



**HAL**  
open science

## Trois essais sur l'économie de la migration

Marlon Seror

► **To cite this version:**

Marlon Seror. Trois essais sur l'économie de la migration. Economies et finances. Université Paris sciences et lettres, 2017. Français. NNT : 2017PSLEE089 . tel-02076678

**HAL Id: tel-02076678**

**<https://theses.hal.science/tel-02076678>**

Submitted on 22 Mar 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres  
PSL Research University

Préparée à l'École Normale Supérieure,  
à l'École d'Économie de Paris et à l'IRD-DIAL

Trois essais sur l'économie de la migration

**Ecole doctorale n°465**

EPS – Économie Panthéon-Sorbonne

**Spécialité** SCIENCES ÉCONOMIQUES

**Soutenue par Marlon SEROR**  
**le 04 décembre 2017**

Dirigée par Flore GUBERT

## COMPOSITION DU JURY :

Mme DÉMURGER Sylvie  
CNRS et GATE Lyon Saint-Étienne,  
Rapporteur

M. FREEMAN Richard  
Harvard University et NBER, Rapporteur

M. GURGAND Marc  
CNRS, École Normale Supérieure et  
École d'Économie de Paris, Membre du  
jury

Mme MESNARD Alice  
City, University of London, Présidente du  
jury



PARIS SCIENCES & LETTRES

# P H. D. T H E S I S

submitted for the degree of

Doctor of Philosophy of Paris Sciences & Lettres in Economics

prepared at the Paris School of Economics, École Normale Supérieure and  
IRD-DIAL, and defended at the Paris School of Economics on Dec. 4, 2017

by

Marlon SEROR

## Three Essays on the Economics of Migration

Thesis Advisor: Flore GUBERT

### **Jury:**

<i>Reviewers:</i>	Sylvie DÉMURGER	-	GATE, CNRS
	Richard B. FREEMAN	-	Harvard University, NBER
<i>Advisor:</i>	Flore GUBERT	-	PSE, IRD-DIAL
<i>Examiners:</i>	Marc GURGAND	-	PSE, ENS, CNRS
	Alice MESNARD (Chair)	-	City, University of London



## Acknowledgments

This thesis would not be in your hands without my supervisor's support. Flore Gubert encouraged me to apply for the Ph.D. program, through all the (very different) versions of my proposal, however remote from her own research interests. She also supported me in my desire to partly shift my focus from Africa to China. I subsequently spent a lot of time far away from PSE and DIAL, in Hong Kong first, thanks to Chen Juan's enthusiasm and trust, and the warm welcome of the Hong Kong Polytechnic University, and then to Harvard University, the unlikely Mecca of Chinese studies—not least owing to Richard Freeman's mentorship of Chinese scholars and support. I am also indebted to Marc Gurgand, my ENS tutor, who oversaw my transition from ancient Greek and classical Chinese to economics, was briefly my Ph.D. supervisor and has always, just as Flore or Richard, been available when I needed help, advice or (always last-minute) recommendation letters. My thanks also go to my reviewers—Sylvie Démurger and Richard—and all the members of my Ph.D. jury—Marc, Alice Mesnard and Flore,—who have already given me (yes, these words *were* part of the original submission) so many comments that I wonder what we will be talking about at the defense. I also received selfless time, advice and (nasty) jokes from Stephan Hebllich, Clement Imbert and Yanos Zylberberg, all co-authors on papers of mine, as well as Zhang Yifan and Xu Hao. (Special kudos to Yanos for being, beside the jury, the first—and hopefully not the last—to read all three chapters of the thesis.) Here I should also express my gratitude to Flore and Jean-Noël Senne as co-authors on the “ghost” chapter of this thesis, the invisible beneficiary of a large portion (most?) of my Ph.D. time. And to all my DIAL colleagues, to whom I bequeath a beautiful *kakemono* that will prove precious on the southern side of the *labo* in the summer: Elçin Sarıkaya, Marion Mercier, Mohammed Ali Marouani, Estelle Mousson Jamel Koussoubé, Delia Visan (office mates first, in order of appearance), Oscar Barrera, Anne Boring, Thomas Calvo, Lisa Chauvet, Virginie Comblon, Raphaël Cottin, Anda David, Esther Delesalle, Danielle Delmas, Axel Demenet, Siyavash Eslami Loutij, Marin Ferry, Yeganeh Forouheshfar, Véronique Gille, Anne Hilger, Hoàng T. Quỳnh for being always game for a coffee break (and drawing cats), Ana Horigoshi, Lê T. Linh, Anne Legendre, Loïc Lepezennec, Sandrine Mesplé-Somps, Björn Nilsson, Christophe Jalil Nordman, Xavier Oudin, Laure Pasquier-Doumer,

Mireille Razafindrakoto, Anne-Sophie Robilliard, François Roubaud, Camille Saint-Macary, Gilles Spielvogel, Marine de Talancé, Jean-Michel Wachsberger and Claire Zanuso. I also want to thank all those who made my stay at Harvard and the NBER such a rich and enjoyable experience: Gökhan Aykaç, Jennifer Amadeo-Holl (who I hope will someday forgive me for all the extra work), Maurice Dalton, Kaushik Ghosh, Len McCain, Song Ran, Susan Stewart, Anna Raute and Camille Terrier (for coffee and endless *promenades digestives* by the Charles River), Vivian Xu, You Yang, and Lorette Baptiste, Elaine Bernard and John “Jack” Trumpbour of the Labor and Worklife Program. Outside the office, but no less important, in Cambridge: Kamran Hakiman, Saurabh Mhatre, Judith Rodríguez, our Huron Village neighbors (except John Malkovich, who wasn’t around much), the construction crew on Blakeslee Street and the Fire Department of Cambridge, MA (without whom this thesis and its author would have vanished into thin air), Mo Lotman and Anna Raute for their hospitality and conversation, Xu Yi’An (whom I forgive for no longer speaking French) and Juliette Yorinski (who shared my taste for the Film Archive). I should add Véronique Guillotin, Béatrice Havet and Éliane Madelaine for their outstanding support. I also enjoyed support from and fruitful discussions with Hiral Arjan (for putting research into political perspective, although I swear I have no responsibility in Brexit), Isabelle Chort, Julian Clarke, Paul Dutronc-Postel, Laure Haggiage, Huang Yang (whom I did not ask for advice for my thesis title, as I was worried he might retaliate), Romain Lafarguette, Sylvie Lambert, Jessica Leight, Mushfiq Mobarak, Cem Özgüzel, Albert Park, Fanny Valeyre, Oliver Vanden Eynde, Margaux Vinez and Zhang Xiaobo. I would like to thank Vincent Fertey and my SciencesPo students for a humbling and fantastic teaching experience, as well as Ted Miguel and UC Berkeley, where the first chapter of this thesis was started, and Yu Jiantuo and the China Development Research Foundation. I benefited for my Ph.D. from funding from the ENS, and additional financial help from the PSE International Mobility Grant and the Arthur Sachs Scholarship at Harvard. I would like to thank my old “Alsatian” friends and overlapping Ph.D. inmates for listening to my academic doubts and yearnings: Michael Murez, David Serfass and Nathan Sperber. This thesis—and the person typing these words, for that matter—could not have happened without my mother, my father and my (many) grandparents, probably all relieved (if everything goes well on December 4<sup>th</sup>) to learn that I am no longer a student.

I also thank my sister and brother, who may not even believe it could ever happen. Last but not least, and until the very last question mark at the very end of this sentence, I have enjoyed the unflinching, supportive, enthusiastic, critical, inspiring, forgiving, loud, quiet, stern, refreshing, exhilarating, artistic, academic, remote, close, indispensable and multifaceted presence of—who but my better half?





# Contents

<b>Introduction</b>	<b>1</b>
<b>I Migrants' Beliefs and Investments</b>	<b>11</b>
1 Introduction . . . . .	12
2 Remittances and information . . . . .	17
3 Theoretical framework . . . . .	18
4 Data . . . . .	26
5 Results . . . . .	28
5.1 Predictions 1 and 2: Positive discrepancies . . . . .	28
5.2 Predictions 3 to 5: Heterogeneous beliefs and learning . . . . .	29
5.3 Robustness checks . . . . .	34
5.4 Alternative interpretations . . . . .	41
6 Discussion and conclusion . . . . .	43
6.1 Policy implications . . . . .	43
6.2 Conclusion . . . . .	44
Tables . . . . .	47
Appendices . . . . .	62
I.A Additional tables . . . . .	62
I.B Remittances and information . . . . .	70
I.C 2SLS strategy . . . . .	73
I.D Measurement error . . . . .	76
I.D.1 Cluelessness . . . . .	77
I.D.2 Measurement error in the test of Prediction 2 . . . . .	78
I.D.3 Focusing on $\delta_O$ . . . . .	80
<b>II Migrants and Firms</b>	<b>81</b>
1 Introduction . . . . .	82
2 Predicting migration to urban areas . . . . .	87
2.1 Migration flows . . . . .	88
2.2 Shocks to agricultural labor returns . . . . .	89
2.3 Predicting rural-to-urban migration flows . . . . .	92
3 Empirical strategy . . . . .	96

---

3.1	Data . . . . .	96
3.2	Accounting framework . . . . .	97
3.3	Empirical strategy . . . . .	100
4	Results . . . . .	101
4.1	The labor supply shift . . . . .	101
4.2	Returns to factors . . . . .	103
4.3	Dynamic effects . . . . .	104
4.4	Reallocation of resources across firms . . . . .	105
4.5	Aggregate effects . . . . .	106
4.6	Composition effects . . . . .	107
5	Conclusion . . . . .	109
	Figures and tables . . . . .	111
	Appendices . . . . .	120
II.A	Migration flows: construction and description . . . . .	120
II.A.1	Elements of context . . . . .	120
II.A.2	Data sources and construction of migration flows . . . . .	122
II.A.3	Description . . . . .	128
II.B	Shocks to rural livelihoods . . . . .	136
II.B.1	Crop suitability and use across Chinese prefectures . . . . .	136
II.B.2	International price variations and domestic prices . . . . .	137
II.B.3	Shocks over time and across regions . . . . .	140
II.B.4	An additional source of variation: rainfall shocks . . . . .	142
II.C	Data description . . . . .	145
II.C.1	Firm data . . . . .	145
II.C.2	UHS data . . . . .	150
II.D	Robustness checks and sensitivity analysis . . . . .	154
II.D.1	Shocks to rural livelihoods . . . . .	154
II.D.2	Main specification . . . . .	155
<b>III Place-Based Policies in the Long Run</b>		<b>159</b>
1	Introduction . . . . .	159
2	Context . . . . .	166
3	Data and empirical strategy . . . . .	170
3.1	Data sources . . . . .	170
3.2	Empirical strategy . . . . .	174
3.3	Descriptive statistics . . . . .	179

---

4	Results . . . . .	180
4.1	The rise and fall of the “156”-program counties . . . . .	181
4.2	Mechanism behind the fall of treated counties . . . . .	186
4.3	Alternative interpretations . . . . .	190
5	Conclusion . . . . .	193
	Figures and tables . . . . .	194
	Appendices . . . . .	203
	III.A Firm comparison within treated places . . . . .	203
	III.B Additional tables . . . . .	206
	<b>Bibliography</b>	<b>213</b>



# Introduction

Migrants by moving remodel the world. They transform places by leaving plagued homes for promised lands. They restructure the economy by preferring firms to farms. They may have in mind images of what their former homes should be and influence them through ideas and money. But the depth of those transformations and their consequences depend on how smoothly factors and information flow, and little is known about how human mobility interacts with frictions in the circulation of information, capital and labor. However, this interaction matters in at least three crucial ways: for understanding the causes of migration, as market imperfections may both trigger and obstruct migrant flows; for minimizing its costs and maximizing its benefits; and because migration itself may improve or disrupt the functioning of markets.

Migration is first and foremost a way to overcome barriers to factor movements that impede development. Migrants further economic growth by pursuing better prospects until returns to labor are equalized across places. The key mechanism in [Lewis's \(1954\)](#) celebrated dual-economy model is that workers move from agriculture to the non-agricultural sector, where labor productivity may be about twice as high ([Gollin et al., 2014](#)). This process of structural transformation spans decades ([Alvarez-Cuadrado and Poschke, 2011](#)), because barriers to mobility slow down the reallocation of labor, and capital is not immediately available to the firms that could hire new hands. There would be large aggregate productivity gains if migration costs could be removed overnight: [Bryan and Morten \(2015\)](#) find that higher mobility costs are responsible for 10% of the labor productivity gap between the United States and Indonesia. Whereas the role of migration in triggering the transition from agriculture to industry has been the focus of a large literature, its part in later stages of economic development remains under-researched.

Secondly, the disruption that migration may entail depends crucially on frictions in capital and labor markets. One of the most hotly debated issues around migration, both in the economic literature and political arena, is its impact at destination, and whether it helps or harms non-migrants. But

even massive flows of immigrants, as experienced in the past two decades by Chinese cities, where as many rural migrants lived in 2010 as there are international migrants worldwide (Chan, 2012b; United Nations, 2015), would be accommodated painlessly in a frictionless economy. If jobs are in short supply, migrants' competition should exert a downward pressure on the wages of previously settled workers. The existence and persistence of an adverse effect of immigration on destination labor markets crucially hinges on frictions in factor movements (Borjas et al., 1997; Card, 2001). No such effect arises in theory if labor markets are fully integrated, as displaced workers export their labor to other places, where competition is less fierce, and thus restore the labor-market equilibrium, or if capital can effortlessly and instantaneously flow in to accommodate the rising labor supply and expand production. If there was no cost or delay in their reactions, (non-migrant) labor and capital flows would adjust until returns are equalized across places and the effect of immigration diffused throughout the economy and hardly felt by the population. But the impact of migration on factor-market frictions themselves is unknown: Large migrant inflows may disrupt factor markets at destination or allow them to function better. It is unclear whether migration leads to a more efficient allocation of factors or favors less productive firms in the receiving economy.

Thirdly, migration itself introduces frictions between migrants and the household of origin they often remain attached to. A key motivation behind migration is to alleviate credit constraints and help origin communities overcome threshold effects that hinder their development. The fact that capital does not flow from rich to poor countries, despite higher returns in the latter (Lucas, 1990; Banerjee and Duflo, 2005), can be interpreted as evidence of imperfections in the international capital market and as an additional powerful justification for migration flows. Migrants, who typically remit a substantial share of their earnings at destination or repatriate their savings, may thus help reduce frictions in access to capital for developing countries and origin places in general. However, migrants typically entrust with close relatives the capital that they send to their home countries. The fact that the investment of migrants' remittances relies on their relatives potentially reintroduces frictions, and remittances need not foster the development of countries of origin if they result in a useful but shortsighted increase in consumption favored by recipients. Close relatives have been shown in the literature on co-resident family

members (Manser and Brown, 1980; Lundberg and Pollak, 1993) to have different preferences, and the apportionment of intra-household resources can significantly deviate from the efficient allocation (Udry, 1996). We would expect those issues to be compounded by distance in the context of migrants' remittances.

Migration is both the sign of spatial frictions and their solution. It is therefore critical to disentangle the imperfections—in labor, capital and information flows—that determine its costs and benefits from the effect that it has on those imperfections. Is migration an opportunity to improve the functioning of the economy, or does disruption prevail, at least in the short run?

This thesis combines three essays in development economics, each of which weaves two main threads together: unexplored aspects of migration and its influence on different forms of friction. Chapter I sheds light on migrants' beliefs and the information asymmetry between remittance senders and recipients that distance aggravates; Chapter II explores the transformation of the receiving economy due to an influx of immigrants, and in particular its impact on the factor-market constraints faced by urban firms; and Chapter III highlights the role of migration in overcoming imperfections in local labor markets and in bringing about the later stage of structural transformation—from heavy industry to consumption goods and services.

Chapter I focuses on the relationship between international migrants and their households of origin and highlights the importance of beliefs and information asymmetry in migrants' remittance and investment decisions.

Migration provides a valuable window into intra-household resource allocations, which a wide literature has focused on. It however differs from the case of co-resident household members as migration implies geographical distance, and the imperfect observability of agents' actions is likely to be a pervasive issue in this context. This friction in the relationship between remittance senders and recipients has potentially far-reaching implications in terms of the ability of remittances to foster development in migrants' places of origin.

On the one hand, migrants show a desire to return and invest in their home countries, of which they enjoy an in-depth and first-hand knowledge. They can moreover rely on a network of relatives to help them carry out

those investments in their absence. They should thus be in a better position than outsiders to exploit business opportunities there. One advantage that investors from outside may have over natives of the country is capital. But migration may in fact offer a solution to credit constraints. Migration therefore shows great promise for the development of origin places.

On the other hand, Chapter I establishes a stylized fact that suggests a different picture thanks to unique data from Senegalese migrants and their households of origin. I show that migrants systematically overestimate assets that they typically invest in and that are held by their households of origin. This puts into question their informational advantage and the possibility of a productive investment of remittances if recipients have a stronger preference for consumption. I rely on the socio-anthropological literature on the Senegalese diaspora to show that those assets are typically financed by migrants, who earmark remittances for their purchase. They are part of the refurbishment or equipment of the family compound, or enter into the creation of a small business, which migrants count on to meet their households' needs and thus alleviate the remittance burden or to meet their own needs upon return to Senegal. This literature also highlights the importance of conflicts and information manipulation about investment remittances, and that migrants resort to various monitoring activities to prevent embezzlement.

To understand whether the stylized fact reflects information manipulation, I hypothesize that frictions in the transmission of information arise between remittance senders and recipients because of distance, and migrants hold heterogeneous beliefs about how trustworthy their relatives are as business partners. Predictions are then derived from a theoretical discussion and confronted with the data. As they contain matched reports from migrants and their own households of origin about asset holdings, the data allow me to directly compare information sets. The fact that discrepancies arise on average means that either migrants are rational but hold differing prior beliefs, and only the most optimistic invest, or they are not fully rational and fail to incorporate in their reports the probability that remittance senders may misrepresent the level of asset holdings, i.e., whether they complied with the migrant's spending instructions. I provide empirical evidence of information manipulation, reject the cognitive bias hypothesis and show that migrants hold heterogeneous beliefs that converge through learning. Since the stylized fact relies on a comparison of reports, measurement error is a serious con-



cern. I therefore develop a flexible measurement error model to show that the predictions supported by the data are robust to various specifications of mismeasurement.

Evidence of the self-selection mechanism and information manipulation implies that there are two major pitfalls to remittances translating into productive investments and thus fostering the development of the country of origin. First, because of information asymmetry, resources are diverted from productive investment to monitoring. The rise of alternative international transfer operators, catering to far-flung diasporas and offering in-kind remittance services or the possibility for migrants to invest directly into their countries of origin without their relatives' help, indeed signals the price that migrants are willing to pay to circumvent information asymmetry issues.

Secondly, it highlights that migrants hold heterogeneous beliefs about the trustworthiness of their remittance recipients as business partners. This is challenging because migrants need not have much power over whom they can partner with. The socio-anthropological literature on the Senegalese diaspora makes it clear that migrants only invest through their (usually, close) relatives, because they trust them and because they will be the ones enjoying the refurbished house or running the migrant's business in her absence, but also because social norms force migrants to invest through their close kin or with their parents' or elders' blessing. Subsequently, many worthwhile investment opportunities may be forsaken because of limited choice in terms of intermediaries and for lack of trust. Conversely, investment remittances may be squandered if unwitting migrants keep on sending money and being deluded about its use.

Chapter II studies the impact of rural-to-urban migration on places of destination. It first sheds light on the effect of immigrants on the labor market and non-migrants' wages, and then investigates its consequences for firms and the reallocation of production factors that migrants induce by choosing one employer over another.

The role of migration in factor reallocation across sectors is well documented and has attracted economists' interest as far back as [Lewis \(1954\)](#), [Kuznets \(1964\)](#) or [Harris and Todaro \(1970\)](#). An aspect of migration that however remains unexplored is that it may also deeply transform the destination economies themselves. In the developing world, rural-to-urban migrants

indeed flock to an urban sector where factor misallocation is pervasive (Hsieh and Klenow, 2009). Since migrants are unlikely to be dispatched randomly to urban firms, we can expect migration to induce a reallocation of factors within the urban sector, which this chapter focuses on.

Because migrants are attracted to destinations that promise higher wages or amenities, it is important to isolate exogenous variation in rural-to-urban migrant flows to estimate their causal effect on urban outcomes. Chapter II relies on international agricultural commodity price shocks interacted with a measure of local suitability for growing different crops to predict emigration from rural areas in China. We find that a higher value of the agricultural portfolio at origin strongly reduces the population's incentive to emigrate, as it reduces the wedge between returns to labor in urban and rural areas, and large adverse fluctuations in labor returns at origin dwarf the costs of migrating. We then combine those exogenous, origin-driven outflows with a gravity model to allocate migrants to destinations. This yields predicted inflows of immigrants that are orthogonal to factor demand in the urban sector and can thus be used to estimate the causal impact of immigration.

This chapter mobilizes various data sets for identification, the measurement of migrant flows and urban outcomes. First, exogenous variation in rural emigration is captured thanks to fluctuations in world demand and supply for agricultural crops reflected in international prices. These prices are then weighted by geo-coded data on harvested area and a model-based measure of potential yields, and collapsed at the level of a prefecture. Prefectures are indeed the finest administrative level at which an exhaustive matrix of bilateral migration flows can be constructed thanks to census data. Urban outcomes can be divided into labor market outcomes for workers, which are studied thanks to a survey of urban non-migrant households, and firm-level variables, which come from a census of large manufacturing firms.

Rural-to-urban migrants are first found to have a strong negative effect on wages in the cities where they settle. This effect is mostly supported by low-skill workers, who are more substitutable to migrant labor, and it is confirmed by highly consistent results from firm-level surveys. The firm data further show that the downward pressure on destination wages persists into the third year after the immigration shock. This implicitly reveals that labor markets in China are relatively insulated from one another, so that the consequences of a large increase in labor supply do not dissipate across space and are strongly

felt locally. As capital does not adjust immediately either, this labor supply shift has a marked effect on the factor mix at the firm level.

However, those transformations are not evenly spread out across urban firms. Private and exporting firms indeed benefit the most from an influx of cheap labor. Firms with a high capital-to-labor ratio also see their labor costs decrease and employment grow more than the average urban firm. The accounting framework developed in the chapter indeed suggests that the capital-to-labor ratio reflects the relative importance of labor- and capital-market constraints. This implies that immigration reduces labor-market frictions for firms that suffer at baseline from difficulties in hiring labor but can mobilize capital to productively employ new workers. Chapter II further investigates the issue of factor allocation within the urban manufacturing sector by estimating the impact of immigration on the distribution of factor productivity in receiving economies. Consistent with the heterogeneity analysis, firms that are more constrained on the labor-market side hire more, which leads to a reduction in the dispersion (and thus misallocation) of labor productivity. Conversely, migration is found to accentuate rather than attenuate the dispersion of capital productivity, so that the overall impact in terms of the efficiency of factor allocation is ambiguous.

This chapter focuses on firms' factor choices and misallocation in the urban manufacturing sector, but the identification strategy it relies on can be exploited to estimate the causal impact of rural-to-urban migration on other urban outcomes, e.g., pollution, congestion, urban sprawl and segregation, etc. This would broaden our understanding of the costs and benefits of the Chinese development path, which heavily relies on migration and labor-intensive manufacturing production, and therefore provide some guidance to the numerous developing countries that try to emulate China's economic success.

Chapter III focuses on the long-term impact of a large industrialization plan in China, and reveals a striking reversal pattern that can be explained by distortions in local labor markets induced by the presence of big plants and the reallocation of resources across places enabled by migration.

Big-push industrialization strategies are common policies in developing countries to accelerate the process of structural transformation and maximize externalities by concentrating production in a few places. As with any place-based policy, a primary concern is the endogeneity of location decisions. Plan-

ners typically select areas that promise greater returns to investment or on the contrary places that are lagging, as another rationale behind place-based policies is the desire to promote even development. To overcome endogeneity issues, this chapter exploits the construction of 156 large plants in late-1950s China and the criteria made explicit by planners for choosing the sites. The “156” program was made possible by scientific and technical cooperation with the U.S.S.R., as well as by Soviet loans and aid. Location decisions were based first and foremost on economic and geographic characteristics that determine the suitability of places for large industrial investments. Historical sources emphasize the importance of access to natural resources and connection to the existing transportation network. Another key determinant of factory locations pertains to geopolitical tensions in East Asia at the time: Chinese and Soviet planners would not risk exposing brand-new, state-of-the-art plants to the threat of enemy attacks.

The identification strategy therefore consists of two steps. First, we define a control group thanks to a matching procedure based on the economic and geographical determinants of suitability singled out by the planners. Second, we create a measure of vulnerability to bombings by calculating the minimum-cost distance for each county in China to U.S. (in Japan and South Korea) and Taiwanese airbases, penalizing travel time for their bombers near U.S.S.R. and North Korean airfields, which act as a protective shield. This vulnerability measure is then used as an instrument for the probability of receiving a “156”-program factory among the set of suitable counties.

The “156” program offers a unique setting to study the long-term impact of large industrial plants thanks to the exogenous dimension of location decisions, but also because it constitutes a gigantic investment in expanding and modernizing the industry in an otherwise backward agrarian economy. The places that hosted at least one of the 156 “Million-Rouble” plants experienced fast growth in the first twenty years after the construction of the factories. Since in this period factor movement was not free, we cannot ascertain that the large increases in population, high growth of economic activity and rapid reallocation of labor from agriculture to the non-agricultural sector are the direct effect of the plants on the local economies or a reflection of the scale of the program itself and complementary investments made by the planner. The identification strategy however allows us to define as the treatment the significant head start that areas surrounding Million-Rouble plants enjoyed

at the onset of the reforms, and investigate whether it helped or hurt those places in the subsequent era of opening-up and loosening of the government's control over the economy. The main finding is a stark reversal of fortune: Treated counties lost in twenty years of reforms much of the gains that they had accumulated in the previous two decades in the wake of the "156" program.

Chapter III does not primarily focus on migration, but migration is crucial for understanding its findings. We develop a theoretical framework that relies on labor-market distortions to explain the rise and fall of treated places. The key mechanism is that the presence of a Million-Rouble plant increases the reservation wage in the local economy, which prevents entrepreneurs from creating firms or entering the market. The initial advantage associated with the treatment unravels when labor mobility is allowed, as migrants flock to areas where the local labor market is less distorted and the economy more vibrant, and where they either become workers or entrepreneurs. Chapter III then tests predictions from the framework and finds support for the hypothesized mechanism.

Heavy manufacturing was well developed in China at the onset of the economic reforms, but light industry and services had been neglected under central planning. Chapter III shows, consistent with the theory presented in the chapter, a widening gap between treated and control counties in terms of the tertiary sector's share of the economy. In the Chinese context, this indicates that firm entry is lower in treated areas and that by moving from treated to control places, migrants shifted economic activity from heavy manufacturing to firms providing services or consumption goods. Chapter III thus highlights the unexplored role of migration in bringing about the later stage of structural transformation.

Following migration and frictions in the circulation of information, labor and capital as guiding threads, this thesis offers insights into how and whether economic development materializes. Chapter I points to migrants' remittances as a valuable source of capital for development but warns us of the frictions introduced in their investments by imperfect information. Chapter II shows that receiving economies grow with immigration, but frictions in factor markets mean that costs and benefits are not evenly distributed among workers. It also provides evidence that immigration reduces the labor-market

distortions that firms face but hints at a possible aggravation of capital misallocation. Chapter III highlights the role of migration in bringing about structural change, by shifting the economy's center of gravity to less distorted labor markets.

# We Only Know What They Tell Us: Migrants' Beliefs and Investments

---

## Abstract<sup>1</sup>

Migrants' investments in their countries of origin should be facilitated by local knowledge and family networks. In this paper, I establish a stylized fact that however questions their informational advantage: Migrants systematically overestimate assets that they typically invest in and that are held by their households of origin. I hypothesize that information asymmetry and heterogeneous beliefs account for the stylized fact and derive predictions from a theoretical discussion. Gaps between migrants' beliefs and true asset holdings can persist if optimistic migrants are more likely to self-select into investment or if migrants fail to adjust for information manipulation. Thanks to unique data with matched reports from migrants and their own households of origin, I provide empirical evidence of information manipulation, reject the cognitive bias hypothesis and show that migrants hold heterogeneous beliefs that converge through learning.

**Keywords:** Asymmetric Information, Belief, International migration, Household Production and Intrahousehold Allocation.

**JEL classification:** D82, D83, F22, O15, D13.

---

<sup>1</sup>I am grateful to Daron Acemoglu, Jenny Aker, Arnab Basu, Cátia Batista, François Bourguignon, Marshall Burke, Juan Chen, Isabelle Chort, Sylvie Démurger, Marcel Fafchamps, Simon Fan, Richard Freeman, Flore Gubert, Marc Gurgand, Émilie Hu, Clément Imbert, Sylvie Lambert, Valentina Mazzucato, Marion Mercier, Alice Mesnard, Ted Miguel, Jean-Noël Senne, Miri Stryjan, Muhamet Yildiz, Roberta Ziparo and Yanos Zylberberg for useful discussions and comments. I also thank participants at the MAFE Project Final Conference, AFSE and CSAE conferences, EUDN workshop in Leibnitz Universität Hannover, AEA meeting in Boston, the Hong Kong Polytechnic University and Lingnan University for helpful comments. All errors are mine.

# 1 Introduction

Remittances are a major opportunity for the development of origin countries. Migrants have indeed access to much larger financial resources than “left-behinds,” and contrary to outsiders, they can rely on their social capital, first and foremost their network of close relatives, to carry out investments. Given migrants’ comparative advantage in investing in their countries of origin, it does not come as a surprise that the rise of international remittances to low- and middle-income countries—from USD 74 to 422 billion in 2000-2015 (World Bank, 2017)—is seen by the development community as showing great promise for development. Migrants’ investments however take place within a transnational household, which may exhibit tensions over the use of remittances and, as any household, suffer from imperfect collective bargaining that leads to an inefficient allocation of resources.

In this paper, I establish a new stylized fact that casts doubt on migrants’ informational advantage and thus on the potential of remittances for economic development. Thanks to matched data on Senegalese migrants and their households of origin, I can compare migrants’ ( $m$ ) and their households’ ( $h$ ) reports of the number of durable assets held by  $h$  in the home country. I find that migrants systematically report a higher number of assets—refrigerators, bicycles, cars, etc.—held by their households of origin than the latter themselves. These assets are typically purchased by migrants through their households of origin as part of investments to carry out on their behalf.

The results of the comparison are exhibited in Table I.1 and establish the stylized fact investigated in this paper. Table I.1 simply regresses the asset discrepancies on a constant. A clear pattern emerges: The constants are systematically positive and significant, which means that migrants overestimate or overstate on average the number of assets held by their households of origin.<sup>2</sup>

This stylized fact suggests a gap between migrants’ and households’ information sets. I provide a theoretical framework to highlight under which condi-

---

<sup>2</sup>Only TV sets and electric fans come across as obvious exceptions, with a negative (and significant, for TV sets) mean. Interestingly, TV sets are the asset that the highest proportion of matched households hold. Only 16% of them do not have any TV set, as against 30% for radios and 43% for electric fans, the second most commonly held assets in the data. Differences between types of assets will be further discussed in Section 3 and shall be critical in distinguishing between interpretations of the results.



tions systematic discrepancies between migrants' and their households' survey reports are possible. A first condition is information asymmetry: Imperfect observability due to geographical distance implies that remittance recipients who disagree with migrants' instructions can use private information to divert resources. Since migrants are likely to have developed mechanisms to monitor the use of their remittances, I include migrants' monitoring technology in the discussion and clarify assumptions under which it may be optimal for recipients to deviate from the migrants' instructions.

But information manipulation alone is not enough for systematic discrepancies to obtain. Rational migrants holding unbiased beliefs about the behavior of their remittance recipients should be able to correctly assess the degree of realization of their asset investments. Although imperfect observability means that discrepancies can arise within individual migrant-household pairs, we would expect them to cancel out *on average* as migrants adjust households' reports based on their knowledge of the probability of a misrepresentation. To account for the stylized fact, it must therefore be assumed that migrants exhibit heterogeneous beliefs—in terms of remittance recipients' trustworthiness as business partners—and optimistic migrants are more likely to invest,<sup>3</sup> or that they suffer from a cognitive bias that prevents them from correctly adjusting the information received.

Based on the theoretical discussion, I derive testable predictions that I then bring to the data. The data contain unique matched information on origin households' asset holdings and thus allow me to contrast migrants' with their own households' reports. I show that the observed discrepancies provide evidence of information manipulation and heterogeneous beliefs. A key insight of the theoretical framework—the selection of migrants with more optimistic prior beliefs into investment—provides me with a test to reject the cognitive bias hypothesis. The data further offer evidence of learning, whereby investor migrants' beliefs converge to “the truth,” i.e., actual asset holding realizations, with migration experience and the number of migrants from the same household. Finally, since the empirics of the paper rely mostly on a comparison of survey reports, measurement error is a potential concern.

---

<sup>3</sup>Note that an alternative framework would be persuasion (Kamenica and Gentzkow, 2011). I abstract from the possibility that households try to change migrants' beliefs because of the importance of monitoring: The migrant's posterior beliefs should be primarily based on (imperfect) evidence from monitoring and prior beliefs, rather than the household's claim that her investments have been carried out as instructed.

Based on a flexible model of measurement error and further empirical tests, I confirm that the gaps in survey reports reflect gaps in information sets and differences in beliefs.

This paper makes several contributions to the literature.

First, this paper sheds light on intra-household resource allocation and provides direct evidence of information asymmetry. The hope that migrants' remittances can be harnessed for development depends on implicit assumptions that run counter to the findings of the microeconomics of the household. Close-knit relationships between members of a (transnational) household are no guarantee that coordination will be smooth and agreements enforced. Relatives may have different preferences (Manser and Brown, 1980; McElroy and Horney, 1981; Lundberg and Pollak, 1993), and intra-household resource allocation is often found to be inefficient by the empirical literature (Udry, 1996; Dercon and Krishnan, 2000; Goldstein et al., 2005; Dubois and Ligon, 2011), which may be a red badge for information asymmetry. The geographical distance inherent in migrant households is likely to aggravate this issue, as agents' actions are not easily observable (Chen, 2006, 2013; De Laat, 2014; Seshan and Yang, 2012; Ambler, 2015). The stylized fact presented in this paper is in line with findings from a growing literature on information asymmetry and control issues in geographically dispersed households. Evidence suggests that migrants seek more control over the use of the remittances they send (Chin et al., 2011; Ashraf et al., 2015; Batista and Narciso, 2016), in particular through in-kind transfers (De Laat, 2014; Batista et al., 2015; Torero and Viceisza, 2015), and transfer recipients alter their behavior based on the observability of their actions (Azam and Gubert, 2005; Chami et al., 2005; Chen, 2006, 2013), which is consistent with the experimental evidence in Jakiela and Ozier (2015).<sup>4</sup> The originality of this paper resides in matched data that enable me to directly contrast migrants' and their households' information sets about assets that migrants typically invest in through remittances. This approach resembles that in De Weerdt et al. (2015), who use cross-reports from domestic migrants and their households of origin in Tanzania to construct a measure of misperception of each other's consumption and shed light on

---

<sup>4</sup>Other papers highlight that *migrants* indulge in strategic behavior—see Chort et al. (2012) on Senegalese migrants, and Ambler (2015). It is worth noting that Ambler (2015) finds evidence of migrants taking advantage of private information but cannot conclude that recipients behave strategically in her setting.

transfer motives. Similarly, [Seshan and Zubrickas \(2015\)](#) rely on reports of Qatar-based migrants' earnings collected from the migrants and their wives in Kerala. They find that wives tend to underestimate their husbands' overseas earnings and that this pattern correlates with lower annual remittances sent home.<sup>5</sup>

Second, this paper provides empirical evidence of an understudied feature of human choices and behavior. Heterogeneous beliefs have been the focus of a mostly theoretical literature ([Van den Steen, 2004, 2011](#); [Sethi and Yildiz, 2012](#); [Acemoglu et al., 2016](#), *inter alios*). The matched data used in this paper allow me to show their existence: Information manipulation creates gaps in information sets, but the systematic discrepancies observed *on average* can be reconciled with rationality only with differing prior beliefs. An alternative explanation could be that migrants suffer from a cognitive bias and deviate from full rationality. This paper however rejects such a hypothesis based on additional tests, and thus confirms the presence of heterogeneous beliefs. Heterogeneous beliefs may have far-reaching consequences, as households of origin are usually migrants' only possible intermediaries for implementing investment projects. A migrant willing to invest in her country of origin needs to trust her relatives to receive the money and follow her instructions on how to spend it. A divergence between migrants' beliefs about their intermediaries' trustworthiness and their true quality as business partners may thus imply inefficient investment decisions.

Third, the empirical results established in this paper indeed constitute crucial elements for assessing the developmental role of migrants' remittances. Developing economies exhibit wide variations in returns to capital. Standard economic theory suggests that returns to capital should be lower in richer than in poorer countries, but it also predicts that capital should flow from low- to high-return areas until marginal products are equalized ([Lucas, 1990](#)). Yet, the microeconomic literature shows persistently high rates of return to capital in developing countries ([Banerjee and Duflo, 2005](#)), not least in the informal sector ([Udry and Anagol, 2006](#); [De Mel et al., 2008](#)). This hints at untapped investment opportunities. International migrants may be in the best

---

<sup>5</sup>[McKenzie et al. \(2013\)](#) also suggest that Tongan migrants to New Zealand misrepresent their earnings *downwards* to their transfer recipients. Other papers focus on information asymmetry between co-resident household members—see [Ashraf \(2009\)](#), [Castilla and Walker \(2013\)](#) and [Ziparo \(2014\)](#).

position to make the most of these unexploited prospects. In particular, they can avail themselves in their origin countries of their social capital, an important explanation for heterogeneity in returns to physical capital (Banerjee and Munshi, 2004). Another key ingredient in migrants' advantage over outsiders in investing in their countries of origin is the trust they have in their business intermediaries. The literature on remittances provides some indications about the impact of information asymmetry in migrants' remittance decisions, which might carry over to investment, but does not discuss the role of beliefs. Information asymmetry may lead to a reduction of remitted amounts (Ashraf et al., 2015; Batista and Narciso, 2016), as migrants expect remittance recipients to hide the actual use of the funds behind the veil of distance. Such a negative effect on remittances could hurt development irrespective of the benefits of migrants' investment projects, as less money irrigates migrant-sending economies. The literature on information asymmetry from migrants' households of origin (Azam and Gubert, 2005; Chami et al., 2005) further suggests that moral hazard is at work and remittances would be more productively employed if migrants could exert more control over their use. Recipients indeed have little incentive to expend effort on reducing their dependence on remittances, whereas migrants want to first alleviate their remittance burden and then prepare their return by setting up means of subsistence back home. It therefore seems likely, as posited in this paper, that migrants' preferences are more conducive to development, and deviations from their instructions are predominantly detrimental to productive investments. Differing prior beliefs, in particular about the trustworthiness of the households of origin as business partners, may induce pessimistic migrants to forsake valuable investment opportunities, while it may lead to a squandering of productive resources if optimistic migrants invest and do not realize that their remittances are being diverted.

Despite the importance of information asymmetry, heterogeneous beliefs and cognitive biases, evidence of these issues is rarely found in the empirical literature. To the best of my knowledge, this is the very first paper to show the presence of heterogeneous beliefs and persistent information asymmetry in the migration context.

The remainder of the paper is organized as follows. Section 2 frames the stylized fact in the socio-anthropological literature on the Senegalese transna-

tional household economy. Section 3 develops a theoretical framework for migrants' remittances in a world of imperfect information and with differing prior beliefs. Section 4 presents the data. Section 5 tests the predictions from the theoretical framework. It also deals with potential alternative interpretations and pays particular attention to measurement error. Section 6 discusses the findings and concludes.

## 2 Remittances and information in Senegalese transnational households

The socio-anthropological literature on the Senegalese diaspora points to strategic behavior in the use of remittances, which constrains migrants in their ability to invest.<sup>6</sup>

Once their families' daily needs are provided for, migrants aspire to set up businesses back home and build a house for their relatives and themselves upon return (Fall, 2003; Dia, 2007; Boltz-Laemmel and Villar, 2014). Remittances sent by migrants for investment are almost exclusively entrusted to their relatives (first and foremost the household head),<sup>7</sup> because they will be the ones enjoying the migrant's investment in her absence and migrants complain they cannot easily find reliable intermediaries (Fall, 2003).

Migrants thus typically earmark remittances for different uses and investments are made by the family back home. A complementary survey I carried out for this study<sup>8</sup> shows that a third of respondents earmark remittances for particular purchases, through in-kind transfers or verbal instructions.

Deviations from the migrants' instructions are however widespread (Dia and Adamou, 2003; Fall, 2003; Marfaing, 2003). The main reasons behind such a diversion of remittances are preference differences, self-control issues and redistributive pressure (kin tax). This suggests information asymmetry and manipulation, whereby relatives back home exploit private information

---

<sup>6</sup>A more in-depth review of the context of Senegalese migrants' remittances based on the socio-anthropological literature is developed in Appendix I.B.

<sup>7</sup>The data presented in Section 4 show that around 70% of Senegalese migrants—both in Europe and in Mauritania—send primarily to the head of their household of origin, usually the migrant's father or mother. See Appendix Table I.A.7 for this and additional descriptive statistics on the relationship between migrants and households of origin.

<sup>8</sup>Please refer to Appendix I.B for a description.

about how remittances are spent to further their own interests. Half the interviewees in the complementary survey thought their earmarking was not followed and information was distorted to extract rents. As one of them put it: “We only know what they tell us.”

Migrants therefore exert themselves to reduce the fungibility of transfers by sending them to shopkeepers so as to allocate them beforehand, remitting in kind and monitoring transfer recipients through phone calls, contacts with other migrants or neighbors from the same community, and visits to Senegal to manage their projects. Should monitoring reveal deviations, migrants explain that the main sanctions at hand are to undermine the recipients’ reputation or to put an end to investments transfers.

### 3 Theoretical framework

The stylized fact hints at gaps between migrants’ and households’ information sets. Based on a solid body of socio-anthropological evidence, we know that households of origin face incentives to bend the information they share with migrants to divert remittances to preferred uses while securing future investment funds. We can also anticipate that they will exaggerate asset holdings for those assets that are favored by migrants.<sup>9</sup> Among those, investments in productive ventures (usually, family-run shops) and housing take pride of place. Nevertheless, we saw that migrants do monitor the use of the remittances they send, which shows that they are aware of potential misrepresentations.

We therefore need a theoretical framework to account for the discrepancies: Albeit a likely culprit, information manipulation alone is not sufficient for discrepancies to obtain *on average*. If agents are rational and aware of the information asymmetry that pervades their relationship, we would expect them to adjust their survey reports accordingly and individual discrepancies to cancel out when we take means over the whole sample.

In this section, I combine elements from the literature on information asymmetry with more recent contributions pertaining to differing prior beliefs

---

<sup>9</sup>Conversely, different motives for remittances can lead to incompatible predictions about information manipulation when it comes to aggregate variables such as earnings or wealth. For instance, we would expect the household of origin, should they behave strategically, to either play down or inflate their wealth depending on whether remitters are believed to be motivated by altruism or inheritance, respectively.

to delineate a theoretical framework. The goal is to shed light on the mechanisms underpinning the stylized fact and to derive predictions that I take to the data in Section 5.

### Theoretical setup

There are two agents, the migrant ( $m$ ) and her household of origin ( $h$ ). Migrants send remittances  $t$  to  $h$  for the purchase of an asset  $A$  that enters the production of a final good (realized much later) that  $m$  wants to invest in.<sup>10</sup> Investments are divided into two types, distinguished by whether all migrants engage in them or selection occurs. I denote  $A_s$  “selection” assets that migrants choose to purchase through remittances and  $A_d$ , or “default” assets, those that all migrants purchase.<sup>11</sup> If the migrant decides to invest into  $A_s$ , she faces a binary choice: invest through the household of origin or save at destination.<sup>12</sup> I abstract from this distinction in the general setup, as it will only become crucial once I introduce differing prior beliefs.

Remittances are part of a contract:  $h$  receives  $t$  along with an instruction for how to spend it,  $A^{*m}$ . Importantly,  $h$  shall enjoy  $A$  and thus always accepts the contract, and for simplicity I assume away any direct cost of implementing  $m$ 's instruction for  $h$ . But geographical distance between  $h$  and  $m$  introduces information asymmetry:  $m$  cannot directly observe the realized  $A^h$  (nor can she observe the final good) and  $h$  can exploit its private information to divert the remittances to its preferred uses.

Households differ in terms of “trustworthiness” as business partners. This trustworthiness  $\theta$  determines the probability that  $h$  deviates from  $m$ 's instructions. It is unknown *ex ante* to both  $m$  and  $h$ . This creates a classic principal-agent problem, where  $m$  is the principal and  $h$  the agent. Moral

---

<sup>10</sup>Migrants often leave with the project of returning. Once the daily needs of their households of origin are provided for, migrants start complementing their remittances with investment funds,  $t$ . Migrants' investment is twofold. First, they wish to expand or refurbish the family compound. Second, they set up businesses to provide their households with a livelihood as well as support themselves upon return.

<sup>11</sup>Reasons why some assets are always purchased include cultural norms, migrants' desire to maintain social ties or aligned preferences between senders and recipients.

<sup>12</sup>A key feature of migrants' investment in their country of origin is the constraint in terms of reliable intermediaries. The household of origin is commonly the only eligible business partner, either because of the family-oriented nature of the investments or because of pervasive trust issues. Migrants also rarely invest in their countries of destination, where they may face legal uncertainty and do not necessarily plan to stay.

hazard emerges as  $h$  privately discovers its type after investment begins and can use this private information to further its own interests at the expense of the migrant's.

For crisp predictions, I assume that both  $m$ 's and  $h$ 's preferences are common knowledge.<sup>13</sup> The migrant thus knows what  $h$  would have done without her investment funds  $t$ . Since  $m$  knows  $h$ 's maximization problem in autarky, she can predict its behavior perfectly, except that  $h$ 's quality as a business partner cannot be observed by  $m$  and, after the remittance is sent,  $h$  is subject to a shock that she cannot observe either, but the distribution of which (conditional on  $\theta$ ) is common knowledge. The household's deviations from  $m$ 's instructions may have a variety of sources. A crucial assumption is that  $m$  may consider shocks to  $h$  (kin tax, self-control problems, emergencies...) that induce them to deviate as endogenous to  $h$ 's type or trustworthiness, which  $m$  tries to assess. Neither  $m$  nor  $h$  has prior information about  $h$ 's type as a business partner because investment from abroad through  $h$  is a completely new game.

After it receives  $t$ ,  $h$  chooses  $A^h$  and reports (sends a message)  $\hat{A}$  to  $m$ , which is always equal to  $m$ 's instruction  $A^{*m}$  if  $h$  is sure to be punished when a deviation is acknowledged. The actual  $A^h$  is unobservable to and unverifiable by  $m$ . However, observability can be restored *ex post* through  $m$ 's monitoring of  $h$ 's behavior. I assume that if  $m$  detects a misrepresentation, the only sanction at hand is to put an end to investment remittances.<sup>14</sup> Monitoring comes at a cost and is imperfect: It allows  $m$  to detect misrepresentations ( $\hat{A} \neq A^h$ ) with a probability that is less than 1 and depends on how much she spent on monitoring.<sup>15</sup> Geographical distance increases the cost of monitor-

---

<sup>13</sup>Preferences do not need to be common knowledge, although long and daily interactions prior to migration make it likely. Unobservable preferences, e.g., because they are not fixed and are affected by migration or remittances, would make gaps in information sets *more* likely.

<sup>14</sup>Because of social norms, the migrant has limited ability to sanction the household of origin for deviations from her instructions. The most credible sanction consists of putting an end to all (investment) transfers, which the migrant should rationally do if she realizes that expected benefits are too low. Such a sanction is however costly to the migrant and she needs to update her beliefs about whether investing through  $h$  is worth the effort. Since  $A$  is unobservable and the final good is realized in the future (upon her return), the migrant needs to engage in monitoring. Typical monitoring activities include phone calls to the household, the collection of information through neighbors who act as her "eyes and ears," trips back home, etc.

<sup>15</sup>A more complex understanding of the costs and benefits of monitoring is possible. Phone calls, for instance, may cut both ways: They enable  $m$  to monitor  $h$  but also give



ing. For simplicity, let us think of this monitoring technology as subject to false negatives but not false positives, i.e.,  $m$  never wrongly concludes that  $h$  misrepresented  $A^h$ .

Imperfect monitoring means that  $h$  may optimally deviate from  $m$ 's instructions, despite much of its behavior being correctly anticipated by  $m$ . Even if  $h$  derives utility from  $A$ ,<sup>16</sup> it would be weakly better off with untargeted remittances. It thus has an incentive to treat  $m$ 's remittances as fungible and increase its utility by purchasing  $A^{*h} \leq A^{*m}$ , where  $A^{*h}$  is the quantity that maximizes  $h$ 's utility subject to a budget constraint relaxed by the amount of the remittance, and  $A^{*m}$  is the migrant's instruction. With monitoring and the threat of a broken investment relationship, which would reduce  $h$ 's intertemporal utility,  $h$  partly abides by  $m$ 's instructions and purchases  $A^{*h} \leq A^h \leq A^{*m}$ .

Thanks to monitoring,  $m$  sometimes observes the true realization of  $A$ ,  $A^h$ . I call  $A^m$  the migrant's survey report, which is equal to  $A^h$  when a misrepresentation is exposed; I detail below what  $A^m$  is when no deviation is detected.

### Prior beliefs

It does not follow from  $h$ 's misrepresentations alone that  $\Delta \equiv A^m - A^h > 0$  on average, as in the stylized fact. Absent monitoring,  $m$  should indeed put little faith in  $h$ 's message  $\hat{A} = A^{*m}$  and replace it in her survey responses with the expected level of asset holding—based on her knowledge of  $h$ 's preferences and of the distribution of shocks that might affect  $h$ 's decisions. Knowing that monitoring is imperfect, she should realize that failure to detect a lie does not mean that the message was truthful and adjust the result of monitoring based on her knowledge of its limitations. More technically, if  $m$  knows the distribution of the likelihood of a misrepresentation conditional on discovering none after monitoring, she holds the right belief about the household's action on average and we should not observe systematic discrepancies. The migrant's belief about  $h$ 's trustworthiness plays a crucial role in her assessment  $A^m$  and thus in the observed discrepancies.

