

Misallocating Misallocation?

Maitreesh Ghatak¹ and Dilip Mookherjee²

¹Department of Economics, London School of Economics, London, United Kingdom;
email: m.ghatak@lse.ac.uk

²Department of Economics, Boston University, Boston, Massachusetts, USA;
email: dilipm@bu.edu

ANNUAL REVIEWS CONNECT

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

Annu. Rev. Econ. 2025. 17:511–38

First published as a Review in Advance on
May 2, 2025

The *Annual Review of Economics* is online at
economics.annualreviews.org

<https://doi.org/10.1146/annurev-economics-091624-051237>

Copyright © 2025 by the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See credit lines of images or other third-party material in this article for license information.

JEL codes: O11, O12, D23, D24, D61, Q15



Keywords

misallocation, welfare, market failure, second-best, developing countries

Abstract

The macrodevelopment literature on misallocation quantifies aggregate productivity losses resulting from microeconomic distortions, relative to a first-best benchmark. However, sources of these distortions are often insufficiently explored. The microdevelopment literature in contrast provides evidence of distortions resulting from market failures owing to asymmetric information, missing markets, transaction costs, and limited state capacity, often without examining the resulting macrolevel consequences. If distortions result from such market failures rather than policies, second-best welfare-improving policies may aggravate productive misallocation. We illustrate these points in the context of manufacturing, agriculture, and rural-urban allocation of labor and land. Hence future research should devote more effort to identifying the source of distortions and using appropriate benchmarks for welfare going beyond productivity.

1. INTRODUCTION

The rapidly growing macrodevelopment literature on misallocation focuses on the role of microeconomic distortions in lowering aggregate productivity and per capita income across countries or over time. It aims to quantify the role of such microeconomic frictions in explaining variations in economic development as distinct from the role of macroeconomic differences in factor endowments or access to technology. Most of the models rely on an exogenous specification of wedges or distortions in an otherwise first-best setting with perfect markets. Some effort has also been devoted to identifying and estimating the role of specific sources of these frictions, such as government policies or market frictions. Apart from their value in isolating the role of purely microeconomic distortions in explaining variations in per capita income, estimated models are frequently used in evaluating productive and welfare effects of counterfactual variations in government policy (Section 2 provides an overview of this literature).

In this article we review this literature from the lens of microdevelopment economics, which studies frictions leading to market failures common in underdeveloped countries, such as asymmetric information, moral hazard, economies of scale, missing or incomplete markets, and weak enforcement of contracts and property rights due to low state capacity. Barring a few exceptions, most macrodevelopment papers abstract from these market frictions. On the other hand, the microdevelopment literature typically abstracts from possible government failures such as suboptimal taxes or regulations that might form an alternative source of observed misallocation patterns. Not many microdevelopment papers explore welfare properties of an explicit underlying model, and ones that do frequently treat policy in purely normative terms. As a consequence, policy implications of models used by the two respective literatures often differ substantially, as we elaborate in some detail in this paper.

Our aim is to highlight these contrasts between the two literatures and discuss resulting implications for fruitful future research directions. Our point of departure is to pose two related conceptual problems when faced with the evidence of some form of misallocation.

The first problem is model identification—in any given setting, we need to discriminate between potential alternative explanations of observed productive misallocation patterns. Are they rooted in weak market institutions that are endemic features of underdevelopment? Should some of the underlying market failures be treated as constraints at par with economic fundamentals such as tastes, technology, and endowments? Or are the distortions more fruitfully viewed as the consequence of suboptimal government policies in an otherwise first-best market economy? Can data patterns suggestive of second-best market frictions also be explained by unobserved heterogeneity in a first-best world? Addressing this identification problem is necessary to obtain a correct diagnosis of the underlying source of productive inefficiency.

This brings us to the second problem, namely, that of welfare implications. This, in turn, is closely related to selecting policies to alleviate misallocation. The reason why the identification problem is important is that in a second-best world, reducing productive misallocation may not necessarily imply higher welfare. As a result, the welfare and policy implications could depend on the specific mechanism at work. This is a version of the theory of the second-best—if there are multiple frictions, a change in some policy (such as removing a distortion) may shift welfare and aggregate productivity in different directions. A policy change that lowers productive misallocation would improve welfare in a first-best world, but in a second-best world it could have the opposite impact. We provide a number of examples of this in the current review.

Other than the differences in the conceptual approaches of micro- and macrodevelopment, the research methodologies commonly employed in the two literatures are also distinct. The empirical microdevelopment literature devotes considerable attention to causal identification of policy effects or underlying mechanisms by using controlled or quasi-experimental methods, using data at

a highly disaggregated level, and utilizing institutional details of the specific context. However, an explicit model is often lacking; general equilibrium, welfare, and dynamic effects or counterfactual analyses are frequently ignored. In contrast, the macrodevelopment literature typically uses an explicit model incorporating dynamic general equilibrium effects. At the same time, relatively little effort is devoted to validating the model via causal identification methods; instead, validation is secured by calibrating the model to fit key moments in the data. The models and data incorporate less fine-grained micro and institutional details.

These assessments suggest the need for future research to blend macrodevelopment and microdevelopment approaches, specifically by enlarging the range of possible alternative explanations for empirically observed patterns, understanding their respective welfare and policy implications, and gauging their relative empirical plausibility in any given setting. Suitable quantitative methodologies need to be developed in order to calibrate or estimate second-best models on par with first-best models and derive welfare effects of counterfactual policies. As we describe below, a number of recent papers have begun to appear along these lines, a very welcome trend that we hope will continue.

The paper is structured as follows. Following a brief overview of the macrodevelopment literature in Section 2, Section 3 recapitulates some of the general welfare properties of second-best economies characterized by distortions or wedges. Section 4 illustrates the central concerns with identification and welfare in a specific context: cross-country comparisons of distributions of firm size, productivity, and wedges in manufacturing. We describe the principal stylized facts documented by Ayerst et al. (2024) and the model they use to explain these patterns by the existence of distorting productivity-based taxes in a setting without any market failure. We show how an alternative model based on capital market frictions based on Buera et al. (2021) in conjunction with local scale economies for small firms generate predictions that are also consistent with these facts. The welfare and policy implications of the two competing models differ sharply: For example, while progressive size-dependent taxes necessarily lower welfare in the Ayerst et al. (2024) model, they raise welfare in the capital friction model in a wide range of circumstances. In a similar vein, Section 5 considers the implications of an alternative source of market failure in an industrial context: the existence of productivity or cost spillovers across firms associated with agglomeration or social networks. We describe a model of entry, firm size, and productivity that is consistent with existing empirical evidence from developing countries. The welfare analysis of the model illustrates how misallocation as measured in the standard way can be a misleading basis for evaluating industrial policy interventions.

To highlight the broader relevance of these considerations, Section 6 discusses microdevelopment literature that focuses on the role of financial market frictions and imperfect enforcement of property rights on misallocation in other settings: agriculture, and allocation of land and labor between rural and urban sectors.

We conclude in Section 7 by reviewing the recent literature that has begun to address the identification and second-best welfare questions. In this section, as well as previous ones, we draw attention to a few select papers that we are familiar with, and make no effort to provide a comprehensive survey. To highlight the contrast between first-best and second-best models, we use a utilitarian welfare measure that abstracts from distributional concerns or long-term sustainability. Such concerns are obviously important, especially for policy, but incorporating them would widen the scope of this review beyond what is currently feasible.

2. OVERVIEW OF MACRODEVELOPMENT LITERATURE

A number of papers provide a broad overview of the macrodevelopment literature, tracing its origin to earlier strands of the growth literature using total factor productivity (TFP) variations

to explain cross-country income differences, and the industrial organization literature on the size distribution of firms with heterogeneous productivity. Hopenhayn (2014) reviews the literature on misallocation building on a model of distribution of firms that differ by idiosyncratic productivity, and focuses on three margins that affect output per capita: the number of firms per capita and the distribution of firm productivity and of factor inputs. These can result from distortions stemming from entry barriers, financial constraints, firm heterogeneity, policies, and institutional constraints. Restuccia & Rogerson (2017) provide an overview of the literature on misallocation focusing on distortions in the allocation of a given amount of capital and labor across heterogeneous producers and evaluate the causes and consequences of misallocation for productivity differences across countries. In an earlier paper (Restuccia & Rogerson 2013), they provide a broader review of the misallocation literature. A recent paper by Buera et al. (2023) offers a perspective on how micro- and macrodevelopment approaches can be combined by using rigorous microlevel experimental evidence to calibrate and validate structural macro models.

In terms of specific approaches, starting with Restuccia & Rogerson (2008) and Hsieh & Klenow (2009), one set of papers embeds idiosyncratic wedges into a structural model without specifying the underlying sources and estimates the resulting (static or dynamic) TFP and output loss in manufacturing or agricultural sectors. Another set of papers focuses on a specific source (ranging from regulations of firms, property rights, trade, or competition to various financial and contractual frictions) and estimates their quantitative impact on productive misallocation. A third set of papers studies robustness of misallocation estimates to measurement errors, ex post shocks, unobserved heterogeneity, and misspecification of technology (adjustment costs or functional form of production functions). These papers usually utilize the panel data structure to control for farm/plot/plant fixed effects, besides minimizing the impact of mismeasurement or idiosyncratic shocks on misallocation estimates. However, analysis of interactions between different types of distortions or an attempt to diagnose the most important distortion is generally lacking. While some papers mention the complementary or amplification effect of removing multiple frictions, second-best arguments are rare (we highlight a few exceptions). Most papers are silent on the welfare or distributional implications of misallocation.

2.1. Misallocation in Manufacturing

Hsieh & Klenow (2009) model misallocation as dispersion of marginal revenue productivity (MRP) across firms, using a first-best benchmark with specific structural assumptions on production functions and joint distribution of heterogeneity and distortive wedges. They compare the resulting measure of misallocation in manufacturing sector across three countries—the United States, China, and India—and find that moving to US efficiency would raise aggregate productivity by 30–50% in China and 40–60% in India. While they are not explicit about the specific source of such misallocation, they examine how measured misallocation covaries with policy distortions, for example, the share of state ownership of plants in China, and delicensing of industry and size restrictions in India. They also discuss a few alternative explanations for MRP dispersion such as measurement error, markups, adjustment costs, other investments, and heterogeneous capital shares.

2.2. Misallocation in Agriculture

Chen et al. (2023) utilize a household dataset in Malawi to measure TFP in farms and quantify the loss of agricultural productivity due to land misallocation. The detailed panel dataset allows them to control for land quality and rainfall shock to agricultural production and include household-farm fixed effects terms to minimize the impact of transitory shocks and measurement errors in TFP calculation. They document that the actual factor allocations are unrelated to farm

productivity, in contrast with the requirement of productive efficiency wherein more productive farms should have higher operational scales of land and capital in order to equate marginal products of factors across farms. Furthermore, they suggest limited land markets in Malawi are a possible source for such misallocation, and argue that a reform of efficient factor allocation would lead to reduction of income inequality and poverty.

Adamopoulos et al. (2022) examine the role of land market distortions in rural China in generating factor misallocation across farmers and affecting patterns of selection of farmers into agriculture. As is common in this literature, they define an efficient factor allocation as one that maximizes total agricultural output, and model farm-specific distortions as idiosyncratic input and output wedges. The panel dataset allows them to calculate farm-level productivity controlling for farm fixed effects and shows that they covary little with factor inputs. They attribute these distortions to egalitarian land allocation institutions and limited land rental markets in China. Furthermore, they embed their estimates of factor misallocation into a model of occupational choice between agriculture and nonagriculture and use it to argue that selection amplifies the effect of distortions in factor allocations on agricultural productivity. Besides reducing interfarm factor misallocation, their model predicts that removing these farm-specific distortions would attract high-ability farmers to work in agriculture—thereby leading to sizable productivity gains.

