INSTITUTE OF DEVELOPING ECONOMIES

IDE Discussion Papers are preliminary materials circulated
to stimulate discussions and critical comments

IDE DISCUSSION PAPER No.959

The paradox of productivity and welfare: The role of institutional markdown

Wenyin Cheng[1*], Bo Meng[1], Tangjun Yuan[2]

Feburary 2025

Abstract
The connection between total factor productivity (TFP) growth and overall welfare has traditionally been regarded as positive. However, limited attention has been given to the fact that certain components of TFP growth may, in fact, have a detrimental impact on welfare. By incorporating markdowns into the decomposition model of revenue-based TFP growth, we identify three key components: the markdown effect, which captures the deviation of sectoral average markdown from unity; the reallocation effect, which reflects variations in markdowns across firms; and the technical effect, which represents physical TFP. Using firm-level data from China from 2000 to 2007, we derive several empirical insights. First, the markdown on labor has been decreasing, but this trend came to a halt after 2004—primarily due to a reduction in institutional power, rather than shifts in market power or the effects of the Lewis turning point. Second, markdowns contribute to approximately one-third of TFP growth, surpassing the impact of the reallocation effect in 2007. Additionally, the welfare loss due to markdowns accounted for 5.65% of total value added in 2007. Finally, using the reallocation effect as a reference, we assess the potential for a domestic Flying Geese pattern across regions. This study advances the theoretical discourse on markdowns and their implications for the productivity-welfare relationship, shedding new light on the complex interactions between distortions, firm dynamics, and economic welfare.

Keywords: Institutional markdown, Reallocation; Productivity growth; Welfare; Flying geese.
JEL classification: F14, L52, R15

1: Institute of Developing Economies, Japan External Trade Organization (Wenyin Cheng: wenyin_cheng@ide.go.jp; Bo Meng: bo_meng@ide.go.jp); 2: Fudan University (e1103@hotmail.com). *: corresponding author.

The Institute of Developing Economies (IDE) is a semigovernmental, nonpartisan, nonprofit research institute, founded in 1958. The Institute merged with the Japan External Trade Organization (JETRO) on July 1, 1998. The Institute conducts basic and comprehensive studies on economic and related affairs in all developing countries and regions, including Asia, the Middle East, Africa, Latin America, Oceania, and Eastern Europe.

INSTITUTE OF DEVELOPING ECONOMIES (IDE), JETRO
3-2-2, WAKABA, MIHAMA-KU, CHIBA-SHI
CHIBA 261-8545, JAPAN

The paradox of productivity and welfare: The role of institutional markdown

Wenyin Cheng Bo Meng Tangjun Yuan

Abstract: The connection between total factor productivity (TFP) growth and overall welfare has traditionally been regarded as positive. However, limited attention has been given to the fact that certain components of TFP growth may, in fact, have a detrimental impact on welfare. By incorporating markdowns into the decomposition model of revenue-based TFP growth, we identify three key components: the **markdown effect**, which captures the deviation of sectoral average markdown from unity; the **reallocation effect**, which reflects variations in markdowns across firms; and the **technical effect**, which represents physical TFP. Using firm-level data from China from 2000 to 2007, we derive several empirical insights. First, the markdown on labor has been decreasing, but this trend came to a halt after 2004—primarily due to a reduction in institutional power, rather than shifts in market power or the effects of the Lewis turning point. Second, markdowns contribute to approximately one-third of TFP growth, surpassing the impact of the reallocation effect in 2007. Additionally, the welfare loss due to markdowns accounted for 5.65% of total value added in 2007. Finally, using the reallocation effect as a reference, we assess the potential for a domestic *Flying Geese* pattern across regions. This study advances the theoretical discourse on markdowns and their implications for the productivity-welfare relationship, shedding new light on the complex interactions between distortions, firm dynamics, and economic welfare.

Key Words: Institutional markdown, Reallocation; Productivity growth; Welfare; Flying geese.

JEL: D24, D41, D61

1. Introduction

The relationship between total factor productivity (TFP) growth and overall welfare has traditionally been perceived as positive. However, little attention has been paid to the possibility that certain components of TFP growth may, in fact, have adverse welfare implications. The markdown on labor, defined as the ratio of wages to the marginal productivity of labor, arises from disparities in bargaining power between labor and management. These disparities can be driven by institutional power (e.g., artificially imposed segregation) in non-market economies or market power (e.g., monopolistic control) in market economies.

While market power is widely cited in the literature as a primary explanation for markdowns (e.g. Lu et al., 2019; Yeh et al., 2022; Alviarez et al., 2023; Rubens, 2023; Xie et al., 2024), institutional power has received comparatively less attention. Notably, migrant labor in China and black labor in South Africa have been recognized as quintessential examples of institutional power (Qin, 2015)[1]. This form of institutional power played a crucial role in driving rapid economic growth in China following the reform and opening-up policy of 1978 and in South Africa during the Apartheid era, particularly between the 1950s and early 1970s.

Markdown stems from distortions; however, it differs from misallocation that also arises from distortions. Reallocation is primarily influenced by firm behaviors within market competition, whereas weak labor rights is shaped by government policies or monopoly. Therefore, it is necessary to distinguish between the two types of distortions. Existing literature on the components of TFP growth primarily focuses on reallocation effect or technical effect (e.g. Baily et al., 1992; Griliches and Regev, 1995; Foster et al., 2001; Cheng et al., 2025). Reallocation primarily refers to the movement of resources from firms with low technological efficiency to those with high technological efficiency, given the existing technology. The technical effect, on the other hand, refers to improvements in technology itself. However, this dichotomy blurs the critical role of markdown.

Markdown of labor (MDL) in China stems from its dual urban-rural structure, while in South Africa, it arose from the Apartheid regime. The urban-rural divide was institutionalized through the *hukou* system, which emerged in the 1950s. The *hukou* system categorized citizens as either rural or urban, and the mobility of rural populations to urban areas was highly restricted. Since 1978, these restrictions have been gradually eased, leading to a sharp increase in the number of migrant workers, particularly after China's accession to the WTO (see Figure 1). However, before 2004[2], large numbers of surplus rural workers migrating to urban can only obtain living wages, far lower than their marginal contributions to products, i.e. marginal labor productivity. This can be considered as markdown.

The MDL can be quantified by a comparison between wages and marginal labor productivity. If labor is underpaid relative to its contribution to output, it can be

[1] Qin (2015) uses the term "weak labor rights," which, in economic terms, corresponds to the institutional markdown on labor. Refugee populations are also a consequence of institutional power rather than market power.

[2] Cai (2007) and Zhang *et al*. (2011) take the year 2004 as the Lewis turning point of China.

considered a manifestation of markdown. Therefore, we utilize the concept of MDL, the ratio of wage to marginal labor productivity, to measure labor rights. Then, we incorporate MDL into the TFP model to quantify the effect of markdown on TFP growth. Furthermore, we will calibrate the welfare implications of markdown. While both reallocation and markdown can contribute to TFP growth, their welfare implications differ significantly: reallocation driven by market competition tends to enhance welfare, whereas markdown stemming from government policies are more likely to undermine it.

In this paper we extend the framework of Petrin and Levinsohn (2012) by introducing the markdown. Similar to MDL, we have markdown of capital (MDK), the ratio of capital price to marginal capital productivity. We break down aggregate productivity growth into three components: technical effect, reallocation effect and rights effect. Reallocation effect stems from the resource mobility among firms with different markdown, while rights effect comes from the part that markdown lower than one. Whereas our primary focus is on the effects of labor rights, it is inevitable to include the effects of capital rights in the model unless we make the strong assumption that the unit cost of capital is fixed at 1. Then we take this model to China's firm-level data to investigate the contributions of the three components to productivity growth and social welfare.

We make contributions to at least three strands of literature. First, this is the first attempt to introduce markdown into the framework of TFP decomposition. This approach allows us to break down TFP growth into three distinct components, moving beyond the traditional two-part framework. In the past two decades there has been greatly increased interest in aggregating and decomposing productivity growth using micro data. The most common approach is by aggregating the plant-level technological changes with output or input share weight, and then decomposing it into reallocation effect and technical effect with the method proposed by Baily, Hulten and Campbell (1992), Griliches and Regev (1995), Olley and Pakes (1996), Foster, Haltiwanger and Krizan (2001), and Melitz and Polanec (2012). This kind of shift-share method based on physical TFP (TFPQ) is problematic because they neglect the fact that some reallocations have nothing to do with technical changes but are only induced by the changes of final demand. Granted that technology remains constant, changes in input and output prices are also likely to result in reallocation. Another most widely used approach is developed in turn by Hulten (1978), Basu and Fernald (2002), Basu *et al.* (2009), Petrin and Levinsohn (2012), which associates aggregate productivity growth with aggregate final demand, and thus solves the above problem. Besides, this framework based on revenue TFP (TFPR) manifests the productivity's welfare meaning more directly. However, none of these alternative approaches informs us the contribution of markdown to TFP growth.

Second, we enrich the literature on the productivity-welfare nexus by highlighting that not all components of productivity growth would lead to welfare improvements; for instance, productivity gains achieved at the expense of markdown can result in adverse welfare outcomes. However, the existing literature (e.g., Basu et al., 2009; Petrin and Levinsohn, 2012; Basu et al., 2022; Liu et al., 2023) primarily examines the positive relationship between TFP and welfare. Magee (1971) argues that both price differentials and price rigidity have traditionally been considered forms of factor market distortions. While the former is associated with reallocation effects, the latter pertains to labor rights effects. Nevertheless, these two effects have often been treated as equivalent in the literature (e.g., Hsieh and Klenow, 2009; Zhang et al., 2023).