---

$h$  the opportunity, in a model with transfers for other purposes than investment in  $A$ , to request more money. Such a risk would increase the cost of monitoring.

<sup>16</sup>The household of origin should indeed derive positive utility from  $A$  as it enjoys the migrant's house and assets in her absence or earns income from the family-run business.

As in every new game, the common prior assumption that the economic literature usually takes as a starting point need not be legitimate. I assume that each individual migrant's prior belief about her own household's type is drawn from a distribution centered on the truth (Van den Steen, 2004), so that each migrant may be too optimistic or pessimistic about  $h$ 's trustworthiness, but migrants on average hold correct beliefs. Importantly, priors mean that there is open disagreement: If several migrants' beliefs could be drawn from a household-specific distribution, they would "agree to disagree" (Aumann, 1976). Each migrant would see the distribution of beliefs as information on how mistaken the others are (Van den Steen, 2011). This distribution can be symmetric, so that migrants are equally likely to be optimistic or pessimistic about  $h$ 's trustworthiness.<sup>17</sup>

The migrant's prior belief in turn determines whether she invests in  $A_s$  but not in  $A_d$ , which is always purchased.<sup>18</sup> This selection is a consequence of the migrant's optimizing behavior (Van den Steen, 2004). Since she can choose whether to invest in  $A_s$  or not (contrary to  $A_d$ ), she determines a course of action based on her prior, and she is more likely to choose an action about which she is more optimistic, a mechanism similar to the "winner's curse" (ibid.). This induces a systematic bias in the beliefs of the migrants who invest, as pessimistic migrants are more likely to select out and optimistic ones to select in, whereas beliefs remain aligned with the truth on average when it comes to  $A_d$ .<sup>19</sup>

A key issue with differing prior beliefs is that of learning. A typical justification of the common prior assumption in the literature is that agents learn through their own experiences. A Bayesian individual who does not assign

---

<sup>17</sup>Following Aumann's (1976) theorem, "if two people have the same priors, and their posteriors for [some event] are common knowledge, then these posteriors are equal." Subsequently, the stylized fact suggests that either posterior beliefs are not common knowledge or migrants hold different priors. In the proposed theoretical framework, the migrant is the only Bayesian individual and tries to establish the truth about  $h$ 's trustworthiness. She can catch a glimpse of it through monitoring, which corresponds to the (partially) public beliefs posited in Aumann's theorem. This leaves us with differing priors, which is the modeling approach adopted here.

<sup>18</sup>Strictly speaking,  $m$ 's threat of putting an end to investment remittances has teeth only when it comes to  $A_s$ . As we saw in Section 2, migrants however have other tools at hand to ensure that recipients follow their instructions, e.g., leveraging reputation for sanctions.

<sup>19</sup>It is however possible that members of  $h$  select into migration based on their prior beliefs about  $h$ 's trustworthiness. Then some bias in migrants' beliefs could also be observed as far as  $A_d$  is concerned, but it should be muted compared to  $A_s$ .

zero probability to “the truth” and observes signals that do provide information about it should eventually hold the right belief (Savage, 1954). Learning may also occur through communication. If beliefs are correlated and observable, agents' posterior beliefs should converge (Sethi and Yildiz, 2012).<sup>20</sup> On the other hand, an equally rational Bayesian individual should “overweight information of which they overestimate the precision and underweight in the opposite case” (Van den Steen, 2011), which leads to an “overestimation of the precision of the final estimate, which tends to increase as agents get more data” (ibid.).

The literature provides reasons why beliefs might diverge even if agents have access to the same information. First, migrants may exhibit a confirmatory bias (Rabin and Schrag, 1999), which might alter the way they monitor  $h$  or lead them to interpret evidence so as to reinforce their beliefs. Second, bounded memory (Wilson, 2014) might limit their ability to learn. Third, perhaps more relevant to this setting, migrants may face uncertainty in the signals they acquire through monitoring, as in Acemoglu et al. (2016).<sup>21</sup> Uncertainty may lead to the migrant's subjective probability distribution of events to have full support, so that two migrants from the same household would “never agree, even after observing the same infinite sequence of signals” (ibid.). Subsequently, an optimistic migrant may downplay evidence of misrepresentations or not translate it into more pessimistic beliefs about  $h$ 's type. The household may realize this and enjoy a rent.

Whether learning leads to a convergence of posterior beliefs to the truth in the context of migrants and their households of origin or leads to hetero-

---

<sup>20</sup>Migrants' priors may indeed be correlated: After observing that a fellow migrant fell victim to her household's manipulations,  $m$  revises her beliefs about  $h$ —unless they live in separate worlds, as each migrant can invest only through her own family. But if socialization patterns are endogenous to investment behavior (migrants willing to set up businesses back home tend to get along and meet), beliefs would converge but still remain misaligned with the truth.

<sup>21</sup>Monitoring is indeed unlikely to provide clear-cut binary results. Phone calls in particular rather produce “soft” information that is difficult to interpret. For instance,  $h$  might reply to  $m$ 's inquiries about  $A^h$  by stating evasively that “All is in order” or “It's as good as done.” Moreover, through monitoring,  $m$  tries to understand  $h$ 's type. As such, “harder” evidence, e.g., physical visits to  $h$ , do not necessarily put an end to uncertainty but might simply provide more compelling evidence that  $h$  misbehaved, although  $h$  could still put forward shocks that prevented it from abiding by  $m$ 's instructions. It is then up to  $m$  to interpret the signal—decide whether the kin tax that  $h$  was subject to was indeed unavoidable or endogenous to  $h$ 's behavior,—which she rationally does based on her priors.

geneous posterior beliefs is an open question.<sup>22</sup> But one key characteristic of beliefs—as opposed to a cognitive bias, see below—is that they are constantly updated by agents on the basis of new information.

Importantly, the stylized fact could not obtain with only information asymmetry or differing priors. Migrants may wrongly estimate  $A^h$  on particular occasions. But if migrants’ beliefs about their households’ quality as business partners were always perfectly aligned with the truth, migrants would correctly adjust monitoring results, so that there would be no discrepancies on average, although migrants would be over- or underestimating  $A^h$  depending on whether monitoring led to a false negative or not. Even if each migrant holds a misaligned belief but beliefs are unbiased on average, discrepancies would arise only if migrants self-select into investment based on their beliefs.

### Cognitive bias

An alternative explanation for the stylized fact is that migrants fail to adjust monitoring results for the possibility of false negatives. There are several potential reasons behind this. Migrants could find it computationally demanding to perform that adjustment and simply share with the enumerator what their recipients reported—although they know it is likely an exaggeration. Migrants might also not be perfectly rational. Adjusting might inflict a psychological cost upon them, as it is tantamount to accusing the household of origin of misrepresentation without evidence. A third possibility is that they suffer from a cognitive bias that prevents them from drawing the consequences of the limitations of their monitoring technology, although they are aware of it. This cognitive bias is similar to [Eyster and Rabin’s \(2005\)](#) “cursed equilibrium,” whereby individuals do not realize the information embedded in other agents’ actions.

Such a cognitive bias has two main consequences that can help us distinguish it from differing priors in the data. First, failure to adjust should be observed whatever the type of assets considered,  $A_s$  or  $A_d$ , since it originates from the way the migrant processes information, not from the selection of migrants into an action. Second, comparative statics should differ: It is in-

---

<sup>22</sup>Note that in a one-shot game, migrants would be unable to benefit from the experience that they gain and no learning would appear in the data.

deed unclear why the cognitive bias should increase or attenuate over time,<sup>23</sup> whereas migrants can be expected to update their beliefs based on the additional information that they gather—through monitoring, with experience or through interactions with fellow migrants. As shown in Acemoglu et al. (2016), the updating of beliefs need not lead to convergence with the “truth”; it may exacerbate the gap between migrants' beliefs  $A^m$  and the true  $A^h$ .

### Predictions

This theoretical discussion allows us to derive predictions that can be tested empirically.

First,  $h$ 's incentive to regard  $m$ 's remittances as fungible means that  $\Delta \equiv A^m - A^h > 0$ . This holds true if  $h$  manipulates information and if either more optimistic migrants select into investing in  $A_s$  or migrants have correct beliefs on average but fail to adjust for false negatives.

Second,  $h$ 's incentives to deviate from  $m$ 's instructions should be stronger when observability is lower. This may originate from a higher cost of monitoring, in particular due to geographical distance. This prediction holds whether discrepancies are due to the selection of optimistic migrants into investment or to a cognitive bias in adjustment. The effect of observability indeed acts through information asymmetry, which is at work under both hypotheses.

Third, the sign of  $\Delta_d \equiv A_d^m - A_d^h$  depends on  $m$ 's beliefs and cognitive behavior. If differing priors is the relevant framework, then we should have  $\Delta_d = 0$ , as there is no selection for this type of investment: Unbiased beliefs on average are reflected in aligned survey reports. On the other hand, if cognitive bias is at work, we should observe no difference between  $A_s$  and  $A_d$  since in both cases  $m$  would fail to adjust, so that  $\Delta_s > 0$  and  $\Delta_d > 0$ . Note that measurement error complicates this pattern—I deal with this issue in Section 5.2 and Appendix I.D.

Fourth, learning should be a gradual process. Subsequently, discrepancies need not survive in equilibrium, as  $m$  gathers information about misrepresentations through monitoring and revises accordingly her beliefs about  $h$ 's trustworthiness. We would thus expect discrepancies to be larger at the beginning of the migration experience and attenuate over time.

<sup>23</sup>Discrepancies due to such a cognitive bias could be *observed* to decrease over time in the data because they also require information manipulation, which improved monitoring could reduce.

Fifth, the speed of convergence is a function of the size of the migrant's network if  $m$  sees her peers' experiences with their households of origin as relevant to her own situation. On the other hand, a lack of age or migration experience gradient (or a positive relationship) would be suggestive evidence of asymptotic divergence. Note that evidence of learning would reinforce the conclusion that migrants hold different beliefs, as cognitive biases should be independent of the amount of information gathered (Van den Steen, 2004).<sup>24</sup>

## 4 Data

To establish the stylized fact presented in the introduction and test predictions from the theoretical framework, I draw on matched data from the MIDDAS project (2009-2010).<sup>25</sup>

MIDDAS first contacted migrants in the host country, aspiring to be nationally representative of the Senegalese immigrant population in France, Italy and Mauritania.<sup>26</sup> Chort and Senne (2013) show thanks to French and Italian Census data that the MIDDAS data collection process was indeed successful at providing a faithful picture of the Senegalese diaspora.

Second, the migrants were asked to put the survey team in touch with their households of origin, who were then interviewed at home in Senegal and presented with a thorough questionnaire that painstakingly describes and follows the complex structure of Senegalese households, made up of several subgroups or "cells."<sup>27</sup> Table I.2 summarizes information about sample size

---

<sup>24</sup>There might be cutoffs in  $m$ 's use of monitoring. We would expect very optimistic migrants not to bother with monitoring. Very pessimistic ones should not invest in  $A_s$  at all. In between, migrants who are optimistic enough to invest but not blindly trusting should engage in monitoring. If monitoring leads to an alignment of beliefs with the truth, migrants with intermediate beliefs may gradually sort into pessimism (and select out) or optimism (and abandon monitoring). Monitoring may also survive in equilibrium if  $m$  believes that  $h$  would deviate from her instructions as soon as the threat of sanctions (credible only under monitoring) is lifted. Conversely, if  $m$  tends to disconnect evidence of misrepresentations from her belief about  $h$ 's type—by believing that shocks are exogenous to  $h$ 's behavior, for instance,—monitoring need not reduce  $m$ 's bias and may survive in equilibrium to curb  $h$ 's incentives to deviate.

<sup>25</sup>"Migration and development in Senegal: an empirical analysis using matched data on Senegalese migrants and their origin households." For a detailed presentation of the project, see <http://www.dial.ird.fr/projets-de-recherche/projets-anr/middas>.

<sup>26</sup>For an overview of the sampling procedure, please refer to Chort et al. (2012).

<sup>27</sup>The household questionnaire was based on "Pauvreté et Structure Familiale" (PSF)—see De Vreyer et al. (2008) for a description.

and composition.<sup>28</sup>

It is worth noting that although not all migrants' households of origin could be matched, resulting in small sample sizes, sample selection seems rather mild. Appendix Tables I.A.5 and I.A.6 explore the determinants of selection into the matched sample. We see that most variables are insignificant and coefficients small. Some variables are however important predictors of inclusion in the matched sample, e.g., enumerator characteristics and income, and shall be studied as potential confounders of the stylized fact in Section 5. [Chort and Senne \(2013\)](#) further show based on a nationally representative survey of the Senegalese population ([De Vreyer et al., 2008](#)) that matched migrant households are similar to the average migrant household in Senegal. Moreover, information asymmetry is likely to be less of an issue in the matched than in the unmatched sample, as migrants and households of origin should exhibit stronger ties in the former. Evidence of information asymmetry and manipulation should therefore be understood as a lower bound compared to the population.

Matched data shall enable us to compare migrants' and households' reports (denoted  $A^m$  and  $A^h$ , respectively) of the number of durable assets  $A$  held by the household of origin, and thus *measure* discrepancies ( $\Delta \equiv A^m - A^h$ ) that potentially coincide with gaps between migrants' and households' information sets.<sup>29</sup>

The data encompass a wide variety of assets typically financed by  $m$ , who earmarks remittances to such goods and exhibits a higher preference for them than their households of origin, as highlighted in the socio-anthropological literature presented in Section 2. As can be seen from Table I.2,  $h$  overwhelmingly resides in urban areas; migrants are thus unlikely to invest in agriculture.<sup>30</sup>

---

<sup>28</sup>Only 4% of migrants remit to more than one household. The stylized fact presented in Section 1 and all the following analyses thus focus on the migrant's household of origin—and almost unique remittance recipient—as  $h$ .

<sup>29</sup>The data also contain information on the characteristics of  $h$ 's dwelling. The results however rely on survey reports and are therefore subject to measurement error, which is discussed in detail in Appendix I.D. Because of the more complex structure of measurement error in housing quality variables, this paper focuses on asset ownership.

<sup>30</sup>The MIDDAS data are cross-sectional and do not provide information on how remittances are apportioned between (intended or actual) uses. Based on the literature and characteristics of the population, we however expect migrants' remittances—excluding those sent to finance daily needs—to be mostly targeted at housing expansion or refurbishment

## 5 Results

I now turn to evidence of information manipulation and heterogeneous beliefs. This section first focuses on the role of observability (Predictions 1 and 2) and then moves on to distinguishing between differing prior beliefs and cognitive bias as an explanation for the discrepancies (Predictions 3 to 5). A final subsection investigates the robustness of the results to various measurement error models and confounding factors, and exploits a quasi-natural experiment to reinforce the identification of the test of Prediction 2.

### 5.1 Predictions 1 and 2: Positive discrepancies

Based on the theoretical framework, we predict that (i)  $h$  might derive benefits from information manipulation, leading to a possible overestimation of  $A^h$  by  $m$ , and (ii) these benefits are reduced by  $m$ 's monitoring. The stylized fact presented in Table I.1 suggests that mechanism (i) is indeed at work. This section focuses on Prediction 2.

The test of Prediction 2 relies on the assumption that geographical proximity reduces the cost of monitoring and increases observability. We indeed find that proximity is strongly correlated with monitoring activities. Table I.3 shows that Senegalese migrants living in Mauritania are significantly more likely to visit  $h$  frequently and to remit in kind, which can be construed as proxies for the frequency and ease of monitoring.<sup>31</sup>

The data support the prediction that geographical distance increases the cost of monitoring and thus widens the gaps in reports observed in the data. Table I.4 reproduces the stylized fact in Column 1 and tests the second prediction in Column 2, i.e., that  $\Delta \equiv A^m - A^h \geq 0$  and  $\Delta$  decreases with observability. Following evidence from Table I.3, “ $m$  lives in Mauritania” is used as a proxy for higher observability. Column 1 simply regresses the asset discrepancies on a constant and Column 2 introduces the high-observability proxy. In most cases, the constant is positive and significant, while the slope

---

and household businesses. Education is mostly public in Senegal and there are no tuition fees. Health expenditures might nevertheless be an important use of transfers.

<sup>31</sup>Conversely, migrants' places of residence within Mauritania do not show much variation along these proxies—results available upon request. Therefore, I argue that a dummy variable equal to 1 if  $m$  lives in Mauritania rather than in Europe is a better proxy for geographical distance and variation in monitoring costs.



coefficient is negative, significant and smaller than or equal to the constant in magnitude, which lends support to the prediction that geographical proximity reduces the potential for information manipulation and thus leads to a smaller  $\Delta$ .

Another way of showing  $m$ 's overestimation of  $A^h$  is to regress the ratio of  $m$ 's to  $h$ 's (average) report ( $A^m/\overline{A^h}$ ) on a constant and the “ $m$  in Mauritania” dummy.<sup>32</sup> This exercise is carried out in Appendix Table I.A.1 and yields results that are consistent with those in Table I.4.

We can further shed light on the magnitude of the discrepancies by expressing them in prices. The MIDDAS household survey indeed asked  $h$  the current sale value of the  $A^h$  held. Appendix Table I.A.2 multiplies the discrepancies by the average sale value and reproduces the approach of Table I.4.<sup>33</sup> I return to the interpretation of these results in Section 6.

Migrants select into different host countries based on unobservable characteristics. It must therefore be noted that the results from Column 2 in Table I.4 do not lend themselves to a causal interpretation. I choose to keep the specifications used to investigate  $\Delta$  as pared down as possible since inadequate control variables can exacerbate rather than cure omitted variable bias if they are correlated with the error term. Controls might also reduce the endogeneity in the variable of interest but at the cost of netting out some of its interesting effects in terms of information asymmetry. Endogeneity shall be further discussed in Section 5.3.

## 5.2 Predictions 3 to 5: Heterogeneous beliefs and learning

The next predictions that I bring to the data pertain to the fact that we observe positive discrepancies *on average* and aim at shedding light on how mi-

<sup>32</sup>I take the average of  $A^h$  in the denominator to keep in the sample households that declare  $A^h = 0$ . Results are however similar when restricting the estimation to cases where  $A^h > 0$ .

<sup>33</sup>The average rather than  $h$ 's valuation is used because assets are valued by  $h$  only if  $A^h > 0$ . The average is computed separately for migrants in different countries of destination and is restricted to assets acquired less than 5 years prior to the survey. This restriction of the sample is meant to overcome the issue of asset value depreciation, as historical prices or the year of purchase are unfortunately unknown. The result is thus a lower bound on the value of the discrepancy at the time that the remittance  $t$  was sent to acquire  $A$ .

grants process information and form beliefs about their intermediaries' trustworthiness.

Prediction 3—i.e., contrasting  $\Delta_s$  with  $\Delta_d$  to test for heterogeneous beliefs against a cognitive bias—requires us to distinguish between  $A_s$  and  $A_d$ . Classifying assets in the data is no easy task and the socio-anthropological literature provides little information on the selection of migrants into different types of investment. Ideally, we would like to distinguish between “consumption” and “investment” assets. But in a poor country like Senegal, where the informal sector accounts for a large share of the economy, most of the assets for which information is available in the MIDDAS data are consumer durables that can be put to productive use. For instance, refrigerators certainly have a direct impact on  $h$ 's living standard but can also be used to store refreshments to be sold in front of the family compound.

Since the key feature of  $A_s$  in the theoretical framework is selection, a tentative test of Prediction 3 consists of identifying  $A_d$  empirically with the most commonly held asset, TV sets. This focus is further justified by the literature, which describes this asset as a fixture of migrants' households in Senegal—see [Dia \(2007\)](#).<sup>34</sup> The discussion below however holds if one extends the  $A_d$  category to include electric fans— $h$  holds on average 1.4 fan. More generally, we shall see that evidence of misrepresentations is stronger for assets that are expensive or seldom held by migrants' households, which we interpret as evidence of selection into investment—see Appendix Table I.A.8 for statistics on average asset holdings in  $h$ .

### Measurement error

Since the stylized fact relies on a comparison of survey reports, measurement error is a potential concern. TV sets stood out in Tables I.1 and I.4 as the only asset for which we observe a negative discrepancy between migrants' and households' reports on average.<sup>35</sup> This negative  $\Delta$  is attenuated by geographical proximity and not statistically significantly different from 0 in Mauritania.

<sup>34</sup>Fatou Diome's novel *Le Ventre de l'Atlantique* (2003) features a character who had been a migrant in Paris and returned to Senegal, and who makes a point of putting his TV set outside his house for the whole village to watch football games—and witness his socio-economic success as a former migrant.

<sup>35</sup>Electric fans also display a negative, albeit insignificant, discrepancy in Tables I.1 and I.4.

To explain this pattern, we need to introduce the issue of measurement error, which is modeled and discussed in detail in Appendix I.D. I simply summarize the intuitions relevant to Prediction 3 in this section.

The key intuition is that  $A^m$  as elicited from the survey is likely to be a poorer measure of  $A$  than  $A^h$  even if there were no information asymmetry and every migrant held a correct belief about  $A$ , i.e.,  $A^m = A^h$ . The reason for this is that distance introduces noise in migrants' beliefs or answers, whereas  $h$  lives in daily proximity to the asset  $A$ , which ensures cleaner measurement. Subsequently, computing  $\Delta$  and regressing it on a constant may give rise to a negative estimate if the true  $\Delta$  is 0 but  $A^m > 0$  is measured with noise and thus attenuates to 0 on average. This is likely to explain the negative  $\Delta$  that we observe for TV sets.

Non-classical measurement error is also possible and may be exacerbated by distance; I however find that it is unlikely to drive the results. Migrants may be “clueless” about  $A^h$  and provide a random response. Now, if the random response is on average lower than  $A^h$ , cluelessness can lead to artificially negative discrepancies. I provide more detail about measurement error and cluelessness in robustness checks (Section 5.3) and Appendix I.D, but TV sets are found to require a much lower level of cluelessness to explain non-zero discrepancies—see Table I.9. This means that the negative  $\Delta$  for TV sets, which are a commonly held asset and thus have a high  $A^h$  on average, are most likely zeros pulled downwards by measurement error.<sup>36</sup>

Contrasting discrepancies for commonly held assets and for the others thus provides suggestive evidence in favor of heterogeneous beliefs: Whereas migrants seem to hold the right beliefs on average for TV sets ( $A_d$ ), for which we expect little selection, most other assets ( $A_s$ ), which probably give rise to selection, exhibit positive discrepancies as more optimistic migrants are more likely to self-select.

## Learning

Evidence of heterogeneous beliefs can be further buttressed by studying patterns of learning and updating of beliefs.

---

<sup>36</sup>Note that cluelessness should be more of a concern in the European sample. The discrepancy in terms of TV sets is indeed negative in Europe, where the required level of cluelessness to account for  $\Delta < 0$  is low (equivalently, cluelessness is likely to explain the non-zero  $\Delta$ ). In Mauritania,  $\Delta_d$  is close to 0 and cluelessness found to be less of an issue.

The theoretical framework posited that migrants' beliefs are drawn from a distribution of potential, on average unbiased beliefs about the trustworthiness of their own households of origin. These differing priors can gradually realign with the "truth" or they may diverge into heterogeneous posterior beliefs. This updating pattern should however not be observed under the cognitive bias hypothesis, as biased migrants would always fail to adjust monitoring results for the probability of a false negative.

If learning leads to a realignment of migrants' beliefs with the "truth," discrepancies should decrease as migrants gain more experience and gather information through monitoring about the use of their remittances and thus the trustworthiness of their households of origin. We would also expect the learning curve to be less steep in Mauritania, where information asymmetry is less of an issue and the additional information gathered through monitoring is less likely to lead to dramatic belief updates.

Consistent with differing prior beliefs, we see that migrants learn over time and discrepancies gradually disappear. Tables I.5 and I.6 provide empirical support for the heterogeneous-belief hypothesis by highlighting evidence of updating. They carry out a heterogeneity analysis in terms of migration experience: The discrepancies are regressed on a constant, the " $m$  in Mauritania" dummy, a variable capturing the length of the current migration spell and an interaction of these two variables. Because of the small sample size, results are not always significant, but all significant coefficients on the migration experience variable are negative, while the interaction is positive. Similar results obtain when age is used as a proxy for migration experience.<sup>37</sup>

Learning may also obtain through interactions with fellow migrants with relevant experiences. By "relevant," I mean that migrants consider news of information manipulation in other transnational households as informative about their particular situation, which leads them to update their beliefs about  $h$ . Since migrants cannot change investment intermediaries, they might however consider fellow migrants' experiences irrelevant. Another potential issue is that migrants' socialization patterns may be endogenous to their beliefs about  $h$ : For instance, if migrants gather to discuss the progress of their productive ventures in Senegal, they will only meet investor migrants, who are on average more optimistic about their intermediaries' trustworthiness,

---

<sup>37</sup>Results available upon request.

which might reinforce their optimism.

To circumvent these two pitfalls,<sup>38</sup> I rely on the number of migrants from the same household of origin. Although (migrant) siblings may differ in many respects and households that send multiple migrants abroad need not be comparable to others, migrants from the same household offer a close empirical counterpart to the idea from the theoretical framework of multiple draws from the same distribution of beliefs about  $h$ 's trustworthiness. If they do communicate about remittances to  $h$ , which both quantitative and qualitative data (from the complementary survey) support, migrants' beliefs should converge—and since I assumed that those beliefs are on average unbiased, the higher the number of migrants from the same households, the more likely it is that they converge to the “truth.” This follows from [Savage \(1954\)](#): Even if they hold different prior beliefs, two individuals who observe (through monitoring) the same sequence of signals informative about the “truth” and communicate about it will eventually hold the same posterior beliefs.

The data comply with Savage's theorem, as migrants' beliefs converge faster to the “truth” when there are more migrants from the same  $h$ . Tables I.7 and I.8 implement a similar heterogeneity analysis as Tables I.5 and I.6, except that migration experience is replaced by the number of migrants from the same household as  $m$ . Results are quite noisy, but Tables I.7 and I.8 provide suggestive evidence that the greater the number of fellow migrants, the lower the discrepancies. This effect is more muted in Mauritania, as one would expect. Tables I.5 through I.8 thus show that migrants update their beliefs. They moreover resolve the empirical issue of whether updating leads to a realignment or aggravates the misalignment of beliefs with the truth: Despite the “soft” nature of the signals obtained through monitoring, discrepancies are found to gradually vanish.<sup>39</sup>

---

<sup>38</sup>Note that the size of  $m$ 's network may also act as a proxy for monitoring activities as migrants often entrust fellow migrants—from the same village, city or region—with in-kind remittances or ask them to gather information on  $h$ .

<sup>39</sup>The point estimates however seem to suggest that convergence is rather slow—around 1-2 percentage point(s) per year for most assets.

## 5.3 Robustness checks

### 5.3.1 Linking theory to empirics

The test of the prediction that  $\Delta_d = 0$  called for a discussion of measurement error. The link between the matched variables presented in Section 4 and the theoretical constructs they are supposed to embody however deserves a more general treatment. In particular, the stylized fact that  $\Delta > 0$  for most assets ( $A_s$ ) could be driven by non-classical measurement error leading to a systematic upward bias that we would then wrongly attribute to migrants' beliefs and investment behavior.

The discrepancies being survey reports, they are indeed likely to capture measurement error. This is especially true of  $A^m$ , as distance should lead to a degradation of the quality of  $m$ 's information. Conversely, the household survey was carried out at  $h$ 's home, where the enumerators could see the assets or at least infer likely asset holdings and double-check with the respondent if the quantity stated sounded outlandish. Nevertheless, the following conclusions remain valid if I allow for measurement error in  $A^h$  as well, provided it is less severe than in  $A^m$ .

A small measurement error model is presented in Appendix I.D. I summarize its key results here.

A first concern about the interpretation of the empirical  $\tilde{\Delta}$  as a reflection of the real gap between  $m$ 's belief and the true asset holding ( $\Delta$ ) is that migrants may be "clueless" about  $A^h$  and provide a random response. Now, if the random response is on average larger (smaller) than  $A^h$ , cluelessness can lead to an artificially positive (negative)  $\tilde{\Delta}$ .<sup>40</sup> To gauge the meaningfulness of the test of Prediction 1 (i.e., of the stylized fact), I compute for each asset the propensity  $c$  of migrants to provide a clueless answer that we *would have* to assume to account for the discrepancies highlighted in Tables I.1 and I.4. Put differently, we are looking for the hypothetical  $c$  that would balance perfectly informed ( $A^m = A^h$ ) and clueless answers. The procedure is explained in Appendix I.D.1. Estimates of  $c$  are displayed in Table I.9: The higher the estimated  $c$ , the more unlikely it is that the observed  $\tilde{\Delta}$  can be attributed to cluelessness.

---

<sup>40</sup>I am indebted to Marcel Fafchamps for suggesting this source of non-classical measurement error.

Cluelessness is not sufficiently serious to jeopardize the stylized fact. Table I.9 shows that for all assets (except TV sets) one would have to assume a share of clueless migrants in excess of 50% to account for observed discrepancies in the European sample, which is unrealistic given the regular contacts between  $m$  and  $h$ , as illustrated in Table I.3, and migrants' well-documented desire to invest through  $h$ . Migrants' cluelessness is thus an unlikely confound of Prediction 1. I abstract from cluelessness in the rest of the discussion of measurement error.

A second concern pertains to two other sources of non-classical measurement error, also explored in Appendix I.D: mean reversion (Bound and Krueger, 1991) and systematic factors leading to an overreporting of  $A^m$ .

Mean-reverting measurement error does not jeopardize our test of Prediction 1 ( $\Delta \geq 0$ ) based on the reports obtained from the surveys, as it can only bias the estimate of  $\Delta$  downwards. Artificial support for Prediction 1 can thus only originate from confounding factors leading to a systematic inflation of  $A^m$ , studied in Section 5.3.2.

Prediction 2 invites us to further develop the measurement-error model. It suggests proxying  $h$ 's room for information manipulation by determinants of the cost of monitoring, such as geographical distance. The issue is that measurement error may not only affect the point estimate but also lead us to reject too often the null hypothesis that distance has no effect. Based on the proxy used in Section 5 (the “ $m$  in Mauritania” dummy), Appendix I.D.2 shows that, if anything, mean-reverting measurement error biases the point estimates and t-statistics toward 0; it thus offers a conservative test of Prediction 2. On the other hand, the confounding factors mentioned above have the opposite effect and require a more in-depth treatment—see below.

### 5.3.2 Competing interpretations of the discrepancies

Although the predictions are borne out by the data and not easily accounted for outside the framework of information asymmetry and heterogeneous beliefs, competing explanations for  $\Delta > 0$ —relying on  $\delta_O > 0$  rather than  $\delta_A > 0$ , in the notation of Appendix I.D—must be investigated.

I scrutinize the impact of several confounding factors on the asset discrepancies—see Tables I.10 through I.12. Potential culprits are the fol-

lowing.

First, the migrant and household surveys were implemented on average 4, 5 and 8 months apart for the French, Mauritanian and Italian samples, respectively. Although during this time gap Senegal experienced no major negative economic shock that led households to massively divest, the role of the time gap must be taken to the test as migrants' households of origin might follow a different growth path from the rest of the country. As the timing of the household interview was determined randomly within regions, I can use the exogenous time gap between the two surveys, standardized on the basis of  $m$ 's and  $h$ 's places of residence, to test this potential confounding factor.

Second, social desirability is a serious concern if  $m$  is more sensitive to it than  $h$  or if they react to different norms.<sup>41</sup> Insofar as social desirability bias in survey response is a product of the interaction with an enumerator, we can control for this issue thanks to a dummy for whether  $m$ 's enumerator was Senegalese or of Senegalese origin and another for whether  $m$  and her enumerator were of the same sex. Since enumerators were randomly assigned to interviewees, the effects of these variables are well identified.

Third, because of the structure of the household questionnaire, the respondent for  $h$  may have been asked to review the possessions of each "cell" of  $h$ , and some might have been more easily overlooked than in the migrant survey. If such were the case, we would however expect the number of members in  $h$  to be trimmed in a similar way in  $h$ 's report. Hence, I use the discrepancy in the size of  $h$  as a control.<sup>42</sup>

Evidence confirming the predictions of the theoretical framework is found to be robust to controlling for these likely confounding factors. Tables I.10 through I.12 show that the proxies for confounding factors usually have the expected signs when significant. " $\Delta h$  size" and "Enumerator Senegalese"

<sup>41</sup>Indeed, some respondents to the complementary survey deemed plausible that  $m$  exaggerated her role in improving  $h$ 's living standard, and subsequently  $A$ .

<sup>42</sup>It is also possible that  $h$  assumed a more restrictive definition of " $h$ 's ownership of  $A$ " than  $m$ , especially as  $A$  may belong to  $m$  although it is used by  $h$ . Transfer discrepancies raise a similar issue: Whereas the migrant questionnaire includes investment funds from  $m$  to  $h$  in remittances,  $h$  was asked about remittances received by members of  $h$ , but whether  $m$ 's investment funds should be included in those remittances was not specified. Tables I.A.3 and I.A.4 in Appendix I.A use  $\Delta t$  as a proxy (albeit endogenous) for inconsistencies in the delimitation of ownership. The transfers received and reported by  $h$  can unfortunately be unambiguously attributed to  $m$  only in the French sample but suggest that the issue of the delimitation of ownership does not explain the discrepancies.



seem the most serious confounds. As expected, a large gap in reported  $h$  sizes corresponds to a large  $\Delta A$ , and migrants faced with a Senegalese enumerator seem to have been tempted to overstate the benefits brought to  $h$  by their transfers, reacting to strong remitting norms<sup>43</sup> and competition for prestige in the Senegalese diaspora. Compared to Table I.4, we see that the confounds are seldom strong enough to knock the constant and “ $m$  in Mauritania” dummy out of significance.<sup>44</sup>

### 5.3.3 Dealing with endogeneity in the high-observability proxy

Although the previous tables lend little credence to alternative explanations for  $\Delta > 0$ , the second prediction of the model still needs to be buttressed, as the significant and robust effect of the Mauritania dummy could be due to a host of unobservables.

Migrants choosing Mauritania over Europe indeed differ along characteristics that may be correlated with trust between  $m$  and  $h$  or with migrants' ability to monitor their households of origin. Appendix Table I.A.7 shows in particular that Senegalese migrants in Mauritania stand out as being less educated and originating from poorer households. More generally, migrants could be attracted to Mauritania *because* it offers lower monitoring costs, which would induce reverse causality in the regressions of  $\Delta$  on migration destination.

I deal with the endogeneity of migrants' destinations in two ways. First, I control for variables that are likely to be correlated with the Mauritania binary variable while influencing the error terms. Second, I resort to a Two-Stage Least Squares strategy, using proximity to Mauritanian refugee camps in Senegal as an instrument for migrants' likelihood to choose Mauritania over Europe.

It is important to note that a well-identified “ $m$  in Mauritania” dummy would still lump together two different determinants of  $\Delta > 0$ : Living just across the Senegal river means that monitoring is cheaper but also that remitting capacity and transfer embezzlement are more constrained. Both channels

---

<sup>43</sup>In the complementary survey, a majority of participants feel it is a “moral obligation” to remit to the household of origin.

<sup>44</sup>Tables I.A.3 and I.A.4 do not display significant results, which could be blamed on very small sample sizes. Other potential confounds, such as delays in information transmission and  $m$ 's underestimating living costs in Senegal, were studied and found not to play a role.

operate through  $h$ 's strategic behavior: *ex post* through the monitoring of  $h$ 's actions and *ex ante* since  $m$  knows that higher remittances create stronger incentives for  $h$  to deviate from her instructions. Subsequently, this does not jeopardize the information-asymmetry hypothesis per se, but it does mean that the only way to isolate the *monitoring* channel is to control for remitting capacity directly.

### **Selection on observables**

Before introducing control variables into the tables previously presented, it is useful to summarize potential channels through which the distance proxy may affect the discrepancies and the direction of the effect. This is done in the table below. The channels consistent with the theoretical framework are in bold.

Since we are interested in the sign and significance of the high-observability proxy rather than its point estimate, an omitted variable threatens the results if the bias it induces is of the same sign as the predicted effect of distance on discrepancies, i.e., if it is negative. This means that factors such as prospective migrants' dreading information manipulation by  $h$  and thus selecting into migration to Mauritania are not serious concerns, since they bias the coefficient of interest toward 0. Similarly, if migrants who are less educated than  $h$  are both more likely to go to Mauritania than Europe and not to detect misrepresentations, a positive bias obtains. Conversely, we would expect mutual affection between  $m$  and  $h$  to both lead them to favor nearby Mauritania as  $m$ 's destination country and reduce the likelihood of nonzero  $\Delta$ . Finally, it is also plausible that poorer households are less likely to be able to send migrants to distant Europe and perhaps more inclined to follow  $m$ 's instructions, inducing a negative bias.

The results of our test of Prediction 2 are mostly robust to controlling for observable determinants of Senegalese migrants' location decisions. Tables I.13 and I.14 revisit evidence on Prediction 2 with controls for obvious confounds.<sup>45</sup> Tables I.13 and I.14 regress the discrepancies for each asset variable on the Mauritania dummy and different controls (as indicated by row

<sup>45</sup>Unfortunately, these controls cannot be argued to be well identified. As a consequence, the results should be taken with a pinch of salt and are merely presented here to scrutinize the robustness of the effect of the high-observability proxy.

Effect of  $m$  being in Mauritania on  $\Delta$ : Potential channels of impact

Channel	Sign
<b>Monitoring cost</b>	–
<b>Remitting capacity</b>	–
$m$ less educated than $h$	+
Info. asymm. problems likely	+
$h$ 's wealth	–
Mutual affection	–

titles), all interacted with the high-observability proxy. The controls are the following: A variable equal to 1 if the migrant received some schooling, while the head of her household of origin—who is usually entrusted with the remittances, as shown in Appendix Table I.A.7—received none, 0 when they have the same level of education and  $-1$  when the migrant is less educated than  $h$ 's head; a dummy equal to 1 if the migrant's spouse lives with  $h$ ; a dummy equal to 1 when  $m$  has at least one child living with  $h$ ; and an index of  $h$ 's wealth, constructed thanks to principal component analysis—two versions of the index are tested, one based on asset ownership at the time of the survey and the other five years before, to partly alleviate reverse causality concerns, albeit at the cost of reduced sample size (275 instead of 310). Migrant's income is finally controlled for to try to disentangle the monitoring and remitting capacity channels, although this approach entails a risk of netting both effects out, since monitoring is costly.

We can see from Tables I.13 and I.14 that when significant the Mauritania dummy remains robustly negative.<sup>46</sup> Besides, the controls do not lead to statistically significantly different point estimates, even though they often enter the regressions significantly—especially the wealth index and proxies for affective ties. We can further note that the results are more robust, the lower the average number of such assets held by  $h$ , which is an indication of migrants' selection into investment.

When controlling for  $m$ 's income, coefficients tend to be smaller and less often significant. When they are, they however remain of similar magnitude and sign. This suggests that monitoring is the main channel through which

<sup>46</sup>The only exception is TV sets, as in previous tables.

distance affects the size of discrepancies.

### Two-Stage Least Squares

As an additional check of the robustness of the effect of distance on asset discrepancies, I propose to use the quasi-natural experiment of Mauritanian refugee camp locations to isolate exogenous variation in Senegalese migrants' destination choices. The historical context and criteria that presided over the location of refugee camps are presented in Appendix I.C.

The key identifying assumption is that the proximity of Senegalese households to refugee camps affects misrepresentations and beliefs only through their members' destination decisions. Any direct effect of refugee camp locations on  $h$ 's trustworthiness or on trust within Senegalese households would constitute a violation of the exclusion restriction. Conversely, if (as seems plausible) the instrument influenced prospective migrants' choice of a destination by providing information on opportunities and costs (in particular, due to racism) of settling in Mauritania without altering intra-household relationships, using proximity to camps to remove the endogeneity of migration decisions is a valid strategy.

The discrepancies that constitute the stylized fact and the role of distance are mostly robust to isolating plausibly exogenous variation in migrants' destination decisions. Table I.15 regresses the asset discrepancies on a dummy for Mauritania, instrumented by the number of camps and refugee population in a 100-kilometer radius around a migrant's geo-located household of origin. To take into account the rules that presided over refugees' assignment to camps, I control for population density<sup>47</sup> and distance to a national road. Because of the small sample size and further loss of power due to the 2SLS strategy, as well as the fact that only the northern part of Senegal was exposed to refugee camps, results are presented without controls in Column 1 and with controls in Column 2. Results are qualitatively similar, albeit less precisely estimated, when additionally controlling for distance to the Mauritanian bor-

---

<sup>47</sup>I compute the average population density within 25-km radius buffers around migrants' households of origin. Density is measured thanks to a 2009 image from WorldPop—available at <http://www.worldpop.org.uk/data/files/index.php?dataset=SEN-POP&action=group>. WorldPop's spatial data sets combine satellite imagery and Census data to generate gridded predictions of population density at a 100-meter resolution.

der or the number of refugee origin villages within a 100-km radius to proxy for pre-existing networks or conflicts.<sup>48</sup> We can see from Table I.15 that the discrepancies are mostly robust: The constant and distance proxy remain of similar magnitudes or increase; they never flip sign if significant.

Looking at the first stage reported in Table I.16, we see that both the number of refugee camps and total refugee population in those camps affect migration destination choice. They work in the same direction when included separately, which suggests that the dominant effect of proximity to the camps was to foster emigration. But when controlling for their sizes, the number of camps reduces the propensity of local Senegalese households to send migrants to Mauritania. This may reflect conflicts over land with refugees or the poor image of the Mauritanian society (because of racial discrimination) that exposure to camps must have inspired—in particular if  $h$  was not living in their immediate vicinity and was thus affected by them in no other way. Conversely, the surrounding refugee population enhances the potential for contacts and information diffusion, increasing the probability to migrate to Mauritania.<sup>49</sup>

## 5.4 Alternative interpretations

The results suggest that migrants overestimate  $A_s$  but not  $A_d$ , which is consistent with information manipulation by  $h$  and heterogeneous beliefs—rather than information asymmetry and cognitive bias. Alternative interpretations are however possible. I discuss them here briefly.

The finding that  $\Delta_d = 0$  can be safely attributed to a lack of selection into investment in  $A_d$ , and thus as evidence that  $\Delta_s > 0$  reflects selection into productive asset investment, if migrants are indeed concerned with the amount of  $A_d$  that  $h$  purchases. If  $m$  derived no utility from  $A_d$ , she would still estimate it correctly on average since there is no selection. Such a focus would mean that  $\Delta_d$  provides no information about  $m$ 's beliefs:  $A_d$  would be irrelevant ( $h$  does not send a message about it) and thus we could not rule out a cognitive bias as I have defined it. The anthropological literature however suggests that

<sup>48</sup>Coefficients are also unaltered when I control for  $m$ 's age or the timing of migration, or if I restrict the sample to migrants who left Senegal after 1989.

<sup>49</sup>Based on Sargan's statistic, reported at the bottom of Table I.16, we cannot reject the null hypothesis that the overidentification restrictions are valid. Note that both the first- and second-stage results remain similar if we choose either the number of refugees or the number of refugee camps as the only excluded instrument.

migrants do send remittances for the purchase of TV sets and other goods, e.g., radios and fans, that are commonly seen in migrants' households (*post-migration*). It thus seems that migrants do care about  $A_d$ , which means that comparing  $\Delta_s$  and  $\Delta_d$  does allow us to establish that migrants hold different beliefs. Besides, as shown in Section 5, learning is at work, which is consistent with migrants' updating beliefs and converging to the "truth" despite differing priors.

Evidence that  $\Delta_d = 0$  might also simply reflect shared preferences: Whereas  $h$  has a lower preference for  $A_s$  than  $m$ , their preferences about  $A_d$  are aligned. Then  $h$  does not lie about  $A_d$  (no information manipulation) and  $\Delta_s$  cannot be unambiguously attributed to differing priors or cognitive bias. However, the literature provides us again with some indication that this is not the mechanism at work. Migrants are indeed often perceived as imposing their standards of success on their households of origin. Their insistence on the purchase of consumption and status goods, such as TV sets, is sometimes frowned upon by their left-behinds as a waste of money—see for instance [Dia \(2007\)](#). This implies that  $m$ 's target level of  $A_d$  is in excess of the optimal level according to  $h$ , i.e., of the quantity that  $h$  would purchase in autarky with an increase in income equivalent to the remittance received.

The same pattern would obtain if we observed only satiated migrants: Perhaps  $A_s$  and  $A_d$  differ also in terms of income elasticity and assets purchased early in the migration spell no longer give rise to remittances and therefore information manipulation. Satiation cannot be easily ruled out. Assuming that migrants become richer over time, we would expect a negative relationship between migration experience and  $\Delta_d$ , which we would confuse with learning. But this is not what we see in Tables I.5 and I.6 for TV sets: We observe a rather cleanly estimated 0, a pattern that reinforces the conclusion that  $A_d$  is not subject to selection based on optimism and does not exhibit satiation. The same argument can be used to reject the possibility that Tables I.5 and I.6 capture improvements in the monitoring technology, which would lead to a decrease in information manipulation and thus in the observed discrepancies over time. The fact that  $A_d$  does not display such a pattern, although we expect monitoring to also play a role in its case, confirms that the time gradient in the observed discrepancies can be attributed to learning and the updating of beliefs.

Finally, if  $h$  has a stronger preference for  $A_d$  than  $m$ , diverts money to

purchase  $A_d$  rather than  $A_s$  and  $m$  monitors  $A_d$  as a second source of information on  $h$ 's type, we would have  $\Delta_s > 0$  and  $\Delta_d < 0$  despite an absence of investment in  $A_d$  by  $m$ . Since  $\Delta_d$  would in this context be the mirror image of  $\Delta_s$ , comparing these two discrepancies would not enable us to conclude that migrants have differing priors. The discussion of measurement error (Appendix I.D) however makes it clear that a negative observed  $\tilde{\Delta}$  is likely when the true  $\Delta$  is 0 but  $A^m$  is measured with more noise than  $A^h$ , which is probably the case. Table I.9 further shows that a significant (negative)  $\Delta_d$  is most likely due to the cluelessness of part of the migrant sample (28% in Mauritania, for instance), while most migrants (72%) correctly estimate  $A_d^h$ .

## 6 Discussion and conclusion

### 6.1 Policy implications

Information asymmetry and migrants' heterogeneous beliefs may have far-reaching consequences for the volume and use of remittances, which many developing countries rely on.

A back-of-the-envelope calculation indeed suggests that the discrepancies are large. Their magnitude can be assessed by summing the average discrepancies (at their current sale value)—see Appendix Table I.A.2—across assets, dividing by the number of years  $m$  has spent at destination, and then comparing this quantity with annual remittances.<sup>50</sup> On average, Senegalese migrants in Europe lose 10.70% of their total annual remittances to what we interpret as deviations from their spending instructions. The effect is smaller in Mauritania, but still represents 5.71% of annual flows on average.

Furthermore, as migrants anticipate embezzlement from their indispensable intermediaries, they may be reluctant to invest in profitable businesses. Since only the most optimistic migrants select into productive investments, worthwhile projects may go unfunded while others are never realized as funds are diverted to the recipients' preferred uses.

A first set of possible interventions would target information asymmetries by facilitating migrants' control over remittances. New international transfer operators, such as Niokobok or Afrimarket in Senegal, offer alongside tradi-

<sup>50</sup>This crude assessment does not consider selection in and out of investment because of learning.

tional money transfer services the opportunity to make in-kind remittances through their websites.<sup>51</sup> Encouraging the use of such services, through preferential taxation or by simplifying administrative approval procedures, might help migrants overcome information asymmetry issues. Information asymmetries can also be reduced by a matching grants scheme, such as the “Tres Por Uno” program in Mexico, which facilitates investments by migrants’ hometown associations in social infrastructures and productive projects, and thus bypasses a potential diversion of remittances by individual recipients—see for instance Aparicio and Meseguer (2012).

Another set of policies would diversify migrants’ options in terms of business partners. Van den Steen (2004) notes that “overoptimism increases with the number of alternatives from which the agent can choose.” Offering migrants alternatives to investing through their households of origin can thus be expected to increase the share of migrants who invest back home and the volume of money remitted for investment purposes. Although mobile banking has been hailed as an opportunity for leapfrogging in developing countries, services traditionally offered by banks (and which could be phone- or Internet-based) such as real estate programs or shares in productive ventures might help migrants avoid falling back on the option of saving at destination. Banks may also appear as less likely to embezzle migrants’ funds, whereas with higher-quality monitoring technologies such as in-kind transfers, signals remain soft as the household could collude with shopkeepers, resell the goods purchased thanks to vouchers or fail to make complementary (e.g., maintenance) investments. Moreover, solutions to information manipulation by investment intermediaries do not suffice when it comes to reducing migrants’ selection into investment projects based on optimism.

## 6.2 Conclusion

The thread of this paper is a stylized fact emerging from matched data on Senegalese migrants interviewed in France, Italy and Mauritania and their households of origin in Senegal: Migrants tend to systematically overestimate or overstate the number of assets held by their households of origin.

---

<sup>51</sup>Niokobok and Afrimarket offer a similar range of products, from food to durable assets and construction materials, but the former delivers goods at the household’s doorstep whereas the latter sells vouchers that can be redeemed at partner stores.



This pattern fits nicely into the picture of the Senegalese diaspora painted in the socio-anthropological literature. Senegalese migrants are indeed keen on sending remittances to invest in productive endeavors or real estate back home, or refurbish or equip their households' compounds. Remittance recipients on the other hand may have different preferences and be vulnerable to stronger pressures to share or spend, thereby causing conflicts and strategic behavior over the use of remittances.

Based on socio-anthropological evidence, I delineate a theoretical framework for migrants' remittance behavior with two key features: information asymmetry and differing prior beliefs about the trustworthiness of their households of origin as investment intermediaries. This framework allows me to derive a few testable predictions: (i) As suggested by the stylized fact, even if much of the two parties' maximization problems is common knowledge, and allowing for monitoring, remittance recipients may still find it optimal to divert targeted funds. (ii) Information asymmetry should be exacerbated by reduced observability of the household's actions, primarily because of distance. (iii) Gaps between remittance senders' and recipients' reports should be observed in the data only when migrants select into the purchase of an asset based on their beliefs, i.e., if more optimistic migrants are more likely to invest. (iv) If differing priors are at work, we would expect migrants to update their beliefs based on new information from monitoring. The last two predictions help me distinguish between two possible causes of the discrepancies: heterogeneous beliefs and a cognitive bias in migrants' adjustment for the limitations of the monitoring technology.

The empirical part of the paper tests the predictions from the theoretical framework. The findings support the prediction of increased information manipulation by remittance recipients when observability is lower and provide suggestive evidence of heterogeneous beliefs rather than a cognitive bias. Furthermore, migrants appear to learn over time and through interactions with fellow migrants, which gradually leads to smaller discrepancies.

