IARIW 2021

Monday 23 – Friday 27 August

Statistisk sentralbyrå
Statistics Norway

Frontier Firms, Inefficiency and Productivity Dynamics

Shipei Zeng

(University of New South Wales and Shenzhen Research Institute of Big Data)

Paper prepared for the 36th IARIW Virtual General Conference
August 23-27, 2021
Session 24: Productivity
Time: Thursday, August 26, 2021 [16:30-18:30 CEST]

Frontier firms, inefficiency and productivity dynamics

Shipei Zeng

University of New South Wales and Shenzhen Research Institute of Big Data

August 11, 2021

Abstract

Productivity dynamics occur when firms enter and exit a market. Contributions from firms to industry productivity can be decomposed into effects from entrants, exits and incumbents. As opposed to productivity dynamics, productivity can also be decomposed into explanatory factors regarding efficiency and technical progress. These two patterns of decomposition provide different perspectives about the driving components of productivity. I propose a framework that merges them and produces a cross dimension. Industry productivity can not only be allocated as firm contributions, but also its explanatory factors can be illustrated analogously. It is developed by specifying firms that are on production frontiers, measuring the deviation from frontiers, and integrating explanatory factors with firm dynamics. A difference-in-differences approach is proposed that validates the firm dynamics from the counterfactual perspective. As an empirical exercise, the framework is applied to Australian firm-level data and reveals the dominant contribution of incumbent firms to industry productivity and industry efficiency.

Keywords: productivity decomposition; firm dynamics; production frontiers; firm observations

1 Introduction

Firm dynamics can promote the development of the economy. New firms enter industries with innovative perspectives and experimental practices. They also serve as an essential source that intensifies competition. With the inflow of new firms, updated techniques, especially digital services, are frequently generated to meet in-demand requirements from customers. Jobs may be created due to the impact of new firms on industries and hence contribute to economic growth over the long term. This is a creative destruction process where competition is incited and less capable firms are replaced by competitive entrants. Old-fashioned firms that lack the ability to adapt are forced to exit the industry. Industry performance is improved and aggregate productivity is enhanced due to the innovation and competition from dynamic turnover.

Economics studies have established the contributing components of firm dynamics. Firm dynamics reflect the input and output reallocation over different firms. Specifically, new firms effectively receive resources from disappearing firms and take up part of the corresponding market shares. This also happens for continuing firms when declining firms operate in a similar way to exiting firms, and growing firms operate as entering firms. Productivity change of incumbent firms, entering firms and exiting firms is an driving factor of aggregate productivity growth, though it is not the only reason for the aggregate change over periods. The market shares of incumbents, entrants and exiters also affect aggregate productivity when the distribution of firm-level productivity is fixed. Aggregate productivity change is driven by both the individual firm productivity and the industry composition.

This paper seeks to decompose productivity change through firm dynamics, and proposes an economic interpretation of the within, between, entry and exit effects. Two different patterns of decomposition are mentioned for productivity. To clarify them, the decomposition involving explanatory factors such as efficiency and technical progress is termed "productivity decomposition" while the decomposition involving firm contributions to aggregate productivity from entry and exit effects is termed "productivity dynamics".

The fundamental framework of firm dynamics was initiated by Baily et al. (1992), which decomposed productivity differences into the within, between, entry and exit effects. The within effect and the between effect measure productivity growth and market share changes of continuing firms while the entry effect and the exit effect highlight contributions from entering and exiting firms. Griliches and Regev (1995) revised this basic framework and added weighting counterparts for the balancing terms. An average industry productivity was also introduced as a reference productivity so that productivity growth of incumbents, entrants and exiters can be compared with a common benchmark. Following the purpose of setting up benchmarks, Baldwin and Gu (2006) proposed another productivity benchmark, that is, using the aggregate productivity of exiting firms as a reference productivity. Diewert and Fox (2010) argued that a reasonable reference is the aggregate productivity of incumbents. This benchmark allows for productivity comparisons in separate periods since entrants are compared with incumbents in the current period while exiters are compared with incumbents in the base period. All these decomposition methods follow the framework of Baily et al. (1992) where the within effect, the between effect, the entry effect and the exit effect are displayed. In addition, Foster et al. (2001) constructed a cross term that is productivity growth multiplied by market share changes. Melitz and Polanec (2015) replaced the within effect and the between effect by a mean term and a covariance term. Their decomposition methods have expanded the boundary of firm dynamics frameworks.

The decomposition framework proposed by Baily et al. (1992) is used to specify firm contributions to productivity differences over periods. It is a framework for productivity dynamics, but not for productivity decomposition into explanatory factors such as efficiency and technical progress. An intuitive impulse is to decompose firm productivity into explanatory factors and apply these factors to firm dynamics. By combining productivity decomposition and firm dynamics, contributing components of productivity differences will be found to include more than the productivity change of incumbents, entrants and exiters. Explanatory factors of productivity for incumbents, entrants and exiters can also be specified in the same approach as productivity dynamics. However, such a combination is not simply a stacking of productivity decompositions over firm dynamics. Although the existing literature has developed many methods of productivity decomposition, they cannot be directly applied to firm dynamics because productivity decomposition methods typically adopt productivity indexes that are incompatible with entering firms and exiting firms.

The economic interpretation is another concern for firm dynamics because the causal specification of contributing components to productivity growth is still insufficient. The framework of separating firm dynamics into the within, between, entry and exit effects is mathematically correct, but there remains a paucity of interpretation on whether these effects truly measure the impact of incumbents, entrants and exiters. For example, the specification of the entry effect will be unsatisfactory if it is only based on a straightforward algebra operation. The algebra operation is mathematically feasible but may not necessarily measure the real impact, or causal effect of entering firms. The entry effect needs to be decided by comparing industry productivity with entering firms and industry productivity without entering firms. Current studies of firm dynamics provide insufficient support on this comparison and hence cannot reach a causal specification.

This paper attempts to contribute to the research of productivity growth and firm dynamics in the following aspects. First, I develop a new method of productivity decomposition to decompose firm-level productivity into explanatory factors. It is built upon the industry-level value added decomposition of Diewert and Fox (2018), but adapted to firm-level data by revising some major assumptions of the model. Second, this paper fills a gap in the literature by combining the productivity decomposition method with productivity dynamics. It decomposes productivity into efficiency and technical progress so that contributions of incumbents, entrants and exiters to explanatory factors of productivity can be specified separately. Third, this paper aims to enhance a causal understanding of productivity dynamics. An difference-in-differences approach is introduced to measure the causal effect of incumbents, entrants and exiters. I assume a pseudo market and compare it with the real market to specify the within, between, entry and exit effects. These effects derived from counterfactual inference are mathematically correct and analytically reasonable.

The overall structure of this paper is as follows. Section 2 clarifies key elements of measuring production output, labour and capital that are required for firm productivity estimation. Section 3 introduces a new method of firm-level productivity decomposition. A number of productivity dynamics methods are reviewed, followed by a difference-in-differences perspective that examines the causal specification. Productivity decomposition and productivity dynamics are combined to produce a cross dimension of decomposition. Firm-level evidence using the productivity decomposition and a cross dimension is demonstrated in Section 4. The final section concludes this paper by summarising the main points.

2 Firm productivity measurement

Prior to the analysis of firm dynamics, firm-level productivity measurement is required where three elementary factors need to be confirmed: production output,

Table 1: Key elements of measuring production output, labour and capital for firm productivity

References	Production output	Labour	Capital
Maré et al. (2017)	- Value of sales, purchases and stocks - Producer price indexes	- Number of FTE (full-time equivalent) workers	- Cost of capital services: depreciation, rental and leasing cost, user cost - Capital goods price indexes
Decker et al. (2017)	- Real gross revenue - BEA price deflators	- Number of workers	
Foster et al. (2017)	- Real gross revenue - BEA price deflators	- Number of workers	
Braguinsky et al. (2015)	- Physical output	- Number of factory operatives: excluding white-collar workers	- Number of capital equipment in operation
Riley et al. (2015)	- Real gross value added	- Aggregate employment	- Machinery and equipment capital stocks - Building and structure capital stocks - Investment deflators
Brandt et al. (2012)	- Real value added	- Employee compensation: wages, employee supplementary benefits, unemployment insurance, etc. - Total annual employment	- Capital stock - Brandt-Rawski deflators
Pierpaolo Parrotta (2012)	- Total sales - Intermediate costs	- Number of employees	- Capital stock (fixed assets)

Table 1: Key elements of measuring production output, labour and capital for firm productivity (continued)

References	Production output	Labour	Capital
Tian and Twite (2011)	- Operating revenue	- Cost of employees	- Net property, plant and equipment
Foster et al. (2008)	- Physical output	- Production-worker hours adjusted	- Capital stocks of equipment and structures - Sector-specific deflators
Breunig and Wong (2008)	- Value of sales - Change in inventories - Intermediate inputs and other operating expense	- Number of FTE workers	- Value of non-current assets - Leasing stock
Brynjolfsson and Hitt (2003)	- Deflated sales less deflated materials	- Number of employees - Labour and related expenses - Price index for total compensation	- Value of capital stock - Capital investment deflator
Delgado et al. (2002)	- Annual gross production of goods and services - Cost of intermediate inputs - Individual price indexes	- Number of effective yearly hours of work	- Capital stock of equipment
Scarpetta et al. (2002)	- Real value-added	- Total employment (or hours worked)	- Real capital stock
Konrad and Mangel (2000)	- Value of sales	- Number of workers	

labour and capital. Sometimes intermediate materials are included to specify gross output based productivity. The measurement of these elements has been a controversial issue as little agreement is reached on what should serve as appropriate indicators for firm-level performance. For example, either revenue productivity or physical productivity is reasonable depending on the specification of the output variable (Foster et al., 2008; Braguinsky et al., 2015). Due to the paucity of firm-level price information, the output measurement of a firm is generally accomplished by deflating the firm revenue with a common price index for the industry to which the firm belongs. The firm revenue deflated by price indexes provides a convenient method of measuring production output, but brings in a new challenge. For each individual firm, sales revenue is closely connected with product prices within the industry. The productivity that reflects efficiency and technical progress can be misleading if product prices are affected by the change of output or the change of market power. The significance of profit selection is then highlighted while the contribution of productivity performance is relatively neglected. This is when physical productivity is preferred because the measurement of physical quantities is not affected by product prices.