Finally, our study advances the literature on markdown by shifting the focus from market power to institutional power, offering a quantitative evaluation of its effects on productivity growth and welfare. Although markdown has only recently garnered attention in the industrial organization literature—where it is primarily examined through the lens of market power—the phenomenon in China is largely driven by institutional power rather than conventional market forces.

The remainder of this paper proceeds as follows: In the next section, we review relevant literature. Section 3 lay out the data and our theoretical framework for productivity decomposition. In section 4, we take our model to data and analyses the three elements of aggregate productivity growth. The welfare losses brought by misallocation and markdown are calculated. We then further measure the potential labor flows and capital flows to identify the pattern of domestic flying geese in Section 5. Section 6 provides certain concluding remarks.

2. Markdown in China

2.1 Measure of markdown

A lower MDL suggests that workers receive wages significantly below their marginal labor productivity. This disparity reflects the imbalance of bargaining power between firms and employees, which may stem from an artificially imposed segregation system or monopoly power. In an economy without any distortion, the MDL should be equal to one. Therefore, we use the gap between sectoral average MDL and one as an indicator of markdown. Ideally, a zero gap implies that workers are compensated exactly in line with their marginal contributions to output, indicating the presence of full labor rights.

The concept of MDL has been widely used in economic research on such topics as inflation (Sbordone, 2002; King and Watson, 2012), unemployment (Rebitzer, 1988), international competitiveness (Wu and Yu, 2022), etc. Nonetheless, its significance in comprehending TFP and welfare growth has been underestimated. In fact, a widening gap between MDL and one leads to higher TFP growth as lower costs result in increased output. However, it leads to lower welfare growth because labors are increasingly underpaid as well.

The level of MDL focus on the average markdown of a sector or the whole economy. Yet, variations in MDLs across firms within a sector or the whole economy would highlight the presence of misallocations. If MDLs were uniform across all sectors, the misallocation, or the potential for realizing a positive reallocation effect, would be non-existent. Labors moving from sectors with lower labor productivity (higher MDL) to those with higher labor productivity (lower MDL) would bring positive reallocation effect. Therefore, by introducing the concept of MDL, we are able to differentiate between the labor rights effect and reallocation effect.

Numerous works have analyzed misallocation based on the variations in marginal product, as evidenced by studies like Hsieh and Klenow (2009) and Banerjee and Duflo (2005). There are some other papers that analyze the potential for reallocation based solely on the variations in input cost. The wages in China have been on the rise since the beginning of the new century. Consequently, some scholars contend that China is gradually losing its comparative advantage in manufacturing, leading to a potential shift of factories to other Southeast Asian countries like Vietnam and Malaysia. However, the complete essence of comparative advantage is not captured solely by the marginal product or the input price; a concurrent consideration of both is essential.

Without markdown, China would have resembled Eastern Europe; without the reform and opening-up policy, it would have been more like North Korea. Markdown can be traced back to the 1950s when China established the Hukou system. However, it was not until the launch of the reform and opening-up policy in 1978 that markdown became a source of comparative advantage for China's economic growth.

After 1978, the restrictions on labor mobility between rural and urban areas were gradually eased. During the 1980s, most migration was limited to "leaving the land without leaving the county." It was not until the early 1990s that large-scale migration of workers beyond county boundaries began. The number of outmigrant workers increased significantly following China’s accession to the WTO. (see Figure 1).

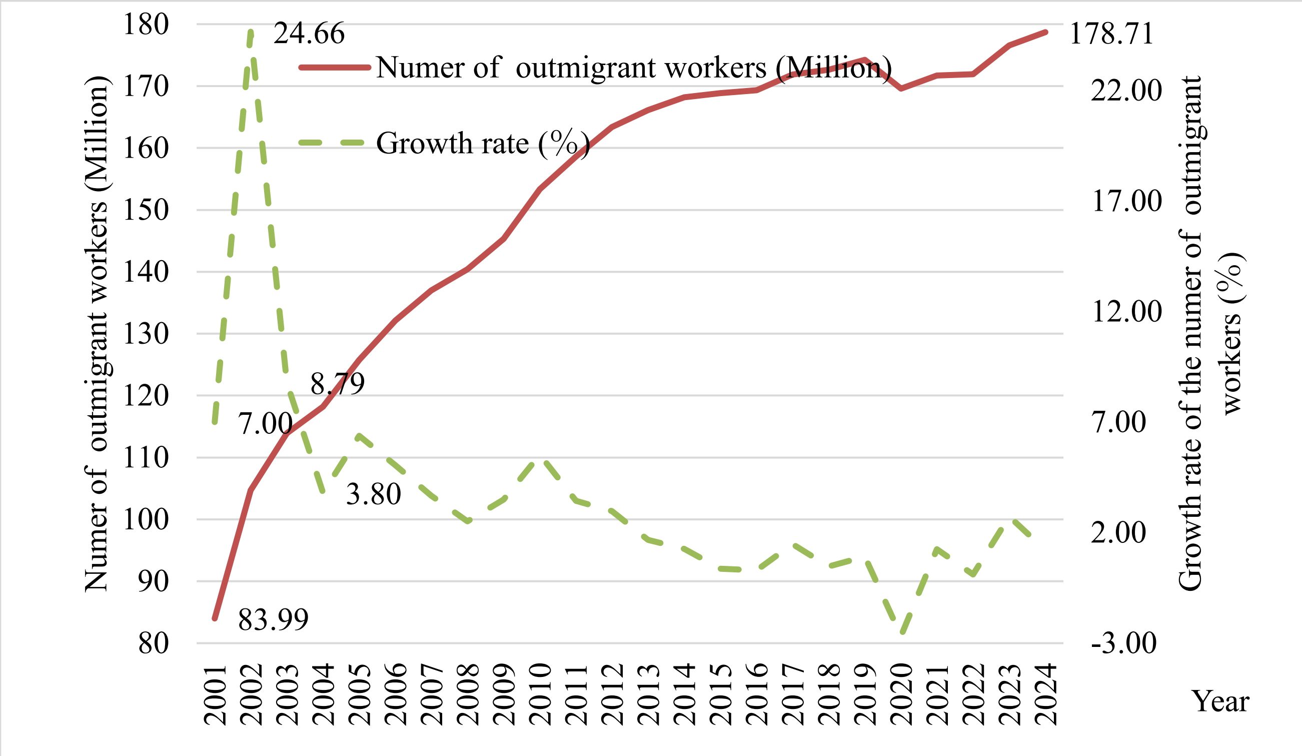

Figure 1. Laborers working outside the county

Source: Data from 2001 to 2007 is derived from Cai (2010); data from 2008 to 2024 is obtained from the annual *Migrant Worker Monitoring and Survey Report* released by the National Bureau of Statistics of China.

Weak labor rights have long been used as a diplomatic excuse by foreign critics of China. According to data from the World Justice Project (WJP) Rule of Law Index, China ranks 139th among the 142 countries included in the index in terms of labor rights in 2024.[3] However, whenever China questions how such rapid development could be achieved without human rights, foreign critics often struggle to provide an answer. Obviously, foreign critics overlook the fact that when markdown meets with the opening-up policy, they become a comparative advantage in driving economic growth. Foreign critics also significant improvements in labor rights since 1978, as evidenced by large-scale rural-to-urban labor migration and rising wages. Markdown may contribute to economic growth for a certain period; however, such growth is unsustainable without labor rights improvements, as evidenced by rising wages of outmigrant workers in China.

2.2 Reduction in institutional markdown or Lewis turning point?

Using survey data from three counties in Gansu Province, Zhang et al. (2011) show that the average wage of outmigrant workers remained around 17.0 RMB per day from

[3] The labor rights indicator in the index encompasses sub-indicators such as equal payment, labor discrimination, unionization and collective bargaining, as well as child and forced labor https://worldjusticeproject.org/rule-of-law-index/country/2024/China/Fundamental%20Rights/

1993 to 2003 but surged to 26.8 RMB in 2006, and the share of outmigrant workers increased by approximately 12 percentage points in both periods (1993–2003 and 2003–2006). This pattern is further confirmed by data from two nationwide surveys. Evidently, the markdown stemming from the urban-rural divide was particularly pronounced before 2003.

Figure 2 presents the markdowns of three selected countries. As shown, China's MDL is significantly lower than that of Japan, the U.S. and even Mexico, suggesting that MDL in China were more serious during that period. The MDL has been declining since China's accession to the WTO, driven by rising labor productivity while wages remained stable before 2004. However, after 2004, the MDL stopped declining, likely due to wage increases, as shown above.