Since a comparison of reports may capture measurement error, I show that the results hold under different measurement error models, and that there are no systematic biases in survey responses. This allows me to conclude that transfer recipients do manipulate information and observed gaps in survey reports are evidence of gaps in information sets and heterogeneous beliefs.

The existence and persistence of information asymmetries within transnational households, despite repeated contacts prior to migration and the use of monitoring, and the misalignment of migrants' beliefs with the "truth" implies that the developmental promises of remittances may fail to materialize.

Whether improving migrants' information and beliefs is beneficial to the transnational household as a whole and development in general remains an open question that we reserve for future research. But evidence suggests that migrants are more inclined to invest their money productively—with the durable assets studied in this paper as inputs into small, family-run enterprises—and thus secure a livelihood upon return, than more consumption-oriented remittance recipients.

Migrants' control over the use of their remittances and better assessment of the profitability of their business projects can thus be expected to foster productive investment in developing countries.

## Tables

Table I.1 – Discrepancies between  $m$ 's and  $h$ 's reports of  $A$  (number of assets)

	Automobile (1)	Bicycle (2)	CD-player (3)
Constant	0.14*** (0.04)	0.41*** (0.07)	0.27*** (0.07)
Observations	316	316	314
	DVD-player (4)	Fan (5)	Freezer (6)
Constant	0.37*** (0.06)	-0.08 (0.14)	0.21*** (0.03)
Observations	315	317	317
	Motorcycle (7)	Radio (8)	Refrigerator (9)
Constant	0.22*** (0.05)	0.23** (0.12)	0.18*** (0.04)
Observations	317	317	318
	TV set (10)		
Constant	-0.22** (0.08)		
Observations	316		

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . This table displays the results from a regression of  $\Delta \equiv A^m - A^h$ , the discrepancy between  $m$ 's and  $h$ 's reports of asset holding  $A$ , on a constant. The constant represents the mean  $\Delta$  for each asset.

Table I.2 – Sample size and composition by country

	France	Italy	Mauritania	All
<hr/> Stage 1: Migrant samples <hr/>				
Number of surveyed migrants	302	303	327	932
... % of women	24.2	22.8	36.4	28.1
<hr/> Stage 2: Origin household samples <hr/>				
Matching rate (%)	29.5	20.1	53.2	34.9
Number of matched households	89	61	174	324
... % in urban areas	71.9	91.8	62.6	70.7

The table displays the number of observations by country in the MIDDAS sample (“Stage 1”), as well as statistics on the number and share of migrant households that could be successfully surveyed in Senegal (“Stage 2”).

Table I.3 – Correlation between migration destination and monitoring proxies

	(1)	(2)	(3)
	Nb of visits to $h$ per year	Days since last visit to $h$	% of total remittances to $h$ in kind
$m$ in Mauritania	0.65*** (0.14)	-190.38*** (61.24)	8.90*** (1.51)
Constant	0.82*** (0.05)	794.68*** (43.72)	2.49*** (0.36)
Observations	895	711	757

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The constant corresponds to the European mean dependent variable.  $m$  in Mauritania captures the effect of  $m$  being in Mauritania instead of Europe.

Table I.4 – Discrepancies between  $m$ 's and  $h$ 's reports of  $A$  (number of assets)  
– Distance gradient

	Automobile		Bicycle		CD-player	
	(1)	(2)	(1)	(2)	(1)	(2)
Constant	0.14*** (0.04)	0.27*** (0.07)	0.41*** (0.07)	0.72*** (0.14)	0.27*** (0.07)	0.50*** (0.10)
$m$ in Mauritania		-0.24*** (0.07)		-0.59** (0.15)		-0.42*** (0.13)
	316	316	316	316	314	314

	DVD-player		Fan		Freezer	
	(1)	(2)	(1)	(2)	(1)	(2)
Constant	0.37*** (0.06)	0.35*** (0.10)	-0.08 (0.14)	-0.07 (0.25)	0.21*** (0.03)	0.30*** (0.06)
$m$ in Mauritania		0.04 (0.11)		-0.03 (0.28)		-0.17** (0.07)
Observations	315	315	317	317	317	317

	Motorcycle		Radio		Refrigerator	
	(1)	(2)	(1)	(2)	(1)	(2)
Constant	0.22*** (0.05)	0.34*** (0.09)	0.23** (0.12)	0.34** (0.20)	0.18*** (0.04)	0.24*** (0.07)
$m$ in Mauritania		-0.22** (0.10)		-0.22 (0.24)		-0.11 (0.08)
	317	317	317	317	318	318

	TV set	
	(1)	(2)
Constant	-0.22** (0.08)	-0.43*** (0.12)
$m$ in Mauritania		0.41** (0.17)
	316	316

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The constant in Column 1 (2) corresponds to the (European) mean  $\Delta$ .  $m$  in Mauritania captures the effect on  $\Delta$  of  $m$  being in Mauritania instead of Europe.

Table I.5 – Heterogeneity analysis – Migration experience (1/2)

	Automobile		Bicycle	
	(1)	(2)	(1)	(2)
Constant	0.267*** (0.053)	0.293*** (0.055)	0.722*** (0.105)	0.761*** (0.110)
$m$ in Mauritania	-0.237*** (0.073)	-0.266*** (0.079)	-0.595*** (0.145)	-0.647*** (0.159)
Years of migration		-0.147** (0.060)		-0.138 (0.120)
Years of migration * $m$ in Mauritania		0.140* (0.085)		0.108 (0.171)
Observations	316	309	316	309

	CD-player		DVD-player	
	(1)	(2)	(1)	(2)
Constant	0.497*** (0.096)	0.520*** (0.100)	0.353*** (0.082)	0.343*** (0.082)
$m$ in Mauritania	-0.424*** (0.132)	-0.483*** (0.142)	0.041 (0.113)	0.001 (0.118)
Years of migration		-0.153 (0.107)		-0.159* (0.089)
Years of migration * $m$ in Mauritania		0.069 (0.153)		0.046 (0.127)
Observations	314	307	315	308

	Fan		Freezer	
	(1)	(2)	(1)	(2)
Constant	-0.066 (0.198)	0.185 (0.199)	0.305*** (0.050)	0.331*** (0.053)
$m$ in Mauritania	-0.030 (0.274)	-0.254 (0.423)	-0.172** (0.069)	-0.239*** (0.075)
Years of migration		-0.686*** (0.216)		-0.065 (0.057)
Years of migration * $m$ in Mauritania		0.749** (0.309)		-0.028 (0.081)
Observations	317	310	317	310

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The constant in Column 1 (2) corresponds to the (European) mean  $\Delta$ .  $m$  in Mauritania captures the effect on  $\Delta$  of  $m$  being in Mauritania instead of Europe. *Years of migration* is the standardized number of years spent by the migrant at destination during the current migration spell.

Table I.6 – Heterogeneity analysis – Migration experience (2/2)

	Motorcycle		Radio	
	(1)	(2)	(1)	(2)
Constant	0.338*** (0.072)	0.110** (0.050)	0.344** (0.169)	0.431** (0.177)
<i>m</i> in Mauritania	-0.223** (0.100)	0.016 (0.072)	-0.218 (0.233)	-0.288 (0.255)
Years of migration		0.084 (0.054)		-0.390** (0.193)
Years of migration * <i>m</i> in Mauritania		-0.100 (0.078)		0.428 (0.275)
Observations	317	310	317	310

	Refrigerator		TV set	
	(1)	(2)	(1)	(2)
Constant	0.238*** (0.057)	0.250*** (0.060)	-0.430*** (0.122)	-0.351* (0.053)
<i>m</i> in Mauritania	-0.113 (0.079)	-0.154* (0.086)	0.412** (0.168)	0.373** (0.179)
Years of migration		-0.041 (0.065)		0.020 (0.135)
Years of migration * <i>m</i> in Mauritania		-0.027 (0.093)		0.074 (0.193)
Observations	318	311	316	309

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The constant in Column 1 (2) corresponds to the (European) mean  $\Delta$ . *m* in Mauritania captures the effect on  $\Delta$  of *m* being in Mauritania instead of Europe. *Years of migration* is the standardized number of years spent by the migrant at destination during the current migration spell.

Table I.7 – Heterogeneity analysis – Number of migrants from same  $h$  (1/2)

	Automobile		Bicycle	
	(1)	(2)	(1)	(2)
Constant	0.192 (0.053)	0.251*** (0.081)	0.722*** (0.105)	0.626*** (0.161)
$m$ in Mauritania	-0.237*** (0.0725)	-0.307*** (0.112)	-0.595*** (0.145)	-0.449** (0.224)
# migrants from $h$		-0.018 (0.012)		0.018 (0.023)
# migrants from $h$ * $m$ in Mauritania		0.014 (0.015)		-0.026 (0.030)
Observations	316	315	316	315

	CD-player		DVD-player	
	(1)	(2)	(1)	(2)
Constant	0.497*** (0.096)	0.460*** (0.147)	0.353*** (0.082)	0.553*** (0.124)
$m$ in Mauritania	-0.424*** (0.132)	-0.428** (0.205)	0.041 (0.113)	-0.253 (0.172)
# migrants from $h$		0.007 (0.021)		-0.037** (0.018)
# migrants from $h$ * $m$ in Mauritania		0.000 (0.028)		0.054** (0.023)
Observations	314	313	315	314

	Fan (electric)		Freezer	
	(1)	(2)	(1)	(2)
Constant	-0.066 (0.198)	-0.187 (0.305)	0.305*** (0.050)	0.357*** (0.077)
$m$ in Mauritania	-0.030 (0.274)	0.184 (0.422)	-0.172** (0.069)	-0.313*** (0.106)
# migrants from $h$		0.023 (0.043)		-0.010 (0.011)
# migrants from $h$ * $m$ in Mauritania		-0.037 (0.057)		0.025* (0.014)
Observations	317	316	317	316

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The constant in Column 1 (2) corresponds to the (European) mean  $\Delta$ .  $m$  in Mauritania captures the effect on  $\Delta$  of  $m$  being in Mauritania instead of Europe. # migrants from same  $h$  is the number of international emigrants (excluding  $m$ ) who originate from the same household  $h$  as  $m$ .



Table I.8 – Heterogeneity analysis – Number of migrants from same  $h$  (2/2)

	Motorcycle		Radio	
	(1)	(2)	(1)	(2)
Constant	0.338*** (0.072)	0.482*** (0.110)	0.344** (0.169)	0.807*** (0.257)
$m$ in Mauritania	-0.223** (0.100)	-0.373** (0.153)	-0.218 (0.233)	-0.528 (0.356)
# migrants from $h$		-0.027* (0.016)		-0.086** (0.036)
# migrants from $h$ * $m$ in Mauritania		0.028 (0.021)		0.062 (0.048)
Observations	317	316	317	316

	Refrigerator		TV set	
	(1)	(2)	(1)	(2)
Constant	0.238*** (0.057)	0.132 (0.087)	-0.430*** (0.122)	-0.275 (0.187)
$m$ in Mauritania	-0.113 (0.079)	-0.031 (0.119)	0.412** (0.168)	0.378 (0.258)
# migrants from $h$		0.020 (0.012)		-0.029 (0.026)
# migrants from $h$ * $m$ in Mauritania		-0.017 (0.016)		0.009 (0.035)
Observations	318	317	316	315

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The constant in Column 1 (2) corresponds to the (European) mean  $\Delta$ .  $m$  in Mauritania captures the effect on  $\Delta$  of  $m$  being in Mauritania instead of Europe. # migrants from same  $h$  is the number of international emigrants (excluding  $m$ ) who originate from the same household  $h$  as  $m$ .

Table I.9 – “Cluelessness” ( $c$ ) necessary to account for observed discrepancies

	Automobile (1)	Bicycle (2)	CD-player (3)
$m$ in Mauritania	0.13 [0.06, 0.21]	0.33 [0.21, 0.44]	0.63 [0.49, 0.77]
$m$ in Europe	0.73 [0.58, 0.89]	0.79 [0.64, 0.95]	0.91 [0.75, 1.07]
Observations	316	316	314
	DVD-player (4)	Fan (5)	Freezer (6)
$m$ in Mauritania	0.78 [0.63, 0.93]	0.69 [0.54, 0.83]	0.42 [0.30, 0.55]
$m$ in Europe	0.80 [0.64, 0.96]	0.60 [0.45, 0.74]	0.75 [0.60, 0.91]
Observations	315	317	317
	Motorcycle (7)	Radio (8)	Refrigerator (9)
$m$ in Mauritania	0.34 [0.22, 0.55]	0.73 [0.59, 0.88]	0.56 [0.43, 0.70]
$m$ in Europe	0.58 [0.44, 0.73]	0.66 [0.51, 0.81]	0.66 [0.51, 0.81]
Observations	317	317	318
	TV set (10)		
$m$ in Mauritania	0.36 [0.25, 0.48]		
$m$ in Europe	0.28 [0.17, 0.39]		
Observations	316		

95% confidence intervals in brackets. The table displays the level of  $c$  (Equation I.2 in Appendix I.D) needed to explain the  $\Delta$  observed in Europe and in Mauritania.

Table I.10 – Impact of confounding factors on estimated asset discrepancies (1/3)

	Automobile				Bicycle			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Constant	0.14*** (0.04)	0.11 (0.07)	0.27*** (0.07)	0.22** (0.10)	0.41*** (0.07)	0.30*** (0.08)	0.72*** (0.14)	0.67*** (0.14)
$m$ in Mauritania			-0.24*** (0.07)	-0.19** (0.09)			-0.59*** (0.15)	-0.61*** (0.14)
Standardized time gap		-0.01 (0.04)		-0.01 (0.04)		-0.07 (0.10)		-0.07 (0.09)
Enumerator and interviewee same sex		0.07 (0.09)		0.05 (0.09)		0.24* (0.14)		0.16 (0.13)
Enumerator Senegalese		-0.05 (0.09)		-0.02 (0.09)		0.00 (0.17)		0.09 (0.16)
$\Delta h$ size		0.01* (0.01)		0.01* (0.01)		0.02** (0.01)		0.02* (0.01)
Observations	316	273	316	273	316	273	316	273
F all confounds = 0		1.59		1.05		1.66		1.08
p-value		0.18		0.38		0.16		0.37

	CD-player				DVD-player			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Constant	0.27*** (0.07)	0.08 (0.12)	0.50*** (0.10)	0.37** (0.16)	0.37*** (0.06)	0.36*** (0.08)	0.35*** (0.10)	0.32*** (0.12)
$m$ in Mauritania			-0.42*** (0.13)	-0.48*** (0.15)			0.04 (0.11)	0.06 (0.12)
Standardized time gap		0.03 (0.09)		0.02 (0.08)		-0.01 (0.06)		-0.01 (0.06)
Enumerator and interviewee same sex		0.13 (0.14)		0.06 (0.14)		-0.28** (0.12)		-0.27** (0.12)
Enumerator Senegalese		0.45*** (0.16)		0.52*** (0.16)		0.45*** (0.14)		0.44*** (0.14)
$\Delta h$ size		0.02** (0.01)		0.01* (0.01)		0.01** (0.01)		0.01** (0.01)
Observations	314	271	314	271	315	272	315	272
F all confounds = 0		3.23		3.50		4.35		4.11
p-value		0.01		0.01		0.00		0.00

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The constant in Columns 1–2 (3–4) corresponds to the (European) mean  $\Delta$ .  $m$  in Mauritania captures the effect on  $\Delta$  of  $m$  being in Mauritania instead of Europe.

Table I.11 – Impact of confounding factors on estimated asset discrepancies (2/3)

	Fan (electric)				Freezer			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Constant	-0.08 (0.14)	-0.08 (0.21)	-0.07 (0.25)	-0.05 (0.33)	0.21*** (0.03)	0.13** (0.05)	0.30*** (0.06)	0.23*** (0.09)
$m$ in Mauritania			-0.03 (0.28)	-0.04 (0.31)			-0.17** (0.07)	-0.17* (0.09)
Standardized time gap		0.20 (0.17)		0.20 (0.17)		-0.03 (0.04)		-0.04 (0.04)
Enumerator and interviewee same sex		-0.26 (0.27)		-0.27 (0.28)		0.19** (0.08)		0.17** (0.08)
Enumerator Senegalese		0.49* (0.29)		0.50 (0.30)		0.02 (0.10)		0.04 (0.10)
$\Delta h$ size		0.04** (0.02)		0.04** (0.02)		0.00 (0.00)		0.00 (0.00)
Observations	317	274	317	274	317	274	317	274
F all confounds = 0		2.37		2.34		1.86		1.40
p-value		0.05		0.06		0.12		0.24

	Motorcycle				Radio			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Constant	0.22*** (0.05)	0.05 (0.06)	0.34*** (0.09)	0.24** (0.11)	0.23** (0.12)	0.15 (0.19)	0.34* (0.20)	0.20 (0.30)
$m$ in Mauritania			-0.22** (0.10)	-0.31** (0.13)			-0.22 (0.24)	-0.09 (0.29)
Standardized time gap		-0.03 (0.05)		-0.03 (0.05)		0.12 (0.14)		0.12 (0.14)
Enumerator and interviewee same sex		0.20* (0.11)		0.16 (0.11)		-0.04 (0.22)		-0.05 (0.24)
Enumerator Senegalese		0.39** (0.16)		0.43*** (0.16)		0.35 (0.30)		0.37 (0.31)
$\Delta h$ size		0.00 (0.01)		-0.00 (0.01)		0.05*** (0.02)		0.05*** (0.02)
Observations	317	274	317	274	317	274	317	274
F all confounds = 0		3.61		3.78		3.19		3.08
p-value		0.01		0.01		0.01		0.02

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The constant in Columns 1–2 (3–4) corresponds to the (European) mean  $\Delta$ .  $m$  in Mauritania captures the effect on  $\Delta$  of  $m$  being in Mauritania instead of Europe.

Table I.12 – Impact of confounding factors on estimated asset discrepancies (3/3)

	Refrigerator				TV set			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Constant	0.18*** (0.04)	0.19*** (0.06)	0.24*** (0.07)	0.27*** (0.09)	-0.22** (0.08)	-0.17 (0.13)	-0.43*** (0.12)	-0.45** (0.18)
<i>m</i> in Mauritania			-0.11 (0.08)	-0.14 (0.09)			0.41** (0.17)	0.46** (0.20)
Standardized time gap		0.09* (0.05)		0.09* (0.05)		-0.12 (0.10)		-0.12 (0.10)
Enumerator and inter- viewee same sex		0.04 (0.09)		0.02 (0.09)		-0.14 (0.18)		-0.08 (0.19)
Enumerator Senegalese		-0.09 (0.09)		-0.07 (0.09)		0.16 (0.18)		0.09 (0.18)
$\Delta h$ size		0.00 (0.00)		0.00 (0.00)		0.03*** (0.01)		0.04*** (0.01)
Observations	318	275	318	275	316	273	316	273
F all confounds = 0		1.20		1.00		3.97		4.88
p-value		0.31		0.41		0.00		0.00

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The constant in Columns 1–2 (3–4) corresponds to the (European) mean  $\Delta$ . *m* in Mauritania captures the effect on  $\Delta$  of *m* being in Mauritania instead of Europe.

Table I.13 – Impact of geographical distance on  $\Delta$ , controlling for likely confounding factors (1/2)

	(1) Auto- mobile	(2) Bicycle	(3) CD- player	(4) DVD- player	(5) Fan (electric)
No controls	-0.24*** (0.07)	-0.59*** (0.15)	-0.42*** (0.13)	0.04 (0.11)	-0.03 (0.28)
<i>m</i> more educated than <i>h</i> 's head	-0.20** (0.08)	-0.50*** (0.14)	-0.46*** (0.17)	0.05 (0.13)	-0.13 (0.33)
Spouse resident in household of origin	-0.23*** (0.09)	-0.64*** (0.17)	-0.27** (0.13)	0.01 (0.13)	-0.03 (0.31)
At least one child lives in household of origin	-0.25** (0.10)	-0.68*** (0.19)	-0.28* (0.15)	0.03 (0.15)	0.02 (0.34)
<i>h</i> 's wealth index, <i>h</i> 's report	-0.25*** (0.07)	-0.62*** (0.15)	-0.44*** (0.12)	0.02 (0.12)	-0.11 (0.25)
<i>h</i> 's wealth index (5 y. bef. survey), <i>h</i> 's report	-0.25*** (0.08)	-0.60*** (0.17)	-0.37*** (0.14)	0.05 (0.13)	-0.09 (0.30)
All controls but income	-0.25** (0.11)	-0.61*** (0.16)	-0.30* (0.15)	0.02 (0.15)	-0.20 (0.35)
<i>m</i> 's total income (€)	-0.34*** (0.12)	-0.43* (0.23)	-0.21 (0.18)	-0.01 (0.20)	0.02 (0.47)
All controls	-0.39** (0.16)	-0.52** (0.25)	-0.01 (0.21)	0.03 (0.24)	0.09 (0.50)

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The table displays the Mauritania dummy coefficients, controlling for different factors in each row. Interactions between the Mauritania dummy and the controls are not reported.

Table I.14 – Impact of geographical distance on  $\Delta$ , controlling for likely confounding factors (2/2)

	(1)	(2)	(3)	(4)	(5)
	Freezer	Motor- cycle	Radio	Refrige- rator	TV set
No controls	-0.17** (0.07)	-0.22** (0.10)	-0.22 (0.24)	-0.11 (0.08)	0.41** (0.17)
<i>m</i> more educated than <i>h</i> 's head	-0.17** (0.07)	-0.23** (0.10)	-0.29 (0.26)	-0.18** (0.09)	0.22 (0.21)
Spouse resident in household of origin	-0.13 (0.08)	-0.18* (0.11)	-0.12 (0.26)	-0.15 (0.10)	0.42** (0.19)
At least one child lives in household of origin	-0.13 (0.09)	-0.22* (0.12)	-0.26 (0.28)	-0.15 (0.11)	0.41* (0.21)
<i>h</i> 's wealth index, <i>h</i> 's report	-0.19*** (0.07)	-0.24** (0.10)	-0.26 (0.22)	-0.10 (0.08)	0.40*** (0.14)
<i>h</i> 's wealth index (5 y. bef. survey), <i>h</i> 's report	-0.15** (0.07)	-0.28*** (0.11)	-0.19 (0.23)	-0.03 (0.08)	0.51*** (0.16)
All controls but income	-0.15* (0.09)	-0.28*** (0.13)	-0.40 (0.28)	-0.24** (0.11)	0.15 (0.24)
<i>m</i> 's total income (€)	-0.08 (0.12)	-0.11 (0.17)	-0.07 (0.33)	0.01 (0.14)	0.74*** (0.28)
All controls	-0.07 (0.15)	-0.27 (0.20)	-0.05 (0.38)	-0.05 (0.17)	0.66** (0.32)

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The table displays the Mauritania dummy coefficients, controlling for different factors in each row. Interactions between the Mauritania dummy and the controls are not reported.

Table I.15 – Discrepancies and distance – Instrumenting for  $m$ 's destination

	Automobile		Bicycle		CD-player	
	(1)	(2)	(1)	(2)	(1)	(2)
Constant	0.15 (0.10)	0.03 (0.16)	0.78*** (0.18)	1.07*** (0.30)	0.59*** (0.17)	0.72*** (0.28)
$m$ in Mauritania	-0.05 (0.17)	0.12 (0.24)	-0.72** (0.31)	-1.18*** (0.44)	-0.58** (0.28)	-0.71* (0.41)
Estimation	IV	IV	IV	IV	IV	IV
Controls	No	Yes	No	Yes	No	Yes
Observations	288	288	288	288	286	286

	DVD-player		Fan (electric)		Freezer	
	(1)	(2)	(1)	(2)	(1)	(2)
Constant	0.40*** (0.15)	0.24 (0.24)	-0.18 (0.36)	-0.14 (0.59)	0.38*** (0.09)	0.48*** (0.15)
$m$ in Mauritania	-0.06 (0.25)	0.10 (0.36)	0.17 (0.61)	0.07 (0.87)	-0.32** (0.16)	-0.40* (0.22)
Estimation	IV	IV	IV	IV	IV	IV
Controls	No	Yes	No	Yes	No	Yes
Observations	287	287	289	289	289	289

	Motorcycle		Radio		Refrigerator	
	(1)	(2)	(1)	(2)	(1)	(2)
Constant	0.42*** (0.12)	0.35* (0.19)	0.23 (0.33)	0.25 (0.53)	0.18* (0.10)	0.08 (0.16)
$m$ in Mauritania	-0.37* (0.20)	-0.30 (0.28)	0.02 (0.55)	0.00 (0.78)	-0.03 (0.17)	0.08 (0.24)
Estimation	IV	IV	IV	IV	IV	IV
Controls	No	Yes	No	Yes	No	Yes
Observations	289	289	289	289	290	290

	TV set	
	(1)	(2)
Constant	-0.11 (0.24)	0.21 (0.39)
$m$ in Mauritania	-0.16 (0.40)	-0.43 (0.57)
Estimation	IV	IV
Controls	No	Yes
Observations	288	288

Standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The constant corresponds to the European mean  $\Delta$ .  $m$  in Mauritania captures the effect on  $\Delta$  of  $m$  being in Mauritania instead of Europe. The instruments are the total number of refugees living in camps and number of camps of at least 100 people within a 100-km radius around  $h$ . Controls include the minimum distance between  $h$  and a national road, and the average population density within a 25-km radius around  $h$ .



Table I.16 – Instrumenting for  $m$ 's destination – First stage

	$m$ in Mauritania
# of refugees (1,000)	0.17*** (0.03)
# of refugee camps	-0.04*** (0.01)
Distance to road	0.00 (0.00)
Population density (1,000/km <sup>2</sup> )	-0.02*** (0.00)
Observations	289
F-statistic	20.59
Sargan's statistic	2.36

Standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The endogenous variable  $m$  in Mauritania is equal to 1 if the migrant resides in Mauritania and 0 if  $m$  lives in Europe. The excluded instruments are # of refugees, which represents the total number of Mauritanian refugees (in thousands) living in camps of 100 inhabitants or more within a 100-km radius of  $h$ , and # of refugee camps, based on the same buffer definition. Distance to road is the minimum distance in kilometers between  $h$  and a national road. Population density is the average number of people per square kilometer within a 25-km radius buffer centered on  $h$ . The F-statistic displayed is the Sanderson-Windmeijer multivariate F-statistic.

## ONLINE APPENDIX

### I.A Additional tables

Table I.A.1 – Ratio of  $m$ 's and  $h$ 's reports of  $A$  (number of assets) – Distance gradient

	Automobile		Bicycle		CD-player	
	(1)	(2)	(1)	(2)	(1)	(2)
Constant	1.88*** (0.37)	2.10*** (0.27)	3.77*** (0.47)	4.81*** (0.64)	1.68*** (0.14)	2.12*** (0.20)
$m$ in Mauritania		-0.42 (0.70)		-1.96** (0.92)		-0.83*** (0.27)
	322	322	322	322	322	322

	DVD-player		Fan		Freezer	
	(1)	(2)	(1)	(2)	(1)	(2)
Constant	2.36*** (0.17)	1.60*** (0.14)	0.91*** (0.08)	0.92*** (0.11)	2.78*** (0.30)	2.45*** (0.25)
$m$ in Mauritania		1.42*** (0.32)		-0.03 (0.15)		0.54 (0.57)
Observations	322	322	322	322	322	322

	Motorcycle		Radio		Refrigerator	
	(1)	(2)	(1)	(2)	(1)	(2)
Constant	4.17*** (0.63)	5.75*** (1.15)	1.21*** (0.08)	1.35** (0.14)	1.41*** (0.10)	1.34*** (0.09)
$m$ in Mauritania		-2.95** (1.30)		-0.25 (0.16)		0.11 (0.19)
	322	322	322	322	322	322

	TV set	
	(1)	(2)
Constant	0.88*** (0.04)	-0.76*** (0.05)
$m$ in Mauritania		0.22*** (0.08)
	322	322

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The constant in Column 1 (2) corresponds to the (European) mean ratio of  $m$ 's report to  $h$ 's average report of  $A$ ,  $A^m/A^h$ , where  $A^h$  is the average  $A^h$  computed by destination country of  $m$ .  $A^m/A^h > 1$  ( $< 1$ ) means that  $m$  over- (under-) estimates  $A^h$  and thus likely over- (under-) estimates  $A^h$ .  $m$  in Mauritania captures the effect on  $A^m/A^h$  of  $m$  being in Mauritania instead of Europe.

Table I.A.2 – Discrepancies between  $m$ 's and  $h$ 's reports of  $A$  valued at current sale prices (in euros) – Distance gradient

	Automobile		Bicycle		CD-player	
	(1)	(2)	(1)	(2)	(1)	(2)
Constant	419.70*** (106.12)	800.83*** (197.74)	11.29*** (2.05)	19.45*** (3.72)	8.24*** (1.83)	15.82*** (3.30)
$m$ in Mauritania		-725.53*** (215.87)		-15.63*** (4.11)		-14.43*** (3.68)
	317	317	316	316	314	314
	DVD-player		Fan		Freezer	
	(1)	(2)	(1)	(2)	(1)	(2)
Constant	8.58*** (1.43)	10.37*** (2.73)	-2.26 (2.23)	-3.49*** (4.39)	34.07*** (5.52)	48.46*** (9.07)
$m$ in Mauritania		-3.42 (2.96)		2.34 (4.63)		-27.47** (11.11)
Observations	315	315	317	317	317	317
	Motorcycle		Radio		Refrigerator	
	(1)	(2)	(1)	(2)	(1)	(2)
Constant	48.39*** (10.72)	77.67*** (20.17)	8.73** (3.89)	15.65** (7.66)	25.77*** (5.71)	40.93*** (10.95)
$m$ in Mauritania		-55.97** (21.94)		-13.21 (8.06)		-28.87** (11.75)
	317	317	317	317	318	318
TV set						
	(1)	(2)				
Constant	-24.20*** (7.39)	-49.48*** (13.29)				
$m$ in Mauritania		48.41*** (14.92)				
	316	316				

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The constant in Column 1 (2) corresponds to the (European) mean  $\Delta$ .  $\Delta$  is reweighted by the average current sale price (in euros) of the asset as stated by  $h$ . This average is computed separately for migrants in different countries of destination and is restricted to assets acquired less than 5 years prior to the survey.  $m$  in Mauritania captures the effect on  $\Delta$  of  $m$  being in Mauritania instead of Europe.

Table I.A.3 – Impact of discrepancies in remittance reports ( $\Delta t$ ) on estimated asset discrepancies ( $\Delta A$ )—French sample only (1/2)

	Automobile				Bicycle			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Constant	0.16*	0.21	0.10	0.15	0.74***	0.87**	0.55***	0.28
	(0.08)	(0.18)	(0.07)	(0.13)	(0.23)	(0.43)	(0.19)	(0.36)
$\Delta t$ (€1000)	-0.01	-0.00	-0.01	-0.00	0.09	0.10	0.07	0.04
	(0.03)	(0.02)	(0.02)	(0.02)	(0.08)	(0.10)	(0.07)	(0.08)
$m$ 's total income (incl. social benefits)		-0.00		-0.00		-0.00		0.00
		(0.00)		(0.00)		(0.00)		(0.00)
Observations	50	49	75	72	50	49	75	72

	CD-player				DVD-player			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Constant	0.28*	0.35	0.26**	0.15	0.04	0.10	0.07	0.16
	(0.14)	(0.26)	(0.12)	(0.20)	(0.16)	(0.29)	(0.13)	(0.22)
$\Delta t$ (€1000)	0.01	0.02	-0.00	-0.02	0.00	0.01	-0.01	0.00
	(0.03)	(0.04)	(0.03)	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)
$m$ 's total income (incl. social benefits)		-0.00		0.00		-0.00		-0.00
		(0.00)		(0.00)		(0.00)		(0.00)
Observations	50	49	75	72	50	49	75	72

	Fan (electric)				Freezer			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Constant	-0.96***	-0.81	-0.58*	-0.38	0.21**	0.26	0.19**	0.33*
	(0.25)	(0.49)	(0.31)	(0.48)	(0.09)	(0.21)	(0.08)	(0.19)
$\Delta t$ (€1000)	0.09	0.10	0.05	0.07	0.04*	0.04	0.02	0.03
	(0.09)	(0.11)	(0.08)	(0.08)	(0.02)	(0.03)	(0.02)	(0.03)
$m$ 's total income (incl. social benefits)		-0.00		-0.00		-0.00		-0.00
		(0.00)		(0.00)		(0.00)		(0.00)
Observations	50	49	75	72	50	49	75	72

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . For each variable  $\Delta A$ , Columns 1–2 (3–4) define  $\Delta t$  as missing (0) when no transfer by  $m$  was reported by  $h$ .

Table I.A.4 – Impact of discrepancies in remittance reports ( $\Delta t$ ) on estimated asset discrepancies ( $\Delta A$ )—French sample only (2/2)

	Motorcycle				Radio			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Constant	0.24 (0.15)	-0.04 (0.32)	0.17 (0.13)	0.06 (0.26)	-0.37 (0.27)	-0.04 (0.45)	-0.30 (0.26)	-0.25 (0.35)
$\Delta t$ (€1000)	-0.00 (0.05)	-0.04 (0.06)	0.06 (0.06)	0.04 (0.07)	0.04 (0.09)	0.09 (0.14)	0.02 (0.09)	0.02 (0.11)
$m$ 's total income (incl. social benefits)		0.00 (0.00)		0.00 (0.00)		-0.00 (0.00)		-0.00 (0.00)
Observations	50	49	75	72	50	49	75	72

	Refrigerator				TV set			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Constant	0.09 (0.07)	0.15 (0.16)	0.03 (0.08)	-0.16 (0.18)	-0.68*** (0.16)	-0.89*** (0.31)	-0.58*** (0.14)	-0.63** (0.25)
$\Delta t$ (€1000)	0.04 (0.02)	0.05 (0.04)	0.01 (0.02)	-0.01 (0.03)	0.01 (0.05)	-0.01 (0.08)	0.01 (0.04)	0.01 (0.06)
$m$ 's total income (incl. social benefits)		-0.00 (0.00)		0.00 (0.00)		0.00 (0.00)		0.00 (0.00)
Observations	50	49	75	72	50	49	75	72

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . For each variable  $\Delta A$ , Columns 1–2 (3–4) define  $\Delta t$  as missing (0) when no transfer by  $m$  was reported by  $h$ .

Table I.A.5 – Selection into the matched sample (1/2)

	(1) France	(2) Italy	(3) Mauritania
<b>Migrant characteristics:</b>			
Female	-0.030 (0.061)	-0.036 (0.053)	-0.047 (0.058)
Age	0.001 (0.002)	0.003 (0.003)	0.004 (0.003)
Son/Daughter of <i>h</i> 's head	0.053 (0.057)	-0.013 (0.052)	0.070 (0.057)
No formal education	0.045 (0.071)	-0.091 (0.062)	-0.042 (0.057)
University degree	0.013 (0.061)	0.065 (0.055)	0.136 (0.274)
<b>Migration history:</b>			
Years since arrival in host country	-0.006** (0.003)	0.001 (0.004)	-0.000*** (0.000)
<b>Economic situation:</b>			
<i>m</i> 's total annual income (incl. social benefits, €)	-0.036 (0.026)	0.066*** (0.023)	0.069*** (0.026)
Unemployed	0.044 (0.078)	-0.018 (0.056)	0.127 (0.087)
<b>Family situation:</b>			
At least one child lives with <i>h</i>	0.023 (0.068)	0.022 (0.051)	0.111** (0.057)
Spouse resident in <i>h</i>	-0.086 (0.071)	0.003 (0.052)	0.108 (0.067)
<b>Characteristics of the household of origin:</b>			
<i>h</i> 's wealth score, <i>m</i> 's report	0.011 (0.029)	0.098*** (0.021)	0.058 (0.039)

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The dependant variable is a dummy equal to 1 if *h* could be surveyed and 0 otherwise. It is regressed on each variable separately. Income and wealth variables are standardized by country to allow for different living standards.

Table I.A.6 – Selection into the matched sample (2/2)

	(1)	(2)	(3)
	France	Italy	Mauritania
<b>Remittance behavior:</b>			
Total transfers to $h$ in the past 12 months (in kind incl., €)	0.018 (0.034)	0.050 (0.036)	0.038 (0.034)
Total money transfers to $h$ in the past 12 months, €	0.015 (0.033)	0.053 (0.036)	0.020 (0.035)
<b>Enumerator characteristics</b>			
Enumerator Senegalese citizen	0.032 (0.075)	0.279** (0.131)	0.137** (0.062)
Enumerator female	0.016 (0.060)	-0.140*** (0.051)	0.106 (0.067)

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The dependant variable is a dummy equal to 1 if  $h$  could be surveyed and 0 otherwise. It is regressed on each variable separately. Transfer variables are standardized by country to allow for different living standards. Remittance amounts are computed on the subsample of migrants with non-zero transfers.  $m(h)$  refers to the migrant (household of origin).

Table I.A.7 – Selection into destinations

	(1) Europe	(2) Mauritania	(3) Difference	(4) p-value
<b>Migrant characteristics:</b>				
Female	0.24	0.37	-0.13***	0.00
Age	36.49	35.85	0.64	0.37
Son/Daughter of <i>h</i> 's head	0.66	0.56	0.10***	0.00
No formal education	0.14	0.41	-0.26***	0.00
University degree	0.26	0.01	0.25***	0.00
<b>Migration history:</b>				
Years since arrival in host country	11.68	6.27	5.41***	0.00
<b>Relationship to <i>h</i>:</b>				
Main remittance recipient is:				
<i>m</i> 's father/mother	0.59	0.59	0.00	0.93
<i>m</i> 's spouse	0.08	0.10	-0.02	0.48
<i>h</i> 's head	0.69	0.74	-0.04	0.28
a woman	0.53	0.51	0.02	0.54
<i>m</i> more educated than <i>h</i> 's head	0.31	0.24	0.07	0.29
At least one child lives with <i>h</i>	0.29	0.40	-0.11***	0.00
Spouse resident in <i>h</i>	0.22	0.22	0.01	0.85
<b>Household characteristics:</b>				
<i>h</i> 's wealth score (5 years ago)	0.23	-0.20	0.43***	0.00

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . This table displays t-tests. *m* (*h*) refers to the migrant (household of origin). Variables pertaining to the relationship between *m* and *h* all refer to *m*'s household of origin. *h*'s wealth score is the first principal component of a vector of assets acquired at least 5 years prior to the survey; it is based on *h*'s asset reports.



Table I.A.8 – Asset holdings in household of origin

	(1) Average $A$ holdings	(2) Share with $A > 0$	(3) Unit value of $A$ (€)
Automobile	0.13 (0.37)	0.12 (0.32)	2108.70 (1121.65)
Bicycle	0.17 (0.67)	0.09 (0.29)	25.45 (12.14)
CD-player	0.34 (0.86)	0.24 (0.43)	19.59 (8.29)
DVD-player	0.37 (0.62)	0.30 (0.46)	19.99 (11.26)
Fan (electric)	1.40 (1.90)	0.57 (0.50)	12.30 (8.53)
Freezer	0.14 (0.40)	0.12 (0.33)	94.44 (39.70)
Motorcycle	0.13 (0.45)	0.09 (0.29)	150.21 (94.73)
Radio	1.21 (1.37)	0.70 (0.46)	20.94 (31.43)
Refrigerator	0.45 (0.63)	0.40 (0.49)	82.59 (43.17)
TV set	1.46 (1.39)	0.84 (0.36)	60.60 (38.64)

Standard deviations in parentheses. Column 1 displays the average number of asset  $A$  held by  $h$ , Column 2 the share of households that hold at least one item of  $A$ , and Column 3 the average unit price in euros of  $A$  (restricted to items acquired less than 5 years prior to the survey) at the current sale value. All statistics in the table are computed based on  $h$ 's reports.

## I.B Remittances and information in Senegalese transnational households

The Senegalese diaspora offers a suitable setting for the study of information asymmetry in the context of international migration and remittances. Emigration has a long history there, both to neighboring African countries such as Mauritania and to distant Europe. It is also a major support of the Senegalese economy: 11.9% of its GDP was accounted for by international remittances in 2015 (World Bank, 2017).<sup>52</sup>

Therefore, even if geographical dispersion is a hotbed of information asymmetry and manipulation, we would expect Senegalese transnational households to have developed all possible hedges against it.

Nevertheless, the socio-anthropological literature on the Senegalese diaspora points to strategic behavior and conflicts over the use of remittances, which constrains migrants in their ability to invest through their households of origin.

Once their families' daily needs are provided for,<sup>53</sup> migrants aspire to set up businesses back home and build a house for their relatives and themselves upon return.<sup>54</sup>

Crucially, the remittances sent by migrants for investment are almost exclusively entrusted to their relatives. This is because close kin will be the ones enjoying the migrant's investment in her absence and because migrants feel unable to find reliable intermediaries outside the family (Fall, 2003).

Recipients are however not given a free hand to spend the transfers. Money is sent back home and the onus is on relatives living in the household of origin to buy inputs into the business or purchase the materials needed for the construction of a new house (Dia, 2007). More generally, a complementary

---

<sup>52</sup>It is interesting to note that 76% of Senegalese households' transfer receipts originate from family members (De Vreyer et al., 2008).

<sup>53</sup>Senegalese migrants call the irreducible, and usually flat, monthly remittance that they send outside investment funds the "DQ," which stands for *dépense quotidienne* (daily expenditure).

<sup>54</sup>Building a house back home is a major concern for Senegalese migrants (Fall, 2003; Dia, 2007). It can be seen as an investment in social capital. Expanding, refurbishing and equipping the family compound through remittances indeed embodies migrants' desire to return to their households of origin and offers a tangible signal of their efforts to maintain ties (Boltz-Laemmel and Villar, 2014).

survey I carried out for this study<sup>55</sup> shows that a third of respondents earmark remittances for particular purchases, through in-kind transfers or verbal instructions.

Invested funds are nevertheless subject to diversion by those in charge in Senegal. Migrants often “bemoan the lack of people worthy of trust among their kith and kin” (Fall, 2003),<sup>56</sup> and according to Marfaing (2003) the majority of migrants have experienced the failure of their business projects in Senegal as business funds are regularly swallowed up in the event of emergencies.

The main reasons behind such a diversion of remittances are preference differences, self-control issues and redistributive pressure (kin tax).

Migrants' preferences over the use of remittances need not be in line with their relatives' back home. To the contrary, conflicts with relatives in Senegal over money, its use and the implementation of investment projects creates a household economy characterized by “a struggle for controlling the resources from migration” (Dia and Adamou, 2003). Migrants on the other hand fear that members of the household of origin embezzle the money,<sup>57</sup> do not purchase all that is necessary or fail to expend effort to maintain the assets purchased.

Different members of the transnational household are also subject to different pressures. Most of the time, migrants are not directly solicited for financial assistance; but their “left-behinds,” first and foremost their spouses, are pressured into redistributing as requests from closer relatives are more difficult to turn down (Boltz-Laemmel and Villar, 2014).

---

<sup>55</sup>Because it was not the sole focus of the matched data that I use for the empirics, a short questionnaire was developed to delve further into information asymmetry, and implemented in October 2012 in Château Rouge. This area of Paris is densely populated by and attracts (from the rest of the conurbation) many Senegalese migrants. The survey assesses whether information asymmetry is a concern for the migrant with questions on earmarking of transfers, suspicion, monitoring and sanctions. At the end of the interview, respondents were presented with the stylized fact identified in the matched data and asked to put forward the most plausible hypotheses. Great care was taken in letting the interviewees express themselves and in gathering qualitative information. Due to the qualitative dimension of the survey, only 20 migrants could be interviewed.

<sup>56</sup>Quotes from Dia and Adamou (2003) and Fall (2003) are translated by the author.

<sup>57</sup>Respondents often mentioned fancy clothes or participation in “baby naming ceremonies” (*ngénte* in Wolof) as the destination of the embezzled transfers.

Conflict over the use of remittances is not the only instrument that recipients and senders can resort to in order to impose their views. Information manipulation is another.<sup>58</sup>

The complementary survey asked respondents to speculate on the stylized fact emerging from the matched data: Most maintain that the discrepancies observed are due to the household of origin lying to the migrant to extract more or secure remittances.<sup>59</sup> The most frequent story pertains to durable goods not being purchased, contrary to the migrant's wishes, or sold if bought by the migrant directly.

This suggests information asymmetry and manipulation, whereby relatives back home exploit private information about how remittances are spent to further their own interests. Half the interviewees in the complementary survey had doubts about the information received from their main transfer recipients and thought their earmarking was not followed or information was distorted to extract rents. As one of them put it: "We only know what they tell us."

Migrants however do exert themselves to improve the information they receive. This may mean reducing the fungibility of transfers by sending them to shopkeepers so as to allocate them beforehand (Dia, 2007), remitting in kind<sup>60</sup> and monitoring transfer recipients through phone calls, and contacts with other migrants or neighbors from the same community. Migrants also make the most of their visits to Senegal to manage their projects (Fall, 2003; Dia, 2007).

If information manipulation is detected, migrants can rely on two main

---

<sup>58</sup>It is worth noting here that migrants may use private information about their earnings or their situation in the destination country to reduce the amounts they have to remit (Chort et al., 2012; McKenzie et al., 2013; Ambler, 2015). Migrants—the ones actually making remittance decisions—being more inclined to invest in productive ventures back home, I shall focus on information asymmetry about the household of origin's actions.

<sup>59</sup>Out of 20 respondents, 15 indicted the household of origin for the observed discrepancies; 7 of those understood the stylized fact immediately and spontaneously answered without the interviewer listing possible options; and 12 came up with an anecdote or an example to illustrate their answers.

<sup>60</sup>Customers' comments posted on the website of Niokobok, a firm specializing in in-kind transfers to Senegal, eloquently reflect the relationship between control issues and in-kind remittances: For instance, one customer explains that it allows her/him to "make sure the money is really used by the household" as "you are always afraid that the money is not managed well when you are far away," according to another—see <https://www.niokobok.com/reviews-7.html>.

channels to mete out sanctions: future investments and reputation. In the complementary survey, respondents explained that they would heap opprobrium on their households of origin in their extended networks or put an end to investment remittances to punish them for deviating from their instructions.

## I.C 2SLS strategy

Section 5.3 provides evidence of the role of distance in creating information asymmetries in the remittance context based on a quasi-natural experiment that offers us an instrument for Senegalese migration destination choices. In this Appendix, I briefly (i) present the historical context that led to the establishment of Mauritanian refugee camps in Senegal, (ii) describe the criteria that presided over camp location decisions, and (iii) discuss potential mechanisms behind the first stage.

On April 9, 1989 a violent clash between Senegalese farmers and Mauritanian herders triggered deadly riots in the capitals of Senegal and Mauritania. In a context of extreme racial tensions in Mauritania, only two years after a failed coup led by “Negro-Mauritanian” military officers to topple a government dominated by the Arab-Berber community, the incident was used as an excuse to expel the “Black Moors” from southern Mauritania to Senegal (while Mauritanian immigrants fled in the opposite direction). Almost overnight, in an escalation that surprised the international community and Senegalese government, 371 “Black Moor” villages were emptied of their populations, their inhabitants haphazardly loaded onto trucks—families being sometimes separated in the confusion—and deported to the border with Senegal, where the refugees crossed the river, quite low in this season (Trémolières and Gnisci, 2004).

The refugees were then dispatched to 276 camps of various sizes—from less than 100 people to several thousands (ibid.).<sup>61</sup> Assignment to a camp depended on where the refugees crossed the river (Santoir, 1990). Camps were set up at a distance (usually about 200 meters) from villages and close to a

---

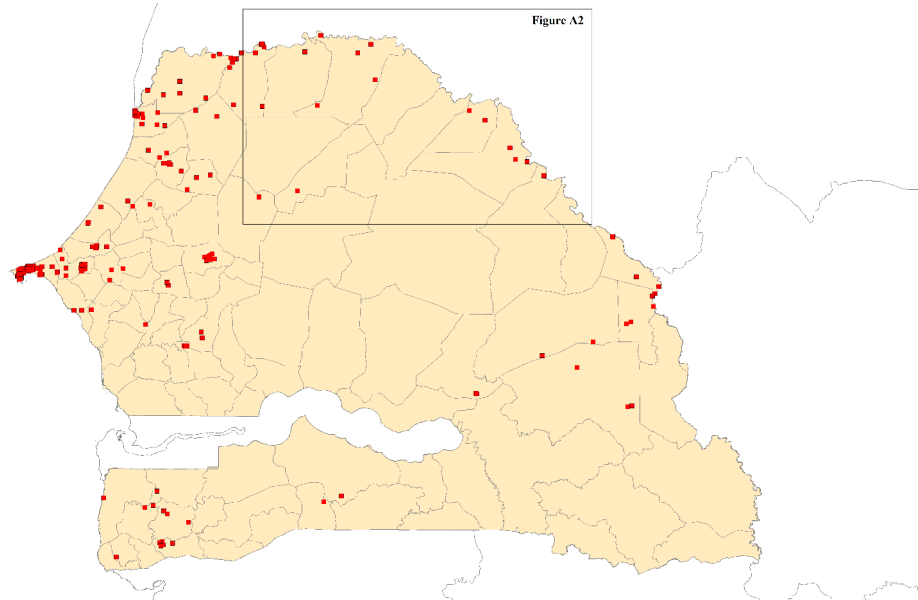
<sup>61</sup>There are no official estimates of the number of Mauritians who took refuge in Senegal in 1989. A widely cited figure is 70,000.

national road to facilitate food distribution (Trémolières and Gnisci, 2004). Figure I.C.2 below represents the camps in Senegal and origin villages of the refugees in Mauritania. The map focuses on the Senegal river valley, where all refugee camps were located. Data on the camps and origin villages of refugees comes from Fresia (2006). They cover larger camps with at least 100 refugees. The total population of those 170 camps is at least 47,000.<sup>62</sup>

Mauritanian refugee camps are likely to have influenced the attitudes of the neighboring Senegalese, including migrants' origin households, whose spatial distribution is displayed in Figure I.C.1. The direction of the effect is however an empirical issue. Relationships with locals were at times difficult, in particular because of intense competition for land (Trémolières and Gnisci, 2004). This might have fueled resentment against the Mauritanian people, which can be expected to have had an adverse effect on prospective migrants' likelihood to choose Mauritania as a destination. Such an adverse effect is however unlikely to have played a major role at longer distances from camps.<sup>63</sup> Conversely, contacts with refugees may have weaved transnational networks after refugees were repatriated between the mid-1990s and late 2000s, which could have fostered Senegalese migration to Mauritania. Finally, proximity to refugee camps may have simply improved information about the risks and prospects of working in Mauritania. Knowledge of the institutionalized racism that led to the expulsion of "Black Moors" from the northern bank of the Senegal river may have deterred some Senegalese from settling there. On the other hand, emigration may have been catalyzed as households living near a refugee camp should have had a better idea of the situation in Mauritania, e.g., economic growth, sectors with buoyant labor demand, or the re-opening of the border with Senegal. Table I.C.1 provides suggestive evidence that the effect of the "treatment" goes through information (or intolerance) by showing that the instrument has no effect on the size and composition (i.e., Senegalese vs. Mauritanian members) of migrants' networks.

