# The Aggregate Productivity Effect of Labour and Capital Market Distortions in Canada


Yu Chen
University of Calgary

Trevor Tombe*
University of Calgary


This Version: December 2024


**Abstract**

How efficiently are workers, investment capital, and production distributed across firms in Canada? And how have they varied over time, across regions, and between sectors? To answer these questions, we present novel measures of the degree of resource misallocation over time and space using uniquely detailed firm-level data (T2-LEAP) between 2001 and 2015. We find the dispersion of marginal returns to both capital and labour across firms have increased significantly during this period, suggesting allocative efficiency is deteriorating. Using a rich but tractable multi-sector model of heterogeneous firms, we find that had the misallocation of labour and capital not worsened over our period of analysis, aggregate Canadian productivity would be significantly higher. Worsening allocations between sectors alone accounts for approximately half of the widening productivity gap between Canada and the United States over this period.



*JEL Classification: D24, E24, O4, O51*

*Keywords*: Misallocation; aggregate productivity; labour and capital market distortions



*Tombe (corresponding author): Professor, Department of Economics, University of Calgary, 2500 University Drive NW, Calgary, Alberta, T2N1N4. Email: `ttombe@ucalgary.ca`. Phone: 1-403-220-8068. Chen: Assistant Professor, Department of Economics, University of Calgary, 2500 University Drive NW, Calgary, Alberta, T2N1N4. Email: `yusonja.chen@ucalgary.ca`. Phone: 1-403-220-7357.


We gratefully acknowledge the financial support of the Social Science and Humanities Research Council (IG 435-2019-0100 and 435-2023-0255). We also benefited from the generosity of staff at Statistics Canada's Prairie Regional Research Data Centre at the University of Calgary and the exceptional research assistance from Chi-Man Yip and Abraham Saucedo Cepeda.

# 1 Introduction

The Canadian economy experienced several macroeconomic shocks since 2000, including the financial crisis, energy booms and busts, auto sector disruptions, rising imports from China, and more. The impact of these shocks, however, varies across sectors and regions, and therefore causes migration of labour and movement of capital across those sectors and regions. Barriers like migration costs, interprovincial trade costs and financial market frictions, however, may both distort the allocation of resources and disrupt their efficient adjustment to shocks. In this paper, we quantify the magnitude and consequences of potential resource misallocation across firms in Canada between 2001 and 2015 using a detailed data of the (near) universe of Canadian firms in Statistics Canada's T2-LEAP. We explore specifically how the dispersion in factor returns varies over time, across regions, and across sectors. And with the aid of a rich multi-sector model of Canada's economy featuring a continuum of heterogeneous firms and factor market frictions, we quantify the overall effect of allocative inefficiencies on Canada's aggregate productivity.

Before providing details, let us provide some intuition. In a well-functioning market, the allocation of workers, inputs and investment, are determined by prices. Productive firms or sectors that can pay higher wages, for example, will naturally hire more workers and expand in size relative to other firms or sectors. The resulting equilibrium allocation will tend to maximize productivity and incomes. What does such an allocation look like? Consider a world with two sectors. Workers are free to move between them, but each additional worker will add less to the firm's value. That is, there's diminishing returns to labour and other inputs. If labour has a different value, at the margin, across firms, then there is scope to move workers to higher value uses to increase overall output. Thus, the optimal allocation of labour equalizes the marginal value of each worker across all firms and sectors. Measured deviations from this may therefore be evidence of misallocation. Our analysis starts by measuring the distance between the Canadian economy and such an efficient benchmark.

Overall, we find worsening allocative efficiency in Canada. From 2001 to 2015, the variances of both capital and labour returns between firms have increased substantially. Much of the increase is driven by the more dispersed firm productivity in Western provinces and Ontario and it is widely spread across most significant industries in the Canadian economy. We demonstrate through a simple model that rising dispersion in returns of this kind is a cause of lower aggregate productivity. To quantify the effect more precisely, we develop a model of heterogeneous firms based on Hsieh and Klenow (2009) that can be mapped to key moments in the data. We estimate that a 23 percent reduction in Canada's aggregate TFP can be attributed to the increased labour and capital misallocation across firms. Our estimates of worsen distortions within sectors, however, are highly volatile over time, but our estimates of between-sector distortions are much more precise. We estimate worsening misallocation between sectors alone accounts for a 9 percent reduction in aggregate TFP over this period. This is large. It accounts approximately half of the widening productivity gap between Canada and the United States.

Our estimates are derived from Statistics Canada's administrative firm-level dataset T2-LEAP and covers the years 2001 through to 2015. This dataset contains longitudinal information on every incorporated Canadian firm hiring employees, including firms' revenue, capital stock, payroll, industry and province. With minimal assumptions that we detail in the next section, we measure the dispersion of capital and labour returns across firms by computing the variance of (log) marginal revenue products of labour and capital. We find substantial worsening of misallocation in Canada during 2001-2015, with a 20 percent rise in the dispersion of capital returns and a 15 percent increase in the labour market. The measured distortion in the capital market first declined in the early 2000s and started to increase shortly before the Great Recession and accelerated after 2011. In terms of labour misallocation, we observe an overall worsening trend although there is variation from this trend in particular years.

There are important differences between regions and sectors. Specifically, we find that Western provinces drive most of the increased distortion in the labour market. Ontario also contributes with a notable increase in measured capital misallocation after the financial crisis. We find Quebec and the Atlantic provinces see more moderate change in the dispersion of returns between firms in those regions. Different sectors also show distinct patterns in the changes in the measured between-firm misallocation. The energy sector, some manufacturing sectors (including the auto industry), as well as finance, insurance and real estate all experienced significant increases in the capital returns dispersion across firms. Not all large sectors experienced worsening allocative efficiency, however. Construction, wholesale and retail trade, for example, see little measured change in the returns to capital or labour between firms in those sectors. These within-sector effects are particularly important since it is only with our detailed administrative data that such measures may be computed. The sectoral results also lend themselves to additional quantitative analysis to estimate the aggregate implications of this rising misallocation.

To that end, we further estimate dispersion in labour and capital market wedges between firms within sectors. This requires a slightly more aggregated set of sectors, due to vetting requirements within the Statistics Canada Research Data Centre. In terms of the quantitative model, we construct a monopolistic competition model with heterogeneous firms based on Hsieh and Klenow (2009), where individual firms face wedges within both the labour and capital markets. Firm-level distortions to the allocation of labour and capital can cause a deviation of aggregate TFP from a hypothetically undistorted economy through both within-sector and between-sector effects. The within-sector distortion captures the distorted allocation of factors across firms within each sector, which reduces sectoral TFP. The between-sector distortion captures the distorted allocation of labour and capital across sectors, which can be summarized by the average firm-level distortions within each sector.

Counterfactual experiments in the model that hold measured misallocation fixed at its level in the initial years of our sample suggest that rising misallocation of labour and capital caused a significant reduction in Canada's average aggregate TFP between 2011 and 2015. And while labour market misallocation contributes most of the between-sector effect in levels, the falling efficiency in the capital market accounts fully for the deterioration of productivity caused by the between-sector distortion from 2001 to 2015. Our finding of significant productivity losses from worsening of misallocation is robust when excluding firm entry and exit. Our analysis also explores the potential for measurement error, increasing dispersion of firm-specific shocks, and other factors, to materially influence our results. We conclude that our estimates of worsening labour and capital allocations, between sectors in particular, are robust.

Our work builds on a large and growing literature on resource misallocation and productivity. In particular, Banerjee and Duflo (2005), Restuccia and Rogerson (2008), Hsieh and Klenow (2009), Brandt et al. (2013), Bartelsman et al. (2013), Restuccia (2019), and Tombe and Zhu (2019) all find the economic costs from misallocation of capital, labour, and/or output can be substantial. Our method closely follows this literature. To the best of our knowledge, though, this paper is the first comprehensive attempt to quantify the magnitude and consequences of resource misallocation across firms in Canada over time. Our study is also related to an empirical literature that identifies changes of misallocation within a country as an important source of changes in productivity over time.[1] Unlike these developing countries, the Canadian economy is a highly developed market economy, with little restriction on labour migration and a well-established financial market. Yet, we find resource misallocation could still play a critical role in the productivity and growth of the Canadian economy. Most recently and perhaps most notably related to our result is Bils et al. (2020), who find rising misallocation and worsening allocative efficiency in the United States over a roughly similar timeframe as ours. They attribute much, but not all, of this deterioration to measurement error in firm-level data. To the extent that the Canadian data may also suffer from this, our results should then be viewed as an upper-bound as measurement error lessens the magnitude of measured misallocation.

For policymakers, Canada's economy-wide productivity has been a growing concern. Based on the most recent data from the Penn World Table, the Canadian economy's productivity relative to the U.S. has fallen 11 percent from 2000 to 2015. Many have investigated possible factors behind it. For example, Leung et al. (2008) and Baldwin et al. (2014)) study how firm size distribution matter for the measured productivity gap, Tang (2017) and Almon and Tang (2011) focus on the impact of industrial structure, while Ranasinghe (2017) examines the importance of innovation spending on the productivity differences. Our study provides a new perspective by examining the changes in efficiency with which inputs are distributed across firms. We find that rising misallocation between sectors, especially in the capital market, contributes heavily to the lower aggregate productivity growth between 2001 and 2015. But further research is necessary to discover the specific sources of the labour and capital wedges we measure and to explore specific policy interventions to mitigate them or their effects.

Finally, our study is related to a group of papers that assesses changes in misallocation over the business cycle. For example, Oberfield (2013), Sandleris and Wright (2014), and Ziebarth (2013) examine how misallocation in an economy change during crises and recessions. While we do not

[1] For example, Calligaris (2015), Fujii and Nozawa (2013), Reis (2013), and Gopinath et al. (2017) study misallocation changes in countries experiencing slow productivity growth, while Chen and Irarrazabal (2015) examines misallocation change in Chile during a period of fast productivity growth.

measure the cyclical effects of misallocation on productivity explicitly, we discover some interesting changes pre- and post-2008 financial crisis. From 2001 to 2008, the within-sector distortion has worsened persistently. However, this trend stopped and revised after the financial crisis. Until 2015, half of the TFP loss caused by the increased within-sector misallocation during 2001-2008 was erased. On the other hand, the between-sector misallocation had only a slight net effect on the aggregate TFP from 2001 to 2008. Nevertheless, it deteriorates rapidly post-crisis, particularly after 2010. This suggests that how misallocation changes after a recession might be quite different at the firm level versus the industry level.

The rest of the paper is organised as follows. In Section 2, we first establish the connection between misallocation and marginal returns across firms using a simple model, then introduce the data and how we use it to measure these marginal returns. We also present the changes in the dispersion of labour and capital returns over time. In Section 3, we construct a model of heterogeneous firms that maps marginal product variation to aggregate productivity. In Section 4, based on numerous counterfactual experiments in the model, we quantify the changes of misallocation from 2001 to 2015 and their effects on aggregate productivity based on our model. We also examine the potential impact of measurement error in this section. In Section 5, we discuss the policy implications of our results. Section 6 concludes.

