2. China’s economic development: A perspective on capital misallocation

Among the abundant and growing literature aiming to understand the miracle of China’s economic development, the contribution of capital reallocation to aggregate total factor productivity (TFP) gain has offered an important perspective. This chapter first describes the general framework that links resource misallocation with TFP loss. It then surveys the empirical evidence of capital misallocation in China and discusses the quantitative implications for aggregate efficiency loss within such a framework. Two potentially interesting sources of capital misallocation are investigated: financial frictions and policy distortions. These sources highlight the remaining challenges in capital allocation that China faces after 40 years of reform.

Framework: Capital misallocation and TFP loss

Why was real gross domestic product (GDP) per capita in China only one-fortieth of that in the United States 40 years ago? And more generally, why are some countries so rich while others are so poor? An important finding economists have reached in the past two decades is that the differences in aggregate TFP are the dominant source of differences in GDP per capita. According to a growth accounting framework in Zhu (2012), growth in aggregate TFP has contributed more than 70 per cent to per capita GDP growth in China. What, then, has caused the substantial improvement in the aggregate TFP in China in the past four decades?

A new and growing literature, as surveyed in Restuccia and Rogerson (2013), argues that resource misallocation across heterogeneous firms in an economy lowers its aggregate TFP. For example, McKinsey Global Institute country and sector studies found large differences across firms within the same sector in many less-developed countries (South Korea, Brazil, Turkey and India). In fact, in many of these cases, the most productive firms within most sectors have productivity levels comparable with those in Western Europe and the United States, but there is a long tail of very low-productivity firms. If resources could be reallocated from the low-productivity firms to high-productivity firms within a country, the aggregate output would increase even with the same amount of production factors, which would lead to an increase in the measured aggregate TFP. This means the efficiency of resource allocation offers a new perspective with which to understand cross-country income differences: the aggregate TFP loss in poor countries relative to the first-best aggregate TFP
they could achieve may arise from the resource misallocation in those countries. Although full liberalisation is a thought experiment, since there are many frictions that impede perfect reallocation in reality, one way to interpret this framework is to gauge the potential gain in aggregate TFP if such friction could be removed under a first-best situation.

Among various production factors, capital misallocation has been documented as a prevailing empirical phenomenon, both in less-developed economies in general (for example, Banerjee and Moll 2010) and in China in particular. In fact, the reallocation of capital within the manufacturing sector is a focal point in explaining the miracle of China’s economic development in Song et al. (2011).

Evidence of capital misallocation

Evidence of capital misallocation in China takes many forms. There are two ways to summarise this evidence: first, a direct approach to studying the cost of capital or return on investment and, second, an indirect approach to infer the misallocation from the dispersion of marginal product of capital.

We start to present the evidence using the first approach. Although the aggregate marginal product of capital in China is about 20 per cent, as in Bai et al. (2006)—a relatively high but still reasonable value by international standards for developing economies—there is a very different picture when we look at the evidence for the cost of capital or return on investment at more disaggregated levels.

In a cross-country comparative study, Allen et al. (2005) find that China’s corporate governance, accounting standards and investor protection systems are poor at best, measured by existing standards in the literature, while its banking system is not well developed and is to a large degree inefficient. The Shanghai Stock Exchange and Shenzhen Stock Exchange have been growing rapidly since their inception in 1992, but their scale and importance are still not comparable with other channels of financing—in particular, the banking sector—for the entire economy. In this environment, Chinese firms must rely heavily on retained earnings to finance investment and operational costs. However, financial repression is far from uniform: Chinese banks, which are mostly state-owned, tend to offer easier credit to state-owned enterprises (SOEs), with less screening, higher lines of credit, lower interest rates and fewer collateral requirements. The Chinese stock market is disproportionately dominated by the SOEs and large semiprivatised SOEs. Not surprisingly, firms in the informal sector are subject to strong discrimination in credit markets. Many successful non-SOEs do not use any channel of formal financing during development (Allen et al. 2005: Figs 3-A and 3-B).
Consistent with this institutional background and based on a survey covering a stratified random sample of 12,400 firms in 120 cities in China with firm-level accounting information for the period 2002–04, Dollar and Wei (2007) examine the presence of systematic distortions in capital allocation that result in uneven marginal returns to capital across firm ownership type, region and sector. They provide a systematic comparison of investment efficiency among wholly and partially state-owned, wholly and partially foreign-owned and domestic privately owned firms, conditioned on their sector, location and size characteristics. They find that, even after a quarter of a century of reforms, SOEs still have significantly lower returns to capital, on average, than domestic private or foreign-owned firms. Similarly, certain regions and sectors have consistently lower returns to capital than other regions and sectors. The findings in this pioneering work have been confirmed by many subsequent studies, as surveyed below.