<sup>62</sup>Data on the number of refugees per camp is truncated above 1,000 in Fresia (2006).

<sup>63</sup>The specification used in Table I.15 uses as instruments the number of camps and refugee population within a 100-km radius of  $h$ .

Figure I.C.1 – Migrants' households of origin ( $h$ ).

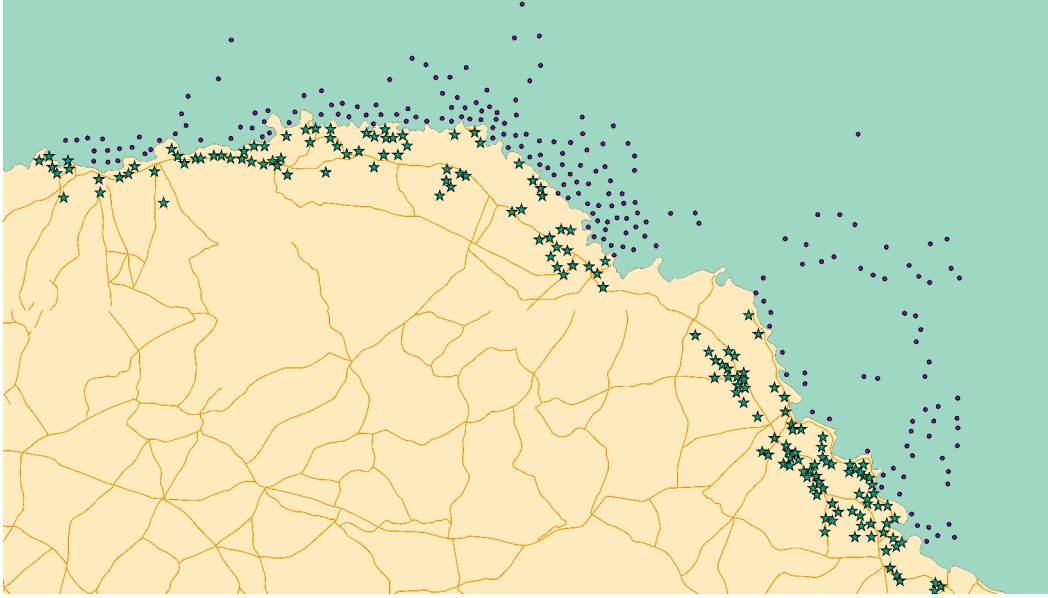
Source: Map by the author, based on MIDDAS data. Notes: Migrants' origin households are represented within Senegal (salmon pink) by red squares if surveyed. GPS coordinates were measured by the enumerators at  $h$ 's home. Grey lines within Senegal are the borders of districts (*départements*).

Table I.C.1 – Effect of proximity to refugee camps on migrants' networks in Mauritania

	Total network size (1)	Share of Mauritians in network (2)
# refugees (1,000)	-0.033 (0.104)	-0.026 (0.026)
# refugee camps	0.009 (0.031)	0.010 (0.007)
Observations	135	135

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The sample is restricted to migrants living in Mauritania. The dependent variables are *Total network size*, which is the number of people in  $m$ 's network in Mauritania, and the *Share of Mauritians in  $m$ 's network* at destination. The main regressors are # of refugees, which represents the total number of Mauritanian refugees (in thousands) living in camps of 100 inhabitants or more within a 100-km radius of  $h$ , and # of refugee camps, based on the same buffer definition.

Figure I.C.2 – Mauritanian refugee camps and origin villages.



Source: Map by the author, based on [Fresia \(2006\)](#). Notes: This map zeroes in on the rectangle drawn on Figure I.C.1. For clarity, district borders and migrants' origin households are ignored, and Mauritanian territory is painted in teal. Mauritanian refugee camps in Senegal (salmon pink) are represented by green stars; their origin villages in Mauritania are materialized by purple diamonds. Roads in Senegal are shown in orange. All camps and origin villages appear on the map.

## I.D Measurement error

I model  $h$ 's and  $m$ 's survey answers as follows:

$$\widetilde{A}^h = A^h \tag{I.1}$$

$$\widetilde{A}^m = (1 - c)(1 + \zeta + \delta_A + \delta_O)A^h + cr + \nu \tag{I.2}$$

where  $\widetilde{A}^i$  denotes the (potentially error-ridden) report of the true information  $A^i$  held by  $i = h, m$ ;  $\zeta \in [-1, 0]$  captures a negative correlation between the report and the error-free variable, which [Bound and Krueger \(1991\)](#) call “mean-reverting measurement error” and is particularly likely for bounded variables such as  $A$ ;  $\delta_A \geq 0$  corresponds to the inflation in  $h$ 's message to  $m$  possible under information asymmetry and observable in survey reports under differing priors *cum* selection (see Section 3);  $\delta_O \geq 0$  stands for factors leading to a systematic inflation of the migrant's report;  $\nu$  is such that  $E[\nu] = 0$  and  $E[A^h\nu] = 0$ ; and  $c$  is the migrant's propensity to provide a “clueless” random



answer  $r$ . I shall deal with cluelessness first and then discuss the impact of the other sources of measurement error.

Two points are worth noting. First,  $A$  in Equations I.1 and I.2 refers to durable asset holdings. Asset holdings  $A$  are strongly unimodal ratio variables stated by  $m$  and  $h$  themselves. Second, Equation I.2 immediately tells us that Prediction 1 ( $\Delta \geq 0$ ) can be tested with  $\tilde{\Delta}$  as long as  $\delta_O = 0$ . Indeed,  $\zeta$  being negative it can only bias the estimate of  $\Delta$  downwards. As far as Prediction 2 is concerned, I show in Appendix I.D.2 that we obtain a conservative estimate of the role of the  $m$  in Mauritania indicator variable even if the effect of  $\zeta$  is allowed to vary with distance. Given our assumptions about  $\nu$ , artificial support for Prediction 1 can only originate in the positive  $\delta_O$ , studied in Section 5.3.2, or from cluelessness—see below.

### I.D.1 Cluelessness

Considering for simplicity that  $A^i$  is a binary variable, i.e., equal to 1 if  $h$  owns at least one unit of  $A$  according to  $i$  and 0 otherwise, an intuitive way of modeling clueless migrants' behavior is to assume that their answers follow a Bernoulli distribution of mean .5, responding 0 or 1 with equal probability. I assume that the probability  $c$  of answering in a clueless manner is the same for all respondents but may differ across assets.<sup>64</sup>

Rewriting Equation I.2 to focus on cluelessness, we have:

$$\tilde{A}^m = (1 - c)A^h + cr \quad (\text{I.3})$$

$$\therefore c = \frac{\tilde{\Delta}}{r - A^h} \quad (\text{I.4})$$

Since we assumed that  $r = .5$ , we can replace  $c$  by  $\tilde{\Delta}/(.5 - A^h)$ . Averaging over all individuals in the sample yields the desired quantity  $c$ .

Estimates of  $c$  are displayed in Table I.9.

---

<sup>64</sup>Assuming the converse yields qualitatively similar results—available upon request.

## I.D.2 Measurement error in the test of Prediction 2

Rephrasing the measurement error model to focus on discrepancies in reports, we get:

$$\begin{aligned}\tilde{\Delta} &= \tilde{A}^m - \tilde{A}^h = (1 + \zeta + \delta_A + \delta_O)A^h - A^h + \nu \\ &= \Delta + (\zeta + \delta_O)A^h + \nu\end{aligned}$$

where  $\Delta \equiv \delta_A A^h$  represents the true discrepancy due to information asymmetry and differing priors, which I intend to highlight empirically.<sup>65</sup>

### I.D.2.1 Focusing on $\zeta$

Focusing on  $\zeta$ , we have  $\tilde{\Delta} = \Delta + \zeta A^h + \nu$  and we want to estimate  $\Delta = \beta MAU + \varepsilon$ , where  $MAU$  is a dummy variable equal to 1 if  $m$  is in Mauritania and 0 if she lives in Europe, and  $\varepsilon$  is assumed to be uncorrelated with  $MAU$  and to have zero expectation. Since being in Mauritania both increases observability and reduces potential remittances, which leads to a decrease in  $\Delta$ —see Section 5.3,—we expect  $\beta \leq 0$ .

Using OLS, we obtain:

$$\hat{\beta} = \frac{\text{cov}(MAU, \tilde{\Delta})}{\text{var}(MAU)} = \frac{\text{cov}(MAU, \Delta + \zeta A^h + \nu)}{\text{var}(MAU)} \quad (\text{I.5})$$

$$= \frac{\text{cov}(MAU, \beta MAU + \zeta A^h + \nu + \varepsilon)}{\text{var}(MAU)} = \frac{\beta \sigma_{MAU}^2 + \zeta \sigma_{MAU, A^h}}{\sigma_{MAU}^2} \quad (\text{I.6})$$

$$= \beta + \zeta \beta_{A^h, MAU} \geq \beta \quad (\text{I.7})$$

where  $\beta_{A^h, MAU}$  denotes the coefficient on  $MAU$  in a regression with  $A^h$  as the dependent variable. Given the lower living standard in Mauritania than in Europe,  $\beta_{A^h, MAU}$  is expected to be negative and the data support this conjecture. Therefore,  $\hat{\beta}$  overestimates  $\beta$ , which goes counter to Prediction 2.

What about the effect of  $\zeta$  on the significance of  $\hat{\beta}$ ? From Equation I.7,

<sup>65</sup>The results are robust to introducing information manipulation into Equation I.2 additively rather than as a fraction of  $A^h$ ,  $\delta_A A^h$ , because they do not rely on the correlation between  $\Delta$  and  $A^h$ —see below.

we have:

$$\text{plim}\widehat{\beta} - \beta = \zeta\beta_{A^h,MAU} \quad (\text{I.8})$$

$$\text{and} \quad \widehat{\varepsilon} = \widetilde{\Delta} - \widehat{\beta}MAU = \Delta + \zeta A^h + \nu - \widehat{\beta}MAU \quad (\text{I.9})$$

$$= \varepsilon - (\Delta - \beta MAU) + \Delta + \zeta A^h + \nu - \widehat{\beta}MAU \quad (\text{I.10})$$

$$= \varepsilon - (\beta - \widehat{\beta})MAU + \zeta A^h + \nu \quad (\text{I.11})$$

Calling  $\sigma_\varepsilon^2$  the variance of  $\varepsilon$  and  $\widehat{\sigma}_\varepsilon^2$  its estimator (and using a similar notation for  $\nu$ ), we have:

$$\text{plim}\widehat{\sigma}_\varepsilon^2 = \sigma_\varepsilon^2 + \zeta^2\beta_{A^h,MAU}^2\sigma_{MAU}^2 + \zeta^2\sigma_{A^h}^2 + \sigma_\nu^2 + \zeta^2\beta_{A^h,MAU}\sigma_{A^h,MAU} \quad (\text{I.12})$$

$$\begin{aligned} \therefore \text{plim}\widehat{s} &\equiv \text{plim}\frac{\widehat{\sigma}_\varepsilon^2}{\sigma_{MAU}^2} \\ &= \frac{\sigma_\varepsilon^2 + \zeta^2\beta_{A^h,MAU}^2\sigma_{MAU}^2 + \zeta^2\sigma_{A^h}^2 + \sigma_\nu^2 + \zeta^2\beta_{A^h,MAU}\sigma_{A^h,MAU}}{\sigma_{MAU}^2} \end{aligned} \quad (\text{I.13})$$

$$= s + \zeta^2\beta_{A^h,MAU}^2 + \zeta^2\frac{\sigma_{A^h}^2}{\sigma_{MAU}^2} + \frac{\sigma_\nu^2}{\sigma_{MAU}^2} + \zeta^2\beta_{A^h,MAU} \quad (\text{I.14})$$

$$= s + 2\zeta^2\beta_{A^h,MAU} + \zeta^2\frac{\sigma_{A^h}^2}{\sigma_{MAU}^2} + \frac{\sigma_\nu^2}{\sigma_{MAU}^2} \quad (\text{I.15})$$

$$\therefore \frac{\text{plim}t}{\sqrt{n}} = \frac{\text{plim}\widehat{\beta}}{\text{plim}\sqrt{\widehat{s}}} = \frac{\beta + \zeta\beta_{A^h,MAU}}{\sqrt{s + 2\zeta^2\beta_{A^h,MAU} + \zeta^2\frac{\sigma_{A^h}^2}{\sigma_{MAU}^2} + \frac{\sigma_\nu^2}{\sigma_{MAU}^2}}} \quad (\text{I.16})$$

$$\therefore \left| \frac{\text{plim}t}{\sqrt{n}} \right| \leq \left| \frac{\beta}{\sqrt{s}} \right| \quad (\text{I.17})$$

Since the numerator in Equation I.16 is greater than  $\beta$  and  $\beta$  is expected to be negative,  $\zeta$  biases the t-statistics toward 0 (and may even yield the wrong sign on  $\beta$ ). The denominator is clearly greater than  $\sqrt{s}$ , which also biases t-statistics downwards and thus provides a conservative test of Prediction 2.

### I.D.2.2 Allowing mean-reverting measurement error to be exacerbated by distance

It is likely that the element in  $\tilde{A}^m$  that is negatively correlated with  $A^h$  increases with the geographical distance between  $m$  and  $h$ . I model this as follows:

$$\tilde{\Delta} = \Delta + \zeta(1 - \eta MAU)A^h + \nu$$

where  $\eta \in [0; 1]$ .

Adapting Equation I.5 yields:

$$\hat{\beta} = \frac{\text{cov}(MAU, \Delta + \zeta(1 - \eta MAU)A^h + \nu)}{\text{var}(MAU)} \quad (\text{I.18})$$

$$= \frac{\text{cov}(MAU, \beta MAU + \zeta(1 - \eta MAU)A^h + \nu + \varepsilon)}{\text{var}(MAU)} \quad (\text{I.19})$$

$$= \frac{\beta \sigma_{MAU}^2 + \zeta \sigma_{MAU, A^h} - \zeta \eta \sigma_{MAU, A^h MAU}}{\sigma_{MAU}^2} \quad (\text{I.20})$$

$$= \beta + \zeta \frac{\sigma_{MAU, A^h} - \eta \sigma_{MAU, A^h MAU}}{\sigma_{MAU}^2} \quad (\text{I.21})$$

Based on the binary nature of  $MAU$ , it can be shown that  $\sigma_{MAU, A^h MAU} \geq 0$ . Since  $\zeta \leq 0$ ,  $\eta \geq 0$  and  $\sigma_{A^h, MAU} \leq 0$ , it still holds that  $\hat{\beta} \geq \beta$ , i.e., it is biased towards 0.

### I.D.3 Focusing on $\delta_O$

Focusing on  $\delta_O$ , we now have:

$$\hat{\beta} = \beta + \delta_O \beta_{A^h, MAU} \leq \beta \leq 0 \quad (\text{I.22})$$

because  $\delta_O \geq 0$  and  $\beta_{A^h, MAU} \leq 0$ .  $\delta_O$  therefore represents a serious confound in the test of Prediction 2, considered in Section 5.3.2.

CHAPTER II

# Migrants and Firms: Evidence from China

---

## Abstract<sup>1</sup>

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

**Keywords:** Migration, Economic Development, Structural Change, Transitional Economies, China.

**JEL classification:** D24, J23, J61, O15.

---

<sup>1</sup>This paper is joint work with Clement Imbert (Warwick), Yifan Zhang (CUHK) and Yanos Zylberberg (Bristol). We are grateful to Samuel Bazzi, Loren Brandt, Holger Breinlich, Gharad Bryan, Juan Chen, Pierre-Philippe Combes, Giacomo De Giorgi, Maëlys De La Rupelle, Sylvie Démurger, Taryn Dinkelman, Christian Dustmann, Giovanni Facchini, Greg Fischer, Richard Freeman, Albrecht Glitz, Doug Gollin, Andre Groeger, Flore Gubert, Naijia Guo, Marc Gurgand, Marieke Kleemans, Michael Kremer, Jessica Leight, Florian Mayneris, David McKenzie, Alice Mesnard, Dilip Mokherjee, Joan Monras, Albert Park, Sandra Poncet, Markus Poschke, Simon Quinn, Mark Rosenzweig, Gabriella Santangelo, Michael Song, Jon Temple, Christine Valente, Thomas Vendryes, Chris Udry and Chris Woodruff for useful discussions and comments. We also thank participants in Bristol, City (University of London), CUHK, Harvard, Manchester, Oxford, Nottingham, St Andrews, St Gallen, Warwick, the Yale Agri-Devo Conference, the China Economic Summer Institute, the CEPR “Labor Markets in Developing Countries” Workshop, NEUDC, the BGSE Summer Forum and the European Economic Association meetings in Geneva for helpful comments. The usual disclaimer applies.

# 1 Introduction

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

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

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

Providing empirical evidence on the impact of rural-to-urban migration at destination is challenging because it requires us to identify large, systematic and exogenous immigrant flows into cities. Our methodology proceeds in two steps. In the first step, we isolate exogenous variation in agricultural labor returns *at origin* from the interaction of (i) an exogenous time-varying factor

(world prices for agricultural commodities) and (ii) time-invariant cropping patterns in each rural prefecture. Interacting price variations and prefecture-specific cropping patterns, we produce a measure of *residual agricultural income*. This measure of return to the traditional sector exhibits large year-to-year variation due to fluctuations in world demand and supply for agricultural products, but also wide cross-sectional differences due to the variety of cropping patterns across prefectures. Fluctuations in the residual agricultural income generate significant variations in outflows from rural areas, as measured in population censuses.<sup>2</sup> An excess value of 10% for the agricultural portfolio—about 1 standard deviation—is associated with a 1 p.p. lower outmigration incidence. In the second step, we combine the exogenous changes in outmigration due to international commodity prices with a gravity model, including travel time between origin and destination and population at destination, in order to generate exogenous fluctuations in migrant flows into urban areas.<sup>3</sup> This method yields fluctuations in immigrant inflows that satisfy three important properties: They are orthogonal to factor demand in the urban sector, they generate economically significant changes in migrant inflows, and they exhibit substantial variation across years and destinations.

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

First, we quantify the labor supply shift at destination. We find that the vast majority of migrants are hired by medium and large manufacturing firms: An increase of one percentage point in the ratio of migrants to manufacturing employment at destination is associated with a percentage increase in employment among the “above-scale” firms surveyed by the National Bureau of

---

<sup>2</sup>We use retrospective information on migration spells from the 2000 Population Census, the 2005 1% Population Survey (or “Mini-Census”) and the 2010 Population Census to reconstruct migration flows.

<sup>3</sup>Our approach is similar in that respect to [Boustan et al. \(2010\)](#).

Statistics. This shift has also implications for labor costs at destination. We find that migrant inflows exert a downward pressure on the average compensation per worker. This effect may be due to a composition effect, if firms were to replace urban workers with rural migrants who are less productive. Using the Urban Household Surveys (2002–2008), we find that the wages of “natives” decline at the same rate as firms’ labor costs, and the implied wage elasticity with respect to migration is about 0.9. Finally, wage employment among urban permanent residents slightly declines, indicating that migrants displace urban “native” workers.

Second, we look at the effect of migrant inflows on factor use and factor productivity in the average manufacturing firm. The increase in employment induces a large decrease in the capital-to-labor ratio. Capital does not adjust immediately to changes in employment, and there is little indication of such an adjustment in the following years. As for returns to factors, we find a drop in the marginal productivity of labor of the same order of magnitude as the decrease in labor wages. This outcome would be inconsistent with migrants inducing additional distortions on labor markets. By contrast, the slow or inexistent capital adjustment underlines possible frictions in access to capital for the average urban firm, and migrants do not seem to relax these constraints (at least in the short run).

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

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



suggest that migration favors manufacturing sector growth by alleviating labor market distortions.

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

The research closely relates to the nascent literature studying the impact of shifts in labor supply on the structure of firms and their relative factor intensities (Peri, 2012; Accetturo et al., 2012; Dustmann and Glitz, 2015; Kerr et al., 2015; Olney, 2013).<sup>4</sup> While many features of our empirical analysis are similar to these papers, especially Dustmann and Glitz (2015), the context we study—a developing economy with massive disparities in productivity and relative factor intensities—is very different. To the best of our knowledge, this is the first microeconomic paper to investigate and provide evidence of the effect of labor supply shocks on firm outcomes in a developing economy.

The empirical investigation sheds light on the importance of granular disparities in productivity and factor allocation across firms in developing economies in general, and in China in particular (Hsieh and Klenow, 2009; Brandt et al., 2013; Hsieh and Song, 2015). A few recent contributions have focused on the role of credit market imperfections in generating the wide dispersion in factor use and factor returns across firms (Buera et al., 2011; Midrigan and Xu, 2014; Moll, 2014). As developed theoretically in Song et al. (2011), factor market imperfections may generate large disparities in returns to capital and labor, and structural transformation should imply a constant reallocation of resources across production units. Our contribution is to provide empirical evidence of this reallocation and to show it is influenced by rural-to-urban migration, a major feature of the process of structural transformation.

A large literature has studied the process of structural transformation, and the aggregate reallocation of resources from the traditional and rural sector to the modern and urban sector.<sup>5</sup> We relate to this literature in three distinct ways.

---

<sup>4</sup>Giesing and Laurentsyeva (2015) provide evidence of firm adaptation to a different type of shifts in labor supply, i.e., the emigration of skilled workers in Eastern European countries.

<sup>5</sup>See Herrendorf et al. (2013) for a review.

First, we study China, which experienced a remarkably rapid structural transformation since the 1980s, with a sharp fall in the share of agriculture, a symmetric rise in manufacturing and services and massive migration flows from rural to urban areas. In our empirical exercise, we study migration and manufacturing growth with coherent data sources spanning a significant part of the structural transformation period, from 1996 to 2010. Much attention has been given to the patterns of Chinese growth (Song et al., 2011; Hsieh and Song, 2015). However, while the role of rural-to-urban migration in fueling growth finds a large echo in the policy debate in China,<sup>6</sup> the economic literature has so far shown less interest in the topic.

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

Third, we relate to “labor pull” models, in which structural change—and thus migration—is prompted by increased productivity in manufacturing. In China, there are non-negligible labor market frictions across space, possibly explaining the large productivity gaps between rural and urban areas. Our period of interest coincides with large changes in these productivity gaps, due to a take-off of the urban sector in presence of mobility restrictions.<sup>7</sup> Consistent with this labor pull interpretation, Facchini et al. (2015) and Cheng and Potlogea (2017) show that trade shocks increase demand for labor in the manufacturing sector and stimulate internal migration in China. We take the opposite approach to these papers and study how changes in migration can

---

<sup>6</sup>See Meng and Zhang (2010) for a survey.

<sup>7</sup>See Bryan and Morten (2015) for an analysis of the implications of labor market frictions for the spatial distribution of activity in the context of developing economies, and Au and Henderson (2006), Whalley and Zhang (2007), Vendryes (2011), Bosker et al. (2012), and Tombe and Zhu (2015) for studies of the specific Chinese context.

affect manufacturing growth.

The paper also contributes to the large literature on the effects of immigration on labor markets (Borjas et al., 1997; Card, 2001; Borjas, 2003), and more specifically to studies of internal migration.<sup>8</sup> In China, the evidence is mixed: While De Sousa and Poncet (2011) find that migration has alleviated upward pressures on Chinese wages between 1995 and 2007, Meng and Zhang (2010) find a modestly positive or zero effect of rural migrants on native urban workers' labor market outcomes, and Combes et al. (2015) find a strong positive impact on local wages. In a more structural approach, Ge and Yang (2014) use wage decomposition methods and a simple calibration to show that migration depressed unskilled wages in urban areas by at least 20% throughout the 1990s and 2000s. Our findings indicate a wage elasticity to unskilled labor comparable to those found in the literature on developing economies.

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

## 2 Predicting migration to urban areas

This section presents the data sources and construction of shocks to agricultural livelihoods. We describe how we generate exogenous rural-to-urban migrant flows based on our variations in labor returns at origin. Our strategy closely follows Boustan et al. (2010).<sup>9</sup>

---

<sup>8</sup>Boustan et al. (2010), Feng et al. (2015a), El Badaoui et al. (2014), Imbert and Papp (2016), and Kleemans and Magruder (2017), among others, study the labor market effects of migration in the United States during the Great Depression and more recently, the U.S. Corn Belt, Thailand, India, and Indonesia respectively.

<sup>9</sup>A similar approach is adopted by El Badaoui et al. (2014), Feng et al. (2015a) and Kleemans and Magruder (2017).

## 2.1 Migration flows

Migration flows are measured thanks to the 2000 and 2010 Censuses, and the 2005 1% Population Survey, also called “2005 Mini-Census,” a representative survey with a similar coverage to that of a census. Rural migrants and all types of non-locally registered individuals are interviewed. The censuses contain crucial information on migration history; depending on the wave, they may also include information on occupation, industry, income, ethnicity, education level and housing characteristics.<sup>10</sup>

Unlike most studies, which compute migration flows as a difference between population stocks, we directly observe migration spells.<sup>11</sup> First, we observe the household registration type (agricultural or non-agricultural) and places of both registration and residence, which are available down to the prefecture level. This information allows us to construct the migrant stock at destination from any possible origin. Second, migrants are asked the main reason for leaving their place of registration and when they did so.<sup>12</sup> The information on places of origin and residence can be combined with retrospective data on the year that the respondent left her place of registration to create a matrix of yearly migration flows between all Chinese prefectures between 1996 and 2010.

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

---

<sup>10</sup>A detailed description of these data sources, as well as of the institutional context of migration in China, is provided in Appendix II.A.

<sup>11</sup>Because of their quality and degree of detail, the census data collected by the National Bureau of Statistics are widely used in the literature (Combes et al., 2015; Facchini et al., 2015; Meng and Zhang, 2010; Tombe and Zhu, 2015, among others), although few authors have been able to use all of the three census data sets used in this paper.

<sup>12</sup>The year of emigration is censored above five years prior to the interview.

these possible biases and describe our adjustments in Appendix II.A.<sup>13</sup>

Census data allow us to recover yearly migration flows from each prefecture of origin to each prefecture of destination. The main explanatory variable in our analysis is immigrant inflows into each prefecture for every year. To identify its causal impact on urban outcomes, however, we need to isolate exogenous variation in these migration inflows. For this, we combine exogenous variation in migration outflows with fixed migration patterns between origins and destinations. We describe these steps in the next two sections.

## 2.2 Shocks to agricultural labor returns

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

### Potential Agricultural Output

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

First, we extract from the 2000 World Census of Agriculture the geo-coded map of harvested area for each crop (in a 5-arc-minute resolution, approximately 10km). We then overlay this map with a map of prefectures, and we construct total harvested area  $h_{co}$  for a given crop  $c$  and a given prefecture  $o$ .<sup>14</sup>

Second, we use a measure of potential yield per hectare as computed in the Global Agro-Ecological Zones (GAEZ) Agricultural Suitability and Potential Yields dataset. This is a model-based measure that uses information on crop requirements (e.g., the length of yield formation period and the stage-specific crop water requirements) and soil characteristics (i.e., the ability of the soil to retain and supply nutrients) to generate a potential yield for a given crop

<sup>13</sup>Return migration is substantial, while step migration is negligible. The results presented in Section 4 are corrected for return migration but remain robust to using non-adjusted flows.

<sup>14</sup>We collapse our analysis at the prefecture level to match migration data but agricultural shocks can be constructed at a 5 arc-minute resolution all over the country.

and a given soil under 5 scenarios: rain-fed (high/intermediate/low water input), and irrigated crop (high/intermediate water input). For each crop  $c$  and prefecture  $o$ , we use the high-input scenarios and information on the share of rain-fed or irrigated harvest in 2000 to construct potential yield  $yi_{co}$ .

The interaction between harvested area and potential yield,  $h_{co}yi_{co}$ , is our measure of potential agricultural output for each crop in each prefecture. By construction,  $h_{co}yi_{co}$  is time-invariant and supposed to capture long-term cropping patterns at origin.<sup>15</sup> Figure II.1 displays the cross-sectional variation in potential output  $h_{co}yi_{co}$  for rice and cotton, and illustrates the large geographic disparity in agricultural portfolios. We provide a more detailed description of the existing variation across prefectures and regions in Appendix II.B.

### Price fluctuations

As a measure of exogenous changes in international demand for crops, we use the World Bank Commodities Price Data (“The Pink Sheet”).<sup>16</sup> We consider prices in constant 2010 USD and per kg between 1980 and 2009 for the following commodities: banana, cassava, coffee, cotton, an index of fodder crops, groundnut, maize, millet, potato, pulses, rapeseed, rice, sorghum, soybean, sugar beet, sugar cane, sunflower, tea and wheat.<sup>17</sup> These crops account for the lion’s share of China’s agricultural production over the period of interest: They represent 90% of total agricultural output in 1998 and 80% in 2007.<sup>18</sup> We also collect producer prices, exports and production as reported by the FAO between 1991 and 2013 for China (and other countries) to check that international price variations translate into producer price variations.

To identify shocks in international prices, we use a percentage deviation from long-term trend,  $hp_{ct}$ , by applying a Hodrick-Prescott filter to the logarithm of nominal prices. Fluctuations in  $hp_{ct}$  capture short- and medium-run

<sup>15</sup>Note, however, that the measure is computed in 2000 and may be partly affected by fluctuations in prices. This measurement error may potentially bias downward our estimates of the effect of returns to agricultural labor on outmigration.

<sup>16</sup>The data are freely available online at <http://data.worldbank.org/data-catalog/commodity-price-data>.

<sup>17</sup>We exclude from our analysis one crop, tobacco, for which (i) China has a dominant position and directly influences international prices and (ii) the China National Tobacco Corporation, a state-owned enterprise, has a monopoly.

<sup>18</sup>This information can be obtained from the website of the National Bureau of Statistics of China at <http://data.stats.gov.cn/english/easyquery.htm?cn=C01>.

fluctuations in international crop prices. We provide in Appendix II.B descriptive statistics about the magnitude of fluctuations across crops. On the whole, fluctuations in prices behave as an Auto-Regressive process. The magnitude of innovation shocks is non-negligible: The market value of rice production decreased by 40% between 1998 and 2001 and increased by 70% between 2007 and 2008. In both instances, the rice-producing regions of China experienced large fluctuations in the value of agricultural production.

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

$$p_{ot} = \left( \sum_c h_{co}y_{i_{co}}\bar{p}_c hp_{ct} \right) / \left( \sum_c h_{co}y_{i_{co}}\bar{p}_c \right) \quad (\text{II.1})$$

The price shocks  $p_{ot}$  exhibit some time-varying volatility coming from world demand and supply, but there are also large cross-sectional differences. A prefecture is only exposed to variations in the prices of crops that it produces. The wide variety of harvested crops across China guarantees a large cross-sectional variance in prices  $p_{ot}$  that will be exploited in our empirical strategy. Appendix Figure II.B.3 displays the spatial dispersion in price shocks  $p_{ot}$  in 2001, just after farmers experienced a crisis across China due to a sharp decrease in the price of rice, and in 2004, after recovery.

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

### 2.3 Predicting rural-to-urban migration flows

Let  $M_{odt}$  denote migrant flows between origin  $o$  (rural areas of a prefecture  $o$ ) and destination  $d$  (a “city,” i.e., urban areas in a prefecture  $d$ ) in a given year  $t = 1996, \dots, 2010$ , which we construct using retrospective questions from the 2000 and 2010 Censuses and the 2005 Mini-Census.<sup>19</sup> We construct the outmigration rate in year  $t$ ,  $m_{ot}$ , by dividing the sum of migrants who left  $o$  (for labor reasons) in year  $t$  by the number of adults who still reside in  $o$ , which we denote with  $R_o$ . Formally, we have:

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

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

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

To estimate the causal effect of migrant inflows on urban destinations, we need variations in migration that are unrelated to potential destination outcomes. Our empirical strategy follows [Boustan et al. \(2010\)](#) and interacts two sources of exogenous variation. First, we use origin variations in returns to agriculture as exogenous determinants of migrant outflows in each rural prefecture. Second, we use a time-invariant gravity model that includes travel time between prefectures and urban population in 1990 to allocate rural migrants to urban destinations.<sup>20</sup> This two-step method yields a prediction of migrant inflows to urban areas that is exogenous to fluctuations in urban labor demand.

<sup>19</sup>There is some debate on how well urbanization is captured in Chinese data—see [Chan \(2007\)](#). However, our population of interest, i.e., rural migrant workers, overwhelmingly settles in urban areas.

<sup>20</sup>Prefecture populations in 1990 are obtained from the 1990 Population Census.



### Exogenous variations in migrant outflows

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

$$m_{ot} = \beta_0 + \beta_s \bar{s}_{ot} + \delta_t + \nu_o + \varepsilon_{o,t}, \quad (\text{II.2})$$

where  $o$  indexes the origin, and  $t$  indexes time  $t = 1996, \dots, 2010$ .  $m_{o,t}$  and  $s_{ot}$  denote the outmigration rate and agricultural residual income at origin  $o$ , respectively.  $\nu_o$  denotes origin fixed effects and captures any time-invariant characteristics of origins, e.g., barriers to mobility. We use 1990 population at origin as a weight to generate consistent outmigration predictions in the number of migrants.

As a measure of shock  $s_{ot}$ , we use the average residual agricultural income in  $t - 1$  and  $t - 2$ . A migration spell recorded at date  $t = 2005$ , for instance, corresponds to a migrant worker who moved between October 2004 and October 2005. Hence, given the timing of the growing cycle for most crops in our sample, migration spells in period  $t$  are most likely to be impacted by variations in  $t - 1$  and before—especially if there are lags in the decision to migrate.<sup>21</sup>

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

$$\widetilde{m}_{ot} = \widetilde{\beta}_0 + \widetilde{\beta}_1 s_{ot} + \widetilde{\nu}_o + \widetilde{\delta}_t$$

where  $\widetilde{\delta}_t$  is the average of the time effect.<sup>22</sup> We then multiply the migration rate by rural population at origin  $R_o$  to compute predicted migration flows from  $o$ :

$$\widetilde{M}_{ot} = \widetilde{m}_{ot} \times R_o$$

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

<sup>21</sup>Incorporating contemporary price/rainfall shocks in the analysis does not change the results. We also estimate the same specification using forward shocks, i.e., the average residual agricultural income in  $t + 1$  and  $t + 2$ , to show that shocks are not anticipated.

<sup>22</sup>We remove time variation from our predictions to avoid correlation between our migrant flows and destination trends in outcomes.

additional standard deviation in the value gap decreases migration incidence by 0.19 standard deviation. The corresponding plot in Figure II.2 shows that the relationship is globally linear. In Appendix II.D, we test whether shocks are anticipated and find that forward variations in prices do not predict migration outflows. The estimated coefficients on the lags and forwards of our constructed shocks in the joint regression are similar to those in the separate specifications.

Fluctuations in returns to agriculture may have two opposite effects on migration (Bazzi, 2016). On the one hand, if rural workers compare the return to labor in rural areas with that in urban areas, then a low return to agriculture should push them toward urban centers (*substitution* or *opportunity cost* effect). On the other hand, low returns to agriculture may affect household wealth and the ability to finance migration to urban centers (*liquidity* or *wealth* effect). In the Chinese context, where (i) migrants typically migrate without their families, and (ii) low-skill jobs in cities are easy to find, the fixed cost of migration may be relatively low.<sup>23</sup> Since wealth accumulation happens over time, it may also be less affected by short-term fluctuations in agricultural prices. The negative relationship between the value of the agricultural portfolio and migration suggests that migration decisions are driven by the opportunity cost of migrating.<sup>24</sup>

### Exogenous variations in origin-destination migration flows

We next estimate the following equation:

$$p_{od} = f(dist_{od}) + \gamma Pop_d + \mu_o + \varepsilon_{od}, \quad (\text{II.3})$$

where  $p_{od}$  is the share of migrants from prefecture  $o$  who went to prefecture  $d$ ,  $dist_{od}$  is the distance between  $o$  and  $d$ ,  $f$  is a parametric function of distance,  $\mu_o$  is an origin fixed effect, and  $Pop_d$  is the total urban population of prefecture  $d$  in 1990. Equation II.3 yields  $\widehat{p}_{od}$ , the predicted probability for migrants from

<sup>23</sup>In the period we study, although migration restrictions still exist through the *hukou* or registration system, they tend to make long-term settlement of rural migrants into urban areas difficult rather than impede rural-to-urban migration itself. See Appendix II.A for more details.

<sup>24</sup>In Appendix II.D, we show that negative rainfall shocks have independent positive effects on migration outflows, which is also consistent with the substitution effect.

prefecture  $o$  to go to prefecture  $d$  based on distance, a fixed and exogenous characteristic of the pair  $(od)$ , and the attractiveness of  $d$  captured by its lagged population. The specifications are weighted by  $Pop_d$ .

We report the estimation of equation II.3 in Panel B of Table II.1. We choose the inverse of distance for function  $f$ .<sup>25</sup> As apparent in this table, distance and population at destination are very strong predictors of the destination choice.

### Predicted migration flows

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

$$\widetilde{M}_{dt} = \sum_{o \neq d} \widetilde{M}_{ot} \times \widehat{p}_{od}, \quad (\text{II.4})$$

where  $\widetilde{M}_{dt}$  is the migrant inflow to destination  $d$  in year  $t$ ,  $\widetilde{M}_{ot}$  the predicted migrant outflow from origin  $o$  in year  $t$ , and  $p_{od}$  the predicted probability that a migrant from  $o$  goes to  $d$ . To prevent migrant inflows from being correlated with destination outcomes, we exclude from  $\widetilde{M}_{dt}$  inflows originating from rural areas of prefecture  $d$ .

This two-step procedure yields synthetic migrant inflows to prefectures of destination that are exogenous to destination outcomes. We provide some intuition about the nature of these exogenous variations in Appendix Figure II.B.4 (measure  $\widetilde{M}_{dt}$  as predicted by price variations). We report these measures cleaned of cross-sectional time-invariant factors in 2001 (left panel) and 2004 (right panel). As shown in Appendix Figure II.B.4, there is some spatial auto-correlation in these measures arising from the spatial auto-correlation of crop composition across prefectures and the transformation of outflows into inflows involving distance between prefectures. There is some auto-correlation across periods as international prices exhibit persistence in their fluctuations. Fortunately, there are also large cross-sectional and time-varying fluctuations that we can use for our analysis—see Appendix II.B for a variance decompo-

<sup>25</sup>The inverse function gives a better fit than a linear or quadratic specification (see Appendix Figure III.4).

sition of the shocks.

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

### 3 Empirical strategy

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

#### 3.1 Data

The empirical analysis mostly relies on establishment-level data spanning 1998-2007 from the National Bureau of Statistics (NBS).<sup>26</sup> The NBS implements every year a census of all state-owned manufacturing enterprises and all non-state manufacturing firms with sales exceeding RMB 5 million or about \$600,000. While small firms are not included in the census, the firms we consider account for 90% of total gross output and 70% of total employment in the manufacturing sector. The firms can be matched across years, so that we can use either the total sample of firms, which represents on average over 200,000 firms per year, or the balanced sample of 45,000 firms.<sup>27</sup> The NBS census contain information on each firm's location, industry, ownership type, exporting activity, number of employees and a wide range of accounting variables (e.g., output, input, value added, wage bill, fixed assets, financial assets,

<sup>26</sup>The following description partly borrows from [Brandt et al. \(2014\)](#), and a more detailed discussion is provided in Appendix II.C.

<sup>27</sup>The selection implied by focusing on the balanced panel is discussed in Appendix II.C—see in particular Appendix Tables II.C.1 and II.C.2.

etc.). We use these data to construct firm-level capital, employment and labor costs, which is defined as total wage bill divided by employment. We also construct measures of factor productivity, which we describe below.

There are four main challenges with using the NBS census. First, matching firms over time is difficult because of frequent changes in firm identifiers. In order to match “identifier-switchers,” we extend the fuzzy algorithm (using—among other information—name, address or phone number) developed by Brandt et al. (2014) to cover the period 1992–2009. Second, although we will use the terms “firm” and “enterprise” interchangeably in the remainder of the paper, the NBS data cover “legal units” (*faren danwei*). Subsequently, different subsidiaries of the same enterprise may be surveyed separately if they are separate legal entities and if they are financially independent. Note however that the vast majority of “legal units” in the data (97% in 2007) are single-plant entities, which brings the data closer to more familiar establishment surveys. Third, the RMB 5 million threshold that defines whether a non-publicly owned firm belongs to the NBS census was not perfectly implemented. In effect, some firms may have entered the database a few years after having reached the sales cutoff. Conversely, some private firms continue to participate in the survey even though their annual sales fall short of the threshold. However, the share of firms below the threshold is negligible, as shown in Figure II.3, and dropping them does not affect the results. Fourth, the censored nature of the data implies that our firm-level analysis cannot capture the impact of immigration on private-sector firms below the RMB 5 million cutoff. This is a potentially serious drawback as migrants are overwhelmingly employed in private firms—see Appendix Table II.A.4 and Démurger et al. (2009), *inter alios*. However, the NBS census covers most (70%) of manufacturing employment and, as can be seen from Appendix Table II.A.4, 51% of rural-to-urban migrants worked in the manufacturing sector in 2005. We would thus expect the above-scale firms surveyed by the NBS to be major migrant employers.

### 3.2 Accounting framework

In this section, we develop a simple model of firm production to guide our empirical analysis. As in Hsieh and Klenow (2009), the economy is divided into sectors in which there is monopolistic competition between heterogeneous

firms. A final good is produced from the combination of sectoral outputs, and sectoral output itself is a CES aggregate of firm-specific differentiated good (where  $\sigma \equiv \frac{1}{1-\rho}$  denotes the elasticity of substitution between differentiated final goods).

Each firm  $i$  in sector  $s$  produces according to a CES production function:

$$Y_{is} = A_{is} [a_s K_{is}^\rho + (1 - a_s) L_{is}^\rho]^{\frac{1}{\rho}}, \quad (\text{II.5})$$

where  $a_s$ , governing the sectoral capital share, is assumed to be constant within each sector and  $\rho$ , governing the elasticity of substitution between factors, is constant across all firms.<sup>28</sup>

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

$$\pi_{is} = P_{is} Y_{is} - (1 + \tau_{is}^L) w L_{is} - (1 + \tau_{is}^K) r K_{is}. \quad (\text{II.6})$$

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

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

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

The marginal revenue product of capital and labor, and the revenue pro-

---

<sup>28</sup>The Cobb-Douglas production function corresponds to  $\rho$  converging toward 0.

ductivity verify:

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

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

Some quantities governing production at the firm level are not directly observed in the data, and we must estimate or calibrate them. We will proceed as follows. As in [Hsieh and Klenow \(2009\)](#), we use U.S. data at the firm level from the 2015 Annual Survey of Manufactures to calibrate the sector-specific capital shares  $a_s$ . As for the constant elasticity of substitution between capital and labor, we use the recent estimates in [Oberfield and Raval \(2014\)](#), i.e., we set  $\sigma = 0.7$ . The implied residual  $\tau_{is}^L$ , capturing the distance to the unconstrained allocation of labor (for a certain firm  $i$ ), will be computed following II.8 and based on the quantities observable in the data.

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

An important assumption of this framework is that labor is homogeneous, which implies no productivity difference between migrant and resident workers. This assumption is driven by data limitations, since the NBS census does not break down firm employment by skill or by migrant status.<sup>29</sup> It can be relaxed as long as an efficient unit of labor, whether provided by a resident or a migrant, is equally costly to the firm. However, any discrepancy between the productivity of natives and migrants would generate a downward bias when estimating the effect of migrant inflows on returns to capital, and total factor productivity.

### 3.3 Empirical strategy

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

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

with standard errors clustered at the level of the prefecture of destination.<sup>30</sup>

We will perform a series of robustness checks to provide support to the main identification assumptions. A first cause of concern is the failure of the exclusion restriction assumption, i.e., if the agricultural commodity price

<sup>29</sup>The census data presented in Appendix II.A do in fact suggest that resident workers are more skilled than migrant workers.

<sup>30</sup>Because the regressor of interest, the migration rate, is itself predicted, correct inference requires us to bootstrap the first stage. The standard errors in the second stage are, however, correctly estimated through 2SLS.



shocks have a direct effect on firms. We shall test this in two ways: by focusing on migration between provinces (instead of prefectures),<sup>31</sup> and by excluding industries that process agricultural products. A second concern is that predicted migrant flows, which are constructed using distance and destination population, capture market access, which has an independent effect on firm growth. To test this, we estimate Equation II.9 controlling for the (log of the) destination population interacted with year, to allow destinations to follow different trends depending on their size. A third and more general concern is that our estimation may be capturing different sectoral trends, which could be correlated with migration patterns through the geographical distribution of manufacturing activities or the diffusion of agricultural price shocks. To alleviate this concern, we include industry×year fixed effects into Equation II.9, so that our estimates only capture within-industry variation in firm outcomes. Finally, we perform a standard placebo check and test whether future migration shocks have any effect on firm outcomes.

## 4 Results

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

### 4.1 The labor supply shift

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

---

<sup>31</sup>Provinces are the second administrative level in China, immediately below the central government and above prefectures.

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

As can be seen from Columns 1 and 2 in Table II.2, the inflow of rural migrants has a strong negative effect on labor costs. A one percentage point increase in the immigration rate is associated with a 0.45% decrease in wages. The IV estimates are negative and larger in magnitude: If migrants are attracted to cities that offer higher wages, OLS estimates should indeed be biased upwards. These estimates suggest that, once cleaned for demand-driven fluctuations, an influx of rural migrants depresses urban labor costs. Following Borjas (2003), we can recover the elasticity of urban wages with respect to migration by multiplying the coefficient by  $\frac{1}{(1+m)^2}$ , where  $m$  is the ratio of migrants to natives. In our context, the migration rate is about 20%, hence  $\frac{1}{(1+m)^2} \approx 0.69$ . The implied wage elasticity from our estimates is  $-0.9$ , which is markedly higher than Borjas's (2003) own estimates of  $-0.4$ . It is, however, comparable to other studies on internal migration in developing countries that use a similar strategy (Kleemans and Magruder, 2017). It may be that immigrants and native are much more substitutable in the case of internal than international migration. It may also be that the labor market for unskilled labor in urban China in the period we study was relatively unregulated, which made it easier for firms to adjust wages and employment.<sup>32</sup>

In parallel to this decrease in labor costs, firms expand employment and absorb the excess labor force. A one percentage point increase in the migration rate increases employment by 0.72% to 1.1% (Columns 3 and 4 of Table II.2). As shown in Columns 5 and 6 of Table II.2, the increase in employment translates into a strong fall in the capital-to-labor ratio, as expected if capital is fixed in the short run. However, capital-to-labor ratio decreases slightly

<sup>32</sup>Minimum wage regulations only came into force toward the end of our observation period (Mayneris et al., 2014).

more than employment increases for the average firm in our sample. Capital may be decreasing due to higher substitutability between capital and labor in the Chinese than in the benchmark U.S. economy. Firms may also sell some assets to finance the hiring of migrant workers.

To interpret these effects as a pure labor supply shift, we provide in Appendix II.D a series of robustness checks that isolate the indirect impact on urban firms through the arrival of workers from potential direct effects of agricultural shocks (through demand for non-tradable goods or the provision of intermediate goods). First, we exclude industries that process agricultural products to show that our results are not driven by the direct effect of agricultural price shocks on manufacturing units. Second, we allow firms in larger urban areas to experience different trends. Third, we control flexibly for industry-specific trends to show that we are not simply capturing urban dynamics linked to sectoral specialization or market power. Fourth, we change our definition for migrant workers restricting them to extra-provincial flows. None of these changes affect our results.<sup>33</sup>

## 4.2 Returns to factors

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

---

<sup>33</sup>The results are also robust to estimating the effect of immigration with the unbalanced panel of firms, which alleviates the sample selection issue but prevents us from including firm fixed effects in the regressions. Results are available upon request.

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

### 4.3 Dynamic effects

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

$$\left\{ \begin{array}{l} \mathbf{M}_{d\tau} = \mathbf{b}_0 + \mathbf{b}_m \widetilde{\mathbf{M}}_{d\tau} + \mathbf{c}_i + \mathbf{n}_\tau + \mathbf{e}_{id\tau} \\ y_{idt} = \beta_0 + \sum_{\tau=t-3}^t \beta_m^\tau \widehat{M}_{d\tau} + \gamma_i + \nu_t + \varepsilon_{idt} \end{array} \right. \quad (\text{II.10})$$

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

Table II.4 presents the results from the estimation. As shown in Columns 1 to 3, the negative effect of migration shocks on the labor costs faced by urban firms persists for at least two years, while the capital-to-labor ratio slowly returns to the mean. Interestingly, the negative effects on total factor productivity quickly fade away. On the whole, these results suggest that a positive migration shock depresses wages for a few years. The firm readjusts its factor mix after the first year, and the negative effects on total factor

productivity on impact are short-lived. These findings are consistent with a reorganization of firms to alleviate the capital-to-labor imbalances.

#### 4.4 Reallocation of resources across firms

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

$$\begin{cases} M_{dt} = b_0 + b_m \widetilde{M}_{dt} + c_i + n_t + e_{idt} \\ y_{idt} = \beta_0 + \beta_m \widehat{M}_{dt} + \beta_x \widehat{M}_{dt} \times X_i + \gamma_i + \nu_t + \mu_t \times X_i + \varepsilon_{idt} \end{cases} \quad (\text{II.11})$$

The results shown in Table II.5 explore the heterogeneity in firms' responses to the labor supply shock induced by migrant inflows. We first see that public-sector firms experience a slightly (not significantly) lower decrease in labor costs, but do not expand their employment when the migration rate increases. This could be due to restricted access to migrant labor for public-sector firms, which were at the time massively laying off urban resident workers (Naughton, 2007). Next, Table II.5 shows that firms with high capital-to-labor ratio experience a higher increase in employment, lower decrease in the capital-to-labor ratio and smaller decrease in the marginal revenue productivity of capital and total factor productivity. These firms may have faced a relative shortage of workers, and now expand their workforce with little negative effect on their productivity. Finally, the results of Table II.5 show that a sizeable share of the increase in employment is due to firms that export

part of their production. Taken together, these findings suggest that private, capital-intensive firms with good access to international markets are the ones that benefit from an influx of migrants.