Production output indicators can be mainly classified into five types: value of sales, operating revenue, value added, value of production and physical output. (1) Value of sales. Konrad and Mangel (2000) adopted logarithmic sales per worker as a measure of productivity, indicating that the value of sales can represent the output quantity. The use of sales as production levels was followed by Breunig and Wong (2008), while the change in inventories, intermediate inputs and other operating expenses was specifically considered in their research. Brynjolfsson and Hitt (2003) and Maré et al. (2017) also accepted that gross output can be captured by the value of sales deflated by price indexes. (2) Operating revenue. This was selected by Tian and Twite (2011) to examine how corporate governance could affect firm productivity. Foster et al. (2017) and Decker et al. (2017) utilised the same dataset and chose gross revenue deflated by price indexes as the output measure. Foster et al. (2017) proposed that revenue productivity may be naturally endogenous but it is still acceptable due to the limitation on data access. (3) Value added. As the compensation for labour and capital, value added is frequently used as the production output in the field of productivity measurement, especially for the national economic accounts. Scarpetta et al. (2002) constructed logarithmic real value added to calculate total factor productivity. A level measure of gross value added was chosen by Riley et al. (2015) to include firms whose value added are zero and negative that a logarithmic measure would otherwise exclude. Brandt et al. (2012) used intermediate inputs such as items for resale and indirect taxes to flesh out real value added from deflated output. (4) Value of production. Delgado et al. (2002) investigated firm productivity using real production value of goods and services. Apart from labour and capital, the cost of intermediate inputs was also included in their production function. A comparison of gross output production function with value added production function has been carried out by Brandt et al. (2012), and a higher productivity growth was found in value added production function. (5) Physical output. It is likely that revenue productivity is correlated with product prices. Foster et al. (2008) directly used physical output to investigate firm productivity in industries so that the effect of product prices could be removed. Braguinsky et al. (2015) also highlighted the difference between physical productivity and revenue productivity, where the physical units of output were obtained and analysed in their research.

Researchers engaged in firm productivity research pay particular attention to labour productivity. Four common measures for labour input are the number of employees, the number of hours worked, the number of full time equivalent workers, and the cost of employees. (1) The number of employees. Similar terms include "the number of workers", "total employment" and "aggregate employment". They can be used interchangeably. The number of employees is one of the most popular measures of labour input because of its clarity and availability. It can be counted clearly as an exact number and is available in firm annual reports. The number of employees has been extensively adopted for the analysis of work-life programs (Konrad and Mangel, 2000), labour diversity (Pierpaolo Parrotta, 2012), firm dynamics (Scarpetta et al., 2002; Decker et al., 2017) and productivity computation (Brynjolfsson and Hitt, 2003; Riley et al., 2015; Foster et al., 2017). A sub-group of total employment can also be selected. Braguinsky et al. (2015) narrowed the definition of employee counts to the sample of factor operatives where white-collar workers are excluded. (2) The number of hours worked. The benefit of using hours worked is that this time measure captures the real working status by adding overtime hours to the total workload and deducting non-working hours from the total workload. Delgado et al. (2002) applied the number of effective yearly hours of work to the measurement of labour input. Foster et al. (2008) adjusted the production-worker hours with the payroll ratio of total compensation to the compensation for these production workers so that the performance of non-production workers is covered. (3) The number of full-time equivalent workers. The number of hours worked can be converted to the number of standard employees given the total hours of a full-time working employee. It leads to a standard measure of employee counts. The estimate of firm-level labour quantity in Maré et al. (2017) is an example of the number of full-time equivalent workers. Regarding part-time working individuals, a flexible option is Breunig and Wong (2008) that deflated part-time employees as 0.426 full-time equivalent persons. The deflation ratio can be determined on a case-by-case basis. (4) Cost of employees. The cost of labour is useful for constructing some indexes such as the Törnqvist index; Brandt et al. (2012) constructed labour value shares in Törnqvist indexes of labour input for productivity computation. Another use of employment cost can be seen in Tian and Twite (2011) that directly defined labour input as the cost of employees in their research about corporate productivity in Australia, though the labour input is not typically defined in this way.

The measurement of capital input is challenging for productivity researchers. The cost of capital services is typically treated as a proportion of the capital stock while the corresponding proportion is vague and difficult to confirm. Maré et al. (2017) categorized the cost of capital services into depreciation cost, rental and leasing cost, and the user cost. The capital input can be specified if each component of the cost of capital services is clearly measured. This poses difficulty in measuring these components separately. Alternatively, some research studies considered the measurement of capital stock rather than the cost of capital services. The growth rate of capital stock mostly results in the same productivity growth indexes as the cost of capital services does when the cost of capital services is a constant proportion of the capital stock. The value of capital stock can be obtained by the perpetual inventory method (Scarpetta et al., 2002; Delgado et al., 2002; Brandt et al., 2012; Riley et al., 2015). It provides an estimate of capital stock by setting up an initial value, accumulating purchased capital and subtracting the capital depreciation. The scope of capital stock varies depending on research purposes. Brynjolfsson and Hitt (2003) aggregated the amount of ordinary capital and computer capital as the capital stock. Breunig and Wong (2008) classified capital stock into non-current assets and leasing stock. Other options include deflated equipment and structures (Foster et al., 2008), net property, plant and equipment (Tian and Twite, 2011), fixed assets (Pierpaolo Parrotta, 2012), and the number of machines in operation (Braguinsky et al., 2015).

3 Dynamics decomposition

3.1 Productivity decomposition

I propose a new framework that decomposes firm productivity into explanatory factors, which is compatible with the decomposition for productivity dynamics. Within the new framework, productivity decomposition can be integrated with productivity dynamics, which means that not only can firm productivity be allocated to the entry, exit, between and within effects, but also the explanatory factors of productivity can be illustrated analogously.

Searching for frontier firms

Productivity decomposition is built on a potential maximum production function. To capture the potential maximum production, Diewert and Fox (2018) established a cost constrained value added function in a longitudinal structure. It searches over periods up to and including the current one to figure out the frontier observation that would potentially have the largest value added at the industry level. However, the longitudinal approximation to the cost constrained value added function cannot be directly applied to firm-level data. The potential maximum output value calculated with the historical performance of an individual firm does not determine the production frontier of the whole industry. Additionally, indexes of productivity explanatory factors in Diewert and Fox (2018) are derived from growth values but growth values of new firms and disappearing firms are not well defined.

The drivers of firm-level productivity change are not fully revealed in the Diewert-Fox method that focuses on the industry level performance. To better demystify the micro drivers of productivity change, I extend the productivity decomposition of Diewert and Fox (2018) to the context of firm dynamics. Some key changes have been made to generalise the productivity decomposition method so that it can be applied to firm-level data sets. First, frontier firms are confirmed by searching firms in periods up to and including the current period. This searching technique explores the production frontier of firms that constitute a group or an industry rather than the production frontier of an individual firm. Second, the concept of output includes more than value added. Gross output measures are available and even preferable for firm-level data. Value added is computed by subtracting intermediate materials from gross output, which may yield negative values at the firm level. Gross output measures avoid this problematic approach as the value of firm-level gross output is non-negative. Third, indexes are constructed by comparing all firms to a benchmark observation and rolling windows are adopted when new observations become available. Fourth, firms in one period are assumed to share the same input and output price levels. I include this assumption because in practice firm-level prices are almost always unavailable in data sets. But the decomposition can still work even with firm-level prices. With these key changes, a generalised productivity decomposition method is developed. It applies to firm-level data and is equivalent to the decomposition of Diewert and Fox (2018) when firm-level data is collapsed into industry-level data.

Consider the production possibilities set S^t in which (y, z) is the element denoting any feasible output and input in a pair, and then the cost constrained output value function is defined as:

$$R^t(p, w, x) = \max_{y,z}\{p \cdot y : (y, z) \in S^t; w \cdot z \leqslant w \cdot x\} \tag{1}$$

where p refers to the output price, w denotes the input price, and x is the input quantity. The cost constrained output value function captures the maximum output value subject to the budget constraint. For any observed input x^t, a feasible solution is the observed output y^t, though it is not necessarily the optimal solution. Therefore, the cost constrained output value function is greater than, or equal to, the observed output value $p^t \cdot y^t$.