2.3. Financial Frictions

Buera et al. (2011) focus on effects of financial frictions combined with sector-specific fixed costs of entrepreneurship on aggregate TFP, prices, output, and firm size distribution. Financial frictions affect not only the allocation of capital among entrepreneurs but also the selection of entrepreneurs. Their quantitative results suggest that, owing to these frictions, low-ability but wealthy entrepreneurs remain in business—resulting in lower average talent among active entrepreneurs. Furthermore, there are too few entrepreneurs and excessively large establishments in the manufacturing sector, while the service sector exhibits opposite patterns. These findings are consistent with the missing middle phenomenon documented in developing economies and empirical facts on establishment size and scale in Mexico and the United States.

Cavalcanti et al. (2024) document high average levels and high dispersion in credit spreads among Brazilian firms and use this to motivate a model in which financial frictions include borrowing limits (as in Buera et al. 2011), intermediation costs (arising from screening, monitoring, and collection costs incurred by lenders, which vary across borrowers of differing productivity and asset levels), and market power of the financial intermediaries. They calculate the loss of output per capita, TFP, and labor wage due to the calibrated financial frictions. Their simulation exercises show that market power plays a less important role than intermediation costs in generating high interest rate spreads. Eliminating one friction has smaller impacts when other frictions are present. They argue that interest rate spreads are a key source of credit market imperfections and generate larger aggregate impacts compared to a model where borrowing constraints constitute the only financial friction. Moreover, younger firms are more constrained by these frictions.

Gopinath et al. (2017) document increased dispersion of returns to capital and decline in TFP across Spanish manufacturing firms between 1999 and 2012. They explain this using a small open economy model with size-dependent financial frictions. Declining real interest rates induce capital inflows that are disproportionately directed to less productive but financially unconstrained firms, resulting in a larger dispersion of MRP of capital between financially constrained and unconstrained firms.

Bau & Matray (2023) exploit the staggered liberalization of access to foreign capital across industries in India to explore its effect on reducing capital and labor misallocation across firms and increasing Solow residuals in treated industries. Modeling misallocation as wedges of input

prices, they find liberalization reforms led to reductions in labor and capital misallocation. These effects were the strongest in areas with less developed local banking sectors, suggesting the role of domestic banking sector inefficiencies in generating capital misallocation.

2.4. Other Sources of Misallocation

Boehm & Oberfield (2020) discuss the implications of weak contract enforcement (due to court congestion) in the organization of production and aggregate productivity in the Indian manufacturing sector. They document that in states with more congested courts and in industries that typically rely on relationship-specific intermediate inputs, cost shares of intermediate inputs will be lower, these input bundles are more tilted toward standardized inputs, and plants tend to have larger vertical spans of production. With a general equilibrium model featuring input-output linkages and enforcement distortion, they estimate that reducing court congestion (to the level of the least congested state) would lead to a 4% increase of aggregate productivity.

Baqae & Farhi (2019) estimate the misallocation resulting from firm-level price markups in US manufacturing between 1997 and 2014 and find that it accounts for about 50% of aggregate TFP growth during this period. With a quantitative model of economic geography, Fajgelbaum et al. (2019) show that tax dispersion under the decentralized tax system in the United States led to aggregate output and welfare losses (compared to the harmonized state taxes benchmark) as workers and firms reallocated in response to these dispersions.

2.5. Misspecification and Measurement Error

A number of papers have highlighted the role of specific assumptions regarding technology and demand conditions made by Hsieh & Klenow (2009) to derive their misallocation measure, and how the measure may be erroneous when these assumptions are violated. Asker et al. (2014) argue that capital adjustment costs in a dynamic setting imply that cross-sectional MRP dispersion can arise owing to firm-specific productivity shocks even in a frictionless economy. Using data spanning 40 countries, they show that industries exhibiting greater time series volatility of productivity shocks have greater MRP dispersion. A structurally estimated investment model with adjustment costs turns out to explain 80–90% of observed cross-industry and cross-country MRP dispersion. Haltiwanger et al. (2018) show that the validity of the Hsieh & Klenow (2009) measure depends on knife-edge assumptions of unit demand elasticity and flat marginal cost curves, which imply factor revenue products do not vary in response to TFP differences. These assumptions are empirically rejected from evidence from various product markets and different countries. Consequently MRP dispersion can arise even in frictionless economies. These papers have sparked efforts in the literature to use different measures of misallocation that are valid under weaker model assumptions. For instance, the Hsieh & Klenow (2009) approach is substantially generalized by Baqae & Farhi (2019) to obtain measures of misallocation that do not depend on specific functional forms or distributional assumptions, besides incorporating arbitrary input-output network linkages.

The role of data measurement error in generating biased misallocation measures has been emphasized by various authors. Gollin & Udry (2020) utilize plot-level panel data from farms in Tanzania and Uganda to distinguish production shocks, measurement errors, and unobserved characteristics such as land quality (which altogether account for 70% of the productivity dispersion) from misallocation, based on the presumption that farmers should not face any constraints on allocating resources among their own plots.¹ Corrections for late-season agricultural shocks,

¹However, they consider a few within-farmer frictions that constrain allocation of factors among plots cultivated by the same farmer, such as varying input costs for different plots, land tenure restrictions, and joint farming. But they find their results are mostly robust to these frictions.

measurement error, and heterogeneity in input (land) quality result in substantially lower estimates of TFP dispersion and of output gains that could be realized from hypothetical reallocations. Aragon et al. (2024) also address concerns about measurement error in the context of African agriculture. They find reallocation gains in the Ugandan agriculture sector would be substantially higher if these are predicted on the basis of plot- rather than farm-level data. However, using different methods from Gollin & Udry (2020), they obtain a lower estimate of dispersion that can be attributed to measurement error, implying that large output losses result from interfarm factor misallocation.

In manufacturing contexts, Rotemberg & White (2021) highlight the role of data processing methods employed by statistical agencies responsible for collecting and processing data. For instance, the US Census edits outliers and imputes missing values, resulting in thinner upper and lower tails in measured firm revenue products in the reported data. They show this results in a significant drop in measured total factor revenue productivity (TFPR) dispersion in US manufacturing. Consequently, measured misallocation in reported data can vary across countries owing to differences in data cleaning and processing procedures by their respective statistical agencies.

2.6. Disentangling the Role of Different Sources of Misallocation

David & Venkateswaran (2019) provide an empirical decomposition of sources of variations in average revenue product of capital, including capital adjustment costs, information frictions (imperfect knowledge about firm fundamentals), and firm-specific distortions such as unobserved heterogeneity in markups and technology, size-dependent policies, and financial frictions. This is one of very few papers that attempt to empirically distinguish different sources of misallocation. Using a recursive framework with the firm's capital investment decision where the firm-specific distortions consist of a component correlated with productivity, besides transitory and permanent components, they show that the key parameters determining dispersion in average revenue products are uniquely identified by a set of empirical moments. Variations in markups and technologies explain a significant amount of measured misallocations in the United States, whereas institutional distortions (e.g., size/productivity-dependent factors) are the dominant driver of misallocation in China. However, the data they use do not permit them to distinguish between the respective roles of size-dependent government policies and financial market imperfections. While they address the robustness of their estimates to nonconvex adjustment costs, distortions in labor choice, and measurement error, their methodology does not allow them to study welfare effects of alternative policies.

3. WELFARE ECONOMICS OF THE SECOND-BEST: RECAPITULATION

The earliest formal analyses of optimal government policies in a second-best setting go back to Diamond & Mirrlees (1971). They consider a government with distributional objectives embodied in a Paretian social welfare function, in a context where it does not have the information required to achieve distributional goals via lump-sum transfers. This necessitates commodity and income taxes that drive a wedge between producer and consumer prices. Under some strong assumptions (capacity to tax every sector, constant returns to scale, and a first-best economy in all other respects) they show that second-best welfare optima involve production efficiency, that is, an absence of wedges between firms.² The result is driven by the property that a move toward the production frontier allows the government to use its fiscal policy tools to achieve a Pareto improvement. In this setting, reductions in productive misallocation translate into higher levels of welfare.

²The Diamond & Mirrlees (1971) result holds in the presence of linear taxes. It is generalized further by Atkinson & Stiglitz (1976) in a context where nonlinear income taxes can be imposed. Assuming utility is

Subsequent work by Stiglitz & Dasgupta (1971) showed that the Diamond-Mirrlees Production Efficiency Theorem no longer holds when the government's capacity to tax certain sectors or agents is restricted. They show that when commodity or factor taxes cannot be imposed in certain industries, optimal taxation usually implies differential taxes (e.g., depending on the elasticity of substitution) and the abandonment of productive efficiency. For example, when there are pure profits from production and the government cannot impose 100% taxes on these profits, production efficiency is no longer desirable: An optimal taxation structure would imply differential factor taxes because it can serve as a substitute for a profit tax. Stiglitz & Dasgupta (1971) also discuss the cases of monopolistic industries and uniform commodity taxes (owing to administrative costs of distinguishing between different types of income) where production efficiency might not be desirable. In the same spirit, Emran & Stiglitz (2005) show that with a large informal economy where value-added taxes (VATs) cannot be collected by the government, replacement of trade taxes by VATs could reduce welfare. While reducing trade taxes would reduce production distortions between tradable and nontradable sectors, the corresponding increase in VATs would increase the distortion between formal and informal sectors; the welfare cost of the latter could overwhelm the benefits of the former.

The preceding literature works with models that are first-best in all respects, apart from the restrictions on the government's capacity to use lump sum transfers to achieve distributional goals. In particular, they assume absence of asymmetric information, a full set of markets, price-taking behavior, convex technology, and the absence of externalities. When insurance markets are missing, Newbery & Stiglitz (1984) provide a striking example where opening up to free trade might be Pareto inferior to autarky, as this may raise risks borne by producers that induce them to shift toward less risky products that hurt consumers. While their example relies on very special assumptions, it serves to highlight the broader point that, in the absence of insurance markets, trade restrictions that generate production misallocation may yield insurance benefits whose welfare effects need to be traded off against the costs of lower productivity.

More general results concerning the failure of the First Welfare Theorem in incomplete market economies are provided by Geanakoplos & Polemarchakis (1986), who show that competitive equilibria are generically constrained Pareto inefficient in the following sense. If assets are traded ex ante before states of nature are realized, followed by spot commodity markets, and asset markets are incomplete, there exist asset reallocations that would generate ex ante Pareto improvements for almost all configurations of endowments and household utility functions.³ A similar result is provided by Stiglitz (1982) for an economy with stock markets but lacking a complete set of Arrow securities. The results are driven by pecuniary externalities generated by the effect of asset reallocations on spot market commodity prices, which vary across agents with heterogeneous endowments and/or risk attitudes (assuming nonhomothetic utility). The diversity of these welfare impacts allows a social planner to design asset reallocations that generate a Pareto improvement.

Greenwald & Stiglitz (1986) use a general framework to illustrate that in a second-best world with some general forms of externalities, government interventions with commodity taxes typically generate Pareto improvements. The optimal tax rule equates the marginal gain from reducing externalities through taxes to the marginal deadweight loss from distortions caused by the taxes. This framework is applied to settings where the externality could be generated by adverse selection, signaling and screening, moral hazard, or incomplete markets.

weakly separable between leisure and consumption, they show second-best optima can be attained by nonlinear income taxes alone, with no consumption or production taxes.

³Formally, they assume asset returns do not have the spanning property that enables agents to reallocate purchasing power across all states of nature. The result requires there be enough contingent commodities relative to the number of households.

