The Causes and Costs of Misallocation

Diego Restuccia Richard Rogerson
University of Toronto Princeton University

1. The Origin of the Misallocation Literature

One of the long-standing issues in economics is to understand why living standards differ so much across countries. One diagnostic tool that researchers have found to be useful in this context is that of development accounting. This tool posits that aggregate output per worker in country j at time t, denoted Y_{jt}, can be represented as $Y_{jt} = A_{jt} \cdot F(K_{jt}, H_{jt})$, where F is a constant returns to scale production function, K_{jt} is input of physical capital per worker, and H_{jt} is human capital per worker. A_{jt}, termed TFP, is a measure of the productivity with which resources are used.

Given cross-sectional data on Y, K, and H, and a choice of function F, this method can measure the extent to which the large differences in output per worker across countries are accounted for by differences in the factors of production, K and H, versus TFP, A. The consensus in this literature is that differences in A are a large, if not dominant source of cross-country differences in living standards. (See, for example, Prescott (1998), Hall and Jones (1999), and Hsieh and Klenow (2010).) One possible interpretation of the low values of A in poor countries is that frontier technologies and best practice methods are slow to diffuse to these countries. But the recent literature on misallocation, which is the focus of this article, offers another interpretation of the lower values of A: that the factors of production are not efficiently allocated across heterogeneous producers.

A simple example will serve to fix ideas and facilitate exposition. Central to the example is the reality that aggregate output is produced by many different and heterogeneous producers, that in particular differ greatly in their individual levels of productivity (Syverson (2011)). Specifically, assume that there are N potential producers of a homogeneous good and that producer i has a production function $y_i = A_i \cdot f(h_i, k_i)$, where y_i is output, h_i is labor input, k_i is input of capital, f is a strictly increasing and strictly concave production function, and the A_i reflects differences in productivity across producers. Assume also that there is a fixed cost of operation for any producer who operates, measured in units of output and denoted by c. Given an aggregate amount of labor and capital, denoted by H and K respectively, there is a unique choice of which producers should operate and how labor and capital should be allocated across them in order to maximize total output net of fixed operating costs. In this example these choices have the following form: a threshold rule determines which producers operate (i.e., producers operate if A>Abar) and conditional upon operation, producers with higher values of A will be allocated a greater amount of labor and capital. Assuming that the A_i's are heterogeneous, this efficient allocation will induce a distribution of producer sizes, something that we observe in reality, even within narrowly defined sectors.

Now consider two economies that use the same amount of labor and capital in aggregate but produce different levels of output. The above setting offers three conceptually distinct channels through which one can account for the different levels of aggregate output. One is that the distribution of the A_i's differ. A second is that even if the distribution of the A_i's is the same for both economies, they could choose different producers to operate. Third, even conditional on the two economies having the same distribution for the A_i's and choosing the same producers to operate, they may allocate inputs (i.e., capital and/or labor) differently across these producers. The first channel reflects *technology*. The second channel reflects *selection*, and the third channel reflects what the literature has referred to as *misallocation*. Note that all three of these channels will generate differences in aggregate TFP in a standard development accounting exercise. Conceptually, one might view selection effects as a special case of misallocation, but from an empirical perspective an important issue is that one does not observe potential producers who do not produce, making it more difficult to measure selection effects. It is important to note that while we can identify three distinct channels, any given difference in policies across economies may well generate effects along all three channels. This same property holds in the context of development accounting: it isolates three different channels (K, H, and A) through which output is affected, but specific policies may well create effects along all three channels.

Casual empiricism suggests that both technology and misallocation channels are potentially relevant. (Selection seems more immune to casual empiricism.) A visit to any less developed country reveals that much production seems to take place using outdated methods, whether in agriculture, manufacturing, or services. And many studies and anecdotes detail how corrupt business practices, regulation, limited competition, or direct government involvement distort the allocation of resources from their most efficient use, particularly so in poorer economies. More generally, the idea that the allocation of inputs across establishments is an important component of aggregate productivity is further reinforced by studies in the United States and other countries that find that the reallocation of inputs from less productive to more productive establishments is an important component of annual aggregate productivity growth. (See, for example, Bailey et al (1992) and Foster et al (2008).)

Given this context, three key questions follow immediately. First, how important is this type of misallocation as a source of aggregate productivity differences across countries? Second, if the effects of misallocation are substantial, what are its key causes? And third, aside from the direct cost of lower contemporaneous output, are there additional costs associated with misallocation? In this article we provide our perspective on where the literature stands with regard to answering these three questions. It is not our intention in this article to survey the literature and, as a result, we inevitably neglect many important references and contributions. We instead refer the reader to available survey articles of this literature, for instance Restuccia and Rogerson (2013) and Hopenhayn (2014).

2. Potential Sources of Misallocation

It is useful to begin with a discussion of prominent potential sources of misallocation. This discussion serves two purposes. First, it helps us understand the scope of the challenge that one faces when trying to assess misallocation. Second, it will help us understand the evolution of the literature and emergence of different approaches. It is important to emphasize that the nature of misallocation that we are focused on is quite specific. Economists routinely study distortions that affect allocations along many dimensions, but in this article we are specifically interested in factors that distort the distribution of inputs *across* producers of a given good. For example, in the context of the standard neoclassical growth model, a proportional tax on capital income distorts the consumption/savings decision and hence may be described as causing misallocation along this margin. But this policy is not a source of misallocation as we have defined it, as its sole effect in this context is to decrease the aggregate amount of capital in the economy. In contrast, if we tax capital income derived from different producers at different rates, then in addition to having an effect on aggregate capital there will also be an effect on how aggregate capital is allocated across producers. It is this second effect that reflects the misallocation that we are interested in.

Implicitly then, we are looking for factors that differentially affect producers so as to distort the allocation of resources across producers. In the example presented in the introduction, the allocation of inputs that maximizes output will equate the marginal products of labor and capital across all producers with positive inputs. Thinking about factors that interfere with equalization of marginal products is thus a good way to identify possible sources of misallocation. With this in mind we can proceed to list some prominent examples. Rather than provide a long laundry list of very specific potential sources of misallocation, we instead emphasize a few general categories of factors with a small number of specific examples.

The first category reflects legislated provisions, including features of the tax code, and regulations. Specific examples would include provisions of the tax code that vary with firm characteristics (e.g., size or age), tariffs applied to particular categories of goods, labor market regulations such as employment protection measures, product market regulations such as those that restrict size or limit market access, and land regulations that may particularly affect agricultural activity. Some regulations might generate misallocation because they do not apply uniformly across all producers; many regulations exempt firms below some size threshold. But even a regulation that applies uniformly to all firms within an industry may generate misallocation within the industry. For example, an employment protection measure that applies to all firms within some industry impacts differently on firms that are desiring to expand versus firms that are desiring to contract.

The second category reflects discretionary provisions made by the government or other entities (e.g., banks) that favor or penalize specific firms. Often referred to collectively as “crony capitalism” or even “government corruption”, the most obvious examples are subsidies, tax breaks or low interest rate loans granted to specific firms, with subsidies to government run enterprises being one such case in point. Less obvious examples include unfair bidding practices for government contracts, preferential market access, or selective enforcement of taxes and regulations.

The third category is market imperfections. One example of this is monopoly power. For example, an efficient allocation might imply that a highly productive producer should have a high level of output. But if this producer has market power it may choose to restrict output and charge a higher price. A second example is various forms of market incompleteness. One important example is financial markets. A large literature has documented substantial differences in financial market development across countries as measured by various indicators. If firms face restrictions on borrowing, due to collateral constraints for example, then productive firms may operate below their efficient scale, while at the same time some less productive firms with access to capital may operate at too large of a scale. More generally, there may be restrictions on the types of contracts or arrangements that can be sustained. The work of Bloom et al (2013) suggests that the size of productive firms in India is restricted by the inability to delegate management outside of the family on account of poor enforcement of property rights. Another important example is how lack of land titling affects markets for land.

There are three points that we want the reader to take away from this discussion. First, the relevant set of factors that might plausibly be generating misallocation is wide-ranging. Second, many of the specific factors will be very narrow in scope. For example, many regulations are particular to specific sectors and even regions. Third, many of these factors, in particular those in the second category of discretionary provisions, are not easily amenable to systematic measurement.

3. Measuring Misallocation: Methodology

Consistent with the fact that the potential sources of misallocation are wide-ranging, the work that helps to shed light on the importance of misallocation is itself wide-ranging in both approach and focus, spanning the fields of development, industrial organization, international economics, labor economics, finance, and macroeconomics. The size and scope of these efforts makes it impossible for us to do justice to all the relevant work given our limited space. Nonetheless, we do want to highlight a few illustrative examples of the different approaches and results, and discuss how they relate to the three questions we posed in the introduction.