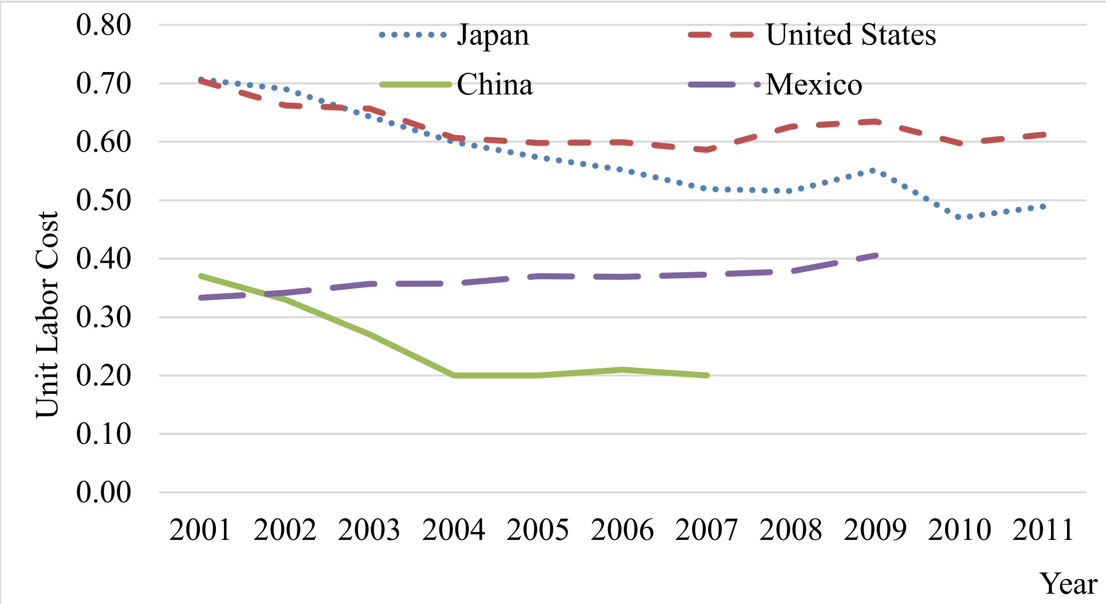

Figure 2. Unit labor cost in China, Mexico, Japan and the U.S.

Source: The markdowns for the manufacturing sectors in Mexico, Japan and the U.S. are obtained from the OECD database (https://data.oecd.org/lprdty/unit-labour-costs.htm), while the markdowns for industrial firms in China are calculated by the authors using the Annual Survey of Industrial Firms database.

Many studies attribute the onset of wage growth to the Lewis turning point (Cai, 2007; Zhang et al., 2011). However, according to Lewis's dual economy theory, wages should not rise until the surplus rural labor force has been fully absorbed (Lewis, 1954; Basu, 1997). Yet, even after 2003, the migration of labor from rural to urban areas continued (see Figure 1).

Markdown provides a more compelling explanation for this phenomenon. Firstly, the reduction in MDL driven by the relaxation of urban-rural mobility restrictions leads to the decline of surplus labor, resulting in the wage increase starting in 2003. Secondly, the labor rights in urban areas were improved. The lack of labor union protections in urban areas led to widespread wage arrears among migrant workers after they moved to cities. This issue gained significant media attention in the early 2000s. In 2004, the government work report explicitly addressed the need to ensure wage payments for rural migrants working in cities, improve labor conditions, and initiate institutional reforms to resolve these issues. These measures helped improve labor rights, thereby contributing to wage growth. Thirdly, labor rights in rural areas improved during this period. In 2003, China launched a nationwide pilot program for rural tax and fee reform, and by 2006, agricultural taxes were abolished nationwide. The rise in rural incomes led to higher wage expectations among migrant workers in urban areas.

3. Theoretical Model

3.1 Productivity decomposition model incorporating markdowns

Now we derive our decomposition framework by integrating markdown into the established framework for productivity decomposition. We relax the two key assumptions embedded in the Solow (1957) canonical theory—constant return to scale and perfect competition, and take account of plant-level heterogeneity. For brevity's sake, our model is initially formulated within a continuous time framework, and we will omit the time index where feasible. Aligning with empirical studies, our assumption is that each firm in the economy specializes in the production of a single product, with each product indexed by i.

The production function for good i is $v_i = v_i(l_i, k_i, z_i)$, v_i, l_i, k_i, z_i denotes real value added, labor inputs, real capital inputs, and technology changes respectively. By differentiating the production function, we get:

$$\Delta v_i = \frac{\partial v_i}{\partial l_i}\Delta l_i + \frac{\partial v_i}{\partial k_i}\Delta k_i + \frac{\partial v_i}{\partial z_i} z_i \tag{1}$$

According to Petrin and Levinsohn (2012), aggregate productivity growth can be computed by subtracting the growth in overall input costs from the growth in aggregate final demand (aggregate value added):

$$\Delta A = \sum_i P_i \Delta v_i - \sum_i W_i \Delta l_i - \sum_i R_i \Delta k_i \tag{2}$$

Substitute (1) into (2), and normalize $\frac{\partial v_i}{\partial z_i} = 1$, we obtain:

$$\Delta A = \sum_i \left(P_i \frac{\partial v_i}{\partial l_i} - W_i\right)\Delta l_i + \sum_i \left(P_i \frac{\partial v_i}{\partial l_i} - R_i\right)\Delta k_i + \sum_i P_i \Delta z_i \tag{3}$$

Petrin and Levinsohn (2012) made significant strides by conceptualizing the 'wedge' between the marginal product and input cost. Yet, a firm's decisions regarding liquidation and input adjustments are often influenced by the rate of profit rather than the absolute profit itself. Thus, it is the proportion of input cost to marginal product (i.e. markdown), not the mere difference between them, that is instrumental in guiding resource reallocation and productivity. Therefore, we express the aggregate productivity growth rate (hereafter APG) as:

$$\dot{A} = \sum_i \omega_i \varepsilon_{li}(1-u_{li})\dot{l}_i + \sum_i \omega_i \varepsilon_{ki}(1-u_{ki})\dot{k}_i + \sum_i \omega_i \dot{z}_i \qquad (4)$$

where $\omega_i = \frac{P_i v_i}{\sum_i P_i v_i}$ is the value-added share weight. $u_{li} = \frac{W_i}{P_i} / \frac{\partial v_i}{\partial l_i}$ and $u_{ki} = \frac{R_i}{P_i} / \frac{\partial v_i}{\partial k_i}$ represent the MDL and capital, respectively, that is, the ratio of factor price to marginal productivity. ε_{li} and ε_{ki} are the output elasticity of labor and capital estimated with value added production function. Whereas z_i refers to the physical TFP at firm level, A refers to the revenue TFP at the aggregate level.

By introducing $\bar{u}$—the weighted average of u , we can rewrite equation (4), and obtain three elements of productivity growth based on markdown: reallocation effect (RE), weak factor rights effect (WR) and technical effect (TE):

$$\dot{A} = RE + MD + TE = \left[\sum_i \omega_i \varepsilon_{li}(\bar{u}_l - u_{li})\dot{l}_i + \sum_i \omega_i \varepsilon_{ki}(\bar{u}_k - u_{ki})\dot{k}_i\right]$$

$$+\left[\sum_i \omega_i \varepsilon_{li}(1-\bar{u}_l)\dot{l}_i + \sum_i \omega_i \varepsilon_{ki}(1-\bar{u}_k)\dot{k}_i\right] + \sum_i \omega_i \dot{z}_i \qquad (5)$$

$RE = \sum_i \omega_i \varepsilon_{li}(\bar{u}_l - u_{li})\dot{l}_i + \sum_i \omega_i \varepsilon_{ki}(\bar{u}_k - u_{ki})\dot{k}_i$ is a measure of reallocation effect. It echoes the basic idea of Foster et al. (2001) as a refinement to the TFP decomposition approach originally presented by Baily et al. (1992) and Griliches et al. (1995), that is, incorporating a reference average level into the reallocation term. This term also reveals the distinct reallocation impacts of labor and capital. Productivity growth can be realized when there is a reduction of inputs in firms bearing higher UFCs alongside an expansion of inputs in firms with lower UFCs, which means the resource reallocation from firms with a higher UFCs to those with a lower UFCs, or from lower marginal productivity to higher marginal productivity.

$MD = \sum_i \omega_i \varepsilon_{li}(1-\bar{u}_l)\dot{l}_i + \sum_i \omega_i \varepsilon_{ki}(1-\bar{u}_k)\dot{k}_i$ measures effect of markdown effct of labor and capital. When markdown is depressed to the level lower than one, which is especially serious in developing countries[4], an additional input use would increase APG. But, as we will show in the following, this will lead to more welfare losses. Institutional markdown can be assumed to be uniform within a given sector. Therefore, we use the sectoral average markdown here.

$TE = \sum_i \omega_i \dot{z}_i$, this is the same old technical efficiency term, which enhances productivity growth without the increment of inputs.

The distinction between RE and WR is straightforward, but it has not been previously noted because no one has incorporated markdown into the productivity decomposition framework, and thus the difference between the two effects is always

4 During 2001-2011, the MDL of developed countries such as America, Germany, Japan, South Korea, ranges from 0.6 to 0.8, but that in Mexico is about 0.4). We get the data from

concealed.[5] For example, it's not easy to find the reference level if we take the form of wedge instead of ratio. By bringing in the markdown we have successfully distinguished between RE and WR. Since innovation has been analyzed fully in previous literatures, our main goal in this paper is to study the productivity and welfare meaning of weak factor rights. Appendix II gives the detailed evolution of productivity decomposition framework. One advantage of our model is that it successfully accommodates non-neoclassical settings by considering markdown.