The non-parametric approximation to $R^t(p, w, x)$ is based on a unit cost function that minimises the input value per output value given all elements in the production possibilities set S^t during period t:

$$c^t(w, p) = \min_s \left\{ \frac{w \cdot x_s}{p \cdot y_s} : (y_s, x_s) \in S^t \right\} \tag{2}$$

With a set of output and input bundles (y_n, x_n) among all N observations from period 1 to period t, the estimation of the unit cost function is:

$$\hat{c}^t(w, p) = \min_n \left\{ \frac{w \cdot x_n}{p \cdot y_n} : n = 1, \ldots, N \right\} \tag{3}$$

This estimation of $c^t(w, p)$ involves observed output and input bundles of all observations throughout the research periods. The frontier firm is searched over a broader scope than the estimation by a direct application of Diewert and Fox (2018) where output and input bundles of only an individual firm could be searched. It facilitates productivity comparisons between firm-level observations.

Suppose production follows constant returns to scale, that is, $(\lambda y_s, \lambda x_s) \in S^t$ for $\lambda > 0$. The cost constrained output value function can be rewritten as:

$$\begin{aligned} R^t(p, w, x) &= \max_{y,z} \{p \cdot y : w \cdot z \leqslant w \cdot x; (y, z) \in S^t\} \\ &= \max_{\lambda} \{p \cdot \lambda y_s : w \cdot \lambda x_s \leqslant w \cdot x; \lambda \geqslant 0\} \\ &= \max_s \left\{ p \cdot y_s \frac{w \cdot x}{w \cdot x_s} : (y_s, x_s) \in S^t \right\} \\ &= \frac{w \cdot x}{c^t(w, p)} \end{aligned} \tag{4}$$

Based on the estimate of the unit cost function, the cost constrained output value function is estimated as:

$$\hat{R}^t(p, w, x) = \frac{w \cdot x}{\hat{c}^t(w, p)} \tag{5}$$

Explanatory factors of output ratios

The cost constrained output value function is employed to divide productivity into separate components. I conduct the productivity decomposition by comparing firms in two consecutive periods: firm i in period t and firm j in period $t-1$. Firms in the same period are assumed to share the same input price level and the same output price level. With firm j in period $t-1$ as the base unit, the output value ratio for firm i in period t can be decomposed into the following explanatory factors: output price inflation (deflation), input quantity ratios, input mix measures, returns to scale, output value efficiency ratios, and technical progress. These explanatory factors will be illustrated separately.

Output price inflation (deflation) is measured by the ratio of cost constrained output value functions with different output price levels p^t and p^{t-1}:

$$\alpha(p^{t-1}, p^t, w, x, s) = \frac{R^s(p^t, w, x)}{R^s(p^{t-1}, w, x)} \tag{6}$$

w is the input price level and x is the input quantity. $R^s(p, w, x)$ is involved with a family of indexes for α. A Laspeyres-type output price index α_L^t is produced when $R^{t-1}(p^t, w^{t-1}, x_j^{t-1})$ is divided by $R^{t-1}(p^{t-1}, w^{t-1}, x_j^{t-1})$, and a Paasche-type output price index α_P^t is produced when $R^t(p^t, w^t, x_i^t)$ is divided by $R^t(p^{t-1}, w^t, x_i^t)$. These names of output price indexes are consistent with a typical Laspeyres index that takes a fixed basket of quantities for the base period, and also consistent with a Paasche index that takes a fixed basket of quantities for the current period. The geometric mean of α_L^t and α_P^t generates an overall output price index α^t that is a

Fisher-type output price index. A family of price indexes will facilitate the productivity decomposition, which will be revealed later. The definition of α implies that output price inflation (deflation) is identical for all firms in the same period.

The input quantity ratio is computed by taking the ratio of input quantities while using input prices as weights:

$$\beta(x_j^{t-1}, x_i^t, w) = \frac{w \cdot x_i^t}{w \cdot x_j^{t-1}} \tag{7}$$

x_i^t is the input quantity of firm i in period t, and x_j^{t-1} is the input quantity of firm j in period $t-1$. The input price level w determines the index type of β. A Laspeyres input quantity index β_L^t is produced when $w^{t-1} \cdot x_i^t$ is divided by $w^{t-1} \cdot x_j^{t-1}$, and a Paasche input quantity index β_P^t is produced when $w^t \cdot x_i^t$ is divided by $w^t \cdot x_j^{t-1}$. The geometric mean of β_L^t and β_P^t generates an overall input quantity index β^t that is a Fisher input quantity index.

Input mix measures are also centred on input series. Input mix indexes capture the effect of input price levels w^t and w^{t-1} on cost constrained output value functions:

$$\gamma(w^{t-1}, w^t, p, x, s) = \frac{R^s(p, w^t, x)}{R^s(p, w^{t-1}, x)} \tag{8}$$

p is the output price level and x is the input quantity. $R^s(p, w, x)$ is involved with a family of indexes for γ. A Laspeyres-type input mix index γ_{LPP}^t is produced when $R^t(p^{t-1}, w^t, x_i^t)$ is divided by $R^t(p^{t-1}, w^{t-1}, x_i^t)$, and a Paasche-type input mix index γ_{PLL}^t is produced when $R^{t-1}(p^t, w^t, x_j^{t-1})$ is divided by $R^{t-1}(p^t, w^{t-1}, x_j^{t-1})$. The geometric mean of γ_{LPP}^t and γ_{PLL}^t generates an overall input mix index γ^t. γ is defined as the input mix index rather than the input price index. If w is a one-dimensional vector, γ would always be unity and not affected by w^t and w^{t-1} because $R^s(p, w, x)$ is homogeneous of degree 0 in the input price w. Therefore, γ should be explained as the effect of inputs on $R^s(p, w, x)$ when w results in the change in relative proportions of inputs, or the mix of inputs.

Returns to scale are typically defined as the output ratio divided by the input ratio when technology is constant for input quantities x_i^t and x_j^{t-1}. The output ratio is denoted by the change of cost constrained output value functions and the input ratio noted by the change of input values. The index of returns to scale is expressed as:

$$\delta(x_j^{t-1}, x_i^t, p, w, s) = \frac{R^s(p, w, x_i^t)/R^s(p, w, x_j^{t-1})}{w \cdot x_i^t / w \cdot x_j^{t-1}} \tag{9}$$

p is the output price level and w is the input price level. A Laspeyres-type index of returns to scale δ_L^t is produced when $R^{t-1}(p^{t-1}, w^{t-1}, x_i^t)/R^{t-1}(p^{t-1}, w^{t-1}, x_j^{t-1})$ is divided by $w^{t-1} \cdot x_i^t / w^{t-1} \cdot x_j^{t-1}$, and a Paasche-type index of returns to scale δ_P^t is produced when $R^t(p^t, w^t, x_i^t)/R^t(p^t, w^t, x_j^{t-1})$ is divided by $w^t \cdot x_i^t / w^t \cdot x_j^{t-1}$. The cost constrained output value function implicitly demonstrates constant returns to scale. This can be also proved by expanding the expression of cost constrained output value functions and cancelling out equivalent terms. So the Laspeyres case δ_L^t and the Paasche case δ_P^t are both equal to unity. The overall index of returns to scale that is generated by the geometric mean of δ_L^t and δ_P^t is then also equal to unity.

Output value efficiency ratios are derived from output value efficiency. Since cost constrained output value functions measure maximum output values that can be feasibly obtained, observed output values are less than or equal to the values of cost constrained output value functions. This matches the concept of efficiency that compares the observed performance with the optimal performance. Output value efficiency for firm i in period t is defined as:

$$e_i^t = \frac{p^t \cdot y_i^t}{R^t(p^t, w^t, x_i^t)} \leqslant 1 \tag{10}$$

and output value efficiency for firm j in period $t-1$ is defined as:

$$e_j^{t-1} = \frac{p^{t-1} \cdot y_j^{t-1}}{R^{t-1}(p^{t-1}, w^{t-1}, x_j^{t-1})} \leqslant 1 \tag{11}$$

The output value efficiency ratio is the ratio of e_i^t and e_j^{t-1} that measures the efficiency difference between firms:

$$\varepsilon^t = \frac{e_i^t}{e_j^{t-1}} \tag{12}$$

Technical progress examines the difference of cost constrained output value functions in periods t and $t-1$:

$$\tau(t-1, t, p, w, x) = \frac{R^t(p, w, x)}{R^{t-1}(p, w, x)} \tag{13}$$

p is the output price level, w is the input price level, and x is the input quantity. A Laspeyres-type technical progress index τ_L^t is produced when $R^t(p^{t-1}, w^{t-1}, x_i^t)$ is divided by $R^{t-1}(p^{t-1}, w^{t-1}, x_i^t)$, and a Paasche-type technical progress index τ_P^t is produced when $R^t(p^t, w^t, x_j^{t-1})$ is divided by $R^{t-1}(p^t, w^t, x_j^{t-1})$. The geometric mean of τ_L^t and τ_P^t generates an overall technical progress index τ^t.