Many studies in the micro-development literature document the presence of misallocation in various contexts using distinct approaches. Banerjee and Duflo (2005) is an important early example of work that documents credit market imperfections among manufacturers in India. De Mel et al (2009) establish wedges between the marginal product of capital and borrowing rates among small producers in Sri Lanka using experimental methods. McKenzie and Woodruff (2008) establish the same fact for small retailers in Mexico, while Banerjee and Duflo (2008) confirm it for medium size firms in India. Karlan et al (2014) and Beamon et al (2015) carry out randomized control trials to demonstrate imperfections in credit and insurance markets facing farmers in less developed countries. Field (2003) shows that land titling for urban land in Peru affects the allocation of labor between the informal and formal sector. Besley and Ghatak (2010) survey work that demonstrates the role of property rights in creating misallocation.

The papers just cited offer compelling evidence that credit, insurance and land markets are either non-existent or do not function well in specific less developed countries. This type of evidence is critical to motivate the possibility that misallocation has important aggregate effects and to provide descriptive evidence on the specific nature of market imperfections. But at the same time they do not allow one to assess the aggregate consequences of misallocation.

The industrial organization literature includes case studies that shed light on the effect of particular types of regulations in specific industries. For example, Olley and Pakes (1996) study the effect of regulation in the US telecommunications industry and find an important role for misallocation. Kirwan et al (2012) show how repeal of the federal quota system on tobacco farms led to large productivity gains from reduced misallocation. Similarly, there are studies that document misallocation in other specific contexts. For example Peek and Rosengren (2005) and Caballero et al (2008) document "zombie lending" practices in Japan, a process by which banks continue to extend credit to poorly performing businesses in order to avoid writing down bad loans.

These types of studies are able to assess the implications of narrowly targeted regulations on the extent of misallocation in very specific and sometimes very small sectors. But they leave unanswered the question of how important these effects are at an aggregate level and how generalizable they are across countries.

Heckman and Pages (2004) refers to several papers documenting the effects of labor market regulations on various labor market outcomes using micro data from Latin America and the Caribbean. While these studies document the prevalence of labor market regulations and shows that they do affect outcomes, they do not assess the implied effects on productivity. There are many papers that document specific aspects of product market regulations in various countries. Guner et al (2008) describes specific examples from Japan and India.

To a large extent we view the body of work just discussed as laying the empirical foundation for the misallocation literature. Specifically, it serves to document the extent to which a case can be made for several of the various factors listed in Section 2 as being empirically relevant. This work is clearly very important in building a case for the importance of misallocation. But it does not address the broader question of whether misallocation is an important factor behind cross-country differences in TFP. It is this question that we now turn our attention to.

There are two different approaches that this branch of the literature has adopted. For reasons that will become clear shortly, we will label one of these approaches as the direct approach and the other as the indirect approach.

The essence of the direct approach is to focus on a specific source of misallocation. In some applications of this approach, the researcher measures the underlying source of misallocation, and then builds a quantitative structural model that can be used to quantify its effects. There is a long tradition of following this approach in the public finance literature as a way to measure the distortions associated with various features of tax systems. A virtue of this approach is that by directly linking the extent of misallocation with the source of the misallocation, any finding of substantial effects due to misallocation will necessarily identify the source of the misallocation as a key byproduct of the analysis. Alternatively, one might try to find quasi-natural experiments that can shed light on the importance of a particular source of misallocation.

To the extent that broad based evidence from quasi-natural experiments is difficult to come by, the direct approach focuses on using structural models. One issue with this approach is that it requires a fully specified structural model, and variation in model specification may have important effects on the findings. While we acknowledge this as an important issue, and will devote significant attention to it subsequently, we also want to stress that one cannot avoid the need for some structure in this context. The reason for this is that evaluating the extent of misallocation necessarily requires one to compute a counterfactual--how much additional output could be generated by reallocating inputs among producers—and computing counterfactuals of this sort must require some structure.

But the direct approach also faces another challenge. Implementing it implicitly requires that one obtain quantitative measures of the underlying source of misallocation. If we think that easily identified regulations, or features of statutory tax codes are the key source of misallocation then this is perhaps not a problem. But if we think that discretionary deals in the form of tax breaks, low interest loans, preferential market access, awarding of contracts at beneficial terms, or lax enforcement of regulations (i.e., crony capitalism or corruption) are the most important sources of misallocation, then the direct approach faces a steep limitation. While there are various attempts to produce rankings of corruption across countries, there are no widespread quantitative measures of the extent of corruption. If many of the quantitatively important factors causing misallocation are indeed difficult to measure, then the direct approach will be of limited use. More generally, even if regulation is an important source of misallocation at the aggregate level, if the nature of the regulation is highly specialized across specific industries it may still be very challenging to use the direct approach.

Motivated by the possibility that many important sources of misallocation may be very difficult to measure or come from very specialized and diffuse sources, the indirect approach seeks to identify the extent of misallocation without identifying the underlying source of the misallocation. To understand this approach, consider the simple model described in the introduction. As noted earlier, the efficient allocation of inputs across producers will necessarily equate marginal products across all active producers. It follows that in this simple model, variation in the marginal products is evidence of misallocation, so that directly examining variation in marginal products provides the opportunity to measure the amount of misallocation without specifying the underlying source of misallocation.

This approach also requires some structure. In our simple example, given cross-section data on output, labor and capital for each of the operating producers, it is sufficient to specify the production function f to directly compute the implied amount of misallocation. To see why, note that given data on y, k and h for each producer and a production function f we can infer the values of the A_i. Given the production function f and the values of A_i we can directly solve for the efficient allocation of inputs among these producers and hence the maximal output. Comparing this to the observed amount of output provides an assessment of the extent of misallocation. Note that because this exercise takes the set of producers as given, it necessarily does not address selection effects. So as noted earlier, even if one views selection effects as conceptually akin to what we have called misallocation, this procedure will only isolate the misallocation effect.

Although the indirect approach still requires some structure, it does not require one to specify a full model, as required in the case of the direct approach. While one might be tempted to conclude that the indirect approach is therefore more powerful than the direct approach, it is important to note two caveats to temper this conclusion. First, if the firm level data contains measurement error, then one might infer variation in marginal products that are nothing more than measurement error. Second, it is easy to write down models in which efficient allocations do not entail equality of marginal products at each point in time. For example, if there are shocks and inputs must be chosen before the realization of the shocks, there will be ex post variation in marginal products. And closely related, in a dynamic setting the presence of adjustment costs can have similar implications. We discuss these in more detail in the next section.

4. How Important is Misallocation? Results Using the Indirect Approach

A useful starting point for our discussion is the paper by Restuccia and Rogerson (2008). They used a version of the Hopenhayn (1992) industry equilibrium model calibrated to match features of the US economy to explore the extent to which misallocation caused by firm specific taxes and subsidies would impact aggregate TFP. These firm specific taxes and subsidies were chosen as a simple representation of the many different factors that might be generating misallocation. Their analysis led to two key findings. The first is that, although randomly taxing some establishments and subsidizing others, thereby reshuffling inputs among establishments, lowers aggregate TFP, the effect of this form of misallocation in their calibration is somewhat modest, being less than 10%.[1]

The second result in Restuccia and Rogerson (2008) concerned the case in which high productivity establishments are systematically taxed and low productivity establishments are systematically subsidized so that the policy entails a systematic redistribution of resources from more to less productive establishments. In this case they found that the effects on aggregate TFP could be several times larger. A key message is that for a source of misallocation to have large effects it is important that it systematically depress inputs at high productivity producers. In particular, studies that focus on identifying misallocation in relatively small and less productive enterprises may not be particularly relevant in terms of aggregate effects.

The Indirect Approach

The Restuccia-Rogerson paper illustrated the potential for factors that generate misallocation to have significant impacts on aggregate TFP. But it was silent on how to connect their hypothetical policy configurations with data to estimate the nature and amount of misallocation present in actual economies. Two subsequent papers—Hsieh and Klenow (2009) and Bartelseman et al (2013)—addressed this issue.

Whereas Restuccia and Rogerson imposed ad hoc specifications of distortions to input decisions, Hsieh and Klenow (2009) noted that the distortions could be uncovered directly given appropriate micro data. Their procedure essentially follows the strategy noted in the previous section, but extended to a setting in which each producer produces a distinct variety, with varieties aggregated via a constant elasticity of substitution aggregator. They also assume that each producer behaves as a monopolistic competitor when choosing capital and labor. The demand structure implied by the CES aggregator is important in allowing them to infer TFP when the data set only includes total revenue as opposed to physical output.

[1] Some discussion of this magnitude is in order. On the one hand, if a country such as the US were to experience a one-time drop in TFP equal to ten percent, no one would consider such a change to be modest, as in terms of output changes this would generate losses that rival those associated with the Great Depression. But if one is trying to account for differences in TFP levels between rich and poor countries as uncovered by development accounting exercises, one requires TFP differences on the order of a factor of 4 or more. From this perspective, a 10% change in TFP is indeed modest.