3.2 Welfare implication of markdown

The welfare implications of markdown are straightforward. A lower MDL indicates that workers receive less than what they should, leading to welfare losses for them. The Harberger triangle characterize the deadweight loss from such distortions. However, it is challenging to provide precise functions for the supply and demand of labor. Therefore, we make certain assumptions to quantify the social welfare loss resulting from markdown. Specifically, the integral of a continuous function can be approximated by the summation of discontinuous function. Then the areas of the Harberger triangle for the whole society can be expressed as follows:

$$\int_0^1 (P \cdot MP_l - W)\, dl + \int_0^1 (P \cdot MP_k - R)\, dk$$
$$= (P \cdot MP_l - W)\Delta l + (P \cdot MP_k - R)\Delta k \tag{6}$$

The ratio of welfare loss to value added provides a more intuitive measure than the absolute welfare loss:

$$\begin{aligned} MD^{welfare} &= \frac{(P \cdot MP_l - W)\Delta l}{P \cdot v} + \frac{(P \cdot MP_k - R)\Delta k}{P \cdot v} \\ &= \tfrac{P \cdot MP_l}{P \cdot v}(1 - \bar{u}_l)\Delta l + \tfrac{P \cdot MP_k}{P \cdot v}(1 - \bar{u}_k)\Delta k \\ &= \varepsilon_l(\bar{u}_l - 1)\dot{l} + \varepsilon_k(\bar{u}_k - 1)\dot{k} \\ &= \textstyle\sum_i \omega_i \varepsilon_{li}(1 - \bar{u}_l)\dot{l}_i + \sum_i \omega_i \varepsilon_{li}(1 - \bar{u}_k)\dot{k}_i = MD \end{aligned} \tag{7}$$

where $\varepsilon_l \cdot \dot{l} = \sum_i \omega_i \varepsilon_{li} \dot{l}_i$

It is obvious that the ratio of welfare loss to value added equals the contribution of weak rights to aggregate TFP growth.

---

5 To our knowledge, only Basu and Fernald(2002) and Basu *et al*. (2009) have mentioned the aggregate distortion. But they assume that MDL and MDK equals with each other, and both of them are equated to output markup, which seems to be unconvincing. See appendix II for details.

4. Empirical analysis

4.1 Description of data and variables

The dataset we utilize is a comprehensive firm-level panel derived from the *Annual Survey of Industrial Firms* (ASIF) in China (2000–2007), which has been widely used in the literature (Cheng et al., 2024). This dataset includes all non-state-owned manufacturing firms with annual revenues exceeding 5 million RMB (approximately $600,000 during this period) as well as all state-owned enterprises. We focus exclusively on the manufacturing sector, where the raw sample size ranges from over 150,000 firms in 2000 to approximately 400,000 establishments in 2007. Given the significant reclassification of industries in 2003, we adjust industry codes using the 2002 and 2007 input-output tables to ensure comparability before and after the reclassification.

To maintain data quality and consistency, we apply several filters. We retain only firms with more than eight employees and exclude observations with negative values for key financial variables, including gross output, intermediate input, value added, capital stock, wages, firm age, and depreciation rate. Missing observations for critical variables such as output and input are removed, and extreme values are winsorized at the 0.1% level to mitigate the influence of outliers. The final cleaned dataset comprises 1,625,595 manufacturing firms.

For production function estimation, we employ the Wooldridge-Levinsohn-Petrin (WLP) approach, which not only addresses the common issues of selection bias and simultaneity but also mitigates collinearity problems inherent in the Olley-Pakes (OP) and Levinsohn-Petrin (LP) methods. Details on the estimation methodology and variable construction can be found in *Appendix I*.

Table 1 provides a summary of the dataset. As evident from the table, while data cleaning slightly reduces the sample size, it remains sufficiently large for robust empirical analysis. The average values of key variables are reported, with all variables transformed into logarithmic form to facilitate direct estimation of labor and capital elasticities.

Table 1 Description of Data and Variables

Year	Raw Sample	Final Sample	lnRVA	lnL	lnL2	lnL3	lnK
2000	148277	134054	8.26	4.96	7.03	7.13	8.32
2001	155410	142843	8.37	4.89	7.02	7.12	8.24
2002	166868	153137	8.49	4.87	7.08	7.18	8.24
2003	181186	172305	8.73	4.84	7.12	7.22	8.21

2004	259412	231909	8.86	4.65	7.10	7.16	7.95
2005	251499	238043	9.09	4.71	7.25	7.32	8.12
2006	279282	262104	9.26	4.67	7.35	7.42	8.16
2007	313046	291200	9.40	4.63	7.44	7.51	8.16

Note: RVA refers to real value added, L means labor, K is real capital stock; L2 designates real labor payment including only wage, welfare and unemployment insurance. L3 is real labor payment including wage, welfare, unemployment insurance, housing fund, endowment and medical insurance. The indices of L2, L3 are used as indicators of human capital in our robustness tests.[6]

4.2 Evolution of markdown

By calculating factor prices and the marginal productivity of factors using the formulas outlined above, we derive markdowns. Here, we provide an overview of their overall trends. As shown in Table 2, the MDL declined prior to 2004 and subsequently stabilized, aligning with the findings of Cai (2007) and Zhang et al. (2011), who identify 2004 as China's Lewis turning point. Regardless of the measurement approach, real wages have nearly doubled over this period. Meanwhile, MDK exhibits a persistent downward trend throughout the sample period, suggesting an increasing degree of capital market distortion.

Even in developed economies, perfect competition—where markdowns equal one—is rarely achieved. Given that all countries experience some level of weak rights, defining a precise benchmark for markdown is challenging. However, as previously demonstrated, China’s markdown is significantly lower than that observed in developed economies, indicating relatively weak rights. This implies that a portion of China's productivity growth during this period may be fueled by weak rights. However, the decline in the MDL stalled after 2004, suggesting that government initiatives aimed at strengthening labor rights may have played a role in stabilizing labor costs.

Table 2 Markdown of Labor and Capital

Year	Labor					Capital				
	MPL	realW	realW_all	MDL	MDL_all	MPK	realz	realz_all	MDK	MDK_all
2000	28.22	10.81	12.16	0.38	0.43	0.15	0.11	0.13	0.73	0.87
2001	34.18	11.30	12.67	0.33	0.37	0.18	0.11	0.13	0.62	0.74
2002	42.01	12.26	13.76	0.29	0.33	0.20	0.11	0.13	0.55	0.65
2003	55.54	13.22	14.81	0.24	0.27	0.27	0.11	0.13	0.42	0.49
2004	78.66	14.50	16.12	0.18	0.20	0.38	0.11	0.12	0.28	0.32
2005	86.73	15.73	17.51	0.18	0.20	0.43	0.12	0.15	0.28	0.34
2006	99.58	18.40	20.45	0.18	0.21	0.51	0.12	0.14	0.23	0.27
2007	118.9	21.12	23.29	0.18	0.20	0.57	0.12	0.14	0.21	0.24

6 Due to the absence of data on housing funds, endowment, and medical insurance for the period 2000–2003, we calculate L2 and L3 using available data from 2004–2007. We then derive a payment multiplier by computing the ratio of L3 to L2. Assuming that this multiplier remains consistent over time, we apply it retroactively to estimate L3 for the years 2000–2003.

Note: els_ L denotes elasticity of labor; MPL denotes marginal product of labor; realW means real wage per capita, in thousand yuan per year, it equals the ratio of L2 to L; realz means real capital price. realW (MDL) and realW_all (MDL_all) are calculated with real labor payment based on L2 and L3 respectively. els_K denotes elasticity of capital, MPK denotes marginal product of capital. realz (MDK) and realz_all (MDK_all) denotes items calculated considering and without considering accounts payable respectively.

4.3 Components of aggregate TFP growth

Our purpose of calculating markdown serves to decompose productivity growth into reallocation effect, aggregate distortion, and technical effect. In order to apply our measure to data, we provide Tornqvist-Divisia approximations to make our framework suitable for discrete-time setting. The APG can be rewritten as follows:

Our objective in calculating markdown is to decompose productivity growth into three key components: the weak rights effect, reallocation effect and technical effect. To effectively implement this decomposition in an empirical setting, we employ Tornqvist-Divisia approximations, ensuring that our framework is compatible with a discrete-time context. The aggregate productivity growth (APG) can be reformulated as follows:

$$\dot{A} = RE + WR + TE = \left[\sum_i \omega_i^a \varepsilon_{li}^a (\bar{u}_l^a - \bar{u}_{li}^a)\dot{l}_i + \sum_i \omega_i^a \varepsilon_{ki}^a (\bar{u}_k^a - \bar{u}_{ki}^a)\dot{k}_i\right]$$

$$+\left[\sum_i \omega_i^a \varepsilon_{li}^a (1 - \bar{u}_l^a)\dot{l}_i + \sum_i \omega_i^a \varepsilon_{ki}^a (1 - \bar{u}_k^a)\dot{k}_i\right] + \sum_i \omega_i^a \dot{z}_i \quad (8)$$

where the superscript "a" denotes the average between the current period and the prior year, s represents the ratio of factor cost to value added. For clarity and transparency, we consolidate the contributions of both input factors into a single term, simplifying the interpretation of our decomposition framework.

Table 3 provides clear evidence that the primary driver of productivity growth during the sample period is technological progress. Meanwhile, the reallocation effects of both labor and capital have declined sharply, approaching near zero. This indicates that the efficiency gains from resource mobility within the existing economic structure have been largely exhausted. Consequently, the focus should now shift toward structural reform to alleviate misallocation and enhance social welfare. A detailed discussion on structural transformation will be provided in Part Six.

The rapid increase in the *weak rights effect* serves as a critical warning that China's market system requires urgent reform. Only by undertaking necessary market adjustments can factor suppliers receive compensation more closely aligned with their actual contributions, thereby improving social welfare. It is essential to recognize the distinct policy implications of reallocation and weak rights effects. Previous studies have often conflated these two, leading to an overestimation of the reallocation effect. By isolating aggregate distortion (AD), we can obtain a more accurate measure of the *sheer* reallocation effect (RE). This distinction is particularly crucial for developing countries, where markdowns are significantly lower than in developed economies. The failure to separate these effects may misguide policymakers, shifting their focus disproportionately toward reallocation while delaying essential market reforms.