## 2 Misallocation in Canada

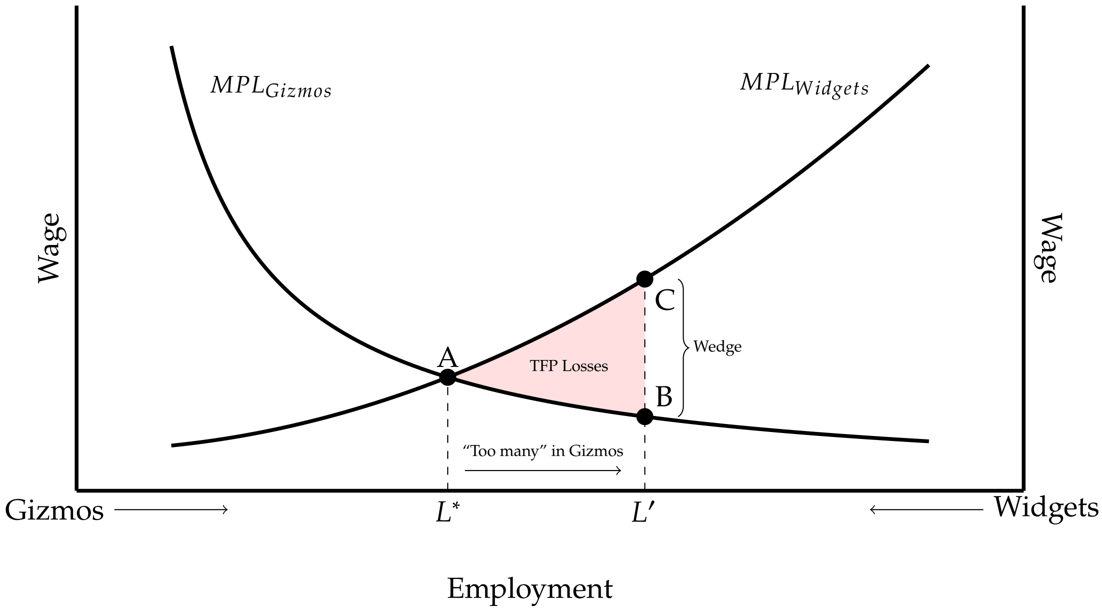
In this sector, we first demonstrate the connection between misallocation and marginal returns across firms using a simple model. We then describe the data and introduce our measure marginal returns across firms, sectors, and regions using firm-level information. Finally, we discuss general patterns we uncover before turning to a full quantitative model.

### 2.1 Misallocation: A Primer

Misallocation results when marginal returns are not equalised across firms. If an additional worker would increase the value of output at Firm A by two dollars but by only one dollar at Firm B, then aggregate output would increase by shifting workers from Firm B to Firm A. We illustrate the intuition behind this in Figure 1. As is clear, the allocative inefficiency is summarized by the area of the deadweight loss triangle shaded between the optimal allocation of labour $L^*$, the inefficient allocation of labour $L'$, and the firms' marginal products of labour. Importantly, which we will show more formally shortly, the magnitude of the inefficiency is related to the size of (1) the elasticity of labour demands and (2) the wedge between marginal products.

In more general terms, consider a simple model where labour is the only input. Firms differ in their productivity and there is diminishing returns to production. Specifically, let $y_i = \varphi_i l_i^\alpha$, where $\alpha \in (0, 1)$ captures the degree of diminishing returns. And, further, let output from all firms be perfectly substitutable for all others, and therefore aggregate output is $Y = \sum_i y_i$. To maximize this aggregate, the marginal product of labour across all firms will equalize and

Figure 1: Wedges Between Marginal Returns Lowers Productivity




Note: Illustrates the efficiency consequences of wedges between marginal returns across two firms. The black lines plot marginal revenue products of labour. If a wedge exists between returns such that labour is more productive making widgets than gizmos then the allocation of labour is inefficient. Equivalently, given diminishing marginal returns, there are "too many" workers producing gizmos. Total losses to the economy equal the shaded region.

thus the optimal allocation of labour across firms is $l_i^* \propto \varphi_i^{1/(1-\alpha)}$ and aggregate output will be $Y^* = \left(\sum_i \varphi_i^{1/(1-\alpha)}\right)^{1-\alpha}$. But if there are distortions to firms' hiring decisions, marginal products will not equalize. Denote the wedge between any given firm $i$'s marginal product and the average $\tau_i$. That is, $\alpha\varphi_i l_i^{\alpha-1}\tau_i = \alpha\varphi_j l_j^{\alpha-1}\tau_j$ for any $(i,j)$. In this case, the equilibrium allocation of labour across firms is

$$l_i \propto \left(\varphi_i \tau_i\right)^{1/(1-\alpha)}, \tag{1}$$

which is larger than $l_i^*$ if $\tau_i > 1$ and smaller if $\tau_i < 1$. Intuitively, if one firm receives a subsidy that others do not, then $\tau_i > 1$ and this firm will have an inefficiently large level of employment while other firms experience the reverse. Aggregating across firms yields total output

$$Y = \sum_i \varphi_i \left( \frac{\left(\varphi_i \tau_i\right)^{1/(1-\alpha)}}{\sum_j \left(\varphi_j \tau_j\right)^{1/(1-\alpha)}} \right)^{\alpha}. \tag{2}$$

Taking the ratio of this and the first-best output $Y^*$, one can show that

$$\hat{Y} \equiv \frac{Y}{Y^*} = \frac{\sum_i l_i^* \tau_i^{\alpha/(1-\alpha)}}{\left(\sum_j l_j^* \tau_j^{1/(1-\alpha)}\right)^{\alpha}} \leq 1, \tag{3}$$

where the last line follows from Jensen's inequality and $\alpha < 1$. The inequality will be strict unless all firms face the same distortion $\tau$. This illustrates the general intuition that any set of wedges that vary across firms will lower aggregate output.

We can approximate the aggregate loss from such wedges using Harbinger triangles to show that the efficiency loss due to differences in marginal products between firms is proportional to the variance of those differences. To see this, note that the size of the deadweight loss triangle in Figure 1, expressed as a share of total payroll, is $\epsilon t_i^2/2$ where $\epsilon$ is the elasticity of employment with respect to labour costs and $t_i = \tau_i - 1$ is the wedge to firm-$i$'s marginal product. Summing across all firms, the aggregate efficiency consequence of all wedges is $\sum_i l_i \epsilon t_i^2/2$ or, more intuitively,

$$DWL \approx \frac{\epsilon}{2} \times var\left(t_i\right), \tag{4}$$

since $\sum_i l_i t_i = 0$.[2] In the case of the simple model described above, we have $\epsilon = \alpha/(1-\alpha)$ and therefore if $\alpha = 2/3$, for example, then $\epsilon = 2$ and the aggregate efficiency loss is simply given by the variance of wedges alone. And regardless of the elasticity, changes in the variance of distortions will lead to equal proportional changes in the magnitude of the aggregate inefficiency. The variance in marginal products across firms will therefore be a central component of the quantitative analysis to come. And although this is an approximation of the aggregate inefficiency, in Section 3 we develop a richer model with multiple factors of production, multiple sectors, and a continuum of firms that, under certain conditions, results is an equivalent expression.

## 2.2 Detailed Data on Canadian Firms

We use a uniquely detailed dataset from Statistics Canada: T2-LEAP. This administrative micro-dataset links annual corporate income tax information (T2 forms) with the Longitudinal Employment Analysis Program (LEAP), which includes firm-specific payroll data. T2-LEAP provides longitudinal information on all statistical enterprises in Canada that employ staff, covering all sectors of the economy and firms of various sizes.[3] The dataset offers annual firm-level data on capital stock, revenue, employment, payroll, and industry affiliation, as well as information on firm entry/exit and provincial location. A more detailed description of the T2-LEAP is presented in the Appendix. Our analysis spans the period from 2001 to 2015.

We create two samples for our analysis. The first sample, referred to as the clean sample, applies the minimal data cleaning necessary for our analysis. Specifically, we exclude any firm-year observations with missing or non-positive values for payroll, revenue, and capital, where capital is calculated using total tangible assets, net of working capital. We also exclude firm-year observations with missing or unclassified NAICS codes. Finally, firms with missing province information and firms located in the territories (Yukon, Northwest Territories, and Nunavut) are also excluded from the clean sample. Our baseline analysis is conducted based on the clean sample

[2] The condition that $\sum_i l_i t_i = 0$ may be interepreted as an aggregate balanced-budget condition if wedges are explicit taxes or subsidies. Alternatively, for non-monetary distortions, it reflects that differences in marginal products $\tau_i$ are all expressed relative to the average and therefore $\sum_i l_i \tau_i = 1$.

[3] It excludes self-employed individuals or partnerships where participants do not draw salaries.

set. The number of firms included in the clean sample grew steadily over time, from 441,060 in 2001 to 548,480 in 2015. The second sample, referred to as the balanced sample, excludes firms that are not present in every sample year from 2001 to 2015, identified using the unique longitudinal identifier in T2-LEAP. This sample restriction excludes firms that entered after 2001 or exited before 2015. We use this sample in the analysis of firm entry and exit in Section 4.5. The sample size of the balanced sample consists of 113355 firms. Additional statistics, including the firm size distribution for both samples, are available in Appendix C.

### 2.3 Measuring Labour and Capital Market Distortions

The key variables of our study are the dispersion of the marginal revenue products of capital and labour across firms. Under the standard assumption of Cobb-Douglas production technologies, firms produce output according to $y_j(i) = \varphi_j(i)k_j(i)^{1-\alpha_j}l_j(i)^{\alpha_j}$, where $\alpha_j$ is the labour share of output in sector $j$ that might differ across industries but not across firms within an industry. It is straightforward to show that firm-$i$'s marginal revenue product equals the factor share times the average product. That is,

$$\begin{aligned} MRPK_j(i) &= (1-\alpha_j)\frac{p_j(i)y_j(i)}{k_j(i)}, \\ MRPL_j(i) &= \alpha_j\frac{p_j(i)y_j(i)}{L_j(i)}. \end{aligned} \tag{5}$$

And under the additional assumption that factor prices are common for all firms, the marginal revenue product of capital of a firm is proportional to the ratio of revenue over the value of its capital stock and the marginal revenue product of labour is proportional to the revenue-payroll ratio. This allows us to rearrange equation (5), to obtain

$$ln\left(\frac{p_j(i)y_j(i)}{Rk_j(i)}\right) = -ln(R) - ln(1-\alpha_j) + ln(MRPK_j(i)), \tag{6}$$

and,

$$ln\left(\frac{p_j(i)y_j(i)}{\omega l_j(i)}\right) = -ln(\omega) - ln(\alpha_j) + ln(MRPL_j(i)), \tag{7}$$

where $R$ and $\omega$ are the factor prices for capital and labour, respectively. For clarify, this approach infers marginal returns from observed average returns and therefore the assumption of a Cobb-Douglas production function is important.

These expressions allow us to measure the dispersion of marginal returns across firms using a regression method. Consider first the return to capital. The left-hand side of equation (6) is an individual firm's revenue over capital stock ratio, which can be calculated directly since both a firm's revenue and its capital stock can be obtained from our data. It can be decomposed into three components: $ln(\omega)$ that is constant for all firms, $ln(1-\alpha_j)$ that is sector-specific, and $ln(MRPL_j(i))$ which is the logarithm of the marginal revenue product of capital of an individual firm. Based on equation (6), we then run the following regression for each sample year

$$ln\left(\frac{p_j(i)y_j(i)}{Rk_j(i)}\right) = \beta_0 + \sum_{s=1}^{M}\beta_s\gamma_s + \epsilon_j(i), \tag{8}$$

using three-digit industrial codes as our sector dummies $\gamma_s$. The coefficient $\beta_s$ captures any sectoral specific factor that contributes to the variation of firms' revenue over capital, such as the different capital-labour input share across sectors. The constant term $\beta_0$ captures the common factors that are constant across all firms. According to equation (6), the residual from the regression is the log marginal revenue product of capital of each individual firm. We run this regression for each year and record the residuals of these regressions. We then compute the variance of these residuals for each year as our measure of the dispersion of marginal revenue product of capital. We can also compare the dispersion of capital returns across regions and across sectors using the variances of the residuals grouped by sector or by region. We can similarly compute the log marginal revenue product of labour with the following regression

$$ln\left(\frac{p_j(i)y_j(i)}{\omega l_j(i)}\right) = \beta_0 + \sum_{s=1}^{M}\beta_s\gamma_s + \epsilon_j(i), \tag{9}$$

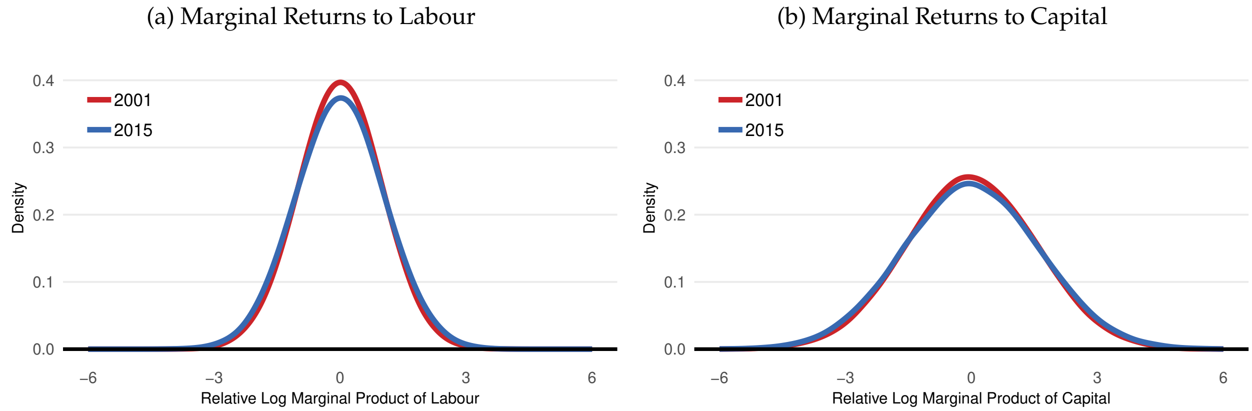
where $\omega l_j(i)$ is the total payroll of firms that can be observed directly from our data. Similar to the captial returns, we run the above regression for each year to obtain the residuals, and calculate the variances of the residuals as our measure of dispersion of labour returns across firms.