The preceding discussion highlights other dimensions relevant to welfare assessments in second-best economies such as insurance, informational externalities, and learning spillovers that need to be incorporated in welfare analyses apart from effects on productive efficiency. This indicates that exclusive focus on production distortions may be too narrow in assessing welfare impacts of counterfactual government policies.

Nevertheless, these arguments are abstract and sometimes driven by special assumptions that may not be empirically relevant. They could be subject to the broader concern that in second-best contexts, anything can happen, with the absence of definite results concerning whether and how productive misallocation diverges from welfare loss. This indicates a need to consider specific settings and examine welfare properties of models suitable for those settings. The following sections focus on these issues in a number of specific settings.

4. MISALLOCATION IN MANUFACTURING: POLICY DISTORTIONS VERSUS CAPITAL MARKET FRICTIONS

Much of the existing macrodevelopment literature has focused on misallocation in the manufacturing sector, in particular on cross-country comparisons of misallocation and their quantitative significance. In this section we describe some stylized facts documented by Ayerst et al. (2024) and a first-best model of policy distortions they use to explain these facts. We then show that the same stylized facts are also consistent with a second-best model of capital frictions of the kind proposed by Buera et al. (2011), in conjunction with technology nonconvexities. We also contrast the welfare and policy implications of the two models. Since our purpose is to highlight these qualitative contrasts in a simple and stark fashion, we describe static versions of the two models.

4.1. The Stylized Facts

Ayerst et al. (2024) use evidence from data on manufacturing firms in 28 countries covering a wide range of the world income distribution over the period 2000–2019 to document the following stylized facts:

1. Average firm size is lower, and firm-level TFP is more dispersed in less developed countries (LDCs).
2. Larger TFP dispersion arises mostly due to greater prevalence of low-productivity firms in LDCs.
3. There is a higher dispersion of distortions [or wedges, measured by average product of labor, closely related to the Hsieh & Klenow (2009) misallocation measure] in LDCs. This pattern is similar to the specific cross-country comparisons in Hsieh & Klenow (2009).
4. Wedges are more highly correlated with firm productivity in LDCs.

4.2. Misallocation and Policy Distortions

Ayerst et al. (2024) explain these facts using a competitive equilibrium model without any market frictions but characterized by distorting tax policies that discriminate against more productive firms. We provide a simplified static and deterministic version of their model to contrast it with the capital friction model in Section 4.3.

There is a large set of agents (or potential entrepreneurs) with varying innovation ability θ . An agent of ability θ invests in productivity z at investment cost $\psi \frac{z^\phi}{\theta}$, where $\psi > 0$ and $\phi > 1$. The agent then gains access to a decreasing returns production function

$$y = z^{1-\gamma} n^\gamma, \quad 1.$$

where y is output and n is employment. The firm pays each worker a given wage w and incurs a fixed cost of operation c_f denominated in labor units. The firm is a price taker and the product price is normalized to unity. It is taxed on sales at a rate τ that depends on its productivity:

$$\tau(z, \epsilon) = 1 - z^{-\rho(1-\gamma)} \epsilon^{1-\gamma}, \quad 2.$$

where the parameter $\rho > 0$ represents progressivity of the tax system, and ϵ is an idiosyncratic firm-specific shock drawn from a given i.i.d. distribution. The realization of ϵ is observed by the entrepreneur before it decides how much to invest in innovation or whether to operate the firm.

If the agent with productivity z faces a tax rate of τ and decides to operate the firm, it would select employment n to maximize operating profits:

$$\pi(z; \tau) \equiv \max_{n \geq 0} [(1 - \tau)z^{1-\gamma} n^\gamma - wn - c_f w] = \Omega(1 - \tau)^{\frac{1}{1-\gamma}} z - c_f w, \quad 3.$$

where $\Omega \equiv \left[\frac{\gamma}{w}\right]^{\frac{\gamma}{1-\gamma}} [1 - \gamma]$. Inserting Equation 2 for the tax rate, the employment level is

$$\begin{aligned} n(z, \tau(z, \epsilon)) &= \left[\frac{\gamma}{w}\right]^{\frac{1}{1-\gamma}} [1 - \tau(z, \epsilon)]^{\frac{1}{1-\gamma}} z \\ &= \left[\frac{\gamma}{w}\right]^{\frac{1}{1-\gamma}} z^{1-\rho} \epsilon. \end{aligned}$$

A higher progressivity parameter ρ therefore results in a flatter slope of employment with respect to its productivity, which represents a form of static misallocation.

Anticipating these outcomes, productivity is chosen at the point of entry by an agent with innovation ability θ and tax shock ϵ to solve

$$\begin{aligned} V(\theta, \epsilon) &= \max_{z \geq 0} \left[\Omega(1 - \tau(z; \epsilon))^{\frac{1}{1-\gamma}} z - \psi \frac{z^\phi}{\theta} \right] - c_f w \\ &= \max_{z \geq 0} \left[\Omega z^{1-\rho} \epsilon - \psi \frac{z^\phi}{\theta} \right] - c_f w. \end{aligned}$$

This results in productivity

$$z(\theta, \epsilon) = \left[\frac{\Omega(1 - \rho)\theta\epsilon}{\psi\phi} \right]^{\frac{1}{\phi+\rho-1}}, \quad 4.$$

and a higher value of ρ lowers productivity, a form of dynamic misallocation.

Finally, the agent decides to operate the firm if its productivity and tax shock are such that its operating profits will be nonnegative:

$$\Omega z^{1-\rho} \epsilon \geq c_f w, \quad 5.$$

which, using Equation 4, reduces to

$$\Gamma(w, \rho) \theta^{\frac{1-\rho}{\phi+\rho-1}} \epsilon^{\frac{\phi}{\phi+\rho-1}} \geq c_f w, \quad 6.$$

where $\Gamma(w, \rho) \equiv \Omega^{\frac{\phi}{\phi+\rho-1}} \left[\frac{1-\rho}{\psi\phi}\right]^{\frac{1-\rho}{\phi+\rho-1}}$. The left-hand side of Equation 6 is increasing in θ , ϵ and falling in ρ . Hence, Equation 6 is an entry condition that can be restated as follows. Define the entry threshold for the tax shock parameter $\hat{\epsilon}(\theta; \rho, w)$ to be the value of ϵ at which Equation 6 holds as an equality. Clearly this threshold is decreasing in θ and rising in ρ . The firm operates if and only if

$$\epsilon \geq \hat{\epsilon}(\theta; \rho, w), \quad 7.$$

so the likelihood of the firm being active is increasing in the entrepreneur's ability and falling in ρ . Holding the wage rate fixed, a higher ρ would reduce entry rates and result in positive selection (i.e., higher average productivity) of active firms.

On the other hand, a higher ρ would lower the demand for labor; with an inelastic aggregate supply of labor the wage rate would fall. This would raise Γ and lower the entry cost, which would provide a countervailing increase in entry rates and adverse selection. The net effect on entry rates and selectivity patterns is therefore theoretically ambiguous. In their empirical calibrations of the model to cross-country data, Ayerst et al. (2024) find that the wage effect dominates, so the net effect of higher ρ is higher entry and adverse selection of active entrepreneurs.

How does this model explain the stylized facts? Ayerst et al. (2024) focus on the role of cross-country variations in ρ (the progressivity parameter), with a higher ρ in LDCs, while all other parameters are the same. The model is consistent with Facts 1 and 2 owing to lower wages in LDCs resulting from higher ρ , which encourage entry of more firms with less able entrepreneurs and lower investments in productivity enhancement. Hence, average firm size and average productivity are lower in LDCs. Firm-level TFP and wedge W given productivity z and tax shock ϵ are given by

$$TFP \equiv \frac{y}{n^\gamma} = z^{1-\gamma}$$

$$W \equiv \frac{y}{n} = \frac{w}{\gamma(1 - \tau(z, \epsilon))} = \frac{wz^{\rho(1-\gamma)}}{\gamma\epsilon^{1-\gamma}},$$

implying the following relationship between TFP and W :

$$\log W = \log \frac{w}{\gamma} - (1 - \gamma) \log \epsilon + \rho(1 - \gamma) \log TFP. \quad 8.$$

Hence $\rho(1 - \gamma)$ equals the elasticity of W with respect to TFP, which explains Fact 4 if LDCs have more progressive tax policies.

Fact 2 pertains to comparisons of dispersion of TFP across countries. From Equation 4 we have

$$\log z = \frac{1}{\phi + \rho - 1} \left[\log \frac{\Omega}{\psi\phi} + \log(1 - \rho) + \log \theta\epsilon \right]. \quad 9.$$

Therefore, in the absence of selection effects, the model predicts a lower dispersion of TFP in LDCs, contrary to Fact 2. This is consistent with the model if the adverse selection effect and higher entry rates in LDCs raise TFP dispersion sufficiently to overcome the opposite prediction of the model in the absence of selection effects.

Finally, higher dispersion of W in LDCs (Fact 3) is explained by Equation 8 if TFP exhibits higher dispersion in LDCs.

4.3. Misallocation and Capital Frictions

We now consider an alternative model with credit market frictions but no tax distortions. The model is based on the formulation of capital frictions by Buera et al. (2011, 2021), which we extend to incorporate scale economies over an initial range of the production function. As in our exposition of the model by Ayerst et al. (2024) above, we focus on a static version of the model. We provide an informal exposition of the main features; technical details and proofs are provided by Ghatak & Mookherjee (2024a). To facilitate comparison with the Ayerst et al. (2024) model, we use the same notation for outputs and input variables as far as possible.

Agents have ex ante heterogeneous ability θ and decide whether to operate a firm or become a worker and earn the going wage. Analogous to the Ayerst et al. (2024) model, entrepreneurs choose how much to invest in productivity enhancement, besides the level of employment. On the other hand, an important difference pertains to assumptions concerning returns to scale in the production function. Firm size or scale of operation S depends on labor employed n and investment in productivity enhancement z :

$$S = z^\gamma n^{1-\gamma}, \quad (10)$$

with $\gamma \in (0, 1)$. Output depends on the owner's ability and firm size according to

$$y = \theta f(S), \quad (11)$$

where

$$\begin{aligned} f(S) &= S_e^{1-\mu} S^\mu & \text{if } S \leq S_e, \\ &= S_e^{1-\delta} S^\delta & \text{if } S > S_e, \end{aligned}$$

and $S_e > 0$ is the technically efficient size. We assume $\mu > 1 > \delta > 0$: Hence, there is an initial phase of increasing returns up to S_e (where $\frac{f(S)}{S}$ is maximized), followed by decreasing returns (see **Figure 1**).

A possible underlying story is that all firms have the same production capacity S_e , which is under- or overutilized if S is below or above S_e . If $u \equiv \frac{S}{S_e}$ denotes the utilization rate, $f(S) = S_e u^\mu$ if $u \leq 1$ and $f(S) = S_e u^\delta$ if $u \geq 1$. Note that this specification reduces to a conventional neoclassical production function with decreasing returns throughout if $\mu = \delta < 1$, and with constant returns throughout if $\mu = \delta = 1$.

Labor is hired at wage rate w and productivity-enhancing investments are incurred at a price r . Both factor prices are exogenously given. Besides variable inputs, every firm incurs a fixed cost c to operate. Output price is normalized to unity. We consider a single period, at the beginning of which input costs are incurred. Output and sales are realized at the end of the period.