The weak rights effect remained relatively low in 2001 and 2002, even falling below the reallocation effect. However, it surged following China's accession to the World Trade Organization (WTO). The post-WTO manufacturing boom accelerated the movement of migrant workers from agriculture to industry. Even if unit labor costs remained stable, the rapid expansion of the industrial labor force likely intensified aggregate distortions. Additionally, the persistent surplus of rural labor suppressed wage growth, further reducing unit labor costs. These distortions have now reached a level where delaying market reforms is no longer viable. Immediate structural adjustments are necessary to correct these inefficiencies and ensure sustainable economic development.

Table 3 Decomposition of APG(%)[7]

Year	**APG**	**TE**	**RE**	RE_L	RE_K	**WR**	WR_L	WR_K
2000	-	-	-	-	-	-	-	-
2001	5.00	3.39	0.97	0.62	0.35	0.64	0.19	0.46
2002	1.14	-2.32	1.82	0.82	1.00	1.64	0.88	0.76
2003	16.35	12.35	0.42	0.32	0.10	3.59	2.48	1.11
2004	24.20	20.03	1.43	0.71	0.72	2.75	1.31	1.44
2005	11.20	4.24	-0.44	0.02	-0.46	7.40	4.76	2.64
2006	10.56	5.95	0.09	0.07	0.02	4.52	2.18	2.34
2007	16.25	10.96	-0.37	0.14	-0.51	5.65	2.85	2.80

Note: RE is reallocation effect. RE_L and RE_K designate reallocation of labor and capital. WR is aggregate distortion. Likewise, AD_L and AD_K represent weak rightsof labor and capital market. dtfp=TE+RE+AD; RE=RE_L+RE_K; AD=AD_L+AD_K.

5. Further discussion: Domestic *flying geese*

The contribution of the reallocation effect to TFP growth is quite low, much smaller than that of markdown. This suggests the presence of substantial misallocation within the economy. In other words, the potential TFP gains from improved resource reallocation could be considerable. Therefore, we assess the potential reallocation gains by comparing the actual economic conditions with an optimal scenario in which the variations in markdown are eliminated. By assuming zero variation across regions, we

7 For the lack of data about housing fund, endowment, medical insurance, and accounts payable during 2000-2003, all the related calculation results listed in our paper are computed by neglecting these variables. Of course, we have made some robustness checks by adding these variables, and the results seem robust.

estimate the potential for regional reallocation, effectively illustrating the domestic *flying geese* pattern.

Table 4 presents the distribution of markdowns and potential factor flows. Overall, the patterns in markdown align with the expected direction of factor movements. Specifically, regions with high markdowns need to shed resources, whereas those with lower unit costs should absorb additional inputs. Furthermore, MDL across all regions reflect the emergence of the turning point around the year 2004.

Turning to regional analysis, the eastern region exhibits high MDL but low MDK, suggesting that future factor mobility should favor labor outflows alongside capital inflows. This dynamic underscores the necessity for the eastern region to shift from labor-intensive to capital-intensive industries, a transition that represents industrial upgrading. In contrast, the northeastern region faces the opposite challenge, with low MDL and high MDK, necessitating a process of "industrial downgrading." As a historically significant industrial hub, the northeast is rich in capital-intensive industries. However, these industries have largely been policy-driven, particularly under initiatives such as the *Revitalizing Northeast Old Industrial Base Strategy*, representing a form of leapfrogging development that runs counter to the region's comparative advantage. As Table 4 illustrates, the northeast's MDL is significantly lower than that of other regions, suggesting that labor-intensive industries hold a stronger comparative advantage and may yield greater efficiency gains. The stark contrast in potential factor flows between the eastern and northeastern regions highlights the misallocation of resources, offering a clear perspective on the future trajectory of industrial transfer and upgrading.

In the central region, both labor and capital appear to be scarce, aligning with the widely discussed "*Central Sunken*" hypothesis—the idea that China's development strategies have prioritized other regions while largely neglecting the middle. This region requires a significant influx of labor. Although the potential capital outflow remains negative, it is closer to zero compared to the west and northeast. A more concerning trend emerges in 2007, when potential capital flow turns positive, indicating an emerging capital shortage. These patterns suggest that the issue of "*Central Sunken*" was not resolved during the sample period; instead, it deteriorated. This calls for greater policy attention to ensure that resource allocation follows principles of comparative advantage. Only by doing so can the central region attract more workers without suffering capital depletion.

Perhaps unexpectedly, the western region exhibits a labor "surplus." A plausible explanation is that inadequate infrastructure has limited the development of manufacturing facilities capable of absorbing these workers. However, data also reveal that the surplus labor pool in the west is shrinking, suggesting an improving capacity to integrate these workers into productive employment. This shift is likely linked to the *Great Western Development Strategy* initiated in 2000, which has contributed to enhanced infrastructure and industrial expansion in the region.

Table 4 Markdown and Potential Labor Flows in Different Areas[8]

Year	East		Middle		West		Northeast	
	MDL	PLF	MDL	PLF	MDL	PLF	MDL	PLF
2000	0.39	-95.5	0.36	691.4	0.38	-1338.7	0.39	742.8
2001	0.34	-3.6	0.31	-23.8	0.33	-1185.3	0.30	1212.7
2002	0.3	-1311.2	0.28	1474.6	0.31	-1352	0.26	1188.5
2003	0.24	-989.4	0.22	1287.8	0.29	-1600.1	0.21	1301.6
2004	0.18	-1473.5	0.17	1270.6	0.21	-986.9	0.18	1189.8
2005	0.18	-1360.7	0.17	665.0	0.21	-1014.5	0.16	1710.1
2006	0.19	-1338.9	0.17	599.7	0.21	-1066.1	0.15	1805.4
2007	0.19	-1469.6	0.15	1392.9	0.18	-900.5	0.16	977.2
Year	**East**		**Middle**		**West**		**Northeast**	
	MDK	PKF	MDK	PKF	MDK	PKF	MDK	PKF
2000	0.65	22.60	0.78	-3.84	0.82	-6.6	1.07	-12.10
2001	0.56	21.30	0.67	-4.43	0.66	-4.51	0.90	-12.30
2002	0.50	21.80	0.58	-3.78	0.57	-3.80	0.82	-14.20
2003	0.38	27.90	0.43	-2.70	0.49	-10	0.64	-15.20
2004	0.25	40.00	0.32	-8.51	0.38	-16.2	0.44	-15.30
2005	0.25	31.90	0.29	-4.18	0.33	-12.9	0.41	-14.80
2006	0.21	33.60	0.26	-7.14	0.26	-10.6	0.34	-15.80
2007	0.20	20.10	0.21	0.97	0.24	-8.92	0.27	-12.10

Note: PLF means potential labor flows, in thousand, PLK means potential capital flows, in billion Yuan. The negative sign indicates outflow, while the positive one means inflow.

According to the classical international *Flying Geese* theory (Akamatsu, 1962; Kojima, 2000), industries relocate across countries once a nation loses its comparative advantage. Historically, we have observed this pattern with the global manufacturing center shifting from the United States to Japan, then to the Four Asian Tigers (Hong Kong, Singapore, South Korea, and Taiwan), and eventually to China. Given that

8 East area includes 10 provinces: Beijing, Tianjing, Hebei, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong, Hainan. The middle region contains 6 provinces: Shanxi, Anhui, Jiangxi, Henan, Hubei, Hunan. There are 11 provinces in the west area: Inner Mongolia, Guangxi, Chongqing, Sichuan, Guizhou, Yunnan , Shaanxi, Gansu, Qinghai, Ningxia, Xinjiang. Northeast zones include: Heilongjiang, Jilin, Liaoning . Tibet is not included in our sample since there are so many problems with the data of this province. South-China Coast includes Guandong, Fujian and Hainan. East-China Coast includes Shanghai, Jiangsu and Zhejiang. North-China Coast includes Beijing, Tianjing, Shandong and Hebei.

China’s labor costs have been rising after 2004, it is often argued that labor-intensive industries should now migrate to other developing economies in Southeast Asia, such as India, Vietnam, and Malaysia. However, our analysis suggests that China still possesses substantial untapped structural advantages, which can be realized by reallocating resources across regions and industries in line with comparative advantage. A particularly significant opportunity lies in redistributing labor from the eastern region to the middle and northeastern regions while simultaneously facilitating industrial upgrading in the east. This internal restructuring mirrors the *Flying Geese* model but on a domestic scale, illustrating China’s potential for continued industrial transformation and efficiency gains (Cai et al., 2009).

6. Concluding remarks

Foreign critics of China's weak labor rights fail to explain the coexistence of weak labor rights and rapid economic growth. Clearly, they overlook the fact that when weak labor rights intersect with the country's opening-up policy, they serve as a comparative advantage in fueling economic expansion. Low labor costs attract substantial domestic and foreign investment, while rising international demand compensates for insufficient domestic consumption caused by low wages. In addition, they overlook the significant progress made in strengthening labor rights, particularly since 2004. This study highlights the paradoxical role of markdown in fostering productivity growth while simultaneously exacerbating welfare losses. By integrating markdown into a comprehensive productivity decomposition framework, we distinguish between the contributions of reallocation, technical change, and markdown.

The empirical evidence from China’s manufacturing sector demonstrates that weak rights have played a significant role in boosting TFP growth following China’s accession to the WTO. However, the persistence of such serious markdowns suggests an urgent need for policy reforms. The findings underscore the necessity of transitioning towards a labor market that balances productivity gains with fair compensation for workers. Addressing these distortions through policy measures—such as strengthening labor protections, improving wage bargaining mechanisms, and enhancing social security—can lead to more equitable and sustainable economic growth. Moreover, the study’s insights have broader implications for developing economies that rely on low-wage labor as a comparative advantage.

