include migrants who were 15 to 64 years at the time of migration, and exclude migrants who study or migrated to study (less than 5% of the total). The size and the speed of rural-urban migration during the study period are unprecedented. We estimate that 45 million rural workers migrated to cities between 2000 and 2005, equivalent to 16% of the urban population in 2000. Since migrants converged towards manufacturing centers, the average firm in our sample is exposed to an even

[11] These data have also been used, among others, by Combes, Démurger and Li (2015); Facchini et al. (2015); Meng and Zhang (2010); Tombe and Zhu (2019).

[12] During our period of interest, barriers to mobility come from restrictions due to the registration system. These restrictions do not impede rural-urban migration but limit the benefits of rural migrants' long-term settlement in urban areas. See Appendix B.1 for more details about how mobility restrictions are applied in practice and the rights of rural migrants in urban China.

[13] A raw measure of migration flows may not account for two types of migration spells: step and return migration. *Step migration* occurs when migrants transit through another city before reaching their destination, so that the date of departure from the place of registration differs from the date of arrival at the current destination. *Return migration* occurs when migrants leave their places of registration after 2000 but come back before 2005, so that they do not appear as migrants in the Mini-Census. Appendix B.2 shows that return migration is substantial while step migration is negligible, and explains how we adjust migration flows to account for return migration.

higher immigration rate of 33% in only five years (Table 1).

The relocation of workers across prefectures is driven by preferences, migration frictions but also, especially during the study period, by labor demand and labor supply shocks.[14] Since our objective is to identify the adjustment of production to migrant inflows in cities, our setting lends itself to the use of a shift-share design (as in Card, 2001, for instance), in which labor supply shocks across origins ("shifts") are combined with historical migration patterns ("shares") into an instrument for migrant inflows.[15]

I.2 A shift-share instrument

In this section, we construct a shift-share instrument, z_d, for migrant flows to a destination d by combining an exogenous shock s_o to agricultural income at origin $o \in \Theta$ (where Θ represents the set of prefectures) with settlement patterns of past migration waves, λ_{od}:

$$z_d = \sum_{o \in \Theta \setminus \{d\}} \lambda_{od} s_o.$$

Agricultural income shocks The agricultural income shock, s_o, is obtained from interacting origin-specific cropping patterns and innovations in commodity prices.

We first construct a measure of cropping patterns in each prefecture by combining a baseline measure of harvested area with potential yield.[16] We use a geo-coded map of harvested areas in China from the 2000 World Census of Agriculture, in order to construct the total harvested area h_{co} for a given crop c in a given prefecture o. Information on potential yield per hectare is extracted from the Global Agro-Ecological Zones Agricultural Suitability and Potential Yields (GAEZ) and collapsed at the prefecture level, y_{co}.[17] We compute potential agricultural output for each crop in each prefecture as the product of total harvested area and average potential yield, i.e., $q_{co} = h_{co} \times y_{co}$. By construction, the potential agricultural output, q_{co}, is time-invariant and is measured at the beginning of the study period. Figure 2 displays potential output q_{co} for rice and cotton, and illustrates the wide cross-sectional variation in agricultural portfolio across Chinese prefectures.

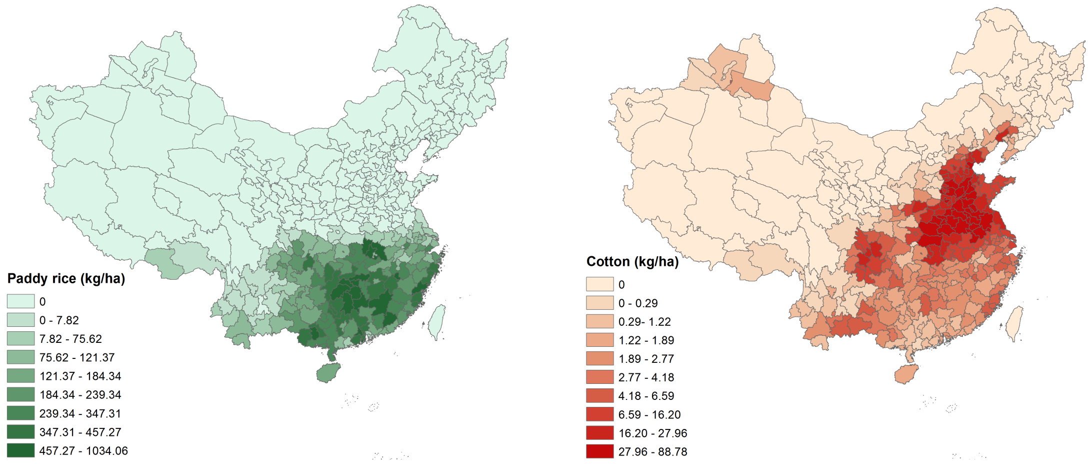
[14]We provide descriptive statistics on migration patterns across regions, and we discuss the selection of migrants in Appendix B.3.

[15]Appendix C.1 develops a stylized theoretical model to explain the economic mechanisms behind the shift-share instrument.

[16]Appendix C.2 describes in more detail how we construct the agricultural income shock and provides summary statistics about the variation in cropping patterns across prefectures and regions.

[17]These maps are provided by the Food and Agriculture Organization (FAO) and the International Institute for Applied Systems Analysis (IIASA), and they are available online from http://www.fao.org/nr/gaez/about-data-portal/en/.

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



(a) Paddy rice. (b) Cotton.

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

To measure innovation in commodity prices, we use Agricultural Producer Prices (APP, 1991–2016) from the FAO, which reports yearly prices "at the farm gate" in each producing country per tonne and in USD between 1991 and 2016. We focus on 21 commodities/crops, which represent 80%-90% of total agricultural output over the period.[18] We construct the international price of each crop as the average price across countries (excluding China) weighted by their baseline share in global exports.[19] Our measure of the year- and crop-specific innovation in the logarithm of nominal prices, p_{ct}, is the residual $\hat{\varepsilon}_{ct}$ from the following AR(1) specification:

$$\log(p_{ct}) = \theta \log(p_{ct-1}) + \eta_t + \nu_c + \varepsilon_{ct}$$

where η_t captures average nominal food prices in each year.[20]

[18] These 21 commodities/crops are banana, cassava, coffee, cotton, fodder crops (barley), groundnut, maize, millet, other cereals (oats), potato, pulses (lentil), rapeseed, rice, sorghum, soybean, sugar beet, sugar cane, sunflower, vegetables (cabbage), tea and wheat. We exclude from our analysis tobacco, for which China has a dominant position on the international market.

[19] For international prices to affect agricultural incomes in rural China, there needs to be a sufficient pass-through to domestic prices. In Appendix C.2, we check that fluctuations in international prices strongly affect producer prices at the farm gate in China.

[20] In Appendix C.3 (effect of price shocks on outmigration) and Appendix F.1 (causal effect of migrant inflows at destination), we test the robustness of our main results to alternative specifications of the price shock: (i) we use commodity prices on international markets from the World Bank Commodities Price Data ("The Pink Sheet"); (ii) we restrict the agricultural portfolio to 17 commodities/crops for which the match between commodity prices and harvested area is immediate; (iii) we isolate price innovations using an AR(2) specification; (iv) we use a Hodrick-Prescott

We combine innovations in crop prices with cropping patterns to construct the "shift" of our shift-share design. The *agricultural income shock* in prefecture o and year t, denoted by s_{ot}, is the average of the percentage deviation in crop prices, $\hat{\varepsilon}_{ct}$, weighted by the expected share of each crop in the agricultural revenue of prefecture o:

$$s_{ot} = \left(\sum_c \bar{p}_c q_{co} \hat{\varepsilon}_{ct} \right) / \left(\sum_c \bar{p}_c q_{co} \right) \tag{1}$$

where $\bar{p}_c$ denotes the nominal international price for each crop at baseline. s_{ot} varies over time, due to fluctuations in world demand and supply, and across space, due to the wide variety of harvested crops across China.[21] The "shift" of our shift-share instrument, s_o, is the sum of s_{ot} over the period 2000–2005.