Hsieh and Klenow apply their method to four-digit manufacturing industries in China, India and the United States. They find large effects of misallocation on TFP. In particular, if all misallocation were eliminated, manufacturing TFP would increase by 86-110% in China, 100-128% in India, and 30-43% in the United States. Two features stand out: first, even in the United States they find large productivity losses associated with misallocation. Second, they find that the less developed economies of China and India have much greater productivity losses associated with misallocation than does the United States. Taken at face value, the message is clear: misallocation is quantitatively important and it is an important factor in accounting for differences in measured TFP across rich and poor countries.

In interpreting these magnitudes it is important to note that these estimates are for the manufacturing sector and not the overall economy. Available evidence suggests that TFP differences in manufacturing tend to be much smaller than aggregate TFP differences. Hsieh and Klenow estimated that TFP differences in manufacturing between the United States and China and India during the relevant period are on the order of 130 and 160%, in contrast to differences on the order of 300 and 600% at the aggregate level.

The Hsieh and Klenow approach measures misallocation without identifying the source of the misallocation. Nonetheless, their analysis does allow them to examine how the nature of misallocation is correlated with various observables. One of their strongest findings in this regard is that state ownership in China is strongly correlated with misallocation in that state-owned firms are much larger than efficiency would dictate. Another important finding is that the identified wedges are strongly correlated with plant-level productivity and that the correlation is stronger in India than in the United States.[2]

A large number of other studies have applied the Hsieh and Klenow (2009) methodology to estimate misallocation in a variety of settings. But before we discuss these it is important to revisit some of the caveats noted at the end of the previous section.

Limitations of the Indirect Approach

The indirect approach essentially assumes a production structure and then uses the data to estimate wedges in the first order conditions that characterize an efficient allocation. This approach interprets the measured wedges as reflecting distortions to real allocations. But there are two other interpretations that are a priori plausible. The first is that the wedges might be interpreted as reflecting model

---

[2] Bento and Restuccia (2016) corroborate this finding for a larger set of developing countries: the extent to which more productive plants face greater implicit taxes is strongly related to GDP per capita across countries. This property of empirical distortions may be important in helping identify the fundamental sources of misallocation and is critical for the broader implications of misallocation we discuss in Section 6.

misspecification—that is, maybe the allocation is efficient but the researchers have incorrectly specified the production structure. In fact, one could always find production functions that would rationalize the observed cross-sectional distribution of capital and labor as an efficient allocation. Second, given that the wedges are estimated using actual data, they also reflect measurement error in the data. Addressing these alternative interpretations in a compelling fashion is a key challenge for this literature.

To be sure, Hsieh and Klenow were aware of these issues and devote considerable effort to addressing them. For example, their benchmark results assume that all producers within a sector use the same Cobb-Douglas production function. It follows that in an efficient allocation, capital to labor ratios will be equated across all producers, implying that any variation in capital to labor ratios will be interpreted as misallocation. An alternative interpretation is that capital shares in the Cobb Douglas production function are heterogeneous across producers. In the extreme, all differences in capital to labor ratios reflect heterogeneity in producer level production functions rather than misallocation. But Hsieh and Klenow show that even with this alternative interpretation there are still large effects of misallocation on productivity.

Another issue is adjustment costs. A voluminous literature on estimating adjustment costs using establishment-level data routinely finds evidence that they are substantial, both for labor and capital. Optimal allocations in a model that features separable adjustment costs and transitory shocks will necessarily find that marginal products of capital and labor in production are not equated in the cross-section in a static sense. This raises the possibility that the wedges found in Hsieh and Klenow are simply picking up the effects of adjustment costs and transitory shocks. Being mindful of this issue, Hsieh and Klenow's preferred interpretation of their findings is to focus on the differences in misallocation across economies rather than the levels per se. The idea is that some amount of "base level" misallocation is best understood as the result of adjustment costs or some other misspecification, and that a reasonable starting point is to assume that this level is the same across economies. This moderates their estimates of the amount of misallocation somewhat—if China and India were to reduce misallocation to the level found in the United States, TFP would increase by 31-51% and 40-59% respectively. While smaller than the earlier values, these magnitudes are still large. In particular, they imply that misallocation can account for almost half of the observed TFP differences in manufacturing.

But is it reasonable to think that all economies would have some common level of measured misallocation that should be ignored in this context? Asker et al (2014) argues in the negative. They argue that differences in misallocation as measured by Hsieh and Klenow can be understood as being consistent with optimal allocations in a world of adjustment costs on capital coupled with transitory firm level shocks in which idiosyncratic shock processes are more variable in poorer countries. We believe two important caveats are in order here. First, one needs to think about why idiosyncratic shocks are more variable in poorer countries—if these shocks reflect randomness in factors such as policy or regulation, then one would want to think of the greater uncertainty as a component of misallocation.

Second, this interpretation highlights the need to examine misallocation using panel data at the establishment level instead of a single cross-section or repeated cross-sections without a panel component. The reason is that measured misallocation that is due to adjustment costs will generate very specific time series patterns. More generally, with panel data one could carry out the Hsieh-Klenow type exercise on specifications that explicitly include adjustment costs. We believe that this type of analysis is an important priority for future research in the area.

Hsieh and Klenow also carry out several calculations to assess the possibility that their results on differences in misallocation across countries are not dominated by differences in measurement error across countries. While not conclusive in ruling out this possibility, several of their calculations do not support such an interpretation.

Against the backdrop of these issues we think this is a useful point to describe the method of Bartlesman et al (2013). The key insight of their paper is that firm-specific taxes and subsidies (that are again viewed as representing a wide array or idiosyncratic distortions) will affect various cross-sectional moments. In particular, they focus on the covariance between size and productivity, a moment that is intuitively very much affected by firm-specific taxes and subsidies. A key feature of their analysis is to start with a specification that implies cross-sectional differences in marginal products even in an efficient allocation, so that moments of the US cross-sectional data on revenue productivity dispersion and employment are consistent with efficiency. They use their model calibrated to the United States to assess the amount of misallocation in manufacturing in a sample of eight economies (the United States, the United Kingdom, France, Germany, Netherlands, Romania, Hungary, and Slovenia) during the 1990s. Rather than inferring the actual distortions faced by each firm in their dataset they instead focus on inferring a statistical representation of these distortions that can match salient moments. Relative to the United States, they find that the effect of misallocation on TFP ranges from 3% in Germany to 12% in Romania. Their choice of countries was dictated by the desire to have data that was consistently collected across countries. Unfortunately, this implies that their sample is almost exclusively represented by developed countries. This makes it difficult to conclude whether misallocation is an important factor in accounting for differences between rich and poor countries. Romania is the one relatively poor country in their sample and it did have the largest effect of misallocation, though the size of the effect is somewhat modest. Nonetheless, we believe that this alternative method for assessing the amount of misallocation is potentially interesting, and think it would be very relevant to compare results across methods for a given country.

Misallocation in Different Countries and Sectors

The basic message that Hsieh and Klenow found from their analysis of manufacturing in China, India and the United States has been confirmed by several studies that include other countries. Busso, Madrigal, and Pages (2013) carry out a comparable analysis of manufacturing in ten Latin American countries and conclude that differences in misallocation between these economies and the United States is an important source of TFP gaps in manufacturing.

Kalemli-Ozcan and Sorensen (2012) study misallocation of capital among private manufacturing firms in 10 African countries. The sample sizes in their analysis are quite small and so not so directly comparable to the other results presented. Their sample also includes firms from India, Ireland, Spain, and South Korea that can be used as benchmarks. Subject to the caveat of small sample sizes, they find that capital misallocation in Africa is significantly higher than in developed countries, though not as severe as in India.

It is important to note that all of the above results pertain to the manufacturing sector. While manufacturing is a significant sector in many economies, it is particularly small in the world's poorest economies, including many in Africa. While one might suspect that finding high misallocation in manufacturing is suggestive of high misallocation in other sectors, there is also reason to think that there could be large differences across sectors. With this in mind it is of interest to ask what we have learned about differences in misallocation across countries in other sectors.

There have been relatively few papers that address misallocation of inputs in the service sector. The study by Busso et al (2013) mentioned above did include some analysis of specific service sectors, such as retail. Their basic finding was that misallocation in service sectors were much larger than in manufacturing. Devries (2010) finds very large misallocation in the retail sector in Brazil. Dias et al (2015), studies misallocation in manufacturing and services in Portugal an also find that misallocation is much larger in services. One limitation of these studies is that they do not include a benchmark such as the US economy. If misallocation measures for the US are also larger in service sectors than in manufacturing then it is not clear if misallocation differences are indeed more important in service sectors. Also, we think it is important to note that an important measurement issue in the context of the service sector is that we have much worse measures of output. Prominent examples include education, health care, and the banking sector.

The agricultural sector is of particular importance in comparing the world's richest and poorest economies because the poorest countries are much less productive in agriculture than in non-agriculture compared to the richest countries and most of their labor is allocated in agriculture (Gollin et al 2002; Restuccia et al 2008). In particular, Caselli (2005) reports that aggregate output per worker differences between rich and poor countries are dominated by differences in agriculture. Looking at 90/10 ratios, factor differences in output per worker were 22 at the aggregate level, 4 in non-agriculture, and 45 in agriculture. In addition, there is a long list of specific policies and institutions in the agricultural sector in developing countries that can potentially create misallocation. Adamopoulos and Restuccia (2014) document and analyze some of these policies in connection to farm size differences across countries. They found that misallocation could potentially account for farm size and productivity differences in agriculture between rich and poor countries.