In addition, our findings suggest that previous studies may have overestimated the reallocation effect by failing to distinguish it from the markdown effect. Overlooking this distinction risks diverting policymakers' attention from essential reforms to strengthen labor rights, even though reallocation-based strategies also play a crucial role in addressing fundamental inefficiencies in the economy. The potential for a domestic *Flying Geese* pattern highlights structural inefficiencies in China's historical development trajectory, underscoring the urgent need for reforms to optimize resource allocation and enhance economic efficiency. A balanced approach that integrates both aspects is necessary for sustainable economic improvement.

Due to the absence of the intermediate input variable in the ASIF database after 2008, estimating factor elasticity becomes challenging; therefore, our analysis is limited to the year 2007. Future research could expand this framework by incorporating more recent data or extending the analysis to global datasets for cross-country comparisons, providing deeper insights into international patterns of weak rights and resource allocation.

References

Ackerberg, D., Caves, K., & Frazer, G. (2006). Structural identification of production function. *MPRA Paper*, 38349. University Library of Munich, Germany.

Akamatsu, K. (1962). A historical pattern of economic growth in developing countries. *The Development Economies*, *Preliminary Issue*, (1), 3–25.

Alviarez, V. I., Fioretti, M., Kikkawa, K., & Morlacco, M. (2023). *Two-sided market power in firm-to-firm trade* (No. w31253). National Bureau of Economic Research.

Baily, M. N., Hulten, C., & Campbell, D. (1992). Productivity dynamics in manufacturing plants. *Brookings Papers on Economic Activity (Microeconomics)*, 187–249.

Banerjee, A., & Duflo, E. (2005). Growth theory through the lens of development economics. In P. Aghion & S. Durlauf (Eds.), *Handbook of Economic Growth* (Vol. 1, Part A, pp. 473–552). North-Holland.

Basu, K. (1997). *Analytical development economics: The less developed economy reconsidered.* MIT Press.

Basu, S., & Fernald, J. G. (2002). Aggregate productivity and aggregate technology. *European Economic Review*, *46*, 963–991.

Basu, S., Pascali, L., Schiantarelli, F., & Servén, L. (2009). Productivity, welfare, and reallocation: Theory and firm-level evidence. *NBER Working Paper No. 15579*.

Basu, S., Pascali, L., Schiantarelli, F., & Servén, L. (2022). Productivity and the welfare of nations. *Journal of the European Economic Association, 20*(4), 1647–1682.

Bosworth, B., & Collins, S. (2008). Accounting for growth: Comparing China and India. *Journal of Economic Perspectives, 22*(1), 45–66.

Brandt, L., Van Biesebroeck, J., & Zhang, Y. (2012). Creative accounting or creative destruction? Firm-level productivity growth in Chinese manufacturing. *Journal of Development Economics, 97*(2), 339–351.

Cai, F. (2007). The Lewis turning point of China's economic development. *Reports on China's Population and Labor* (No. 8). Social Sciences Academic Press. (In Chinese).

Cai, F. (2010). Shortage of migrant labor: Causes and policy implications. *China Opening Herald, 149*(2), 5–10. (In Chinese).

Cai, F., Wang, D., & Qu, Y. (2009). Flying geese within borders: How China sustains its labor-intensive industries. *Economic Research Journal, 9*, 4–14. (In Chinese).

Cheng, W., Liu, S., Meng, B., & Gao, Y. (2025). Impact of countercyclical fiscal policy on total factor productivity in state-owned enterprises: A corporate governance perspective. *Economic Analysis and Policy*. (Accepted).

Cheng, W., Meng, B., Gao, Y., & Dollar, D. (2024). The paradox of decelerated technology importation and accelerated innovation in China: Insights from national technology development zones. *China Economic Review*, *88*, 102303.

Domar, E. D. (1961). On the measurement of technological change. *Economic Journal, 71*, 709–729.

Fabricant, S. (1942). *Employment in manufacturing, 1899–1939.* NBER.

Foster, L., Haltiwanger, J., & Krizan, C. J. (2001). Aggregate productivity growth: Lessons from microeconomic evidence. *NBER Working Paper No. 10129.*

Griliches, Z., & Regev, H. (1995). Firm productivity in Israeli industry, 1979–1988. *Journal of Econometrics, 65*, 175–203.

Hall, R. E., & Jorgenson, D. W. (1967). Tax policy and investment behavior. *American Economic Review, 57*, 391–414.

Hall, R. E., & Jorgenson, D. W. (1969). Tax policy and investment behavior: Reply and further results. *American Economic Review, 59*, 388–401.

Hall, R. E., & Jorgenson, D. W. (1971). Application of the theory of optimum capital accumulation. In G. Fromm (Ed.), *Tax incentives and capital spending*. North-Holland.

Hsieh, C. T., & Klenow, P. J. (2009). Misallocation and manufacturing TFP in China and India. *The Quarterly Journal of Economics, 124*(4), 1403–1448.

Hulten, C. (1978). Growth accounting with intermediate inputs. *Review of Economic Studies, 45*, 511–518.

Jorgenson, D. W. (1963). Capital theory and investment behavior. *American Economic*

Review, 91, 1–32.

King, R. G., & Watson, M. W. (2012). Inflation and unit labor cost. *Journal of Money, Credit and Banking, 44*, 111–149.

Kojima, K. (2000). The *Flying Geese* model of Asian economic development: Origin, theoretical extensions, and regional policy implications. *Journal of Asian Economics, 11*, 375–401.

Krugman, P. (1994). *The age of diminished expectations.* MIT Press.

Levinsohn, J., & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies, 70*(2), 317–341.

Lin, J. (2011). New structural economics: A framework for rethinking development. *The World Bank Research Observer, 26*(2), 193–221.

Lu, Y., Sugita, Y., & Zhu, L. (2019). *Wage and markdowns and FDI liberalization* (No. HIAS-E-83). Hitotsubashi Institute for Advanced Study, Hitotsubashi University.

Kleinman, B., Liu, E., & Redding, S. J. (2024). International friends and enemies. *American Economic Journal: Macroeconomics, 16*(4), 350–385.

Magee, S. P. (1971). Factor market distortions, production, distribution, and the pure theory of international trade. *The Quarterly Journal of Economics, 85*(4), 623–643.

McMillan, M., & Rodrik, D. (2011). Globalization, structural change, and productivity growth. *NBER Working Paper No. 17143.*

Melitz, M. J., & Polanec, S. (2012). Dynamic Olley-Pakes productivity decomposition with entry and exit. *NBER Working Paper No. 18182.*

OECD. (2009). *Measuring capital* (2nd ed.). OECD Publications.

Olley, S., & Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica, 64*(6), 1263–1297.

Petrin, A., & Levinsohn, J. (2012). Measuring aggregate productivity growth using plant-level data. *RAND Journal of Economics, 43*(4), 705–725.

Petrin, A., & Sivadasan, J. (2011). Estimating lost output from allocative inefficiency, with an application to Chile and firing costs. *NBER Working Paper No. 17373.*

Qin, H. (2015). Dilemmas of twenty-first century globalization: Explanations and solutions, with a critique of Thomas Piketty's *Capital in the Twenty-First Century. Leaders, 63.*

Qu, Y., Cai, F., & Zhang, X. (2013). Has the *Flying Geese* paradigm occurred in China? *China Economist, 8*(6).

Rebitzer, J. B. (1988). Unemployment, labor relations, and unit labor costs. *The American Economic Review, 78*(2), 389–394.

Ren, R. E. (1996). A preliminary study on the international competitiveness of Chinese manufacturing. *China Soft Science, 9*, 74–82. (In Chinese).

Rubens, M. (2023). Market structure, oligopsony power, and productivity. *American Economic Review*, *113*(9), 2382-2410.
Sbordone, A. M. (2002). Prices and unit labor costs: A new test of price stickiness. *Journal of Monetary Economics, 49*(2), 265–292.
Schumpeter, J. A. (1982). *The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle.* Transaction Publishers.
Solow, R. (1957). Technical change and the aggregate production function. *Review of Economics and Statistics, 39*(3), 312–320.
Wang, Y. (2013). New viewpoint to new structural economics. *Journal of Translation from Foreign Literatures of Economics, 2*, 99–106. (In Chinese).
Weitzman, M. L. (1984). *The share economy: Conquering stagflation.* Harvard University Press.
Wooldridge, J. M. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters, 104*(3), 112–114.
Wu, H. X., & Yu, C. (2022). The impact of the digital economy on China's economic growth and productivity performance. *China Economic Journal, 15*(2), 153–170.
Xie, E., Xu, M., & Yu, M. (2024). Trade liberalization, labor market power, and misallocation across firms: Evidence from China's WTO accession. *Journal of Development Economics*, *171*, 103353.
Yeh, C., Macaluso, C., & Hershbein, B. (2022). Monopsony in the US labor market. *American Economic Review,* 112(7), 2099-2138.
Young, A. (2003). Gold into base metals: Productivity growth in the People's Republic of China during the reform period. *Journal of Political Economy, 111*(6), 1220–1261.
Zhang, S., Luo, J., Huang, D. H., & Xu, J. (2023). Market distortion, factor misallocation, and efficiency loss in manufacturing enterprises. *Journal of Business Research, 154*, 113290.
Zhang, X., Yang, J., & Wang, S. (2011). China has reached the Lewis turning point. *China Economic Review, 22*(4), 542–554.