A second source of evidence involves fitting a production function to firm-level data and directly estimating the distribution of marginal products. A key paper in advancing this structural approach is Hsieh and Klenow (2009). They assume that there are firm-specific wedges affecting total production and capital—essentially modelled as ‘taxes’. As a result of these wedges, firms produce amounts different to what would be dictated by their productivity and also may have different capital–labour ratios. If there are no firm-specific distortions and all firms within a sector have the same markup, the TFP revenue (TFPR) will be equalised across firms within a sector. In general, variation of TFPR within a sector will be a measure of misallocation.

Hsieh and Klenow (2009) quantify the potential extent of misallocation using microdata on plants in the manufacturing sector in China, India and the United States. First, they find that there is greater dispersion of TFPR in India and China than in the United States. For example, for TFPR, the 90–10 ratio is 1.59 in China, 1.60 in India and 1.19 in the United States. Second, they estimate that this could account for lower aggregate productivity. In particular, their estimates suggest that this type of misallocation could increase TFP in China by 30 to 50 per cent and in India by 40 to 60 per cent. Finally, they also find evidence for more rapid reallocation towards firms with higher TFPR in China than even in the United States, possibly reflecting rapid reallocation as less-efficient SOEs are being weeded out there. But reallocation away from less-efficient firms seems slower in India.

As pointed out by Restuccia and Rogerson (2013), the finding in Hsieh and Klenow (2009) remains the single strongest piece of evidence to support the idea that misallocation is an important component of cross-country differences in TFP. However, the highly parametric assumptions on preferences and production technology in this approach also face methodological challenges. The validity of the inference in Hsieh and Klenow (2009) hinges on two conditions: 1) average and marginal revenue products have the same dispersion; and 2) the dispersion
of marginal revenue products—a mirror image of price heterogeneity—reflects the magnitude of misallocation. Both conditions are strict. Condition (1) applies only to environments with homogeneous output and demand elasticities. Condition (2) will not necessarily hold in a dynamic environment with frictions such as adjustment costs. When it comes to the data, this approach needs another condition—that measurement errors do not add to the dispersions. Violation of any of the conditions would lead to biased estimations.

To address these concerns, Song and Wu (2015) develop a new method of identifying capital misallocation in a more general environment, where none of the conditions has to hold. The new method has a distinctive feature of matching a set of first and second moments of both the revenue–capital ratio and the profit–revenue ratio. The profit–revenue ratio, which has not yet been explored in the misallocation literature, plays an important role in identification. Specifically, Song and Wu (2015) match the variance of the revenue–capital and profit–revenue ratios and the cross correlation between the two ratios. The three empirical moments allow them to back out the three parameters governing the magnitude of the misallocation and unobserved heterogeneities in output and demand elasticities. In addition, while the original approach uses cross-sectional data, the new method explores intergroup variations in panel data, which can effectively mitigate the bias caused by capital adjustment costs and measurement errors.