An early article that focused on one dimension of misallocation in the agricultural sector in a poor country is Udry (1996). He studied the extent to which land was misallocated within families in Burkina Faso. He found that the productivity losses from this type of misallocation were 6%. There was no rich country to be used as a benchmark, but given the large differences in agricultural TFP across countries, this difference is quite modest.

Restuccia and Santaeulalia-Llopis (2015) offer a much broader assessment of misallocation in agriculture in a poor country. They study misallocation across household farms in Malawi, a poor country that has a large share of its labor force engaged in agriculture. They have data on outputs and inputs as well as measures of transitory shocks and so are able to measure farm level TFP. They find that the allocation of inputs is relatively constant across farms despite large differences in measured TFP, suggesting a large amount of misallocation. In fact, they found that aggregate agricultural output would increase by a factor of 3.6 if inputs were allocated efficiently. These magnitudes are several times larger than those which Hsieh and Klenow found for manufacturing in China and India.

They also carry out some analysis to suggest that institutional factors that impact on land allocation are likely playing a key role. Specifically, they can compare misallocation within different groups of farmers that are differentially influenced by restrictive land markets. Whereas most farmers in Malawi operate a given allocation of land, other farmers have access to marketed land (in most cases informal rentals). This allows Restuccia and Santaeulalia-Llopis to contrast the amount of misallocation within different groups. They find that misallocation is much larger for the group of farmers without access to marketed land so that the output gains are 2.6 times larger in this group relative to the gains for the group of farms with rented land.

Other studies also document misallocation in agriculture. For instance, Adamopoulos et al (2016) study the case of China, where the land market is severely restricted by the household responsibility system. This system is based on an individual registration "hukou" in a village. Land ownership and allocation resides with the collective village and use rights are distributed uniformly among the household members registered in the village. While there are no explicit restrictions on rentals in China, implicit "use or lose it" rules prevail for fear of redistribution and this, in addition to other provisions in the system, severely limits the mobility of individuals and households across sectors and space. In this context, farm operational scales are essentially limited to the use rights of land for each household and hence, not surprisingly, the authors find that land allocations are unrelated to farm productivity. In particular, eliminating misallocation in this context is found to increase agricultural productivity by a factor of 1.84 fold, with 60% of this gain arising from reallocation of factors across farms within villages. Exploiting the panel dimension of the data to remove potential transitory variation in farm productivity, the authors show that reallocation gains are still substantial, representing 81-86% of the cross-sectional productivity gains.[3]

Chen et al. (2016) study the case of Ethiopia where land is also customary and households are allocated use rights at the local level. The authors exploit regional variation in the extent of rented land that was created by differential implementation of a land certification program that started in the early 2000s. Land certification allows a more flexible scale of operation for farmers. The authors found that regions with more land rentals are associated with substantially less misallocation: a one percent higher share of land rental is associated with a 1.7 percentage point increase in agricultural productivity.

Misallocation over Time

The results that we have described so far focused on assessing the potential for differences in misallocation across countries at a point in time to account for differences in productivity across countries at a point in time. This fits well with the motivation of assessing the role of misallocation in accounting for differences in productivity levels across countries in the cross-section. But it is also of interest to ask whether changes in misallocation over time within a country are an important source of changes in productivity over time within a country. This is akin to connecting misallocation with growth accounting. In the time series domain one might be interested in either high frequency changes or low frequency changes. Several papers have pushed in this direction in a variety of different contexts.

The literature has identified changes in misallocation as an important component of low frequency movements in TFP in three distinct contexts: the rapid TFP growth in Chile following the crisis of the early 1980s, and the low TFP growth in Japan starting in the early 1990s and in Southern Europe following the adoption of the Euro. Chen and Irrazabal (2015) show that misallocation decreased during Chile's decade long period of growth following the crisis of the early 1980s and was an important part of productivity growth during this time. Fujii and Nozawa (2013) show that capital misallocation in manufacturing became more pronounced after 1990 in Japan, a period characterized by poor productivity growth. Gopinath et al (2015) study the evolution of input misallocation over time using micro firm-level data from several European countries to show increased capital misallocation and

[3] Adamopoulos et al (2016) also analyze how the distortions implicit in the land market institution affect the sector choice of individuals. Because the more productive farmers are especially hindered by the restricted land markets, as it severely limit their operational scale, distorted occupational choices further depress productivity in the agricultural sector via selection effects. For instance, the authors find that removing distortions in the agricultural sector lead to an increase in agricultural productivity of more than 10 fold compared to the static gain of 1.8 fold from static misallocation.

roughly constant labor misallocation in Southern European countries at the time these countries joined the Euro in 1999.[4] Note that TFP changes in the time series dimension tend to be much smaller than those in the cross-section, so in these cases even modest effects of misallocation are sufficient for misallocation to play a dominant role in the context of the time series changes observed in the data. We think Ziebarth (2013) is an interesting analysis of long run changes in the context of the United States. In particular, he found that misallocation levels among US manufacturers in the late 19th century were similar to those in present day India and China.

We think a promising avenue for further study in the context of time series changes is to focus on changes in misallocation during periods in which important policy or regulatory changes occurred that one might reasonably believe have important effects on misallocation. Using the indirect approach in such settings provides an opportunity to produce suggestive evidence regarding the importance of specific underlying sources. Hsieh and Klenow (2009) took a first step in this direction. For China they found a decrease in misallocation during the period of 1998 to 2005. The finding that misallocation improved in China is consistent with the view that various reforms enacted during this period served to lessen the importance of distortions. Time series patterns can differ by sector.[5]

In contrast, Hsieh and Klenow found that misallocation in India worsened over the period from 1987 to 1994, a result which seems puzzling given the nature of reforms enacted there. An important reform in India involved the elimination of the license "raj" system, a system of controls on the entry of firms into the manufacturing sector, which arguably would have contributed to reallocation and productivity growth in India. Bollard et al (2013) pursued this further, though focusing only on very large firms. Although this period witnessed rapid productivity growth for their sample of firms, they find that little of the productivity growth was due to changes in misallocation. There of course multiple interpretations of this finding; perhaps the raj system is not an important source of misallocation among large firms, or perhaps it is not even an important source of misallocation overall. Alternatively, as noted earlier, the Hsieh-Klenow method might not be isolating true misallocation. More work is needed to further refine our conclusions in this regard.

The paper by Bartleseman et al (2013) described earlier also included a time series component. They found that misallocation decreased over the period of the 1990s in the transition economies of Eastern Europe. This is also consistent with the notion that increased market reforms were leading to less misallocation, but the extent of

---

[4] See also the discussion in Reis (2014) and the analysis in Dias et al (2015) for the case of Portugal, and Calligaris et al (2015) for Italy.

[5] Interestingly, despite widespread reform in other sectors, land market institutions have remained essentially the same in China, and Adamopoulos et al (2016) found that misallocation in the agricultural sector in China has remained roughly constant for the period of study (1993-2002).

the changes is somewhat modest, on the order of TFP gains of a few percentage points.

Several papers have assessed the extent to which misallocation changes at business cycle frequencies, focusing on fairly dramatic episodes such as crises or protracted recessions. Oberfield (2013) studies misallocation changed in Chile during the crisis of the early 1980s, Sandleris and Wright (2014) examine misallocation in Argentina during its crisis in the early 2000s, and Ziebarth (2015) assessed misallocation during the US Great Depression. All of these authors find that misallocation increased sharply in each of these episodes and accounted for a large part of measured drops in aggregate TFP.

In our view changes in misallocation measures at business cycle frequency need to be treated with extreme caution. As emphasized earlier, these measures can be heavily influenced by adjustment costs that may give rise to factor hording. To us it remains very much an open question of whether true misallocation of resources increases during these types of episodes.

5. Causes of Misallocation: The Direct Approach

Taken at face value, the findings based on the indirect approach indicate that misallocation is an important factor behind low TFP observed in less developed countries. We also noted some reasons to regard these findings as somewhat tentative. The case for the importance of misallocation would undoubtedly be bolstered by compelling analyses that isolate the cause of the misallocation. Moreover, this is exactly the type of analysis that is required if one wants to offer advice to policy makers regarding how to boost TFP by decreasing the amount of misallocation. In this section we discuss the efforts that have been taken to provide this type of analysis. All of the studies described in this section necessarily follow the direct approach. Our focus is on analyses that address aggregate consequences. Rather than examine very context-specific sources of misallocation, the studies we describe focus on somewhat broad sources of misallocation, following the categorization laid out in Section 2. While we think it is important for this literature to make contact with context-specific sources of misallocation, we regard these analyses as a very important first step in helping us assess rough orders of magnitude.