To fix ideas and provide intuition, we display the distribution of marginal returns to both labour and capital in Figure 2 based on a log-normal distribution approximation using the standard deviations as measured in our data with these regression methods. Both panels show a notable degree of dispersion in returns and both show increased dispersion across firms from 2001 to 2015. This suggests an overall rise of resource misallocation during this period, but we will show this more precisely shortly. Comparing panels (a) and (b), it is evident that the distribution of marginal returns to capital exhibits greater dispersion across firms than does the distribution of marginal returns to labour. The larger dispersion in labour and capital markets between 2001 and 2015, however, is more difficult to discern although both see increases. Before we quantify these changes and their potential effect on aggregate productivity in Section 4, we first investigate how the dispersion of marginal returns across firms has evolved during the period 2001 to 2015.

### 2.4 Changes in Marginal Product Dispersion Over Time

We begin by reporting the change in the dispersion of labour and capital returns over time across Canada's economy as a whole. Specifically, we estimate the variance in log labour returns $var(log(\tau_j(i)^l))$ and log capital returns $var(log(\tau_j(i)^k))$ and display the change of each in Figure 3 from 2001 to 2015. We find the variance of both factors' marginal returns increased substantially, especially after 2006. The variance of log capital returns increases by over 20 percent

Figure 2: Dispersion of Labour and Capital Returns Across Firms



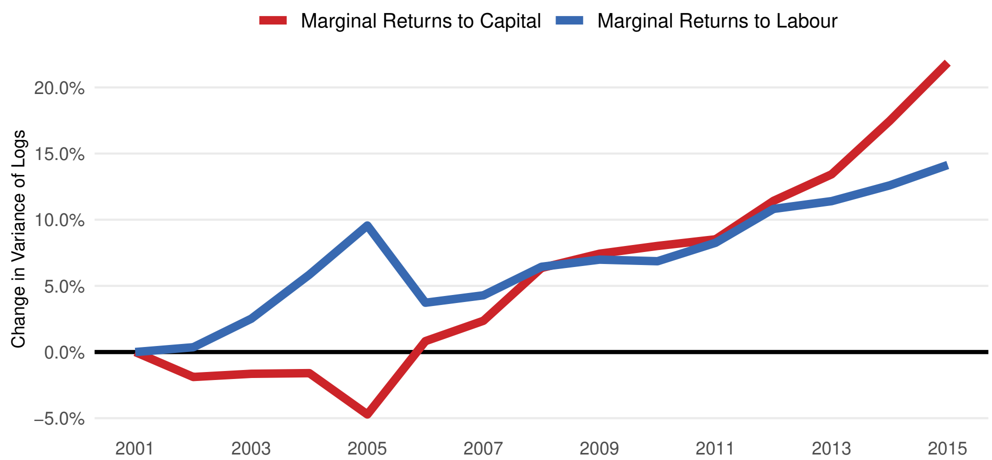

Displays the distribution of (log) marginal products of labour and capital in 2001 and 2015. Each are normalized relative to the average marginal products for that year.

(from 2.39 in 2001 to 2.61 in 2015) while the variance of log labour returns increases by nearly 15 percent (from 0.94 in 2001 to 1.09 in 2015). While there are year-to-year variations in the change, the overall trend towards higher variance in log returns is clear. These changes are economically significant. If we presume an elasticity of $\epsilon = 3$, then the change in aggregate deadweight loss would be between 25 and 30 percent. This does not imply TFP would decline by that amount, only that the aggregate deadweight loss in 2015 would be that much larger than it was in 2001. In Sections 3 and 4, we more precisely quantify the effect of distortions on aggregate productivity using a richer model.

There are notable differences across regions as well. We separate the changes in the variance of marginal returns across the five broad regions of Canada by running equation 9 separately for each region. We further aggregate the prairie provinces and Atlantic provinces into their respective regions. We find much of the increase is driven by larger variation within the western provinces of British Columbia and the three prairie provinces. Ontario also experiences notable increases in the dispersion of capital returns across firms within that province, especially following the financial crisis. We display these changes in Figure 4. And again, these are economically meaningful increases. They may imply that strong demand for labour and capital in economically expanding regions of Canada over this time, which indeed were the four western provinces and Ontario , was not associated with a sufficiently large movement of workers or investment capital towards those regions.

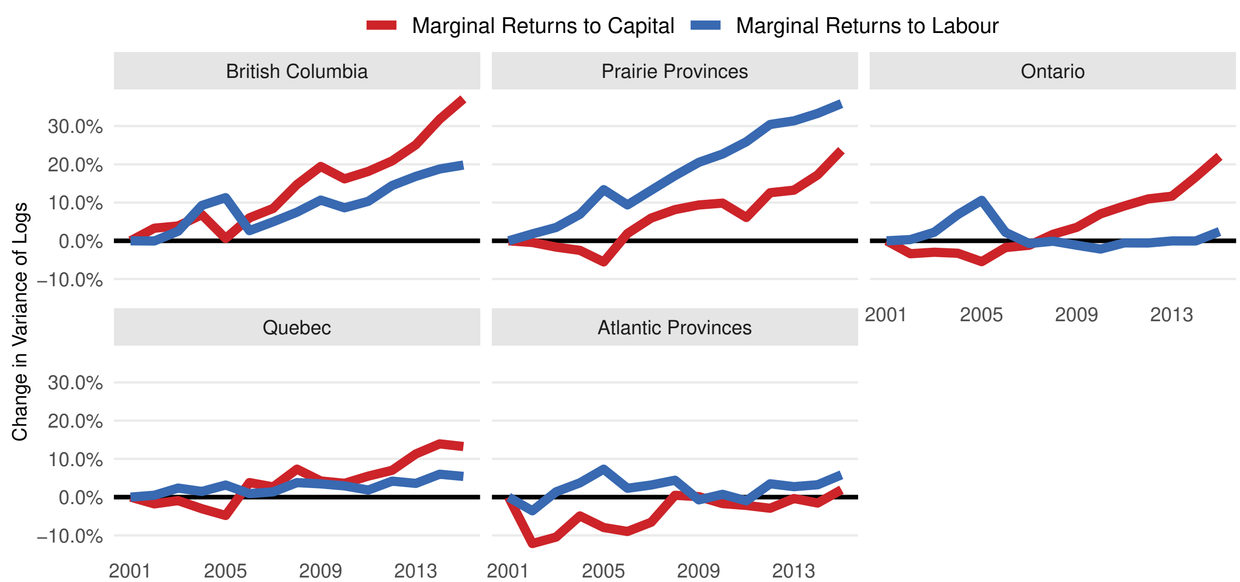
Finally, and most notably for the quantitative analysis to come, we disaggregate the change in variance in marginal returns by sector. For the 30 sectors for which we are able to estimate such factor returns in the data, we display the change between 2001 and 2015 in Figure 5. It is evident in this figure that many large and influential sectors of Canada's economy saw increases in our measure of between-firm misallocation. Manufacturing of paper products, primary metals, and transport equipment (i.e., autos) saw large increases. The latter three in particular experi-

Figure 3: Change in the Variance of Marginal Revenue Products of $L$ and $K$




Displays the change in the variance of log marginal revenue products of capital and labour in Canada from 2001 to 2015.

Figure 4: Change in the Variance of Marginal Revenue Products of $L$ and $K$, by Region




Displays the change in the variance of log marginal revenue products of capital and labour within five broad regions of Canada from 2001 to 2015.

enced meaningful increases in capital returns dispersion across firms. Mining, oil and gas also experienced large increases – again, perhaps reflecting an expanding activity that saw insufficient movement of labour and investment towards the sector. Not all large sectors experienced changes, to be sure. Construction and wholesale and retail trade saw little change in the dispersion of their factor returns across firms.

Changes in the variance of log factor returns of the magnitude we document here suggest factor misallocation has increased. This has potentially important implications for Canada's aggregate economy. Merely documenting the change in dispersion, however, does not shed light on by how much aggregate productivity might be affected. To answer that question requires significantly more structure be imposed on the data. To that end, we turn to our main quantitative modeling analysis that maps to key moments of our firm-level data. With this model, we examine how factor market distortions misallocate labour and capital between sectors and within them. The within-sector dimension to this analysis is unique and only possible with our detailed firm-level data.

## 3 A Quantitative Model of Misallocation

To begin, consider a continuum of firms that produce differentiated goods within a set of $J$ sectors. To produce output, firm $i$ in sector $j$ uses labour and capital within a constant returns to scale production technology

$$y_j(i) = \varphi_j(i) k_j(i)^{1-\alpha_j} l_j(i)^{\alpha_j}. \tag{10}$$

Output in sector $j$ is a composite good produced according to

$$Y_j = \left( \int y_j(i)^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}}, \tag{11}$$