Song and Wu (2015) then apply the structural estimation to firm-level panel data from the industrial survey conducted by China’s National Bureau of Statistics (NBS), using the generalised approach. They find that a full correction of capital misallocation would increase China’s manufacturing output by 20 per cent. In contrast, the original Hsieh and Klenow (2009) approach implies a much larger efficiency gain, of 35 per cent. This suggests that, first, it is indeed important to control for the unobserved heterogeneities in output and demand elasticities, investment frictions and measurement errors, to infer the correct magnitude of TFP loss from dispersion in marginal products. Second, even after controlling all these factors, one still identifies substantial TFP loss from capital misallocation in China, which confirms the main finding in Hsieh and Klenow (2009) and the misallocation literature at large.

The seminal framework of Hsieh and Klenow (2009) has inspired many important studies. For example, Brandt et al. (2013) focus on factor misallocation at a more aggregate level, between provinces and between the state and nonstate sectors in China’s non-agricultural economy, which includes both manufacturing and services, and covering a longer period, from 1985 to 2007. They also decompose the overall TFP loss into the losses due to interprovincial and intraprovincial intersectoral distortions. They find, first, on average, the misallocation of factors across provinces and sectors resulted in a reduction of non-agricultural TFP of at least 20 per cent, with the intraprovincial distortions accounting for more than half of the total loss.
Second, TFP losses from interprovincial distortions were relatively constant over the entire period. Third, despite significant interprovincial labour flows, the TFP loss from interprovincial labour market distortions remains high due to an increase in the cross-province dispersion in TFP. Fourth, the measure of intraprovincial distortions declined sharply between 1985 and 1997, contributing to 0.52 per cent of non-agricultural TFP growth per year, but then increased significantly in the past 10 years, reducing the non-agricultural TFP growth rate by 0.5 per cent a year. Finally, almost all of the intraprovincial distortions were due to the misallocation of capital between the state and nonstate sectors, which increased sharply in recent years.

**Sources of capital misallocation**

The qualitative significance and quantitative importance of capital misallocation in China raise one pertinent research question: what are the underlying factors that cause the misallocation? Two natural candidates have attracted increasing interest in recent literature: capital market imperfections due to financial frictions and nonmarket distortions induced by government policies.

It is obviously difficult to make a clear distinction between financial frictions and policy distortions. They are not necessarily very different in conceptual terms, and they may also overlap with a number of other frictions and distortions that are very similar. Therefore, in this chapter, we narrow the definition of financial friction strictly to those factors that are due to imperfect information or imperfect enforcement in the capital market that would cause capital misallocation even in developed economies. We ask to what extent such factors contribute to the observed capital misallocation. This implies that all those nontypical financial frictions are labelled as policy distortions in a very broad sense.

**Financial frictions**

Quantifying how much of the observed capital misallocation can be attributed to financial frictions is the central theme of some recent literature—for example, Moll (2014) and Midrigan and Xu (2014), among many others. While modelling details and estimated magnitudes differ, these studies share a common methodology: they develop theoretical models and gauge the size of TFP loss by calibrating model parameters to match the distribution and dynamics of output across production units. Despite the financial underdevelopment and substantial capital misallocation in China, there have been relatively few studies that directly link financial frictions with capital misallocation and TFP loss in China. Ek and Wu (2018) and Wu (2018) are two attempts in this direction. Both papers offer novel identification
strategies that differ from the existing literature and apply such strategies to detailed firm-level panel data from China’s Annual Survey of Industrial Enterprises, covering the period 1998 to 2007.

Ek and Wu (2018) directly estimate the effect of financial frictions on capital misallocation using an inference from investment–cash flow sensitivity. Although investment–cash flow sensitivity is frequently used as an indicator of financing constraint, and financing constraint is a consequence of financial frictions, there has not been any research that connects capital misallocation directly to investment–cash flow sensitivity. This chapter fills this gap by providing a simple yet consequential theoretical model, which links the heterogeneity in investment–cash flow sensitivity—a common indicator of financing constraint—to the dispersion of the marginal revenue product of capital (MRPC), which is a direct measure of allocative inefficiency.

The validity of this new approach, of course, depends crucially on the answers to two methodological questions. First, is investment–cash flow sensitivity a reliable indicator of financing constraint? Even in perfect capital markets, cash flow sensitivity may result from measurement errors in Tobin’s q, from imperfect competition and/or decreasing returns to scale, from the presence of capital adjustment costs or from a combination of measurement errors in Q and identification problems. Furthermore, a firm’s cash flow position is endogenous to its productivity shocks and may contain information about its investment opportunities.