## 4.5 Aggregate effects

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

Table II.6 presents the effects of migration on the firm size distribution. As Column 2 shows, a one percentage point increase in the migration rate increases the number of firms present in the census for the average industry-prefecture cell by 16 firms (4% of an average of 400).<sup>34</sup> In Columns 3 to 6, we consider the number of firms with revenues above RMB 10 million and 20 million. We find that migration increases the number of firms by the same proportion at all thresholds. In other terms, migration shifts the entire firm size distribution to the right. Although, in the absence of Economic Census data, we cannot ascertain that this effect holds for the whole firm size distribution, this is an indication that the impact of migration on firm size is probably felt below the RMB 5 million threshold.

We now turn to the effect of immigration on the distribution of marginal returns to factors across firms in a given industry-prefecture. Let us first review what impact we expect migration to have on the factor productivity distributions. First, the effect on the dispersion of the marginal revenue product of labor is an empirical issue as two mechanisms may be at work in opposite directions. Whereas the lower wages implied by an influx of immigrants should allow less productive firms to survive (*firm entry*), which thickens the lower tail and thus exacerbates productivity dispersion, firms that face constraints on the labor market and have a high MRPL should hire more, leading to a

---

<sup>34</sup>For this exercise, we limit our analysis to firms above the entry threshold, and drop the few firms below.

contraction of the distribution from the upper tail and a reduction in misallocation (*heterogeneity*). Second, we would expect the effect on the dispersion of MRPK to be unambiguously positive, although different mechanisms may lead to this result. If more productive firms, i.e., those with a high MRPK, hire, migration increases the productivity of capital even further. On the other hand, if an influx of cheap labor allows low-MRPK firms to survive or enter, then we should witness an increase in dispersion, but from an expansion of the lower tail of the distribution.

In Columns 1 and 2 of Table II.7, we present our estimates for the causal effect of migration on the dispersion of the marginal revenue product of labor. Panels A, B and C use alternative measures of this dispersion: the standard deviation, interquartile range and difference between the 90<sup>th</sup> and 10<sup>th</sup> percentiles, respectively. Across measures, we find that a positive labor supply shock decreases the dispersion of labor productivity: A one percentage point increase in migration decreases the 10<sup>th</sup>-90<sup>th</sup> percentile range by 0.7%. This finding is consistent with the heterogeneous effects observed in the firm-based analysis: Firms with a relatively high capital-to-labor ratio (and thus a high marginal revenue product of labor) are more likely to hire thereby reducing their marginal revenue product of labor. This implies that the *heterogeneity* mechanism clearly dominates a *firm entry* effect whereby less productive firms would enjoy a higher rate of survival or entry. By contrast, Columns 3 and 6 show that migration slightly increases the dispersion of the marginal revenue of capital and total factor productivity. These findings suggest that an inflow of migrants alleviates the effect of labor market distortions but may reinforce that of capital market imperfections.

## 4.6 Composition effects

A major shortcoming of our analysis of firm outcomes is that migration inflows may have two effects: They may change the quantity of labor supplied and the composition of the labor force in urban areas. To shed light on these issues, we use seven cross-sections of the Urban Household Survey (2002–2008), a representative survey of urban “natives”—see Appendix II.C for a description.

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

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

$$\begin{cases} M_{dt} = b_0 + b_m \widetilde{M}_{dt} + b_x X_j + c_d + n_t + e_{dt} \\ y_{jdt} = \beta_0 + \beta_m M_{dt} + \beta_x X_j + \gamma_d + \nu_t + \varepsilon_{jdt} \end{cases} \quad (\text{II.12})$$

Since unskilled urban residents are more likely to be competing for jobs with migrant workers, they may experience larger changes in wages and occupation in response to migration inflows. To test this, we estimate the same specification interacting the migration shock with a *LowSkill<sub>j</sub>* binary variable equal to one if the worker has lower secondary education or less. Table II.8 presents the results.

We first consider the impact on natives' real wages. The estimates are very similar to the estimates for labor costs using establishment-level data: A one percentage point increase in the immigration rate is associated with a 1.5% decrease in wages (Column 2 in Panel A). As expected, the wage results are stronger for urban residents with at most lower secondary education, who are more substitutable to migrant workers (Column 2 in Panel B). These results suggest that the decline in labor costs estimated using firm data did reflect a change in equilibrium wages due to a labor supply shock and not only a composition effect.

We next consider the effect of rural-to-urban migration on the occupation status of urban residents (Column 3 to 8 in Panel A). The OLS estimates are close to zero and mostly insignificant. IV estimates however show that a one percentage point increase in migration decreases wage employment for urban residents by 0.3 percentage points (the average participation to wage employment is above 90% among the labor force). Interestingly, self-employment seems to increase in the same proportion (by 0.4 percentage points). These results provide suggestive evidence that migrants displace urban residents, pushing them into self-employment.

The estimates presented in Panel B Column 4 suggest that the negative



effect on wage employment is driven by high-skill workers: The (statistically insignificant) coefficient of the interaction with the low-skill indicator is of opposite sign and of similar magnitude to the main coefficient. These findings suggest that less skilled urban residents suffer large wage cuts but do not have the possibility to become self-employed. They also provide modest support to the idea that migration changes the composition of wage workers by pushing out more skilled natives, but this displacement may only marginally affect the above-scale firms surveyed by the NBS.

## 5 Conclusion

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

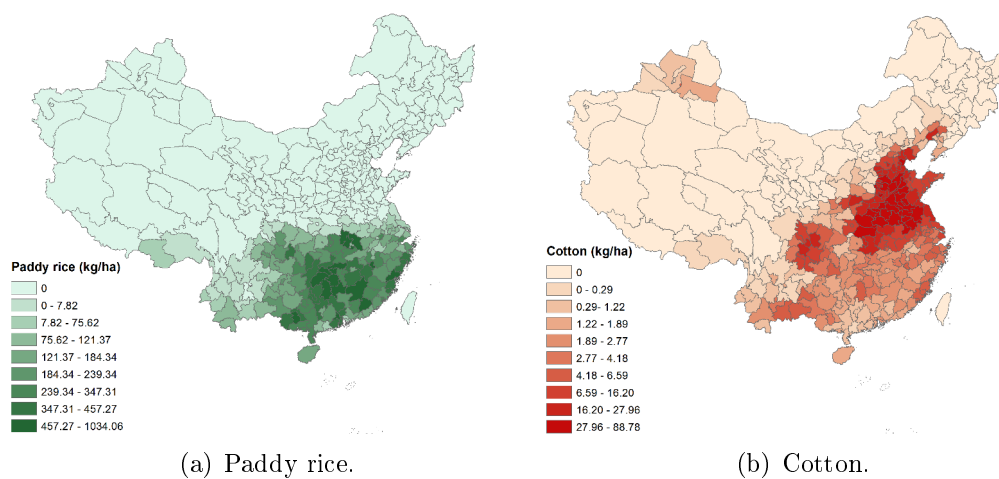
We find that the average firm experiences a large increase in employment together with a marked decrease in labor costs, which indicates that migration changes the urban labor market equilibrium. The magnitudes suggest that migrants and natives are close substitutes (the wage elasticity with respect to migration is about 0.9) and that labor demand is highly elastic (with a demand elasticity close to 0.8). These labor market effects are independently confirmed by a representative survey of urban workers. As a response to this positive labor supply shock, firms increase employment so that the marginal product of labor decreases proportionally to the wage change. As capital does not adjust, capital and total factor productivity decreases. These effects are temporary and suggest costly factor adjustment in the short run.

The analysis of the heterogeneous impact of migration confirms the importance of factor market distortions. The increase in employment is con-

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

## Figures and tables

Figure II.1 – Potential output in China for rice and cotton (2000).



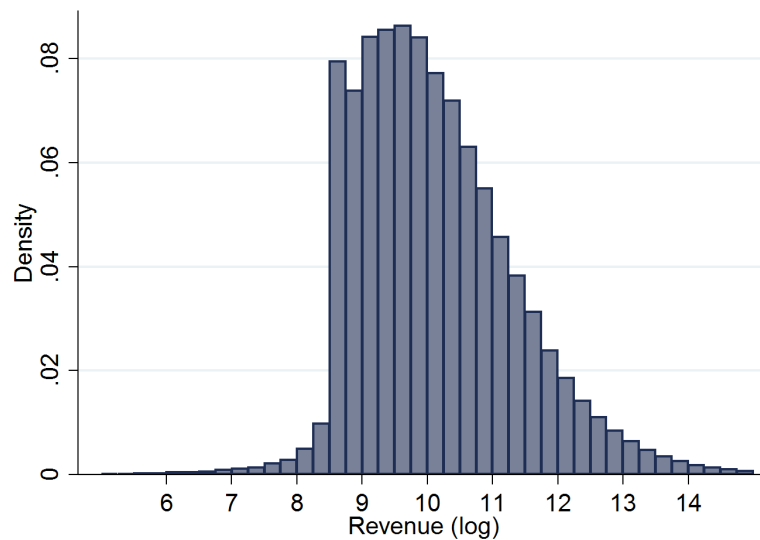
Notes: These two maps represent the potential output constructed with harvested areas in 2000 and potential yield (GAEZ model) for 2 common crops in China, paddy rice (left panel) and cotton (right panel).

Figure II.2 – Value of agricultural portfolio at origin and outmigration rates.



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

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



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

Table II.1 – Migration Predictions

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

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

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

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

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

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

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

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

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

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

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

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

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

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



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

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

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

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

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

Notes: Standard errors are clustered at the prefecture level and are reported between parentheses. \*\*\*:  $p < 0.01$ , \*\*:  $p < 0.05$ , \*:  $p < 0.1$ . The aggregation is performed using all firms present at any time in the NBS firm census between 1998 and 2006. The unit of observation is an industry  $\times$  prefecture  $\times$  year. The outcomes are different measures of the dispersion of MRPL, MRPK and TFPR across firms. In Panel A, the outcomes are standard deviations; in Panel B, they are differences between the log of the 75<sup>th</sup> and 25<sup>th</sup> percentiles; and in Panel C, they are differences between the log of the 90<sup>th</sup> and 10<sup>th</sup> percentile. MRPL, MRPK and TFPR are defined in the Section 3. All specifications include industry\*prefecture and year fixed effects. In the IV estimation, the instrument is the migration rate predicted using price shocks at origin, distance between origin and destination, and destination population. All estimates are weighted by the number of firms in the industry\*prefecture in 1998.

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

	Real Monthly Wage		Employee		Unemployed		Self-employed	
	OLS (1)	2SLS (2)	OLS (3)	2SLS (4)	OLS (5)	2SLS (6)	OLS (7)	2SLS (8)
<i>Panel A: Main</i>								
Migration Rate	-0.234* (0.122)	-1.479* (0.800)	-0.007 (0.038)	-0.351** (0.177)	-0.028** (0.014)	0.041 (0.065)	0.036 (0.034)	0.309** (0.146)
Observations	244,020	244,020	271,600	271,600	271,600	271,600	271,600	271,600
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>Panel B: Unskilled</i>								
Migration Rate	-0.181 (0.117)	-1.344 (0.826)	-0.017 (0.036)	-0.405** (0.177)	-0.034** (0.013)	0.054 (0.066)	0.051 (0.033)	0.350** (0.145)
MR × Unskilled	-0.260** (0.118)	-1.191*** (0.458)	0.042 (0.051)	0.391 (0.249)	0.024 (0.018)	-0.030 (0.053)	-0.067 (0.049)	-0.361 (0.238)
Observations	244,020	244,020	271,600	271,600	271,600	271,600	271,600	271,600
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors are clustered at the prefecture level and are reported between parentheses. \*\*\*: p<0.01, \*\*: p<0.05, \*: p<0.1. All specifications include year and destination fixed effects. Wages are deflated by a consumer price index based on consumption expenditure data available in the Urban Household Surveys. Unskilled workers are defined as workers with no education, primary education or lower secondary education. *Employee* is a dummy for receiving a wage, while *Self-employed* is a dummy equal to one for individuals that are self-employed workers or employers.

## ONLINE APPENDIX

### II.A Migration flows: construction and description

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

#### II.A.1 Elements of context

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

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

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

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

## II.A.2 Data sources and construction of migration flows

### Data description

To measure migration flows, we use the 2000 Population Census, the 2005 1% Population Survey, also called “2005 mini-census,” and the 2010 Population Census.

After the beginning of the reforms and loosening of restrictions on mobility, there was a growing disconnect between census data focusing on *hukou* location and the rising “floating population” (*liudong renkou*) of non-locally registered citizens. The 2000 Population Census was the first census to acknowledge this gap and record migrants’ places of residence—provided they had been living there for more than 6 months (Ebenstein and Zhao, 2015). In addition to the place of residence (at the prefecture level in our data), *hukou dengji di* or place of registration (prefecture level)<sup>35</sup> and *hukou* type, the 2000 and 2010 Population Censuses contain retrospective information on the place of residence five years before the survey (at the province level), the year of arrival at destination (if within the past 5 years) and the reason for departure.

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

A few caveats are in order. First, the sampling frame contained only information on population by registration. High-immigration areas could thus be under-sampled. Comparing the flows for 2000 according to the 2000 Census and 1% Survey, we indeed find a small discrepancy that we attribute to coverage issues. Second, the 2005 1% Population Survey offers a set of variables similar to standard censuses, but some differences are worth bearing in mind: (i) Both data sources provide prefecture-level information on the place of residence, but it is defined as “current residence” in 2005 and thus also captures

<sup>35</sup>The censuses record both the places of residence and registration at the county level but our version of the data only contains residence codes down to the prefecture level.

migrants who have been established at destination for less than 6 months. (ii) The 2000 and 2010 Censuses contain prefecture-level information on the place of residence prior to departure, while the 1% Survey records *hukou* location at the prefecture level. These two places are one and the same if there is no step migration, i.e., if rural dwellers move directly to their final destination. Along the same lines, the 1% Survey records the timing of *departure* from a migrant's place of registration rather than of *arrival* at destination, so that figures need not exactly coincide—unless, again, there is no step migration. Third, although we do not need this assumption for the empirics as our analyses are carried out at the prefecture level, it is important to bear in mind that the data do not allow us to determine whether a migrant is residing in a rural or urban area. The census and mini-census data do not record the place of residence at high enough resolution to unambiguously infer whether the destination is urban or rural. Nevertheless, it is clear from the literature that rural-to-rural migration represents a small share of outmigration from rural areas, not least because most of it is explained by marriages, which usually give right to local registration (Fan, 2008; Chan, 2012b).<sup>36</sup> Fourth, we cannot account for migrants who changed their *hukou* location or type. This assumption is quite innocuous given that *hukou* conversion is marginal.

### Migration flow construction

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

We construct annual migration flows between each prefecture of origin and destination by combining information on the current place of residence (the destination), the place of registration (the origin) and the year in which the migrant left the origin. One advantage of working with Censuses is that they cover—or are representative of, in the case of the Mini-Census—the whole population: All individuals, irrespectively of their *hukou* status, are inter-

---

<sup>36</sup>Only 4.7% of agricultural *hukou* inter-prefectural migrants in the 2005 Mini-Census reported having left their place of registration to live with their spouses after marriage.

viewed in 2000, 2005 and 2010. However, not all migration spells are observed in the Census data. We describe below (i) which migration spells are directly observed and which migration spells are omitted, and (ii) how we can infer some of these unobserved spells and adjust the raw migration flows.

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

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

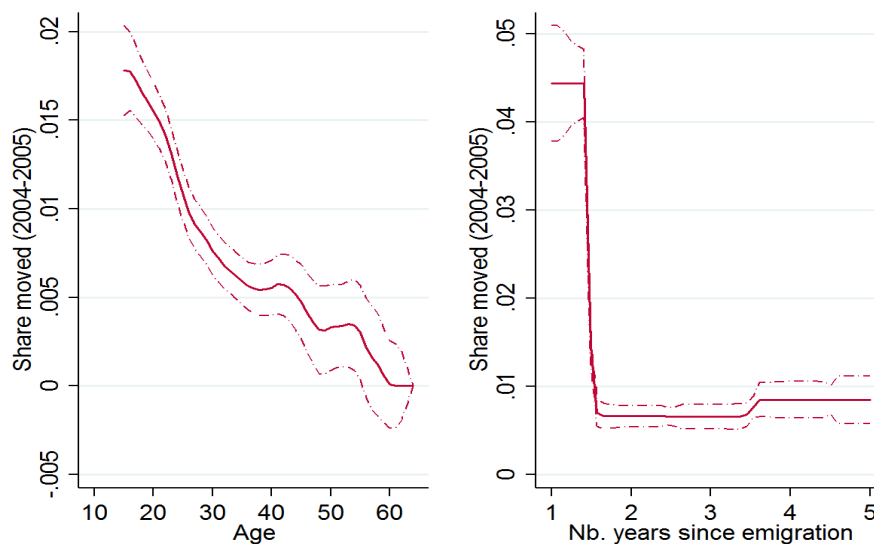
A fraction of the *step migration* and *return migration* spells can, however, be observed in the restricted sample of individuals interviewed in the 2005 1% Population Survey and the 2000 and 2010 Censuses. Indeed, the 2005 1% Population Survey includes the locations in which the individuals were living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

We first study the importance of step migration. Among all migrants



who lived in their province of registration in 2000 and were living in another province in 2005, we compute the fraction that lived in yet another province in 2004. As Appendix Figure II.A.1 shows, only a minority of migrants have changed provinces of destination in the last year. Step migration is not only small, but concentrated in the very first year after departure from the place of origin. In other words, step migration induces errors in arrival and departure dates that are also quite small. As adjusting for step migration would require strong assumptions about the intermediate destination, which is not observed in the data, we do not correct the raw flows for step migration.

Figure II.A.1 – Share of step migrants as a function of age and time since departure.

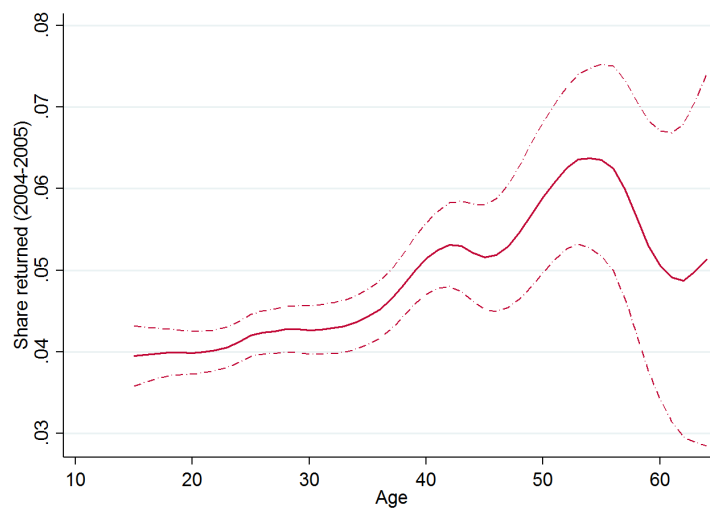


Sources: 2005 Mini-Census.

We then consider the extent of return migration. Among all migrants from rural areas who lived in their province of registration in 2000 and who lived in another province in 2004, we compute the fraction that had returned to their province of registration by 2005. As Appendix Figure II.A.2 shows, this share is not negligible: In a given year, between 4 and 6% of rural migrants who have left their province of registration in the last six years go back to their *hukou* location. Return migration is hence an important phenomenon,

which leads us to underestimate true migration flows and the effect of shocks on outmigration. Because of the retrospective nature of the data, past flows, for instance in 2000 for an individual interviewed in 2005, are mechanically underestimated. In contrast to step migration, however, it is possible—under reasonable assumptions—to adjust migration flows and account for return migration. We provide below a description of these adjustments.

Figure II.A.2 – Share of return migrants by age.



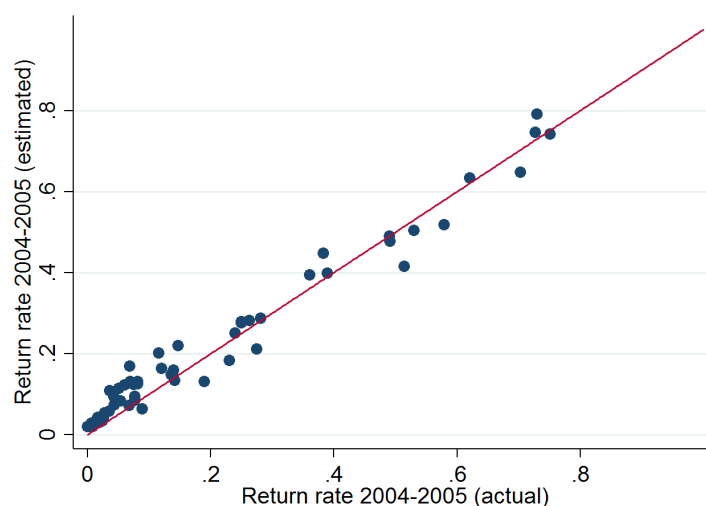
Sources: 2005 Mini-Census.

Adjusting for return migration requires us to observe the destination and duration-specific yearly rate of return. Indeed, there is a wide disparity across destinations in return rates, and, as in any survival analysis with censoring, there are large compositional adjustments along the duration of the migration spell. Specifically, the probability for a migrant to return home sharply decreases with the length of the migration spell, mostly reflecting heterogeneity across migrants in their propensity to return. To capture these differences across destinations and length of the migration spell, we make the following assumptions. (i) Each migrant is characterized by a constant Poisson rate  $f$  of returning. (ii) We suppose that there is a fixed and destination-specific distribution of migrant types  $H(\cdot)$  upon arrival. We also allow the distributions to differ across *hukou* type, i.e., rural or urban. (iii) To fit the observed return rates as a function of migration duration, we further assume

that  $h(f) = \lambda_p^2 f e^{-\lambda_p f}$  where  $\lambda_p$ , the province-specific exponential parameter, will be calibrated using actual return rates within five years, for individuals at destination five years before the survey.

Under these assumptions, the evolution of the pool of migrants with duration can easily be computed. In the cross-section (i.e., with all cohorts and not only newly-arrived migrants), the distribution of migrant types is  $h_c(f) = \lambda_p e^{-\lambda_p f}$ , such that the average return rate in the cross-section is  $1/\lambda_p$ . The targeted moment, i.e., the return rate over 5 years, is  $1/(\lambda_p + 5)$ . We calibrate the *hukou*- and province-specific exponential parameter  $\lambda_p$  to match the actual return rate, and we perform this calibration *in each survey* so as to flexibly allow for long-term fluctuations in these province-specific distributions. The correction term that we apply to migration spells is more complicated to compute and depends on the return rate over  $k$  years for newly-arrived migrants,  $1 - \lambda_p^2/(\lambda_p + k)^2$ . We carry out this exercise for the 2000 and 2010 Censuses and 2005 Survey data.<sup>37</sup>

Figure II.A.3 – Over-identification test for the return migration correction.



Sources: 2005 Mini-Census.

One concern with this methodology is that we may not precisely capture

<sup>37</sup>We also adjust for coverage issues in the 1% Population Survey due to its sampling design to obtain consistent figures across data sets. We calculate the prefecture-level and *hukou*-specific adjustment term that matches the flows (corrected for return migration) in 2000, the only year for which the Census and survey overlap.

the duration-dependence in return rates, and thus over- or underestimate return rates for individuals arriving immediately before the interview. Using the 2005 1% Population Survey, we can provide an over-identification test by computing the return probability between 2004 and 2005 for recently-arrived migrants (i.e., between 2000 and 2004) and comparing it with the empirical moment. As shown in Figure II.A.3, the prediction—computed with the cross-sectional return rate over 5 years—matches quite well the actual annual return rate for recently-arrived migrants.

### II.A.3 Description

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

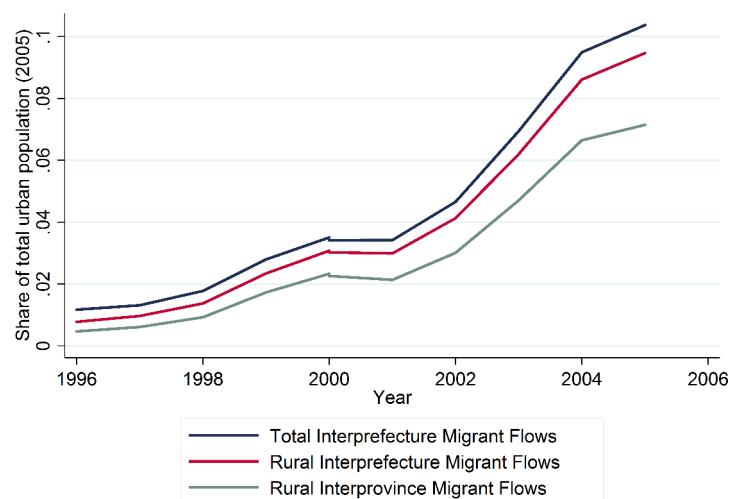
#### Migration patterns over time and across regions

Figure II.A.4 illustrates the rise in inter-prefectural migrant flows between 1996 and 2005 as a share of the population registered in urban areas in 2005. The rising trend and magnitude of migration flows is striking: In 2005, the annual inflow of migrants from other prefectures was around 10% of the destination population, as against less than 2% in 1996. Two interesting facts pertain to the composition of the incoming migrants. First, about 80% of the yearly migrant inflows consist of agricultural *hukou* holders (“rural” migrants), the remainder being accounted for by urban dwellers originating from other prefectures. Second, about 80% of inter-prefectural rural-to-urban migrations recorded over the period 1996-2005 involved the crossing of a provincial border.

As far as variation in migration outflows across space is concerned, Table II.A.1 compares outmigration (top panel) and immigration (middle panel) rates in 2000 and 2005 across China’s six broad regions. The table also distinguishes between the types of administrative border crossed by migrants: inter-prefectural, inter-provincial and inter-regional.<sup>38</sup> Two obvious patterns emerge from Table II.A.1: There is significant variation in terms of both emigration and immigration rates across regions but no region is left aside from

<sup>38</sup>China is divided into 6 broad regions.

Figure II.A.4 – Evolution of migration rates between 1996 and 2005.



Sources: 2000 Census and 2005 Mini-Census.

the migration phenomenon.

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

Table II.A.2 displays the shares of rural-to-urban migrants, defined as agricultural *hukou* holders who crossed a prefecture boundary and belong to working-age cohorts (15-64), in the total urban population of prefectures.

The upper panel of Table II.A.2 distinguishes between inter-prefectural migrants and those who left their provinces of origin. We see that inter-prefectural migrants represented 19% of a prefecture's total number of urban residents on average in 2005, while inter-provincial migrants accounted for 15% of it, which confirms that a majority (77%) of inter-prefectural migrations

Table II.A.1 – Descriptive statistics of migration flows by region.

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

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

imply the crossing of a provincial boundary.

The high share of inter-provincial migrations among inter-prefectural moves already sheds light on the distance traveled by internal migrants in China. Figure II.A.5 and Table III.4 offer additional evidence on the impact of distance on migrants' destination choices. We see that there is a strong and significant inverse relationship between the share of migrants from origin  $o$  to destination  $d$  (among all migrants from  $o$ ) and distance between  $o$  and  $d$ .

Table II.A.2 – Descriptive statistics from the 2005 mini-census.

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

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

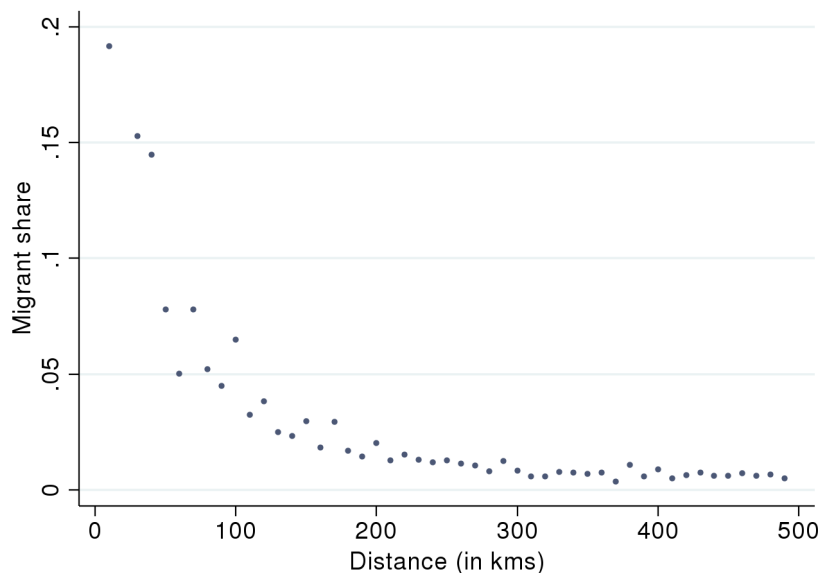
### Selection of migrants

We now provide some descriptive statistics on the profiles of internal migrants in China—in terms of education, demographics and labor market situation. To understand the effects of our shocks on outmigration and the impact of rural-to-urban migrants on the urban labor market and firms, it is useful to know the motives behind migration spells and describe the profile of rural migrant workers relative to both non-migrant rural dwellers and “native” urbanites.

The bottom panel of Table II.A.2 presents the reasons put forward by inter-prefectural agricultural *hukou* migrants for leaving their places of registration. A vast majority (82%) moved away in order to seek work (“Work or business”), mostly as laborers, while all other rationales attract much smaller shares.<sup>39</sup>

<sup>39</sup>The only other reasons that display shares in excess of 1% are “Education and training,” “Other,” “Live with/Seek refuge from relatives or friends,” which Fan (2008) based on metadata from the Population Census Office dubs “Migration to seek the support of relatives or friends,” “Following relatives,” which should be understood as “Family members following the job transfer of cadres and workers”, and “Marriage.”

Figure II.A.5 – Origin-destination migration predictions—the role of distance.



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

Table II.A.3 – Distance and migration flows between origins and destinations (2000-2005).

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

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



Migrants are not a representative sample of the Chinese population with an agricultural *hukou*. As can be seen from Table II.A.4, migrants tend to be younger, more educated and more often single than the non-migrant rural population. They are also more likely to be self-employed or employees and to work in the private sector. Their total monthly income is more than twice as large. As expected, a majority (78%) of rural dwellers work in agriculture, as against 5% among migrants.

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

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

Migrant selection into destinations based on economic prospects poses a serious threat to the identification of the effect of labor inflows on urban areas. Our empirical strategy however relies on exogenous variation in agricultural prices at migrants’ places of origin, which drive emigration decisions, thus allowing us to estimate the causal effect of immigration on urban labor markets

---

<sup>40</sup>Results available upon request.

Table II.A.4 – Migrant selection (2005 mini-census).

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

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

and firms.

## II.B Shocks to rural livelihoods

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

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

### II.B.1 Crop suitability and use across Chinese prefectures

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

The top and middle panels of Table II.B.1 show the variation in potential yields and harvested areas by crop and region. We focus on the four most important crops—rice, wheat, maize and soy—and on the high-input scenarios, which are better tailored to China and therefore the ones we use to compute the weights. It is obvious from the table that, as expected, some crops are concentrated in particular regions. This is especially true of rice, which is absent from the colder and drier northern regions. Nonetheless, it is also apparent that there is substantial regional variation in terms of crops and that no crop is cultivated in a single region, nor a region characterized by a single crop. A large part of the cross-sectional variation that we exploit does not come from regional differences but from more local and granular disparities

across prefectures.

Table II.B.1 – Descriptive statistics of price shocks, potential yields and harvested areas by region.

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

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

An illustration of these regional differences is also provided in Figure II.1 of the paper.

## II.B.2 International price variations and domestic prices

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

One important assumption behind our empirical strategy is that China's

agricultural prices are not insulated from world market fluctuations. Table II.B.2 confirms that international price variations do translate into price fluctuations in the Chinese domestic market. The first column provides the correlation between Chinese domestic prices for different crops in different years and international prices. We find that a 10% increase in the latter yields a 4% hike in the former, which constitutes a substantial pass-through from the international to the domestic market. The second column looks at the logarithm of output as the dependent variable and explains it by international and domestic prices. We can see that both prices are positively associated to crop production over the period of interest. While output and local prices are both determined by local demand and supply, international prices better explain the variation in local output than local prices. One explanation could be that local demand and local supply have opposite effects on the comovement of output and prices, while international price shocks should be perceived as a pure demand shock from the viewpoint of Chinese producers.

Table II.B.2 – Correlation between international crop prices and local Chinese crop prices/production.

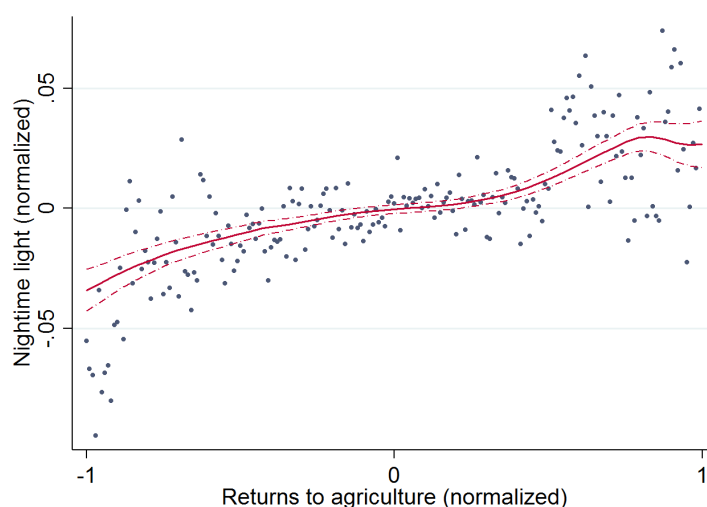
VARIABLES	Prices	Output
Price (International)	.402*** (.0861)	.201** (0.062)
Price (China)		.0824* (.0432)
Observations	210	210
R-squared	.579	.337
Trend	Yes	Yes
Crop FEs	Yes	Yes

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

Figure II.B.1 further shows that the effect of international price shocks can be felt in origin prefectures across China. We regress night-time luminosity on residual returns to agriculture in the *same* prefecture. We find that a higher value of the local agricultural portfolio has a strong positive effect on local population or economic activity, which night-time lights proxy for.

Given the large year-to-year variation in the price shocks—see Figures II.B.2 and II.B.3 below,—the good fit visible on the graph, with no obvious group of outliers, suggests that no prefecture, however poor or remote, is impervious to international price shocks.<sup>41</sup>

Figure II.B.1 – Effect of price deviations from trends on night-time luminosity.



Notes: This Figure illustrates the relationship between the standardized value of the prefecture-specific agricultural portfolio as predicted by international prices (x-axis) and night-time luminosity as measured by satellite (y-axis). We consider the residuals of all measures once cleaned by prefecture and year fixed effects. For the sake of exposition, we group prefecture $\times$ year observations, create 100 bins of observations with similar price shock and represent the average night-time luminosity within a bin. The lines are locally weighted regressions on all observations.

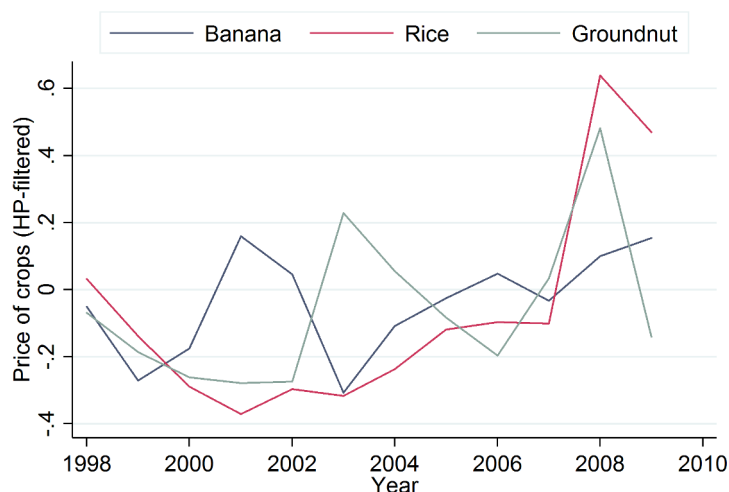
Another important assumption is that there are relevant short-term fluctuations in international crop prices. Figure II.B.2 plots the evolution of international prices for a selection of crops.<sup>42</sup> We interpret these short-term fluctuations as random shocks on the international market due to fluctuations in world supply and demand for each crop. As can be seen from Figure II.B.2, the HP-filtered price series resemble AR(1) processes with jumps. There are large swings followed by a gradual return to the mean. Importantly, all crops

<sup>41</sup>Note that households within an origin prefecture may react differently to international price fluctuations. This does not jeopardize identification as we are interested in the effect of immigration on urban areas: As long as agricultural price shocks trigger large outmigration flows while having no direct effect on urban firms, this strategy is a legitimate one.

<sup>42</sup>As in the main empirical specification, we have filtered out the long-term trend component following the procedure designed by Hodrick and Prescott (1997).

display such large fluctuations over time, and the fluctuations need not coincide across crops.

Figure II.B.2 – Price deviations from trends on International Commodity Markets 1998-2010.



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

### II.B.3 Shocks over time and across regions

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

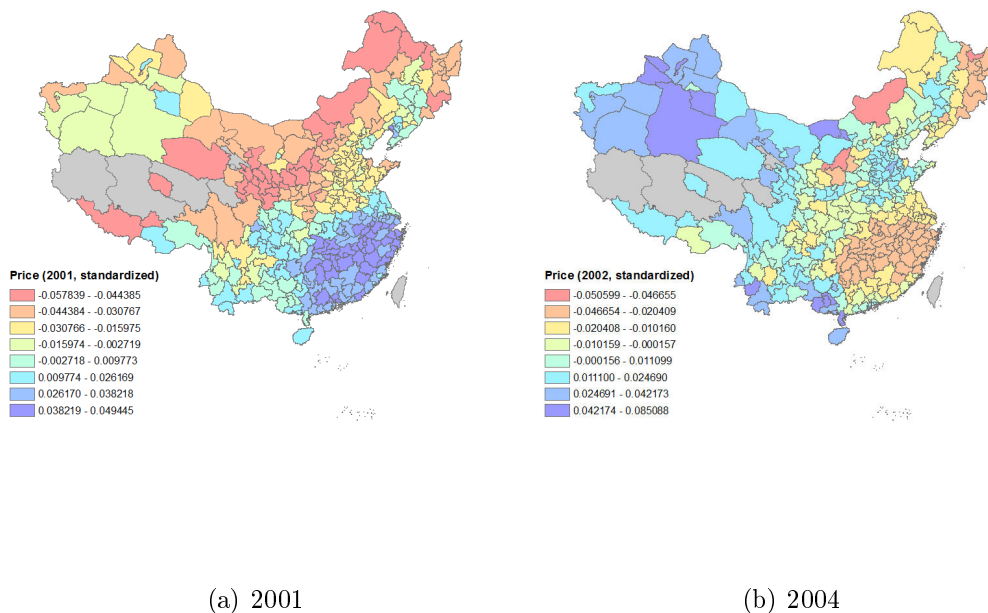
The bottom panel of Table II.B.1 provides between- and within-region variation in the price shock for China's six major regions. Between variation is measured in 2000. Reassuringly for our identification strategy, we can see that all regions display fluctuations in the price shocks, both across prefectures and over time. No region stands out as being particularly subject to such shocks or immune to them either.

Evidence from Table II.B.1 can be illustrated by some maps. Figure II.B.3 displays the price shocks in 2001 (left panel) and 2004 (right panel). Despite some spatial correlation due to the underlying cropping patterns, we see that



there is substantial variation across prefectures (delimited by dark lines) and that the picture also changes noticeably from one year to the other.

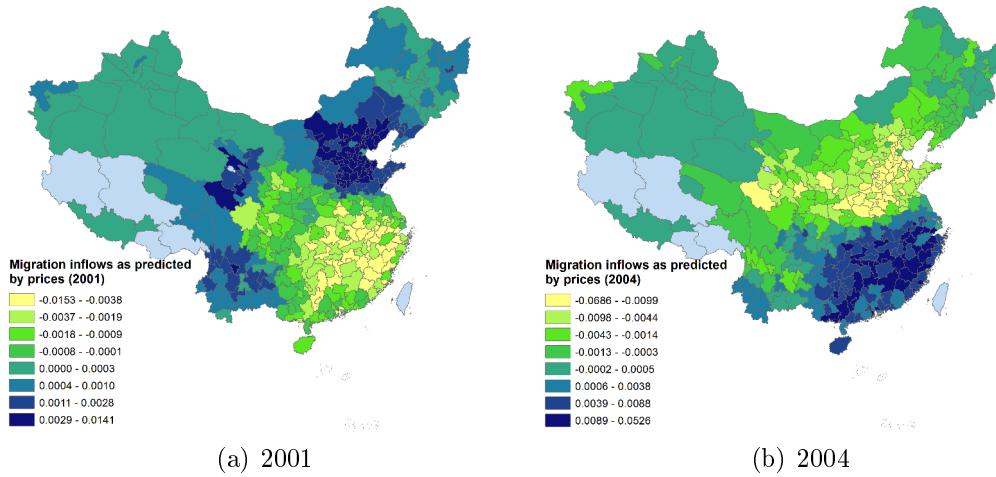
Figure II.B.3 – Shocks to rural livelihoods across Chinese prefectures in 2001 and 2004.



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

These cross-sectional and time variations carry over from the price shocks to the predicted flows of immigrants. Figure II.B.4 represents immigration rates at the prefecture level in 2001 (left panel) and 2004 (right panel), as predicted by agricultural price shocks in migrants' prefectures of origin. Here again, we see that predicted immigration varies widely both across prefectures and over time.

Figure II.B.4 – Measure  $\widehat{M}_{dt}$  of immigrant inflows to cities as predicted by prices in 2001 and 2004.



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

## II.B.4 An additional source of variation: rainfall shocks

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

The monthly precipitation measure (0.5 degree latitude  $\times$  0.5 degree longitude precision) covers the period 1901-2011 and mostly relies on the Global Historical Climatology Network.<sup>43</sup> Once collapsed at the prefecture level, this provides us with a measure  $ra_{omt}$  of rainfall for prefecture  $o$  in month  $m$  and year  $t$ .

We refine this rainfall measure to account for the growing cycle of each crop, i.e., (i) the harvest season and (ii) the crop-specific rainfall requirements. For a given year, there are several sources of variation across Chinese prefectures in actual yields due to rainfall. First, different locations receive different levels of rainfall. Second, exposure to rainfall depends on the growing cycle of the different harvested crops (winter, spring or summer/fall crops). In addition, some crops are resistant to large water deficits, while others immediately

<sup>43</sup>UDEL\_AirT\_Precip data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, was obtained from their website at <http://www.esrl.noaa.gov/psd/>.

perish with low rainfall. The large cross-sectional variation in each year may come from (i) a direct effect of local rainfall, (ii) an indirect effect coming from the interaction with the crop-specific growing cycle and the variety of crops grown across China.

We rely on the measure  $ra_{omt}$  of rainfall for prefecture  $o$  in month  $m$  and year  $t$  and construct for each crop a measure  $wr_c$  of the minimum crop-specific water requirement during the growing season  $M_c$  as predicted by the yield response to water.<sup>44</sup> We then generate

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

This measure has a very intuitive interpretation. The quantity  $\max\{\sum_{m \in M_c} wr_c - ra_{omt}, 0\}$  is the deficit between actual rainfall and the minimum crop water requirement  $wr_c$  during the growing season. We then penalize this deficit with a factor  $\alpha$  capturing potential non-linearities in the impact of rainfall deficit. In our baseline specification, this penalization parameter  $\alpha$  is set equal to 3.<sup>45</sup> A high ratio  $\max\{\sum_{m \in M_c} wr_c - ra_{omt}, 0\}/wr_c$  would be associated with a bad harvest for the corresponding crop. We then weight these ratios by potential output for each crop in each prefecture.

There is large year-to-year variation in rainfall deficits. Also, for a given year, because of differences in cropping patterns across prefectures, the spatial auto-correlation of rainfall shocks is much lower than the correlation of rainfall itself. While the exogeneity of rainfall shocks is not questionable, we still need to assume that urban labor demand is not directly affected by rainfall in order to use it as an instrument for rural-to-urban migration.<sup>46</sup>

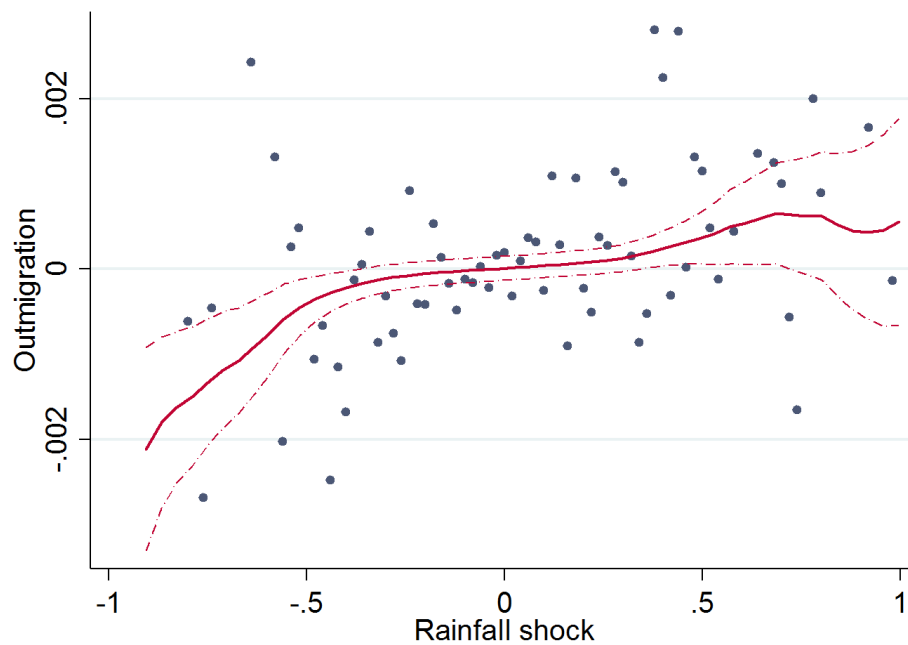
The relationship between rainfall shocks and rural emigration is illustrated in Figure II.B.5. Remember that a high value of the rainfall shock is associated with severe water deficit during the growing seasons of locally grown crops. We see from Figure II.B.5 that low rainfall in a given year pushes rural dwellers out of their prefectures of origin.

<sup>44</sup>Information on crop-specific water requirements can be found here: <http://www.fao.org/nr/water/cropinfo.html>.

<sup>45</sup>The results are robust to more conservative values for  $\alpha$ , e.g.,  $\alpha = 1$  or  $\alpha = 2$ .

<sup>46</sup>Our results hold when controlling for local rainfall shocks. Results available upon request.

Figure II.B.5 – Rainfall deficits relative to water requirements at origin and outmigration rates.



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

## II.C Data description

### II.C.1 Firm data

We present here in greater detail our firm survey data. We first summarize the main characteristics of the data, along with some descriptive statistics, and then discuss some possible issues and how we tackle them.<sup>47</sup>

#### Description

The firm data that we rely on for the better part of the empirics come from the National Bureau of Statistics (NBS). The NBS implements every year a census of all state-owned manufacturing enterprises and all non-state manufacturing firms with sales exceeding RMB 5 million, or about \$600,000 over that period.<sup>48</sup> This threshold gives the data their common name of “above-scale” manufacturing firm surveys (“*xian’e*” or “*guimo yishang*” *gongye qiye diaocha*), despite the fact that the data constitute a census of SOEs irrespective of their size.

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

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

---

<sup>47</sup>Please refer to Brandt et al. (2014) for an exhaustive treatment. This section partly summarizes the challenges that they highlight.

<sup>48</sup>The average exchange rate over the period of interest was RMB 8.26 to the USD, so RMB 5 million represents about \$605,000. Note that the threshold was raised to RMB 20 million in 2011.

data corresponded to single plants.

We restrict ourselves to the balanced panel of firms over the period in most of our analysis, and we only study entrants and exiters separately. In contrast with firm data in some developed countries, matching firms over time in the NBS is difficult because of frequent changes in firm identifiers. To match “identifier-switchers,” we use the fuzzy algorithm developed by [Brandt et al. \(2014\)](#), which uses slowly-changing firm characteristics such as name, address or phone number. While total sample size ranges between 150,000 and 500,000 per year, we end up with 45,000 firms when we limit the sample to the balanced panel.

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

Tables [II.C.1](#) and [II.C.2](#) compare firms in the unbalanced and balanced panels along key characteristics in 1998 and 2006, respectively. Given the large sample size, most t-tests yield statistically significant differences. The unbalanced and balanced samples however look very similar in terms of industrial sectors. The sectors that account for the largest shares of the sample are Equipment, Mineral, Chemical, Food, and Textile.<sup>49</sup> As expected, firms in the balanced panel are larger and perform generally better than the average firm in the sample, advantages that they retain until the end of our period. The balanced sample also contains a higher proportion of private-sector firms in 1998.<sup>50</sup> It is however less affected by the rising share of private (mostly domestic) firms, so that the pattern is reversed by 2006.<sup>51</sup> These two major

<sup>49</sup>The footnote to Table [II.C.1](#) details the contents of each category.

<sup>50</sup>Ownership type is defined based on official registration (*qiye dengji zhuce leixing*). Out of 23 exhaustive categories, Tables [II.C.1](#) and [II.C.2](#) use three categories: (i) state-owned or hybrid, (ii) domestic private and (iii) foreign private firms, including those from Hong Kong, Macau and Taiwan. Collectively owned firms are included in the private sector.

<sup>51</sup>In terms of time patterns, private firms still accounted for less than half of total real capital stock, value added, sales revenues, compensation and employment in 1998 but rep-

differences reflect (i) the reforms imposed on state-owned enterprises in the 1990s, which involved privatization in some cases and led many of them to fall below the RMB 5 million threshold and thus exit the data set, and (ii) the “Grasp the Big, Let Go of the Small” (*zhua da, fang xiao*) policy that concentrated state ownership on the largest and most promising (or strategic) SOEs.

### Possible issues

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

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

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

Third, although each “legal unit” in each year contains a unique identifier, changes were introduced over time, so that identifiers need not be consistent from one year to the next. Matching firms over time in the NBS data is

---

resented over 80% of the total under all five indicators by 2007. The evolution in terms of employment is particularly striking: Whereas only 32% of total employment could be attributed to private firms in the NBS firm sample in 1998, they accounted for 89% of it in 2007.

Table II.C.1 – Comparison of the unbalanced and balanced panels of firms at baseline (1998).