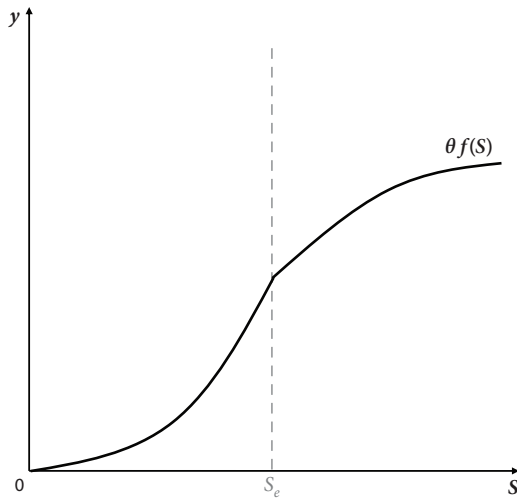


Figure 1

S-shaped production function.

Agents differ on two dimensions: ability θ and wealth consisting of collateralizable assets a . $G(\theta)$ denotes the distribution function of θ , and $H(a|\theta)$ the distribution over a conditional on θ . Wealth plays a role analogous to the tax shock ϵ in the Ayerst et al. (2024) model, a source of friction affecting firm decisions and performance. Wealth matters because the scale of production is limited by the working capital available to the agent, described by the borrowing constraint that we now explain. While agents face a limit on how much they can borrow, the interest rate on borrowing and lending is the same. Let $i \equiv 1 + r$ denote the resulting interest factor.

To keep the model simple, we abstract from how factor prices w, i are determined and treat them as fixed, corresponding to a small open economy facing a perfectly elastic supply of labor and loanable funds. It can be extended to closed economies and incorporate endogenously determined factor prices, but these do not play an important role [unlike the Ayerst et al. (2024) model].

The profit of an entrepreneur of ability θ selecting inputs n, z is $[\theta f(z^\gamma n^{1-\gamma}) - i(wn + rz) - c]$. Given any scale S of operation, n, z will be chosen to minimize $wn + rz$ subject to $S = z^\gamma n^{1-\gamma}$. Normalize units so that $[\frac{z}{\gamma}]^\gamma [\frac{w}{1-\gamma}]^{1-\gamma} = 1$. The solution is $n = \frac{1-\gamma}{w} S, z = \frac{\gamma}{r} S$, resulting in total cost $S + c$ incurred at the beginning of the period. Profit at the end of the period equals $\pi(S; \theta) - ic$ where

$$\pi(S; \theta) \equiv \theta f(S) - iS \quad 12.$$

denotes operating profits (excluding the fixed cost).

Producing at scale S thus requires a financial outlay of $S + c$ at the beginning of the period. The agent would need to borrow if its own wealth a is smaller than $S + c$. Without loss of generality, the borrower borrows $S + c$ and posts its assets as collateral. In the event of a default, the lender can seize the end-of-period value of the borrower's assets ia and a fraction ϕ of profits $(\pi(S, \theta) - ic)$. The borrower will not default if the default cost $\phi[\pi(S; \theta) - ic] + ia$ exceeds the repayment due $i(S + c)$. This gives rise to the financing constraint

$$ia + \phi[\pi(S, \theta) - ic] \geq i(S + c). \quad 13.$$

Finally, for the agent to want to become an entrepreneur, it must be able to finance a scale S that generates a profit of at least w :

$$\pi(S; \theta) - ic \geq w. \quad 14.$$

The (hypothetical) first-best allocation corresponds to choice of $S \geq 0$ by each agent conditional on entering, to maximize profit $\pi(S; \theta)$ without any constraint. Let $S^*(\theta)$ denote the first-best scale, which equals zero if $\theta < i$, S_c if θ lies between i and $\frac{i}{\delta}$, and $S_c [\frac{\delta\theta}{i}]^{\frac{1}{1-\delta}}$ otherwise. Corresponding first-best operating profits $\pi^*(\theta)$ equal 0 if $\theta < i$, $(\theta - i)S_c$ if θ lies between i and $\frac{i}{\delta}$, and $S_c \theta^{\frac{1}{1-\delta}} [(\delta)^{\frac{\delta}{1-\delta}} - (\delta)^{\frac{1}{1-\delta}}] i^{-\frac{\delta}{1-\delta}}$ otherwise. Hence, the agent will enter if and only if $\pi^*(\theta) \geq ic + w$, or $\theta \geq \underline{\theta}^F$ defined by the property that $\pi^*(\underline{\theta}^F) = ic + w$. In what follows, we restrict attention to agents of ability at least $\underline{\theta}^F$, since those of lower ability will never find it worthwhile to become an entrepreneur either with or without the borrowing constraint.

In the second-best economy, scale S maximizes $\pi(S; \theta)$ subject to the borrowing constraint given in Equation 13, and the agent enters if Equation 14 holds. Observe that the first-best scale $S^*(\theta)$ is feasible if the agent's wealth lies above the threshold $\bar{a}(\theta)$ defined by:

$$\bar{a}(\theta) = \max\{0, S^*(\theta) + c - \frac{\phi}{i} [\pi(S^*(\theta); \theta) - ic]\}. \quad 15.$$

Those with wealth below this threshold are constrained to a smaller range of scales owing to the borrowing constraint.

For any agent of type (a, θ) with $\theta \geq \underline{\theta}^F$, the second-best allocation turns out to be the following. The agent enters if and only if $a \geq \hat{a}(\theta)$ given by

$$\hat{a}(\theta) \equiv \max \left\{ 0, \frac{1+\phi}{i}c - \frac{1+\phi}{i}\pi \left(\underline{S}(\theta), \frac{\phi\theta}{1+\phi} \right) \right\}, \quad 16.$$

where $\underline{S}(\theta) (\leq S^*(\theta))$ defined by the condition $\pi(\underline{S}(\theta); \theta) = ic + w$ is the minimum scale at which the entrepreneur would earn at least w . Those with $a \in [\hat{a}(\theta), \bar{a}(\theta)]$ are credit-constrained and select scale $S(a, \theta)$, which is the value of S where the borrowing constraint in Equation 13 just binds. The second-best scale $S(a, \theta)$ is locally increasing in a and θ . It ranges from $\underline{S}(\theta)$ to $S^*(\theta)$ as a ranges from $\hat{a}(\theta)$ to $\bar{a}(\theta)$. Those with $a \geq \bar{a}(\theta)$ are unconstrained and select first-best scale $S^*(\theta)$, which is locally independent of a .

Some properties of the second-best allocation can be noted (we hereafter refer to scale S as firm size). Fixing ability θ , the support of the conditional firm size $S(a, \theta)$ distribution is $[\underline{S}(\theta), S^*(\theta)]$ if the wealth distribution conditional on θ has full support. Allowing θ to also vary, higher θ values correspond to a wider range of firm sizes $[\underline{S}(\theta), S^*(\theta)]$ since $\underline{S}(\theta)$ is decreasing while $S^*(\theta)$ is increasing in θ . Unlike the first-best allocation, firms operated by poor entrepreneurs may select scales below S_e in the second-best. To see this, consider any ability $\theta > i + \frac{ic+w}{S_e}$. For such an agent, $\underline{S}(\theta) < S_e$, because operating at scale S_e ensures the agent will earn operating profits $(\theta - i)S_e$, which exceeds $ic + w$. The range of active firm sizes for any such ability will therefore include scales below S_e .

To simplify the exposition in what follows, we focus on economies where the following parameter restrictions hold: (a) $\frac{i}{\delta} < i + \frac{ic+w}{S_e}$, which implies there is no bunching in the first-best; (b) ability and wealth are either independent, or positively correlated (in the sense that the conditional wealth distribution at higher ability levels first-order stochastically dominates those at lower levels); and (c) $\underline{\theta}^F > i + \frac{ic+w}{S_e}$, which ensures that for every relevant ability the minimum scale of operation $\underline{S}(\theta)$ is smaller than S_e . Define the wealth threshold $\bar{a}(\theta) (> \hat{a}(\theta))$ at which the second-best scale is S_e , that is, $S(\bar{a}(\theta), \theta) = S_e$.

Active firms can then be classified into three groups:

1. Small firms: those with scale $S < S_e$ owing to the wealth of their owners, which falls below $\bar{a}(\theta)$.
2. Medium firms: those with scale $S \in [S_e, S^*(\theta)]$ whose owners have wealth above $\bar{a}(\theta)$ but not large enough to attain the first-best.
3. Large firms: those achieving scale $S^*(\theta)$ owing to their owners' wealth exceeding $\bar{a}(\theta)$.

Figure 2a shows entry and firm size category outcomes for different combinations of ability and wealth. **Figure 2b** shows variations in firm size induced by variations in wealth, holding ability fixed.

Among small firms, we can compute logs of output y , TFP $\frac{y}{n^{1-\gamma}\mu}$, and wedge $W \equiv \frac{y}{n}$:

$$\begin{aligned} \log y &= (1 - \mu) \log S_e + \log \theta + \mu \log S(a, \theta), \\ \log TFP &= \gamma \mu \log \left(\frac{\gamma}{r} \right) + (1 - \mu) \log S_e + \log \theta + \gamma \mu \log S(a, \theta), \\ \log W &= \log C + (1 - \mu) \log S_e + \log \theta + (\mu - 1) \log S(a, \theta), \end{aligned} \quad 17.$$

where $C \equiv \left[\frac{\gamma w}{(1-\gamma)r} \right]^{\gamma \mu} \left[\frac{1-\gamma}{w} \right]^{\mu-1}$. Note that $\mu - 1 > 0$ implies the wedge (labor productivity) is increasing in firm size resulting from higher wealth (holding ability fixed), owing to local scale economies over the range of small firm sizes. Hence, wealth effects induce positive comovement

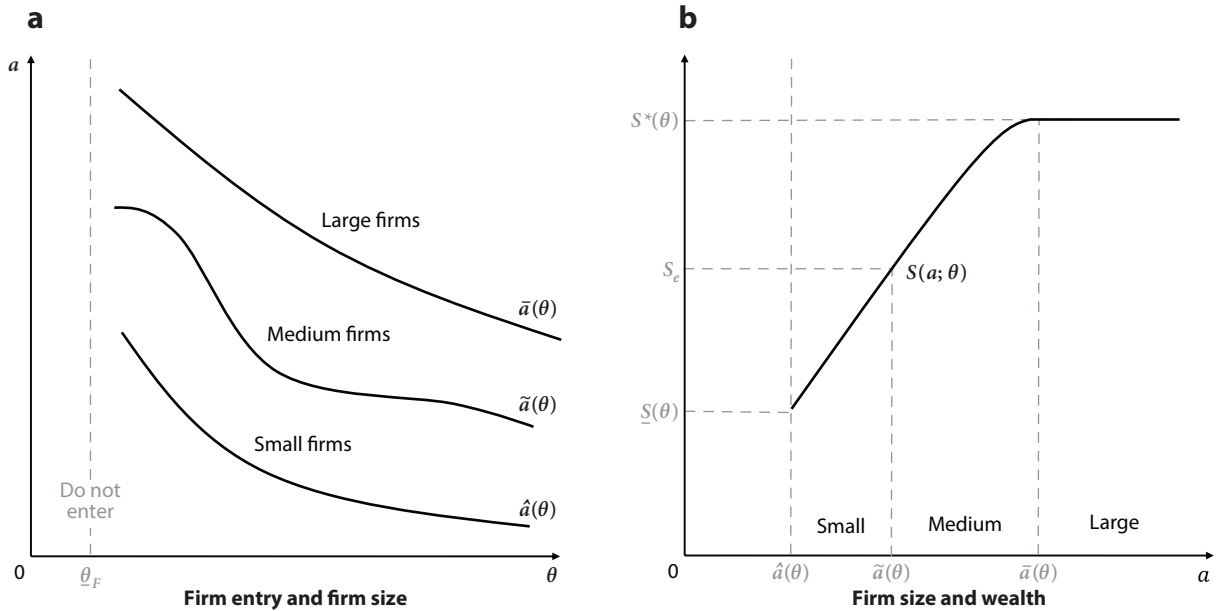


Figure 2

Second-best allocation of firm entry, firm size, and surplus as a function of ability and wealth.

of output, TFP, and the wedge. Moreover, an increase in wealth dispersion among owners of small firms generates higher wedge and TFP dispersion within this group.