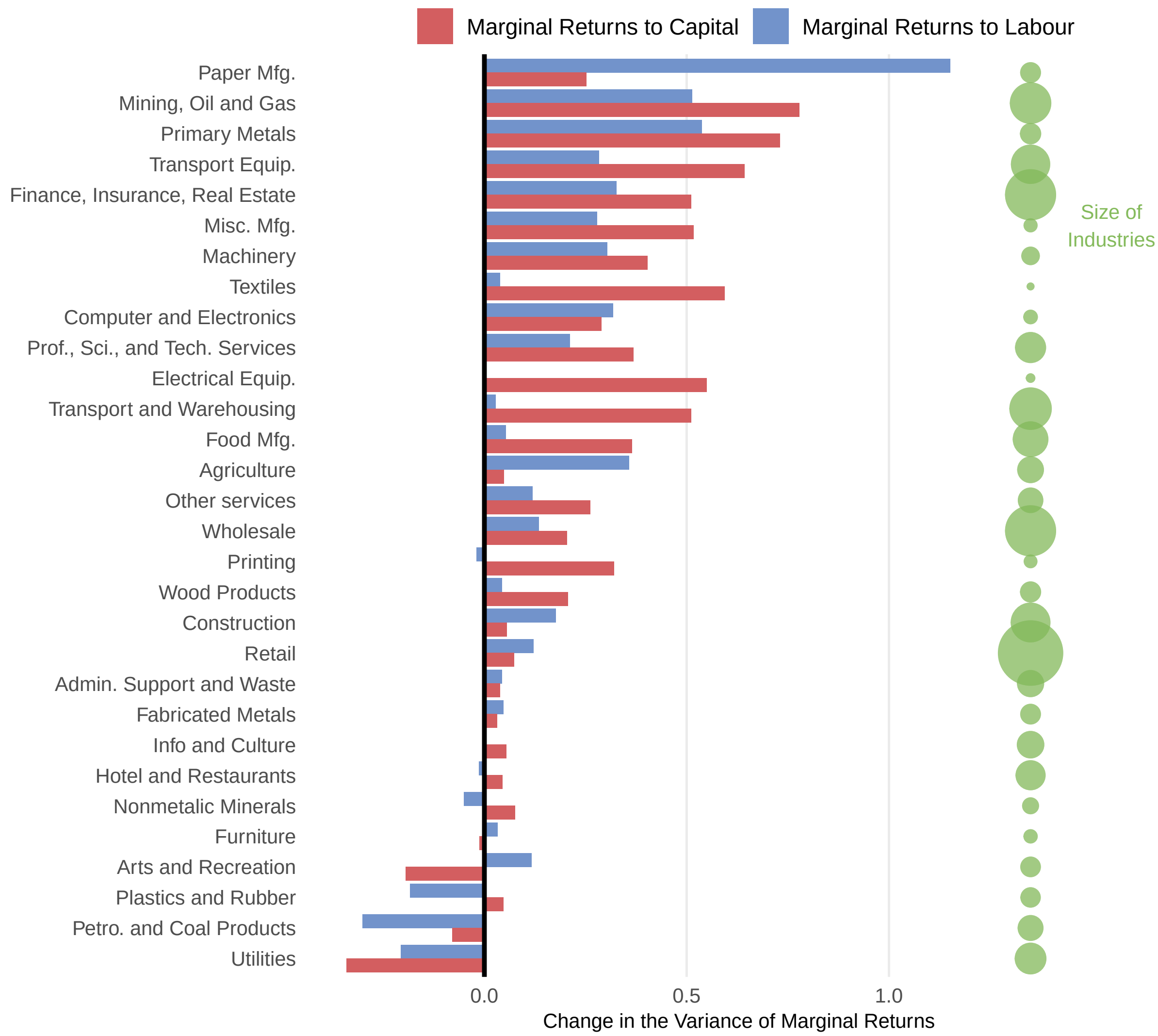
where $\sigma$ is the elasticity of substitution across firm output. Finally, we presume aggregate output is

$$Y = \prod_{j=1}^{J} Y_j^{\beta_j}, \tag{12}$$

where $\sum_j \beta_j = 1$.

Solving for the equilibrium allocations of labour and capital across firms is straightforward. Given the CES aggregation of firm output within each sector, there will be a markup $m = \sigma/(\sigma - 1)$ over marginal costs. Total spending on labour and capital will therefore be $p_j(i)y_j(i)/m$, of which $\alpha_j$ is allocated to labour and $1 - \alpha_j$ to capital. Thus, given wedges within the labour and capital markets – $\tau_j(i)^l$ and $\tau_j(i)^k$, respectively – individual firms optimally choose their input quantities

Figure 5: Change in the Variance of Marginal Revenue Products Between 2001 and 2015, by Sector




Displays the change in the variance of log marginal revenue products of capital and labour between 2001 and 2015 within each of our 30 sectors. The total size of each sector, in terms of its total revenue, is illustrated by the column of points to the right of the figure.

to satisfy

$$\begin{aligned} \tau_j(i)^k \cdot R \cdot k_j(i) &= (1-\alpha_j)p_j(i)y_j(i)/m, && (13) \\ \tau_j(i)^l \cdot w \cdot l_j(i) &= \alpha_j p_j(i)y_j(i)/m. && (14) \end{aligned}$$

Here we consider distortions as equivalent to taxes on the purchase of labour and capital. Absent factor market wedges, the marginal revenue products of labour and capital would equalize across all firms and sectors.

### 3.1 Between-Sector Distortions

As discussed earlier, variation in marginal revenue products will reflect variation in both $\tau_j(i)^k$ and $\tau_j(i)^l$. And since undistorted wages, capital costs, and markups are common to all firms, one can use these expressions to determine the equilibrium allocation of employment and capital across sectors. In particular,

$$\begin{aligned} k_j &= \frac{(1-\alpha_j)\beta_j/\bar{\tau}_j^k}{\sum_{i=1}^J (1-\alpha_i)\beta_i/\bar{\tau}_i^k}, && (15) \\ l_j &= \frac{\alpha_j\beta_j/\bar{\tau}_j^l}{\sum_{i=1}^J \alpha_i\beta_i/\bar{\tau}_i^l}, && (16) \end{aligned}$$

where $\bar{\tau}_j$ denotes the revenue-weighted harmonic mean of capital and labour distortions, respectively. In an undistorted economy, labour and capital would be allocated only according to the intensity with which these inputs are used in production and the importance of each sector in final demand. That is, $l_j^* \propto \alpha_j\beta_j$ and $k_j^* \propto (1-\alpha_j)\beta_j$. It may be instructive to write equilibrium allocations in terms of optimal allocations and distortions as follows,

$$\begin{aligned} k_j &= k_j^*/\bar{\tau}_j^k, && (17) \\ l_j &= l_j^*/\bar{\tau}_j^l, && (18) \end{aligned}$$

where we implicitly use a balanced-budget restriction on factor distortions to ensure $\sum_j f_j \bar{\tau}_j^f = 1$ for both factors $f \in \{l, k\}$.[4] Sectors with high distortions will therefore employ fewer factors relative

[4]Although our model does not feature intersectoral linkages through intermediate inputs, but our measure of between-sector distortions is not affected by this modeling choice. The effect of input-output linkages on optimal labour and capital allocations would be fully captured within each sector's share of total sales. Specifically, given a direct requirements matrix $\boldsymbol{A}$ total sales would be $\boldsymbol{\gamma} = (1-\boldsymbol{A})^{-1}\boldsymbol{\beta}$, where $\boldsymbol{\gamma}$ is a vector of sectoral sales $\gamma_j$, $(1-\boldsymbol{A})^{-1}$ is the Leontief Inverse Matrix, and $\boldsymbol{\beta}$ is a vector of final demand shares $\beta_j$ from equation 12. Given this, optimal labour and capital allocations would be proportional to $\alpha_j\gamma_j$ and $(1-\alpha_j)\gamma_j$, respectively. Since we use total revenue from data, we would infer the same between-sector distortion measures in a model with full input-output linkages as in this model without them. Importantly, however, the aggregate effect of these between-sector distortions would be larger in a model with such linkages between $\sum_j \gamma_j > 1$ while $\sum_j \beta_j = 1$. Our results should therefore be viewed as underestimating the potential effect of between-sector distortions. See Jones (2011) for a fuller discussion of misallocation in the context of a model with input-output linkages.

to an undistorted economy. This matters for aggregate TFP since,

$$A = \prod_{j=1}^{J} \left( A_j l_j^{\alpha_j} k_j^{1-\alpha_j} \right)^{\beta_j} \tag{19}$$

and therefore, denoting the ratio of distorted equilibrium values of a variable $x$ to the optimal values in a counterfactual equilibrium $x^*$ as $\hat{x} \equiv x/x^*$, we have

$$\hat{A} = \prod_{j=1}^{J} \left[ \frac{\hat{A}_j}{\left(\bar{\tau}_j^l\right)^{\alpha_j} \left(\bar{\tau}_j^k\right)^{1-\alpha_j}} \right]^{\beta_j} . \tag{20}$$

Between-sector distortions are captured by the denominator while within-sector distortions are captured by the numerator. Sectoral TFP will change due to distortions altering the within sector allocation of factors and by distorting the firm size distribution. We turn to these next.

### 3.2 Within-Sector Distortions

Within a sector, distortions to labour and capital allocations can affect the firm size distribution, aggregate sectoral output, and aggregate sectoral TFP. Given sectoral output from equation 11, total sales of firm $i$ in sector $j$ is

$$p_j(i) y_j(i) = P_j Y_j \left( \frac{p_j(i)}{P_j} \right)^{1-\sigma} . \tag{21}$$

And since prices are a markup over marginal costs,

$$c_j(i) \propto \frac{1}{\varphi_j(i)} \left( \tau_j(i)^k \right)^{1-\alpha_j} \left( \tau_j(i)^l \right)^{\alpha_j} , \tag{22}$$

we have total sales given by

$$p_j(i) y_j(i) \propto \left( \frac{\varphi_j(i)}{\left(\tau_j(i)^k\right)^{1-\alpha_j} \left(\tau_j(i)^l\right)^{\alpha_j}} \right)^{\sigma-1} . \tag{23}$$

All else equal, higher productivity firms have higher sales. But firms facing higher capital or labour distortions will therefore be smaller than their optimal firm size, which is proportional to $\varphi_j(i)^{\sigma-1}$. A distorted firm size distribution will also distort the allocation of labour and capital across firms, since the number employed depends on input purchases which itself depends on sales, as we saw in equations 13 and 14. And these allocations determine sectoral TFP through

$$A_j = \left( \int \left[ \varphi(i) l_j(i)^{\alpha_j} k_j(i)^{1-\alpha_j} \right]^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}} . \tag{24}$$

To proceed further, summarize labour and capital distortions within a single variable $\tau_j(i) \equiv \left(\tau_j(i)^k\right)^{1-\alpha_j}\left(\tau_j(i)^l\right)^{\alpha_j}$. Using optimal allocations for labour and capital, one can show

$$A_j = \left(\int \left(\varphi_j(i) \cdot \frac{\bar{\tau}_j}{\tau_j(i)}\right)^{\sigma-1} di\right)^{1/(\sigma-1)}, \tag{25}$$

as in Hsieh and Klenow (2009), and therefore $\hat{A}_j = A_j/A_j^*$ where

$$A_j^* = \left(\int \left(\varphi_j(i)\right)^{\sigma-1} di\right)^{1/(\sigma-1)}. \tag{26}$$

With these expressions in hand, combined with out firm-level data on labour, capital, and output, we may proceed to quantify the effect of misallocation on Canada's TFP.

# 4 Quantitative Analysis

The variance in marginal returns to labour and capital across firms is central to our analysis. As we've seen, increases in this dispersion may correspond to increases in the inefficiency with which factors of production are allocated across firms, sectors, or regions in Canada. In this section, we highlight various measures of this dispersion and quantify the extent to which Canada's aggregate productivity may be reduced as a result.

## 4.1 Model Calibration

There are relatively few parameters in the model. But they are important. Estimating optimal labour and capital allocations across sectors requires sector-specific values for two parameters:

1. **Labour's share of value-added $\alpha_j$:** We use the Canadian national accounts information reported in the symmetric input-output tables for 2015 in Statistics Canada data table 36-10-0001-01. This parameter is held constant over time. To be clear, the main results concerning changes in misallocation are not sensitive to using a value for $\alpha_j$ calculated from the data, but these are not vetted for public release.

2. **Each sector's share of total expenditures $\beta_j$:** For our main results, we use total sales, by sector, from our firm-level data. This share is highly correlated with industry output shares from the national accounts. Indeed, the correlation between the two is nearly 0.99 in 2015, although this share varies over time.

To estimate within-sector distortions, we also require each sector's share of aggregate expenditure but, in addition, we require the elasticity of substitution across products $\sigma$. For our main results we use $\sigma = 3$ as in Hsieh and Klenow (2009). As this is low relative to estimates in the literature, and as larger values of this parameter correspond to larger efficiency costs, we view this as conservative.

## 4.2 Between-Sector Distortions

We begin by quantifying the effect of firm level distortions on the allocation of labour and capital across sectors. The aggregate effect may be flexibly quantified using equation 20 and, in particular, is the inverse of the weighted geometric average of labour distortions $\bar{\tau}_j^l$ and capital distortions $\bar{\tau}_j^k$. Specifically,

$$\hat{A}^{\text{between sector}} = \prod_{j=1}^{J} \left[ \left(\bar{\tau}_j^l\right)^{\alpha_j} \left(\bar{\tau}_j^k\right)^{1-\alpha_j} \right]^{-\beta_j}. \tag{27}$$

Sectors where distortions result in a higher level of employment or capital stock than is optimal to maximize aggregate productivity will have $\bar{\tau}_j^l < 1$ or $\bar{\tau}_j^k < 1$. Intuitively, this reflects distortions that lower the marginal cost of labour or capital for firms within that sector, thereby resulting in higher use of either (or both) factors. We report our measures of labour and capital allocations, both observed and optimal, in Table 1. We also report the industry share of total revenue $\beta_j$ and each sector's labour input share $\alpha_j$.