The Role of Regulation and Discretionary Provisions

We begin with studies that focus on misallocation due to regulation. One of the earliest examples is the analysis of firing costs in Hopenhayn and Rogerson (1993). Firing costs are a specific type of adjustment cost, and as discussed in the previous section, adjustment costs in general will lead one to measure misallocation of factors in a given cross-section. In contrast to the previous discussion, however, in this case the adjustment costs are themselves due to policy, and the resulting variation in marginal products does reflect true misallocation. Using a quantitative version of the model in Hopenhayn (1992), these authors find that firing costs equal to one year's wages will lead to steady state productivity losses of roughly 2%. While this magnitude is significant in the context of differences among OECD economies, it is relatively small in the context of differences between rich and poor economies.[6]

Guner et al (2008) study the misallocation effects of what they call size-dependent policies. This refers to a range of policies that implicitly tax larger firms, where size might be measured in sales, labor or capital. This specification covers examples of both labor and product market regulations, including, for example, regulations that only become effective beyond some employment threshold, outright restrictions on the number of employees, or restrictions on the amount of physical space that a firm may operate, such as the size of a retail store. They use a Lucas span-of-control model to analyze simple but abstract versions of such policies, calibrated so as to achieve a given difference in average firm size. While they find that such policies can have a large effect on measures such as the number of firms and the firm size distribution, they find relatively small effects on TFP.[7]

State owned enterprises are a relatively easy to identify form of what we called discretionary provisions. Such enterprises are not that prevalent in all economies but one important exception is China. The misallocation of resources between private and state-owned enterprises is a key source of productivity losses in the analysis of Song et al (2011). More recently, Brandt et al (2013) study the importance of misallocation across state and non-state sectors and across provinces over time in China for the non-agricultural sector. They find that misallocation reduces non-agricultural TFP by an average of 20 percent for the period 1985-2007. More than half of this productivity loss is due to within province misallocation of capital between state and non-state sectors. While across province distortions remain fairly constant over time, there is increased state/non-state capital misallocation between 1998 and 2007.

Several papers have assessed how the inability of the government to enforce regulations can lead to the emergence of an informal sector that is effectively unregulated. Leal (2014) calibrates a model using data from Mexico that assumes firms can avoid regulation by choosing to hire capital below a certain threshold. He finds that making enforcement uniform would increase TFP by slightly more than 4%.

[6] In related work, Lagos (2006) studies a Mortensen-Pissarides matching model to study how labor market policies such as unemployment insurance and employment protection affect productivity via selection effects. He finds that small changes in labor-market policy parameters (replacement rate and firing taxes) generate decreases in aggregate TFP on the order of 2-3%.

[7] In related work, Gourio and Roys (2014) and Garicano et al (forthcoming) study the effects of size-dependent labor regulations using plant-level data from France where firms with 50 or more employees face substantial additional labor regulations.

We previously noted the importance of restrictions on land markets in the work of Adamopoulos and Restuccia (2014) and Restuccia and Santaeulalia-Lopis (2015). There is a long tradition in development economics emphasizing property rights as a key institution shaping resource allocation and productivity, for example the classic work of Besley (1995) linking property rights institutions and investment incentives. In particular, Besley and Ghatak (2010) survey the work on property rights and development, emphasizing the importance of property rights institutions for resource allocation and productivity growth. Insecure and ill-defined property rights discourage productive investments and impose severe barriers of resource reallocation to best uses. A specific regulation in the land market relates to land reforms which have been prevalent in developing countries (see, for instance, De Janvry 1981; Banerjee 1999; and Deininger and Feder 2001). Land reforms are often associated with a maximum cap to farm size and explicit restrictions to land markets aimed at redistributing the excess land from the cap to landless and smallholder households. Adamopoulos and Restuccia (2015) study a comprehensive land reform with such characteristics in the Philippines using a quantitative model and panel micro data of farms that cover the period before and after the reform. They find that on impact the land reform had a substantial negative effect on farm size and agricultural productivity (reductions of 34% and 17% respectively). The negative productivity effects stem from both a selection effect and the misallocation of resources across heterogeneous farmers. Government intervention in the land market had a substantial impact on these outcomes, as a market-oriented reform would have generated only 1/3 of the reductions in farm size and productivity. They also note that full enforcement of the farm size cap would have resulted in a doubling of the reduction in agricultural productivity.[8]

Market Imperfections

Next we consider studies that emphasize market structure. One extensive literature within this category is that of the role of trade on resource allocation and growth. But market structure also broadly relates to industrial policies, patent protection, and competition policies. The trade literature has studied the effect of trade policy on aggregate productivity through the lens of model structures that extend the work of Eaton and Kortum (2002) and Melitz (2003). Other studies have tackled the issue of trade and productivity directly exploiting relevant variation in the data. A prominent example of this literature is Pavcnik (2002). She studies the effect of liberalized trade on aggregate productivity in Chile during an episode of substantial reductions in trade barriers that exposed plants to foreign competition using microeconomic panel data. Pavcnik isolates the contribution of trade to productivity growth by exploiting the variation in outcomes between plants in the import competing/export oriented sectors and plants in the non-traded sectors. She finds that trade had a substantial positive effect on plant growth: plants in import-

[8] de Janvry et al (2015) study a different land reform in Mexico in the 90s whereby farmers were given ownership certificates of land, removing the pre-existing link between land rights and land use, and show substantial labor and land reallocations associated with the reform.

competing sectors grew 3-10% more than plants in the non-traded sector. Reallocation of resources across producers contributed substantially to aggregate growth via reallocation of resources from less to more efficient plants and through plant exit because exiting plants were on average 8% less productive than continuing plants. Her analysis stresses both misallocation and selection.

A few studies have investigated the quantitative impact of heterogeneity in markups on misallocation. Epifani and Gancia (2011) show that dispersion of markups across manufacturing industries is significantly greater in poorer countries than in richer countries, though they do not carry out a quantitative assessment of what this implies for cross-country differences in TFP. Edmonds et al (2015) calibrate a model to Taiwanese manufacturing data and find that moving from autarky to free trade decreases markup heterogeneity and leads to an increase in TFP of slightly more than 12%. Note that moving to free trade merely serves to reduce markup heterogeneity and does not eliminate them. So this calculation isolates the potential for free trade to reduce misallocation of this type.

Lastly, we consider the case of incomplete markets. Financial market imperfections are a prominent example within this group, and these imperfections are perhaps the single most studied source of misallocation. A robust empirical finding is that financial market development is positively correlated with output per capita. (See, for example, the survey article of Levine (1997).)

The literature on financial market development and economic development is too large for us to discuss in any detail. Our focus in this article is on the subset of papers in this literature that have focused on quantifying the misallocation of capital across producers due to credit constraints. Recent contributions include Buera et al (2011), Greenwood et al (2013), Midrigan and Xu (2013), and Moll (2014). This literature has generated a range of estimates, some of them quite large. It is now well understood that the effects depend on various model features, specifically the scope for individuals to accumulate assets in order to grow out of financial constraints. This in turn is heavily influenced by the persistence of productivity (or demand) at the producer level. As the literature has made more attempts to model this feature and discipline it using micro data, the steady state effects of capital misallocation on TFP have diminished. Midrigan and Xu find that the magnitude of this effect is no more than about ten percent. Quantitatively the largest effects on productivity from financial frictions seem to come from the distortion in occupational choices made by entrepreneurs with differential amounts of collateral capital, which reflect what we had called selection effects.

We previously noted that Gopinath et al (2015) provide empirical evidence of increased capital misallocation in Southern European countries following the introduction of the Euro in 1999. They show that a large part of this misallocation can be accounted for by their quantitative firm dynamics model with financial frictions and capital adjustment costs when subjected to a drop in the real interest rate. The magnitude of the effects that they generate is on the order of a 3 percent drop in TFP.

Other relevant institutional frictions include imperfect information and imperfect insurance. For example, David et al (2016) study the role of information frictions in generating misallocation and productivity loses. They identify information frictions combining production and stock market data of firms and find that these types of frictions can reduce aggregate productivity by 7-10% in China and India. Imperfect insurance and credit restrictions have also played prominently in development economics (see, for instance, Udry 2012, for a survey of the relevant literature). Munshi and Rosenzweig (2016) emphasize risk and differential insurance arrangements between rural and urban sectors in restricting labor mobility, therefore potentially generating labor misallocation across space.

In addition to misallocation within narrowly defined industries, misallocation of inputs can occur across sectors or across space. There are fewer studies in this area including the work of Munshi and Rosenzweig just described. Hsieh and Moretti (2015) study misallocation of individuals across 220 US metropolitan areas from 1964 to 2009. They document a doubling of the dispersion in wages across US cities during the sample period. Using a simple model of spatial reallocation, they show that the increase in wage dispersion across US cities contributed to a loss in aggregate GDP per capita of 13.5%. They argue that across-city labor misallocation is directly related to housing regulations and the associated constraints in housing supply. Fajgelbaum et al (2015) study the role of state taxes in spatial misallocation in the United States. Using a spatial reallocation framework and changes in taxes across US states over time, they estimate how spatial allocation (workers and firms) responds to US state taxes. They find that eliminating tax dispersion across US states produces modest increases in output although the output loses from greater dispersion in taxes than observed in the US system can be large.