Among medium firms, we obtain analogous expressions with δ replacing μ . Output and TFP continue to be positively correlated, but the sign of the TFP-wedge correlation is now ambiguous (as $\delta - 1 < 0$). Holding ability fixed, an increase in wealth raises investment in productivity enhancement, which raises TFP but lowers W owing to decreasing returns to scale. On the other hand, increasing ability while holding wealth fixed raises both TFP and W . The net effect can go either way. Owing to decreasing rather than increasing returns over the range of medium firm sizes, the TFP-wedge correlation is likely to be smaller over the medium firm size range compared with the small firm size range.

Finally, among large firms,

$$\log y = \log \theta + \delta \log S^*(\theta) + (1 - \delta) \log S_e, \quad 18.$$

$$\log TFP = \gamma \delta \log \left(\frac{\gamma}{r} \right) + \log \theta + \gamma \delta \log S^*(\theta) + (1 - \delta) \log S_e, \quad 19.$$

while the wedge is constant, since large firms select first-best scales where the factor marginal products are equal. Hence, within the large firm group, output and TFP are positively correlated, but the wedge is uncorrelated with either.

Figure 3 shows how TFP and W covary with firm size as wealth is varied, holding ability fixed at two different levels $\theta' > \theta > i + \frac{w}{S_e}$.

We now explain how this model can explain the stylized facts, under the hypothesis that developed countries (DCs) and LDCs differ only with respect to the wealth distribution. Specifically, suppose that the DC distribution dominates the LDC distribution in both the first- and second-order sense (higher mean, lower dispersion).

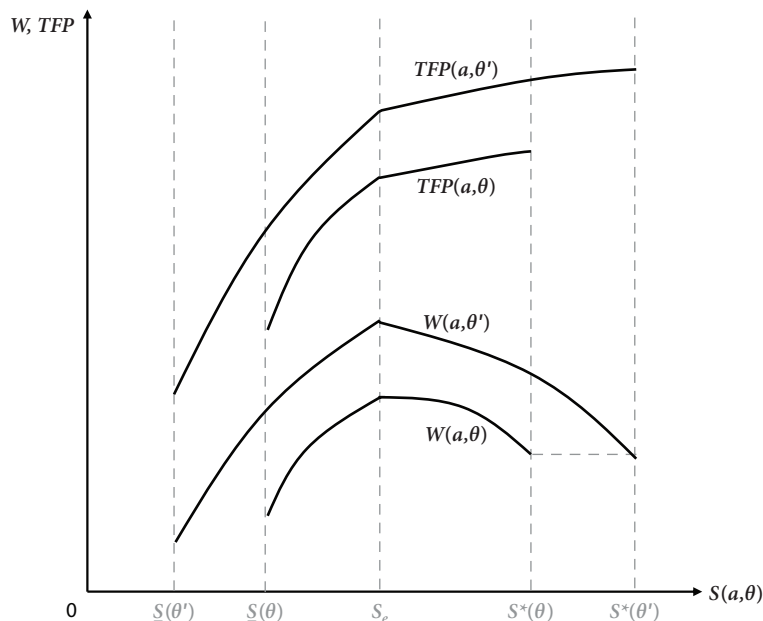


Figure 3

Total factor productivity (TFP) and wedge (W) variation with wealth, holding ability fixed.

Consistent with Fact 1, average firm size would be lower in LDCs, since firm size is increasing in entrepreneur wealth. In particular there would be more small enterprises and fewer large enterprises. Firm level size, output, and TFP would be more dispersed in LDCs owing to higher wealth dispersion.

Moreover, consistent with Fact 2, higher wealth dispersion in LDCs generates a higher weight in the lower tail of the TFP distribution, resulting in a preponderance of small enterprises with lower TFP compared to medium or large enterprises.

Fact 3 states that the dispersion of wedges is larger in LDCs. As shown in **Figure 3**, holding ability fixed, the wedge is rising in wealth among small enterprises, falling in wealth among medium enterprises, and constant for large enterprises. Hence, the wedge-TFP relationship exhibits an inverted-U shape, which eventually flattens out at large scales. If most DC firms are large while most LDC firms are small, wedge dispersion in DCs would be smaller.

Finally, the wedge and TFP would be positively correlated within the small firm category, while the correlation within the medium category is likely to be smaller, and is zero among large firms. Hence, the estimated elasticity of the wedge with respect to TFP could be positive and larger in LDCs, consistent with Fact 4.

4.4. Contrasting Welfare Implications of the Two Models

In the Ayerst et al. (2024) model, the progressive firm-specific taxes create static and dynamic misallocation, both of which reduce aggregate output and welfare. This is due to the fact that in the absence of these policies their economy is characterized by no frictions, so classical welfare theorems for first-best economies apply.

Since this is not true for the capital friction model, and there are no general results concerning welfare properties of such a second-best economy in the literature, we need to examine the welfare

impact of progressive firm-specific taxes. In the second-best allocation described in the previous section, welfare equals aggregate income:

$$\begin{aligned}
 W = & wG(\underline{\theta}^F) + \int_{\underline{\theta}^F}^{\bar{\theta}} \left[wH(\hat{a}(\theta)|\theta) \right. \\
 & + \int_{\hat{a}(\theta)}^{\bar{a}(\theta)} \{ \pi(S(a, \theta), \theta) - ic \} dH(a|\theta) \\
 & \left. + (1 - H(\bar{a}(\theta))) \{ \pi(S^*(\theta), \theta) - ic \} \right] dG(\theta),
 \end{aligned} \tag{20}$$

where $\bar{\theta}$ (possibly ∞) denotes the upper bound of ability. The first line represents wage earnings of workers, the second line the profits of constrained entrepreneurs E^c , and the third line the profits of unconstrained entrepreneurs E^u . Compared to the first-best, welfare is lower for those who have ability above $\underline{\theta}^F$ and either (a) own wealth less than $\hat{a}(\theta)$, who work instead of becoming an entrepreneur; or (b) own wealth between $\hat{a}(\theta)$ and $\bar{a}(\theta)$, who earn less than first-best profit owing to a suboptimal firm size. Total output in the economy is lower on both extensive and intensive margins owing to credit frictions. Moreover, factors are misallocated between those in E^c and E^u , as factor marginal products vary between entrepreneurs in E^c and E^u , and also between different entrepreneurs with varying wealth within E^c .

Now consider the welfare impact of size-dependent policies of the following form: Firms producing output that exceeds a threshold q^* are required to pay a tax t , while those producing below q^* receive a subsidy s .⁴ The policy balances the government's budget if the tax revenues collected from the high-output firms are sufficient to cover the subsidies paid to the low-output firms.

It turns out a balanced-budget, welfare-enhancing, size-dependent policy always exists if enforcement institutions are of intermediate strength. This is irrespective of specific production parameters, wealth, or ability distributions. By intermediate strength we mean the following. Define $\phi^*(\theta) \equiv \frac{i(S^*(\theta)+c)}{\pi(S^*(\theta), \theta) - ic}$, which is a threshold for the enforcement parameter ϕ such that all agents of ability θ can borrow enough to attain their respective first-best allocations, irrespective of their wealth. A necessary condition for the first-best allocation to be unattainable for the economy as a whole is that enforcement institutions are not too strong in the sense that ϕ is smaller than the threshold $\phi^*(\underline{\theta}^F)$ for the lowest active ability level in the first-best economy. We add to this the condition that ϕ is not too low, in the sense that $\phi > \frac{\delta}{1-\delta}$, which happens to be the limiting threshold for agents of arbitrarily high ability. Formally, we say that enforcement institutions are of intermediate strength if

$$\phi^*(\underline{\theta}^F) > \phi > \frac{\delta}{1-\delta}. \tag{21}$$

Here is a sketch of the argument for the existence of a welfare-enhancing, balanced-budget, size-dependent policy whenever Equation 21 holds. The condition given in Equation 21 implies the existence of ability level $\tilde{\theta}$, above which agents are able to finance their first-best scale of production irrespective of their wealth ($\bar{a}(\theta) = 0$), while those of lower ability need a positive amount of wealth ($\bar{a}(\theta) > 0$) to do so. In the laissez-faire equilibrium, those with ability above $\tilde{\theta}$ will operate at a larger scale than all those with lower ability, owing to both their superior ability and the absence of financing constraints. The policy sets the output threshold for the tax at $q^* = q^F(\tilde{\theta})$, where $q^F(\theta) \equiv \theta f(S^*(\theta))$ denotes the first-best output for ability θ . Any firm that produces q^* or

⁴Analogous results can be shown for productivity-dependent policies of the type considered in the Ayerst et al. (2024) model.

less receives the subsidy s , while any higher output invites the fixed tax t . This creates a sharp disincentive for firms to raise their output above q^* , amounting to an effective fiscal distortion of $\nu \equiv s + t$.

The effect of this policy on different groups of agents ordered by ability is as follows. There exists $\epsilon(\nu)$, an increasing function of the policy distortion ν , satisfying $\epsilon(0) = 0$, such that the following types of firms are created.

1. Largest firms (no size effect): Those with $\theta > \tilde{\theta} + \epsilon(s + t)$ produce $q^F(\theta)$ as before and pay the tax.
2. Large firms (shrinkage): Those with $\theta \in [\tilde{\theta}, \tilde{\theta} + \epsilon(s + t)]$ produce q^* instead of $q^F(\theta)$ and receive the subsidy.
3. Large firms (no size effect): Those with $\theta \in [\underline{\theta}^F, \tilde{\theta})$ with assets $a \geq \bar{a}(\theta)$ continue to produce $q^F(\theta)$ at scale $S^*(\theta)$ and receive the subsidy.
4. Medium and small firms (expansion): For those with $\theta \in [\underline{\theta}^F, \tilde{\theta})$ with assets $a \in [\hat{a}(\theta, 0), \bar{a}(\theta))$, size expands from $S(a, \theta)$ to $\min\{S(a + s, \theta), S^*(\theta)\}$ and they receive subsidy s .
5. New entry: Those with $\theta \in [\underline{\theta}^F, \tilde{\theta})$ and assets $a \in [\hat{a}(\theta, s), \hat{a}(\theta, 0))$ enter;⁵ these new firms select size $S(a + s, \theta)$ and receive subsidy s .

Group 2 agents respond to the policy by shrinking their scale and bunch at the output threshold q^* , which is below their first-best output, a new distortion created by the policy that lowers welfare. Group 4 agents respond by expanding their scale owing to the positive wealth effect generated by the subsidy. This neutralizes the preexisting underproduction distortion owing to the market friction, which raises welfare. Group 5 comprises new entrants that start small firms of low productivity, resulting in ambiguous welfare effects.⁶

The size of the subsidy is set to ensure that their costs (which depend on the size of groups 2–5) are financed by the taxes collected (which depends on the size of group 1). If the tax and subsidies are small, the welfare gains achieved by group 4 are first-order; these overwhelm the welfare effects generated by groups 2 and 5, which are second-order. The first-order welfare gains achieved by group 4 are due to the role of subsidies that relieve the underproduction of their firms under laissez-faire owing to credit constraints. The welfare losses generated by the shrinkage of firms in group 2 are comparatively negligible because those firms were achieving their first-best outcome where marginal products of factors were equal to their costs. So the wedges created by the size-dependent policy create a net welfare gain in the economy as a whole primarily because they offset the preexisting wedges resulting from capital market frictions for group 4 agents. However, the firms in this group that expand are smaller and less productive compared to those that shrink in group 2, conveying the impression to someone viewing these outcomes through the lens of the Ayerst et al. (2024) model that the policy increases misallocation and thereby harms welfare.