To address these concerns, Ek and Wu (2018) present a structural model of costly external finance. Firms in this model are allowed to face imperfect competition and/or use decreasing returns to scale technology. In the absence of any friction, their model generates the same optimal condition as those models in the recent literature: optimal capital stock is only a function of current output, Jorgensonian user cost of capital and production technology. This allows them to develop an empirical specification for investment that does not rely on Tobin’s q. They then consider an autoregressive–distributed lag structure to accommodate the possibility of capital adjustment costs, which yields an error correction specification, as in Bond et al. (2003). Under the null hypothesis of no financial friction, cash flow should not affect investment under this specification. Ek and Wu (2018) allow for the potential endogeneity of cash flow in their estimation using generalised method of moments techniques. They test whether the cash flow terms show significantly different predictive powers across those samples that produce significantly different investment–cash flow sensitivities.

The second concern about investment–cash flow sensitivity and financing constraint is the well-known Kaplan and Zingales critique. Kaplan and Zingales (1997) argue that investment–cash flow sensitivities do not always monotonically increase as firms become more financially constrained. Thus, one cannot in general use estimates
of investment–cash flow sensitivities to proxy the severity of financial frictions. The theoretical model in Ek and Wu (2018) shows that the relationship between investment–cash flow sensitivities and the severity of financial frictions indeed depends on the curvature of the profit function and the cost function of external finance. However, even though more financially constrained firms do not necessarily exhibit higher sensitivity, it remains the case that unconstrained firms should display no investment–cash flow sensitivity. Therefore, finding that one group of firms has positively significant sensitivity while the other group shows no sensitivity is a sufficient, though not necessary, condition of capital misallocation, which is indeed the general pattern of the empirical findings. Given that Ek and Wu’s identification strategy only relies on investment–cash flow sensitivities and not excess investment–cash flow sensitivities, it is not subject to the Kaplan and Zingales critique.

When they apply the error correction investment model to a 10-year balanced panel of US Compustat database firms, Ek and Wu (2018) do not detect any investment–cash flow sensitivity. In contrast, there are significant sensitivities for a 10-year balanced panel of Chinese firms. Within Chinese firms, when splitting the sample using any criterion based on age, size, ownership or political connection, and for both the balanced and the unbalanced panels, they obtain significant cash flow effects for firms that are young, small, non-SOEs and without political connections. The resulting aggregate TFP loss implied by these investment–cash flow sensitivities is 4 to 5.2 per cent for the balanced panel and 10 to 15.2 per cent for the unbalanced panel.

While Ek and Wu (2018) take a structural approach, the identification strategy in Wu (2018) is more of a reduced-form approach. In particular, it uses a program evaluation perspective to quantify the effects on capital misallocation of both financial frictions and policy distortions.

**Financial frictions versus policy distortions**

As mentioned earlier in this chapter, a stylised fact of the Chinese economy is that the cost of capital, the return to investment or the proxies for the MRPC differs significantly across firms with different ownership types. This distinct phenomenon has often been taken as direct evidence of policy distortions in capital allocation. After all, ownership should not matter for MRPC in a world without policy distortions, were ownership orthogonal of other firm characteristics that may affect MRPC. However, as a less-developed economy, China also has a less-developed capital market with a lagged legal, auditing and contracting environment. If firms with different ownership types systematically differ in other characteristics, such
as age and size, and if such characteristics do affect MRPC because of financial frictions, even in the absence of policy distortions, these firms could still have different MRPC under an imperfect capital market.