With explanatory factors defined above, the output value ratio can be straightforwardly decomposed as:

$$\begin{aligned} \frac{p^t \cdot y_i^t}{p^{t-1} \cdot y_j^{t-1}} &= \alpha_P^t \cdot \beta_L^t \cdot \gamma_{LPP}^t \cdot \delta_L^t \cdot \varepsilon^t \cdot \tau_L^t \\ &= \alpha_L^t \cdot \beta_P^t \cdot \gamma_{PLL}^t \cdot \delta_P^t \cdot \varepsilon^t \cdot \tau_P^t \\ &= \alpha^t \cdot \beta^t \cdot \gamma^t \cdot \delta^t \cdot \varepsilon^t \cdot \tau^t \\ &= \alpha^t \cdot \beta^t \cdot \gamma^t \cdot \varepsilon^t \cdot \tau^t \end{aligned} \tag{14}$$

Denote the ratio of output values as $g(y_j^{t-1}, y_i^t, p^{t-1}, p^t)$. It measures the relative output gains by comparing two observations, that is, firm i in period t and firm j in period $t-1$. The decomposition shows that the output ratio is directly explained by contributing components. The index of returns to scale is equal to unity and is omitted in the equation. With the productivity decomposition, it can be specified how output prices, input quantities, input prices, efficiency and technical progress contribute to the output ratio quantitatively. Since α^t is the overall output price index and β^t is the overall input quantity index, the decomposition of the output value ratio further leads to the decomposition of the productivity ratio. The productivity ratio is the output value ratio divided by the output price index and the input quantity index:

$$\begin{aligned} TFPR_{ij}^t &= \frac{p^t \cdot y_i^t / p^{t-1} \cdot y_j^{t-1}}{\alpha^t \cdot \beta^t} \\ &= \gamma^t \cdot \varepsilon^t \cdot \tau^t \end{aligned} \tag{15}$$

The decomposition of the output ratio and the productivity ratio reveals driving components of performance differences between two observations. I develop fixed base output indexes and fixed base productivity indexes from these value ratios by setting a benchmark and applying a rolling window. Suppose firm 1 in period 1 is the reference unit, and the window length is N. For firms in periods 1, 2, $\cdots$, N, fixed base output indexes are computed as:

$$\frac{p^t \cdot y_i^t}{p^1 \cdot y_1^1} = A^t \cdot B_i^t \cdot C_i^t \cdot E_i^t \cdot T^t \tag{16}$$

A^t, B_i^t, C_i^t, E_i^t and T^t are fixed base indexes of explanatory factors, that is, measures of α, β, γ, ε and τ when firm 1 in period 1 is the benchmark unit. A^t and T^t are independent of firm subscripts because firms in the same period share the same output price level and the same industry production frontier. Then fixed base productivity indexes are defined as fixed base output value indexes divided by fixed base output price indexes and fixed base input quantity indexes:

$$\begin{aligned} TFP_i^t &= \frac{p^t \cdot y_i^t}{p^1 \cdot y_1^1 \cdot A^t \cdot B_i^t} \\ &= C_i^t \cdot E_i^t \cdot T^t \end{aligned} \tag{17}$$

Rolling windows

When new periods are included, a rolling window allows the output decomposition and the productivity decomposition to update indexes without changing previous records. A new benchmark unit is set up, for example firm 1 in period 2 in the new window. Note firm 1 in period 2 is not necessarily the same firm as in period 1. It is denoted as the first firm in the new window by order. A new set of fixed base output indexes for firms in periods 2, 3, $\cdots$, $N+1$ are computed as:

$$\frac{p^t \cdot y_i^t}{p^2 \cdot y_1^2} = A^{t^*} \cdot B_i^{t^*} \cdot C_i^{t^*} \cdot E_i^{t^*} \cdot T^{t^*} \tag{18}$$

The next step is to consider how to link this new set of indexes with the set of indexes originated from the initial window. A linking observation is required to make indexes based on firm 1 in period 2 comparable with indexes based on firm 1 in period 1. In the context of rolling window methods, researchers have considered different linking observations. If the last but one observation in the new window serves as the linking observation, it is called a movement splice (Ivancic et al., 2011); if the first observation in the new window serves as the linking observation, it is called a window splice (Krsinich, 2016); if the middle observation serves as the linking observation, it is called a half splice (de Haan, 2015). Diewert and Fox (2017) generalised the splice method by taking a geometric mean of indexes from all possible linking observations, which produces a mean splice. This means that each observation has equal importance in determining the linking observation. I follow the mean splice method by Diewert and Fox (2017) and choose the geometric mean of indexes from common observations appearing in both windows as the linking point, that is, I construct the linked indexes by comparing the geometric mean of indexes of firms that the initial window and the new window have in common. For example, the linked efficiency indexes for firms in period $N+1$ can be expressed as:

$$E_i^{N+1} = E_i^{N+1^*} \cdot \left(\prod_{k,t} \frac{E_k^t}{E_k^{t^*}} \right)^{\frac{1}{M}} \tag{19}$$

$E_i^{N+1^*}$ is the efficiency index of firm i in period $N+1$ in the new window, E_k^t is the efficiency index of linked observations based on firm 1 in period 1, and $E_k^{t^*}$ is the efficiency index of linked observations in the new window. M is the number of linked observations (the sum of the number of firms from periods 2 to N in this case). The linked indexes for other explanatory factors are computed in the same way. With these linked indexes A^{N+1}, B_i^{N+1}, C_i^{N+1}, E_i^{N+1} and T^{N+1}, linked productivity indexes are inherently the linked output value indexes divided by linked output price indexes and linked input quantity indexes:

$$\begin{aligned} TFP_i^{N+1} &= \frac{p^{N+1} \cdot y_i^{N+1}}{p^1 \cdot y_1^1 \cdot (A^{N+1} \cdot B_i^{N+1})} \\ &= \frac{p^{N+1} \cdot y_i^{N+1}}{p^2 \cdot y_1^2 \cdot (A^{N+1^*} \cdot B_i^{N+1^*})} \cdot \frac{p^2 \cdot y_1^2}{p^1 \cdot y_1^1 \cdot \left(\prod_{k,t}(A^t \cdot B_k^t)/(A^{t^*} \cdot B_k^{t^*})\right)^{\frac{1}{M}}} \\ &= C_i^{N+1^*} \cdot E_i^{N+1^*} \cdot T^{N+1^*} \cdot \left(\prod_{k,t} \frac{(p^2 \cdot y_1^2/p^t \cdot y_k^t) \cdot (A^{t^*} \cdot B_k^{t^*})}{(p^1 \cdot y_1^1/p^t \cdot y_k^t) \cdot (A^t \cdot B_k^t)} \right)^{\frac{1}{M}} \\ &= C_i^{N+1^*} \cdot E_i^{N+1^*} \cdot T^{N+1^*} \cdot \left(\prod_{k,t} \frac{C_k^t \cdot E_k^t \cdot T^t}{C_k^{t^*} \cdot E_k^{t^*} \cdot T^{t^*}} \right)^{\frac{1}{M}} \\ &= C_i^{N+1} \cdot E_i^{N+1} \cdot T^{N+1} \end{aligned} \tag{20}$$

This demonstrates that the linked productivity indexes are the product of linked input mix indexes, linked efficiency indexes and linked technical progress indexes, which facilitates productivity decomposition. Given the window length at 2, this new approach of productivity decomposition is identical to the decomposition method in Diewert and Fox (2018) when firm-level data sets are aggregated as industry-level data as if there is only one "firm" in each period.

Multilateral indexes

Multilateral indexes are typically used in making international comparisons, and can also be used in time series contexts (Ivancic et al., 2011). They satisfy the circularity test from the axiomatic approach to index number choice. [1]

Fixed base indexes compare every observation with the base observation. When observations are compared consecutively and all results are multiplied, chained indexes are produced. Obviously, fixed base indexes are not necessarily equal to chained indexes, which is biased as chain drift. Chain drift can be resolved by a multilateral index method: the GEKS index proposed by Gini (1931), Eltetö and Köves (1964) and Szulc (1964). It takes the geometric mean of index ratios with rolling base observations. For example, the GEKS-type output value index between firm 1 in period 1 and firm i in period t is expressed as:

$$\Omega_i^t(R) = \prod_{k,s} \left(R(y_k^s, y_i^t, p^1, p^t) / R(y_k^s, y_1^1, p^1, p^t) \right)^{\frac{1}{N}} \tag{21}$$

y_k^s is the output of firm k in period s, and N is the number of observations. The base observation y_k^s rolls over N observations that are included in research periods. Using the circularity test and the time reversal test, it can be demonstrated that the GEKS index does not involve chain drift. It means the GEKS-type fixed base

[1] Although multilateral indexes can be constructed for the productivity decomposition in this paper, I choose not to use this approach in order to be consistent with the decomposition of Diewert and Fox (2018). It is also computationally infeasible to calculate multilateral indexes for big data sets when the number of firms becomes overly large in each window. Therefore, I show that multilateral indexes can be adopted theoretically but do not use them in the application of this paper.

index of output values is equal to the GEKS-type chained index of output values. Apply the GEKS method to the decomposition of output value indexes:

$$\Omega_i^t(R) = \Omega^t(A) \cdot \Omega_i^t(B) \cdot \Omega_i^t(C) \cdot \Omega_i^t(E) \cdot \Omega^t(T) \tag{22}$$

and to the decomposition of productivity indexes:

$$\Omega_i^t(TFP) = \Omega_i^t(C) \cdot \Omega_i^t(E) \cdot \Omega^t(T) \tag{23}$$

$\Omega_i^t(\cdot)$ of explanatory factors (A, B, C, E and T) and productivity indexes (TFP) follows the same pattern as $\Omega_i^t(R)$.