Itskhoki & Moll (2019) show that in the presence of capital market frictions, Ramsey-optimal policies involve wage repression at early stages of development. Such policies result in higher entrepreneurial profits and faster wealth accumulation, which relaxes borrowing constraints in the future, leading to higher labor productivity and wages. In the long run, the optimal policy reverses and becomes proworker. Such policies can raise the long-run welfare of workers as well as

⁵We use $\hat{a}(\theta, s)$ to denote the wealth entry threshold for agents with type θ that receive subsidy s .

⁶The largest among these small new firms earn profits that exceed the wages they previously earned, even when these profits are calculated excluding the subsidy they receive. For the smallest among them, their profit excluding the subsidy falls below the wage. The former (respectively, latter) category generates welfare gains (respectively, losses). Whether the sum of these two conflicting welfare effects is positive or negative depends on the relative frequency of the two categories, which in turn depends on the shape of the wealth distribution in a neighborhood of the entry threshold.

entrepreneurs. In Ghatak & Mookherjee (2024a) we show that wage repression policies are generally welfare improving (though not Pareto improving) even in a static context, owing to the first-order welfare gains generated by relieving underproduction in small and medium enterprises. By contrast, in the Ayerst et al. (2024) model, such policies are likely to aggravate misallocation and lower welfare by increasing the adverse selection resulting from additional entry of low-productivity entrepreneurs.

5. INTERFIRM SPILLOVERS: MISALLOCATION AND WELFARE IMPLICATIONS

A large literature on endogenous growth (Lucas 1988, Romer 1986) and urban economics (Henderson 1974) is based on the existence of productivity or learning spillovers across firms. Contemporary arguments for soft industrial policy or place-based policies are primarily based on such agglomeration spillovers across entrepreneurs located in close physical proximity (Harrison & Rodriguez-Clare 2010, Rodrik 2004, Stiglitz 2017). Empirical evidence of such spillovers has been provided by a number of authors, mainly in the context of developed countries (Bloom et al. 2013, Combes et al. 2012, Greenstone et al. 2010, Moretti 2004).

In developing countries, the literature on industrial clusters and trade relations stresses the importance of social networks, which help overcome problems of trust and cooperation faced by small- and medium-size entrepreneurs in accessing credit, insurance, knowhow, and reliable input supply in environments with weak market and state institutions (Casella & Rauch 2002; Dai et al. 2021; Fafchamps 2001; Gupta et al. 2022; McMillan & Woodruff 1999; Munshi 2011, 2014). These network relationships generate interfirm spillovers whose domain is restricted to firms owned by entrepreneurs belonging to a social network defined by ethnicity or social origin. In many of these contexts, ethnic groups differ considerably with respect to internal cohesion, trust, and cooperation, resulting in wide disparities in entry, levels, and growth rates of firm size and productivity. Network spillovers differ from agglomeration spillovers whose domain is instead defined by physical proximity, that is, across entrepreneurs at a common location, irrespective of social identity/origin. Empirical evidence of network-specific spillovers is available for caste networks in India (Banerjee & Munshi 2004, Gupta et al. 2022, Munshi 2011) and hometown networks in China (Dai et al. 2023).

The existence of interfirm spillovers implies a departure from a first-best environment. Despite growing evidence of such spillovers, their implications for productive misallocation and welfare have not been explored in the literature. In a model of heterogeneous networks, Mookherjee & Munshi (2024) show that standard measures of productive misallocation are an unreliable indicator of welfare effects of government policies. In their model, agents are partitioned into multiple networks, where spillovers occur across firms belonging to the same network, with zero cross-network spillovers. Agents differ on two dimensions: individual ability, which is drawn from an i.i.d. distribution, and the cohesiveness of the social network they belong to. More socially cohesive networks are characterized by stronger spillovers. Each agent decides whether or not to become an entrepreneur rather than pursue an alternative occupation in which returns to ability are positive but lower than in entrepreneurship. Firm TFP increases with individual ability, network cohesiveness, and size (i.e., how many agents from the same network decide to enter in equilibrium). Conditional on entering, each agent decides how much capital to invest, where the cost of capital is decreasing in network cohesiveness and size.

Nash equilibria of this model exhibit productive misallocation measured by cross-network dispersion of MRP if networks vary in cohesiveness. Entry is more attractive for members of a more cohesive network owing to the stronger spillovers; so they achieve a larger network size, lower

cost of capital, and MRP. The TFP and firm size of the marginal entrant in a more cohesive network also turn out to be lower. These disparities suggest that aggregate productivity would rise if there were a reallocation on either extensive or intensive margins in favor of less cohesive networks. However, this may not be true owing to the associated spillover effects. It turns out that on the intensive margin, (capital) allocation is efficient.⁷ Inefficiencies arise instead on the extensive margin: In every network, the entry rate is inefficiently low owing to the intranetwork spillover, which each individual agent ignores in a Nash equilibrium. Hence, the social planner can raise aggregate surplus by providing a network-specific entry subsidy (financed by taxes imposed on consumers). Since more cohesive networks are characterized by stronger spillovers, the optimal subsidy is typically higher for a more cohesive network. Compared to the decentralized laissez-faire equilibrium, the welfare optimal policy will then aggravate cross-network MRP dispersion. MRP dispersion is therefore an unreliable welfare criterion in this setting. A researcher that ignores the interfirm spillovers may then erroneously interpret the allocation resulting from optimal policy interventions as exhibiting a welfare loss relative to the first-best and infer that this welfare loss is a consequence of the policy.

6. MISALLOCATION AND WELFARE IMPLICATIONS OF MARKET FRICTIONS IN OTHER SECTORS

6.1. Agriculture

The microdevelopment literature on land allocation studies the role of different kinds of frictions in land, labor, credit, and insurance markets in explaining commonly observed agrarian practices associated with productive misallocation. This includes sharecropping tenancy, the presence of family or cooperative farms that may not pay people according to market principles, agency costs associated with hired labor, dispersion of interest rates unrelated to borrower risks or productivity despite competition across lenders, or bundling of credit and insurance with land or labor transactions in rural economies. Many of these have been explained by agency costs, informational asymmetries, and various contracting frictions (Stiglitz 1988).

The empirical microdevelopment literature has also generated a number of stylized facts about agriculture in LDCs pertaining to productive misallocation. First, smaller farms generally exhibit higher productivity than larger ones, a phenomenon well-documented in the literature and with the general consensus being it is not driven by unobserved land quality (Banerjee 1999, Foster & Rosenzweig 2022).⁸ Theoretical explanations include that by Eswaran & Kotwal (1986), which is rooted in a combination of frictions in credit and labor markets. Second, land that is owner-cultivated tends to be more productive than land under sharecropping tenancy as predicted by Marshallian theories (Burchardi et al. 2017, Shaban 1987). In addition, this literature also examined other factors such as social norms relating to gender that may generate misallocation. For example, Udry (1996) finds that in Burkina Faso, plots controlled by women are farmed less intensively than similar plots within the same farming households controlled by men; his estimates suggest that about 6% of output is lost because of inefficient factor allocation within the household.

These facts suggest the need for corrective policies that could boost agricultural productivity and welfare by offsetting the effect of distorting market frictions. For instance, the superior productivity of small farms compared to large farms, or of owner-cultivated farms compared to

⁷This owes to differences in underlying assumptions between this model and Hsieh & Klenow (2009), where the aggregate amount of capital and the set of firms are exogenously fixed.

⁸However, Foster & Rosenzweig (2022) find a U-shaped pattern in India, where farms of intermediate size are the least productive, while the smallest and largest farms are equally productive.

sharecropped farms, has generated arguments for land reforms that redistribute land from large landowners to landless peasants and from landlords to tenants. Such policies have been viewed as a win-win situation, as they have been expected to raise aggregate productivity as well as reduce inequality.

However, empirical evidence on the effects of land reforms on agricultural productivity is limited, and the available evidence is somewhat mixed (e.g., Banerjee et al. 2002, Bardhan & Mookherjee 2011, Besley & Burgess 2000, Ghatak & Roy 2007). Adamopoulos & Restuccia (2020) show that these reforms were often coupled with restrictions on land sales and rentals, which had a significant impact on average farm sizes. Using data from the Comprehensive Agrarian Reform Program in the Philippines, they argue that these restrictions on landholdings and land markets induced misallocation and loss of overall productivity. They conclude that although land reforms might aim to improve productivity by reallocating land to more productive farmers, the way these reforms are implemented can undermine these potential benefits, particularly when land is redistributed based on size rather than productivity.⁹ A functioning rental market for land could help mitigate these adverse effects by separating land use from land rights.

The scope of this argument obviously does not extend to economies with functioning land rental markets. Moreover, it is unclear whether unfettered markets for private land rights or rentals would necessarily eliminate productive misallocation in the presence of transaction costs, informational asymmetries, and incentive problems that create frictions in the land, labor, and credit markets, which lead to the emergence of second-best arrangements such as family-operated farms and sharecropping tenancy. For example, if tenancy involves the loss of productivity due to incentive problems, would freeing up restrictions on land markets eliminate this loss? Presumably the landlord leases the land to the tenant on account of the latter being a more productive farmer, in which case an efficient market solution would require tenants to buy the land from their landlords. Yet this often does not happen, and sharecropping tenancy persists even in the absence of regulations forbidding sale of land rights by landlords to their tenants. This persistence can itself be explained by credit market imperfections, which prevent poor tenants from being able to borrow enough to be able to buy out the land (Mookherjee 1997). Further, some land rental regulations such as minimum crop shares accruing to tenants or protection from eviction can be justified by consequent incentive effects for tenants arising from agency problems and credit market imperfections, consistent with empirical studies in India (Banerjee et al. 2002, Bardhan & Mookherjee 2011).

Agency problems may also undermine the productive efficiency of private property rights in land compared to agricultural cooperatives under certain circumstances. Montero (2022) studies a unique setting of land reform in El Salvador, which specifies a threshold of landholding size so that properties owned by individuals with landholding above such a threshold need to be reorganized as cooperatives, while those below such threshold can remain as outside-owned properties. The paper presents a theoretical model in which cooperatives and haciendas enter into contracts for cash crops but not staple crops with workers owing to moral hazard problems (cash crops cannot be directly consumed by individual workers because they require centralized processing). Consequently, in haciendas, the owner faces a trade-off between incentivizing workers and extracting

⁹A recent paper (Wang & Kim 2024) examines Taiwan's significant 1950s land reform, often credited as key to its economic success. One of their key findings resonates with why the existing evidence of the productivity impact of land reform is mixed. They find that the earlier phase of the reform that redistributed former Japanese-owned public lands reduced tenancy and improved rice yields. A later phase of reform, which broke up larger estates to reduce tenancy, did not boost agricultural productivity and may have resulted in farms that were too small to be viable.

rent, which leads to production inefficiencies. Specifically, owners offer sharecropping contracts with suboptimal incentives, as stronger incentives for workers would necessitate paying them rents that would reduce the owner's profits. On the other hand, cooperatives may experience inefficiencies due to the desire to redistribute earnings among workers with varying abilities. Consistent with this model, the empirical findings confirm that cooperatives devote less land to cash crops and more land to staple crops. Cooperatives tend to be more productive in staple crops because members are full residual claimants on their earnings. Income distribution under cooperatives is more equitable than that under outside-owned properties.