We find distortions shift the allocation of labour and capital across sectors in quantitatively meaningful ways. In our data, averaged over all years, we observe 8.5% of labour and 15.7% of capital is allocated, for example, to the mining, oil, and gas sector. Given sectoral revenues and labour intensities, however, the optimal share of labour and capital for that sector is 5.7% and 13.4%, respectively. In other sectors, such as most manufactured goods sectors, the optimal allocation is larger than the observed allocations. These differences are large and imply roughly one-quarter of overall employment and capital stock would need to be reallocated across sectors in order to achieve the first-best allocations $l_j^*$ and $k_j^*$. We display the implied measure of average distortions, which is the ratio of observed to optimal allocations, in panel (a) and (b) of Figure 6. While these measures are averaged over all years in our sample (as in Table 1), we separately estimate these distortions for each year to construct the effect on aggregate productivity. We display this in panel (c) of Figure 6. We find between-sector distortions, on average, lower Canada's aggregate productivity by 22% during our period of study. We also find that such distortions have been gradually worsening since 2004. From 2004 to 2015, worsening misallocation between sectors has lowered aggregate productivity by nearly 12%. Given the period of improvement between 2001 and 2004, we find that aggregate productivity was 6% lower in 2015 than it would have otherwise have been had between-sector distortions remained constant since 2001.

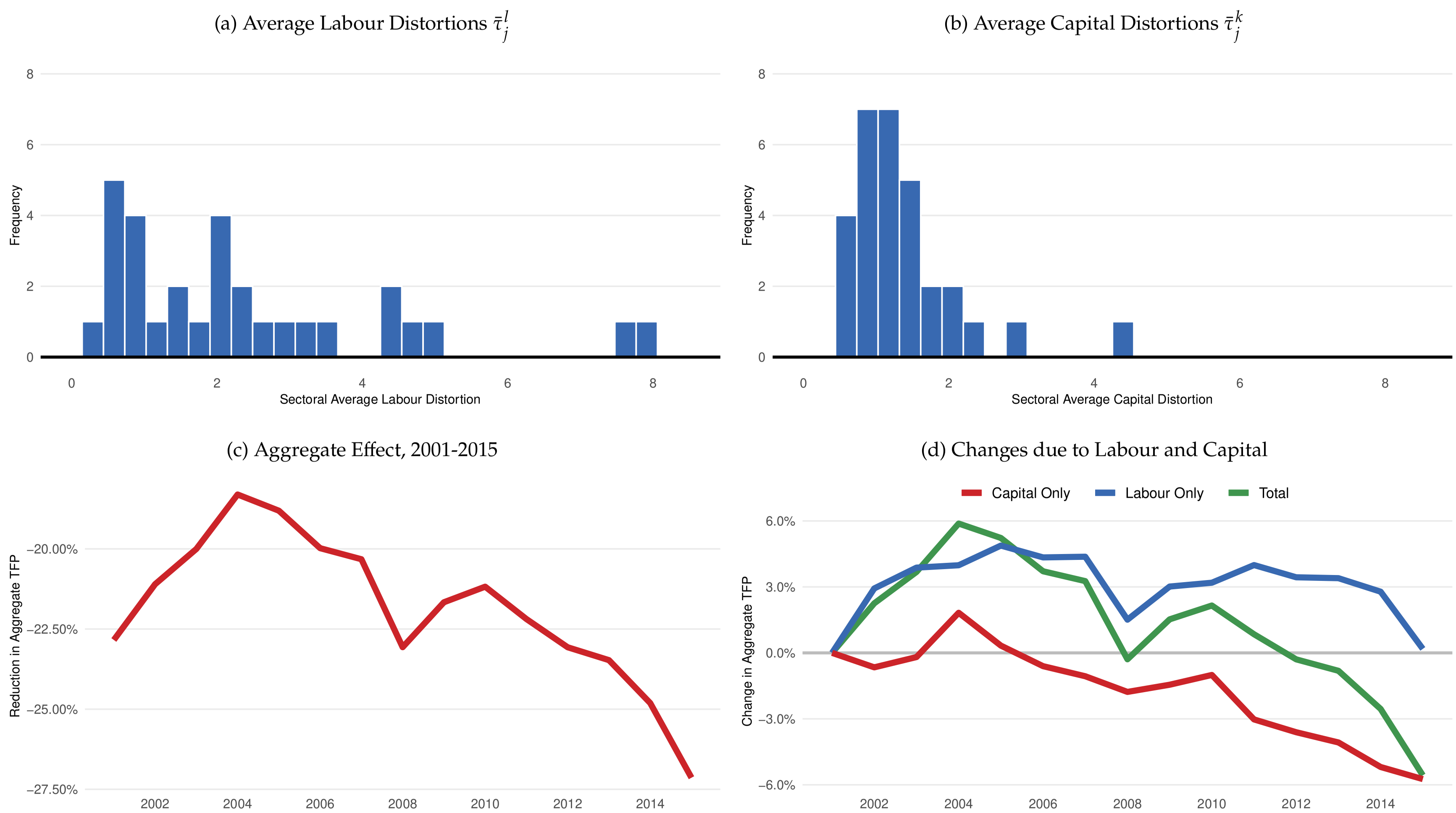
The aggregate effect of between-sector misallocation may be further decomposed into contributions from labour and capital distortions. Comparing panel (a) and (b) of figure 6 makes clear the dispersion in average labour distortions exceed those for capital. We quantify each factor's individual contribution by exploiting the multiplicative form of equation 27. Specifically, we can capture the contribution of labour misallocation from $\prod_{j=1}^{J} \left(\bar{\tau}_j^l\right)^{-\beta_j \alpha_j}$ and capital misallocation from $\prod_{j=1}^{J} \left(\bar{\tau}_j^k\right)^{-\beta_j (1-\alpha_j)}$. Of the 22% overall effect, we find nearly two-thirds is accounted for labour mis-

Table 1: Allocations and Industry Parameters for Between-Sector Distortions (%)

| | Actual Allocations | | Optimal Allocations | | Revenue Shares | Labour Shares |
|---|---|---|---|---|---|---|
| Sector | Labour | Capital | Labour | Capital | $\beta_j$ | $\alpha_j$ |
| Agriculture | 0.2 | 2.5 | 1.0 | 3.2 | 1.9 | 30.0 |
| Mining, Oil and Gas | 8.5 | 15.7 | 5.7 | 13.4 | 8.8 | 37.7 |
| Utilities | 2.3 | 6.6 | 1.7 | 4.5 | 2.9 | 35.3 |
| Construction | 2.9 | 2.4 | 6.4 | 3.8 | 5.3 | 70.7 |
| Food Mfg. | 3.6 | 2.0 | 3.9 | 4.5 | 4.1 | 55.0 |
| Textiles | 0.0 | 0.1 | 0.3 | 0.1 | 0.2 | 77.3 |
| Wood Products | 0.7 | 1.0 | 1.5 | 0.9 | 1.3 | 71.6 |
| Paper Mfg. | 2.0 | 1.7 | 1.7 | 1.5 | 1.6 | 62.4 |
| Printing | 0.9 | 0.3 | 0.5 | 0.3 | 0.4 | 73.6 |
| Petro. and Coal Products | 0.5 | 1.4 | 1.3 | 4.2 | 2.5 | 29.9 |
| Plastics and Rubber | 0.3 | 0.7 | 1.4 | 0.9 | 1.2 | 69.8 |
| Nonmetalic Minerals | 0.9 | 0.5 | 0.7 | 0.5 | 0.6 | 66.7 |
| Primary Metals | 0.7 | 1.8 | 1.5 | 1.4 | 1.5 | 60.0 |
| Fabricated Metals | 0.5 | 0.5 | 1.5 | 0.6 | 1.2 | 77.6 |
| Machinery | 0.3 | 0.4 | 1.0 | 0.5 | 0.8 | 73.4 |
| Computer and Electronics | 0.5 | 0.3 | 0.7 | 0.5 | 0.6 | 68.6 |
| Electrical Equip. | 0.2 | 0.2 | 0.4 | 0.3 | 0.3 | 70.0 |
| Transport Equip. | 3.9 | 2.8 | 6.3 | 5.6 | 6.0 | 61.6 |
| Furniture | 0.1 | 0.1 | 0.6 | 0.2 | 0.4 | 84.3 |
| Misc. Mfg. | 0.1 | 0.1 | 0.4 | 0.2 | 0.3 | 77.4 |
| Wholesale | 3.6 | 2.5 | 12.3 | 11.0 | 11.8 | 61.4 |
| Retail | 29.8 | 7.4 | 23.0 | 10.1 | 17.7 | 76.6 |
| Transport and Warehousing | 9.5 | 7.4 | 6.4 | 6.4 | 6.4 | 58.9 |
| Info and Culture | 9.9 | 3.4 | 2.1 | 3.7 | 2.8 | 44.2 |
| Finance, Insurance, Real Estate | 9.2 | 31.1 | 5.0 | 14.2 | 8.8 | 33.7 |
| Prof., Sci., and Tech. Services | 3.9 | 1.2 | 3.2 | 2.3 | 2.9 | 66.3 |
| Admin. Support and Waste | 1.2 | 1.3 | 2.5 | 1.7 | 2.2 | 68.1 |
| Arts and Recreation | 0.5 | 1.1 | 1.2 | 0.8 | 1.0 | 68.1 |
| Hotel and Restaurants | 2.6 | 2.4 | 3.5 | 1.3 | 2.6 | 79.1 |
| Other services | 0.5 | 1.2 | 2.4 | 1.3 | 1.9 | 72.8 |

Reports the average observed labour and capital allocations $l_j$ and $k_j$, the optimal allocations $l_j^*$ and $k_j^*$, the shares of total revenue $\beta_j$, and the labour input share $\alpha_j$ by sector between 2001 and 2015.

Figure 6: Between-Sector Distortions in Canada




Displays our measure of average between-sector distortions $\bar{\tau}_j^l$ and $\bar{\tau}_j^k$ across industries in panels (a) and (b). These are averaged across all years. The aggregate effect of these distortions, and how that effect has changed over time, is displayed in panel (c). Finally, panel (d) displays the contribution of labour and capital distortions to changes in the overall aggregate effect.

allocation while the remaining one-third is accounted for by capital misallocation.[5] But changes in each factor's contribution, which we plot in panel (d) of Figure 6 reveals that capital markets fully accounts for the deterioration between 2001 and 2004. Labour markets, for the most part, improved in many years. From 2001 to 2014, between-sector misallocation of labour eased and contributed 3% to aggregate TFP growth. But in 2015, this gain was fully reversed. We find a majority of this drop is due to a rising measure of labour misallocation in Canada's transport manufacturing sector. Specifically, the optimal share of labour allocated to that sector rises but the observed share rises more slowly. In any case, these broad results suggest the efficiency with which capital is allocated has fallen and has negatively contributed to Canada's overall productivity growth.

### 4.3 Within-Sector Distortions

Firm distortions not only affect allocations between sectors but also affect allocations between firms within each sectors. Indeed, the between sector distortions of the previous section captured the average effect of firm-level distortions $\tau_j(i)^k$ and $\tau_j(i)^j$. Such firm-level distortions also change firms' size and the share of a sector's employment and capital allocated to each firm. It can therefore lower sectoral TFP, which has implications for aggregate productivity. Specifically, from equation 20, we have

$$\hat{A}^{\text{within sector}} = \prod_{j=1}^{J} \hat{A}_j^{\beta_j}, \tag{28}$$

where $\hat{A}_j = A_j/A_j^*$ using equations 25 and 26. Due to disclosure limitations on the firm-level data, however, we must report the sectoral results at a higher level of aggregation than was possible for the between-sector results. Specifically, we aggregate sectors with one-digit NAICS codes 2 together, as well as those with 5. We aggregate manufacturing sectors into their respective two-digit NAICS, with the exception of food products (311) and textiles (313). With this data, we estimate equations 25, 26, and 28 and list our results using the more aggregate sectoral groupings in Table 2.