A weighted aggregation approach

Since firm productivity and its explanatory factors have been clarified, these factors can be further aggregated for an industry. The study of firm dynamics is centred on the contribution of continuing firms, new firms and disappearing firms to the whole industry. A simple way to construct an aggregate industry productivity is dividing the sum of output quantities by the sum of input quantities. But the explanatory factors of firm productivity are not clearly evaluated in such aggregation. The purpose of defining C_i^t, E_i^t and T^t is to decompose productivity into explanatory factors and meanwhile contributing components of firms are specified within the context of firm dynamics. A suitable aggregation is required to separate input mix indexes, efficiency indexes and technical progress from firm productivity.

I consider a firm weighted aggregation derived from Törnqvist (1936) indexes because the product of C_i^t, E_i^t and T^t can be transformed into the sum of $\ln C_i^t$, $\ln E_i^t$ and $\ln T^t$ by taking logarithmic forms. The original Törnqvist index is computed with averaged value shares over two periods for corresponding logarithmic components. Based on this principle, a firm weighted aggregation of productivity is similarly conducted by using the firm weight w_i^t and logarithmic productivity $\ln TFP_i^t$. The firm weight w_i^t for firm i in period t is defined as the firm input quantity weight in the industry:

$$w_i^t = \frac{x_i^t}{\sum_i x_i^t} \tag{24}$$

where x_i^t is the input quantity for firm i in industry l, $l = 1, \cdots, L$. The logarithmic industry productivity for industry l can be expressed as:

$$\begin{aligned} \ln TFP_l^t &= \ln \frac{\sum_i y_i^t}{\sum_i x_i^t} \\ &= \ln \sum_i \left(\frac{x_i^t}{\sum_i x_i^t} \cdot \frac{y_i^t}{x_i^t} \right) \\ &\approx \sum_i \frac{x_i^t}{\sum_i x_i^t} \ln \frac{y_i^t}{x_i^t} \\ &= \sum_i w_i^t \ln TFP_i^t \end{aligned} \tag{25}$$

TFP_i^t is the productivity for firm i in industry l, $l = 1, \cdots, L$. The approximation sign holds when a similar productivity is observed for each firm. It can be easily checked by assuming an identical productivity for each firm, that is, $y_i^t / x_i^t = \eta^t$. Then $\ln \sum w_i^t \eta^t = \ln \eta^t = \sum w_i^t \ln \eta^t$, indicating a linear operation on η^t. For firms with multiple inputs, I use the share of input quantity indexes as the firm weight, though the share of output values is frequently used in aggregate productivity measures. Since input quantity shares are adopted to link firm productivity to industry productivity for a one-dimensional input vector, it is reasonable to follow this principle and to take the share of input quantity indexes as the firm weight for a multi-dimensional input vector. The share of output values is easier to obtain, but it is inconsistent with the relationship between firm productivity and industry productivity.

3.2 Productivity dynamics

It is a necessary operation for firm dynamics to take the difference of industry productivity since continuing firms, new firms and disappearing firms need to be distinguished across periods.[2] Denote U as the set of all firms in period t and period $t-1$. I decompose the change of industry productivity into the changes of industry input mix indexes, industry efficiency indexes and industry technical progress indexes for industry l:

$$
\begin{aligned}
\Delta \ln TFP_l^{t,t-1} &= \sum_{i\in U} w_i^t \ln TFP_i^t - \sum_{i\in U} w_i^{t-1} \ln TFP_i^{t-1} \\
&= \sum_{i\in U} w_i^t (\ln C_i^t + \ln E_i^t + \ln T^t) \\
&- \sum_{i\in U} w_i^{t-1} (\ln C_i^{t-1} + \ln E_i^{t-1} + \ln T^{t-1}) \\
&= \sum_{i\in U} w_i^t \ln C_i^t - \sum_{i\in U} w_i^{t-1} \ln C_i^{t-1} \\
&+ \sum_{i\in U} w_i^t \ln E_i^t - \sum_{i\in U} w_i^{t-1} \ln E_i^{t-1} \\
&+ \sum_{i\in U} w_i^t \ln T^t - \sum_{i\in U} w_i^{t-1} \ln T^{t-1} \\
&= \Delta \ln C_l^{t,t-1} + \Delta \ln E_l^{t,t-1} + \Delta \ln T_l^{t,t-1}
\end{aligned} \tag{26}
$$

The changes of aggregate productivity, input mix indexes, and efficiency share a symmetric function structure. I generalise the decomposition of firm dynamics for these three items by taking $\Psi \in \{\ln TFP_l, \ln C_l, \ln E_l\}$. In the following review of productivity dynamics, the subset U_C is the set of continuing firms that exist in both periods, U_N is the set of new firms that enter in period t, and U_D is the set of disappearing firms that exit in period t

[2] A similar productivity among all firms in an industry is a relatively strong assumption considering the firm heterogeneity in real production. The difference of $\ln TFP_l^t$ may also be useful to remove the accompanying bias because the approximation bias can be possibly cancelled if the estimates of industry productivity over two periods are biased in the same direction.

BHC decomposition

Given U, U_C, U_N and U_D, a basic expansion of $\Delta\Psi^{t,t-1}$ is:

$$\begin{aligned}\Delta\Psi^{t,t-1} &= \sum_{i\in U} w_i^t \Psi_i^t - \sum_{i\in U} w_i^{t-1}\Psi_i^{t-1} \\ &= \underbrace{\sum_{i\in U_C} w_i^{t-1}(\Psi_i^t - \Psi_i^{t-1})}_{within} + \underbrace{\sum_{i\in U_C}(w_i^t - w_i^{t-1})\Psi_i^t}_{between} \\ &+ \underbrace{\sum_{i\in U_N} w_i^t\Psi_i^t}_{entry} - \underbrace{\sum_{i\in U_D} w_i^{t-1}\Psi_i^{t-1}}_{exit}\end{aligned} \tag{27}$$

This is the BHC decomposition proposed by Baily et al. (1992), which serves as a basic framework for productivity dynamics. Differences in industry productivity are decomposed into the within, between, entry and exit effects. The within term highlights the performance improvement in each firm separately across periods while the between term captures the change of industry shares of firms. These two terms are used to describe the contribution of continuing firms, or incumbents in both periods. For new firms and disappearing firms, the entry term and exit term in the equation are defined, reflecting a weighted aggregation of performance impact from non-continuing firms.

The basic expansion of $\Delta\Psi^{t,t-1}$ in the BHC decomposition needs to be revised due to two concerns. First, the within effect only includes the firm weight in period $t-1$ and the between effect only includes the performance in period t. The firm weight w_i^{t-1} in the within term and the performance Ψ_i^t in the between term can be replaced by w_i^t and Ψ_i^{t-1} simultaneously so that the equation still holds. An average value after the replacement allows factors in both periods to enter the equation. Second, a reference performance is absent in the BHC decomposition. The entry and exit effects are absolute measures of Ψ_i^t and Ψ_i^{t-1} while a relative measure of performance is essential for the comparison between continuing firms, new firms and disappearing firms.

GR decomposition

Griliches and Regev (1995) suggested using the average industry productivity as a reference for each firm's productivity performance. Denoting the reference performance $\bar{\Psi} = \frac{1}{2}(\Psi^t + \Psi^{t-1})$, the GR decomposition is conducted as:

$$\begin{aligned}
\Delta\Psi^{t,t-1} &= \sum_{i\in U} w_i^t(\Psi_i^t - \bar{\Psi}) - \sum_{i\in U} w_i^{t-1}(\Psi_i^{t-1} - \bar{\Psi}) \\
&= \underbrace{\sum_{i\in U_C} \frac{1}{2}(w_i^t + w_i^{t-1})(\Psi_i^t - \Psi_i^{t-1})}_{within} \\
&+ \underbrace{\sum_{i\in U_C} \frac{1}{2}(w_i^t - w_i^{t-1})(\Psi_i^t + \Psi_i^{t-1} - 2\bar{\Psi})}_{between} \\
&+ \underbrace{\sum_{i\in U_N} w_i^t(\Psi_i^t - \bar{\Psi})}_{entry} - \underbrace{\sum_{i\in U_D} w_i^{t-1}(\Psi_i^{t-1} - \bar{\Psi})}_{exit}
\end{aligned} \tag{28}$$

The within effect measures productivity improvement of each firm weighted by an average market share and the between effect denotes the compositional shift weighted by an average productivity comparison. They include more information than components in the BHC decomposition since the base period and the current period are involved simultaneously. Additionally, the reference performance $\bar{\Psi}$ allows the entry effect and the exit effect to be more reasonable; in the BHC decomposition, the effect of new firms is always positive and the effect of disappearing firms is always negative. The GR decomposition indicates that the entry effect can be negative if it is below the reference productivity and the exit effect can be positive if it is below the reference productivity, which provides a more reasonable measurement of the relative productivity contribution.