The unique institutional feature on firm ownership type and MRPC dispersion in China inspires us to design the following identification strategy. Imagine an investment-promoting program that offers favourable treatment to some firms. The treatment status of a firm depends on its ownership type. The exact treatment may take various forms—for example, an investment tax credit or a special bank loan with a low interest rate. The effect of the treatment is to lower the generalised user cost of capital, or the mirror image of the MRPC of those treated firms. Firms differ not only in their treatment status, but also in a set of firm characteristics, known as covariates in an evaluation problem, through which financial frictions operate to affect their MRPC. The average treatment effect of the program on MRPC dispersion can then be decomposed into the average treatment effect on treated (ATT) and the selection bias. The ATT is the difference between the actual MRPC of those treated and the counterfactual MRPC of those treated had they not received the treatment, which identifies the effect of policy distortions on the average MRPC dispersion across ownership type. The selection bias is the MRPC difference between the treated and the untreated in the absence of treatment, which captures the effect of financial frictions on the average MRPC dispersion across ownership.

In applying this identification strategy, Wu (2018) considers a structural model with both policy distortions and financial frictions. To nest those microfoundations that have been the most common building blocks in the recent literature on financial frictions and aggregate TFP, Wu allows for two types of highly synthesised reduced-form financial constraints. The aggregate TFP loss in the model economy depends on the dispersion of the firm-specific MRPC, which is determined by some joint distribution of a set of parameters. These parameters govern the magnitude of firm-specific policy distortions and financial frictions and characterise the state of firm productivity and internal funds. Wu then conducts propensity score matching based on a set of covariates that are suggested by the model through which financial frictions may affect MRPC, even in the absence of policy distortions. These covariates are exactly the same as those that appear in the vast theoretical and empirical literature on financial frictions. However, by matching firms that have different treatment status but are otherwise similar in terms of these covariates, one does not have to take a stand on the functional relations among these observed covariates and MRPC; nor does one need to specify the exact causal direction between the treatment status and the observed covariates.

The detailed firm-level panel data from China’s Annual Survey of Industrial Enterprises is employed to obtain point effect estimates on the ATT and the selection bias across different firm ownership types. For example, not surprisingly, an SOE, on average,
has an MRPC 42 per cent lower than that of a domestic privately owned firm, where policy distortions and financial frictions lower its MRPC by 22 and 20 per cent, respectively. More interestingly, the average MRPC of a foreign-owned firm is 2 per cent lower than that of a domestic privately owned firm. But without policy distortions, its MRPC would be 20 per cent higher than that of a domestic privately owned firm due to financial frictions. This suggests that foreign-owned firms in fact receive similar levels of favourable policy distortions to state-owned firms.

Although such estimates are interesting in their own right, what truly helps us answer our research questions is a byproduct of the matching procedure—a counterfactual MRPC of those treated firms had they not received the treatment. Using this information, Wu (2018) calculates the aggregate TFP losses in a hypothetical economy without policy distortions, which turn out to vary from 7.3 to 9.4 per cent over the period 2000–07. Thus, her estimates of the effect of financial frictions on aggregate TFP loss are in line with Ek and Wu’s (2018), based on the same dataset but using a completely different approach. The annual average aggregate TFP loss in the actual economy reaches 27.5 per cent. This implies that 70 per cent of the aggregate TFP loss can be attributed to policy distortions. It also finds that the policy distortions have reduced the average MRPC in China by 15.5 per cent, which provides one possible explanation for China’s unusually high investment rate.

Policy distortions

Although a large literature has identified the policy distortions as an underlying source of capital misallocation and aggregate TFP loss in China, the policy distortions used in most papers are modelled in only an abstract and generic way. To offer specific policy implications, it is important to identify specific institutional factors that have distorted capital allocation.

Using matched samples of firms with balanced covariates, Wu (2018) evaluates several popular hypotheses on why the Chinese Government has introduced various rules, regulations and institutions that favour certain firms. From the public finance perspective, the first possible reason for a government to favour a firm is that the firm contributes significant tax revenue.