Appendix

I. Variables and Method

(1) Construction of Variables

We compute nominal value added by subtracting intermediate input from gross output and then adding to VAT payable. The real value added is calculated by separately deflating output, input and VAT (value added tax). The gross output and VAT payable are deflated by 2-digit PPI (Producer Price Index), while the intermediate input deflator is PPIRM (Purchasing Price Index of Raw Material, Fuel and Power) classified by province.[9] The two deflators are obtained from *2008 China Urban Life and Price Yearbook*. Since the variable of gross output is absent in 2004, we replace it with the sum of sales and net inventory growth.

We are able to directly obtain labor input and nominal capital stock with the variable employee and net value of fixed asset in our dataset, respectively. The calculation of real capital stock, however, should resort to perpetual inventory method. The deflator of labor payment and investment are Consumer Price Index (CPI) and Fixed-asset Investment Price Index respectively. Both deflators are from *China Statistical Yearbook*. All deflators we use are transformed into fixed base index with 2000 as the base year.

In order to calculate markdown, we should first compute indices of factor prices and marginal productivity of factors, the ratio between these two indices amounts to markdown. Therefore, we need to estimate production function to obtain elasticity of factor so as to compute marginal productivity. We will estimate the Cobb-Douglas production function separately for each 4-digit industry year by year, and control province.[10]

$$lnY_{it} = \alpha lnK_{it} + \beta lnL_{it} + \varepsilon_{it}$$

Y represents real value added, K means real capital stock, L denotes labor, i refers to 4-digit industry, t is year 2000-2007.

Marginal productivity of labor and capital can therefore be expressed as *MPK* and *MPL,* respectively:

$$\mathrm{MPK}_{it} = \frac{\alpha_{it} Y_{it}}{K_{it}} \quad \text{(A1)}$$

$$MPL_{it} = \frac{\beta_{it} Y_{it}}{L_{it}} \quad \text{(A2)}$$

The calculation of labor price is quite easy since the data of payment for labor are

[9] Some works (e.g. Nie *et al*. (2011) replace gross output with sales, which neglect changes in inventories. Others, such as Young (2003) deflate value added directly instead of deflating gross output and intermediate input separately, which is likely to underestimate the real value added considering the faster increase of input price compared to that of output price with the progressing of market reform.

10 Technology choice is endogenous to endowment structure, which in turn goes hand in glove with industrial structure, neglecting which is likely to underestimate the role of capital accumulation to TFP, especially in developing countries. (Wang, 2013) Besides, the difference of factor elasticity among industries and regions would result in biases of reallocation were it not taken account of.

always available. When it comes to the capital price, however, it seems a little delicate. While it might be argued that we are able to obtain capital price from capital share, which in turn can be computed simply by deducting labor share from one.[11] However, this result holds only in neoclassical settings. Since we have taken account of imperfect competition and distortion, economic profits cannot be considered to be zero, and thus chances are high that the sum of labor share and capital share isn't equal to unity.

Petrin and Levinsohn (2012) replace APG with aggregate labor productivity growth since they are not able to obtain the data of capital rental price, which is likely to overestimate APG and reallocation. We solve this problem by using Hall-Jorgenson formula to calculate the rental price and capital cost share. By introducing a perfectly competitive capital renal market, Jorgenson (1963) derives a capital rental price calculation formula with a capital asset pricing model.[12] Then Hall and Jorgenson (1967, 1969, 1971) make some refinements to this formula. The typical formula can be written as:

$$z_t = r + \delta - \frac{(b_t - b_{t-1})}{b_{t-1}} \tag{A3}$$

Where r is rate of return, δ denotes depreciate rate, b means capital purchasing price, and $\frac{(b_t - b_{t-1})}{b_{t-1}}$ designates capital price increase rate. Capital price increase is deducted from the formula because it belongs to part of capital return but not part of rental price. Since there is no firm-level capital purchasing price, the price increase rate is calculated by deducting one from price index of fixed asset investment. δ equals the quotient of depreciation and net value of fixed assets. There are two ways to calculate r: internal rate of return and external one. The two rates of return might differ greatly due to the difference of data origins. Following the suggestion of OECD(2009, P140), we use external rate of return, which can be calculated with variables in our data.[13]

$$r = \frac{financialfee}{totaldebt - accountspayable} \tag{A4}$$

(2) Estimation Method: WLP

It is critical to obtain consistent and valid estimates for the parameters of the production function when analyzing productivity. Therefore, choosing appropriate method

[11] In fact, we can get capital income share by deducting labor share from one. But what we need here is capital cost share, which is generally less than capital income share considering the existence of profit.

[12] In fact, we have derived the same formulation with a dynamic general equilibrium model. For brevity's sake, we don't show this process.

[13] Since accounts payable is not interest bearing, it ought to be subtracted from total debt. For the lack of data about accounts payable during 2000-2003, we use the similar method as that used to calculate L3 to estimate the accounts payable in the year 2000-2003.

according to the feature of data matters a lot. The traditional OLS method fails to solve the selection problem induced by shutdown decision and the simultaneity biases generated by endogenous input demands. The fixed-effect estimates (FE) and instrument variable estimates (IV), to some extent, solve the simultaneity problem.[14] Olley and Pakes (1996) (hereafter OP) circumvents both biases by using investment as a proxy for productivity. However, the assumption of OP that investment is increasing in productivity seems inapt in our dataset. Because the state-owned enterprises, which accounts for a large proportion in our sample, seldom if ever make their investment decisions according to productivity but motivated by rent-seeking or policy. Moreover, the OP method requires investment to be nonnegative and non-missing, which may lead to 42% sample losses in our data. Levinsohn and Petrin (2003) (hereafter LP) replace investment with materials as the proxy and relieve part of the drawbacks of OP method.

In fact, both OP and LP methods assume that firms respond to productivity shocks by adjusting capital and intermediate input, and consider labor to be exogenous. However, in a country with low wages like China, chances are high that firms adjust labor inputs when productivity changes. Ackerberg *et al*.(2006) argue that labor is also a function of state variable and thus LP is likely to generates a collinearity problem, which may bias downwards the estimate of the labor coefficient, or labor may drop out of the estimation. To avoid this drawback in the LP, Wooldridge (2009) (hereafter WLP) provides an alternative way that simultaneously estimates the first and second-stage equations through GMM. Besides, it is easy for WLP to obtain fully robust standard errors by efficiently using the moment conditions implied by the OP and LP assumptions. In a word, our preferred method is WLP, which not only solves the common selection problem and simultaneity biases, but also addresses the collinearity problem.

II. Evolution of Productivity Decomposition Framework

(1) APG based on weighted sum of technological changes

We provide a brief review of APG decomposition approaches. The review is brief since these methods have been discussed in more detail in previous works. Our goal is to find out the evolution path of these alternatives. Let's first look at the decompositions based on weighted sum of producer-level technological changes. Baily, Hulten and Campbell (1992) propose their decomposition model (hereafter BHC) as follows:

[14] Since the requirements of FE and IV method are somewhat rigorous, chances are that they cannot solve the simultaneity biases appropriately. For example, FE requires that the unobservable variables are constant across time. A valid instrument which is correlated with observable inputs while exogenous to unobservable productivity is also not easy to find in our data.

$$\Delta tfp_t = \Delta \sum_i \omega_{it} z_{it} = \sum_{i \in S} \omega_{it-1} \Delta z_{it} + \sum_{i \in S} \Delta \omega_{it} z_{it} + \sum_{i \in N} \omega_{it} z_{it} - \sum_{i \in X} \omega_{it-1} z_{it-1} \tag{A5}$$

where Δ is the first difference operator from $t-\tau$ to t. S, E and X denotes the sets of stayers, entrants and exiters, respectively. Stayers denote the plants that are operating both in t and the prior year $t-1$. Plants existing in t but not in $t-1$ are designated the entrants. Plants existing in $t-1$ but no longer in t are defined as exiters. The first two terms are TE and RE of continuing firms, respectively. TE measures the variation of technology holding value added share constant, while RE measures the share changes of establishments with different technology levels. The last two terms illustrate the contribution of the entry of new producers and the exit of old ones to APG.

Griliches and Regev (1995) put forward an alternative framework (hereafter GR), which relieves part of the measurement concerns of BHC by introducing the cross-time means as weights.

$$\Delta tfp_t = \Delta \sum_i \omega_{it} z_{it} = \sum_i \omega^a_{it} \Delta z_{it} + \sum_i \Delta \omega_{it} z^a_{it} \tag{A6}$$

where the superscript "a" represents the average across time. The two terms can be explained in like manner as BHC.

Olley and Pakes (1996) make further improvement as to reallocation effect. They have taken account of the importance of cross-sectional mean in calculating RE. But what they analyze is productivity instead of productivity growth, or put it another way, they neglect the dynamics of productivity.

$$tfp_t = \sum_i \omega_{it} z_{it} = \sum_i \bar{\omega}_{it} \bar{z}_{it} + \sum_i \left(\omega_{it} - \bar{\omega}_t\right)\left(z_{it} - \bar{z}_t\right) \tag{A7}$$

where a bar over a variable indicates the unweighted mean (across all plants). The first term is the technical effect and the second is reallocation effect, which means that reallocation of more share to more productive firms enhances industry productivity.

Foster, Haltiwanger and Krizan (2001) propose an extension of BHC and GR approaches by introducing industry mean as a benchmark. What should be mentioned here is that the assumption of closed economy is a prerequisite for the introduction of industry mean.