	Unbalanced	Balanced	Difference	p-value
<b>A: Size and production</b>				
Total number of employees	247.89	450.84	-202.95*	0.000
Real capital stock	18,142	47,073	-28,931*	0.000
Real sales revenue	24,503	65,920	-41,416*	0.000
Real value added	6,686	18,417	-11,732*	0.000
Real wages and benefits	1,969	4,840	-2,871*	0.000
<b>B: Ownership</b>				
Public sector	0.68	0.48	0.19*	0.000
Private sector	0.32	0.51	-0.19*	0.000
Domestic private	0.27	0.39	-0.13*	0.000
Foreign private	0.05	0.12	-0.07*	0.000
<b>C: Industrial sector</b>				
Chemical	0.16	0.17	-0.02*	0.000
Equipment	0.22	0.28	-0.05*	0.000
Food	0.16	0.10	0.06*	0.000
Handicraft	0.04	0.05	-0.01*	0.000
Mineral	0.19	0.18	0.01*	0.001
Textile	0.14	0.14	0.00	0.646
Tobacco	0.00	0.00	0.00	0.823
Wood	0.09	0.08	0.02*	0.000
Number of “legal units”	100,666	44,899		

Notes: \*:  $p < 0.01$ . This table displays t-tests. All variables in Panel A are in RMB 1,000, except Total number of employees. Benefits include housing subsidies, pensions, medical insurance and welfare payments. Sales revenue, value added, and wages and benefits are deflated by a province-level consumer price index. Real capital stock is created as described in Appendix II.C. Panel C displays the proportion of firms in each 1-digit industrial sector. As far as the main categories are concerned, Chemical covers the petroleum processing, coking and nuclear fuel processing industry, chemical materials and chemical products manufacturing, pharmaceutical manufacturing, chemical fiber manufacturing, rubber products industry and plastic products industry; Equipment includes general equipment manufacturing, special equipment manufacturing, transportation equipment manufacturing, electrical machinery and equipment manufacturing industry, and communications equipment, computers and other electronic equipment manufacturing; Food refers to the agricultural food processing industry, food manufacturing and beverage manufacturing; Mineral corresponds to the non-metallic mineral products industry, ferrous metal smelting and rolling processing industry, non-ferrous metal smelting and rolling processing industry, and metal products industry; and Textile consists of the textile industry, garment, footwear and headgear manufacturing, and leather, fur, feathers (down) and related products industry.

therefore a challenge. To match “identifier-switchers,” we extend the fuzzy algorithm (using firm identifier, firm name, name of the legal representative, address, phone number, founding year and main products) developed by [Brandt et al. \(2014\)](#). While total sample size ranges between 150,000 and 500,000 per year, we end up with 45,000 firms when we limit the sample to the fully balanced panel between 1998 and 2006.



Table II.C.2 – Comparison of the unbalanced and balanced panels of firms at endline (2006).

	Unbalanced	Balanced	Difference	p-value
<b>A: Size and production</b>				
Total number of employees	208.80	475.64	-266.84*	0.000
Real capital stock	22,668	67,100	-44,432*	0.000
Real sales revenue	79,946	207,975	-128,030*	0.000
Real value added	21,282	57,135	-35,852*	0.000
Real wages and benefits	4,300	11,760	-7,460*	0.000
<b>B: Ownership</b>				
Public sector	0.06	0.20	-0.14*	0.000
Private sector	0.94	0.80	0.14*	0.000
Domestic private	0.83	0.67	0.16*	0.000
Foreign private	0.11	0.13	-0.02*	0.000
<b>C: Industrial sector</b>				
Chemical	0.17	0.18	-0.01*	0.000
Equipment	0.27	0.27	-0.01*	0.002
Food	0.09	0.10	-0.00*	0.001
Handicraft	0.05	0.05	-0.00	0.957
Mineral	0.17	0.18	-0.01*	0.000
Textile	0.17	0.14	0.02*	0.000
Tobacco	0.00	0.00	-0.00*	0.000
Wood	0.08	0.08	0.01*	0.000
Number of “legal units”	439,511	44,899		

Notes: \*:  $p < 0.01$ . This table displays t-tests. All variables in Panel A are in RMB 1,000, except Total number of employees. Benefits include housing subsidies, pensions, medical insurance and welfare payments. Sales revenue, value added, and wages and benefits are deflated by a province-level consumer price index. Real capital stock is created as described in Appendix II.C. Panel C displays the proportion of firms in each 1-digit industrial sector. For the contents of the main industrial categories, see the footnote to Table II.C.1.

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

only since 2003 and 2004, respectively but accounted for only 3.5% of total worker compensation in 2007.

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

## II.C.2 UHS data

### Description

In order to study labor market outcomes from the workers’ point of view, we use the national Urban Household Survey (UHS) implemented by the National Bureau of Statistics. The UHS is a survey of urban China, with a consistent questionnaire since 1986 but considered representative from 2002 onward, and our description will correspond to this latter period. The survey is based on a three-stage stratified random sampling. Its design is similar to that of the Current Population Survey in the United States ([Ge and Yang, 2014](#); [Feng et al., 2015b](#)) and includes 18 provinces and 207 prefectures.<sup>52</sup> The data we use

<sup>52</sup>The provinces are Beijing, Shanxi, Liaoning, Heilongjiang, Shanghai, Jiangsu, Zhejiang, Anhui, Jiangxi, Shandong, Henan, Hubei, Guangdong, Chongqing, Sichuan, Yunnan, Shaanxi and Gansu.

for our analysis are annual cross-sections, with a sample size that ranges from 68,376 in 2002 to 94,428 individuals in 2008. Our analysis will be restricted to the locally registered urban population.<sup>53</sup>

The UHS is a very rich dataset with detailed information on individual employment, income—including monthly wages, bonuses, allowances, housing and medical subsidies, overtime and other income from the work unit—and household-level characteristics—see [Feng et al. \(2015b\)](#) for a comprehensive description of the survey. Our measure of real wages relies on monthly wages divided by a prefecture- and year-specific consumer price index, which we computed using the detailed household-level consumption data available in the UHS. We also construct three employment outcomes: wage employment, unemployment and self-employment (which also includes firm owners).<sup>54</sup> Appendix Table II.C.3 provides some descriptive statistics of key variables over the period 2002-2008.

### Empirical strategy

Let  $y_{jdt}$  be the labor market outcome of worker  $j$  in urban destination  $d$  in year  $t$ . We regress  $y_{jdt}$  on predicted migration that year,  $\widetilde{M}_{dt}$ , and a vector of individual characteristics  $\mathbf{X}_j$ . The vector  $\mathbf{X}_j$  includes dummy variables for individual  $j$ 's marital status, gender, education level (primary, lower secondary, upper secondary and tertiary) and age. We also include year/occupation dummies to control for workers' skills and sector-specific fluctuations in labor costs.<sup>55</sup> To control for labor market conditions at destination and aggregate fluctuations in labor market outcomes, we also include destination fixed effects and year fixed effects. The effect of  $M_{dt}$  on  $y_{jdt}$  is estimated through

<sup>53</sup>While all households living in urban areas are eligible, sampling still ignores urban dwellers living in townships and in the suburban districts of Beijing, Chongqing, Shanghai, and Tianjin ([Park, 2008](#)). Rural-to-urban migrants, who are more likely to live in peripheral areas of cities, are therefore under-represented.

<sup>54</sup>Working hours in the month preceding the survey were also recorded in UHS 2002-2006. However, as pointed out by [Ge and Yang \(2014\)](#), they vary within a very narrow range, which means that the UHS measure might understate actual variations in working hours. For this reason, we do not use hours of work in our analysis.

<sup>55</sup>UHS occupation categories are “Head of organization,” “Professional skill worker,” “Staff,” “Commercial and service worker,” “Agriculture,” “Production operator,” “Soldier” and “Other occupations.” Since occupation itself may be an outcome of immigration, we check that our results are robust to excluding it from the vector of controls.

Two-Stage Least Squares (2SLS), using  $\widetilde{M}_{dt}$  as an instrument:

$$\begin{cases} M_{dt} = b_0 + b_m \widetilde{M}_{dt} + b_x \mathbf{X}_j + c_d + n_t + e_{dt} \\ y_{jdt} = \beta_0 + \beta_m M_{dt} + \beta_x \mathbf{X}_j + \gamma_d + \nu_t + \varepsilon_{jdt} \end{cases}, \quad (\text{II.14})$$

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

Table II.C.3 – Descriptive statistics from the UHS data (2002-2008).

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

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

## II.D Robustness checks and sensitivity analysis

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

### II.D.1 Shocks to rural livelihoods

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

The results presented in Table II.D.1 show that rainfall shocks are also strong predictors of rural outmigration. As expected, a more severe rainfall *deficit* reduces farmers' expected output and leads to more outmigration. This effect is consistent with that of price shocks and thus reinforces our interpretation of fluctuations in international agricultural commodity prices as shocks to peasants' livelihoods. Table II.D.1 further shows that prices and rainfall constitute two independent sources of variation in rural outmigration, as can be seen from their independent effects in Column 3.

Table II.D.1 – Comparison of actual and predicted immigration rates in urban areas (Census, 2000-2005).

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

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

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

Table II.D.2 – Predicting Outmigration – Forward Shocks

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

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

## II.D.2 Main specification

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

Tables II.D.3 and II.D.4 control for channels through which price shocks might have a direct effect on firms and thus imperil the causal interpretation of the effects reported in Tables II.2 and II.3, respectively. Panel A reproduces the baseline results of Tables II.2 and II.3. Panels B to E present the results

of four robustness checks.

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

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

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

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



Table II.D.3 – Impact of migration inflows on urban firms – robustness checks (1/2).

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

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

Table II.D.4 – Impact of migration inflows on urban firms – robustness checks (2/2).

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

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

# Place-Based Policies in the Long Run: Evidence from Million-Rouble Plants in China

---

## Abstract<sup>1</sup>

This paper studies the short- and long-term impacts of Million-Rouble plants built thanks to U.S.S.R. economic and scientific cooperation on the spatial distribution of economic activity in China. We exploit exogenous variation in location decisions due to the threat of active U.S. airbases and find that between the 1950s and the end of the command-economy era, the “156” program led to an expansion of urban areas and GDP per capita, and accelerated structural transformation in treated counties. This significant advantage however eroded in the subsequent period and the later stage of structural transformation—from manufacturing to services—failed to materialize. We hypothesize and find support for the idea that the “156” factories distorted local labor markets and stifled entrepreneurship, which prevented treated areas from benefiting from the economic reforms.

**Keywords:** Local Economic Development, Foreign Investment, China, Place-Based Policies.

**JEL classification:** R11, R53, J24, N95.

## 1 Introduction

Many countries in an early development stage try to leapfrog by setting up large industrialization plans. A first rationale behind this is that primary

---

<sup>1</sup>This paper is joint work with Stephan Heblich (Bristol, CESifo, IfW Kiel, IZA, SERC), Hao Xu (PKU) and Yanos Zylberberg (Bristol). We are grateful to Sylvie Démurger, James Fenske, Richard Freeman, Flore Gubert, Sylvie Lambert, Simon Quinn, Jon Temple and Liam Wren-Lewis for useful discussions and comments. We also thank participants at the LSE, in Oxford and in Bristol for helpful comments. The usual disclaimer applies.

sector development is hampered by large fixed costs, which calls for state intervention to accelerate the process of structural transformation. Another is that those place-based policies offer the opportunity of reorganizing the spatial distribution of production to optimize externalities.

The rapid industrialization of China in the past decades overshadows previous attempts at hastening the pace of structural transformation and economic development. The Great Leap Forward (1958–1962) or the Third Front Movement launched in 1964 involved a large reallocation of resources from agriculture to industry (Lin, 1990; Tsui, 1991; Li and Yang, 2005). They are however put forward as disastrous or highly inefficient industrial experiences. Contrary to such policies, a major central planning program had already been developed through a series of deals struck by the U.S.S.R. and Chinese governments in the 1950s to boost China’s industrialization. The program, which spanned a wide range of locations, consisted of about 150 industrial projects that benefited from Soviet technology transfers, aid and loans.

After four decades of warfare and political turmoil, Chinese leaders studied the possibility of international economic cooperation to foster the development of heavy industry and transform China’s agrarian economy. The regime’s revolutionary agenda, American support for the Nationalist government in Taiwan in the aftermath of the civil war, the Western embargo (Zhang, 2001) and then the Korean War, in which China directly participated by sending troops, reinforced links between China and the Soviet Union—which Chairman Mao called “leaning to one side” (*yi bian dao*) in a famous speech.<sup>2</sup> The U.S.S.R. agreed to cooperate and assist China in the creation of state-of-the-art industrial sites.<sup>3</sup> A rapid ideological divergence however precipitated a Sino-Soviet split (1960–1989) that ended all industrial cooperation between the two countries. By 1960, 156 projects had nevertheless benefited from the latest technological development and expertise of Soviet engineers. Of these 156 projects, only 150 were fully operational at the time of the schism and were immediately regarded as a cornerstone of early Chinese industrialization. The iconic “156” program (“156” *xiang gongcheng*) is still well known in China

---

<sup>2</sup>This speech, entitled “On the People’s Democratic Dictatorship,” was delivered on June 30, 1949.

<sup>3</sup>In contrast to the U.S., economic aid to “satellite” countries was not a pillar of the Soviet Union’s policy to maintain influence, neither in Southeast Asia, nor in Eastern Europe, Africa or South America. This program may thus be considered an exception.

today.

This paper exploits the “156” factories to study the short- and long-term impacts of an early industrialization program on the spatial distribution of economic activity in China between 1950 and 2015. The analysis goes through the following stages:

(i) We first collect data on the early development of Chinese industrialization between the 1950s and 1980-90s and document the evolution of the places that hosted a “Million-Rouble” plant under central planning. In contrast to many regional development programs, these place-based policies were direct constructions of industrial sites rather than large-scale infrastructure investments as in the well-known example of the Tennessee Valley (Kline and Moretti, 2014). While the Million-Rouble plants were located in areas that were not initially developed, they enjoyed a natural advantage for industrialization (because of access to natural resources and connectedness). They were responsible for most of the early boom in industrial activity before the 1980s and constituted the pillars of their local economies.

(ii) We then analyze the role of the Sino-Soviet factories during the massive structural transformation of the Chinese economy and show a reversal effect: The neighborhoods of these factories benefited less from the general economic take-off than similar counties, even though the factories themselves were still productive.

(iii) Finally, we delineate a conceptual framework to understand why treated areas, despite a significant head start at the onset of the reforms, were hurt by the presence of a Million-Rouble plant. The framework relies on distortions in factor markets leading to high entry costs for entrepreneurs to explain the demise of the counties that hosted the regime’s flagship industries. Empirical evidence supports the hypothesized mechanism, as average wages are higher in areas surrounding Million-Rouble plants, and results are consistent with the predictions of the framework: Industrial employment and output are concentrated among a smaller number of firms, and the tertiarization of the economy is weaker in treated places, which can be construed as signs of lower firm entry.

In the ideal setting, the decision to locate actual projects would have

a natural set of counterfactual industry locations, e.g., a list of candidate locations as in [Kline and Moretti \(2014\)](#) or [Greenstone et al. \(2010\)](#), and a well-identified exogenous component in the selection process among these locations. We emulate this setup by proceeding in two steps.

In a first step, we rely on criteria described in [Bo \(1991\)](#) to define a set of suitable locations  $C = \{c_1, \dots, c_N\}$ , collapsed at the county level.<sup>4</sup> These criteria, described in Section 2, included first and foremost access to natural resources through the existing transportation network. We thus define a county as treated,  $(T_c)_{c \in C} = 1$ , if it actually hosted a factory and consider all other counties *within the set of suitable locations* as the control group ( $T_c = 0$ ).<sup>5</sup>

The decision-making process thus described however leaves us with many possible counties, including those hosting actual factories, and we cannot rule out that unobserved heterogeneity influenced the choice of locations among suitable places. In a second step, we therefore leverage another dimension of the decision process, also made explicit by [Bo \(1991\)](#), to isolate exogenous variation in the selection of actual hosting counties. As the planning period coincided with the immediate aftermath of the Korean War, in which China intervened directly, it was too risky to build state-of-the-art factories within reach of U.S. or Taiwanese bombers. To incorporate this crucial component of the decision process,<sup>6</sup> we construct a measure of vulnerability to air strikes from major rear U.S. Air Force (U.S.A.F.) and Taiwanese bases, accounting for the presence of allied bases acting as a shield. More precisely, we compute, for each county  $c$  and major rear enemy base the length of the minimum-distance route for a bomber avoiding the strike zones of U.S.S.R. and North Korean bases. We then create for each county a vulnerability measure  $(V_c)_{c \in C}$  that is the minimum of those penalized distances across all enemy bases. We then implement a two-stage strategy instrumenting the treatment  $T_c$  by the vulnerability measure  $V_c$  among the set of suitable locations and controlling for an extensive set of variables that might both have entered location decisions

---

<sup>4</sup>Although more than 156 projects had been agreed on by the time the Sino-Soviet split was completed ([Dong and Wu, 2004](#)), planned projects had not yet been awarded a definitive location, which prevents us from using abandoned projects as a control group.

<sup>5</sup>We shall relax this definition in the empirics by considering as treated counties that did not host a factory but lie within a certain radius of a “156”-program plant.

<sup>6</sup>Senior generals were directly involved in the planning decisions.

and determine our outcomes of interest, to identify the causal effect of the “156” projects on local economies.

We construct measures of economic activity between 1950 and 2015 from a variety of sources. The first data source that is readily available in the early stages of industrialization is census data, covering the period 1953–2010. A second data source is the National Bureau of Statistics “above-scale” annual firm survey (1992–2009), which constitutes a census of firms with annual sales in excess of RMB 5 million and of all state-owned firms regardless of size. Our data further include information gathered from the China City Statistical Yearbooks (*Zhongguo Chengshi Tongji Nianjian*), which cover the period 1994–2013, and the geo-coded information needed to create our instrument, control variables and some of our outcomes of interest (e.g., night-time lights as a proxy for economic activity or population).

The “156” program is a unique experiment to study agglomeration economies and place-based policies in the context of a developing agrarian economy. First, central planning in the context of Asia’s “hot” Cold War introduced exogenous variation in the decision to locate the projects. Second, due to the expertise provided by the Soviet Union and the state of China’s economy after the Sino-Japanese and civil wars, the “156” factories—close to the technological frontier—constituted a massive push shock in an otherwise agrarian economy. Third, these investments preceded the massive structural transformation by about 30 years. The “156” program thus presents the features of a large counterfactual experiment off the equilibrium path. Fourth, the “156” consist of many different types of factories, providing in-sample variation in the type of treatment as well as placebo experiments.

This paper makes important contributions to different strands of the literature.

First, it contributes to the literature on place-based policies by studying the impact of an exceptional but ill-timed investment in a growing economy. The effect of place-based policies has recently attracted a lot of attention. They are indeed popular instruments for fostering local economic development but importantly have an ambiguous impact on welfare as diverting resources to one place comes at the expense of the others (Glaeser and Gottlieb, 2008). Contrary to the papers reviewed by Neumark and Simpson (2015), we are

interested in the effect of a place-based policy in a developing country and in the long run. Another point of departure with the existing literature is that the “156” program consisted of turnkey factories that immediately generated jobs and income, which brings this paper closer to the literature on the effects of industrial concentration (Chinitz, 1961; Greenstone et al., 2010; Falck et al., 2013; Glaeser et al., 2015), to which we also contribute by investigating in the long run the impact of big industrial plants under two very different economic regimes—the planned and open economies. The closest paper to ours is Kline and Moretti (2014). They use the boundaries of the Tennessee Valley Authority (TVA) and other proposed projects to study the short- and long-run effects of the TVA, in particular on the transition from agriculture to manufacturing. Similarly, this paper uses exogenous variation in the siting of the “156” factories to causally estimate short- and long-run effects. The Chinese setting further allows us to look at two different transitions: from agriculture to (heavy) manufacturing in the command-economy era and from heavy to light, export-oriented industries and eventually to services in the subsequent reform period.

Second, place-based policies are often justified by threshold effects in economic development, an argument put forward by proponents of big-push industrial policies. This paper sheds light on the *local effects* of a major big-push development plan on structural transformation. Taken as a whole, the “156” program bears more than a passing resemblance to Rosenstein-Rodan’s “Russian model” of industrialization. Rosenstein-Rodan (1943) called for coordinated and large investments across industries to modernize agrarian economies based on the key insight, later formalized by Murphy et al. (1989), of a divergence between private and social returns to investment. Hirschman (1958) opposed Rosenstein-Rodan’s “balanced growth” strategy as too demanding of developing countries and argued that investments in one particular sector, or “unbalanced growth,” would generate an imbalance that would require complementary investments, thus creating further imbalances through which the economy would grow. The “156” program, which built one or a handful of large plants per treated county, is akin to a Hirschman-style industrialization plan at the local level and offers a close test of big-push industrialization theories—subject to the caveat that Hirschman’s theory, as Rosenstein-Rodan’s, per-



tains to macroeconomic policies. To the best of our knowledge, this paper is the first to provide such a test.

Third, the paper provides evidence on path dependence in the location of economic activity across cities and the long-term effects of historical shocks. Contrary to [Bleakley and Lin \(2012\)](#) for instance, evidence of persistence is however incidental in this paper. We indeed strive to show the effect of the “156” plants over time, with particular emphasis on the later reform period. Our use of data from 1953 is thus primarily meant to make explicit the link between the initial shock (the factories that benefited from Soviet cooperation in the 1950s) and the economic advance of treated over control counties when reforms were introduced. We also find evidence of a reversal of fortune. Our findings indeed suggest that the “156” factories distorted local factor markets, crowding out smaller firms and stifling entrepreneurship during the reform era. This result resonates with the sluggish labor market and reduced entry that [Glaeser et al. \(2015\)](#) find in localities where extractive and heavy industries developed because of access to natural resources (coal mines and iron ore deposits).

Fourth, it relates to the literature on agglomeration and spillover effects. The closest paper to ours in this respect is [Greenstone et al. \(2010\)](#), which studies the agglomeration effects of million-dollar manufacturing plants in the United States. Their identification strategy relies on the comparison of actual treated localities and their first runners-up. We develop an alternative strategy based on exogenous variation in how manufacturing plants were assigned to different counties. Another point of departure with [Greenstone et al. \(2010\)](#) is that we focus on a developing agrarian economy and contrast the impact of the “156” program before and after China opened up to international trade and adopted more market-friendly economic policies.

Finally, this paper also sheds light on China’s economic transition, and we hypothesize that frictions in labor markets may explain the effects of the Million-Rouble plants that we find in the data. This paper thus relates to the literatures on transition economies ([Blanchard, 1997](#); [Buera and Shin, 2013](#)), as well as on structural transformation ([Song et al., 2011](#); [Tombe and Zhu, 2015](#)), industrial structure ([Brandt et al., 2016](#); [Hsieh and Klenow, 2009](#); [Hsieh and Song, 2015](#)) and labor market distortions in China ([Brandt et al.,](#)

2013; Mayneris et al., 2016). This paper nicely complements Brandt et al. (2016). They establish among other stylized facts the wide dispersion in output per worker within the non-state sector across localities in China, which they impute to “entry wedges” highly correlated with the share of state-owned enterprises (SOEs) in the local economy. By offering causal evidence of the long-term impact of large, initially state-owned Million-Rouble plants on local growth during the economic take-off and in particular on their adverse effect on firm creation, this paper sheds light on the “entry wedges” estimated by Brandt et al. (2016).

The remainder of the paper is organized as follows. Section 2 describes the historical context of the “156” program. We detail our main data sources and empirical strategy in Section 3. Section 4 presents our empirical findings and conceptual framework. Section 5 concludes.

## 2 Context

Chinese industrial policy in the second half of the 20<sup>th</sup> century was heavily influenced by geopolitics. In this section, we will first see that the “156” program was launched during the Cold War thanks to Soviet aid and that this historical context affected location choices through the threat of enemy attacks. Second, we discuss the abrupt end of Sino-Soviet cooperation, to which the “156” factories nevertheless survived. Third, we describe the major change in economic regime introduced by the reforms and opening-up policy of the 1980s and 1990s and what difference it made for the Million-Rouble plants.

### Sino-Soviet cooperation

In 1949, after decades of destruction through the Sino-Japanese and Chinese civil wars, the Chinese government engaged in economic cooperation with the Soviet Union to industrialize the economy (Lüthi, 2010) and give China its own independent industrial system (Dong, 1999). Even before the Chinese Communist Party eventually won the civil war, Anastas Mikoyan, one of the most prominent Soviet statesmen and future vice-head of government, visited

China on behalf of the U.S.S.R. government and discussed possible economic assistance from the Soviet Union. The purpose for the U.S.S.R. was to extend its influence in the region. Several months later, Liu Shaoqi, future vice-chairman of the People's Republic of China, visited the U.S.S.R.—soon followed by the first foreign visit of Mao Zedong. The possibility of economic cooperation became credible after the Sino-Soviet Treaty of Friendship and Alliance of 1950, which included a large loan and centered on the construction of the “156” factories. In August 1952, Chinese Prime Minister Zhou Enlai visited Moscow to formalize the involvement of the Soviet Union in China's long-delayed First Five-Year Plan (1953–1957). The U.S.S.R. would assist China in the construction of about 50 industrial sites, followed by 91 new projects agreed on in May 1953 and an additional 15 in October 1954.

The economic aid from the U.S.S.R. extended beyond large loans. First, during the peak of the cooperation, between 1953 and 1956, 20,000 scientific, industrial and technical experts from the Soviet Union lived and worked in China to design the construction of factories and rationalize production (Zhang, 2001; Wang, 2003). In order to build capabilities, 80,000 Chinese students were also trained in Soviet universities and technological institutes. Second, the U.S.S.R. provided more than half of the required equipment.<sup>7</sup> Third, while some blueprints were destroyed, the existing technology could be imitated and represented a large shift in the technological frontier for such an agrarian economy as China (Bo, 1991).<sup>8</sup> Chinese scholars (Dong and Wu, 2004; Zhang, 2009; Shi, 2013, *inter alios*) thus credit the “156” program with having (i) invested in basic sectors such as the energy and steel industries and laid the foundations for the development of other industries, (ii) boosted China's production capacity and shifted its technological frontier, and (iii) promoted a more even development by industrializing central and western provinces. The construction of the large “156” factories is also believed to have triggered the rapid development of Chinese cities in the following decades (He

---

<sup>7</sup>As a payment, China was to give 140,000 tons of tungsten concentrate, 110,000 tons of tin, 35,000 tons of molybdenum concentrate, 30,000 tons of antimony, 90,000 tons of rubber and other produce including wool, rice or tea. Some low-skilled workers were also sent to Siberia.

<sup>8</sup>The 15 projects agreed on in 1954, which belong to the “156” program, even benefited from state-of-the-art equipment that few Soviet factories enjoyed (Goncharenko, 2002), allowing China to make the most of Gerschenkron's (1962) “advantage of backwardness” (Tang, 2009).

and Zhou, 2007).

### Location decisions

It is important to bear in mind that the “156” factories were regarded as iconic firms and planners put much thought in siting decisions. They made sure that locations made sense from an economic point of view—the criteria set out to determine locations at the time shall help us identify the relevant counterfactual for treated places, as explained in Section 3. These suitability criteria, detailed for instance in Bo (1991), are: (i) access to natural resources (through existing roads and rail), (ii) connection to the transportation network and (iii) belonging to an agrarian province, as the investments were seen as an opportunity to smooth the spatial distribution of income while developing the country as a whole, a common goal of place-based policies. But as this period was also an era of heightened geopolitical tensions in East Asia that culminated in the Korean War, where U.S. soldiers and Chinese “volunteers” directly confronted, planners were concerned that the brand-new plants might become the target of enemy attacks and involved senior military officials to decide where the factories should be built.

### Sino-Soviet Split

After Stalin’s death, ideological and political tensions started to rise with Khrushchev’s policies of “de-Stalinization” and of “peaceful coexistence with the West.” As China kept encouraging a Stalin-like cult of Mao’s personality, experimented with disastrous grass-roots industrialization during the Great Leap Forward—the polar opposite of orthodox Soviet central planning,—and pursued an aggressive foreign policy with the Second Taiwan Strait Crisis, the normalization of the Soviet regime and prospect of *détente* between the two superpowers could only worry Chinese leaders. The Sino-Soviet split materialized in 1960 with the sudden withdrawal of experts and engineers from China, the repatriation of Chinese students from the U.S.S.R. and the cancellation of ongoing projects. The only remnants of the short-lived Sino-Soviet alliance are the “156” projects. Six factories were not viable and closed. We shall here make use of the 150 plants that had been already completed and were operational by 1960.

### The “156” under central planning

For the first 20 years of their existence, the Million-Rouble plants developed in a planned economy. From the 1950s through the 1980s, the development of the “156-program” factories and of their local economies was fueled by the provisions of the plan: Factor movement was not free, and if more workers or capital could be productively employed, the plan should have allocated more resources. In that sense, we could in theory shed light on agglomeration economies in the purely Communist period. However, we are (at least currently) not well equipped to do that for two reasons: (i) We lack the necessary data on productivity, and (ii) we cannot exclude that the central planner had other objectives in mind than economic growth, i.e., what resembled a Hirschman-style, “unbalanced growth” policy may have gradually evolved into a “balanced growth” big push *à la* Rosenstein-Rodan as complementary investments accompanied the initial treatment. This is why we prefer to view the command-economy era as a whole as the treatment: There was a shock in the late 1950s, which was the creation of the Million-Rouble plants, and the empirical part of this paper shall show that treated counties enjoyed a significant head start at the onset of the second period. The question is then to determine whether that head start helped or hurt treated places.

### The “156” during structural transformation

Starting in the 1980s and accelerating in the 1990s, China introduced reforms to allow the economy to first grow out of the plan and then without a plan. Concretely, private firms could be set up and a dual price system allowed market transactions alongside the old quota requirements. In the 1990s, restrictions on labor mobility were gradually loosened, and migration began to rise as a major feature of Chinese economic growth.

The introduction of economic reforms meant that the “156” factories had to adapt to the market economy. Nine firms went bankrupt but most modernized and still constitute major firms today. The first firm forced into bankruptcy (in the late 1970s) was a coal mine, because of resource depletion. Since then, eight other factories have been closed, all coal or non-ferrous metal mines.<sup>9</sup>

---

<sup>9</sup>Two other firms, a paper mill and a former military electronics plant, were officially closed for bankruptcy but were actually restructured and continued to operate.

Some plants were also relocated nearby. When construction plans were made in the 1950s, most plants were built in the city center. As pollution issues and the need for expansion had not been anticipated, nine plants were moved to the suburbs. In a nutshell, despite the significant transformation that the “156” factories had to undergo in the second, reform period, most of them are still productive today and often remain leaders in their respective industries. They have also diversified their activities, their products now ranging from computer screens to carrier-based aircraft.

Finally, the reform and opening-up policies introduced by Deng Xiaoping starting in the 1980s shifted the center of the Chinese economy toward export-oriented manufacturing industries and later services. But the double-digit growth of the economy required a strong heavy industry to irrigate downstream production. It is therefore *a priori* unclear whether the “156” factories and counties where they were established were to win or lose from the profound changes undergone by the Chinese economy since the 1980-90s. This is the question we aim to answer in the next sections.

### 3 Data and empirical strategy

This section describes data sources, the empirical strategy and provides some descriptive statistics.

#### 3.1 Data sources

Big-push industrialization projects like the factories built under the “156” program can affect economic activity in a number of ways. First, the investments themselves entail a reallocation of factors of production across space and sectors. Second, the hope of big-push projects is to foster growth, which should lead to further factor reallocation in favor of the areas that received the investments and bring about structural transformation. Third, Million-Rouble plants may have spillover effects on other firms within treated areas as well as on neighboring counties.

In this paper, we mobilize six main sources of data to shed light on the short- and long-term effects of the “156” program: Information on the location and timing of the “156” investments, Population Census data, firm surveys,

satellite data, city-level information from statistical yearbooks and additional GIS data to create control variables and the instrument for factory location needed for identification.<sup>10</sup>

### Factory locations

The treatment that this paper studies is defined by the geo-coded location of the factories that constitute the “156” program. The counties containing these Million-Rouble plants are mapped in Figure III.1, where we can see that they span a wide variety of regions. We also collected information on the timing (start and end dates) of the construction of the plants, as well as the value of the investments. These pieces of information were extracted from various sources, primarily Bo (1991) and Dong and Wu (2004).

### Census data

China implemented its first Population Census in 1953. The data, aggregated at the county level, only provide population and household counts. The 1964 Census is richer. It disaggregates the population by sex and agricultural status. At the time of the command economy and strict commune-based provision of public services (health care, education and even clothes and food), the agricultural status or household registration (*hukou*) type, formalized in the wake of the Great Leap Forward, was a faithful reflection of both activity and the urban or rural environment of residence. This piece of information offers us the opportunity to start tracking the evolution of urbanization and economic sectors from 1964 onward. Additional county-level information is available in the 1982 Census: age distribution, birth and death rates, infant mortality rate, distribution of educational attainment, share of employed per-

---

<sup>10</sup>Additional data sources that we plan to tap include Industrial and Economic Censuses and annual data from county gazetteers and Urban Household Surveys. County gazetteers are currently being digitized and harmonized as part of Harvard University’s China Gazetteer Project—see <https://www.chinagazetteer.com/#/>. Variables that would be of particular interest for this paper pertain to industrial and agricultural production in the command-economy era, as well as information on population, broken down by education, age, gender and broad sector of activity. The Urban Household Surveys of the National Bureau of Statistics are a widely used data set. Researchers typically have access to data on the 2002–2008 period for a subset of provinces. The data contain a wealth of information on locally registered urban residents—on employment, income, education, age, marital status, sector of activity, consumption, etc.

sons in the population, literacy and a disaggregation of employment by broad sectors (agriculture and industry). Starting with the 1990 Census, more precise data were collected on the sector and type of employment and occupation, as well as on housing, which became urban households' main asset in the reform period. The following waves provide crucial information on migration, a phenomenon that mostly involved agricultural-*hukou* holders moving to cities in search of better earning opportunities and started to transform the Chinese economy and society in the 1990s. The 1990 Census records the number of locally registered residents as well as the number of people moving in, distinguishing between within-prefecture, within-province and extra-provincial mobility.<sup>11</sup> The 2000 Census further includes information on the place of residence in 1995, timing of the last migration spell, reason for migrating, and place and type of household registration. Similar information can be found in the 2005 1% Population Survey<sup>12</sup> and 2010 Census.

### Firm surveys

As far as firm outcomes are concerned, we shall rely on the National Bureau of Statistics (NBS) “above-scale” manufacturing firm (*“xian’e yishang” gongye qiye*) survey data, which constitute a longitudinal census of all state-owned enterprises (SOEs) and of all non-SOEs as long as their annual sales exceed RMB 5 million.<sup>13</sup> These data cover the manufacturing sector over the period 1992–2009 and contain a wealth of accounting information, in particular on wages. Compared to the firm data used in the relevant literature, e.g., Glaeser *et al.* (2015), the NBS surveys are available at the level of “legal units” (*faren danwei*). Different subsidiaries of the same firm, provided they have their own names, can sign contracts, possess and use assets independently, assume their liabilities and are financially independent (Brandt *et al.*, 2014), will be considered as separate “legal units.” The implication of this definition of units

---

<sup>11</sup>Below the central government, China is divided into provinces (*sheng*—Administrative level 1), which are subdivided into prefectures (*diji shi/qu*—Administrative level 2), also commonly referred to as “cities,” themselves made up of several counties (*xian*—Administrative level 3).

<sup>12</sup>The 2005 1% Population Survey is a nationally representative three-stage cluster sampling survey. The sampling frame is the 2004 population registry of the Public Security Bureau (Ebenstein and Zhao, 2015).

<sup>13</sup>Unique firm identifiers can be retrieved thanks to the algorithm designed by Brandt *et al.* (2014) to construct a panel of firms spanning the period 1998–2007.



of observation is that the concept of *faren danwei* almost perfectly coincides with that of an establishment, the common unit of observation in U.S. data: In 2007, nearly 97% of the units in our data corresponded to single plants.<sup>14</sup>

### City Statistical Yearbooks

We shall conceptualize the rise and fall of treated counties in Section 4 as hinging on a distortion of labor markets induced by the presence of large industrial plants. Data from the China City Statistical Yearbooks (*Zhongguo Chengshi Tongji Nianjian*) will be instrumental in shedding light on the effect of Million-Rouble plants on local (prefecture-level) wages, as well as on city-level employment shares, which we can use to explore the effect of the “156” program on tertiary employment. These data also allow us to highlight an important externality that the “156” factories may inflict on other firms in their localities, the pollution that they emit. Pollution indeed affects workers’ and entrepreneurs’ health and might influence firms’ location decisions at the margin. We measure the effect of the treatment on ground pollution readings averaged yearly at the prefecture level. The City Statistical Yearbooks cover the period 1994–2013.

An important challenge of the data collected by statistical offices—censuses, surveys and yearbooks—is that they rely on official administrative divisions at the time of data collection. Now, county, and to a lesser extent prefecture, boundaries are subject to frequent and sometimes substantial changes in China. To deal with this issue, we use the 2010 administrative map of China as our benchmark and re-weight the data collected in other years to match the 2010 borders. More precisely, we overlay the 2010 map with the map for every other year  $y$  and create a new map with all the polygons defined by the 2010 and year- $y$  divisions. We then compute the value of the variable of interest for each polygon<sup>15</sup> and collapse the values at the level of the 2010

---

<sup>14</sup>The NBS “above-scale” firm surveys can be complemented by the the Industrial Census of 1995 and the Economic Censuses of 2004 and 2008, which contain information on ownership type, industry, capital, investment, the number of employees, sales revenue, profits, wages and benefits, etc. They are made available to researchers in a coarser version than the original micro data and the firm surveys, and do not offer a panel dimension. They nevertheless present the advantage of being faithful snapshots of the economy.

<sup>15</sup>This computation is based on the assumption that the variable in question is evenly distributed across space within the year- $y$  county.

counties, using the share of each polygon in the total area of the 2010 county as a weight.

## GIS data

Finally, we collect geographic information system (GIS) data from various sources. GIS data serve two purposes in this paper. First, they allow us to track the evolution of living standards. The GDP figures provided by the National Bureau of Statistics at the county level shall indeed help us shed light on the treatment effect on economic activity, but they are available at large and irregular intervals (we use the 1982 and 2000 data), and measures tend to be noisy. We therefore resort to a second measure of economic activity (or population) as a complement: night-time luminosity captured between 1992 and 2013 by satellites. This measure is processed from aerial images that can be obtained online from the NASA.<sup>16</sup> Second, GIS data are needed for identification purposes as our instrument for industry locations relies on minimum-cost distances between Chinese counties and military bases. The complete list of airbases is provided in Appendix Tables III.B.5 through III.B.7. We also create control variables (listed in Appendix Tables III.B.1 and III.B.2) derived from precipitation data from the Global Historical Climatology Network, geocoded information on the ruggedness and elevation of the terrain, coal, ore, oil and gas deposits, and the location of the coast, rivers, roads and railroads.

## 3.2 Empirical strategy

In this section, we first describe the construction of the control group, based on variables orthogonal to military motives that determine a county's suitability for hosting investments in heavy industry. We then describe how we construct a measure of vulnerability to enemy bombing and use it to explain the choice of industry locations among suitable places.

---

<sup>16</sup>We use the "Average Visible, Stable Lights, & Cloud Free Coverages" available from <https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html>.

### Treatment and control groups

The empirical strategy consists of (i) generating a list of candidate locations as in [Kline and Moretti \(2014\)](#) and (ii) exploiting a source of exogenous variation in the selection process among these candidates.

The first determinant of eligibility for the “156” program, as described in [Bo \(1991\)](#), is access to raw materials: coal, ore, oil and gas deposits. In what follows, for the sake of exposition, we focus on coal fields to describe our methodology.

Our measure of access to the raw materials required for the industry relies on the plant’s connection to transportation network—[Bo’s \(1991\)](#) second eligibility criterion. We reproduce the transportation network in China at the time of the First Five-Year Plan using the existing railroad and road networks in 1948 and 1962, respectively, and calculate the average cost to coal fields. The railroad map that we use is reproduced in [Figure III.3](#). We create a grid over all of China, allowing for different costs of crossing a cell depending on the means of transportation available in that cell. Relative transport costs are based on [Fogel \(1964\)](#) and [Glaeser and Kohlhase \(2004\)](#).<sup>17</sup> We then calculate the minimum travel cost from the closest field for all points through the existing transportation network and collapse it at the county level. The spatial distribution of minimum distance costs to coal fields (using both rail and road networks) is displayed in [Figure III.4](#). This measure of access to coal fields through the existing transportation network will be the key variable to define the set of eligible counties.

---

<sup>17</sup>We derive the cost of transporting goods on roads by truck relative to rail from [Glaeser and Kohlhase \(2004\)](#), who estimate costs of 28 cents per ton mile for trucks and 3 cents per ton mile for rail in the United States at the end of the 20<sup>th</sup> century. We assume the same ratio holds for 1950s China and normalize transport cost through a cell to 1 if a railroad is present. The relative cost of transporting goods through cells that lie neither on a road nor on a railroad line is set at twice the transport cost by truck. This corresponds roughly to the average between the cost of following a road and that of transporting goods by wagon calculated by [Fogel \(1964\)](#). The rationale is that 1950s China had trucks and could extend roads at the margin, which reduces the transport cost in the absence of any infrastructure, but this cost should nevertheless be much higher than with existing facilities. Waterways are not considered because almost none of the factories were set up near a river in a place that was accessible neither by train or truck. Furthermore, only 20% of total freight traffic was borne by waterways in 1958, and only 13.2% of this (about 2.5% of the total) was carried out by barges, the rest being accounted for by wooden and sail boats adequate for grain but not for the industrial output of the “156” factories ([Rong, 2012](#)).

Although they do not feature among the list of explicit determinants, a visual assessment of the locations of the factories (Figure III.1) suggests that other geographical and economic factors entered siting decisions. We thus create an indicator variable that equals 1 if a county belongs to the provincial capital and 0 otherwise,<sup>18</sup> and variables that capture proximity to Ming-dynasty courier stations—a proxy for connectedness and market access,—county population at baseline (measured by the 1953 Census), county area (in logarithms), distance to the coast and proximity to a railroad hub. The exact definitions of these variables are provided in Appendix Tables III.B.1 and III.B.2. Importantly, our findings do not hinge on this specific set of variables; restricting ourselves to the determinants singled out by Bo (1991) yields similar results. It is worth noting at this stage that siting decisions were certainly informed by little more, and perhaps much less, than what we can control for thanks to detailed GIS data. The lack of a well-functioning statistical administration, which explains the delay in devising the First Five-Year Plan (1953–1957) after Communist victory in 1949, indeed put severe constraints on policy making in the early years of the People’s Republic of China.

We then regress our endogenous treatment on access to natural resources (and the additional determinants described above) to generate a propensity measure  $P_c$  for each county. We define the set of suitable locations  $C = \{c_1, \dots, c_N\}$  by matching the counties actually treated with their nearest neighbors (in terms of the propensity  $P_c$ ) in the control group.<sup>19</sup> We restrict the matching procedure to counties with a measure  $P_c$  in the support of the control group. Because of the potential spatial spillovers of the treatment, we impose that matched control counties be selected outside the immediate vicinity of treated counties. In the baseline, we exclude counties whose centroids lie within a 3 latitude degrees  $\times$  3 longitude degrees rectangle (roughly 2-3 times the size of the average prefecture) centered on a treated county.<sup>20</sup> As shown in Figure III.1, this exercise gives us 223 suitable counties, among

---

<sup>18</sup>We consider a county as part of the provincial capital if it belongs to the prefecture in which the provincial capital was located at the time of the First Five-Year Plan.

<sup>19</sup>In the baseline specification, we use the 5 nearest neighbors.

<sup>20</sup>In the current strategy, very few county characteristics are selected thus leaving many variables available for a balance test. By contrast, many variables could be used to refine the initial matching, thereby leaving few characteristics to compare across treatment and control groups in an “over-identification” check.

which 54 are treated.<sup>21</sup>

### Vulnerability

To isolate exogenous variation in the decision to select counties, we construct a measure of vulnerability to air strikes from major rear U.S. Air Force (U.S.A.F.) and Taiwanese bases, accounting for the presence of allied bases acting as a shield. To this end, we first geo-locate active U.S. Air Force bases and Taiwanese military airfields, as well as U.S.S.R. and North Korean major airbases. These bases are shown in Figure III.5, and the exhaustive list is provided in Appendix Tables III.B.5, III.B.6 and III.B.7. Second, we penalize travel time for enemy bombers in the vicinity of U.S.S.R. and North Korean bases. Intuitively, enemy bombers face a trade-off between minimizing travel time and avoiding the airfields of China's allies, proximity to which increases the risk of being shot down. In the baseline, we use a continuous measure of the cost for enemy airplanes of reaching any given point of the Chinese territory. Third, we compute the travel cost from each active U.S.A.F. or Taiwanese base to each county and define for each county a measure of vulnerability  $(V_c)_{c \in C}$  that is the minimum of the penalized distances across all enemy bases. We shall use this measure of vulnerability to enemy attacks in the empirics as an instrument for factory location decisions. We report an illustration of the minimal cost from U.S.A.F. and Taiwanese bases to control and treated counties in Figure III.5.

Figure III.2 illustrates graphically the definition of the treatment and empirical strategy. In Panel (a), we present a histogram of county density as a function of normalized distance to the closest factory. Treated counties—here, counties whose centroids lie within 20 km of a Million-Rouble plant—and control counties belong to clearly distinct groups. This reflects the restriction that we imposed on the selection of matched control locations and provides reassuring evidence that we should not be concerned about spillover effects into the control group. A similar picture obtains if we use instead a dummy equal to 1 if a county contains a factory and 0 otherwise as our definition of the treatment. Panel (b) provides a kernel density representation of the

<sup>21</sup>In total, 60 counties hosted at least one Million-Rouble plant. Six counties have to be dropped because they fail to meet the common-support requirement.

relationship between proximity to U.S. and Taiwanese airbases and factory location choices. Although we find both treated and control counties at most levels of vulnerability, which is due to the importance for location decisions of other criteria than security, first and foremost economic and geographical suitability, we see that the distribution of travel time across treated counties has a much fatter right tail than that of the control group, which shows that factories were preferably established at a distance from enemy threats.

### Benchmark specification

To estimate the impact of factories on economic activity across Chinese counties, we run a simple two-stage specification at the county level.

Let  $c$  denote a county and  $(T_c)_{c \in C}$  the treatment variable indicating whether a county actually hosted a factory. We estimate the following regressions by Two-Stage Least-Squares (2SLS):

$$\begin{cases} T_c = b_0 + b_1 V_c + \mathbf{b}_x \mathbf{X}_c + e_c \\ Y_c = \beta_0 + \beta_1 \widehat{T}_c + \beta_x \mathbf{X}_c + \varepsilon_c \end{cases} \quad (\text{S})$$

where  $Y_c$  is a measure of economic activity, employment, population or local amenities at the county level. The controls included are the logarithm of base-line population (measured in the 1953 Population Census), logarithm of the county's area, proximity to ports, proximity to 1900 cities, proximity to Ming-dynasty courier stations, distance to the coast, distance to military airfields and market access.<sup>22</sup> Definitions are presented in Appendix Tables III.B.1 and III.B.2, and descriptive statistics provided in Appendix Table III.B.3.

The NBS “above-scale” firm surveys and City Statistical Yearbooks are available at the prefecture (Administrative level 2) rather than county (Administrative level 3) level. We implement a different empirical strategy when analyzing the effect of Million-Rouble plants on prefecture-level outcomes. Because of a lack of power, we do not restrict the estimation to suitable areas selected by matching. Instead, we consider all non-treated prefectures as part of the control, and we ensure that they are observationally identical to treated prefectures by expanding the set of control variables, using the full list pre-

<sup>22</sup>Market access is defined as the sum of the populations of all other Chinese counties normalized by the inverse of distance.

sented in Appendix Tables III.B.1 and III.B.2. We also introduce prefecture fixed effects. This strategy allows us to keep all the prefectures in the data and thus retain sufficient statistical power.

### Identifying assumption

The key assumption, under which 2SLS is a valid strategy, is that the instrument has no effect on our outcomes of interest other than through the location of the Million-Rouble plants. A partial assessment of the identifying assumption is available from Figure III.5: The vulnerability measure does not overlap with the coast-interior divide that characterizes the spatial distribution of economic activity in China. Some factories were indeed set up on the coast, first and foremost in Dalian, but not on the southern shore, too exposed to American or Taiwanese strikes.

A concern with specification (S) is that vulnerability may nevertheless correlate with some unobserved amenities, which would explain both the decision to locate factories and later economic growth. For instance, China's south east is extremely vulnerable but widely benefited from the opening of Chinese ports to trade in the reform era. Such a violation of the exclusion restriction would induce a spurious negative correlation between economic growth and the presence of "156"-program industries in the reform period. To deal with this concern, we add direct controls for alternative channels, such as distance to the main trading ports and proximity to trading routes.<sup>23</sup>

### 3.3 Descriptive statistics

The "156" program represented a comprehensive push to expand and modernize Chinese industry in a wide range of sectors. Table III.1 shows the different industries concerned by the projects and the corresponding number of factories built. The "156" program, in keeping with the Soviet industrial development strategy, awarded the lion's share of funds to heavy, extractive and energy industries, e.g., aviation, coal mining or power plants. Construc-

---

<sup>23</sup>An additional check of the validity of our identification strategy would be to use bases that became active only during the Vietnam War to generate a counterfactual index of vulnerability and show its orthogonality to later economic outcomes among suitable locations. [Work in progress.]

tion started between 1953 and 1955, and was achieved at the latest in the first quarter of 1959. The last two columns of the table show planned and actual investment. The figures attest the scale of the program for a poor country like China in the 1950s: The average planned investment in aviation factories for instance was about 70,000,000 yuan, which amounts to some \$90,000,000 in 2010 U.S. dollars.<sup>24</sup>

About 5% of Chinese counties lie within 20 km of at least of one of the “156” factories—see Table III.2. Comparing treated counties with the rest of China based on this definition, we can confirm the historical accounts—in particular, Bo (1991)—of the decision process that presided over factory location choices and motivate our empirical strategy. As expected from a context of heightened international tensions in Asia following the Korean War, treated counties are located at a much greater distance from U.S.A.F. and Taiwanese bases: The difference in mean penalized distance between treated places and the average Chinese county is about half a standard deviation. Differences in terms of population are small at baseline (1953 Census) and tend to grow over time. Treated counties were less agricultural in 1964, 1982 and 1990. Unfortunately, the 1953 Census does not contain information on sector of employment or urbanization. Part of the difference in 1964 may thus be due to the treatment itself. As far as geographic controls are concerned (bottom of Table III.2), treated counties exhibit lower travel costs to coal, iron ore, oil and gas resources. They are not significantly closer to ports and lie at a similar distance from the coast and markets, but they are noticeably closer to railroads. Finally, factories were more likely to be built in flatter counties at low elevation.