Our goal in this section was to discuss the various efforts made to assess the aggregate importance of misallocation attributed to several categories of distortions, particularly with an eye toward asking whether we could isolate factors that might generate misallocation of the magnitude found using the indirect method of Hsieh and Klenow (2009). In this regard, we view the current state as somewhat disappointing. The existing literature has identified some factors that can account for large effects of misallocation in agriculture. But it has yet to identify any particular factor which can account for the magnitudes of misallocation found in manufacturing. One possibility is that the effects estimated by Hsieh and Klenow are overestimates of the extent of differences in misallocation. Alternatively, it is possible that the aggregate effects are the result of many different individual factors, each of which contributes a small part, so that we will never isolate a single dominant factor. Or perhaps the existing analyses, based on relatively simple models and somewhat generic treatments of potential sources of misallocation, are not adequate for the purposes at hand and more refined analyses will generate larger effects. In this regard we would note as one example that the somewhat generic models of collateral constraints which have dominated the assessment of financial market frictions may not adequately capture the full extent of frictions that are present in less developed counties.

6. Additional Consequences of Misallocation

As stressed earlier, an immediate consequence of misallocation is a decrease in productivity: misallocation implies that less output is produced given the level of inputs. Taking the findings from the literature at face value, estimates of these costs are large.

But in this section we want to describe a recent strand of the literature that finds that the costs of misallocation might indeed be significantly larger than indicated by these direct costs. Recall the simple example from the introduction. At that time we argued that there were three very different rationalizations for cross country differences in TFP: differences in the A_i's across countries (technologies), differences in the set of operating producers (selection), or differences in the extent to which inputs are efficiently allocated across producers with a given profile for the A_i's. This last category was defined to correspond to misallocation. In this section we want to argue that in a dynamic setting in which the evolution of the A_i's is at least partly endogenous, factors that induce misallocation are also likely to have negative implications for the distribution of the A_i's either via changes in the selection of operating producers or the technologies used by those producers or both.

A simple example serves to illustrate the intuition. Imagine a dynamic extension of our simple model that includes entry and exit of producers. A standard assumption in these models of firm dynamics (see, e.g., Hopenhayn (1992)) is that a new entrant has initial productivity that is an *iid* random draw from some distribution with density $g(A)$. A simple extension of this would allow for an effort or resource cost that a potential entrant pays upfront, with greater up front expenditure yielding a draw from a better distribution. That is, if x is the expenditure on entry costs, then the density that one draws an initial productivity from is $g(A,x)$, where it is natural to assume that the distribution displays first order stochastic dominance with respect to expenditure x.

In this setting, the expenditure on x is influenced by the extent to which higher x yields higher values of A and how they translate into profits. Holding all else constant, a higher value of A will imply higher profits. But policies that induce misallocation can impact this relationship. If, for example, more productive firms implicitly or explicitly face higher taxes or more costly regulations then this will serve to diminish the incentives for expenditure on x. This in turn will translate into a worse distribution of A's in the economy. Restuccia (2013) provides an early example of this approach studying the productivity gap between Latin America and the US.

One paper of particular importance in this literature is the recent paper by Hsieh and Klenow (2014) on the life cycle of manufacturing plants in India, Mexico and the United States. This paper begins by noting that the effect of age on size of manufacturing plants is much more pronounced in the United States than it is in either India or Mexico. From a cross-sectional perspective they find that only a small part of this reflects misallocation, i.e., the key difference between the countries is that older plants in India and Mexico are much less productive relative to younger plants than is the case in the United States. Given that older plants in India and Mexico are not so much more productive than their younger counterparts, it is efficient for them to not be much larger. But this leads to the question of why the gradient of productivity with age is less in Mexico and India than it is in the United States. They then go on to show that if one models investment in productivity improvements at the establishment level, then the greater implicit taxes faced by more productive establishments in India and Mexico can potentially account for a large share of the differences in productivity gradients with age.

Bento and Restuccia (2016) study the effect of misallocation on establishment sizes across countries in a model with endogenous productivity that is closely related to Hsieh and Klenow's work. Establishments invest in their productivity at entry and over their life cycle. The authors use their tractable framework to decompose the effects of static misallocation and entry-level and life-cycle productivity investment on aggregate productivity differences across countries. They find that the greater implicit taxes faced by more productive establishments in India compared to the US reduces aggregate productivity by 53% and average establishment size by 86%. This compares to a 15% reduction in establishment size in Hsieh and Klenow with only life cycle investment. The reduction in aggregate productivity is roughly equally shared between static misallocation and entry-level productivity. Life-cycle productivity plays a minor role in amplifying the productivity differences because the reduction in life-cycle productivity growth is offset by its effect on establishment entry.

Da Rocha et al (2016) study the effect of firing costs on productivity in the spirit of the analysis by Hopenhayn and Rogerson (2003) discussed earlier but in a model that includes an endogenous choice for innovation. They find that the dynamic effects on productivity are substantial, increasing the TFP loss from static misallocation of around 2% to an overall effect of 20%.

Peters (2016) studies a model in which limited competition leads to heterogeneity in markups. He embeds this in a model of innovation and shows that the dynamic effect of markup heterogeneity is more than four times larger than the static misallocation effects.

Other papers that have recognized this feedback from misallocation to the determination of firm level TFPs include Bello et al (2011), Bhatacharya et al (2013), Gabler and Poschke (2013), and Ranasinghe (2014).

There are other approaches to study the broader consequences of misallocation. We note in particular two. The first group can be described as an attempt to connect misallocation with barriers to technology adoption and adoption lags across countries. Ayerst (2016) represents early work in this direction. The basic insight is that the policies and institutions that generate misallocation may create disincentives to adopt the most modern and best technologies and so this work provides an explicit connection between the technology and misallocation channels described earlier. The second group can be described as assessing the consequences of misallocation in environments with explicit input-output linkages as emphasized in the survey article of Jones (2013).

In summary, the recent literature has demonstrated that factors that generate misallocation may ultimately have much larger effects on TFP than are reflected in static misallocation effects as measured by Hsieh and Klenow (2009). We believe this has important implications for interpreting existing work and the direction of future work. Factors that generate misallocation may prove to be much more important than indicated by static empirical measures. Whereas much of the existing methodology has focused on measuring the static effects of misallocation, the work described here suggests that we need to focus on measuring dynamic effects. Panel data will be critical to producing compelling empirical evidence.

7. Conclusions/Where to From Here?

Understanding the determinants of the large differences in TFP across countries and their evolution over time is an extremely difficult yet important task. An active area of research in growth economics has emphasized the role of resource allocation across heterogeneous production units as a potentially important factor in accounting for productivity differences.

To take stock, we think is useful to revisit the three questions we posed in the introduction. First, how important is misallocation? Our perspective of the literature is that the answer to this question is that misallocation is a substantial channel in accounting for productivity differences but that the magnitude of the effects depends crucially on the approach followed and the specific context. Our description of the relevant studies has highlighted the advantages and disadvantages of the two broad approaches—that we labeled as the direct and indirect approach—that the literature has followed to assess the impact of misallocation on aggregate productivity and emphasized the substantial gap in the magnitude of the effects found using these two approaches. Whereas the indirect approach encapsulates all the factors that create misallocation without identifying the source, the direct approach starts by identifying and measuring a specific source of misallocation. It is perhaps not surprising then to find that the productivity losses from misallocation reported using the indirect approach are typically an order of magnitude or more larger than the loss associated with specific policies and institutions reported using the direct approach. Part of this magnitude difference may reflect measurement error and model misspecification that can potentially plague the indirect approach method. Second, what are the causes of misallocation? While the literature has identified a number of specific policies and institutions generating misallocation, it is clear from the different analyses that there is not a single source of misallocation that can explain the dominant share of productivity differences in the data, and that each of many specific factors contributes a small part of the overall effect. In this context, the direct approach, which is more amenable to isolating key sources of misallocation, may be somewhat more challenging to implement. But our view is that studies that follow the direct approach are more likely to reach concrete, persuasive, and specific conclusions of practical policy relevance. Nevertheless, we also see the appeal of the indirect approach, especially as it relates to identifying important dimensions of misallocation, whether within industries or across industries; whether related to specific factors of production such as capital, labor or land; among others. Third, are there additional costs to misallocation? The answer is clearly yes and the literature has already looked beyond the misallocation channel into factors that also impact the selection and technology channels we discussed in the introduction. Whereas much of the literature has focused on static misallocation, we think the dynamic effects of misallocation deserve much more attention going forward. Further progress in this area will require new methods and more data.