FHK decomposition

Foster et al. (2001) considered incorporating the industry productivity in the base period as the reference performance. The FHK decomposition is implemented by using Ψ^{t-1} as the benchmark:

$$
\begin{aligned}
\Delta\Psi^{t,t-1} &= \sum_{i \in U} w_i^t(\Psi_i^t - \Psi^{t-1}) - \sum_{i \in U} w_i^{t-1}(\Psi_i^{t-1} - \Psi^{t-1}) \\
&= \underbrace{\sum_{i \in U_C} w_i^{t-1}(\Psi_i^t - \Psi_i^{t-1})}_{within} + \underbrace{\sum_{i \in U_C} (w_i^t - w_i^{t-1})(\Psi_i^{t-1} - \Psi^{t-1})}_{between} \\
&+ \underbrace{\sum_{i \in U_C} (w_i^t - w_i^{t-1})(\Psi_i^t - \Psi_i^{t-1})}_{cross} \\
&+ \underbrace{\sum_{i \in U_N} w_i^t(\Psi_i^t - \Psi^{t-1})}_{entry} - \underbrace{\sum_{i \in U_D} w_i^{t-1}(\Psi_i^{t-1} - \Psi^{t-1})}_{exit}
\end{aligned}
\tag{29}
$$

Apart from the within and between effects, a cross term is newly specified in the FHK decomposition. This covariance-type effect captures the product of productivity improvement and share reallocation. Foster et al. (2001) separated the cross effect for continuing firms because they attempted to split terms in a clear manner; the average firm weight across different periods in the GR decomposition is regarded as an unclear component for the within effect by Foster et al. (2001) because the firm weight in the current period is used and involves share reallocation. In return for the cross term, the average firm weight in the within effect is diminished, as well as the average measure of firm productivity in the between effect. The entry effect and the exit effect are the same components as with the GR decomposition except the change of benchmark productivity.

BG decomposition

If new firms emerge to replace disappearing firms, as Baldwin and Gu (2006) proposed, the reference productivity needs to be the aggregate productivity of disappearing firms. For simplicity, I take the transformation of the GR decomposition as an example. Denote Ψ_D as the productivity of the disappearing firm in period $t-1$, the GR decomposition revised by Baldwin and Gu (2006) is carried out as:

$$\begin{aligned}\Delta\Psi^{t,t-1} &= \sum_{i\in U} w_i^t(\Psi_i^t - \Psi_D) - \sum_{i\in U} w_i^{t-1}(\Psi_i^{t-1} - \Psi_D)\\ &= \underbrace{\sum_{i\in U_C} \frac{1}{2}(w_i^t + w_i^{t-1})(\Psi_i^t - \Psi_i^{t-1})}_{within}\\ &+ \underbrace{\sum_{i\in U_C} \frac{1}{2}(w_i^t - w_i^{t-1})(\Psi_i^t + \Psi_i^{t-1} - 2\Psi_D)}_{between}\\ &+ \underbrace{\sum_{i\in U_N} w_i^t(\Psi_i^t - \Psi_D)}_{entry} - \underbrace{\sum_{i\in U_D} w_i^{t-1}(\Psi_i^{t-1} - \Psi_D)}_{exit}\end{aligned} \tag{30}$$

The same reference productivity Ψ_D can be used to replace Ψ^{t-1} in the FHK decomposition. The feature of this set of BG decomposition methods, regardless of the GR type or the FHK type, is the comparison between new firms and disappearing firms. Baldwin and Gu (2006) employed a counterfactual argument to defend the use of aggregate productivity for disappearing firms. Without new firms, disappearing firms that would have exited in period t will continue to operate, and the sum of market shares of these pseudo firms in period t is exactly the sum of market shares of new firms. New firms enter the industry to take the place of market shares that would have been assigned to disappearing firms if they did not exit. This yields a direct comparison where the benchmark is the aggregate productivity of disappearing firms. If new firms are supposed to be generally small and less productive than continuing firms, an appropriate comparison for new firms should be with disappearing firms rather than with the industry average.

DF decomposition

Note that the preceding decomposition methods adopt the absolute firm weights when splitting contributions into continuing firms, new firms and disappearing firms. Consider the within effect in the BHC decomposition. The sum of market shares of continuing firms is less than 1 because disappearing firms take up the remaining part in period $t-1$. By taking the absolute firm weights among all firms in period $t-1$, the within term, that is, the sum of productivity change weighted by w_i^{t-1}, is likely to be understated. The entry effect and the exit effect suffer from similar biases because their firm weights also do not sum to 1. To ensure the sum of market shares of a certain type of firms is equal to 1, a relative firm weight, which is distinguished from the absolute firm weight, needs to be constructed. Diewert and Fox (2010) defined a micro input share as the firm weight for a type of firms and also defined an aggregate input share as the ratio of aggregate shares for a type of firms to aggregate shares for all firms. Based on the relative firm weight, the DF decomposition is as follows:

$$
\begin{aligned}
\Delta\Psi^{t,t-1} &= \sum_{i\in U} w_i^t \Psi_i^t - \sum_{i\in U} w_i^{t-1}\Psi_i^{t-1} \\
&= \underbrace{\sum_{i\in U_C} \frac{1}{2}(s_{Ci}^t + s_{Ci}^{t-1})(\Psi_i^t - \Psi_i^{t-1})}_{within} \\
&+ \underbrace{\sum_{i\in U_C} \frac{1}{2}(s_{Ci}^t - s_{Ci}^{t-1})(\Psi_i^t + \Psi_i^{t-1})}_{between} \\
&+ \underbrace{s_N^t \sum_{i\in U_N} s_{Ni}^t(\Psi_i^t - \Psi_C^t)}_{entry} - \underbrace{s_D^{t-1} \sum_{i\in U_D} s_{Di}^{t-1}(\Psi_i^{t-1} - \Psi_C^{t-1})}_{exit}
\end{aligned}
\tag{31}
$$

The micro input shares for continuing firms are $s_{Ci}^{t-1} = w_i^{t-1}/\sum_{i\in U_C} w_i^{t-1}$ and $s_{Ci}^t = w_i^t/\sum_{i\in U_C} w_i^t$ in periods $t-1$ and t respectively. For new firms, the micro input share is $s_{Ni}^t = w_i^t/\sum_{i\in U_N} w_i^t$, and for disappearing firms, it is similarly constructed as $s_{Di}^{t-1} = w_i^{t-1}/\sum_{i\in U_D} w_i^{t-1}$. These are relative measures of firm weights among selected firms compared with the absolute firm weights over all firms in an industry. In addition, the aggregate input share of new firms is $s_N^t = \sum_{i\in U_N} w_i^t/\sum_{i\in U} w_i^t$, and that figure of disappearing firms is $s_D^{t-1} = \sum_{i\in U_D} w_i^{t-1}/\sum_{i\in U} w_i^{t-1}$. The use of relative firm weights results in the reference productivity measures in the entry effect and the exit effect; they are Ψ_C^t for the productivity measure of continuing firms in period t and Ψ_C^{t-1} for the productivity measure of continuing firms in period $t-1$. New firms and disappearing firms can be compared with the aggregate productivity of continuing firms in their corresponding periods.

MP decomposition

An alternative basic framework of productivity dynamics is established by Olley and Pakes (1996). The OP decomposition suggests the growth in industry productivity is either due to resource reallocation from low-productivity firms to high-productivity firms, or an increase in the average firm productivity. I take continuing firms as an example. Suppose the number of continuing firms in period t or period $t-1$ is $N(C)$. The unweighted average productivity is:

$$\bar{\Psi}_C^t = \frac{1}{N(C)} \sum_{i \in U_C} \Psi_i^t \tag{32}$$

and the unweighted firm weight is:

$$\begin{aligned} \bar{s}_C^t &= \frac{1}{N(C)} \sum_{i \in U_C} s_{Ci}^t \\ &= \frac{1}{N(C)} \end{aligned} \tag{33}$$

The industry productivity of continuing firms can be decomposed into two terms:

$$\Psi_C^t = \bar{\Psi}_C^t + \sum_{i \in U_C} (s_{Ci}^t - \bar{s}_C^t)(\Psi_i^t - \bar{\Psi}_C^t) \tag{34}$$

The first term represents the unweighted average productivity of continuing firms and the second term is a mixed measurement of firm weights and productivity in a covariance type. The OP decomposition is extended into a dynamic version as the MP decomposition by Melitz and Polanec (2015):

$$\begin{aligned} \Delta\Psi^{t,t-1} &= \sum_{i \in U} w_i^t \Psi_i^t - \sum_{i \in U} w_i^{t-1} \Psi_i^{t-1} \\ &= \underbrace{\bar{\Psi}_C^t - \bar{\Psi}_C^{t-1}}_{mean} \\ &+ \underbrace{\sum_{i \in U_C} \left((s_{Ci}^t - \bar{s}_C^t)(\Psi_i^t - \bar{\Psi}_C^t) - (s_{Ci}^{t-1} - \bar{s}_C^{t-1})(\Psi_i^{t-1} - \bar{\Psi}_C^{t-1}) \right)}_{covariance} \\ &+ \underbrace{s_N^t \sum_{i \in U_N} s_{Ni}^t (\Psi_i^t - \Psi_C^t)}_{entry} - \underbrace{s_D^{t-1} \sum_{i \in U_D} s_{Di}^{t-1} (\Psi_i^{t-1} - \Psi_C^{t-1})}_{exit} \end{aligned} \tag{35}$$

The entry and exit effects in the MP decomposition are the same as the components in the DF decomposition. The MP decomposition mainly differs in decomposing the effect of continuing firms, though Melitz and Polanec (2015) suggested applying such a decomposition to entrants and exiters. The mean term is the change in unweighted mean productivity, and the covariance term captures the joint reallocation of firm weight and firm productivity. Since the unweighted mean productivity of continuing firms is used as the cross-sectional reference, the covariance term is different from the cross term in the FHK decomposition.[3]

3.3 Difference-in-differences specification

The DID (difference-in-differences) estimator is commonly used in econometrics to identify the treatment effect by comparing the treated group and the control group. The treated group is involved with a treatment or a policy change while the control group is not exposed to the impact. By comparing the differential change between the outcome of the treated group and the outcome of the control group, the treatment effect can be estimated with this DID approach.