Migration flows and previous settlement patterns To measure settlement patterns at baseline, we rely on a 1% extract of the 2000 Population Census and use the same definition of migrants as in the 2005 Mini-Census. The emigration rate, n_{ot}, is obtained by dividing the sum of migrants who left origin o (rural areas of prefecture o) in year t by the number of working-age residents in o in 2000, which we denote with P_o. Letting M_{odt} denote the number of workers migrating from origin o to the urban areas of a prefecture d, different from origin o, in a given year $t = 2000, \ldots, 2005$, we have:

$$n_{ot} = \frac{\sum_{d \in \Theta \setminus \{o\}} M_{odt}}{P_o}.$$

The immigration rate, m_{dt}, is obtained by dividing the sum of migrants who arrived in destination d in year t by the number of residents (non-migrants) in d at baseline, in 2000, which we denote with P_d,

$$m_{dt} = \frac{\sum_{o \in \Theta \setminus \{d\}} M_{odt}}{P_d}.$$

The "shares" of our shift-share instrument are the historical settlement patterns from each prefecture of origin to each prefecture of destination, which we measure using the stock of migrants in 2000:

$$\lambda_{od} = \frac{\sum_{t<2000} M_{odt}}{\sum_d \sum_{t<2000} M_{odt}},$$

filter with a parameter of 14,400 to isolate short-run fluctuations in prices. Our main findings are robust to these alternative specifications.

[21] The agricultural income shock, s_{ot}, retains part of the persistence in crop prices: a shock in year t would affect not only revenue in that year, but also expected revenue in the future.

where $\sum_{t<2000} M_{odt}$ is the stock of migrants with a rural hukou from origin o who arrived at destination d before 2000. Historical settlement patterns capture relocation costs (e.g., travel time), idiosyncratic variation in migrant networks, preferences for certain destinations (Kinnan, Wang and Wang, 2018), or permanent differences in labor demand across urban areas. We show in Appendix C.3 that they do predict the subsequent allocation of rural migrants across urban destinations between 2000 and 2005. The relationship is noisy, which may be due to changes in labor demand in Chinese cities after WTO accession (Facchini et al., 2015).[22]

Predicting migrant flows Our shift-share design relies on the premise that agricultural returns affect emigration from rural areas. To test this, we regress the rural emigration rate between 2000 and 2005, n_o, on the agricultural income shock, s_o:

$$n_o = \beta_0 + \beta_1 s_o + \eta_o. \tag{2}$$

Panel A of Table 2 presents the estimates.[23] Emigration between 2000 and 2005 is negatively correlated with the agricultural income shock. A 10% lower agricultural income (about one standard deviation in s_o) is associated with a 1.2 p.p. higher migration incidence.[24] The negative relationship between agricultural income and migration suggests that migration decisions are driven by the opportunity cost of migration, rather than by liquidity constraints (Angelucci, 2015; Bazzi, 2017).[25]

We construct our shift-share instrument by combining the agricultural income shock across origins with earlier migration patterns:

$$z_d = \sum_{o \in \Theta \setminus \{d\}} \lambda_{od} s_o. \tag{3}$$

[22] Appendix C.3 and Appendix F.1 test the robustness of our main results to two alternative definitions of the "shares": (i) we use migration stocks in 1995; (ii) we use predicted settlement patterns based on a gravity model of migration flows (as in Boustan, Fishback and Kantor, 2010, for instance).

[23] In the baseline specifications, we apply a 99% winsorization to emigration and immigration rates. Appendix C.3 and Appendix F.2 test the sensitivity of our findings to the definition of migration: all migrants; males only; low-skilled only; no adjustment for return migration; and including outliers.

[24] This semi-elasticity corresponds to an elasticity of -2.7. We show in Appendix C.1 that the elasticity of the emigration rate to the agricultural revenue may be interpreted as the shape parameter for the distribution of worker preferences for different locations, as is common in models of New Economic Geography (Monte, Redding and Rossi-Hansberg, 2018; Bryan and Morten, 2019; Monras, 2020).

[25] In the Chinese context, workers migrate without their families, low-skill jobs in cities are easy to find, and the fixed cost of migration is relatively low. Chinese households also have high savings, so that the impact of short-term fluctuations in agricultural prices on wealth is small.

Table 2. Origin-based migration predictions.

	Emigration	
	Inter-prefecture	Outside 300-km radius
Panel A: Emigration rate from origin		
Price shock	-0.117	-0.045
	(0.019)	(0.019)
Observations	335	335
	Immigration	
	Inter-prefecture	Outside 300-km radius
Panel B: Immigration rate at destination		
Shift-share instrument	-1.620	-1.305
	(0.425)	(0.368)
Observations	315	315

Notes: Robust standard errors are reported between parentheses. In Panel A, the dependent variable is the number of rural emigrants to urban areas in other prefectures or in prefectures located outside of a 300-km radius around the origin, divided by the number of rural residents at origin. In Panel B, the dependent variable is the number of rural immigrants from other prefectures or prefectures located outside of a 300-km radius around the destination divided by the number of urban residents at destination. See Section I and Equations (2) and (4) for a more comprehensive description of the two specifications.

To check that the instrument is a good predictor of immigration flows, we regress the actual immigration rate between 2000 and 2005, m_d, on z_d:

$$m_d = \alpha_0 + \alpha_1 z_d + \eta_d. \tag{4}$$

Panel B of Table 2 presents the estimates. The relationship is significant and negative: it is the first stage of the empirical strategy that we describe next.

I.3 Empirical strategy and identification

Empirical strategy Our baseline specification considers a specification in difference using the balanced panel of firms present every year between 2000 and 2006. We regress the change in outcomes between 2000 and 2006, Δy_{id}, for a firm i in prefecture d on the sum of yearly migration flows over the period, m_d:

$$\Delta y_{id} = \alpha + \beta m_d + \varepsilon_{id}, \tag{5}$$

where standard errors are clustered at the level of the prefecture of destination and each observation is weighted by firm employment at baseline, e_{id}. Migration flows may reflect a surge in labor demand after opening to trade or local investment in infrastructure and amenities. In order to isolate a supply-driven component in the relocation of workers, we use the shift-share variable z_d (see Equation 3) as an instrument for the immigration rate m_d and estimate Equation (5) with 2SLS. We have already reported a prefecture-level equivalent of the first stage by regressing the immigration rate on the instrument in Equation (4).

Identification and inference A recent literature discusses identification and inference in shift-share designs (Adão, Kolesár and Morales, 2019; Borusyak, Hull and Jaravel, 2018; Goldsmith-Pinkham, Sorkin and Swift, forthcoming). It suggests that consistency can be achieved if either the shares (Goldsmith-Pinkham, Sorkin and Swift, forthcoming) or the shifts (Borusyak, Hull and Jaravel, 2018) are exogenous. In our setting, the shares—previous settlement patterns—reflect the expectations of workers about the evolution of labor demand across destinations; they are likely endogenous to production outcomes in cities. Instead, the validity of our shift-share design relies on the assumption that shifts—the agricultural income shocks—are exogenous to manufacturing outcomes. Specifically, following Borusyak, Hull and Jaravel (2018), a shift-share estimator would be consistent under two conditions: (i) shifts are quasi-randomly assigned across origins, and (ii) there are many uncorrelated shifts. Given our empirical strategy, condition (i) is likely to hold: innovations in the international price of agricultural commodities are driven by world supply and demand and are likely exogenous from the point of view of each Chinese prefecture at baseline.[26] Condition (ii) may not be verified: the shifts are spatially correlated in our setting, since they combine innovations in crop prices with cropping patterns, which are strongly determined by geography (see Figure 2).