Although the existing literature has not yet delivered definitive answers, we do believe that the existing work does suggest an important role for misallocation, and we are optimistic that future work will further refine our estimates of its effects. We see promising work following the direct approach in at least in two areas. First, more work is needed on the various mechanisms that can potentially amplify the effect of misallocation on aggregate productivity and in particular in connecting policies that generate misallocation with observed micro productivity distributions. Second, a great appeal of the direct approach is its connection to changes in policies and institutions over time. The increasing availability of micro data sets, especially panel data sets of firms and households, are likely to yield opportunities to exploit changes in policies and institutions and variations across individuals, firms, regions, and other relevant dimensions, to study the role misallocation in those specific contexts.

Another exciting direction for future work relates to broader notions of misallocation. To serve as illustration of potentially important dimensions, we briefly note recent papers that focus on aspects of misallocation that are somewhat broader than those we have studied, but which we think are also promising new directions. We have focused on how the volume of labor and capital might be misallocated across firms in a given industry. Similarly but closely related is the allocation of individuals across occupations or tasks. In this context, discrimination, culture, and social norms can prevent the best allocation of talent across employment status, occupations, and sectors and thus act as a source of misallocation. Two papers have explored this quantitatively. Hnatkovska et al (2013) document the misallocation of talent in India that arises as a result of the caste system, a system of social stratification that has historically restricted an individual's access to education and occupation opportunities. They document that these barriers have decreased dramatically over the last twenty years.

Hsieh et al (2014) similarly illustrate the disparity in the distribution of talent across occupations in the United States, but go one step further and quantify the contribution of its convergence in the last 50 years for productivity growth. For example, in 1960 around 94 percent of doctors and lawyers were white men whereas by 2008 the share declined to 62 percent. To the extent that innate talent is unlikely to feature such a concentration across gender and races, the occupational distribution in 1960 reflects misallocation of talent and the observed convergence represents an improvement in the allocation. Even for developed economies misallocation can be an important source of macroeconomic growth. They estimate that convergence in the occupational distribution across races and gender can account for 15 to 20 percent of growth in aggregate output per worker in the United States between 1960 and 2008. We think this work suggests a promising direction for much additional work on the allocation of talent and how it differs across economies.

Overall, we are encouraged by the progress the literature has made in its assessment of misallocation as an important factor accounting for productivity differences and the emergence of relevant sources of misallocation. We hope that our perspectives on the progress, challenges, and potential fruitful directions of the literature in this article will facilitate additional work on this very relevant topic.

References

Adamopoulos, Tasso, and Diego Restuccia. "The size distribution of farms and international productivity differences." The American Economic Review 104, no. 6 (2014): 1667-1697.

Adamopoulos, Tasso, and Diego Restuccia. "Land Reform and Productivity: A Quantitative Analysis with Micro Data." Unpublished manuscript, University of Toronto, (2015).

Adamopoulos, Tasso, Loren Brandt, Jessica Leight, and Diego Restuccia. "Misallocation, Selection and Productivity: A Quantitative Analysis with Panel Data from China." Unpublished manuscript, University of Toronto (2016).

Asker, John, Allan Collard-Wexler, and Jan De Loecker. "Dynamic inputs and resource (mis) allocation." Journal of Political Economy 122, no. 5 (2014): 1013-1063.

Ayerst, Stephen. "Policy Distortions and Technology Adoption." Manuscript, University of Toronto, 2016.

Bhattacharya, Dhritiman, Nezih Guner, and Gustavo Ventura. "Distortions, endogenous managerial skills and productivity differences." Review of Economic Dynamics 16, no. 1 (2013): 11-25.

Baily, Martin N., Charles Hulten, and David Campbell. "Productivity dynamics in manufacturing plants." Brookings papers on economic activity. Microeconomics 1992 (1992): 187-267.

Banerjee, Abhijit V. "Land reforms: prospects and strategies." In Conference Paper, Annual World Bank Conference on Development Economics, Washington DC, pp. 99-24. 1999.

Banerjee, Abhijit V., and Esther Duflo. "Growth theory through the lens of development economics." Handbook of Economic Growth 1 (2005): 473-552.

Banerjee, Abhijit V., and Esther Duflo. "Do firms want to borrow more? Testing credit constraints using a directed lending program." The Review of Economic Studies 81, no. 2 (2014): 572-607.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. "Cross-country differences in productivity: The role of allocation and selection." The American Economic Review 103, no. 1 (2013): 305-334.

Beaman, Lori, Dean Karlan, Bram Thuysbaert, and Christopher Udry. Self-selection into credit markets: Evidence from agriculture in mali. No. w20387. National Bureau of Economic Research, 2014.

Bello, Omar D., Juan S. Blyde, and Diego Restuccia. "Venezuela's growth experience." Latin american journal of economics 48, no. 2 (2011): 199-226.

Bento, Pedro, and Diego Restuccia. Misallocation, Establishment Size, and Productivity. No. tecipa-557. 2016.

Besley, Timothy. "Property rights and investment incentives: Theory and evidence from Ghana." journal of Political Economy (1995): 903-937.

Besley, Timothy, and Maitreesh Ghatak. "Property rights and economic development." In Handbook of Development Economics edited by Dani Rodrik and Mark Rosenzweig. Vol 5 (2009).

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. "Does Management Matter? Evidence from India." The Quarterly Journal of Economics 128, no. 1 (2013): 1-51.

Bollard, Albert, Peter J. Klenow, and Gunjan Sharma. "India's mysterious manufacturing miracle." Review of Economic Dynamics 16, no. 1 (2013): 59-85.

Brandt, Loren, Trevor Tombe, and Xiaodong Zhu. "Factor market distortions across time, space and sectors in China." Review of Economic Dynamics 16, no. 1 (2013): 39-58.

Buera, Francisco J., Joseph P. Kaboski, and Yongseok Shin. "Finance and Development: A Tale of Two Sectors." The American Economic Review 101, no. 5 (2011): 1964.

Busso, Matías, Lucía Madrigal, and Carmen Pagés. "Productivity and Resource Misallocation in Latin America." The BE Journal of Macroeconomics 13, no. 1 (2013): 903-932.

Caballero, Ricardo J., Takeo Hoshi, and Anil K. Kashyap. "Zombie lending and depressed restructuring in Japan." The American Economic Review 98, no. 5 (2008): 1943-1977.

Calligaris, Sara. "Misallocation and Total Factor Productivity in Italy: Evidence from Firm - Level Data." Labour 29, no. 4 (2015): 367-393.

Caselli, Francesco. "Accounting for cross-country income differences." Handbook of economic growth 1 (2005): 679-741.

Caselli, Francesco, and Nicola Gennaioli. "Dynastic management." Economic Inquiry 51, no. 1 (2013): 971-996.

Chen, Chaoran, Diego Restuccia, and Raul Santaeulalia-Llopis. Land Markets, Resource Allocation, and Agricultural Productivity. Unpublished manuscript, University of Toronto (2016).

Chen, Kaiji, and Alfonso Irarrazabal. "The role of allocative efficiency in a decade of recovery." Review of Economic Dynamics 18, no. 3 (2015): 523-550.

Da-Rocha, José-Marıa, Marina Mendes Tavares, and Diego Restuccia. Firing costs, misallocation, and aggregate productivity. Working Paper, University of Toronto, 2015.

David, Joel M., Hugo A. Hopenhayn, and Venky Venkateswaran. "Information, Misallocation, and Aggregate Productivity." The Quarterly Journal of Economics 131, no. 2 (2016): 943-1005.

De Janvry, Alain. "The role of land reform in economic development: policies and politics." American Journal of Agricultural Economics 63, no. 2 (1981): 384-392.

De Janvry, Alain, Kyle Emerick, Marco Gonzalez-Navarro, and Elisabeth Sadoulet. "Delinking land rights from land use: Certification and migration in Mexico." The American Economic Review 105, no. 10 (2015): 3125-3149.

De Mel, Suresh, David McKenzie, and Christopher Woodruff. "Returns to capital in microenterprises: evidence from a field experiment." The Quarterly Journal of Economics (2008): 1329-1372.

Deininger, Klaus, and Gershon Feder. "Land institutions and land markets." Handbook of agricultural economics 1 (2001): 287-331.

Dias, Daniel, Carlos Robalo Marques, and Christine Richmond. Misallocation and productivity in the lead up to the Eurozone crisis. No. w201411. 2014.

Dias, Daniel, Carlos Robalo Marques, and Christine Richmond. "Comparing misallocation between sectors in Portugal." Economic Bulletin and Financial Stability Report Articles (2016).

Eaton, Jonathan, and Samuel Kortum. "Technology, geography, and trade." Econometrica 70, no. 5 (2002): 1741-1779.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu. "Competition, markups, and the gains from international trade." The American Economic Review 105, no. 10 (2015): 3183-3221.

Epifani, Paolo, and Gino Gancia. "Trade, markup heterogeneity and misallocations." Journal of International Economics 83, no. 1 (2011): 1-13.

Fajgelbaum, Pablo D., Eduardo Morales, Juan Carlos Suarez Serrato, and Owen M. Zidar. State taxes and spatial misallocation. No. w21760. National Bureau of Economic Research, 2015.