We estimate $\hat{A}_j$ for each sector over time and report the results in Table 2 for selected years. To be clear, there is a material degree of year-to-year volatility in the measure. We find that misallocation exists within all sectors, not surprisingly, and is often large. Such large effects are common in the literature and should not be interpreted as a measure of what is feasible for policy reforms to address. So, we look at changes over time as more informative in that regard. Most sectors see improvements over the first two periods, while four sectors see reductions in within-sector allocative efficiency. Overall, however, aggregate within-sector misallocation worsens by over 15 percent during the sample period. As noted, however, the volatility in this measure suggests one should consider this change not statistically significant. The change over the period 2001 to

[5] The (weighted geometric) average labour distortion across sectors and years is 0.857 and the average for capital distortions is 0.911. And since the total between-sector effect is $\hat{A}^{\text{between sector}} = 0.781$, labour's contribution to that total is $log(0.857)/log(0.781) = 0.62$.

Table 2: Effect of Within-Sector Distortions on Sectoral Productivity $A_j$

| Sector | Effect on Sectoral Productivity $\hat{A}_j$ 2001 | 2008 | 2015 | Change (%) 2001-2015 |
|---|---|---|---|---|
| Agriculture | 0.060 | 0.029 | 0.038 | -37 |
| Mining, Oil, and Gas; Utilities; Constr. | 0.008 | 0.004 | 0.018 | 113 |
| Food Mfg. | 0.048 | 0.099 | 0.100 | 107 |
| Textiles | 0.119 | 0.148 | 0.434 | 265 |
| Manufacturing, NAICS 32 | 0.107 | 0.110 | 0.048 | -55 |
| Manufacturing, NAICS 33 | 0.078 | 0.039 | 0.135 | 72 |
| Wholesale | 0.034 | 0.016 | 0.059 | 74 |
| Retail | 0.024 | 0.024 | 0.020 | -15 |
| Transport and Warehousing | 0.027 | 0.017 | 0.029 | 9 |
| Info; FIRE; Prof., Sci., and Tech. Services; Admin | 0.003 | 0.002 | 0.000 | -93 |
| Arts and Recreation | 0.008 | 0.035 | 0.033 | 305 |
| Hotel and Restaurants | 0.087 | 0.103 | 0.088 | 1 |
| Other Services | 0.002 | 0.125 | 0.073 | 3,374 |
| Aggregate | 0.0194 | 0.0151 | 0.0164 | -15.5 |

Reports the average effect of within-sector distortions on sectoral productivity $\hat{A}_j$. The bottom row reports the aggregate effect of these distortions on aggregate productivity, using sectoral revenue shares and equation 28.

2014, for example, is an improvement of 8.2 percent. And the standard deviation of the absolutely value of the change across all years (relative to 2001) is 15.3 percent. But as we will turn to next, this volatility is not found in the between sector results even using the more aggregated sector groupings. One may therefore wish to put more weight on the aggregate implications of the between-sector results than the within-sector results.

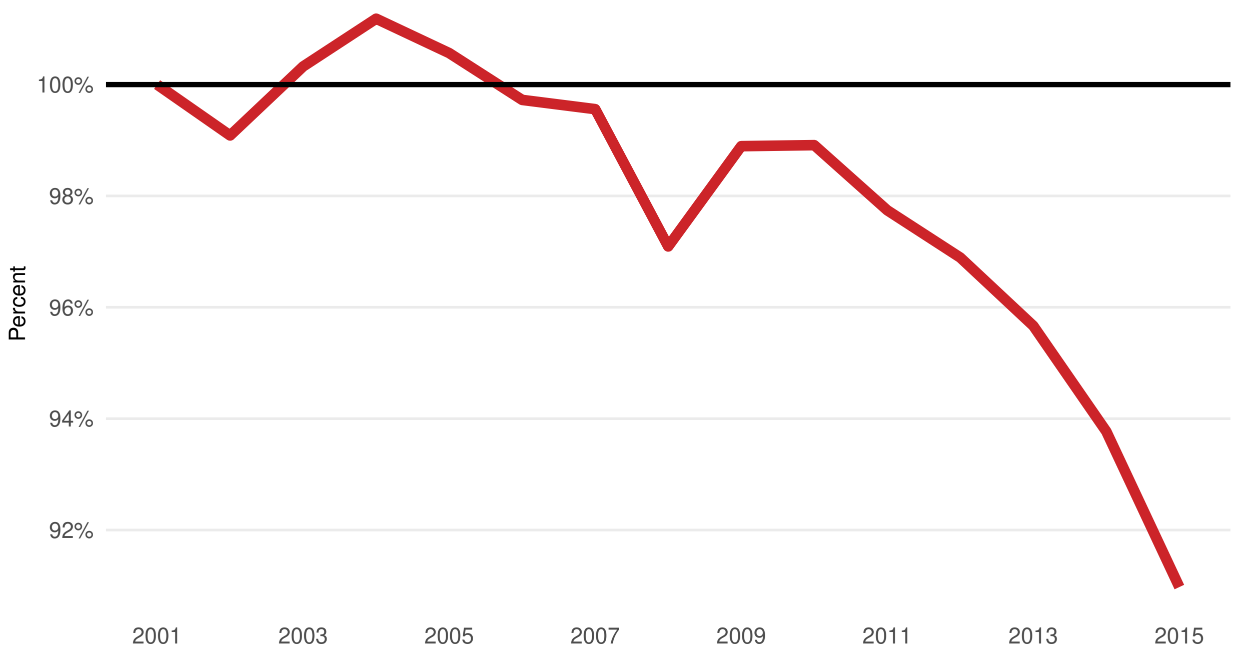
### 4.4 The Aggregate Effect of Misallocation

Both the between- and within-sector distortions aggregate to affect changes in Canada's national productivity. We saw that in the early years of our sample, reductions in the extent of distortions lowered misallocation of labour and capital across sectors. This resulted in improvements in aggregate productivity. After 2004, however, changes in distortions increased the extent of misallocation and detracted from aggregate productivity. We summarize all of these effects using equation 20, which is equivalent to

$$\hat{A} = \left(\hat{A}^{\text{between sector}}\right) \times \left(\hat{A}^{\text{within sector}}\right). \tag{29}$$

To combine the within- and between-sector distortions in this way, however, requires we use a common sectoral classification for both measures of misallocation. To that end, we re-estimate the between-sector results from Table 1 using the same sectors as in Table 2 and display the results in Figure 7 for all years of our data. We find that worsening between-sector distortions in the final period of our analysis leads to a 9 percent reduction in aggregate TFP relative to the initial

Figure 7: Between-Sector Distortions, Using Aggregated Sectors




Displays our measure of the aggregate effect of between-sector distortions at the more aggregated sector groupings.

period. Importantly, and unlike the within-sector results of the previous section, the volatility in these estimates over time is low. These results are also in-line with the previous between-sector distortions estimated using the more detailed sectoral classification.

Combining both within- and between-sector distortions reveals a relatively large negative effect on Canada's aggregate productivity. Using equation 29, we find a 23 percent reduction in aggregate TFP in the final period compared to the initial period when both types of misallocation are combined. Of this overall reduction, nearly two-thirds is accounted for by the change in within-sector distortions and just over one-third by between-sector distortions. But as discussed in the previous section, the year-by-year volatility in the within-sector distortion is large. This imprecision of the estimate suggests the contribution of changes in within-sector distortions should therefore be discounted. The between-sector contribution, however, is much more precisely estimated and accounts for a 9 percent reduction in overall aggregate productivity over the period. While large, this magnitude is similar to some recent results found for some other countries. Calligaris (2015) find that aggregate TFP in Italy, for example, could have been 21 percent larger had misallocation there not worsened between 1997 and 2011.

Compared to Canada's observed productivity gap with the United States, our results suggest that nearly all can be accounted for by worsening misallocation. Specifically, the Penn World Table (*ctfp* in version 10.1) finds Canada's TFP in 2015 was 0.828 of the U.S. level, which implies productivity would increase by nearly 21 percent if Canada matched the U.S. level. In 2001, however, Canada's relative TFP was 0.972 of the U.S. level. Had aggregate productivity kept pace with the U.S. level, then Canadian TFP in the final period of our analysis would have been 17.4 percent higher. Our results therefore imply that worsening between-sector misallocation may account for approximately half of Canada's widening productivity gap with the United States.

### 4.5 Firm Entry and Exit

Changes in a country's aggregate misallocation must come from one of the two sources, either through the change of misallocation across operating firms or through the entry and exit of firms. In the previous analysis, our sample includes all firms that operated for at least one year between 2001 and 2005 with available information on key variables. In this sample, new entrants could enter and existing firms could exit, which may impact estimates related to the within-sector allocation of labour and capital. To assess the contribution of these dynamics to our estimates, we switch to the balanced sample, which includes only firms present in all sample years, thereby eliminating new entrants and exiting firms during this period.

Restricting to continually operating firms has a substantial effect on the measure of within-sector misallocation. We display the estimates for the effect of these distortions on sectoral productivity on $A_j$ in Table 3. Notably, the magnitude of the distortions is smaller than when entry and exit are included (as reported in Table 2) and the change over time between 2001 and 2015 is also smaller. In aggregate, we find that increasing within sector distortions lower sectoral productivity by 12.3 percent in 2015 compared to 2001. The magnitude and direction of these estimates are similar to the baseline results reported previously with entry and exit included. But while the volatility is somewhat lower in this case relative to our baseline results, the standard deviation of aggregate changes relative to 2001 across all years is 13.4 percentage points and therefore one should discount the significance of this change.

In terms of between-sector distortions, our estimates based on the sample of continuous firms is relatively similar, both in magnitude and in the pattern of changes over time. We find little change over the initial years of our sample (as was the case in Figure 7), followed by a systematic decline through to 2015. Without entry and exit, we estimate that the effect of worsening between-sector distortions decreases aggregate productivity by 12.7 in 2015 compared to 2001. This is somewhat larger than our baseline estimates of 9 percent. We conclude that whether one allows for entry or exit has only a limited effect on our overall estimates. And with the within-sector estimates discounted due to substantial volatility in either case, our baseline approach of including entry and exit provides a more conservative estimate of changes in between-sector distortions.

### 4.6 Measurement Error

Recent studies emphasize that much of the apparent misallocation might be due to measurement error. Gollin and Udry (2021) find that much of the observed misallocation in sub-Saharan African agriculture is due to measurement error and heterogeneity, while Bils, Klenow, and Ruane (2021)

Table 3: Effect of Within-Sector Distortions on Sectoral Productivity $A_j$ (No Entry or Exit)

| | Effect on Sectoral Productivity $\hat{A}_j$ | | | Change (%) |
|---|---|---|---|---|
| Sector | 2001 | 2008 | 2015 | 2001-2015 |
| Agriculture | 0.142 | 0.220 | 0.138 | -3 |
| Mining, Oil, and Gas; Utilities; Constr. | 0.214 | 0.227 | 0.068 | -68 |
| Food Mfg. | 0.193 | 0.185 | 0.104 | -46 |
| Textiles | 0.557 | 0.647 | 0.562 | 1 |
| Manufacturing, NAICS 32 | 0.206 | 0.193 | 0.145 | -30 |
| Manufacturing, NAICS 33 | 0.149 | 0.235 | 0.126 | -15 |
| Wholesale | 0.053 | 0.033 | 0.082 | 56 |
| Retail | 0.016 | 0.034 | 0.034 | 116 |
| Transport and Warehousing | 0.107 | 0.065 | 0.093 | -13 |
| Info; FIRE; Prof., Sci., and Tech. Services; Admin | 0.022 | 0.046 | 0.017 | -19 |
| Arts and Recreation | 0.035 | 0.022 | 0.023 | -36 |
| Hotel and Restaurants | 0.130 | 0.156 | 0.182 | 40 |
| Other Services | 0.279 | 0.339 | 0.147 | -47 |
| Aggregate | 0.0690 | 0.0875 | 0.0605 | -12.3 |

Reports the average effect of within-sector distortions on sectoral productivity $\hat{A}_j$, based only on the sample of firms that operate continuously within our sample period 2001 to 2015. The bottom row reports the aggregate effect of these distortions on aggregate productivity, using sectoral revenue shares and equation 28.

find that additive measurement errors explain a significant portion of the increasing TFPR dispersion in the U.S. In this subsection, we discuss how measurement error might impact our results and the tests we conduct to address these concerns.