To discuss these issues, we apply the equivalence result of Borusyak, Hull and Jaravel (2018) at the level of origins, and transform our firm-level specification (5) into a shift-level specification.[27] In our setting, this equivalence result conveys a

[26] We exclude from the analysis the only crop for which China is a price setter on the international markets, tobacco, which also happens to be produced in a specific region of China.

[27] In principle, our shift-share design can be interpreted as the combination of agricultural income shocks (a "shift" at the level of origins) with previous settlement patterns, but it can also be interpreted as the interaction of price innovations (a "shift" at the level of crops) with a combination of previous settlement patterns and agricultural portfolios at typical origins. The equivalence result of Borusyak, Hull and Jaravel (2018) applied at the level of crops would require innovations in agricultural commodity prices to be exogenous and independent, which is a weaker identification assumption than our current one. However, the equivalence result also requires shifts to be numerous, and there are only 21 crops; this is why we apply it at the level of the 335 prefectures

simple intuition. Our shift-share design transforms agricultural income shocks at origin into shocks to firms at destination via a matrix of migration patterns. The equivalence result of Borusyak, Hull and Jaravel (2018) allows us to invert this transformation and estimate, at the origin-level, the effect of agricultural income shocks on firm outcomes in the typical destination, i.e., on a weighted average of firm outcomes using migration patterns as weights. Concretely, we first aggregate firm-level outcomes y_{id} into destination-level outcomes y_d. We then construct the weights $\tilde{\lambda}_{od} = e_d \lambda_{od} / \left(\sum_d e_d \lambda_{od}\right)$, where λ_{od} are historical migration shares and e_d is total employment in prefecture d at baseline. Finally, we compute transformed outcomes $\tilde{y}_o = \sum_d \tilde{\lambda}_{od} y_d$ and transformed immigration $\tilde{m}_o = \sum_d \tilde{\lambda}_{od} m_d$ and estimate the following equation:

$$\Delta \tilde{y}_o = \alpha + \beta \tilde{m}_o + \varepsilon_o. \tag{6}$$

where the agricultural income shock, s_o, is used as an instrument for $\tilde{m}_o$. Following Borusyak, Hull and Jaravel (2018), we use this specification to discuss identification and test for pre-trends in outcomes between 1998–2000. We also check two conditions about the distribution of migration patterns: (i) the Herfindahl index of origin contributions, $\sum_o (\sum_d \tilde{\lambda}_{od})^2$, is small so that the effect is not driven by a few origins; (ii) the Herfindahl indices of settlement patterns, $\sum_o \lambda_{od}^2$, are not too small on average so that shocks do not affect all destinations to the same extent. Finally, outcomes may be correlated across destinations with similar migration shares, which implies that standard inference may be invalid. We use the method proposed by Adão, Kolesár and Morales (2019) to provide valid standard errors in our shift-share design; we also use specification (6) to provide standard errors accounting for spatial correlation across origins (e.g., using Conley, 1999).

Even if the shifts are randomly assigned across origins, we need the exclusion restriction to hold, i.e. that agricultural income shocks in rural hinterlands only affect firm outcomes in cities through migration. This raises a number of concerns. First, changes in the price of agricultural output may affect local industries that use agricultural products as intermediate inputs. We check that our results are robust to excluding firms that process agricultural goods and to controlling for industry-specific trends. Second, cities and their surroundings are also integrated through final goods markets, so that agricultural income in rural hinterlands may affect demand for manufactured products in cities (Santangelo, 2016). We alleviate this concern by (i) excluding migration within a 300km radius, (ii) controlling for the agricultural income shock in neighboring prefectures, (iii) controlling for market access, and (iv) considering only exporting firms, less dependent on local demand.

of origin.

Finally, the shift-share instrument correlates with past migration by construction, which may influence current outcomes (a concern raised by Jaeger, Ruist and Stuhler, 2018). We address this concern by (i) controlling for the stock of immigrants at baseline and (ii) checking that lagged shocks (1993–1998) do not explain later changes in outcomes.

II Migration, labor cost, and factor demand

This section quantifies the effect of migrant labor supply on labor cost, factor demand, and factor productivity at destination.

II.1 Average effect on labor cost, factor demand and factor productivity

A key determinant of firms' structure of production is relative input costs. We estimate specification (5) on the sub-sample of firms present all years between 2000 and 2006 and use log total compensation per employee (including fringe benefits) as a measure of labor cost. The first column of Table 3 displays the OLS estimate (Panel A) and the IV estimate (Panel B); observations are weighted by employment at baseline.[28] An inflow of rural migrants is negatively associated with labor cost at destination: a ten percentage point increase in the immigration rate induces a 1.5% decrease in compensation per employee.[29]

Our findings are in line with recent studies arguing that rural-urban migration has moderated wage growth in urban China (De Sousa and Poncet, 2011; Ge and Yang, 2014). The magnitude of the previous wage response to immigration is comparable to the literature on international migrants in developed countries (see, e.g., Borjas, 2003). However, internal migrants are more substitutable with "natives" than international migrants, and the literature on internal migration in developing economies tends to find larger negative wage effects (see, e.g., Kleemans and Magruder, 2018; Imbert and Papp, 2019). An important difference between our empirical approach and these papers is that we estimate the effects of migration on changes in wages over a six year period, instead of year-on-year changes. This gives time for labor demand to adjust upwards, e.g., through technology and prod-

[28] In the baseline specification, we apply a 99% winsorization to firm outcomes.

[29] Average compensation per employee may decrease due to an outward shift in labor supply, but also to the replacement of native workers by less productive migrants (see Appendix D.3 for a back-of-the-envelope quantification of such an effect). The NBS data do not provide yearly information on the composition of the workforce by skill or migrant status. To shed light on the issue, we exploit the Urban Household Survey (2002–2006), a representative survey of urban "natives," which we describe in Appendix A.3. The estimates are noisy, but confirm that wages decline among native workers, especially less-skilled ones—see Appendix E.3.

Table 3. Impact of migration inflows on urban firms.

	Labor cost (1)	Employment (2)	K/L ratio (3)	Y/L ratio (4)
Panel A: OLS estimates				
Migration	-0.172	0.270	-0.269	-0.339
	(0.057)	(0.035)	(0.044)	(0.071)
Observations	31,886	31,886	31,886	31,886
Panel B: IV estimates				
Migration	-0.147	0.294	-0.431	-0.437
	(0.062)	(0.053)	(0.095)	(0.108)
Observations	31,886	31,886	31,886	31,886
F-stat. (first stage)	23.59	23.59	23.59	23.59

Notes: Standard errors are clustered at the prefecture level and reported between parentheses. The sample is composed of the firms present every year in the NBS firm census between 2000 and 2006. *Migration* is the immigration rate, i.e., the migration flow divided by population at destination and at baseline. *Labor cost* is the (log) compensation per worker including social security and housing benefits. *Employment* is the (log) number of workers. *K/L ratio* is the (log) ratio of fixed assets to employment. *Y/L ratio* is the (log) ratio of value added to employment. See Section I and Equation (5) for a description of the IV specification.

uct adjustments towards more labor-intensive production lines as we will show in Section III, which would attenuate the immediate negative effect on wages. Another possible explanation for the lower wage impacts in our setting is labor market regulation: the minimum wage legislation is gradually implemented during our study period (Mayneris, Poncet and Zhang, 2018; Hau, Huang and Wang, 2018).