Second, a government may also distort capital allocation to pursue specific industrial policies. For example, China is well known for having adopted an export-led growth strategy since the beginning of its ‘reform and opening-up’ policy (Lin 2012). More recently, it has been suggested that China practices a form of state capitalism in a vertical industrial structure: SOEs are explicitly or implicitly allowed to monopolise key upstream industries, while the downstream industries are largely open to private competition (Li et al. 2015). Under these two hypotheses, firms in upstream industries that are exporting can expect to receive favourable policy distortions.
Third, the well-known trade-off between growth and stability facing the Chinese Government has often been taken as an argument to justify policy distortions. To minimise social unease and reduce resistance to reform, the government may have a strong political motivation to maintain employment stability. For example, to avoid laying off workers or shutting factories during an economic downturn, the government usually asks the state-owned banks to bail out loss-making SOEs, which creates a problem known as the ‘soft-budget constraint’ (Qian and Roland 1998; Brandt and Zhu 2001). Under this rationale, we may regard the government as a risk-averse social planner that optimally allocates capital according to the capital asset pricing model. If so, firms that are countercyclical have a smaller beta and only need to offer a lower required rate of return on capital.

Finally, differing from all the above hypotheses, which assume a benevolent government, an alternative hypothesis is that the government prefers firms with political connections. For example, Communist Party membership has been found to help private entrepreneurs obtain loans from banks or other state institutions (Li et al. 2008; Guo et al. 2014). Firms with government-appointed or government-connected chief executive officers are found to face much less severe financial constraints (Fan et al. 2007; Cull et al. 2015). Since there is no information regarding a firm’s entrepreneur or chief executive officer in our dataset, whether the firm has a labour union is adopted as an alternative measure of political connection. Unlike labour unions in most Western countries, which help workers collectively bargain with employers for higher wages and better working conditions, Chinese labour unions pass on the ideology of the Communist Party to the workers and monitor whether the firm is ‘politically correct’ or at least is operating consistently with Communist Party policy.

To test these interesting hypotheses, Wu (2018) implements a regression using the restricted sample of matched firms. The dependent variable is the difference between the actual MRPC of a firm in a year and its counterfactual MRPC had the firm not received the favourable treatment in that year. The independent variables include the six factors inferred from the popular hypotheses. This produces a set of interesting findings. First, the actual MRPC of those firms that contribute high tax revenue is in fact higher than their counterfactual MRPC. This denies the first hypothesis that firms receive favourable policy distortions in capital because they contribute more tax revenue. Instead, it suggests that those firms that have received favourable policy distortions in capital also receive favourable tax treatment, such as tax breaks or direct subsidies. Second, averaging across the years, and all else being equal, a firm that is an exporter, belongs to an upstream industry and has a labour union has an MRPC 14 per cent, 2.6 per cent and 14.9 per cent, respectively, lower than otherwise. Finally, beta is the only variable whose estimates change the signs from significantly positive to insignificantly negative over our sample period. A positive coefficient on beta is consistent with the capital asset pricing model, thus verifying
the motivation of policy distortions as a trade-off between risk and return. The fact that beta becomes irrelevant after 2005 seems to indicate that employment stability was no longer a major concern for the government in more recent years. The same pattern is highlighted in Hsieh and Song (2015) using different evidence.

Three conclusions therefore can be drawn from the empirical exercises. First, favourable policy distortions in capital go hand-in-hand with favourable tax treatment. Second, pursuing an export-led growth strategy and practising state capitalism are two important factors that drive policy distortions. Political connection with the Communist Party is another reason for firms to receive favourable treatment. Finally, concern about the trade-off between return and risk also leads to policy distortions but is relevant only in the early years of the study period.

Conclusions and policy implications

In recognising that both financial frictions and policy distortions have caused capital misallocation and aggregate TFP losses in China, this chapter uses an important perspective to understand China’s economic development in the past four decades, thanks to both the development of the financial environment and the economic reforms correcting overall policy distortions. To the extent that large and persistent capital misallocation still exists after 40 years of reform, the factors identified in this chapter can be interpreted as directions for further reform. The structural reforms that are particularly relevant to capital misallocation are mainly in the enterprise sector and the financial system—for example, those reforms proposed by the World Bank and the Development Research Centre of the State Council (2013).

References