Weak state capacity for enforcing private land rights may be responsible for the prevalence and superior performance of communal land rights in certain LDCs. Rossignol et al. (2024) show that communal property rights tend to be more prevalent in areas that are geographically more suited to crops requiring longer fallow periods to maintain high productivity. Longer fallow periods increase the costs of protecting the land, making communal ownership more advantageous if state enforcement capacity of private property rights is weak. They construct an ecological measure that estimates the optimal fallow duration for the most suitable staple crops across different regions, considering factors such as soil type, temperature, and climate and provide evidence that areas requiring longer fallow periods are more likely to have communal property rights, both in the past and today. Moreover, they provide evidence that private land titling initiatives promoted by the World Bank have been less successful in regions with longer fallow requirements, indicating a potential misalignment between promarket land policies and existing land institutions.

Even if privatization of land property rights may raise agricultural productivity, it may not result in welfare improvements if markets for insurance are missing. Baland & Francois (2005) provide a theoretical argument and many illustrative case studies to argue that the strength of commonly held property lies in its superior insurance capabilities, which help maintain income during periods of negative agricultural shocks. Lack of insurance against covariate weather shocks or natural disasters may also rationalize ownership of land among poor unproductive farmers owing to its value as a hedge against such risks (Ghatak & Mookherjee 2024b). In these settings, small farmers would be unwilling to sell their land at prevailing market prices, resulting in protests and resistance to government efforts to acquire land under powers of eminent domain in order to transfer it to more productive users. Compensations paid to dispossessed landowners need to include risk premia (over and above market prices) that are agent specific and privately known, suggesting the need for auction-like procedures for acquiring land rather than standard eminent domain policies. Moreover, land acquisition policies need to be complemented with public provision (or subsidization of private provision) of insurance against covariate shocks to mitigate the insecurity created by these policies. The cost of these additional compensations and interventions needs to be traded off against the productivity gains from the land reallocation.

A related strand of research examines the role of infrastructure in misallocation in agriculture. Morando (2023) shows this may be driven by poor infrastructure rather than restrictive land policies. Morando finds that areas with poorer transport connectivity in Uganda had more subsistence farming and greater misallocation resulting from an inefficiently large portion of inputs, in particular, land and capital, being utilized by less productive subsistence farmers. The efficiency losses were more pronounced in regions where subsistence farming was more common, primarily due to poor market connectivity. In contrast, the paper finds no significant link between misallocation and access to credit or land market activity. The author concludes that transportation costs can be crucial in shaping efficient allocation of agricultural inputs. Moreover, while land market liberalization is necessary, it would not be sufficient to address the problem of misallocation.

Recall from the discussion in Section 3 above that if insurance markets are missing, the welfare effects of trade liberalization (e.g., by lowering transport costs) may involve trade-offs between

productivity improvements and increased exposure of farmers to market risk (à la Newbery & Stiglitz 1984). Allen & Atkin (2022) estimate these two sets of effects in a detailed quantitative exercise applied to Indian agriculture and find that the efficiency benefits dominate the cost of greater risk borne by farmers. They argue that while trade liberalization boosts average returns by encouraging specialization, it also impacts the volatility of returns by reducing the negative correlation between local prices and productivity shocks. Using forty years of agricultural microdata from India, the authors find that expansions of the Indian highway network lessened the sensitivity of local prices to local rainfall, while increasing the sensitivity of local prices to yields in other regions. In response, farmers not only shifted toward crops for which they had a comparative advantage but also favored crops with less volatile yields. This shift was particularly pronounced among farmers with limited access to formal banking services. Using a structural model, they find that the overall gains from specialization surpassed any losses from risk and that advances in risk-mitigation technologies encouraged farmers to pursue higher-risk, higher-return crops that they might otherwise avoid. However, if rural bank access had remained unchanged, the welfare gains would have been only half as substantial. This implies the necessity of combining improvements in financial access with transport infrastructure.

6.2. Rural-Urban Migration

Lagakos (2020) summarizes the literature on urban-rural wage gaps in developing countries, which indicates a misallocation of labor between rural and urban areas owing to insufficient migration. To account for the urban-rural wage gaps, existing cross-sectional approaches indicate a large role of higher education levels and ability, providing support to explanations based on selection effects (whereby more able and educated workers move to urban areas) rather than misallocation. For instance, panel data estimates, which control for individual fixed effects, yield much smaller wage gaps compared to the cross-sectional results. However, experimental studies from Bangladesh (Bryan et al. 2014) find substantial increases in wages among households incentivized to move with migration incentives, indicating the need to understand why such workers do not migrate. They discuss the potential role of information frictions, financial frictions (borrowing constraints, lack of insurance markets), and land market frictions in restricting migration.

Lagakos et al. (2023) study a model where agents are exposed to idiosyncratic shocks and seasonal income fluctuations and decide whether to stay in the rural area, seasonally migrate, or permanently migrate to urban areas. In their quantitative model of migration estimated to fit the data from the Bangladesh experiment of Bryan et al. (2014), they find households with low assets and adverse transitory shocks are more likely to migrate, which implies that households may use seasonal migration as a form of self-insurance to generate enough income in lean seasons. This contrasts with the hypothesis that binding credit constraints prevent households from migrating. Furthermore, this paper solves a social planner's problem of optimal migration and efficient allocation, which is characterized by lower seasonal migration rates and provision of formal insurance to those with low assets and adverse transitory shocks. Compared to migration subsidies, larger welfare gains would be generated by providing insurance and reducing moves of desperation among vulnerable rural households.

The importance of missing formal markets for insurance in explaining rural-urban wage gaps is also highlighted in the context of India by Munshi & Rosenzweig (2016). Informal insurance networks along caste lines in the village discourage male workers from migrating to cities that lack any such network. This leads to low permanent migration rates among males and large spatial wage gaps between rural and urban areas in India. They develop a model of household migration decision-making and an endogenous income-sharing rule within the rural caste network, which predicts that households that are wealthier or face lower rural income risks tend to benefit less from

the network and hence are more likely to have male migrant members. Reduced-form evidence from Indian household survey data is consistent with these theoretical predictions. Counterfactual analysis suggests that an improvement of formal insurance will more than double migration rates, in contrast to negligible effects of an exogenous increase in income gains from migration.

6.3. Rural-Urban Land Allocation

Besides labor, structural transformation during the process of development also requires transfer of land from rural to nonagricultural uses. Governments in many developing countries accordingly focus on land acquisition policies to hasten growth by coercing or incentivizing rural landowners to sell their land to governments or private firms for industrial or infrastructural purposes. The nature of such policies varies widely across countries such as China and India, and they have frequently generated widespread protests from dispossessed farmers and progressive civil society representatives.

As in the context of intrarural misallocation, rural-urban land misallocation raises questions regarding the underlying sources—are they caused by policy restrictions on land market transactions, or would they arise also in a laissez-faire equilibrium owing to market failures? If so, what is the nature of these market failures, and what do they imply about suitable corrective policies? Examples of market failures include holdout problems arising when landownership is highly fragmented and land acquisition for an industrial project requires acquiring land from a large number of owners. If this is the only source of market failure, suitable corrective policies involve policies of eminent domain where the government coercively acquires land from the owners at market prices, which ensures they are not adversely affected. However, eminent domain policies often generate protests from dispossessed farmers, even when they are compensated at market prices (or even above by a certain proportion), as verified in detailed household surveys from India (Ghatak et al. 2013). This suggests the role of heterogeneous security or collateral value placed on land by these owners that is insufficiently compensated. Additional losses are incurred by tenants and agricultural workers on acquired lands who are typically provided little or no compensation. These losses need to be traded off against the productivity advantages of moving land to nonagricultural uses. If tenants earn incentive rents in sharecropping contracts owing to agency problems, welfare optimal land acquisition policies would mandate that landowners provide their tenants some compensation when they are evicted as a consequence of sale of land by the landlord, as shown by Ghatak & Mookherjee (2014). Such regulations would slow the rate of structural transformation and thus increase land misallocation between rural and urban uses, while increasing agricultural productivity owing to increased ex ante investment incentives of landlords and tenants.

7. CONCLUDING COMMENTS

This review contrasts the macrodevelopment approach to misallocation with the related literature on microdevelopment. The macrodevelopment approach has made important advances in quantifying aggregate productivity losses resulting from microeconomic distortions, and provides insights into the role of heterogeneity, dynamics, and general equilibrium effects in understanding effects of external shocks and policy variations. However, it does so mostly using first-best models that abstract from market or institutional failures. The microdevelopment approach instead focuses on distortions resulting from market failure. In such second-best settings, welfare implications of policies frequently differ from first-best settings; second-best policies are highly context dependent and often involve productive misallocation. The policy implications of the two approaches consequently differ markedly—what may be an inefficient policy in a first-best setting may be a constrained efficient policy response to a market failure in a

second-best one. This suggests the need for future research to identify the underlying source of productive misallocation in any given context. And if market frictions seem relevant, more effort needs to be devoted to studying welfare effects of alternative policies using a second-best model that is appropriate for the given context.

This is challenging, and the literature is nascent, so much remains to be done. Yet there are hopeful signs of progress. To provide some notable instances, David & Venkateswaran (2019) make progress on the model identification front by disentangling the roles of information and financial frictions, unobserved heterogeneity in markups and technology, size-dependent policies, and capital adjustment costs in generating misallocation. With regard to second-best welfare analysis in settings with missing insurance markets, Allen & Atkin (2022) distinguish between welfare effects of transport improvements on productive efficiency and on risk borne by Indian farmers, while Lagakos et al. (2023) and Munshi & Rosenzweig (2016) evaluate welfare effects of alternative policies in a rural-urban migration setting with missing insurance markets. Donald et al. (2024) provide a methodology for calculating welfare effects of productivity or infrastructure changes based on a model of spatial economies with financial and trade frictions, agglomeration, and congestion externalities and apply it to evaluate hypothetical policy changes in the US economy. Dávila & Schaab (2022) develop a general methodology in dynamic stochastic economies with heterogeneous agents for estimating welfare effects of policies or shocks on risk sharing, intertemporal smoothing, and redistribution apart from aggregate productive efficiency.

We hope that continuing along these lines, there will be a wider range of frictions that will be studied, with in-depth exploration of their microfoundations with an eye to diagnosing the main sources of misallocation. Moreover, going beyond productivity and looking at welfare aspects will give us a better understanding of why misallocation exists in the first place and the resulting policy implications. This promises to be a rich and exciting research agenda where some of the classic issues of the causes of underdevelopment can be revisited and new ones can be explored, expanding the sources of frictions from policy distortions and market failures to social norms (e.g., discrimination), intrahousehold resource allocation, and behavioral biases.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Linchuan Xu and Tianyu Zhang provided excellent research assistance. We are grateful to Diego Restuccia and Francisco Buera for useful discussions, and Benjamin Moll and Debraj Ray for helpful suggestions. We alone remain responsible for all errors or omissions.