Following a positive labor supply shock and a decline in wages, one would expect manufacturing firms to hire more workers. Column 2 in Table 3 presents the estimated effect of immigration on (log) employment: a ten percentage point increase in the immigration rate raises employment in the average manufacturing firm by 2.9%. The magnitude of our estimate suggests that a large proportion of migrant workers is not hired by the firms in our sample: they may be hired by smaller firms, work in other sectors (e.g., construction), or transit through unemployment or self-employment (Giulietti, Ning and Zimmermann, 2012; Zhang and Zhao, 2015).

Migrant labor supply shocks strongly affect relative factor use at destination. As column 3 of Table 3 shows, the capital-to-labor ratio decreases by 4.3% following a ten percentage point increase in the migration rate, which suggests that capital does not adjust to the increase in employment.[30] There are three possible reasons for this finding. First, firms that expand may belong to sectors with relatively high substitutability between capital and labor. A moderate (and negative) adjustment of capital could then be an optimal response. Second, credit constraints or adjustment costs may prevent firms from reaching the optimal use of production factors. To shed light on these two channels, we study treatment heterogeneity in Appendix E.1 and estimate migration effects for (i) firms in sectors with high elasticity of substitution between factors and (ii) public sector firms, which have an easier access to credit (Brandt, Tombe and Zhu, 2013). We do not find evidence of heterogeneous effects along these dimensions. A third possible reason for the lack of adjustment of capital may be that firms change the organization of their production lines and adopt new technologies with different factor intensities. We provide evidence in support for this interpretation in Section III.

Finally, the average product of labor falls sharply in response to migrant inflows. A ten percentage point increase in the immigration rate decreases value added per worker by 4.4% (column 4 of Table 3). Since employment increases by 2.9%, the coefficient implies that the labor supply shock has a negative, albeit small, effect on value added at the firm level. Firm expansion may come at a cost; for instance, new hires may need to be trained and production lines may need to be adjusted. In Appendix D.1, we develop and estimate a quantitative framework à la Oberfield and Raval (2014) which allows for sector-specific complementarities between capital and labor, and we compute the marginal revenue of capital and labor, as well as total factor productivity for each firm. We then estimate the effect of migration on these productivity measures, and show that the marginal revenue product of labor falls markedly when immigration increases, the marginal revenue product of capital rises slightly, and total factor productivity decreases moderately (Appendix D.2).

The response of manufacturing firms to internal immigration in China resonates with Lewis's (2011) findings on Mexican immigration to the United States in the 1980s and 1990s: firms choose not to mechanize due to the availability of cheap labor. We provide additional support for this interpretation in Section III, by shedding light on the adoption of new products and new technologies. The rest of this section provides sensitivity analysis of our baseline results, starting with compositional effects and sample choice, and ending with a discussion of identification.

[30]The effect of immigrant flows on capital can be inferred from adding the estimates of columns 2 and 3 of Table 3: the OLS estimate would be 0.001 and the IV estimate is about -0.137, not significantly different from zero at the 5% level.

II.2 Aggregation, firm selection and entry/exit

Immigration may change the allocation of factors across firms: in that case, its effect on aggregate outcomes may differ from its effect on firm-level outcomes.

We first investigate the effect of immigration on aggregate outcomes constructed from the baseline sample of firms present every year in the NBS data between 2000 and 2006. We compute the sum of employment, wage bill, value added and capital across firms, and construct our main outcomes—compensation per worker, employment, capital-labor ratio and value added per worker—at the sector $\times$ prefecture level. We then estimate specification (5), with a sector $\times$ prefecture as the unit of observation. Panel A of Table 4 presents the results. The effects on labor cost, employment, capital-to-labor ratio, and value-added per worker are similar to the within-firm results from Table 3, which suggests that compositional effects within the balanced sample are small. Appendix E.1 provides corroborating evidence: we find little evidence of heterogeneous effects along firm characteristics such as capital intensity or output per worker at baseline. We also present the outcome of an aggregation at the prefecture level in Appendix E.3 and show similar estimates to Table 3. This suggests that the reallocation across sectors is small, which is consistent with the literature on developed countries (Dustmann and Glitz, 2015).

Immigration may also affect firm entry and exit, so that results based on firms present from 2000–2006 would miss part of its aggregate impact. To account for the potential effect on entry into and exit from the sample, we construct outcomes at the sector $\times$ prefecture level using all firms observed at any point in the NBS data between 2000 and 2006. The results are shown in Panel B of Table 4. The wage response to a ten percentage point increase in the immigration rate is -1.7%, close to the estimate using the balanced sample (-1.3%). By contrast, the employment effect is twice as large within the unbalanced sample, suggesting that an important share of migrant workers are absorbed by new entrants or future exiters (as in Dustmann and Glitz, 2015). Accounting for entry into and exit from our sample amplifies the effect of migration on production: the selection effect seems to favor firms with low capital intensity and low productivity per worker.

We provide more direct evidence on firm selection in Appendix E.2, with the caveat that we do not observe all firms but only those above the RMB 5 million sales threshold. We first estimate the effect of migration on profitability and on the probability that an establishment reports net profits in the balanced sample of firms. We find that migration increases the profitability of incumbent firms. Second, we study firm entry and exit: we consider as entrants firms that appeared in the sample and who were created between 2000 and 2006, and as exiters firms that

Table 4. Impact of migration inflows on urban firms—sensitivity analysis with aggregate variables at the prefecture × sector level.

	Labor cost (1)	Employment (2)	K/L ratio (3)	Y/L ratio (4)
Panel A: Balanced sample of firms				
Migration	-0.134	0.352	-0.492	-0.487
	(0.062)	(0.079)	(0.126)	(0.117)
Observations	4,495	4,495	4,495	4,495
F-stat (first)	23.76	23.76	23.76	23.76
Panel B: Unbalanced sample of firms				
Migration	-0.167	0.709	-0.592	-0.660
	(0.074)	(0.145)	(0.136)	(0.196)
Observations	6,424	6,424	6,424	6,424
F-stat. (first)	21.28	21.28	21.28	21.28

Notes: Standard errors are clustered at the prefecture level and reported between parentheses. The unit of observation is a prefecture × sector. In Panel A, the sample is composed of the firms present every year in the NBS firm census between 2000 and 2006. In Panel B, the sample is composed of all firms present in the firm census at any point between 2000 and 2006. Outcomes are aggregated at the prefecture × sector level. *Migration* is the immigration rate, i.e., the migration flow divided by population at destination and at baseline. *Labor cost* is the (log) compensation per worker including social security and housing benefits. *Employment* is the (log) number of workers. *K/L ratio* is the (log) ratio of fixed assets to employment. *Y/L ratio* is the (log) ratio of value added to employment.

disappear from the sample between these dates. We find that migration lowers firm exit *and* entry. Taken together, these results suggest that cheaper labor allows low-productivity, low-profitability incumbent firms to survive, or at least to remain large enough to stay in the sample.

II.3 Sensitivity analysis, identification and inference

We now provide a thorough discussion of the different threats to the validity of our estimates: we first consider potential failures of the exclusion restriction and we discuss identification in our shift-share design (Borusyak, Hull and Jaravel, 2018); we then discuss inference following the recent contribution of Adão, Kolesár and Morales (2019).