## 4 Results

The “156” factories built thanks to Soviet cooperation are celebrated as having laid the foundations for the industrialization and modernization of the Chinese economy. Their influence on local trajectories however spans two very different periods, and it is unclear whether treated counties gained from the “156”

---

<sup>24</sup>In 1957, RMB 70,000,000 represented 12,000,000 Soviet roubles and USD 10,000,000. In 2010 prices, the average investment in the aviation factories of the “156” program would be close to 3.3 billion roubles.



program in the long run.

In what follows, we therefore systematically distinguish between the command-economy and reform periods. We first examine the evolution of local economic outcomes in the counties that hosted a Million-Rouble plant and provide evidence of a rise-and-fall pattern. Second, we present a conceptual framework based on distortions in factor markets as the key mechanism behind the demise of treated counties in the reform era and provide preliminary tests of its predictions.

## 4.1 The rise and fall of the “156”-program counties

**Siting decisions** Before comparing the impacts of the program in the two periods of interest, we present evidence of the explanatory power of vulnerability to enemy bombings, which shall be critical for identification.

Vulnerability to enemy bombings is a strong and significant determinant of the probability for a county to host a Million-Rouble plant. Table III.3 displays the results from the first stage of our main Two-Stage Least Squares strategy, based on four different definitions of the treatment. Column 1 uses as a dependent variable a dummy equal to 1 if a county received a factory under the “156” program and 0 otherwise. Column 2 defines as treated a county whose centroid lies within 20 kilometers of a factory. Column 3 adopts a similar definition, except that the radius is extended to 50 km. Column 4 has on the left-hand side a continuous variable that captures distance to the nearest factory. More precisely, “Distance” is the standardized logarithm of distance to the nearest factory. All columns restrict the estimation to the set of treated and control counties defined by matching on access to natural resources and the additional economic and geographical determinants presented in Section 3. A rich set of variables is also included to control for determinants of local suitability for industrial investments that might have a direct effect on outcomes of interest in the second stage; results are however robust to removing all controls, as well as to focusing the matching procedure on the determinants explicitly used for siting decisions. The control variables are detailed in Appendix Tables III.B.1 and III.B.2.

Moreover, our excluded instrument, “penalized distance,” affects factory location in the expected direction. As can be seen from Table III.3, greater

distance from U.S. and Taiwanese bases (taking into account the protective role of U.S.S.R. and North Korean airfields) reduces the threat of enemy bombing and thus leads to a higher probability for a county to receive one of the Million-Rouble plants.

### **Before the transition**

To study the long-term effects of the “156” program on the spatial distribution of economic activity in the reform period, we first need to assess the impact of the program on treated counties by the time the reforms were introduced.

Industrial investment requires two main inputs: capital and labor. The visible part of the “156” program consists of large capital investments in the form of buildings and machinery. There was also a major human capital input, first through Soviet experts and then through the Chinese engineers and managers that they trained *in situ*, as well as Chinese students who graduated from Soviet universities and technological institutes and were hired on the “156” plants upon their return. However, the factories would not have been operational without workers.

While some free movement of labor still occurred after the advent of “New China” in 1949, mobility was subject to authorization starting in the late 1950s. The government had indeed tightened its grip on labor movement in the wake of the Great Leap Forward, when policies encouraging self-sufficiency at the commune level led to famines and distress migrations that threatened the sustainability of urban food provision systems. Between the construction of the “156” factories and the advent of the reforms, the government would thus instruct workers where to live and where to work to accommodate rising demand for labor in some (treated) counties and ensure the growth of the plants and local economy. Changes in population and the non-agricultural share of employment thus reflect both demographics and the government’s industrial policy. They offer us a clear insight into the scale of “156” program investments—combined with potential agglomeration effects that are difficult to isolate in a command economy. Population can be construed as a first approximation for economic activity and the reallocation of production factors across space induced by the Million-Rouble plants. Non-agricultural employment reflects both industrial activity (at a time when services accounted for a

relatively small share of the labor force) and factor reallocation across sectors, which itself strongly correlates with urbanization.

We find that the reception of industrial investment under the “156” program had a strong positive impact on population in the earlier period. Table III.4 uses a dummy equal to 1 if a county received at least one factory, and 0 otherwise, as our treatment variable, and its first two columns focus on the logarithms of population measured in the 1964, 1982 and 1990 Censuses as dependent variables.<sup>25</sup> The significant OLS effects are confirmed by the 2SLS estimations. The effects are very large: In 1982 for instance, the population of treated counties was on average 1.2 times as large as in the control group. Interestingly (but not significantly), the impact of the Million-Rouble plants seem to have become more muted in the early reform era (between 1982 and 1990), in sharp contrast to the large and significant increases observed between 1964 and 1982.

The “156” program also led to an increase in the share of the population that is not employed in agriculture. The 2SLS strategy again yields coefficients that are consistent with the OLS estimates. As when population is the regressand, we however note that the 2SLS coefficients tend to be larger, and often significantly so. This likely reflects the fact that the places selected to host a Million-Rouble plant had not been the scene of major industrial developments prior to the First Five-Year Plan (Bo, 1991). In 1982, treated counties had on average a share of non-agricultural employment 62 percentage points higher than control counties. This is the equivalent of moving from the median to the top 5% of the distribution in the control group. The impact on the non-agricultural share of the population shows a strong influence of the “156” program on the allocation of labor, which can be seen as evidence of structural transformation and urbanization. It is worth noting that we find in Table III.4 that the effect of the “156” plants still survives in the 1990 Census, as evidenced by Panel C, despite the fact that the national average share of

---

<sup>25</sup>Where to draw the line between the command-economy and reform eras is a vexed issue as the reforms in the non-agricultural sector were introduced gradually. Whereas private firms were allowed to develop and compete with state-owned enterprises (SOEs) from the mid-1980s onward, which was instrumental in introducing market discipline in SOEs, the big privatization wave did not start until the 1990s. We thus extend both the command-economy and reform periods to the maximum possible in order to provide conservative evidence of the stark reversal that we highlight.

non-agricultural employment was much higher in 1990 (at 27%) than in 1964 (15%). Finally, note that the F-statistic reported from the first stage provides reassuring evidence against weak-instrument bias.

The large increases in population and its non-agricultural share associated with the presence of a Million-Rouble plant are also found when we relax the definition of the treatment. We allow neighboring counties whose centroids lie within 20 km of a factory to be considered as treated, even if no factory was established within their borders. A circle with a 20-km radius has about the same area as the median treated county. The results, displayed in Appendix Table III.B.4 are in line with those of Table III.4.

These results show large effects of the “156” program. The increase in the population of a county, for instance, that we can attribute to the treatment, is much larger than the number of workers that we would expect a plant to employ. The change in the non-agricultural share of employment reinforces this impression: The treatment affected local economies beyond the size of the factories constructed thanks to Soviet assistance. Evidence of agglomeration effects in the context of pre-reform China is difficult to establish because of a lack of data on productivity and because of the nature of planned economies: The observed increase in population may well lag behind labor demand in treated counties; it is however impossible to rule out that the government encouraged hiring in its flagship industries to hasten the development of a modern working class. The magnitude of the coefficients and their persistence through 1990 nevertheless suggests that agglomeration effects might have been at work. The estimated effects are very similar when we use the proximity version of the treatment variable (Appendix Table III.B.4).

Additional outcomes can help us shed light on the effect of the “156” program at the onset of the reforms. The overall picture is that hosting a Million-Rouble plant made a county richer and more developed. Table III.5 displays the impact of the factories (according to the first definition of the treatment) on GDP per capita in logarithms, the illiteracy rate and the male-to-female ratio. The last two variables can be seen as proxies for lagging social development, which economic growth and urbanization are likely to pave the way for. All three dependent variables are measured in 1982. To save space, Table III.5 focuses on 2SLS estimates. We find that building one or more factories under

the “156” program in a county had a strong and significant positive effect on GDP per capita: GDP per capita was more than four times as high in the treatment as in the control group, which is the equivalent of moving from the median to close to the top 10% of the control-group distribution. Table III.5 further suggests that the treatment had reduced the illiteracy rate by more than 11 percentage points and the male-to-female ratio by 2 percentage points (not significant) by the time the reforms were introduced. Note that these findings need not be evidence of a change in local norms or human capital accumulation. Since workers were assigned to work units as part of the Plan and an urban job—in particular in one of the “156” plants but more generally in any factory—was considered a good position for semi-skilled workers in this agrarian society, the higher social indicators that we observe in treated counties are partly due to positively selected immigration.

To summarize the evidence on the impact of the “156” program in the command-economy era, receiving a Million-Rouble plant is found to have had a very large impact on a county’s population, its GDP per capita and the speed of its structural transformation by 1990. Treated areas thus entered the subsequent reform period with a substantial head start.

### **During the transition**

We now investigate how the treated counties fared in the transition period. Recent decades allow us to avail ourselves of more direct proxies for economic activity and the level of development. The gradual transition from a command economy to freer market interactions, as well as the development of the tertiary sector, also implies that population and the non-agricultural share of employment are likely to offer poorer approximations of economic activity than in the previous period.

We find that the introduction of economic reforms led to an erosion of the head start of treated counties. To capture economic activity or population, we first use remotely sensed night-time lights and then GDP. The first column of Table III.6 regresses the logarithm of night-time luminosity in 1992, which corresponds to the time when the wave of reforms started to transform the manufacturing sector, on our “factory” treatment. We can see that the effect is strong, positive and significant, which implies that the “156” program

had been largely beneficial to treated counties by 1992. The next column defines the dependent variable as the log-difference in night-time luminosity over the period 1992–2013. The effect is not only negative and significant; it substantially offsets the gains from the previous period. Between 1992 and 2013, treated counties indeed lost about 40% of their initial advantage over the control group. A similar, yet much noisier, picture obtains when one compares GDP in 1982 and in 2000 (Columns 3 and 4 of Table III.6): Whereas the “156” program caused treated counties to enjoy a much higher GDP in 1982, almost half of those gains were wiped out in the following period. A potential concern is that this pattern could be explained by the sheer share of the “156” factories in the local economy. Appendix III.A explores and rejects this idea by comparing firms descended from “156”-program factories with other “above-scale” establishments in the same prefecture.

## 4.2 Mechanism behind the fall of treated counties

We delineate here a conceptual framework to explain the relative decline of the “156”-program counties after the introduction of the reforms and then present further empirical results that test the hypothesized mechanism.

### Environment

We think of the effect of the treatment in the reform period along the lines of a modified Buera and Shin (2013) model with the following ingredients. Agents differ in entrepreneurial ability and can choose each period between becoming entrepreneurs and workers based on the respective expected utilities of these options.<sup>26</sup> Entrepreneurs produce with capital  $K$  and labor  $L$ , subject to a collateral constraint  $K \leq \lambda\Omega$ , where  $\Omega$  is the endowment.<sup>27</sup> Importantly, the model assumes within-period credit, which implies that the only way for entrepreneurs’ start-up firms to cross the so-called “valley of death” is to generate

<sup>26</sup>The first change introduced by the reforms was indeed that entrepreneurs could create firms and employ workers, now free to decide their “work units” (*gongzuo danwei*) without the planner’s interference.

<sup>27</sup>This feature from Buera and Shin (2013) is well suited to the Chinese environment. Chinese entrepreneurs in the 1980s and 1990s, and to a large extent this is still true today, faced severe credit constraints as banks were publicly owned and state-owned banks tend to lend to state-owned firms (Boyreau-Debray and Wei, 2005; Song et al., 2011).

profits immediately.

Besides these ingredients already present in [Buera and Shin \(2013\)](#), we introduce labor-market frictions. In the absence of venture capital, start-up firms are more likely to survive when labor is cheap or easily accessible. Now, the presence of the “156” factories can be supposed to increase the cost for entrepreneurs of hiring labor.<sup>28</sup> The key mechanism is that the cost of entry is higher in treated areas.

### Labor mobility

This main mechanism holds true irrespective of labor mobility, but allowing workers and potential entrepreneurs to move alters its quantitative implications.<sup>29</sup> Without mobility, higher wages in one labor market reduce incentives to create firms, but this adverse effect on entrepreneurship is partly mitigated by the fact that more agents choose to become workers, which exerts a downward pressure on wages and thus ensures that firm creation remains attractive for some.

Labor mobility means that this compensatory effect no longer obtains. If wages are too high in one labor market for agents to become entrepreneurs, they can now avail themselves of better opportunities in less distorted markets, where labor is cheaper. Because labor mobility provides an outlet for would-be entrepreneurs, competition for wage employment decreases, and labor does not become cheaper in the distorted economy. In the well-functioning market, the arrival of new entrepreneurs induces a shock to labor demand. Its effect on wages is however moderate, as marginally higher wages attract workers from

---

<sup>28</sup>This effect could operate through higher wages as the data seem to suggest (see [Table III.8](#) presented below), housing and other benefits, job security or higher local prices. What is needed for the main mechanism to obtain is that workers living near a Million-Rouble plant have a higher reservation wage, which implies that there are fewer entrepreneurs.

<sup>29</sup>Restrictions on factor mobility and particularly labor movements were only gradually lifted in the reform period. Even today, migration is widespread and *de facto* restrictions on hiring are low (at least in the private sector), but barriers outside of the labor market still curb mobility by excluding non-locally registered workers from a variety of public services (health, education, etc.) and rights (e.g., home buying). Migration costs were compounded in the early days of the reform era by the small size of the free market for goods. This prevented migrants from venturing far or a long time away from their places of household registration (*hukou dengji di*) as they could not earn or redeem at destination the food ration coupons (*liang piao*) that remained vital in the early 1980s, were phased out during the decade and were finally abolished in the first half of the 1990s.

outside, which keeps wages down and thus does not stifle entrepreneurship. Subsequently, labor mobility amplifies the effect of distortions, which inhibit firm creation, as all potential entrepreneurs converge to less distorted labor markets.

## Predictions

We divide the reform era into two sub-periods: First, since at the beginning of the reform period labor is still immobile, workers are mostly employed by the Million-Rouble plant, and it is too costly for entrepreneurs to enter. This implies that we should observe lower firm entry and higher industrial concentration in treated areas. Once labor mobility becomes possible, some workers continue to work for the plant, but all other workers move to less distorted labor markets, where they can become either workers in recently set-up firms or entrepreneurs. Subsequently, population dwindles, industrial concentration further increases, etc.

We would first expect lower firm entry and size in treated areas. Unfortunately, we have to defer this test as we would need Economic Census data to capture the entire firm size distribution. The prediction of higher concentration around the “156” factories can however be assessed thanks to the “above-scale” firm data, although in this case as well Economic Censuses would provide a more faithful picture of the treatment effect on the local economy. Finally, in China’s context, where the large plants are incumbents exogenously attributed to counties rather than the firms set up by the most talented entrepreneurs (as in Buera and Shin’s world), the tertiary sector, which was underdeveloped in the first period, should expand much more in control counties.

The hypothesized mechanism first relies on the idea that large plants lead to distortions in local labor markets, for which we find empirical support. The City Statistical Yearbooks indeed provide evidence of significantly higher wages in treated prefectures in the transition period. The estimates are reported in Table III.8. Evidence supporting this conjecture also emerges in Appendix III.A, where we compare factories originating from the “156” program with the rest of the manufacturing sector in treated prefectures. Appendix Table III.A.1 shows that wages are significantly higher in the former and that



this pattern is not due to heterogeneous labor since it survives controlling for high-skill employment shares. These two pieces of evidence suggest that the high average wages paid by the regime's flagship factories have stifled entrepreneurship and the entry of other firms in treated areas.

The data support the first prediction of the theoretical framework: The "156" program caused higher industrial concentration in treated places. We compute a Herfindahl-Hirschman Index (HHI) using shares of local employment or output. We present the results in two sub-periods—between 1994 and 2000, and 2001–2013, i.e., before and after China's entry into the World Trade Organization (WTO). Note that this index is defined across firms, not sectors, so that it captures how much employment or production is dominated by just a few firms. Table III.9 uses the NBS annual firm surveys to test the first prediction. Columns 1 and 2 regress the Employment HHI on the treatment variable for the 1993–2000 and 2001–2007 periods, respectively. Columns 3 and 4 perform a similar exercise but express concentration in terms of output. In both sub-periods and under both definitions, we find that industrial concentration, which we normalize to range between 0 and 1, was substantially higher in treated counties—and significantly so in most cases. The effect is large: Hosting one or more Million-Rouble plants raises the Output HHI by 45 percentage points. This is the equivalent of going from the 25<sup>th</sup> to the 95<sup>th</sup> percentile.

We next provide evidence of the negative treatment effect on the tertiarization of the economy predicted by our theoretical framework. Table III.7 uses the China City Statistical Yearbooks (1994–2013) to show the differential evolution of sectoral employment shares between treated and control prefectures, before and after China's admission into the WTO. Interestingly, treated prefectures in 1994–2000 had an approximately 5 percentage point lower share of employment in services and 5 percentage point higher share of employment in the manufacturing sector, although none of these coefficients is statistically significantly different from zero. Conversely, in the next sub-period the secondary sector commanded a 20-percentage point higher share of employment in treated prefectures, at the expense of tertiary activities, which attracted a 16-p.p. lower share of employees.<sup>30</sup> Table III.7 thus shows that treated ar-

---

<sup>30</sup>Note that prefectures typically contain between 5 and 10 counties. The effect of the

areas continued to be significantly more industrialized in the reform period and largely missed the trend of growing tertiary sector share. Since the tertiary sector in China developed mostly after the introduction of the reforms and its growth was predominantly due to new entrants, Table III.7 can be regarded as suggestive evidence of lower firm entry in areas treated under the “156” program.

The specialization in the secondary sector and failure of treated areas to embark on the second structural transformation—from manufacturing to services—is also consistent with the much higher levels of air and water pollution that they suffer from. Table III.10 regresses industrial SO<sub>2</sub> emissions, industrial discharge of waste water and industrial emissions of smoke and dust on our treatment variable. Data on the dependent variables come from the City Statistical Yearbooks, so the unit of observation is a prefecture×year. Except for waste water, which however exhibits a consistent pattern, we find significant effects of the “156” program on industrial pollution between 2003 and 2013. The magnitude of the effects is large. For instance, being in a treated prefecture doubles total industrial emissions of sulfur dioxide compared to the Chinese mean.

### 4.3 Alternative interpretations

Several interpretations can be put forward to explain the rise and fall of places that benefited from “156”-program investments. The hypothesized mechanism presented in Section 4 hinges on the combined roles of labor market distortions generated by the presence of a Million-Rouble plant and migration restrictions in the early reform period. We now explore potential alternative explanations for the demise of treated counties.

First, it might well be that the big-push industrialization launched in the 1950s eventually failed because it constituted a wasteful diversion of resources to areas that had no comparative advantage in heavy industry. If, for instance, the security imperative that entered siting decisions led to the selection of systematically more remote and less productive places, resources would indeed

---

treatment on employment shares is thus expectedly diluted at the prefecture level. This—along with the different time frames—explains the gap in estimates between Tables III.4 and III.B.4 on the one hand and Table III.7 on the other.

have naturally flowed back to where market forces would have directed them in the first place, absent the government's intervention and heavy-handed management of labor and capital. The matching strategy presented in Section 3 is however designed to select a control group that is virtually identical to treated counties in terms of suitability to receive a large industrial plant. The only parameter that distinguishes treated and control areas within the eligible set of counties  $C$  is vulnerability to enemy attacks. We can therefore be confident that the relative demise of the "156"-program counties is due to the treatment, not to mean reversion. The fact that the results are robust to conditioning on an extensive set of geographical and economic controls reinforces this conviction.

Reversion to the mean is also at odds with other pieces of evidence. Our finding that treated areas have a weaker service sector, for instance, than similar places, while the descendants of the "156" plants represent a relatively small share of local economies, suggests that the treatment affected the industrial landscape beyond the "156"-program firms and altered the way they developed. Treated counties did not just revert to the path followed by identical places in the control group.

It moreover seems difficult to argue that the "156" program constituted a waste of financial and human resources and of the modern technology shared by the U.S.S.R. This is a major point of departure between the "156" projects and the Third Front Movement starting in 1964, which relocated factories to the interior according to this principle: "Close to the mountains, dispersed, and hidden in the caves" (*kao shan, fen san, jin dong*), to avoid air strikes from China's new enemies—the U.S.S.R. and its Asian allies. Apart from the threat of U.S. and Taiwanese bombings, the "156" projects were sensibly located from an economic point of view—near natural resources, transportation networks and markets.

Another interpretation relates to the ownership of the "156" plants. When they were established, all factories were publicly owned. If such emblematic factories as the "156" were more likely to remain state-owned even through the big privatization wave of the late 1990s, or if treated counties were heavily reliant on state-owned enterprises (SOEs), we might expect inefficiencies to aggravate in treated counties where the public sector would predominate,

while state involvement used to be the key to their economic rise before the reforms.

We indeed find that firms descended from the “156” factories were 9 percentage points more likely (controlling for industry, prefecture and year) to be SOEs over the period 1998–2007. A similar picture arises when we consider the share of the state in a firm’s registered (or “paid-up”) capital as our measure of public ownership. Unfortunately, our current firm data do not allow us to compare the share of publicly owned firms across treated and control counties. The NBS “above-scale” firm surveys indeed cover all state-owned firms but do not offer a representative sample of the rest of the manufacturing sector. Economic Census data should help us investigate further the impact of the treatment on the overall industrial landscape.

Lastly, our current data do not enable us to disentangle which channels led to industrial concentration and muted entrepreneurship.

Large firms may crowd entrepreneurs out of local resources, such as inputs and financial capital. Access to capital is particularly critical for new, private firms as all banks in China during our period of interest were state-owned. Conversely, current and former SOEs can avail themselves of long-lasting relationships and frequent mobility between their management, local administration and financial institutions. We can however argue that accessing capital was a challenge for entrepreneurs across China, whereas we would expect Million-Rouble plants to have an effect on local labor markets. Large establishments might also monopolize raw materials and intermediate inputs if high transportation costs or other frictions prevent them from flowing freely across space. They can also outsource inputs directly, which is not conducive to the variegated environment that new firms need to grow. Entrants may find it difficult to hire workers as big plants are typically located on the outskirts of cities, which makes it costly for households to provide workers both to the large factory at the periphery and to small firms in the city center. Finally, since cities around the “156” factories grew dramatically as a result of the treatment, they might not sport a wide range of services and small enterprises catering to the local market.

Besides such static factors, [Chinitz \(1961\)](#) and [Glaeser et al. \(2015\)](#) suggest that parents transmit skills and attitudes to their offspring, so that en-

entrepreneurial spirit may fade away generation after generation in highly concentrated industrial landscapes. Another dynamic explanation could be that large firms are more likely to collude with local governments to erect barriers to entry.

## 5 Conclusion

The big-push investments of the past may have long-lasting effects on local economies. However, this impact need not be linear.

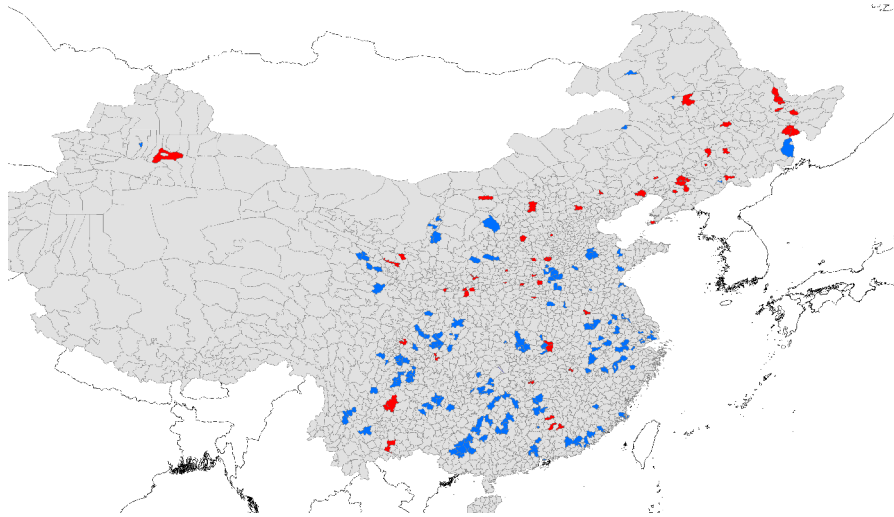
This paper showed that the “156” program was effective in spurring a dramatic structural transformation from agriculture to manufacturing and triggering urbanization. It also raised living standards substantially, so that treated areas enjoyed a much higher level of development than similar control counties when the reforms were introduced.

The reforms however reversed the fate of the places that had received investments under the “156” program. A large share of the GDP gains from the command-economy period vanished in the course of the opening-up era, bringing treated and control counties closer over time. However, treated areas did not merely revert to the path followed by other places in the absence of the “156” program; they failed to embark on the shift from manufacturing to services that similar places experienced in the reform period, which shows that the “156” plants had a persistent, albeit now adverse, influence on local economies.

Our preferred interpretation of the long-term effect of the big-push investments carried out as part of China’s cooperation with the Soviet Union is that the Million-Rouble plants stifled entrepreneurship by distorting labor markets: Treated places were successful in the centrally planned economy but did not provide an environment conducive to entrepreneurship and therefore failed to thrive in the reform era. Additional evidence is needed to buttress this interpretation but it is consistent with the pattern observed. Based on the preliminary results presented in this paper, high wages seem to have crowded out entrepreneurs, which induced industrial concentration and eventually a relative population and economic decline.

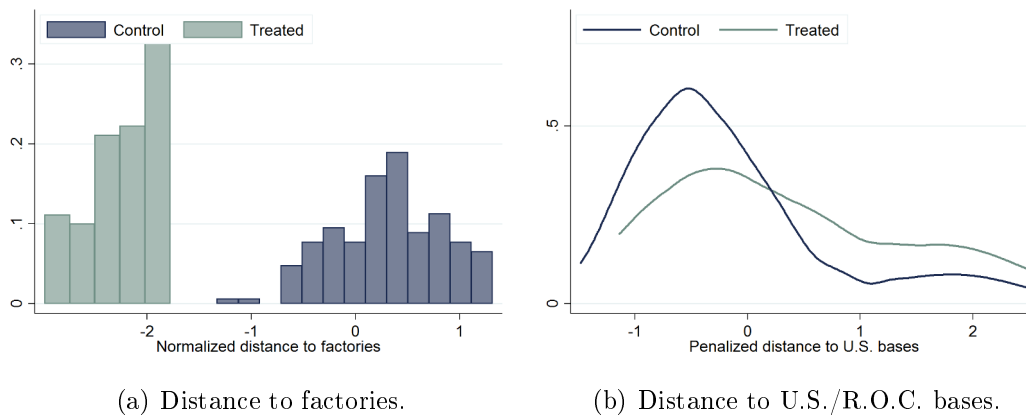
## Figures and tables

Figure III.1 – Selection of a matched group of control counties.



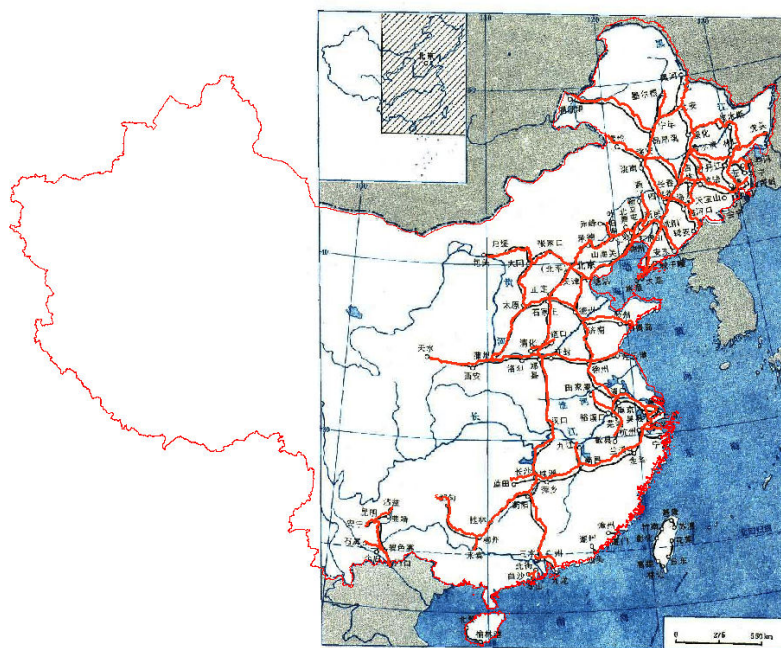
Notes: This map shows in red counties that host at least one “156”-program factory and in blue the control group of counties. The control group is selected through the matching procedure described in Section 3.

Figure III.2 – Treatment and control counties (proximity to factories, and to U.S. bases).



Sources: *Distance to factories* is the standardized distance to the closest factory. *Distance to U.S./R.O.C. bases* is the standardized distance to the main military U.S. or Republic of China (Taiwanese) airfields penalized by the proximity to U.S.S.R. and North Korean airfields. Treatment is defined as a dummy equal to 1 if a county centroid lies within 20 km of a factory and 0 otherwise. The control group is selected through the matching procedure described in Section 3.

Figure III.3 – Railroad map of China – geo-location.



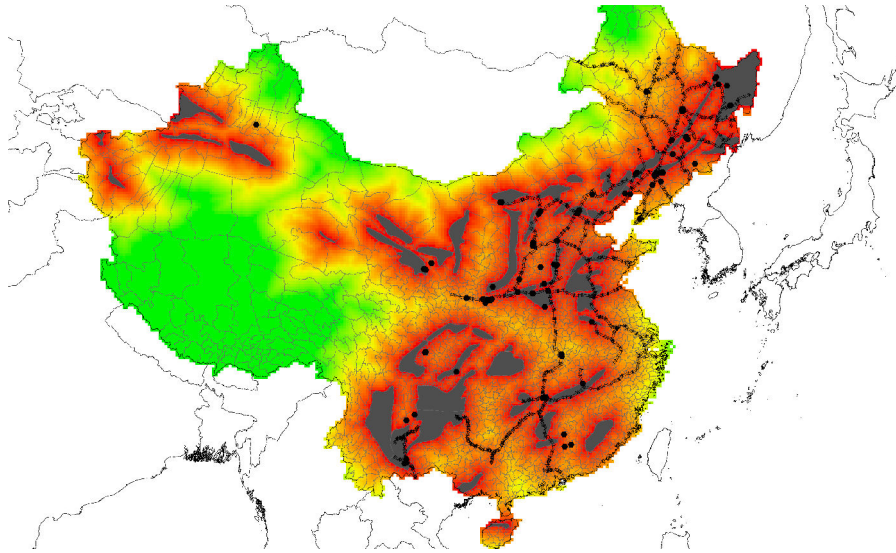
Sources: Railroad Map of China (1948, Joint Intelligence Committee). Black lines are from the original source. Red lines are the inferred geo-located polylines using current geo-coded railroad lines and cities.

Table III.1 – Descriptive statistics.

Industries	Number	Construction		Investment	
		Start	End	Planned	Actual
Aviation	14	1953.9	1957.3	7271	7204
Chemical	7	1955.3	1958.4	15291	15474
Coal mining	25	1954.3	1958.5	5323	5832
Electronic	10	1955.5	1957.9	5661	4752
Iron and Steel	7	1953.9	1959.0	78361	84586
Machinery	23	1954.8	1958.2	9972	10336
Nonferrous Metals	13	1955.1	1959.0	15018	15451
Powerplants	23	1954.0	1957.9	13039	9023
Weapons	16	1955.1	1958.4	13533	12262
Other	12	1955.3	1959.3	11751	12513

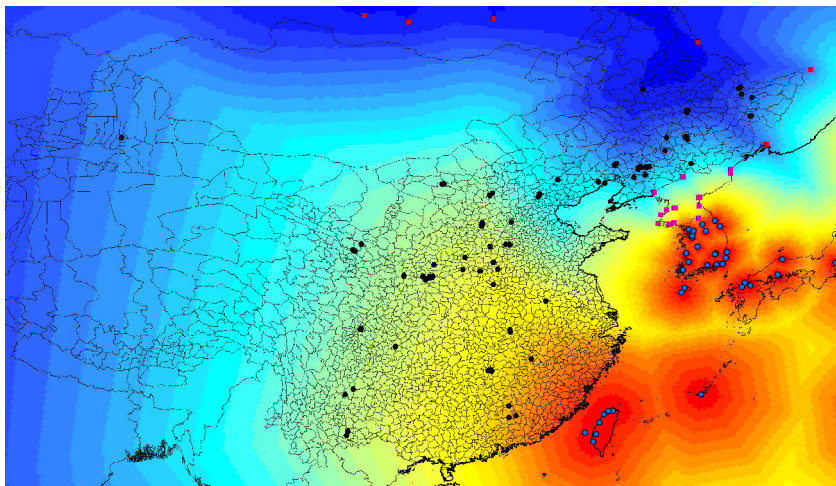
Notes: Other industries are shipbuilding, pharmaceutical and paper-making industries. Investment is in 10,000 yuan (1960) and RMB 1 = USD 0.15 in 1960, so that the average planned investment in aviation factories is about \$10 million in 1960 USD and about \$90 million in 2010 USD.

Figure III.4 – Factory locations and access to natural resources (coal).



Notes: This map represents the minimum travel time to coal-bearing areas using the railroad and road networks (red: low travel time, green: high travel time). Railroads and roads are geo-located from 1948 and 1962 maps, respectively. Factory locations are indicated with black dots, coal-bearing zones are highlighted with gray areas.

Figure III.5 – Factory locations and penalized distance to U.S./Taiwanese airfields and bases.



Notes: This map represents the penalized travel time from U.S. or Taiwanese airfields (red: low travel time, blue: high travel time). U.S. and Taiwanese airbases are indicated with blue dots, while U.S.S.R. and North Korean airbases are indicated with red and purple rectangles, respectively.



Table III.2 – Descriptive statistics.

VARIABLES	Mean	Std dev.	With factory	No factory
Observations	2,408		111	2,297
<i>Treatment</i>				
Factory	0.046	0.209	1.000	0.000
Penalized distance	0.000	1.000	0.467	-0.022
<i>Population</i>				
Population (1953, log)	12.20	1.01	12.41	12.19
Population (1964, log)	12.03	2.19	12.84	11.99
Population (1982, log)	12.62	0.98	13.18	12.60
Population (1990, log)	12.71	0.99	13.27	12.68
<i>Urbanization</i>				
Share non agr. (1964)	0.154	0.210	0.332	0.146
Share non agr. (1982)	0.230	0.194	0.437	0.220
Share non agr. (1990)	0.269	0.215	0.476	0.259
<i>Geographic controls</i>				
Area (log)	7.64	0.957	7.24	7.66
Travel cost to coal mines (log)	14.06	1.63	12.68	14.12
Travel cost to ore (log)	13.97	1.31	13.07	14.01
Travel cost to oil/gas (log)	14.10	1.58	13.39	14.13
Proximity to ports	0.148	0.355	0.153	0.148
Distance to rail (log)	10.80	2.04	9.05	10.89
Distance to the coast (log)	12.57	1.73	12.84	12.56
Distance to airfields (log)	10.71	1.00	10.14	10.73
Market access (log)	16.76	0.471	16.75	16.76
Slope (100 degrees)	0.024	0.024	0.017	0.025
Elevation (m)	846.1	1096	546.0	860.6

Notes: Penalized distance is standardized (mean 0 and variance 1 over the whole sample of counties).

Table III.3 – Factories and penalized distance to U.S./Taiwanese airfields and bases.

VARIABLES	Factory (1)	Proximity (20 km) (2)	Proximity (50 km) (3)	Distance (4)
Penalized distance	.211*** (.045)	.206*** (.045)	.216*** (.046)	-.929*** (.149)
Observations	220	220	220	220
Controls	Yes	Yes	Yes	Yes

Notes: Standard errors are reported between parentheses and clustered at the level of a rectangle of 3 latitude degrees  $\times$  3 longitude degrees. Each cell is the result of a separate regression. The unit of observation is a county (Administrative level 3). *Penalized distance* is the normalized distance to the main military U.S. and Taiwanese airfields penalized by proximity to U.S.S.R. and North Korean airfields. Controls include population in 1953 (log), county area (log), proximity to major ports (through the river network), proximity to Ming-dynasty courier stations, distance to the coast, distance to military airfields and market access (population of neighboring counties normalized by the inverse of distance).

Table III.4 – Factories and employment in 1964, 1982 and 1990 (before the transition).

Panel A: 1964	Population		Share non-agr.	
	(OLS)	(2SLS)	(OLS)	(2SLS)
Factory	.338*** (.086)	.421** (.165)	.279*** (.044)	.560*** (.102)
Observations	220	220	220	220
F-Stat. (first stage)		21.67		21.67
Controls	Yes	Yes	Yes	Yes
Panel B: 1982	Population		Share non-agr.	
	(OLS)	(2SLS)	(OLS)	(2SLS)
Factory	.451*** (.077)	.796*** (.217)	.284*** (.039)	.619*** (.099)
Observations	220	220	220	220
F-Stat. (first stage)		21.67		21.67
Controls	Yes	Yes	Yes	Yes
Panel C: 1990	Population		Share non-agr.	
	(OLS)	(2SLS)	(OLS)	(2SLS)
Factory	.484*** (.095)	.784*** (.237)	.311*** (.048)	.633*** (.116)
Observations	220	220	220	220
F-Stat. (first stage)		21.67		21.67
Controls	Yes	Yes	Yes	Yes

Notes: Standard errors are reported between parentheses and clustered at the level of a rectangle of 3 latitude degrees  $\times$  3 longitude degrees. Each cell is the result of a separate regression. The unit of observation is a county (Administrative level 3). Controls include population in 1953 (log), county area (log), proximity to major ports (through the river network), proximity to Ming-dynasty courier stations, distance to the coast, distance to military airfields and market access (population of neighboring counties normalized by the inverse of distance). The instrument is the distance to the main military U.S. and Taiwanese airfields penalized by the proximity to U.S.S.R. and North Korean airfields. *Population* is the logarithm of total population in the county and *Share non-agr.* the share of the population that has a non-agricultural household registration (*hukou*). The F-statistic displayed is the Sanderson-Windmeijer F-statistic.

Table III.5 – Factories and additional outcomes (1982).

VARIABLES	GDP p.c.	Illiteracy	Male-to-female
Factory	1.648*** (.326)	-11.455** (5.343)	-2.413 (3.210)
Observations	220	220	220
F-Stat. (first stage)	21.67	21.67	21.67
Controls	Yes	Yes	Yes

Notes: Standard errors are reported between parentheses and clustered at the level of a rectangle of 3 latitude degrees  $\times$  3 longitude degrees. The unit of observation is a county (Administrative level 3). Controls include population in 1953 (log), county area (log), proximity to major ports (through the river network), proximity to Ming-dynasty courier stations, distance to the coast, distance to military airfields and market access (population of neighboring counties normalized by the inverse of distance). The instrument is the distance to the main military U.S. and Taiwanese airfields penalized by the proximity to U.S.S.R. and North Korean airfields. *GDP p.c.* is the county's GDP per capita in 1982; it is expressed in logarithms. *Illiteracy* is the illiteracy rate in the county (in percentage points). *Male-to-female* is the number of males for 100 females. The last two dependent variables are measured in the 1982 Census. The F-statistic displayed is the Sanderson-Windmeijer F-statistic.

Table III.6 – Factories and GDP during the transition (1982–2013).

VARIABLES	Sat. (92)	$\Delta_{92-13}$ Sat.	GDP (82)	$\Delta_{82-00}$ GDP
Factory	1.822*** (.498)	-.524*** (.195)	2.444*** (.405)	-.611 (.434)
Observations	220	220	220	214
F-Stat (first stage)	21.67	21.67	21.67	23.25
Controls	Yes	Yes	Yes	Yes

Notes: Standard errors are reported between parentheses and clustered at the level of a rectangle of 3 latitude degrees  $\times$  3 longitude degrees. The unit of observation is a county (Administrative level 3). Controls include population in 1953 (log), county area (log), proximity to major ports (through the river network), proximity to Ming-dynasty courier stations, distance to the coast, distance to military airfields and market access (population of neighboring counties normalized by the inverse of distance). The instrument is the distance to the main military U.S. and Taiwanese airfields penalized by the proximity to U.S.S.R. and North Korean airfields. The F-statistic displayed is the Sanderson-Windmeijer F-statistic. *Sat. (92)* is the logarithm of night-time luminosity measured by satellite in 1992.  $\Delta_{92-13}$  *Sat.* is the log-difference in night-time luminosity between 1992 and 2013. *GDP (82)* is the logarithm of local GDP in 1982.  $\Delta_{82-00}$  *GDP* is the log-difference in GDP between 1982 and 2000.

Table III.7 – Factories and structural change during the transition (1994–2013).

VARIABLES	Primary sector share		Secondary sector share		Tertiary sector share	
	(94-00)	(01-13)	(94-00)	(01-13)	(94-00)	(01-13)
Factory	-.003 (.043)	-.042 (.056)	.049 (.054)	.198** (.091)	-.046 (.032)	-.156*** (.058)
Observations	1,690	3,584	1,690	3,611	1,690	3,611
F-Stat. (first stage)	10.42	6.88	10.42	6.74	10.42	6.74
Fixed effects (prefecture)	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects (year)	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors are reported between parentheses and clustered at the level of a rectangle latitude-degree  $\times$  longitude-degree. The unit of observation is a county (Administrative level 3). Controls include population in 1953 (log), county area, distance to coal mines, ore fields, oil and gas fields (through the 1948 railway network), distance to the major ports (through the river network), distance to railways, distance to 1900 cities, distance to Ming-dynasty courier stations, distance to the coast, distance to civilian airfields, distance to military airfields, market access (population of neighboring counties normalized by the inverse of distance), mean distance to other counties' centroids, mean slope, mean elevation, standard deviation of elevation, longitude, expected yields for 6 major crops, soil characteristics and prefecture fixed effects (Administrative level 2). The instrument is the distance to the main military U.S. and Taiwanese airfields penalized by the proximity to U.S.S.R. and North Korean airfields. The dependent variables are the shares of employment in the *Primary sector*, *Secondary sector* and *Tertiary sector* in the specified time period. The unit of observation is a prefecture $\times$ year.

Table III.8 – Factories and wages during the transition (1994–2013).

VARIABLES	Average wages	
	(94-00)	(01-13)
Factory	2,564.4*** (835.6)	9,636.0* (5,387.3)
Observations	1,690	3,597
F-Stat. (first stage)	10.42	6.93
Fixed effects (prefecture)	Yes	Yes
Fixed effects (year)	Yes	Yes
Controls	Yes	Yes

Notes: Standard errors are reported between parentheses and clustered at the level of a rectangle latitude-degree  $\times$  longitude-degree. The unit of observation is a county (Administrative level 3). Controls include population in 1953 (log), county area, distance to coal mines, ore fields, oil and gas fields (through the 1948 railway network), distance to the major ports (through the river network), distance to railways, distance to 1900 cities, distance to Ming-dynasty courier stations, distance to the coast, distance to civilian airfields, distance to military airfields, market access (population of neighboring counties normalized by the inverse of distance), mean distance to other counties' centroids, mean slope, mean elevation, standard deviation of elevation, longitude, expected yields for 6 major crops, soil characteristics and prefecture fixed effects (Administrative level 2). The instrument is the distance to the main military U.S. and Taiwanese airfields penalized by the proximity to U.S.S.R. and North Korean airfields. The dependent variable is the *Average wage rate* in the prefecture in the specified time period (1994–2000 or 2001–2013). The unit of observation is a prefecture $\times$ year.

Table III.9 – Factories and firm concentration during the transition (1993–2007).

VARIABLES	Employment HHI		Output HHI	
	(93-00)	(01-07)	(93-00)	(01-07)
Factory	.1908 (.2787)	.3081 (.2199)	.4583* (.2687)	.4469** (.2049)
Observations	998	1,005	998	1,005
F-Stat. (first stage)	10.22	10.11	10.22	10.11
Fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Notes: Standard errors are reported between parentheses and clustered at the level of a rectangle latitude-degree  $\times$  longitude-degree. Each cell is the result of a separate regression. The unit of observation is a county (Administrative level 3). Controls include population in 1953 (log), county area, distance to coal mines, ore fields, oil and gas fields (through the 1948 railway network), distance to the major ports (through the river network), distance to railways, distance to 1900 cities, distance to Ming dynasty courier stations, distance to the coast, distance to civilian airfields, distance to military airfields, market access (population of neighboring counties normalized by the inverse of distance), mean distance to other counties' centroids, mean slope, mean elevation, standard deviation of elevation, longitude, expected yields for 6 major crops, soil characteristics and prefecture fixed effects (Administrative level 2). The instrument is the distance to the main military U.S. and Taiwanese airfields penalized by the proximity to U.S.S.R. and North Korean airfields. *HHI* (1993–2000 and 2001–2007) is the average of the yearly Herfindahl-Hirschman indices constructed using employment (Columns 1 and 2) or output (Columns 3 and 4) across firms of a same county:  $HHI_c = \sum_{f \in d} \left( \frac{y_f}{\sum_{f \in d} y_f} \right)^2$ . The unit of observation is a county  $\times$  year.

Table III.10 – Factories and pollution (2003–2013).

VARIABLES	Sulfur	Waste	Smoke
	dioxide	water	and dust
Factory	6.820*** (1.919)	4.821 (3.642)	2.784** (1.377)
Observations	3,050	3,055	3,051
F-Stat. (first stage)	6.854	6.936	6.935
Fixed effects (prefecture)	Yes	Yes	Yes
Fixed effects (year)	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Notes: Standard errors are reported between parentheses and clustered at the level of a rectangle latitude-degree  $\times$  longitude-degree. Each cell is the result of a separate regression. The unit of observation is a county (Administrative level 3). Controls include population in 1953 (log), distance to coal mines, ore fields, oil and gas fields (through the 1948 railway network), distance to the major ports (through the river network), distance to railways, distance to 1900 cities, distance to Ming-dynasty courier stations, distance to the coast, distance to civilian airfields, distance to military airfields, market access (population of neighboring counties normalized by the inverse of distance), mean distance to other counties' centroids, mean slope, mean elevation, standard deviation of elevation, longitude, expected yields for 6 major crops, soil characteristics and prefecture fixed effects (Administrative level 2). The instrument is the distance to the main military U.S. and Taiwanese airfields penalized by the proximity to U.S.S.R. and North Korean airfields. *Sulfur dioxide* refers to total industrial SO<sub>2</sub> emissions (in tons per square kilometer). *Waste water* captures total industrial discharge of liquid waste (10,000 tons/km<sup>2</sup>). *Smoke and dust* refers to total industrial emissions of smoke and dust (in tons/km<sup>2</sup>). The unit of observation is a prefecture  $\times$  year.

## ONLINE APPENDIX

### III.A Firm comparison within treated places

To shed light on the rise and fall of the counties treated under the “156” program, it is instructive to investigate what happened to the firms that originate from the “156” plants and compare them with other firms in the same area. This will help us understand the influence of the “156” firms themselves on other, neighboring firms and thus delineate mechanisms behind the observed rise-and-fall pattern: Are firms descended from the “156” factories themselves still running and thriving? Or did they drag other firms down with them?

Table III.A.1 focuses on prefectures that received at least one of the Million-Rouble plants that we could match in the annual “above-scale” survey of firms. Currently, 72 factories could be identified in the firm survey.<sup>31</sup> We temporarily focus on prefectures instead of counties in this sub-section as county identifiers are not always readily available from the firm survey data and would have to be inferred from more detailed addresses.

The firms descended from “156” plants appear to significantly differ from other firms within treated prefectures. The differences are economically large. First, they exhibit much higher employment and compensation rates (all in logarithms). This result also holds when looking at wages proper instead. Compensation, which includes housing subsidies, pension and medical insurance, and welfare payable, is 53% higher in the “156”-program firms. Part of the higher wages is probably explained by a significantly higher share of high-skilled (i.e., college-educated) employees; however, the result on compensation is robust to adding the share of high-skilled workers as a control (not displayed).<sup>32</sup> These firms also clearly stand out in terms of size: They are 38 times as large as the average non-“156” firm, keeping prefecture, year and industry constant.

Second, when focusing on profits and value added per worker as depen-

---

<sup>31</sup>The tracking of the “156” factories is still work in progress and will be completed thanks to extensive data collection on firm histories. The comparison in Table III.A.1 was obtained thanks to fuzzy matching on firm names. Out of the 72 matched firms, there were 20 exact matches.

<sup>32</sup>The disaggregation of the workforce by educational attainment is available only for 2004 (year of the Economic Census, when additional variables were collected).

Table III.A.1 – Comparison of “156”-program and other manufacturing firms within treated prefectures.

	Employment (total)	Compensation (per worker)	Value added (per worker)	Profits (per worker)	High-skilled (share)
Factory	3.662 (.084)	.427 (.024)	-.230 (.051)	-.739 (.078)	.065 (.017)
Observations	387,679	273,221	209,763	190,592	24,578
Prefecture FEs	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes
Industry FEs	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors are reported between parentheses. All regressions are estimated with Ordinary Least Squares and include industry (2-digit CIC), prefecture and year fixed effects. The main explanatory variable, *Factory*, is a dummy equal to 1 if the firm was originally founded under the “156” program, and 0 otherwise. The sample is restricted to prefectures where at least one firm in the NBS annual “above-scale” surveys was identified as descended from one of the “156” factories. *Employment* is the natural logarithm of the number of workers in the firm (1992–2008, except for 1994 and 1997). *Compensation* (1996–2008, except for 1997) combines wages, housing subsidies, pension and medical insurance, and welfare payable. We take logarithms. *Value added* is available between 1998 and 2007, 2004 excluded. It is expressed in logarithms. *Profits*—also in logarithms—are defined as value added minus total compensation. They are available from 1998 to 2007, except for 2004. *High-skilled* is the share of college-educated employees in the firm’s work force. All monetary variables are deflated by province-level consumer price indices and expressed in per capita terms. The unit of observation is a firm×year.

dent variables (also in logarithms), we see that they are less profitable and less productive than the average firm in the same prefecture, which hints at inefficiencies. Inefficiencies may originate from an outdated production process or capital, poor management or from closer ties with local governments. Closer ties with the government cut both ways: On the one hand, they enable firms to access capital and enjoy protection as local governments can make it difficult for competitors to enter the market (by refusing authorizations or through red tape); on the other, they allow governments to exert pressure on firms to avoid layoffs and maintain living standards. Note that profits, value added and compensation are deflated in the Table by province-level consumer price indices to account for the fact that the cost of living may be higher in treated prefectures.