I introduce the DID approach as it serves to provide a causal effect interpretation of productivity dynamics. Many decompositions of productivity dynamics are centred on the algebra but lack sufficient economic explanations to support them. Each type of these dynamics methods seems mathematically feasible. However, the components in the decomposition may not be the effects as their names reveal. There are concerns about, for example, whether the entry effect term in the decomposition correctly reflects the impact of new firms. The entry effect needs to be specified from the perspective of causal inference, rather than a straightforward algebra op-

[3]The MP decomposition is recorded for the completeness of the methodology summary. I will follow the classic framework of the BHC decomposition and concentrate on the framework where the within effect and the between effect are displayed.

eration. Baldwin and Gu (2006) has employed a counterfactual specification in firm dynamics, which is literally the DID approach. But only the entry effect is specified in their DID structure. I focus on the entry, exit, within and between effects. Additionally, as I adopt the aggregate value for each type of firms, the decomposition components are different from those in Baldwin and Gu (2006). Using the DID specification, I develop firm dynamics in an alternative way, and demonstrate why the DF decomposition proposed by Diewert and Fox (2010) is preferred.

The analysis framework is set up where industry productivity is measured for two markets: a real market and a pseudo market. The real market is affected by firm dynamics, that is, the entry of new firms and the exit of disappearing firms. In period $t-1$, the industry productivity of the real market is expressed as:

$$\begin{aligned}\Psi^{t-1} &= \sum_{i\in U_C} w_i^{t-1}\Psi_i^{t-1} + \sum_{i\in U_D} w_i^{t-1}\Psi_i^{t-1} \\ &= s_C^{t-1}\Psi_C^{t-1} + s_D^{t-1}\Psi_D^{t-1}\end{aligned} \tag{36}$$

where

$$s_C^{t-1} = \frac{\sum_{i\in U_C} w_i^{t-1}}{\sum_{i\in U} w_i^{t-1}} \tag{37}$$

$$s_D^{t-1} = \frac{\sum_{i\in U_D} w_i^{t-1}}{\sum_{i\in U} w_i^{t-1}} \tag{38}$$

$$\Psi_C^{t-1} = \sum_{i\in U_C} s_{Ci}^{t-1}\Psi_i^{t-1} \tag{39}$$

$$\Psi_D^{t-1} = \sum_{i\in U_D} s_{Di}^{t-1}\Psi_i^{t-1} \tag{40}$$

The relative weight s_C^{t-1} is an aggregate period $t-1$ market share of continuing firms that continue to exist from period $t-1$ to period t, and s_D^{t-1} is an aggregate market share of disappearing firms that choose to exit in period t. Ψ_C^{t-1} and Ψ_D^{t-1} measure the aggregate productivity of continuing firms and disappearing firms in period $t-1$ respectively. The micro firm weight for continuing firms is $s_{Ci}^{t-1} = w_i^{t-1}/\sum_{i\in U_C} w_i^{t-1}$

and that for disappearing firms is $s_{Di}^t = w_i^t / \sum_{i \in U_D} w_i^t$. Similarly, in period t, the industry productivity of the real market is expressed as:

$$\Psi^t = s_C^t \Psi_C^t + s_N^t \Psi_N^t \tag{41}$$

The aggregate market shares s_C^t and s_N^t, and the aggregate productivity Ψ_C^t and Ψ_N^t, are constructed in the same way:

$$s_C^t = \frac{\sum_{i \in U_C} w_i^t}{\sum_{i \in U} w_i^t} \tag{42}$$

$$s_N^t = \frac{\sum_{i \in U_N} w_i^t}{\sum_{i \in U} w_i^t} \tag{43}$$

$$\Psi_C^t = \sum_{i \in U_C} s_{Ci}^t \Psi_i^t \tag{44}$$

$$\Psi_N^t = \sum_{i \in U_N} s_{Ni}^t \Psi_i^t \tag{45}$$

With the industry productivity Ψ^{t-1} and Ψ^t, the productivity change of the real market over two periods is calculated as:

$$\Delta \Psi^{t,t-1} = s_C^t \Psi_C^t + s_N^t \Psi_N^t - s_C^{t-1} \Psi_C^{t-1} - s_D^{t-1} \Psi_D^{t-1} \tag{46}$$

Entry effects

As the counterpart, the pseudo market is assumed not to be disturbed by firm dynamics. I will make use of the pseudo market to specify the true entry effect, and then decompose productivity change into explainable components. In the base period $t-1$, the industry productivity in the pseudo market is supposed to be the same as that in the real market:

$$\widetilde{\Psi}^{t-1} = s_C^{t-1}\Psi_C^{t-1} + s_D^{t-1}\Psi_D^{t-1} \tag{47}$$

Consider the treatment in the DID model as the entry of new firms. An intuitive way to interpret the absence of new firms is that only continuing firms exist in the current period t. Without new firms in period t, the industry productivity would be the aggregate productivity of continuing firms:

$$\widetilde{\Psi}^{t} = \Psi_C^{t} \tag{48}$$

Given the industry productivity $\widetilde{\Psi}^{t-1}$ and $\widetilde{\Psi}^{t}$ in the pseudo market, the productivity change without an entry effect is:

$$\Delta\widetilde{\Psi}^{t,t-1} = \Psi_C^{t} - s_C^{t-1}\Psi_C^{t-1} - s_D^{t-1}\Psi_D^{t-1} \tag{49}$$

By subtracting the productivity change in the pseudo market, the productivity change before and after the firm entry in the real market can be written as the DID specification:

$$\begin{aligned}
\Delta\Psi^{t,t-1} - \Delta\widetilde{\Psi}^{t,t-1} &= s_C^{t}\Psi_C^{t} + s_N^{t}\Psi_N^{t} - \Psi_C^{t} \\
&= (1 - s_N^{t})\Psi_C^{t} + s_N^{t}\Psi_N^{t} - \Psi_C^{t} \\
&= s_N^{t}(\Psi_N^{t} - \Psi_C^{t}) \\
&= s_N^{t}\sum_{i\in U_N} s_{Ni}^{t}(\Psi_i^{t} - \Psi_C^{t})
\end{aligned} \tag{50}$$

Note this is exactly the entry effect in the DF decomposition proposed by Diewert and Fox (2010). Hence, this component in the decomposition is strongly supported by the DID specification.

Exit effects

The exit effect can be constructed correspondingly. In the current period t, the industry productivity in the pseudo market is supposed to be the same as that in the real market:

$$\widetilde{\Psi}^t = s_C^t \Psi_C^t + s_N^t \Psi_N^t \tag{51}$$

To specify the exit effect, I consider how the industry productivity responds to disappearing firms. Without the effect of disappearing firms, only continuing firms exist in the base period $t-1$. It allows the aggregate productivity of continuing firms to be the industry productivity:

$$\widetilde{\Psi}^{t-1} = \Psi_C^{t-1} \tag{52}$$

Since the industry productivity $\widetilde{\Psi}^{t-1}$ and $\widetilde{\Psi}^t$ in the pseudo market are obtained, the productivity change without an exit effect is:

$$\Delta\widetilde{\Psi}^{t,t-1} = s_C^t \Psi_C^t + s_N^t \Psi_N^t - \Psi_C^{t-1} \tag{53}$$

By comparing the productivity change in the pseudo market and the productivity change in the real market, the specification of the exit effect is computed by subtracting $\Delta\widetilde{\Psi}^{t,t-1}$ from $\Delta\Psi^{t,t-1}$:

$$\begin{aligned}
\Delta\Psi^{t,t-1} - \Delta\widetilde{\Psi}^{t,t-1} &= -s_C^{t-1}\Psi_C^{t-1} - s_D^{t-1}\Psi_D^{t-1} + \Psi_C^{t-1} \\
&= -(1 - s_D^{t-1})\Psi_C^{t-1} - s_D^{t-1}\Psi_D^{t-1} + \Psi_C^{t-1} \\
&= -s_D^{t-1}(\Psi_D^{t-1} - \Psi_C^{t-1}) \\
&= -s_D^{t-1} \sum_{i \in U_D} s_{Di}^{t-1}(\Psi_i^{t-1} - \Psi_C^{t-1})
\end{aligned} \tag{54}$$

This term is the same as the exit effect in the DF decomposition. Diewert and Fox (2010) considered it as the relative contribution of exiting firms compared with continuing firms, while the DID approach measures the exit effect as the treatment effect of disappearing firms, with the measure being exactly equal to that of Diewert and Fox (2010).