Identification We interpret our estimates as the effect of immigration on manufacturing production. One concern is that the shift-share instrument, which is a combination of agricultural income shocks and migration patterns, may have independent effects on production in cities. In other words, the exclusion restriction may fail. First, destinations could be affected by the commodity price shock through the market for intermediary goods. Rural hinterlands may be producing goods which directly enter the production of final goods in urban centers, e.g., cotton. In Panel A of Table 5, we reproduce Table 3 but exclude industries that use agricultural products as intermediate inputs. In Panel B, we add 2-digit industry fixed effects as controls in specification (5). Second, if rural dwellers consume the final goods produced in urban areas, agricultural income shocks would affect the demand for manufacturing goods. To address this concern, we exclude from the analysis all migration flows between prefectures that are less than 300 km apart (Panel C). We also control for the shocks in neighboring prefectures weighted by the inverse of distance in Panel D. Assuming that trade follows a gravity model, this specification allows us to control for rural-urban spillovers through goods markets, so that the identification only comes from idiosyncratic variation in migration patterns (not related to distance, but due for example to historical events, see Kinnan, Wang and Wang, 2018). We perform two other robustness checks: we control for a measure of market access—the sum of the rural population in all prefectures weighted by the inverse of the distance to the prefecture where the firm is located (Panel E),—and we restrict the sample to exporting firms at baseline, arguably less exposed to variation in local demand for their final product (Panel F). Finally, the shares of our shift-share design reflect historical migration patterns (Jaeger, Ruist and Stuhler, 2018); our estimates may be conflating the effect of current migration shocks with the lagged effect of past shocks. To address this, we control for the stock of immigrants at baseline and allow it to have independent effects on outcomes (Panel G). In all instances, the estimates are similar to the main results, which provides reassurance that our estimates do capture the effect of current migration on production.[31]

A recent literature discusses identification in a shift-share design (Borusyak, Hull and Jaravel, 2018; Goldsmith-Pinkham, Sorkin and Swift, forthcoming). As explained in Section I, we follow Borusyak, Hull and Jaravel (2018): we invert the transformation of shifts into a shift-share variable, and we instead consider transformed outcomes at the level of shifts. In our setting, this amounts to estimating

[31]The effect of immigration on firms itself could be multifaceted. Our preferred interpretation is that new workers affect labor markets and the relative abundance of production factors in cities. However, new workers in cities are also consumers of non-tradable goods, which may benefit firms providing these goods (e.g., housing services) or affect firms relying on these goods or services (e.g., with a highly land-intensive production). We indeed provide some evidence that rural-urban migration affects living standards at destination (see Appendix E.3, exploiting a survey of urban workers).

Table 5. Impact of migration inflows on urban firms—sensitivity analysis.

	Labor cost (1)	Employment (2)	K/L ratio (3)	Y/L ratio (4)
Panel A: Excluding industries that process agricultural goods				
Migration	-0.139	0.276	-0.408	-0.428
	(0.064)	(0.050)	(0.093)	(0.105)
Observations	29,047	29,047	29,047	29,047
Panel B: Controlling for industry fixed effects				
Migration	-0.104	0.274	-0.472	-0.365
	(0.074)	(0.056)	(0.109)	(0.113)
Observations	31,886	31,886	31,886	31,886
Panel C: Excluding migration within a 300-km radius				
Migration	-0.172	0.345	-0.506	-0.513
	(0.071)	(0.065)	(0.115)	(0.129)
Observations	31,886	31,886	31,886	31,886
Panel D: Controlling for shocks in neighboring prefectures				
Migration	-0.241	0.276	-0.838	-0.437
	(0.267)	(0.186)	(0.467)	(0.368)
Observations	31,886	31,886	31,886	31,886
Panel E: Controlling for market access				
Migration	-0.167	0.297	-0.460	-0.440
	(0.066)	(0.053)	(0.102)	(0.112)
Observations	31,886	31,886	31,886	31,886
Panel F: Keeping only exporting firms				
Migration	-0.132	0.257	-0.388	-0.325
	(0.061)	(0.057)	(0.104)	(0.094)
Observations	10,653	10,653	10,653	10,653
Panel G: Controlling for the stock of immigrants at baseline				
Migration	-0.110	0.257	-0.559	-0.497
	(0.088)	(0.072)	(0.144)	(0.147)
Observations	31,886	31,886	31,886	31,886

Notes: Standard errors are clustered at the prefecture level and reported between parentheses. The sample is composed of the firms present every year in the NBS firm census between 2000 and 2006. See Section I and Equation (5) for a description of the IV specification.

Equation (6), which regresses a weighted average of firm outcomes across origins on a weighted average of immigration instrumented by the shifts, using migration patterns as weights. The transformed specification can be interpreted as estimating the effect of push shocks on outcomes at the "typical" destination, across the different origins. We present the OLS estimates in Panel A of Table 6 and the 2SLS estimates in Panel B. Consistent with the equivalence result of Borusyak, Hull and Jaravel (2018), the origin-level point estimates are identical to our destination-level results (Table 3). In Panel C, we show a reduced-form specification where transformed firm outcomes are directly explained by the agricultural income shock. This reduced form approach offers the opportunity to correlate firm outcomes with counterfactual shocks. In Panel D, we provide a test of the parallel trends assumption and regress pre-treatment differences in outcomes (1998–2000) on the shocks computed over the study period (2000–2005). We do not find strong evidence that urban centers whose hinterlands are exposed to agricultural shocks between 2000 and 2005 follow different trends than others before 2000. The estimates on employment and value added per worker are significantly different from zero, but are much smaller in magnitude as compared to the reduced-form estimates. Finally, we perform a placebo regression with lagged agricultural income shocks (computed over 1993–1998) as the independent variable. This specification tests whether past shocks and previous immigration waves predict the future evolution of outcomes at destination through a sluggish restructuring of production (Jaeger, Ruist and Stuhler, 2018). We do find some persistent effect of earlier price (or migration) shocks on urban wages (Panel E), but the effect goes in the opposite direction to our main result (Panel C).

Inference The clustering of standard errors in our main specification does not account for the heteroskedasticity induced by (i) the correlation between outcomes across prefectures with similar exposure to shocks and (ii) the correlation between agricultural income shocks across origins. Appendix F.2 provides a sensitivity analysis for inference. We first compute robust standard errors, standard errors clustered at the level of the province of destination, and use a more continuous modeling of spatial auto-correlation (following Conley, 1999). As argued by Adão, Kolesár and Morales (2019), however, the heteroskedasticity induced by shift-share designs may not be adequately captured by spatial clustering. For instance, migrants from the same origin may join similar industries across different destinations, so that firms in these destinations experience correlated shocks even if they are not geographically close. We thus use the inference method proposed by Adão, Kolesár and Morales (2019) and report the AKM and AKM0 standard errors. Another concern is that the

Table 6. Impact of migration inflows on urban firms—transformation from destinations to origins following Borusyak, Hull and Jaravel (2018).