LITERATURE CITED

- Adamopoulos T, Brandt L, Leight J, Restuccia D. 2022. Misallocation, selection, and productivity: a quantitative analysis with panel data from China. *Econometrica* 90(3):1261–82
- Adamopoulos T, Restuccia D. 2020. Land reform and productivity: a quantitative analysis with micro data. *Am. Econ. J. Macroecon.* 12(3):1–39
- Allen T, Atkin D. 2022. Volatility and the gains from trade. *Econometrica* 90(5):2053–92
- Aragon FM, Restuccia D, Rud JP. 2024. Assessing misallocation in agriculture: plots versus farms. *Food Policy* 128:1026–78
- Asker J, Collard-Wexler A, Loecker J. 2014. Dynamic inputs and resource (mis)allocation. *J. Political Econ.* 122(5):1013–63

- Atkinson AB, Stiglitz JE. 1976. The design of tax structure: direct versus indirect taxation. *J. Public Econ.* 6:55–75
- Ayerst S, Nguyen DM, Restuccia D. 2024. *The micro and macro productivity of nations*. NBER Work. Pap. 32750
- Baland JM, Francois P. 2005. Commons as insurance and the welfare impact of privatization. *J. Public Econ.* 89:211–31
- Banerjee A. 1999. *Land reforms: prospects and strategies*. Work. Pap. 99-24, Dep. Econ., MIT, Cambridge, MA
- Banerjee A, Munshi K. 2004. How efficiently is capital allocated: evidence from the knitted garment industry in Tirupur. *Rev. Econ. Stud.* 71(1):19–42
- Banerjee AV, Gertler PJ, Ghatak M. 2002. Empowerment and efficiency: tenancy reform in West Bengal. *J. Political Econ.* 110(2):239–80
- Baqae DR, Farhi E. 2019. Productivity and misallocation in general equilibrium. *Q. J. Econ.* 135(1):105–63
- Bardhan P, Mookherjee D. 2011. Subsidized farm input programs and agricultural performance: a farm-level analysis of West Bengal's green revolution, 1982–1995. *Am. Econ. J. Appl. Econ.* 3(4):186–214
- Bau N, Matray A. 2023. Misallocation and capital market integration: evidence from India. *Econometrica* 91(1):67–106
- Besley T, Burgess R. 2000. Land reform, poverty reduction, and growth: evidence from India. *Q. J. Econ.* 115(2):389–430
- Bloom N, Schankerman M, Van Reenen J. 2013. Identifying technology spillovers and product market rivalry. *Econometrica* 81(4):1347–93
- Boehm J, Oberfield E. 2020. Misallocation in the market for inputs: enforcement and the organization of production. *Q. J. Econ.* 135(4):2007–58
- Bryan G, Chowdhury S, Mobarak AM. 2014. Underinvestment in a profitable technology: the case of seasonal migration in Bangladesh. *Econometrica* 82(5):1671–748
- Buera F, Kaboski J, Shin Y. 2011. Finance and development: a tale of two sectors. *Am. Econ. Rev.* 101(5):1964–2002
- Buera F, Kaboski J, Townsend RM. 2023. From micro to macro development. *J. Econ. Lit.* 61(2):471–503
- Buera FJ, Kaboski JP, Shin Y. 2021. The macroeconomics of microfinance. *Rev. Econ. Stud.* 88:126–61
- Burchardi KB, Gulesci S, Lerva B, Sulaiman M. 2017. Moral hazard: experimental evidence from tenancy contracts. *Q. J. Econ.* 134(1):281–347
- Casella A, Rauch JE. 2002. Anonymous market and group ties in international trade. *J. Int. Econ.* 58:19–47
- Cavalcanti T, Kaboski JP, Martins B, Santos C. 2024. *Financing costs and development*. STEG Work. Pap. 92
- Chen C, Restuccia D, Raul SL. 2023. Land misallocation and productivity. *Am. Econ. J. Macroecon.* 15(2):441–65
- Combes PP, Duranton G, Gobillon L, Puga D, Roux S. 2012. The productivity advantages of large cities: distinguishing agglomeration from firm selection. *Econometrica* 80(6):2543–94
- Dai R, Mookherjee D, Munshi K, Zhang X. 2023. *Community networks, entrepreneurship and the process of economic development*. Work. Pap., Dep. Econ., Boston Univ.
- Dai R, Mookherjee D, Quan Y, Zhang X. 2021. Industrial clusters, networks and resilience to the COVID-19 shock in China. *J. Econ. Behav. Organ.* 183:433–55
- David JM, Venkateswaran V. 2019. The sources of capital misallocation. *Am. Econ. Rev.* 109(7):2531–67
- Dávila E, Schaab A. 2022. *Welfare assessments with heterogeneous individuals*. NBER Work. Pap. 30571
- Diamond PA, Mirrlees JA. 1971. Optimal taxation and public production I: production efficiency. *Am. Econ. Rev.* 61(1):8–27
- Donald E, Fukui M, Miyauchi Y. 2024. *Unpacking aggregate welfare in a spatial economy*. Work. Pap. 400, Inst. Econ. Dev., Boston Univ.
- Emran MS, Stiglitz JE. 2005. On selective indirect tax reform in developing countries. *J. Public Econ.* 89:599–623
- Eswaran M, Kotwal A. 1986. Access to capital and agrarian production organisation. *Econ. J.* 96(382):482–98
- Fafchamps M. 2001. Networks, communities, and markets in Sub-Saharan Africa: implications for firm growth and investment. *J. African Econ.* 10:109–42
- Fajgelbaum PD, Morales E, Serrato JCS, Zidar O. 2019. State taxes and spatial misallocation. *Rev. Econ. Stud.* 86:333–76
- Foster AD, Rosenzweig MR. 2022. Are there too many farms in the world? Labor market transaction costs, machine capacities, and optimal farm size. *J. Political Econ.* 130(3):636–80

- Geanakoplos J, Polemarchakis H. 1986. Existence regularity and constrained sub-optimality of competitive allocations when asset market is incomplete. In *Uncertainty, Information and Communication: Essays in Honor of Kenneth J. Arrow*, Vol. III, ed. WP Heller, RM Starr, DA Starrett. Cambridge Univ. Press
- Ghatak M, Mookherjee D. 2014. Land acquisition for industrialization and compensation of displaced farmers. *J. Dev. Econ.* 110:303–12
- Ghatak M, Mookherjee D. 2024a. *Capital frictions and misallocation with an s-shaped production function*. Work. Pap., Dep. Econ., Boston Univ.
- Ghatak M, Mookherjee D. 2024b. *Misallocation implications of security and collateral value of land*. Work. Pap. 401, Inst. Econ. Dev., Boston Univ.
- Ghatak M, Roy S. 2007. Land reform and agricultural productivity in India: a review of the evidence. *Oxf. Rev. Econ. Policy* 23(2):251–69
- Ghatak M, Sandip M, Mookherjee D, Nath A. 2013. Land acquisition and compensation in Singur: What really happened? *Econ. Political Wkly.* 58(21):32–44
- Gollin D, Udry C. 2020. Heterogeneity, measurement error, and misallocation: evidence from African agriculture. *J. Political Econ.* 129(1):1–80
- Gopinath G, Kalemli-Ozcan S, Karabarbounis L, Villegas-Sanchez C. 2017. Capital allocation and productivity in South Europe. *Q. J. Econ.* 132(4):1915–67
- Greenstone M, Hornbeck R, Moretti E. 2010. Identifying agglomeration spillovers: evidence from winners and losers of large plant openings. *J. Political Econ.* 118(3):536–98
- Greenwald BC, Stiglitz JE. 1986. Externalities in economies with imperfect information and incomplete markets. *Q. J. Econ.* 101(2):229–64
- Gupta B, Mookherjee D, Munshi K, Sanclemente M. 2022. Community origins of industrial entrepreneurship in colonial India. *J. Dev. Econ.* 159:102973
- Haltiwanger J, Kulick R, Syverson C. 2018. *Misallocation measures: the distortion that ate the residual*. NBER Work. Pap. 24109
- Harrison A, Rodríguez-Clare A. 2010. Trade, foreign investment, and industrial policy for developing countries. In *Handbook of Development Economics*, Vol. 5, ed. D Rodrik, M Rosenzweig, pp. 4039–214. Elsevier
- Henderson JV. 1974. The sizes and types of cities. *Am. Econ. Rev.* 64(4):640–56
- Hopenhayn HA. 2014. Firms, misallocation, and aggregate productivity: a review. *Annu. Rev. Econ.* 6:735–70
- Hsieh CT, Klenow PJ. 2009. Misallocation and manufacturing TFP in China and India. *Q. J. Econ.* 124(4):1403–48
- Itskhoki O, Moll B. 2019. Optimal development policies with financial frictions. *Econometrica* 87(1):139–73
- Lagakos D. 2020. Urban-rural gaps in the developing world: Does internal migration offer opportunities? *J. Econ. Perspect.* 34(3):174–92
- Lagakos D, Mobarak AM, Waugh ME. 2023. The welfare effects of encouraging rural-urban migration. *Econometrica* 91(3):803–37
- Lucas RE. 1988. On the mechanics of economic development. *J. Monetary Econ.* 22(1):3–42
- McMillan J, Woodruff C. 1999. Interfirm relationships and informal credit in Vietnam. *Q. J. Econ.* 114(4):1285–320
- Montero E. 2022. Cooperative property rights and development: evidence from land reform in El Salvador. *J. Political Econ.* 130(1):48–93
- Mookherjee D. 1997. Informational rents and property rights in land. In *Property Relations, Incentives, and Welfare*, ed. JE Roemer. Macmillan
- Mookherjee D, Munshi K. 2024. *Misallocation and welfare implications of entrepreneurial networks*. Work. Pap., Dep. Econ., Boston Univ.
- Morando B. 2023. Subsistence farming and factor misallocation: evidence from Ugandan agriculture. *World Bank Econ. Rev.* 37(4):570–98
- Moretti E. 2004. Workers' education, spillovers, and productivity: evidence from plant-level production functions. *Am. Econ. Rev.* 94(3):656–90
- Munshi K. 2011. Strength in numbers: networks as a solution to occupational traps. *Rev. Econ. Stud.* 78:1069–101

- Munshi K. 2014. Community networks and the process of development. *J. Econ. Perspect.* 28(4):49–76
- Munshi K, Rosenzweig M. 2016. Networks and misallocation: insurance, migration and the rural-urban wage gap. *Am. Econ. Rev.* 106(1):46–98
- Newbery DMG, Stiglitz JE. 1984. Pareto inferior trade. *Rev. Econ. Stud.* 51(1):1–12
- Restuccia D, Rogerson R. 2008. Policy distortions and aggregate productivity with heterogeneous plants. *Rev. Econ. Dyn.* 11(4):707–20
- Restuccia D, Rogerson R. 2013. Misallocation and productivity. *Rev. Econ. Dyn.* 16(1):1–10
- Restuccia D, Rogerson R. 2017. The causes and costs of misallocation. *J. Econ. Perspect.* 31(3):151–74
- Rodrik D. 2004. *Industrial policy for the twenty-first century*. Fac. Res. Work. Pap. 47, Harvard Kennedy School, Cambridge, MA
- Romer PM. 1986. Increasing returns and long-run growth. *J. Political Econ.* 94(5):1002–37
- Rossignol EL, Lowes S, Montero E. 2024. *Fallow lengths and the structure of property rights*. NBER Work. Pap. 32226
- Rotemberg M, White TK. 2021. *Plant-to-table(s and figures): processed manufacturing data and measured misallocation*. Work. Pap., Dep. Econ., New York Univ.
- Shaban RA. 1987. Testing between competing models of sharecropping. *J. Political Econ.* 95(5):893–920
- Stiglitz JE. 1982. The inefficiency of the stock market equilibrium. *Rev. Econ. Stud.* 49(2):241–61
- Stiglitz JE. 1988. Economic organization, information, and development. In *Handbook of Development Economics*, Vol. 1, ed. H Chenery, T Srinivasan, pp. 93–160. Elsevier
- Stiglitz JE. 2017. Industrial policy, learning, and development. In *The Practice of Industrial Policy*, ed. J Page, F Tarp, pp. 23–39. Oxford Univ. Press
- Stiglitz JE, Dasgupta P. 1971. Differential taxation, public goods, and economic efficiency. *Rev. Econ. Stud.* 38(2):151–74
- Udry C. 1996. Gender, agricultural production, and the theory of the household. *J. Political Econ.* 104(5):1010–46
- Wang J-K, Kim O. 2024. *Land reform in Taiwan, 1950-1961: effects on agriculture and structural change*. Work. Pap., Penn State Univ., University Park, PA