Table III.A.2 shows the share of the “treated” firms in the local economy. We see that in the period 1992–2008 they did not constitute a major share of employment, sales or profits in the prefectures where they were located. They represent 1.3% of manufacturing employment and 1.4% of revenue in that



sector.<sup>33</sup> This confirms that the fall of treated places observed above cannot be directly imputed to them. Besides, although they are less productive and profitable than the firms surrounding them, they nevertheless generate value, which should participate in higher GDP and more intense night-time luminosity.

The relative demise observed in Section 4 thus suggests that the differences noted across counties are not the direct effect of a rise and fall of the “156” factories but reflect the impact that those factories and the firms descended from them have had over time on local economies.

Table III.A.2 – Share of the “156” factories in local economies.

	Employment	Compensation	Revenue	Value added	Profits
Share	.013 (.089)	.004 (.034)	.014 (.094)	.014 (.122)	.015 (.220)
Observations	11,087	7,571	8,353	6,937	6,945

Notes: Standard deviations are reported between parentheses. The sample is restricted to prefectures where at least one firm in the NBS annual “above-scale” surveys was identified as descended from one of the “156” factories. For each variable, the table displays the share of such factories, e.g., *Employment* is the share of those factories in local manufacturing employment (1992–2008). *Revenue* refers to total sales. It is available in 1996–2007, except for 1997. *Compensation* (1996–2008, except for 1997) combines wages, housing subsidies, pension and medical insurance, and welfare payable. *Value added* is available between 1998 and 2007, 2004 excluded. *Profits* are defined as value added minus total compensation. They are available from 1998 to 2007, except for 2004. All monetary variables are deflated by province-level consumer price indices and expressed in per capita terms. The unit of observation is a prefecture×year×industry (2-digit Chinese Industrial Classification).

<sup>33</sup>Since not all “156” factories have been matched to firms in the “above-scale” data, these figures are lower bounds. However, there were on average three “156”-program factories per treated prefecture and in the subset of prefectures with at least one matched factory we have an average of 1.15. It therefore seems safe to conclude that they always represent a small share of the local economy.

## III.B Additional tables

Table III.B.1 – Definition of instrument and controls variables (1/2).

VARIABLES	Definition
<b><i>Population</i></b>	
Population (1953)	Total population of the county in the First Chinese Population Census (1953).
<b><i>Resource access controls</i></b>	
Travel cost to coal mines	Distance to coal mines following the 1948 railroad network.
Travel cost to ore	Distance to ore deposits following the 1948 railroad network.
Travel cost to oil/gas	Distance to oil or gas deposits following the 1948 railroad network.
<b><i>Topographic controls</i></b>	
Slope (100 degrees)	Average of the median slope of all cells in the county.
Strong slope	Dummy equal to 1 if the average slope in the county is greater than 10 degrees, and 0 otherwise.
Elevation (mean; m)	Average elevation in the county (in meters).
Elevation (st. dev.; m)	Standard deviation of elevation in the county (in meters).
<b><i>Market access controls</i></b>	
Proximity to ports	Dummy equal to 1 for a county whose centroid is lying within 500 km of a port following navigable waterways, and 0 otherwise.
Dist. to railways	Minimum distance to a railroad (1948 network).
Dist. to the coast	Minimum distance to the coast.
Market access	Market access indicator, equal to the sum of the populations (according to the 1953 Census) of all other Chinese counties normalized by the inverse of distance.
Proximity to courier stations	Dummy equal to 1 if the county centroid is located within 5 km of the closest Ming-dynasty courier station.
Proximity to 1900 city	Dummy equal to 1 if the county centroid is located within 5 km of the closest city as of 1900.
Prefecture capital	Dummy equal to 1 if the county belongs to the capital of the prefecture.
Province capital	Dummy equal to 1 if the county belongs to the capital of the province.
Dist. to other counties (km)	Average distance between the county's centroid and the centroids of all other Chinese counties.

Table III.B.2 – Definition of instrument and controls variables (2/2).

VARIABLES	Definition
<b><i>Geomorphic controls</i></b>	
Lake plain	Share of the county's area that consists of lacustrine plains.
Sand hills	Share of the county's area that consists of sand hills.
Tidal marsh	Share of the county's area that consists of tidal marshes.
<b><i>Agricultural controls</i></b>	
Expected yield: cotton	Average potential yield (kg/ha) of cotton under the high-input scenario (GAEZ model-based).
Expected yield: maize	Average potential yield (kg/ha) of maize under the high-input scenario (GAEZ model-based).
Expected yield: rice	Average potential yield (kg/ha) of rice under the high-input scenario (GAEZ model-based).
Expected yield: sugarcane	Average potential yield (kg/ha) of sugarcane under the high-input scenario (GAEZ model-based).
Expected yield: tea	Average potential yield (kg/ha) of tea under the high-input scenario (GAEZ model-based).
Expected yield: tobacco	Average potential yield (kg/ha) of tobacco under the high-input scenario (GAEZ model-based).
Expected yield: wheat	Average potential yield (kg/ha) of wheat under the high-input scenario (GAEZ model-based).
<b><i>Other geographic controls</i></b>	
Area	Total land area of the county.
Longitude	Longitude of the centroid of the county.
Dist. to civilian airfields	Minimum distance to a Chinese civilian airfield.
Dist. to military airfields	Minimum distance to a Chinese military airfield.

Table III.B.3 – Descriptive statistics - Full list of controls.

VARIABLES	Mean	Std dev.	With factory	No factory
Observations	2,408		111	2,297
<i>Population</i>				
Population (1953, log)	12.20	1.01	12.41	12.19
<i>Resource access controls</i>				
Travel cost to coal mines (log)	14.06	1.63	12.68	14.12
Travel cost to ore (log)	13.97	1.31	13.07	14.01
Travel cost to oil/gas (log)	14.10	1.58	13.39	14.13
<i>Topographic controls</i>				
Slope (100 degrees)	0.024	0.024	0.017	0.025
Strong slope	0.625	0.484	0.595	0.626
Elevation (mean; m)	846.1	1096	546.0	860.6
Elevation (st. dev.; m)	192.4	204.0	155.0	194.2
<i>Market access controls</i>				
Proximity to ports	0.148	0.355	0.153	0.148
Dist. to railways (log)	10.80	2.04	9.05	10.89
Dist. to the coast (log)	12.57	1.73	12.84	12.56
Market access (log)	16.76	0.471	16.75	16.76
Proximity to courier stations	0.074	0.266	0.144	0.074
Proximity to 1900 city	0.467	0.499	0.802	0.451
Prefecture capital	0.078	0.268	0.108	0.077
Dist. to other counties (km)	13.82	4.794	14.09	13.81
<i>Geomorphic controls</i>				
Lake plain	0.436	0.417	0.485	0.433
Sand hills	0.382	0.379	0.441	0.380
Tidal marsh	0.113	0.244	0.037	0.116
<i>Agricultural controls</i>				
Expected yield: cotton	442.4	356.3	396.1	444.6
Expected yield: maize	7971.2	3874.7	8164.9	7961.8
Expected yield: rice	3031.1	3870.8	1585.5	3101.0
Expected yield: sugarcane	1818.5	3547.1	1065.5	1855.0
Expected yield: tea	266.4	407.1	120.0	273.5
Expected yield: tobacco	995.4	554.1	993.3	995.5
Expected yield: wheat	4848.4	2499.1	4626.0	4859.2
<i>Other geographic controls</i>				
Area (log)	7.64	0.957	7.24	7.66
Longitude	111.2	9.539	115.2	111.0
Dist. to civilian airfields (log)	10.71	1.00	10.14	10.73
Dist. to military airfields (log)	10.91	1.05	10.27	10.95

Table III.B.4 – Proximity to factories and employment in 1964, 1982 and 1990.

Panel A: 1964	Population		Share non-agr.	
	(OLS)	(2SLS)	(OLS)	(2SLS)
Proximity (20 kms)	.348*** (.085)	.431*** (.169)	.283*** (.045)	.573*** (.104)
Observations	220	220	220	220
F-Stat. (first stage)		21.44		21.44
Controls	Yes	Yes	Yes	Yes
Panel B: 1982	Population		Share non-agr.	
	(OLS)	(2SLS)	(OLS)	(2SLS)
Proximity (20 kms)	.464*** (.077)	.815*** (.225)	.289*** (.041)	.634*** (.101)
Observations	220	220	220	220
F-Stat. (first stage)		21.44		21.44
Controls	Yes	Yes	Yes	Yes
Panel C: 1990	Population		Share non-agr.	
	(OLS)	(2SLS)	(OLS)	(2SLS)
Proximity (20 kms)	.498*** (.095)	.802*** (.246)	.315*** (.049)	.648*** (.119)
Observations	220	220	220	220
F-Stat. (first stage)		21.44		21.44
Controls	Yes	Yes	Yes	Yes

Notes: Standard errors are reported between parentheses and clustered at the level of a rectangle of 3 latitude degrees  $\times$  3 longitude degrees. Each cell is the result of a separate regression. The unit of observation is a county (Administrative level 3). Controls include population in 1953 (log), county area (log), proximity to major ports (through the river network), proximity to Ming-dynasty courier stations, distance to the coast, distance to military airfields and market access (population of neighboring counties normalized by the inverse of distance). The instrument is the distance to the main military U.S. and Taiwanese airfields penalized by the proximity to U.S.S.R. and North Korean airfields. *Population* is the logarithm of total population in the county and *Share non-agr.* the share of the population that has a non-agricultural household registration (*hukou*). The F-statistic displayed is the Sanderson-Windmeijer F-statistic.

Table III.B.5 – List of U.S. Air Force bases (1/2).

Base name	Country	Operator
Ashiya Air Base	Japan	USAF
Chitose Air Base	Japan	USAF
Fuchu Air Base	Japan	USAF
Itazuke Air Base	Japan	USAF
Johnson Air Base	Japan	USAF
Komaki Air Base	Japan	USAF
Miho Air Base	Japan	USAF
Naha Air Base	Japan	USAF
Central Field, Iwo Jima	Japan	USAF
South Field, Iwo Jima	Japan	USAF
Tachikawa Air Base	Japan	USAF
Tsuiki Air Base	Japan	USAF
K-1 Pusan-West	Korea	USAF
K-2 Taegu (Taegu #1)	Korea	USAF
K-3 Pohang	Korea	USAF
K-4 Sachon	Korea	USAF
K-5 Taejon	Korea	USAF
K-6 Pyongtaek	Korea	USAF
K-7 Kwangju	Korea	USAF
K-8 Kunsan	Korea	USAF
K-9 Pusan-East	Korea	USAF
K-10 Chinhae	Korea	USAF
K-11 Urusan (Ulsan)	Korea	USAF
K-13 Suwon	Korea	USAF
K-14 Kimpo	Korea	USAF
K-15 Mokpo	Korea	USAF
K-16 Seoul	Korea	USAF
K-18 Kangnung (Koryo)	Korea	USAF
K-37 Taegu No. 2	Korea	USAF
K-39 Cheju-do No. 1	Korea	USAF
K-40 Cheju-do No. 2	Korea	USAF
K-46 Hoengsong	Korea	USAF
K-47 Chunchon	Korea	USAF
K-50 Sokcho-ri	Korea	USAF
K-55 Osan-ni	Korea	USAF
Clark Air Base	Philippines	USAF

Notes: "Korea" refers to South Korea. USAF stands for U.S. Air Force.

Table III.B.6 – List of U.S. Air Force bases (2/2).

Base name	Country	Operator
Chiayi Air Base	Taiwan	USTDC
Ching Chuan Kang Air Base	Taiwan	USTDC
Lo Shan Radar Site	Taiwan	USTDC
Makung Island	Taiwan	USTDC
O Laun Pi	Taiwan	USTDC
Shihmen Air Station	Taiwan	USTDC
Shu Linkou Air Station	Taiwan	USTDC
Sung Shan Air Base	Taiwan	USTDC
Tainan Air Base	Taiwan	USTDC
Taipei Air Station	Taiwan	USTDC
Takangshan	Taiwan	USTDC
Taoyuan Air Base	Taiwan	USTDC
Hsinchu Air Base	Taiwan	USTDC
Don Muang Royal Thai Air Force Base	Thailand	USAF
Korat Royal Thai Air Force Base	Thailand	USAF
Nakhon Phanom Royal Thai Navy Base	Thailand	USAF
Takhli Royal Thai Air Force Base	Thailand	USAF
U-Tapao Royal Thai Navy Airfield	Thailand	USAF
Ubon Royal Thai Air Force Base	Thailand	USAF
Udorn Royal Thai Air Force Base	Thailand	USAF
Bien Hoa Air Base	Vietnam	USAF
Binh Thuy Air Base	Vietnam	USAF
Cam Ranh Air Base	Vietnam	USAF
Da Nang Air Base	Vietnam	USAF
Nha Trang Air Base	Vietnam	USAF
Phu Cat Air Base	Vietnam	USAF
Pleiku Air Base	Vietnam	USAF
Phan Rang Air Base	Vietnam	USAF
Tan Son Nhut Air Base	Vietnam	USAF
Tuy Hoa Air Base	Vietnam	USAF

Notes: "Vietnam" refers to South Vietnam. USAF stands for U.S. Air Force, and USTDC for United States Taiwan Defense Command.

Table III.B.7 – List of U.S.S.R. and North Korean airbases.

Base name	Country
Yelizovo	U.S.S.R.
Sokol	U.S.S.R.
Khomutovo	U.S.S.R.
Novy	U.S.S.R.
Knevichi	U.S.S.R.
Yakutsk	U.S.S.R.
Tiksi	U.S.S.R.
Ignatyev	U.S.S.R.
Chulman	U.S.S.R.
Mirny	U.S.S.R.
Kadala	U.S.S.R.
Mukhino	U.S.S.R.
Irkutsk	U.S.S.R.
Khatanga	U.S.S.R.
Bratsk	U.S.S.R.
K-28 Hamhung West	Korea
K-29 Sinanju	Korea
K-30 Sinuiju	Korea
K-31 Kilchu (Kisshu)	Korea
K-32 Oesichon-dong	Korea
K-35 Hoeryong (Kainsei)	Korea
K-36 Kanggye	Korea
K-38 Wonju AB	Korea
K-41 Chungju	Korea
K-42 Andong No. 2	Korea
K-43 Kyongju	Korea
K-44 Changhowon-ni	Korea
K-45 Yoju	Korea
K-48 Iri	Korea
K-49 Yangsu-ri	Korea
K-51 Inje	Korea
K-52 Yanggu	Korea
K-57 Kwangju	Korea

Notes: "Korea" refers to North Korea.



# Bibliography

- Accetturo, Antonio, Matteo Bugamelli, and Andrea Roberto Lamorgese**, “Welcome to the machine: firms’ reaction to low-skilled immigration,” *Bank of Italy Working Paper*, 2012, 846. (Cited on page 85.)
- Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz**, “Fragility of asymptotic agreement under Bayesian learning,” *Theoretical Economics*, 2016, 11 (1), 187–225. (Cited on pages 15, 23 and 25.)
- ADB**, “The Declining Share of Agricultural Employment in the People’s Republic of China: How Fast?,” ADB Economics Working Paper Series 419, Asian Development Bank (ADB) November 2014. (Cited on page 82.)
- Alvarez-Cuadrado, Francisco and Markus Poschke**, “Structural Change Out of Agriculture: Labor Push versus Labor Pull,” *American Economic Journal: Macroeconomics*, July 2011, 3 (3), 127–58. (Cited on pages 1, 82 and 86.)
- Ambler, Kate**, “Don’t tell on me: Experimental evidence of asymmetric information in transnational households,” *Journal of Development Economics*, 2015, 113, 52–69. (Cited on pages 14 and 72.)
- Angelucci, Manuela**, “Migration and Financial Constraints: Evidence from Mexico,” *The Review of Economics and Statistics*, March 2015, 97 (1), 224–228. (Cited on page 86.)
- Aparicio, Francisco Javier and Covadonga Meseguer**, “Collective remittances and the state: the 3 × 1 program in Mexican municipalities,” *World Development*, 2012, 40 (1), 206–222. (Cited on page 44.)
- Ashraf, Nava**, “Spousal Control and Intra-household Decision Making: An Experimental Study in the Philippines,” *The American Economic Review*, 2009, 99 (4), 1245–1277. (Cited on page 15.)
- , **Diego Aycinena, Claudia A. Martínez, and Dean Yang**, “Savings in Transnational Households: A Field Experiment among Migrants from

- El Salvador,” *Review of Economics and Statistics*, 2015, 97 (2), 332–351. (Cited on pages 14 and 16.)
- Au, Chun-Chung and J. Vernon Henderson**, “How migration restrictions limit agglomeration and productivity in China,” *Journal of Development Economics*, 2006, 80, 350–88. (Cited on page 86.)
- Aumann, Robert J.**, “Agreeing to Disagree,” *The Annals of Statistics*, 1976, 4 (6), 1236–1239. (Cited on page 22.)
- Azam, Jean-Paul and Flore Gubert**, “Those in Kayes,” *Revue économique*, 2005, 56 (6), 1331–1358. (Cited on pages 14 and 16.)
- Banerjee, Abhijit V. and Esther Duflo**, “Growth Theory through the Lens of Development Economics,” *Handbook of Economic Growth*, 2005, 1, 473–552. (Cited on pages 2 and 15.)
- and **Kaivan Munshi**, “How Efficiently is Capital Allocated? Evidence from the Knitted Garment Industry in Tirupur,” *The Review of Economic Studies*, 2004, 71 (1), 19–42. (Cited on page 16.)
- Batista, Cátia and Gaia Narciso**, “Migrant Remittances and Information Flows: Evidence from a Field Experiment,” 2016. (Cited on pages 14 and 16.)
- , **Dan Silverman, and Dean Yang**, “Directed giving: Evidence from an inter-household transfer experiment,” *Journal of Economic Behavior & Organization*, 2015, 118, 2–21. (Cited on page 14.)
- Bazzi, Samuel**, “Wealth Heterogeneity and the Income Elasticity of Migration,” *American Economic Journal: Applied Economics*, 2016. (Cited on pages 86 and 94.)
- Blanchard, Olivier**, *The Economics of Post-Communist Transition*, Oxford: Clarendon press, 1997. (Cited on page 165.)
- Bleakley, Hoyt and Jeffrey Lin**, “Portage and Path Dependence,” *The Quarterly Journal of Economics*, 2012, 127 (2), 587–644. (Cited on page 165.)

- Bo, Yibo**, *Review of Several Major Decisions and Events [Ruogan Zhongda Juece yu Shijian de Huigu]*, Beijing: The Chinese Communist Party School Press [Zhonggong Zhongyang Dangxiao Chuban She], 1991. (Cited on pages 162, 167, 168, 171, 175, 176, 180 and 183.)
- Boltz-Laemmel, Marie and Paola Villar**, “Redistribution au sein de la famille étendue au Sénégal : Le rôle des migrants internes et internationaux,” *Document de travail (Docweb) CEPREMAP*, January 2014, (1404). (Cited on pages 17, 70 and 71.)
- Borjas, George J.**, “The Labor Demand Curve Is Downward Sloping: Re-examining The Impact Of Immigration On The Labor Market,” *The Quarterly Journal of Economics*, November 2003, 118 (4), 1335–1374. (Cited on pages 87 and 102.)
- , **Richard B. Freeman, Lawrence F. Katz, John DiNardo, and John M. Abowd**, “How Much Do Immigration and Trade Affect Labor Market Outcomes?,” *Brookings Papers on Economic Activity*, 1997, 1997 (1), 1–90. (Cited on pages 2 and 87.)
- Bosker, Maarten, Steven Brakman, Harry Garretsen, and Marc Schramm**, “Relaxing Hukou: Increased labor mobility and China’s economic geography,” *Journal of Urban Economics*, 2012, 72, 252–66. (Cited on page 86.)
- Bound, John and Alan B. Krueger**, “The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?,” *Journal of Labor Economics*, 1991, 9 (1). (Cited on pages 35 and 76.)
- Boustan, Leah Platt, Price V. Fishback, and Shawn Kantor**, “The Effect of Internal Migration on Local Labor Markets: American Cities during the Great Depression,” *Journal of Labor Economics*, October 2010, 28 (4), 719–746. (Cited on pages 83, 87 and 92.)
- Boyreau-Debray, Genevieve and Shang-Jin Wei**, “Pitfalls of a State-Dominated Financial System: The Case of China,” NBER Working Papers 11214, National Bureau of Economic Research, Inc March 2005. (Cited on page 186.)

- Brandt, Loren, Gueorgui Kambourov, and Kjetil Storesletten**, “Firm Entry and Regional Growth Disparities: the Effect of SOEs in China,” Technical Report May 2016. (Cited on pages 165 and 166.)
- , **Johannes Van Biesebroeck, and Yifan Zhang**, “Challenges of working with the Chinese NBS firm-level data,” *China Economic Review*, 2014, 30 (C), 339–352. (Cited on pages 96, 97, 145, 146, 148, 150 and 172.)
- , **Trevor Tombe, and Xiaodong Zhu**, “Factor market distortions across time, space and sectors in China,” *Review of Economic Dynamics*, 2013, 16 (1), 39–58. (Cited on pages 85 and 165.)
- Bryan, Gharad and Melanie Morten**, “Economic Development and the Spatial Allocation of Labor: Evidence from Indonesia,” *Manuscript, London School of Economics and Stanford University*, 2015. (Cited on pages 1 and 86.)
- Buera, Francisco J. and Yongseok Shin**, “Financial Frictions and the Persistence of History: A Quantitative Exploration,” *Journal of Political Economy*, 2013, 121 (2), 221–272. (Cited on pages 165, 186 and 187.)
- Buera, Francisco J, Joseph P Kaboski, and Yongseok Shin**, “Finance and development: A tale of two sectors,” *The American Economic Review*, 2011, 101 (5), 1964–2002. (Cited on page 85.)
- Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli**, “Agricultural productivity and structural transformation: evidence from Brazil,” 2016, (106), 1320–1365. (Cited on page 86.)
- Card, David**, “Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration,” *Journal of Labor Economics*, 2001, 19 (1), 22–64. (Cited on pages 2 and 87.)
- Castilla, Carolina and Thomas Walker**, “Is Ignorance Bliss? The Effect of Asymmetric Information between Spouses on Intra-Household Allocations,” *The American Economic Review*, 2013, 103 (3), 263–268. (Cited on page 15.)

- Chami, Ralph, Connel Fullenkamp, and Samir Jahjah**, “Are Immigrant Remittance Flows a Source of Capital for Development?,” *IMF Staff Papers*, 2005, 52 (1), 55–81. (Cited on pages 14 and 16.)
- Chan, Kam Wing**, “Misconceptions and Complexities in the Study of China’s Cities: Definitions, Statistics, and Implications,” *Eurasian Geography and Economics*, 2007, 48 (4), 383–412. (Cited on page 92.)
- , “The Chinese hukou system at 50,” *Eurasian geography and economics*, 2009, 50 (2), 197–221. (Cited on page 121.)
- , “The household registration system and migrant labor in China: Notes on a debate,” *Population and Development Review*, 2010, 36 (2), 357–364. (Cited on page 121.)
- , “Crossing the 50 Percent Population Rubicon: Can China Urbanize to Prosperity?,” *Eurasian Geography and Economics*, 2012, 53 (1), 63–86. (Cited on page 121.)
- , “Migration and development in China: trends, geography and current issues,” *Migration and Development*, 2012, 1 (2), 187–205. (Cited on pages 2, 82 and 123.)
- **and Will Buckingham**, “Is China abolishing the hukou system?,” *The China Quarterly*, 2008, 195, 582–606. (Cited on page 121.)
- Chen, Joyce J.**, “Identifying Non-Cooperative Behavior Among Spouses: Child Outcomes in Migrant-Sending Households,” 2006. (Cited on page 14.)
- , “Identifying Non-Cooperative Behavior Among Spouses: Child Outcomes in Migrant-Sending Households,” *Journal of Development Economics*, 2013, 100, 1–18. (Cited on page 14.)
- Cheng, Wenya and Andrei Potlogea**, “Trade Liberalization and Economic Development: Evidence from China’s WTO Accession,” 2017. Manuscript. (Cited on page 86.)
- Chin, Aimee, Leonie Karkoviata, and Nathaniel Wilcox**, “Impact of Bank Accounts on Migrant Savings and Remittances: Evidence

from a Field Experiment,” [http://www.uh.edu/~achin/research/ckw\\_banking\\_june2011.pdf](http://www.uh.edu/~achin/research/ckw_banking_june2011.pdf) 2011. (Cited on page 14.)

**Chinitz, Benjamin**, “Contrasts in Agglomeration: New York and Pittsburgh,” *The American Economic Review*, 1961, 51 (2), 279–289. (Cited on pages 164 and 192.)

**Chort, Isabelle and Jean-Noël Senne**, “Intra-household Selection into Migration: Evidence from a Matched Sample of Migrants and Origin Households in Senegal,” *Document de travail UMR DIAL*, 2013. (Cited on pages 26 and 27.)

—, **Flore Gubert, and Jean-Noël Senne**, “Migrant networks as a basis for social control: Remittance incentives among Senegalese in France and Italy,” *Regional Science and Urban Economics*, 2012. (Cited on pages 14, 26 and 72.)

**Combes, Pierre-Philippe, Sylvie Démurger, and Shi Li**, “Migration externalities in Chinese cities,” *European Economic Review*, 2015, 76 (C), 152–167. (Cited on pages 87 and 88.)

**De Laat, Joost**, “Household allocations and endogenous information: The case of split migrants in Kenya,” *Journal of Development Economics*, 2014, 106, 108–117. (Cited on page 14.)

**De Mel, Suresh, David McKenzie, and Christopher Woodruff**, “Returns to Capital in Microenterprises: Evidence from a Field Experiment,” *The Quarterly Journal of Economics*, 2008, 123 (4), 1329–1372. (Cited on page 15.)

**De Vreyer, Philippe, Sylvie Lambert, and Abla Safir**, “Remittances and Poverty: Who benefits in the Household?,” *Paris School of Economics mimeograph*, 2008. (Cited on pages 26, 27 and 70.)

**De Weerd, Joachim, Garance Genicot, and Alice Mesnard**, “Asymmetry of Information within Family Networks,” Working Paper 21685, National Bureau of Economic Research October 2015. (Cited on page 14.)

- Démurger, Sylvie, Marc Gurgand, Shi Li, and Ximing Yue**, “Migrants as second-class workers in urban China? A decomposition analysis,” *Journal of Comparative Economics*, December 2009, 37 (4), 610–628. (Cited on pages 97, 120 and 121.)
- Dercon, Stefan and Pramila Krishnan**, “In Sickness and in Health: Risk Sharing within Households in Rural Ethiopia,” *Journal of Political Economy*, 2000, 108 (4), 688–727. (Cited on page 14.)
- Dia, Hamidou**, “Les investissements des migrants dans la vallée du fleuve Sénégal: confiance et conflits d’intérêts,” *Revue européenne des migrations internationales*, 2007, 23 (3), 29–49. (Cited on pages 17, 30, 42, 70 and 72.)
- **and Moumouni Adamou**, *Les transferts d’argent des migrants de la vallée du fleuve Sénégal. Processus d’affectation et impact sur la région d’origine* Migrations internationales, recompositions territoriales et développement. Document de synthèse des projets du Programme FSP 2003–74, Paris: Institut de recherche pour le développement, (Cited on pages 17 and 71.)
- Dong, Zhikai**, “On the Establishment of the “156 Program” [Guanyu “156 Xiang” de Queli],” *Researches in Chinese Economic History [Zhongguo Jingji Shi Yanjiu]*, 1999, 4, 93–107. (Cited on page 166.)
- **and Jiang Wu**, *Industry Cornerstone of New China [Xin Zhongguo Gongye de Dianjishi – 156 Xiang Jianshe Yanjiu (1950–2000)]*, Guangzhou: Guangdong Economy Publishing House, 2004. (Cited on pages 162, 167 and 171.)
- Dubois, Pierre and Ethan A Ligon**, “Incentives and nutrition for rotten kids: intrahousehold food allocation in the Philippines,” *CUDARE Working Paper 1114*, 2011. (Cited on page 14.)
- Dustmann, Christian and Albrecht Glitz**, “How Do Industries and Firms Respond to Changes in Local Labor Supply?,” *Journal of Labor Economics*, 2015, 33 (3), 711–750. (Cited on page 85.)

- Ebenstein, Avraham and Yaohui Zhao**, “Tracking rural-to-urban migration in China: Lessons from the 2005 inter-census population survey,” *Population Studies*, 2015, 69 (3), 337–353. (Cited on pages 122 and 172.)
- El Badaoui, Eliane, Eric Strobl, and Frank Walsh**, “The Impact of Internal Migration on Local Labour Markets in Thailand,” *EconomiX Working Papers 2014-12*, University of Paris West - Nanterre la Défense, EconomiX 2014. (Cited on page 87.)
- Eyster, Erik and Matthew Rabin**, “Cursed Equilibrium,” *Econometrica*, 2005, 73 (5), 1623–1672. (Cited on page 24.)
- Facchini, Giovanni, Maggie Y. Liu, Anna Maria Mayda, and Minghai Zhou**, “The impact of China’s WTO accession on internal migration,” December 2015. (Cited on pages 86 and 88.)
- Falck, Oliver, Christina Guenther, Stephan Heblich, and William R. Kerr**, “From Russia with love: the impact of relocated firms on incumbent survival,” *Journal of Economic Geography*, May 2013, 13 (3), 419–449. (Cited on page 164.)
- Fall, P.D.**, “Migration internationale et droits des travailleurs au Sénégal,” *Série UNESCO: Rapports par pays sur la ratification de la Convention des Nations Unies sur les droits des migrants*, 2003. (Cited on pages 17, 70, 71 and 72.)
- Fan, Cindy C.**, *China on the Move*, London: Routledge, 2008. (Cited on pages 123 and 131.)
- Feng, Shuaizhang, Michael Oppenheimer, and Wolfram Schlenker**, “Weather Anomalies, Crop Yields, and Migration in the US Corn Belt,” March 2015. (Cited on page 87.)
- , **Yingyao Hu, and Robert Moffit**, “Long Run Trends in Unemployment and Labor Force Participation in China,” August 2015. (Cited on pages 150 and 151.)



- Fogel, Robert**, *Railroads and American economic growth: Essays in econometric history*, Baltimore, MD: Johns Hopkins University Press, 1964. (Cited on page 175.)
- Fresia, Marion**, “Des “réfugiés-migrants” : les parcours d’exil de réfugiés mauritaniens au Sénégal,” *New Issues in Refugee Research*, 2006, *Research Paper No. 135*. (Cited on pages 74 and 76.)
- Ge, Suqin and Dennis Tao Yang**, “Changes In China’s Wage Structure,” *Journal of the European Economic Association*, 04 2014, 12 (2), 300–336. (Cited on pages 87, 150 and 151.)
- Gerschenkron, Alexander**, *Economic Backwardness in Historical Perspective: A Book of Essays*, Cambridge, MA: Belknap Press of Harvard University Press, 1962. (Cited on page 167.)
- Giesing, Yvonne and Nadzeya Laurentsyeva**, “Brain Drain and Firm Productivity: Evidence from the Sequential Opening of EU Labour Markets,” 2015. Unpublished. (Cited on page 85.)
- Glaeser, Edward L. and Janet E. Kohlhase**, “Cities, regions and the decline of transport costs,” *Papers in Regional Science*, 2004, 83 (1), 197–228. (Cited on page 175.)
- **and Joshua D. Gottlieb**, “The Economics of Place-Making Policies,” Technical Report, National Bureau of Economic Research 2008. (Cited on page 163.)
- , **Sari Pekkala Kerr, and William R. Kerr**, “Entrepreneurship and urban growth: An empirical assessment with historical mines,” *Review of Economics and Statistics*, 2015, 97 (2), 498–520. (Cited on pages 164, 165, 172 and 192.)
- Goldstein, Markus, Alain de Janvry, and Elisabeth Sadoulet**, *Is a Friend in Need a Friend Indeed? Inclusion and Exclusion in Mutual Insurance Networks in Southern Ghana Insurance Against Poverty*, Oxford: Oxford University Press, 2005. (Cited on page 14.)

- Gollin, Douglas, David Lagakos, and Michael E. Waugh**, “The Agricultural Productivity Gap,” *The Quarterly Journal of Economics*, 2014, *129* (2), 939–993. (Cited on page 1.)
- , **Stephen Parente, and Richard Rogerson**, “The Role of Agriculture in Development,” *The American Economic Review*, May 2002, *92* (2), 160–164. (Cited on page 86.)
- Goncharenko, Sergei**, *Beijing and Moscow: From Allies to Enemies [Beijing yu Mosike: Cong Lianmeng Zou Xiang Duikang]*, Nanning: Guangxi Normal University Press [Guangxi Shifan Daxue Chuban She], 2002. (Cited on page 167.)
- Greenstone, Michael, Richard Hornbeck, and Enrico Moretti**, “Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings,” *Journal of Political Economy*, 2010, *118* (3), 536–598. (Cited on pages 162, 164 and 165.)
- Harris, John R. and Michael P. Todaro**, “Migration, Unemployment and Development: A Two-Sector Analysis,” *The American Economic Review*, 1970, *60* (1), pp. 126–142. (Cited on pages 5 and 82.)
- He, Yimin and Mingchang Zhou**, “The 156 Program and the Development of Industrial Cities in New China (1949–1957) [156 Xiang Gongcheng yu Xin Zhongguo Gongye Chengshi Fazhan (1949–1957)],” *Contemporary China History Studies [Dangdai Zhongguo Shi Yanjiu]*, 2007, *14* (2), 70–77. (Cited on page 167.)
- Herrendorf, Berthold, Christopher Herrington, and Akos Valentinyi**, “Sectoral Technology and Structural Transformation,” Technical Report 9386, C.E.P.R. Discussion Papers March 2013. (Cited on page 85.)
- Hirschman, Albert O.**, *The Strategy of Economic Development*, New Haven, CT: Yale University Press, 1958. (Cited on page 164.)
- Hodrick, R. and E. Prescott**, “Post-war U.S. business cycles: An empirical investigation,” *Journal of Money, Credit and Banking*, 1997, *29* (1), 1–16. (Cited on page 139.)

- Hsieh, Chang-Tai and Peter J. Klenow**, “Misallocation and Manufacturing TFP in China and India,” *The Quarterly Journal of Economics*, November 2009, *124* (4), 1403–1448. (Cited on pages 6, 82, 85, 97, 99 and 165.)
- **and Zheng Song**, “Grasp the Large, Let Go of the Small: The Transformation of the State Sector in China,” NBER Working Papers 21006, National Bureau of Economic Research, Inc March 2015. (Cited on pages 85, 86 and 165.)
- Imbert, Clement and John Papp**, “Short-term Migration, Rural Workfare Programs and Urban Labor Markets - Evidence from India,” The Warwick Economics Research Paper Series (TWERPS) 1116, University of Warwick, Department of Economics 2016. (Cited on page 87.)
- Jakiela, Pamela and Owen Ozier**, “Does Africa need a rotten kin theorem? Experimental evidence from village economies,” *The Review of Economic Studies*, 2015, *83* (1), 231–268. (Cited on page 14.)
- Kamenica, Emir and Matthew Gentzkow**, “Bayesian Persuasion,” *The American Economic Review*, October 2011, *101* (6), 2590–2615. (Cited on page 13.)
- Kerr, Sari Pekkala, William R. Kerr, and William F. Lincoln**, “Skilled Immigration and the Employment Structures of US Firms,” *Journal of Labor Economics*, 2015, *33* (S1), S147 – S186. (Cited on page 85.)
- Kleemans, Marieke and Jeremy Magruder**, “Labor Market Changes in Response to Immigration: Evidence from Internal Migration Driven by Weather Shocks in Indonesia,” *Economic Journal*, 2017. Manuscript. (Cited on pages 87 and 102.)
- Kline, Patrick and Enrico Moretti**, “Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority,” *The Quarterly Journal of Economics*, 2014, *129* (1), 275–331. (Cited on pages 161, 162, 164 and 175.)
- Kuznets, Simon**, *Agriculture in Economic Development*, New York, NY: McGraw-Hill Book Company, 1964. (Cited on pages 5 and 82.)

- Lewis, Arthur**, “Economic Development with Unlimited Supplies of Labour,” *The Manchester School*, 1954, 22 (2), 139–191. (Cited on pages 1, 5 and 82.)
- Li, Wei and Dennis Tao Yang**, “The Great Leap Forward: Anatomy of a Central Planning Disaster,” *Journal of Political Economy*, 2005, 113 (4), 840–877. (Cited on page 160.)
- Lin, Justin Yifu**, “Collectivization and China’s agricultural crisis in 1959–1961,” *Journal of Political Economy*, 1990, 98 (6), 1228–1252. (Cited on page 160.)
- Lucas, Robert E.**, “Why Doesn’t Capital Flow from Rich to Poor Countries?,” *The American Economic Review*, 1990, 80 (2), 92–96. (Cited on pages 2 and 15.)
- Lundberg, Shelly and Robert A. Pollak**, “Separate Spheres Bargaining and the Marriage Market,” *Journal of Political Economy*, 1993, 101 (6), 988–1010. (Cited on pages 3 and 14.)
- Lüthi, Lorenz M.**, *The Sino-Soviet split: Cold War in the communist world*, Princeton, NJ: Princeton University Press, 2010. (Cited on page 166.)
- Manser, Marilyn and Murray Brown**, “Marriage and Household Decision-Making: A Bargaining Analysis,” *International Economic Review*, 1980, pp. 31–44. (Cited on pages 3 and 14.)
- Marfaing, Laurence**, *Les Sénégalais en Allemagne : quotidien et stratégies de retour*, Paris: Karthala, 2003. (Cited on pages 17 and 71.)
- Mayneris, Florian, Sandra Poncet, and Tao Zhang**, “The cleansing effect of minimum wages. Minimum wages, firm dynamics and aggregate productivity in China,” Core Discussion Papers 2014/44 October 2014. (Cited on page 102.)
- , —, and —, “Improving or Disappearing: Firm-Level Adjustments to Minimum Wages in China,” IRES Discussion Paper 2016-27 2016. (Cited on page 166.)

- McElroy, Marjorie B and Mary Jean Horney**, “Nash-Bargained Household Decisions: Toward a Generalization of the Theory of Demand,” *International Economic Review*, 1981, pp. 333–349. (Cited on page 14.)
- McKenzie, David, John Gibson, and Stephen Stillman**, “A land of milk and honey with streets paved with gold: Do emigrants have over-optimistic expectations about incomes abroad?,” *Journal of Development Economics*, 2013, *102*, 106–127. (Cited on pages 15 and 72.)
- Meng, Xin and Dandan Zhang**, “Labour Market Impact of Large Scale Internal Migration on Chinese Urban ‘Native’ Workers,” IZA Discussion Papers 5288 October 2010. (Cited on pages 86, 87 and 88.)
- Midrigan, Virgiliu and Daniel Yi Xu**, “Finance and misallocation: Evidence from plant-level data,” *The American Economic Review*, 2014, *104* (2), 422–458. (Cited on page 85.)
- Moll, Benjamin**, “Productivity losses from financial frictions: can self-financing undo capital misallocation?,” *The American Economic Review*, 2014, *104* (10), 3186–3221. (Cited on page 85.)
- Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny**, “Industrialization and the Big Push,” *Journal of Political Economy*, 1989, *97* (5), 1003–1026. (Cited on page 164.)
- Naughton, Barry**, *The Chinese Economy: Transition and Growth*, Cambridge, MA: MIT Press, 2007. (Cited on page 105.)
- Neumark, David and Helen D. Simpson**, “The Handbook of Regional and Urban Economics,” in Gilles Duranton, J.V. Henderson, and William Strange, eds., *The Handbook of Regional and Urban Economics*, Cambridge, MA: Elsevier/Academic Press, 2015. (Cited on page 163.)
- Oberfield, Ezra and Devesh Raval**, “Micro data and macro technology,” Technical Report, National Bureau of Economic Research 2014. (Cited on page 99.)
- Olney, William W**, “Immigration and firm expansion,” *Journal of Regional Science*, 2013, *53* (1), 142–157. (Cited on page 85.)

- Park, Albert**, *China Urbanizes: Consequences, Strategies, and Policies*, Washington, D.C.: The World Bank, 2008. (Cited on page 151.)
- Peri, Giovanni**, “The Effect Of Immigration On Productivity: Evidence From U.S. States,” *The Review of Economics and Statistics*, February 2012, 94 (1), 348–358. (Cited on page 85.)
- Perkins, D.H. and T.G. Rawski**, “Forecasting China’s economic growth,” in L. Brandt and T.G. Rawski, eds., *China’s great economic transformation*, Cambridge: Cambridge University Press, 2008. (Cited on page 150.)
- Rabin, Matthew and Joel L Schrag**, “First Impressions Matter: A Model of Confirmatory Bias,” *The Quarterly Journal of Economics*, 1999, 114 (1), 37–82. (Cited on page 23.)
- Rong, Xinchun**, “On the Development and Changes of the Shipping Business in New China (1949–2010) [Shi lun Xin Zhongguo Haiyun Shiye de Fazhan he Bianqian (1949–2010)],” *Researches in Chinese Economic History [Zhongguo Jingji Shi Yanjiu]*, 2012, 2, 127–137. (Cited on page 175.)
- Rosenstein-Rodan, Paul N.**, “Problems of Industrialisation of Eastern and South-Eastern Europe,” *The Economic Journal*, 1943, 53 (210/211), 202–211. (Cited on page 164.)
- Sanderson, Eleanor and Frank Windmeijer**, “A weak instrument  $F$ -test in linear IV models with multiple endogenous variables,” *Journal of Econometrics*, 2016, 190 (2), 212–221. (Not cited.)
- Santoir, Christian**, “Les Peul “refusés”, les Peul mauritaniens réfugiés au Sénégal,” *Cahiers des Sciences Humaines*, 1990, 29 (1), 577–603. (Cited on page 73.)
- Savage, Leonard J.**, *The Foundations of Statistics*, New York, NY: Wiley, 1954. (Cited on pages 23 and 33.)
- Seshan, Ganesh and Dean Yang**, “Transnational household finance: A field experiment on the cross-border impacts of financial education for migrant workers,” in “Qatar Foundation Annual Research Forum” number 2012 2012. (Cited on page 14.)

- **and Robertas Zubrickas**, “Asymmetric Information about Migrant Earnings and Remittance Flows,” *The World Bank Economic Review*, 2015, p. lhv032. (Cited on page 15.)
- Sethi, Rajiv and Muhamet Yildiz**, “Public Disagreement,” *American Economic Journal: Microeconomics*, 2012, 4 (3), 57–95. (Cited on pages 15 and 23.)
- Shi, Wenjie**, “A Study of the “156 Program” and its Influence on the Early Modernization of New China [“156 xiang gongcheng” ji qi dui Xin Zhongguo zaoqi xiandaihua de yingxiang chutan],” *Financial Information [Caijing Zhengfa Zixun]*, 2013, 1, 42–45. (Cited on page 167.)
- Song, Yang**, “What should economists know about the current Chinese hukou system?,” *China Economic Review*, 2014, 29, 200–212. (Cited on page 121.)
- Song, Zheng, Kjetil Storesletten, and Fabrizio Zilibotti**, “Growing Like China,” *American Economic Review*, 2011, 101 (1), 196–233. (Cited on pages 82, 85, 86, 165 and 186.)
- Sousa, Jose De and Sandra Poncet**, “How are wages set in Beijing?,” *Regional Science and Urban Economics*, 2011, 41 (1), 9–19. (Cited on page 87.)
- Tang, Yanyan**, “An Analysis of the Advantage of Backwardness Based on the Construction of the “156” Program [Cong “156 Xiang Gongcheng” de Jianshe Kan Houfa Youshi de Fahui],” *Theory Monthly [Lilun Yuekan]*, 2009, 20. (Cited on page 167.)
- Tombe, Trevor and Xiaodong Zhu**, “Trade, Migration and Productivity: A Quantitative Analysis of China,” June 2015. (Cited on pages 86, 88 and 165.)
- Torero, Maximo and Angelino Viceisza**, “To remit, or not to remit: that is the question. A remittance field experiment,” *Journal of Economic Behavior & Organization*, 2015, 112, 221–236. (Cited on page 14.)

- Trémolières, Marie and Donata Gnisci**, *Une vie transfrontalière en pointillé : Les migrants involontaires mauritaniens de la Vallée du Fleuve Sénégal*, ENDA-Diapol, 2004. (Cited on pages 73 and 74.)
- Tsui, Kai Yuen**, “China’s regional inequality, 1952–1985,” *Journal of Comparative Economics*, 1991, 15 (1), 1–21. (Cited on page 160.)
- Udry, Christopher**, “Gender, Agricultural Production, and the Theory of the Household,” *Journal of Political Economy*, 1996, 104 (5), 1010–1046. (Cited on pages 3 and 14.)
- **and Santosh Anagol**, “The Return to Capital in Ghana,” *The American Economic Review*, 2006, 96 (2), 388. (Cited on page 15.)
- United Nations**, “Trends in International Migrant Stock: The 2015 revision.” Technical Report, United Nations, Department of Economics and Social Affairs 2015. (Cited on pages 2 and 82.)
- Van den Steen, Eric**, “Rational Overoptimism (and Other Biases),” *The American Economic Review*, 2004, 94 (4), 1141–1151. (Cited on pages 15, 22, 26 and 44.)
- , “Overconfidence by Bayesian-rational agents,” *Management Science*, 2011, 57 (5), 884–896. (Cited on pages 15, 22 and 23.)
- Vendryes, Thomas**, “Migration constraints and development: *Hukou* and capital accumulation in China,” *China Economic Review*, 2011, 722, 669–92. (Cited on page 86.)
- Wang, Fei-Ling**, *Organizing through division and exclusion: China’s hukou system*, Stanford, CA: Stanford University Press, 2005. (Cited on page 121.)
- Wang, Qi**, “An Analysis of the “156 Program” and Sino-Soviet Relations in the 1950s [“156 Xiang Gongcheng” yu 20 Shiji 50 Niandai Zhong-Sulian Guanxi Pingxi],” *Contemporary China History Studies [Dangdai Zhongguo Shi Yanjiu]*, 2003, 10, 110–116. (Cited on page 167.)
- Whalley, John and Shunming Zhang**, “A numerical simulation analysis of (Hukou) labour mobility restrictions in China,” *Journal of Development Economics*, 2007, 83, 392–410. (Cited on page 86.)



- Wilson, Andrea**, “Bounded Memory and Biases in Information Processing,” *Econometrica*, 2014, 82 (6), 2257–2294. (Cited on page 23.)
- World Bank**, “World Development Indicators,” <http://data.worldbank.org/data-catalog/world-development-indicators> March 2017. Accessed March 14, 2017. (Cited on pages 12 and 70.)
- Zhang, Jiuchun**, “A Study of the “156 Program” of Industrial Construction in the 1950s [20 Shiji 50 Niandai Gongye Jianshe “156 Xiang Gongcheng” Yanjiu],” *Journal of Engineering Studies [Gongcheng Yanjiu – Kua Xueke Shiye zhong de Gongcheng]*, 2009, 1, 213–222. (Cited on page 167.)
- Zhang, Li and Li Tao**, “Barriers to the acquisition of urban hukou in Chinese cities,” *Environment and Planning A*, 2012, 44 (12), 2883–2900. (Cited on page 121.)
- Zhang, Shuguang**, *Economic Cold War: America’s Embargo against China and the Sino-Soviet Alliance, 1949–1963*, Stanford, CA: Stanford University Press, 2001. (Cited on pages 160 and 167.)
- Ziparo, Roberta**, “Public good provision and communication in the household. A model and some evidence from Cameroon,” February 2014. (Cited on page 15.)

## Résumé

Une des marques du développement est la résorption du décalage entre la distribution spatiale de l'activité économique et celle de la population. Cette thèse étudie les conséquences d'une telle résorption sous l'effet de la migration ou d'une redistribution géographique de l'activité économique.

Le premier chapitre s'intéresse à la relation entre migrants internationaux et foyers d'origine. Il met en lumière l'importance, pour les envois de fonds et les investissements, des croyances des migrants et de l'asymétrie d'information, aggravée par la distance, entre migrants et destinataires de ces envois.

Le deuxième chapitre explore la transformation de l'économie d'accueil sous l'effet d'un afflux de migrants venus des campagnes en ville, en Chine. Il mesure d'abord leur impact sur le marché du travail à destination, puis examine comment ils affectent l'allocation des facteurs de production et les contraintes rencontrées par les entreprises sur les marchés du travail et du capital.

Le troisième chapitre étudie l'impact sur le long terme d'un vaste programme d'industrialisation réalisé en Chine, et montre un retour de fortune. Ce renversement est imputé aux distorsions introduites sur le marché du travail local par la présence de grands complexes manufacturiers. Ce chapitre met en évidence le rôle de la migration pour surmonter ces imperfections et amener l'étape ultérieure de la transformation structurelle : le passage de l'industrie lourde à la production de biens de consommation et de services.

## Mots Clés

Migration, Information asymétrique, Information et connaissances, Allocation des ressources intra-ménage, Développement économique, Transformation structurelle, Economies en transition, Politique de développement régional.

## Abstract

A key characteristic of the process of economic development is the shrinking spatial mismatch between economic activity and population. This thesis analyzes what happens when this spatial mismatch is reduced, as people's places of residence and work or the geographical distribution of economic activity is altered.

Chapter I deals with the relationship between international migrants and their households of origin. It sheds light on the importance for remittances and investments of migrants' beliefs and the information asymmetry between remittance senders and recipients that distance aggravates.

Chapter II explores the transformation of the receiving economy due to an influx of rural-to-urban migrants in China. It first quantifies the effect of immigrants on the labor market at destination, and then investigates their impact on the reallocation of production factors and factor-market constraints faced by urban firms.

Chapter III focuses on the long-term impact of a large industrialization plan in China. It reveals a reversal pattern due to the distortions in local labor markets induced by the presence of big plants. It highlights the role of migration in overcoming such imperfections and in bringing about the later stage of structural transformation—from heavy industry to consumption goods and services.

## Keywords

Migration, Asymmetric Information, Belief, Household Production and Intrahousehold Allocation, Economic Development, Structural Change, Transitional Economies, Place-Based Policies.