	Labor cost (1)	Employment (2)	K/L ratio (3)	Y/L ratio (4)
Panel A: OLS				
Transformed migration	-0.173	0.295	-0.436	-0.459
	(0.019)	(0.013)	(0.018)	(0.017)
Observations	335	335	335	335
Panel B: IV				
Transformed migration	-0.147	0.294	-0.431	-0.437
	(0.029)	(0.019)	(0.024)	(0.026)
Observations	335	335	335	335
F-stat. (first stage)	102.9	102.9	102.9	102.9
Panel C: Reduced form				
Price shock	0.232	-0.465	0.683	0.692
	(0.053)	(0.049)	(0.084)	(0.084)
Observations	335	335	335	335
Panel D: Parallel trends (1998–2000)				
Price shock	0.004	-0.102	0.025	0.178
	(0.032)	(0.023)	(0.031)	(0.050)
Observations	335	335	335	335
Panel E: Lagged Shocks (1993–1998)				
Price shock	-0.199	-0.150	0.202	0.026
	(0.079)	(0.080)	(0.115)	(0.116)
Observations	335	335	335	335

Notes: Robust standard errors are reported between parentheses. The sample is composed of the 335 prefectures of origin. The dependent variables and *Transformed migration* are variables at destination (outcomes and the immigration rate) transformed into variables across origins, through a combination of (i) migration patterns between origins and destinations, and (ii) employment levels within destinations at baseline (see Section I).

shifts are spatially auto-correlated across prefectures of origin, so that there is less independent variation than a destination-level analysis would suggest. The transformation suggested by Borusyak, Hull and Jaravel (2018) offers the opportunity to better account for spatial correlation across origins. In Appendix F.2, we estimate Equation (6) with standard errors clustered at the province level and standard errors accounting for spatial auto-correlation (Conley, 1999). Our estimates remain precisely estimated regardless of the inference method used.

III Restructuring of production

This section characterizes the restructuring of production following the arrival of new workers. Specifically, we investigate whether manufacturing firms change their output mix, whether they shift patenting away from capital-enhancing and skill-enhancing technologies, and how much of the observed adjustments in factor use and patenting can be explained by endogenous product choice.

Product choice We investigate changes in production lines, which we measure via changes in the (main) end product.[32] We classify products based on their product class, as defined by a unique product description or by the product code assigned by our algorithm (see Section I). We proxy the skill intensity of each product class by the average share of the workforce with a high-school degree among firms whose main products belong to this class at baseline and capital intensity by the average capital-to-labor ratio among firms whose main products belong to this class at baseline. We estimate specification (5) at the establishment level between 2000 and 2006, and we control for fixed effects at the level of the product code at baseline in order to capture the adoption of new products while keeping fixed the initial distribution of products across destinations.

We first use the textual description to detect any change in the (main) end product between 2000 and 2006, and we determine the direction of the change using the characteristics of the average establishment providing the same description at baseline. Panel A of Table 7 presents the results. Establishments in prefectures that experience large immigration flows are more likely to change the textual description of their main product in 2000–2006 (column 1); a ten percentage point increase in the immigration rate raises the probability to provide two distinct descriptions by 2.1 percentage points. The effect is mostly driven by a transition towards products with lower human capital (columns 2 and 3) and lower physical capital intensity (columns 4 and 5). Distinct textual descriptions could refer to the same product

[32] While our baseline analysis focuses on the first product declared by firms, we exploit the reporting of up to *three* products in Appendix E.4. We estimate the effect of immigration on the number of products and changes in their "similarity", as measured with linguistic distance or through input/output accounts. We find little effect on either outcome.

Table 7. Impact of migration inflows on urban firms—production restructuring.

Change in product	Any (1)	High ed. (2)	Low ed. (3)	High K/L (4)	Low K/L (5)
Panel A: Using unique descriptions to define product categories					
Migration	0.205	-0.001	0.206	0.049	0.157
	(0.049)	(0.028)	(0.036)	(0.038)	(0.026)
Observations	27,062	27,062	27,062	27,062	27,062
Panel B: Using HS codes to define product categories					
Migration	0.120	0.001	0.118	-0.011	0.131
	(0.035)	(0.022)	(0.030)	(0.024)	(0.029)
Observations	27,062	27,062	27,062	27,062	27,062

Notes: Standard errors are clustered at the prefecture level and reported between parentheses. The sample is composed of the firms present every year in the NBS firm census between 2000 and 2006 and whose main product was matched to a HS6 code by the NLP algorithm described in Section I. *Migration* is the immigration rate, i.e., the migration flow divided by population at destination and at baseline. The dependent variables are dummy variables equal to one if there is any change in the main product (1), and if this change goes toward products manufactured by establishments with a more (2) or less (3) educated workforce, and by more (4) or less (5) capital-abundant establishments. See Section I and Equation (5) for a description of the IV specification.

class: we next exploit the Natural Language Processing algorithm described in Section I to assign a unique HS-6 product category to each description. We replicate the previous exercise with these newly-defined product categories in Panel B of Table 7. A ten percentage point increase in the immigration rate raises the probability to change HS codes in 2000–2006 by 1.2 percentage points (column 1) a 2% decrease from an average of 61%. The change is towards products that are typically found in low-skilled (column 3) and labor-abundant firms (column 5).

We provide further results on endogenous product choice in Appendix E.4. First, we leverage the linguistic similarity score between two HS product codes, as provided by our textual analysis. We compute the human or physical capital intensity of a given product code as the weighted average of capital intensity across all firms weighted by the similarity score with their own product code. This method allows for a more continuous product characterization than the first classification we use, which attributes a score of 1 when product classes coincide and 0 otherwise. Second, we compute the human or physical capital intensity of each product code using an input/output matrix rather than language similarity. This takes into account the fact that products produced by capital-intensive firms may rely on intermediary inputs provided by labor-intensive firms. Our conclusion remains the same with these alternative measures of product change: there is an adjustment of production lines toward low-skilled, labor-intensive production. We also estimate possible changes in the technological content of products by exploiting a measure of cross-industry patent citations in the United States (Bloom, Schankerman and Van Reenen, 2013). We find evidence that firms re-orient their production towards products that are less reliant on technological innovation (i.e., with fewer and more concentrated citations across industries).

Innovation We now use a more direct observation of technological change within firms, through their patenting behavior. We exploit the match provided by He et al. (2018) between the NBS sample of manufacturing establishments and patents submitted to the State Intellectual Property Office (renamed as China National Intellectual Property Administration in 2018). The description of each patent provides a detailed classification of its technological content. We use this classification to qualify the nature of technological innovation, using average characteristics of firms that submitted a patent within each subcategory at baseline. Specifically, we classify a patent as high-education if the average share of the workforce with a high school degree among firms that submitted patents at baseline was above the median. Similarly we classify a patent as high-capital-to-labor ratio if firms that submitted patents in the same class had above-median capital-to-labor ratio on average. This exercise assumes that: (i) capital-intensive firms primarily patent capital-augmenting technologies; (ii) technology is homogeneous within a patent class.

We estimate specification (5) at the establishment-level between 2000 and 2006, and regress the difference in the probability to submit a patent application between 2000 and 2006 on the immigration rate, instrumented by the shift-share instrument. Panel A of Table 8 shows that a ten percentage point increase in the immigration rate decreases the probability to submit a patent by 0.47 percentage points (from an average of 4.7 percentage points, a reduction of exactly 10%). The effect is mostly driven by the innovation and utility patent categories (see columns 3 to 4), which suggests a decline in fundamental innovation and in the creation of new production processes. Rural-urban migration does not only affect the pace of technological progress, but also its direction. The drop in patenting is most pronounced for skill-enhancing and capital-enhancing technologies (see Panel B of Table 8). In response to an inflow of unskilled labor, manufacturing establishments are less likely to push the technological frontier, especially towards technologies that use more skilled labor

Table 8. Impact of migration inflows on urban firms—technological innovations.