#### 4.6.1 Mismeasurement due to Misreporting

One potential source of mismeasurement may arise from misreporting. For example, Bils et al. (2021) finds that additive measurement error in revenue and inputs—such as omitted revenue or inputs from certain products—accounts for a large portion of the declining trend in U.S. allocative efficiency. Specifically, their approach presumes measurement error could be captured by the degree to which high average revenue product plants exhibit a low elasticity of revenue growth with respect to input growth. To investigate whether additive measurement error contributes to declining allocative efficiency in our data, we follow the approach of Bils et al. (2021) to examine this potential source of mismeasurement. We first tested whether firms with higher TFPR exhibit lower elasticities of revenue growth relative to input growth. Our results show a negative correlation between firms' TFPR and revenue-input elasticity across TFPR deciles for the periods 2001–2005, 2006–2010, and 2011–2015, in both clean and balanced samples. This finding suggests that additive measurement error is present in our data and contributes to the observed allocative inefficiency. However, the absolute value of this negative correlation declined significantly over time, decreasing by approximately 50 % from the first to the last period. As the impact of additive measurement error diminishes over time, it is unlikely to explain the observed worsening of misallocation.

Next, we construct the corrected variance of ln(TFPR) as in Step 4 of Bils et al. (2021)'s procedure, incorporating the impact of measurement errors. Comparing the corrected variance of ln(TFPR) with the measured variance, we find that the gap between these two values declines over time, from the period 2001–2005 to 2011–2015. This suggests that our findings on allocative efficiency are increasingly robust to measurement error over time. Moreover, the increase in the variance of ln(TFPR) is larger for the corrected ln(TFPR) than for the measured ln(TFPR), indicating that, if we could fully correct for measurement error, the decline in allocative efficiency could appear even more substantial. Unfortunately, the final step in Bils et al. (2021)'s procedure—calculating the corrected misallocation—does not replicate successfully with our Canadian administrative dataset. The error term to be added to the firm/plant-level distribution, which should be constructed as lognormally distributed with a variance of $-(\mathrm{Cov}[\ln(\mathrm{TFPR}), \ln(\hat{\beta}_k)] - \mathrm{Var}(\ln(\hat{\beta}_k)))$, fails to yield a positive value for this variance. This issue arises in both clean and balanced samples, regardless of the trimming criteria or variations in the construction of revenue and input data for the regression, preventing us from constructing the error term and completing the procedure. One possible reason why the method failed is that our data does not contain information on intermediary input. Therefore, when we replicate the method of Bils et al. (2021), our input measure likely misrepresents the true input cost. Even with the incomplete replication, however, our examinations from the earlier steps reassure us that additive measurement error is unlikely to affect the robustness of our conclusions regarding changes in misallocation.

#### 4.6.2 Data Cleaning and Trimming

Another source of measurement error stems from the data processing and cleaning process. Rotemberg and White (2021) show that data cleaning significantly affects measured misallocation. They find that the U.S. Census's editing and imputation procedures lead to about 80% of plants in the Census of Manufactures having values in the final dataset that differ from the raw data. This cleaning process substantially reduces the mass in the tails, cutting the standard deviation of TFPR by half. Rotemberg and White (2021)'s calculations indicate that using the original, uncleaned data significantly increases measured misallocation, with differences potentially reaching a factor of one thousand. Unfortunately, the literature provides little insight into how data imputation affects the volatility of measured allocative efficiency, which is more relevant to our study.

Our T2-LEAP data, which links the T2 Corporate Income Tax form with the Longitudinal Employment Analysis Program (LEAP), is administrative in nature. According to Statistics Canada, imputation is not applied to the LEAP data, which contains the payroll information of firms. For the T2 Corporation Income Tax Return, several checks are performed on the data to verify internal consistency and identify extreme values. Imputation is performed to provide values for the units with total nonresponse. The rationale behind this imputation is that even if all tax data have not been received, the T2 database must ensure tax data are available for each unit on annual business surveys frames. These imputed values get overwritten with reported data when they are eventually received by the Agency.[6] According to a technical report by Statistics Canada (2003), while imputation for total non-response was necessary for approximately 36% of enterprises in 2002, the overall impact of imputation on operating revenues across all industries was around 23%. The impact is notably lower than the level of imputation applied by the U.S. Census."

In our data, common reporting errors, such as missing values, are frequently present. We find that over 50% of firm-year entries in the raw data have missing or negative values in at least one of the following variables: payroll, revenue, tangible assets, working capital, or NAICS code. Rather than editing or imputing the data, we exclude these observations. Please note that some of the apparent errors are not actual errors, but rather result from non-operating firms. These may include, but are not limited to, shell companies, holding companies, and special purpose entities designed for non-operational purposes such as finance and tax optimization. Since these companies are not suitable for studying firm productivity, they are excluded from our sample.

Trimming, which drops outliers from the sample, is a popular approach to data cleaning. In Hsieh and Klenow (2009), 1 % the tails of plant productivity and distortions are trimmed in each year, while Bils et al. (2021) trim the 1% tails of both measured TFPR and TFPQ deviations from the industry average in each year. While trimming can increase the magnitude of measured allocative efficiency significantly, it could also drop data that is fairly reliable, creating mismeasurement errors (Rotemberg and White, 2021). In our analysis, we do not apply trimming of extreme values to our sample.[7]

## 5 Policy Discussion

Measuring the magnitude and consequences of resource misallocation, though in some ways may be somewhat abstract, provides important information for policy makers. It can further shed light on a potentially cause of Canada's relatively slower productivity growth in recent years and help direct efforts towards potential reform options. In this section, we discuss some of these issues.

Of course, some of what we measure as productivity might reflect inefficient allocations that result from transitory developments that are less relevant for policymakers. A rising trend in capital misallocation, for example, might result from an increasing dispersion of firm-specific shocks combined with adjustment costs that slow the pace of reallocation. It takes time, after all, for capital stocks to both depreciate in a sector receiving a negative shock and accumulate in a sector receiving a positive one. To investigate this further, we calculated the variance in both input growth and output growth over time. We find that these showed volatility only around a relatively constant level, with no clear upward trend. Additionally, we constructed the average firm marginal revenue product of capital over a series of five-year periods to assess whether dispersion in the

[6] See Brisebois et al. (2011) for methodology used in processing tax data at Statistics Canada.

[7] We are unable to provide a comparison between trimmed and untrimmed results, as done in Rotemberg and White (2021), because Statistics Canada requires both the trimmed sample and the residuals from trimming—i.e., the 1% tails of each industry year, based on the trimming method—to pass confidentiality tests. In our case, the residuals failed to pass these tests.

"permanent" component of firm MRPK has increased over time. We find the log of average MRPK declined consistently across the periods 2001–2005, 2006–2010, and 2011–2015 for both the clean and balanced samples. While not conclusive, this suggests a combination of increasing dispersion of shocks and adjustment costs is not a key driver of our results.

Turning to more policy relevant issues, two important contributors to lower aggregate productivity growth between 2001 and 2015 was rising misallocation of labour and capital between sectors, especially in the capital market. Policy makers have for many years struggled with issues around inter-provincial trade barriers, the efficiency of Canada's banking system, harmonising of securities, and finance regulations across provinces. To the extent that there is significant industry variation across provinces, such barriers may inhibit the efficient allocation of capital across sectors. The volatility of Canada's economy (in particular the oil producing regions) may also matter. As capital is long lived, misallocation may result from the slow speed with which it responds to economic shocks. It may reallocate in an efficient manner in time, but cannot do so instantaneously. As volatility increases, this may become a more important consideration. Our research, however, provides little guidance to policy makers around the specific sources of capital wedges or around specific policy interventions that may alleviate the aggregate costs. Our results do, however, point to the potentially significant macroeconomic implications of distortions in Canada's capital market worsening since 2001. Future research exploring this area would be valuable.

Our results also speak to the potentially important effect of migration costs and labour mobility restrictions. While we do not explicitly measure inter-provincial barriers to labour mobility, our between-sector distortions will partially reflect such barriers. Provinces, after all, differ in the composition of economic activities. Some industries are larger in some regions than others – such as oil and gas in Alberta or finance in Ontario. Policy makers in Canada have long sought to minimize the cost of moving between provinces, though much remains to be done. British Columbia and Alberta, for example, implemented the Trade, Investment, and Labour Mobility Agreement (TILMA) in 2009 in an effort to mutually recognize or harmonize rules and regulations that make trade and migration difficult. The recent Canadian Free Trade Agreement (CFTA), which began in 2017, hopes to make progress between all provinces. Efforts to ensure credentials and occupational licenses are easily transferable across provincial boundaries may help lower labour market distortions and therefore help increase overall productivity.

# 6 Conclusion

The efficiency with which labour and capital are allocated across firms matters for an economy's aggregate productivity. This is well known. But how potential resource misallocation changes over time and how that may contribute to productivity growth in Canada has not previously been investigated. Our analysis exploits access to uniquely detailed firm-level administrative data to measure the magnitude and consequences of labour and capital misallocation across firms not only in levels but also in changes between 2001 and 2015. We find, as most studies in this area do, that misallocation has a significant negative effect on overall productivity. But, importantly, we also find that misallocation has consistently worsened in Canada since 2001. Our estimates suggest this decreasing allocative efficiency of labour and capital markets lowered Canada's aggregate productivity significantly. Worsening between sector allocations alone lowered aggregate productivity by 9 percent over this period. This is meaningful not only because it is large, but also because it accounts for approximately half of the declining relative productivity of Canada compared to the United States. The country's lagging productivity performance is an ongoing area of concern for policy makers and our results point to a potentially important contributing factor. There remains much work to be done to uncover the underlying causes of misallocation and to deploy richer methods that map a full structural model to the firm-level data if it becomes more widely available. Future research in these areas would be valuable.

## References


Michael-John Almon and Jianmin Tang. Industrial structural change and the post-2000 output and productivity growth slowdown: A canada-us comparison. *International Productivity Monitor*, (22), 2011.

John R Baldwin, Luke Rispoli, and Danny Leung. *Canada-United States labour productivity gap across firm size classes*. Statistics Canada, Economic Analysis Division, 2014.

Abhijit V. Banerjee and Esther Duflo. Growth Theory through the Lens of Development Economics. In Philippe Aghion and Steven Durlauf, editors, *Handbook of Economic Growth*, volume 1 of *Handbook of Economic Growth*, chapter 7, pages 473–552. Elsevier, 2005.

Eric Bartelsman, John Haltiwanger, and Stefano Scarpetta. Cross-country differences in productivity: The role of allocation and selection. *American Economic Review*, 103(1):305–34, February 2013.

Mark Bils, Peter Klenow, and Cian Ruane. Misallocation or mismeasurement? *NBER Working Paper*, (26711), January 2020.

Mark Bils, Peter J Klenow, and Cian Ruane. Misallocation or mismeasurement? *Journal of Monetary Economics*, 124:S39–S56, 2021.

Loren Brandt, Trevor Tombe, and Xiaodong Zhu. Factor market distortions across time, space and sectors in china. *Review of Economic Dynamics*, 16(1):39 – 58, 2013.