Their modification to BHC (hereafter BHC-FHK) can be written as:

$$\Delta tfp_t = \Delta \sum_i \omega_{it} z_{it} = \sum_{i \in S} \omega_{it-1} \Delta z_{it} + \sum_{i \in S} \left(z_{it-1} - \bar{z}_{t-1}\right) \Delta \omega_{it} + \sum_{i \in S} \Delta z_{it} \Delta \omega_{it} + \sum_{i \in N} \omega_{it} \left(z_{it} - \bar{z}_{t-1}\right) - \sum_{i \in X} \omega_{it-1} \left(z_{it-1} - \bar{z}_{t-1}\right) \tag{A8}$$

The first term is the same as BHC. The second term is the pure reallocation excluding cross term since it adopts z_{it-1} instead of z_{it} as the weight of reallocation between survivors. The last two terms denotes the contribution of net entry to APG, which are also part of reallocation effect.

Their modified version of GR (hereafter GR-FHK) is given by

$$\Delta tfp_t = \Delta\sum_i \omega_{it} z_{it} = \sum_{i\in S} \omega_{it}^a \Delta z_{it} + \sum_{i\in S}\left(z_{it}^a - \overline{z}_t^a\right)\Delta\omega_{it} + \sum_{i\in N}\omega_{it}\left(z_{it} - \overline{z}_t^a\right) - \sum_{i\in X}\omega_{it-1}\left(z_{it-1} - \overline{z}_t^a\right) \tag{A9}$$

The interpretation of each term is identical to that of BHC-FHK except that there is no cross term, which is the result of using cross-time mean as the weight.

Melitz and Polanec (2012) make some refinements to the decomposition in Olley and Pakes (1996) in order to accommodate dynamics and net entry effects. We designate this framework as OP-MP:

$$\Delta tfp_t = \Delta\sum_i \omega_{it} z_{it} = \Delta\sum_{i\in S}\overline{\omega}_{it}\overline{z}_{it} + \Delta\sum_{i\in S}\left(\omega_{it} - \overline{\omega}_t\right)\left(z_{it} - \overline{z}_t\right) + \sum_{i\in N}\omega_{it}\left(\sum_{i\in N}\frac{\omega_{it}}{\sum_{i\in N}\omega_{it}} z_{it} - \sum_{i\in S}\frac{\omega_{it}}{\sum_{i\in S}\omega_{it}} z_{it}\right) - \sum_{i\in X}\omega_{it-1}\left(\sum_{i\in X}\frac{\omega_{it-1}}{\sum_{i\in X}\omega_{it-1}} z_{it-1} - \sum_{i\in S}\frac{\omega_{it-1}}{\sum_{i\in S}\omega_{it-1}} z_{it-1}\right) \tag{A10}$$

Unlike the preceding two methods, this framework provides different reference levels for the reallocation effects of existing, entering, and exiting establishments, respectively.

From the above literatures, we can come to the conclusion that it would be more precise and logical to introduce a reference level into the reallocation effect. Besides, by exploiting across-time means instead of levels of current or last period, measurement errors are more likely to be reduced.

(2) APG based on aggregate final demand

Then we come to the decompositions using productivity related with final demand. We introduce first the model proposed by Petrin and Levinsohn (2012)—hereafter PL.

$$\Delta tfp_t = \sum_i \omega_{it}\Delta z_{it} + \sum_{J=L、K}\sum_i \omega_{it}\left(\varepsilon_{Jit} - s_{Jit}\right)\Delta j_{it} + F_{it} \tag{A11}$$

where s_{Ji} is the share of input costs in value added, ε_{Ji} is the output elasticity of inputs. The first term is TE and the second RE. The last term denotes distortion. For transparency, we combine the reallocation terms of two inputs into one term.

The most notable advantage of this decomposition, compared with the above ones, is the successful separation of the reallocation effect of each factor. However, from the form of the reallocation term, we cannot tell how to introduce reference level to the effect. Besides, the independent distortion term seems to be a little abrupt since it is related with neither technology nor factors.

To our knowledge, only Basu and Fernald(2002) and Basu *et al*. (2009) have mentioned the aggregate distortion. We designate this method as BF. The formulation in their original papers seems too long and complicated. For simplicity's sake, we make two changes, which won't affect the nature of BF method. The first transformation is about the implication of cost minimization obtained from Hall (1990). We assume that the following equation holds:

$$dq_i = \mu_i \left[s_{Li} dl_i + s_{Ki} dk_i \right] + dz_i \tag{A12}$$

where μ_i is the markup of output i (i.e. P_i / MC_i).

Besides, since the data we use are disaggregated, we calculate APG simply by aggregating productivity growth of each firm, just as Petrin and Levinsohn (2012) do.[15] With some mathematical manipulation, we can write the formula of BF as:

$$\Delta tfp_t = \sum_i w_{it} \Delta z_{it} + \sum_i w_{it} \left(\mu_{it} - \bar{\mu}_{it} \right) \left(s_{Lit} \Delta l_{it} + s_{Kit} \Delta k_{it} \right) + \sum_i w_{it} \left(\bar{\mu}_{it} - 1 \right) \left(s_{Lit} \Delta l_{it} + s_{Kit} \Delta k_{it} \right) \tag{A13}$$

The first term is technical effect. The second and third terms are reallocation effect and aggregate distortion, respectively, from which we can tell apart the effect of labor and capital. Unlike BHC-FHK and GR-FHK, the hypothesis of closed economy is not necessary in this model.

This framework successfully separates out the weak rightsfrom reallocation effect, which, I think, is a great contribution to the productivity decomposition literatures. Nevertheless, they attribute the reallocation of labor and capital to the same driver, namely the difference in output markup.[16] This means both factors are reallocated in the same way, which goes against realities. This drawback of BF comes from the assumption obtained from Hall (1990), and we will explain it briefly. By differentiating the Cobb-Douglas value added production function we get:

$$dq_i = \varepsilon_{Li} dl_i + \varepsilon_{Ki} dk_i + dz_i \tag{A14}$$

Since output elasticity is the quotient of input cost share and its unit cost. The above equation can be further transformed into:

$$dq_i = \frac{s_{Li}}{u_{Li}} dl_i + \frac{s_{Ki}}{u_{Ki}} dk_i + dz_i \tag{A15}$$

15 In fact, there exists the problem of aggregation bias, that is, the result computed with aggregated data differs from that with disaggregated data. Interested readers may find more details from Basu&Fernald(2002).

16 By calculating with aggregated data, Basu and Fernald obtained additional reallocation terms, which are determined by factor prices only. However, as we have analyzed before, it is the joint effect of factor prices and marginal product of inputs that determines reallocation. Therefore, these additional reallocation effects are not precise enough either.

Comparing it with equation (24), we find that the implied hypotheses in BF method are that the unit cost of labor and capital equals with each other, and both of them are equated to output markup. Undoubtedly, the first assumption counters reality. The second hypothesis holds when only when there is only one input. In fact, output price-marginal cost markup is only an indirect driver of input reallocation. It is the unit cost of factors that result in reallocation of inputs directly. In a word, BF needs further improvements. Besides, if we replace equation (24) with equation (26), then the BF method will be identical to PL.

(3) Economic Intuition of Our Framework

Let's first provide the economic intuition of reallocation effect (RE). If the unit cost of a factor in one firm is less than the industry average values, i.e. $\left(\overline{u}_i - u_i\right) > 0$, the mobility of resources to this firm ($dl > 0$ or $dk > 0$) would mean exploiting the unit cost advantages of the plant, and thus make the reallocation effect positive ($RE > 0$).The situation of $\left(\overline{u}_i - u_i\right) < 0$ can be analyzed in the same way. But when it comes to $\left(\overline{u}_i - u_i\right) = 0$, there would be no reallocation, no matter whether the unit cost is less than one or not. Therefore, only the deviations of unit cost of inputs from the industry means can be attributed to reallocation effect, while the gap between unit cost and one cannot be attributed to reallocation effect.

In fact, the gap between unit cost and one is weak rights(WR), whose economic intuition is also evident: only if there are distortions in the input or output market which lead to the gap between $\overline{u}_{Ji}$ and 1, an additional input use increases APG by raising the aggregate final demand, granted that there is no resource reallocation or technological improvement.

Petrin and Levinsohn (2012) introduce distortion by deducting them directly from output, and derive an independent distortion term, which makes it hard to conduct quantitative analysis on distortions. In fact, distortions like market power, sunk costs and adjustment costs are all reflected in the gap between $\overline{u}_{Ji}$ and 1, and thus can be derived directly from the formula without assuming an unique distortion term. A simple illustration is enough to shed light on this argument. Assume an economy without any distortion at the beginning, and thus u =1. Now comes a shock of output tax τ, which reduces marginal output of input, for example L, from MPL to $\frac{1}{1+\tau}MPL$, so firms would choose the inputs and outputs where $u = \frac{w}{MPL} < 1$, which can be captured by the WR term. The other types of frictions and distortions can be analyzed in similar ways.

Another widely used decomposition method proposed by Fabricant (1942) even derives reallocation effect by simply deducting within-sector productivity changes from productivity growth. Both of the two studies are not able to capture the weak rightsproperly. Our framework makes it by introducing unit cost of inputs and then separating out WR from RE.

Generally speaking, following the evolution path of these brilliant works, we construct our framework in a more precise, logical and practical way. First and foremost, we introduce unit cost of inputs into our decomposition, which correctly discloses the driving force of reallocation and origin of aggregate distortion, and help distinguish between reallocation and weak rightsby adding reference levels. Secondly, the computation of our APG considers the aggregate final demand, which relates more closely with welfare. The third one is that weights are calculated in the form of cross-time mean, which relieves measurement issues. Finally, our model can be applied to opened economy and we also include the contribution of net entry to APG.