New patent	Any (1)	Design (2)	Innovation (3)	Utility (4)
Panel A: Patent categories				
Migration	-0.047 (0.016)	-0.006 (0.010)	-0.035 (0.012)	-0.027 (0.015)
Observations	31,886	31,886	31,886	31,886
New patent	High education (1)	Low education (2)	High K/L (3)	Low K/L (4)
Panel B: Patent characteristics				
Migration	-0.045 (0.017)	-0.018 (0.012)	-0.046 (0.014)	-0.014 (0.012)
Observations	31,886	31,886	31,886	31,886

Notes: Standard errors are clustered at the prefecture level and reported between parentheses. The sample is composed of the firms present every year in the NBS firm census between 2000 and 2006. *Migration* is the immigration rate, i.e., the migration flow divided by population at destination and at baseline. The dependent variable is the difference in the probability to submit a patent application between 2000 and 2006. In Panel A (columns 2-4), we distinguish three categories of patents: design (external appearance of the final product); innovation (fundamental innovations in methods); and utility (changes in processing, shape or structure of products). In Panel B, we divide patents into technologies associated with high/low average human capital, and labor-abundant technologies versus capital-abundant ones. See Section I and Equation (5) for a description of the IV specification.

or capital.

Product choice, factor use, and patenting We have documented a restructuring of production through (a) the adjustment of factor use and technology, and (b) the adoption of low-skilled, labor-intensive products. We now quantify the role of product choice in the adoption of more labor-intensive organizational forms. We regress changes in factor use and patenting on a dummy for firms that changed their main product in 2000–2006, on the immigration rate, and on the interaction between product change and immigration. The immigration rate and its interaction are instrumented by the shift-share instrument and by its interaction with the product change dummy. Since product change is affected by immigration and by many unobserved factors (e.g., international demand), the estimates do not have a causal interpretation but are only suggestive of the mediation effect of product choice.

We present the results of this specification in Table 9. The coefficients on the

Table 9. Impact of migration inflows on urban firms—changes in product, and changes in factor use and technology.

	K/L Ratio	Patents		
		Any new	Any innovation	Any utility
	(1)	(2)	(3)	(4)
Product change	0.084	0.053	0.026	0.033
	(0.037)	(0.015)	(0.012)	(0.013)
Migration	-0.353	0.018	-0.023	0.018
	(0.141)	(0.031)	(0.019)	(0.023)
Migration × Product change	-0.123	-0.103	-0.027	-0.078
	(0.111)	(0.035)	(0.023)	(0.028)
Observations	27,062	27,062	27,062	27,062
F-stat. (first)[a]	10.42	10.42	10.42	10.42

Notes: Standard errors are clustered at the prefecture level and reported between parentheses. The sample is composed of the firms present every year in the NBS firm census between 2000 and 2006 and whose main product was matched to a HS6 code by the NLP algorithm described in Section I. *Migration* is the immigration rate, i.e., the migration flow divided by population at destination and at baseline. *Product change* is a dummy variable equal to one for firms that changed their main product in 2000–2006. See Section I and Equation (5) for a description of the IV specification. In columns 3 and 4, we consider two patent categories: innovation (fundamental innovations in methods); and utility (changes in processing, shape or structure of products).
[a] The IV specification uses two endogenous variables and two instruments; the critical value for weak instruments is then 7.03 (at 10%).

product change dummy across columns 1 to 4 suggest that, absent immigration, a change in product is associated with a shift towards capital-intensive technologies. Following an immigration shock, all firms adjust their input mix by using more labor and less capital, whether they change products or not (see the estimate on 'Migration', column 1). In stark contrast, the decline in patenting only occurs among firms that change their output mix, especially for patents related to production processes (see columns 2–4). These findings shed new light on the adjustment of technology to the supply of the different production factors. One margin of adjustment occurs keeping the same output mix (Beaudry and Green, 2003). Another important margin of adjustment, identified by Goldberg et al. (2010) and Bas and Paunov (2019) in the response to new imported inputs, derives from endogenous product choice.

IV Conclusion

This paper provides unique evidence on the causal effect of rural-urban migration on manufacturing production in China. We combine information on migration flows with longitudinal data on manufacturing establishments between 2000 and 2006, a period of rapid structural transformation and sustained manufacturing growth. We instrument immigrant flows using a shift-share design, which combines shocks to agricultural income due to cropping patterns and fluctuations in international crop prices with historical migration patterns between rural and urban areas. We find that migration decreases labor costs and increases employment, and that manufacturing production becomes more labor-intensive, as capital does not adjust. We are also able to document the reorganization of production with unique data on product choice and patent applications. Our results show that the abundance of rural migrant labor induces labor-oriented directed technological change and the adoption of labor-intensive product varieties among manufacturing firms. This mechanism is likely at play in other countries that are currently in the process of structural transformation. Over the last decade, China has experienced a sharp trend reversal with slower rural-urban migration and faster automation in manufacturing (Cheng et al., 2019).

Our empirical setting, which provides a unique opportunity to study the effect of rural migrants on manufacturing firms, also has important limitations. First, we do not directly observe changes in the skill composition of the workforce within firms. Second, the identification of causal estimates relies on unexpected variation in rural-urban migration, which may induce a different adjustment by firms than long-term trends in labor supply.[33] Third, rural migrants may impact other markets than labor markets, e.g., through demand for non-tradable goods at destination. Fourth, our difference-in-difference setting implies that the estimates cannot be extrapolated at the level of the country without understanding the magnitude of general equilibrium effects (Adão, Arkolakis and Esposito, 2019; Beraja, Hurst and Ospina, 2019). Finally, we only observe manufacturing firms; our analysis cannot shed light on the implications of migration on aggregate productivity, or on productivity gaps across sectors. We see these issues as promising avenues for future research.

[33]While shocks to rural-urban migration may trigger a different response, these agricultural shocks have little effect on the characteristics of the average migrant, as we show in Appendix C.3.

References

Abebe, Girum, Stefano Caria, Marcel Fafchamps, Paolo Falco, Simon Franklin, and Simon Quinn. 2016. "Anonymity or Distance? Job Search and Labour Market Exclusion in a Growing African City." CSAE Working Paper Series.

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

Adão, Rodrigo, Costas Arkolakis, and Federico Esposito. 2019. "Spatial linkages, global shocks, and local labor markets: Theory and evidence." National Bureau of Economic Research.

Adão, Rodrigo, Michal Kolesár, and Eduardo Morales. 2019. "Shift-share designs: Theory and inference." *The Quarterly Journal of Economics*, 134(4): 1949–2010.

Akcigit, Ufuk, Salome Baslandze, and Stefanie Stantcheva. 2016. "Taxation and the international mobility of inventors." *American Economic Review*, 106(10): 2930–81.

Alfonsi, Livia, Oriana Bandiera, Vittorio Bassi, Robin Burgess, Imran Rasul, Munshi Sulaiman, and Anna Vitali. 2017. "Tackling Youth Unemployment: Evidence from a Labour Market Experiment in Uganda." STICERD, LSE 64.

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

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

Atkin, David, Azam Chaudhry, Shamyla Chaudry, Amit K Khandelwal, and Eric Verhoogen. 2017. "Organizational barriers to technology adoption: Evidence from soccer-ball producers in Pakistan." *The Quarterly Journal of Economics*, 132(3): 1101–1164.

Banerjee, Abhijit V., and Esther Duflo. 2014. "Do Firms Want to Borrow More? Testing Credit Constraints Using a Directed Lending Program." *Review of Economic Studies*, 81(2): 572–607.

Bas, Maria, and Caroline Paunov. 2019. "What gains and distributional implications result from trade liberalization?"

Bazzi, Samuel. 2017. "Wealth heterogeneity and the income elasticity of migration." *American Economic Journal: Applied Economics*, 9(2): 219–55.

Beaudry, Paul, and David A Green. 2003. "Wages and employment in the United States and Germany: What explains the differences?" *American Economic Review*, 93(3): 573–602.

Beraja, Martin, Erik Hurst, and Juan Ospina. 2019. "The aggregate implications of regional business cycles." *Econometrica*, 87(6): 1789–1833.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen. 2013. "Identifying technology spillovers and product market rivalry." *Econometrica*, 81(4): 1347–1393.