François Brisebois, R. Laroche, and R Manriquez. Processing methodology of tax data at statistics canada. CONFERENCE OF EUROPEAN STATISTICIANS, 2011. URL https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2011/wp.7.e.pdf.

Sara Calligaris. Misallocation and total factor productivity in i taly: Evidence from firm-level data. *Labour*, 29(4):367–393, 2015.

Kaiji Chen and Alfonso Irarrazabal. The role of allocative efficiency in a decade of recovery. *Review of Economic Dynamics*, 18(3):523–550, 2015.

Daisuke Fujii and Yoshio Nozawa. Misallocation of capital during japan's lost two decades. *Development Bank of Japan Working Paper*, 1304, 2013.

Gita Gopinath, Şebnem Kalemli-Özcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez. Capital allocation and productivity in south europe. *The Quarterly Journal of Economics*, 132(4): 1915–1967, 2017.

Chang-Tai Hsieh and Peter J. Klenow. Misallocation and Manufacturing TFP in China and India*. *The Quarterly Journal of Economics*, 124(4):1403–1448, 11 2009.

Charles Jones. Misallocation, economic growth, and input-output economics. *NBER Working Paper*, (16742), 2011.

Danny Leung, Cesaire Meh, Yaz Terajima, et al. Productivity in canada: Does firm size matter? *Bank of Canada Review*, 2008(Autumn):7–16, 2008.

Ezra Oberfield. Productivity and misallocation during a crisis: Evidence from the chilean crisis of 1982. *Review of Economic Dynamics*, 16(1):100–119, 2013.

Ashantha Ranasinghe. Innovation, firm size and the canada-us productivity gap. *Journal of Economic Dynamics and Control*, 85:46–58, 2017.

Ricardo Reis. The portuguese slump and crash and the euro crisis. *Brookings Papers on Economic Activity*, page 172, 2013.

Diego Restuccia. Misallocation and aggregate productivity across time and space. *Canadian Journal of Economics*, 52(1):5–32, 2019.

Diego Restuccia and Richard Rogerson. Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics*, 11(4):707 – 720, 2008.

Martin Rotemberg and T Kirk White. Plant-to-table (s and figures): Processed manufacturing data and measured misallocation. *Working Paper.*, 2021.

Guido Sandleris and Mark LJ Wright. The costs of financial crises: Resource misallocation, productivity, and welfare in the 2001 argentine crisis. *The Scandinavian Journal of Economics*, 116(1): 87–127, 2014.

Statistics Canada. Data Quality, Concepts and Methodology. https://www.statcan.gc.ca/en/statistical-programs/document/2510_D1_T2_V5-eng.pdf, 2003.

Jianmin Tang. Industrial structure change and the widening canada–us labor productivity gap in the post-2000 period. *Industrial and Corporate Change*, 26(2):259–278, 2017.

Trevor Tombe and Xiaodong Zhu. Trade, Migration and Productivity: A Quantitative Analysis of China. *American Economic Review*, 109(5):1843–1872, 2019.

Nicolas L Ziebarth. Are china and india backward? evidence from the 19th century us census of manufactures. *Review of Economic Dynamics*, 16(1):86–99, 2013.

# A Between and Within Sector Misallocation

This section shows how we identify the between-sector and within-sector distortions using firm-level data. We start with the aggregate TFP

$$A \equiv \frac{Y}{L^{\bar{\alpha}} K^{1-\bar{\alpha}}}, \tag{30}$$

where $\bar{\alpha} = \sum_{i=1}^{N_J} \alpha_j \beta_j$. Given that $Y = \prod_{j=1}^{J} Y_j^{\beta_j} = \prod_{j=1}^{J} \left( A_j L_j^{\alpha_j} K_j^{1-\alpha_j} \right)^{\beta_j}$, we have

$$A = \prod_{j=1}^{J} \left( A_j l_j^{\alpha} k_j^{1-\alpha_j} \right)^{\beta_j}, \tag{31}$$

where $l_j$ and $k_j$ are sector $j$'s the labour share and capital share as in equations (16) and (15). The aggregate TFP distortion can be written as

$$\begin{aligned} \hat{A} &= \frac{\prod_{j=1}^{J} \left( A_j l_j^{\alpha} k_j^{1-\alpha_j} \right)^{\beta_j}}{\prod_{j=1}^{J} \left( A_j^* l_j^{*\alpha} k_j^{*1-\alpha_j} \right)^{\beta_j}} \\ &= \left( \frac{\prod_{j=1}^{J} A_j^{\beta_j}}{\prod_{j=1}^{J} A_j^{*\beta_j}} \right) \left( \frac{\prod_{j=1}^{J} l_j^{\alpha} k_j^{1-\alpha_j}}{\prod_{j=1}^{J} l_j^{*\alpha} k_j^{*1-\alpha_j}} \right)^{\beta_j}. \end{aligned} \tag{32}$$

The component in the first parentheses presents the within-sector distortion, where $A_j$ is the sector $j$'s TFP as defined in (25). Below is its discrete version equivalence

$$A_j = \left[ \sum_{i=1}^{N_J} \left[ \phi_j(i) \left( \frac{\bar{\tau}_j^l}{\tau_j^l(i)} \right)^{\alpha_j} \left( \frac{\bar{\tau}_j^k}{\tau_j^k(i)} \right)^{1-\alpha_j} \right]^{\sigma-1} \right]^{\frac{1}{\sigma-1}}. \tag{33}$$

$A_j^*$ is the sector $j$'s productivity in the absence of misallocation as given in equation (26). The corresponding discrete version is

$$A_j^* = \left( \sum_{i=1}^{N_J} \left( \phi_j(i) \right)^{\sigma-1} \right)^{\frac{1}{\sigma-1}}. \tag{34}$$

For the component in the second parentheses, $k^*$ and $l^*$ are optimal capital and labour shares, such that

$$k^* = \frac{K_j^*}{K^*} = \frac{(1-\alpha_j)\beta_j}{\sum_{i=1}^{J}(1-\alpha_i)\beta_i}, \tag{35}$$

$$l^* = \frac{L_j^*}{L^*} = \frac{\alpha_j \beta_j}{\sum_{i=1}^J \alpha_i \beta_i}. \tag{36}$$

The component in the second parentheses in equation (32) can be written as

$$\begin{aligned}
\prod_{j=1}^J \left( \frac{l_j^\alpha k_j^{1-\alpha_j}}{l_j^{*\alpha} k_j^{*1-\alpha_j}} \right)^{\beta_j} &= \prod_{j=1}^J \left[ \left( \frac{\frac{\alpha_j \beta_j / \bar{\tau}_j^l}{\sum_{i=1}^J \alpha_i \beta_i / \bar{\tau}_i^l}}{\frac{\alpha_j \beta_j}{\sum_{j=1}^J \alpha_j \beta_j}} \right)^{\alpha_j} \left( \frac{\frac{(1-\alpha_j)\beta_j / \bar{\tau}_j^k}{\sum_{i=1}^J (1-\alpha_i)\beta_i / \bar{\tau}_i^k}}{\frac{(1-\alpha_j)\beta_j}{\sum_{j=1}^J (1-\alpha_j)\beta_j}} \right)^{1-\alpha_j} \right]^{\beta_j} \\
&= \prod_{j=1}^J \left[ \left( \frac{\sum_{i=1}^J \alpha_i \beta_i}{\sum_{i=1}^J \alpha_i \beta_i / \bar{\tau}_i^l} \frac{1}{\bar{\tau}_j^l} \right)^{\alpha_j} \left( \frac{\sum_{i=1}^J (1-\alpha_i)\beta_i}{\sum_{i=1}^J (1-\alpha_i)\beta_i / \bar{\tau}_i^k} \frac{1}{\bar{\tau}_j^k} \right)^{1-\alpha_j} \right]^{\beta_j},
\end{aligned} \tag{37}$$

where $\bar{\tau}^l$ and $\bar{\tau}^k$ are defined as

$$\bar{\tau}^l \equiv \frac{\sum_{i=1}^J \alpha_i \beta_i}{\sum_{i=1}^J \alpha_i \beta_i / \bar{\tau}_i^l}, \tag{38}$$

$$\bar{\tau}^k \equiv \frac{\sum_{i=1}^J (1-\alpha_i)\beta_i}{\sum_{i=1}^J (1-\alpha_i)\beta_i / \bar{\tau}_i^k}. \tag{39}$$

Therefore, the component in the second parentheses in equation (32) is the between-sector distortion

$$\hat{A}^{\text{between sector}} = \prod_{j=1}^J \left[ \left( \frac{\bar{\tau}^l}{\bar{\tau}_j^l} \right)^{\alpha_j} \left( \frac{\bar{\tau}^k}{\bar{\tau}_j^k} \right)^{1-\alpha_j} \right]^{\beta_j}. \tag{40}$$

# B T2-LEAP Data

This section provides detailed information on the T2-LEAP data, sourced from Statistics Canada. T2-Longitudinal Employment Analysis Program (T2-LEAP) is an enterprise-level database that contains key information on firm entry and exit, demographics, finances, and performance. T2-LEAP is created by linking data from the T2 Corporate Income Tax form to the Longitudinal Employment Analysis Program (LEAP) dataset.

The Longitudinal Employment Analysis Program (LEAP) is an administrative databank maintained by the Economic Analysis Division (EAD) at Statistics Canada. It contains annual employment information for every employer business in Canada. LEAP is constructed from three main sources: T4 administrative data received from the Canada Revenue Agency, the Statistics Canada Business Register and Statistics Canada's Survey of Employment, Payrolls and Hours (SEPH). The target population of LEAP is every employer in Canada. The target population of LEAP includes all employers in Canada, covering both incorporated and unincorporated businesses that issue at least one T4 slip in a given calendar year, with the exception of self-employed individuals or partnerships that do not draw salaries. According to Statistics Canada, imputation is not applied to the LEAP database.

To create T2-LEAP, the LEAP file was linked to the Corporate Income Tax File (T2). The T2 file includes all incorporated firms that file a T2 tax return with the Canada Revenue Agency (CRA). The T2 file provides data on, among other things, sales, gross profits, equity and assets for all incorporated firms in Canada. All resident corporations, including non-profit organizations, tax-exempt corporations, and inactive corporations, have to file a T2 Corporation Income Tax Return for every tax year, with the sole exception of registered charities. While the T2-LEAP dataset covers all incorporated private-sector employers in Canada, public sector enterprises, classified under 2-digit NAICS codes 61, 62, and 91, are excluded.

For T2 Corporate Income Tax form, several checks are performed on the data to verify internal consistency and identify extreme values. Imputation for complete non-response is performed by 2 general methods. The preferred and most common method makes use of historical information about the non-responding unit and current trends in principal characteristics of similar units. When historical information is not available, such as in the case of births, a donor of similar size and industry is substituted for the missing unit.

# C Data Sampling and Subsetting

From the raw data to our clean sample, an average of 57% of observations are dropped each year. In the raw data, over one-third of year-firm observations lack a positive payroll value, while another third lack a positive value for capital, calculated as total tangible assets minus working capital. Applying these two criteria together results in a drop of 56% of total year-firm observations. Additionally, we exclude year-firm observations that lack a positive revenue value, lack an identifiable NAICS code, have missing provincial information, or are located in the territories (Yukon, Northwest Territories, and Nunavut). These three criteria contribute an additional one percent drop from the raw data to the clean sample.

Table 4 below compares the size distribution of firms in the clean and balanced samples. Since the balanced sample includes only firms that were continuously present over the 15-year period from 2001 to 2015, it contains a higher proportion of medium and large firms relative to the clean sample.

Table 4: Average Size Distribution of Firms in the Clean and Balanced Samples

| | Average labour units: 0-9 | Average labour units: 10-100 | Average labour units: > 100 |
|---|---|---|---|
| Clean | 81.2% | 16.8% | 2.0% |
| Balanced | 66.39% | 19.7% | 4.3% |

Reports the firm size distribution of the clean and balanced samples based on national average labour units (NALUS), averaged across 2001-2015.