Field, Erica. "Property rights, community public goods, and household time allocation in urban squatter communities: evidence from Peru." Wm. & Mary L. Rev. 45 (2003): 837.

Foster, Lucia, John Haltiwanger, and Chad Syverson. "Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?." The American Economic Review 98, no. 1 (2008): 394-425.

Fujii, Daisuke, and Yoshio Nozawa. Misallocation of Capital During Japan's Lost Two Decades. Working Paper, 2013.

Gabler, Alain, and Markus Poschke. "Experimentation by firms, distortions, and aggregate productivity." Review of Economic Dynamics 16, no. 1 (2013): 26-38.

Garicano, Luis, Claire Lelarge, and John Van Reenen. Firm size distortions and the productivity distribution: Evidence from France. American Economic Review, forthcoming.

Gollin, Douglas, Stephen Parente, and Richard Rogerson. "The role of agriculture in development." The American Economic Review 92, no. 2 (2002): 160-164.

Gopinath, Gita, Sebnem Kalemli-Ozcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez. Capital allocation and productivity in South Europe. No. w21453. National Bureau of Economic Research, 2015.

Gourio, François, and Nicolas Roys. "Size‐dependent regulations, firm size distribution, and reallocation." Quantitative Economics 5, no. 2 (2014): 377-416.

Greenwood, Jeremy, Juan M. Sanchez, and Cheng Wang. "Quantifying the impact of financial development on economic development." Review of Economic Dynamics 16, no. 1 (2013): 194-215.

Guner, Nezih, Gustavo Ventura, and Yi Xu. "Macroeconomic implications of size-dependent policies." Review of Economic Dynamics 11, no. 4 (2008): 721-744.

Hall, Robert E., and Charles I. Jones. "Why Do Some Countries Produce So Much More Output Per Worker Than Others?." The Quarterly Journal of Economics 114, no. 1 (1999): 83-116.

Heckman, James, and Carmen Pagés. "Introductory Chapter." Law and Employment: Lessons from the Latin America and the Caribbean (2004): 1-108.

Hnatkovska, Viktoria, Amartya Lahiri, and Sourabh Paul. "Castes and Labor Mobility." American Economic Journal: Applied Economics 4, no. 2 (2012): 274-307.

Hopenhayn, Hugo A. "Entry, exit, and firm dynamics in long run equilibrium." Econometrica (1992): 1127-1150.

Hopenhayn, Hugo, and Richard Rogerson. "Job turnover and policy evaluation: A general equilibrium analysis." Journal of political Economy (1993): 915-938.

Hopenhayn, Hugo A. "Firms, misallocation, and aggregate productivity: A review." Annu. Rev. Econ. 6, no. 1 (2014): 735-770.

Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow. The allocation of talent and us economic growth. No. w18693. National Bureau of Economic Research, 2014.

Hsieh, Chang-Tai, and Peter J. Klenow. "Misallocation and Manufacturing TFP in China and India." The Quarterly Journal of Economics 124, no. 4 (2009): 1403-1448.

Hsieh, Chang-Tai, and Peter J. Klenow. "Development accounting." American Economic Journal: Macroeconomics 2, no. 1 (2010): 207-223.

Hsieh, Chang-Tai, and Peter J. Klenow. "The Life Cycle of Plants in India and Mexico." The Quarterly Journal of Economics 129, no. 3 (2014): 1035-1084.

Hsieh, Chang-Tai, and Enrico Moretti. Why Do Cities Matter? Local Growth and Aggregate Growth. No. w21154. National Bureau of Economic Research, 2015.

Jones, Charles I. "Misallocation, Input-Output Economics, and Economic Growth, volume II of Advances in Economics and Econometrics-Tenth World Congress." (2013).

Kalemli-Ozcan, Sebnem, and Bent E. Sørensen. "Misallocation, property rights, and access to finance: evidence from within and across Africa." In African Successes: Modernization and Development. University of Chicago Press, 2014.

Karlan, Dean, Robert Osei, Isaac Osei-Akoto, and Christopher Udry. "Agricultural Decisions after Relaxing Credit and Risk Constraints*." Quarterly Journal of Economics 129, no. 2 (2014).

Kirwan, Barrett E., Shinsuke Uchida, and T. Kirk White. "Aggregate and farm-level productivity growth in tobacco: before and after the quota buyout." American Journal of Agricultural Economics 94, no. 4 (2012): 838-853.

Lagos, Ricardo. "A model of TFP." The Review of Economic Studies 73, no. 4 (2006): 983-1007.

Leal Ordóñez, Julio. "Tax collection, the informal sector, and productivity." Review of Economic Dynamics 17, no. 2 (2014): 262-286.

Levine, Ross. "Financial development and economic growth: views and agenda." Journal of economic literature 35, no. 2 (1997): 688-726.

McKenzie, David, and Christopher Woodruff. "Experimental evidence on returns to capital and access to finance in Mexico." The World Bank Economic Review 22, no. 3 (2008): 457-482.

Melitz, Marc J. "The impact of trade on intra ‐ industry reallocations and aggregate industry productivity." Econometrica 71, no. 6 (2003): 1695-1725.

Midrigan, Virgiliu, and Daniel Yi Xu. "Finance and Misallocation: Evidence from Plant-Level Data." American Economic Review 104, no. 2 (2014): 422-58.

Moll, Benjamin. "Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?." The American Economic Review 104, no. 10 (2014): 3186-3221.

Munshi, Kaivan, and Mark Rosenzweig. "Networks and misallocation: Insurance, migration, and the rural-urban wage gap." The American Economic Review 106, no. 1 (2016): 46-98.

Oberfield, Ezra. "Productivity and misallocation during a crisis: Evidence from the Chilean crisis of 1982." Review of Economic Dynamics 16, no. 1 (2013): 100-119.

Olley, G. Steven, and Ariel Pakes. "The Dynamics of Productivity in the Telecommunications Equipment Industry." Econometrica (1996): 1263-1297.

Pavcnik, Nina. "Trade liberalization, exit, and productivity improvements: Evidence from Chilean plants." The Review of Economic Studies 69, no. 1 (2002): 245-276.

Peek, Joe, and Eric S. Rosengren. "Unnatural selection: Perverse incentives and the misallocation of credit in Japan." The American Economic Review 95, no. 4 (2005): 1144-1166.

Peters, Michael. Heterogeneous mark-ups, growth and endogenous misallocation. Manuscript, Yale University, 2016.

Prescott, Edward C. "Lawrence R. Klein lecture 1997: Needed: A theory of total factor productivity." International economic review (1998): 525-551.

Ranasinghe, Ashantha. "Impact of policy distortions on firm-level innovation, productivity dynamics and TFP." Journal of Economic Dynamics and Control 46 (2014): 114-129.

Reis, Ricardo. "The Portuguese Slump and Crash and the Euro Crisis." Brookings Papers on Economic Activity (2013): 143.

Restuccia, Diego. "The Latin American Development Problem: An Interpretation." Economía 13, no. 2 (2013): 69-100.

Restuccia, Diego, Dennis Tao Yang, and Xiaodong Zhu. "Agriculture and aggregate productivity: A quantitative cross-country analysis." Journal of Monetary Economics 55, no. 2 (2008): 234-250.

Restuccia, Diego, and Richard Rogerson. "Policy distortions and aggregate productivity with heterogeneous establishments." Review of Economic dynamics 11, no. 4 (2008): 707-720.

Restuccia, Diego, and Richard Rogerson. "Misallocation and productivity." Review of Economic Dynamics 16, no. 1 (2013): 1-10.

Restuccia, Diego, and Raul Santaeulalia-Llopis. Land Misallocation and Productivity. No. tecipa-533. 2015.

Sandleris, Guido, and Mark LJ Wright. "The costs of financial crises: Resource misallocation, productivity, and welfare in the 2001 argentine crisis." The Scandinavian Journal of Economics 116, no. 1 (2014): 87-127.

Song, Zheng, Kjetil Storesletten, and Fabrizio Zilibotti. "Growing like china." The American Economic Review 101, no. 1 (2011): 196-233.

Syverson, Chad. "What determines productivity?." Journal of Economic literature 49, no. 2 (2011): 326-365.

Udry, Christopher. "Gender, agricultural production, and the theory of the household." Journal of political Economy (1996): 1010-1046.

Udry, Christopher. "Misallocation, Growth and Financial Market Imperfections." SED plenary lecture (2012).

Vries, Gaaitzen J. "Productivity in a distorted market: the case of Brazil's retail sector." Review of Income and Wealth 60, no. 3 (2014): 499-524.

Ziebarth, Nicolas L. "Are China and India backward? Evidence from the 19th century US Census of Manufactures." Review of Economic Dynamics 16, no. 1 (2013): 86-99.

Ziebarth, Nicolas L. "The Great Depression through the eyes of the Census of Manufactures." Historical Methods: A Journal of Quantitative and Interdisciplinary History 48, no. 4 (2015): 185-194.