Boehm, Johannes, and Ezra Oberfield. 2020. "Misallocation in the Market for Inputs: Enforcement and the Organization of Production." C.E.P.R. Discussion Papers CEPR Discussion Papers 14482.

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

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2018. "Quasi-experimental shift-share research designs." National Bureau of Economic Research.

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

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

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

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

Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli. 2016. "Agricultural productivity and structural transformation: Evidence from Brazil." *American Economic Review*, 106(6): 1320–65.

Bustos, Paula, Juan Manuel Castro Vincenzi, Joan Monras, and Jacopo Ponticelli. 2018. "Structural Transformation, Industrial Specialization, and Endogenous Growth." C.E.P.R. Discussion Papers CEPR Discussion Papers 13379.

Card, David. 2001. "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration." *Journal of Labor Economics*, 19(1): 22–64.

Card, David, and John DiNardo. 2000. "Do Immigrant Inflows Lead to Native Outflows?" *American Economic Review*, 90(2): 360–367.

Cheng, Hong, Ruixue Jia, Dandan Li, and Hongbin Li. 2019. "The Rise of Robots in China." *Journal of Economic Perspectives*, 33(2): 71–88.

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

Conley, Timothy G. 1999. "GMM estimation with cross sectional dependence." *Journal of Econometrics*, 92(1): 1–45.

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

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

El Badaoui, Eliane, Eric Strobl, and Frank Walsh. 2017. "Impact of Internal Migration on Labor Market Outcomes of Native Males in Thailand." *Economic Development and Cultural Change*, 66: 147–177.

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

Feng, Shuaizhang, Yingyao Hu, and Robert Moffitt. 2017. "Long run trends in unemployment and labor force participation in urban China." *Journal of Comparative Economics*, 45(2): 304 – 324.

Fields, Gary S. 1975. "Rural-urban migration, urban unemployment and underemployment, and job-search activity in LDCs." *Journal of development economics*, 2(2): 165–187.

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

Giulietti, Corrado, Guangjie Ning, and Klaus F Zimmermann. 2012. "Self-employment of rural-to-urban migrants in China." *International Journal of Manpower*, 33(1): 96–117.

Goldberg, Pinelopi Koujianou, Amit Kumar Khandelwal, Nina Pavcnik, and Petia Topalova. 2010. "Imported intermediate inputs and domestic product growth: Evidence from India." *The Quarterly journal of economics*, 125(4): 1727–1767.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. forthcoming. "Bartik Instruments: What, When, Why, and How." *American Economic Review.*

Gollin, Douglas, David Lagakos, and Michael E. Waugh. 2014. "The Agricultural Productivity Gap." *The Quarterly Journal of Economics*, 129(2): 939–993.

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

Gröger, André, and Yanos Zylberberg. 2016. "Internal Labor Migration as a Shock Coping Strategy: Evidence from a Typhoon." *American Economic Journal: Applied Economics*, 8(2): 123–53.

Hall, Robert E., and Charles I. Jones. 1999. "Why do Some Countries Produce So Much More Output Per Worker than Others?" *The Quarterly Journal of Economics*, 114(1): 83–116.

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

Hau, Harald, Yi Huang, and Gewei Wang. 2018. "Firm response to competitive shocks: Evidence from China's minimum wage policy." *SFI Research Paper*, , (16-47).

Herrendorf, Berthold, Christopher Herrington, and Akos Valentinyi. 2015. "Sectoral Technology and Structural Transformation." *American Economic Journal: Macroeconomics*, 7(4): 104–33.

He, Zi-Lin, Tony W Tong, Yuchen Zhang, and Wenlong He. 2018. "A database linking Chinese patents to China's census firms." *Scientific data*, 5: 180042.

Hoberg, Gerard, and Gordon Phillips. 2016. "Text-based network industries and endogenous product differentiation." *Journal of Political Economy*, 124(5): 1423–1465.

Howitt, Peter. 2000. "Endogenous Growth and Cross-Country Income Differences." *American Economic Review*, 90(4): 829–846.

Imbert, Clément, and John Papp. 2019. "Short-term Migration, Rural Public Works, and Urban Labor Markets: Evidence from India." *Journal of the European Economic Association*, 18(2): 927–963.

Jaeger, David A, Joakim Ruist, and Jan Stuhler. 2018. "Shift-Share Instruments and the Impact of Immigration." C.E.P.R. Discussion Papers CEPR Discussion Papers 12701.

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

Kinnan, Cynthia, Shing-Yi Wang, and Yongxiang Wang. 2018. "Access to Migration for Rural Households." *American Economic Journal: Applied Economics*, 10(4): 79–119.

Kleemans, Marieke, and Jeremy Magruder. 2018. "Labour Market Responses To Immigration: Evidence From Internal Migration Driven By Weather Shocks." *Economic Journal*, 128(613): 2032–2065.

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

Lewis, Ethan. 2011. "Immigration, Skill Mix, and Capital Skill Complementarity." *The Quarterly Journal of Economics*, 126(2): 1029–1069.

Li, Shen, Zhe Zhao, Renfen Hu, Wensi Li, Tao Liu, and Xiaoyong Du. 2018. "Analogical Reasoning on Chinese Morphological and Semantic Relations." 138–143. Association for Computational Linguistics.

Marden, Sam. 2015. "The agricultural roots of industrial development: "forward linkages" in reform era China." Manuscript.

Mayneris, Florian, Sandra Poncet, and Tao Zhang. 2018. "Improving or disappearing: Firm-level adjustments to minimum wages in China." *Journal of Development Economics*, 135: 20–42.

Meghir, Costas, Renata Narita, and Jean-Marc Robin. 2015. "Wages and informality in developing countries." *American Economic Review*, 105(4): 1509–46.

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

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

Minale, Luigi. 2018. "Agricultural productivity shocks, labour reallocation and rural-urban migration in China." *Journal of Economic Geography*, lby013.

Mitaritonna, Cristina, Gianluca Orefice, and Giovanni Peri. 2017. "Immigrants and firms' outcomes: Evidence from France." *European Economic Review*, 96: 62–82.

Monras, Joan. 2020. "Immigration and wage dynamics: Evidence from the mexican peso crisis." *Journal of Political Economy*, 128(8): 3017–3089.

Monte, Ferdinando, Stephen J Redding, and Esteban Rossi-Hansberg. 2018. "Commuting, migration, and local employment elasticities." *American Economic Review*, 108(12): 3855–90.

Moser, Petra, Alessandra Voena, and Fabian Waldinger. 2014. "German Jewish émigrés and US invention." *American Economic Review*, 104(10): 3222–55.

Oberfield, Ezra, and Devesh Raval. 2014. "Micro data and macro technology."

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

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

Santangelo, Gabriella. 2016. "Firms and Farms: The Impact of Agricultural Productivity on the Local Indian Economy." Manuscript.

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

Tombe, Trevor, and Xiaodong Zhu. 2019. "Trade, Migration, and Productivity: A Quantitative Analysis of China." *American Economic Review*, 109(5): 1843–72.

Ulyssea, Gabriel. 2018. "Firms, Informality, and Development: Theory and Evidence from Brazil." *American Economic Review*, 108(8): 2015–47.

Verhoogen, Eric. 2020. "Firm-Level Upgrading in Developing Countries." Mimeo.

Verhoogen, Eric A. 2008. "Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector." *The Quarterly Journal of Economics*, 123(2): 489–530.

Zhang, Junfu, and Zhong Zhao. 2015. "Social-family network and self-employment: evidence from temporary rural-urban migrants in China." *IZA Journal of Labor & Development*, 4(1): 4.