Within effects

The entry effect and the exit effect have been specified in the DID approach so that the industry productivity can be rewritten to highlight the contribution from continuing firms, new firms and disappearing firms. In the current period t, the industry productivity in the real market is:

$$\Psi^t = \Psi_C^t + s_N^t(\Psi_N^t - \Psi_C^t) \tag{55}$$

where the first term refers to the contribution of continuing firms to the overall productivity and the second term refers to the contribution of new firms to the overall productivity (the entry effect specified above). Similarly, in the base period $t-1$, the industry productivity in the real market is:

$$\Psi^{t-1} = \Psi_C^{t-1} + s_D^{t-1}(\Psi_D^{t-1} - \Psi_C^{t-1}) \tag{56}$$

where the firm term indicates the contribution from continuing firms and the second term indicates the contribution from disappearing firms. Note the second term is slightly different from the exit effect specified above: the negative sign in the exit effect has now been removed. This is because the exit effect is generated when this group of firms exit from the market in period t. When these firms are still operating in period $t-1$, the negative sign needs to be removed to measure their contribution to the overall productivity. By separating the contribution of different firms, I investigate the within effect and the between effect of continuing firms.

The within effect is defined as the contribution from continuing firms of which the firm productivity improves with fixed market shares. If the within effect did not occur, the firm productivity would be fixed at the productivity level in period $t-1$ or at the productivity level in period t.

For the first case, in the base period $t-1$, I suppose the industry productivity in the pseudo market is the same as that in the real market:

$$\widetilde{\Psi}^{t-1} = \Psi_C^{t-1} + s_D^{t-1}(\Psi_D^{t-1} - \Psi_C^{t-1}) \tag{57}$$

While in the current period t, the industry productivity without the within effect in the pseudo market would be:

$$\widetilde{\Psi}^{t} = \widetilde{\Psi}_C^{t} + s_N^{t}(\Psi_N^{t} - \Psi_C^{t}) \tag{58}$$

The contribution of continuing firms without productivity improvement is $\widetilde{\Psi}_C^t = \sum_{i \in U_C} s_{Ci}^t \Psi_i^{t-1}$. It shows that continuing firms would not have gained productivity improvement from period $t-1$ to period t without the within effect. By comparing the productivity change in the real market and the productivity change in the pseudo market, the within effect is specified as:

$$\begin{aligned}
\Delta\Psi^{t,t-1} - \Delta\widetilde{\Psi}^{t,t-1} &= \Psi_C^t - \widetilde{\Psi}_C^t \\
&= \sum_{i \in U_C} s_{Ci}^t \Psi_i^t - \sum_{i \in U_C} s_{Ci}^t \Psi_i^{t-1} \\
&= \sum_{i \in U_C} s_{Ci}^t (\Psi_i^t - \Psi_i^{t-1})
\end{aligned} \tag{59}$$

For the second case, in the current period t, I suppose the productivity in the pseudo market is the same as that in the real market:

$$\widetilde{\Psi}^{t} = \Psi_C^{t} + s_N^{t}(\Psi_N^{t} - \Psi_C^{t}) \tag{60}$$

While in the base period $t-1$, the industry productivity without the within effect in the pseudo market would be:

$$\widetilde{\Psi}^{t-1} = \widetilde{\Psi}_C^{t-1} + s_D^{t-1}(\Psi_D^{t-1} - \Psi_C^{t-1}) \tag{61}$$

The contribution of continuing firms without productivity improvement turns to be $\widetilde{\Psi}_C^{t-1} = \sum_{i \in U_C} s_{Ci}^{t-1} \Psi_i^t$. It demonstrates that continuing firms would not have gained productivity growth over periods without the within effect. The within effect is specified by comparing the productivity change in the real market and that in the pseudo market:

$$\begin{aligned} \Delta\Psi^{t,t-1} - \Delta\widetilde{\Psi}^{t,t-1} &= -\Psi_C^{t-1} + \widetilde{\Psi}_C^{t-1} \\ &= -\sum_{i \in U_C} s_{Ci}^{t-1}\Psi_i^{t-1} + \sum_{i \in U_C} s_{Ci}^{t-1}\Psi_i^t \\ &= \sum_{i \in U_C} s_{Ci}^{t-1}(\Psi_i^t - \Psi_i^{t-1}) \end{aligned} \tag{62}$$

The within effect in the first case is based on the assumption that productivity is fixed on the level in period $t-1$ and in the second case the productivity is assumed to be fixed on the level in period t. An average of the within effects for these two cases leads to the balanced within effect, which is also the within effect in the DF decomposition:

$$\Delta\Psi^{t,t-1} - \Delta\widetilde{\Psi}^{t,t-1} = \frac{1}{2}\sum_{i \in U_C}(s_{Ci}^t + s_{Ci}^{t-1})(\Psi_i^t - \Psi_i^{t-1}) \tag{63}$$

Between effects

The between effect captures the contribution from continuing firms of which the market shares improve with fixed productivity levels. If the between effect did not occur, the market shares would be fixed at the firm weights in period $t-1$ or at the firm weights in period t. Since the analysis of the between effect is symmetric with the within effect, I skip the repeated explanation of settings in the real market and in the pseudo market. The DID specification is almost the same except the specific expressions of $\widetilde{\Psi}_C^t$ and $\widetilde{\Psi}_C^{t-1}$.

If the market shares are assumed to be fixed at the base period $t-1$, then the contribution of continuing firms without market share improvement would be $\widetilde{\Psi}_C^t = \sum_{i \in U_C} s_{Ci}^{t-1} \Psi_i^t$. The DID specification of the between effect is:

$$\begin{aligned} \Delta\Psi^{t,t-1} - \Delta\widetilde{\Psi}^{t,t-1} &= \Psi_C^t - \widetilde{\Psi}_C^t \\ &= \sum_{i \in U_C} s_{Ci}^t \Psi_i^t - \sum_{i \in U_C} s_{Ci}^{t-1} \Psi_i^t \\ &= \sum_{i \in U_C} (s_{Ci}^t - s_{Ci}^{t-1}) \Psi_i^t \end{aligned} \tag{64}$$

If the market shares are assumed to be fixed at the current period t, then the contribution of continuing firms without market share change would be $\widetilde{\Psi}_M^{t-1} = \sum_{i \in U_M} s_{Mi}^t \Psi_i^{t-1}$. The DID specification of the between effect is:

$$\begin{aligned} \Delta\Psi^{t,t-1} - \Delta\widetilde{\Psi}^{t,t-1} &= -\Psi_C^{t-1} + \widetilde{\Psi}_C^{t-1} \\ &= -\sum_{i \in U_C} s_{Ci}^{t-1} \Psi_i^{t-1} + \sum_{i \in U_C} s_{Ci}^t \Psi_i^{t-1} \\ &= \sum_{i \in U_C} (s_{Ci}^t - s_{Ci}^{t-1}) \Psi_i^{t-1} \end{aligned} \tag{65}$$

Taking the average of the between effects in these two situations, the same term as the between effect in the DF decomposition can be found:

$$\Delta\Psi^{t,t-1} - \Delta\widetilde{\Psi}^{t,t-1} = \frac{1}{2} \sum_{i \in U_C} (s_{Ci}^t - s_{Ci}^{t-1})(\Psi_i^t + \Psi_i^{t-1}) \tag{66}$$

3.4 Firm replacement

By incorporating the DID specification, I have demonstrated that the DF decomposition is not only mathematically correct but also consistent with causal inference. Alternatively, I take the perspective that the entry of new firms occurs so that the disappearing firms can be replaced. This replacement concept in firm dynamics has been proposed by Baldwin and Gu (2006), but I reach a different decomposition result from the BG decomposition.[4]

Without new firms entering the market, firms that should have disappeared would remain in the current period. So disappearing firms are technically matched with new firms, resulting in the process of firm replacement. Firm replacement improves the industry productivity through two channels: enhancing firm productivity and enlarging market shares. These effects are not the entry effect or the exit effect that have been elaborated in previous decomposition methods. I define them as a capacity effect and a portion effect caused by firm replacement. The capacity effect measures productivity improvement when disappearing firms are replaced by new firms, while the portion effect measures the change of market shares caused by the replacement. I use the DID specification to compute the capacity and portion effects. The industry productivity without a capacity effect in the pseudo market would be:

$$\widetilde{\Psi}^t = s_C^t \Psi_C^t + s_N^t \Psi_D^{t-1} \tag{67}$$

In this simulation case, the disappearing firms remain in period t with aggregate productivity Ψ_D^{t-1} and occupy the market share that would have been taken by new firms. Without the capacity effect of new firms, the productivity change in the

[4]Because the decomposition result is a deviation from the classic framework that consists of entry effect and exit effect, I list it here as an innovative idea but not for empirical exercises.

pseudo market is expressed as:

$$\Delta\widetilde{\Psi}^{t,t-1} = s_C^t \Psi_C^t + s_N^t \Psi_D^{t-1} - s_C^{t-1}\Psi_C^{t-1} - s_D^{t-1}\Psi_D^{t-1} \tag{68}$$

Using the DID specification, I subtract the productivity change in the pseudo market from the productivity change in the real market, and figure out the capacity effect:

$$\Delta\Psi^{t,t-1} - \Delta\widetilde{\Psi}^{t,t-1} = s_N^t(\Psi_N^t - \Psi_D^{t-1}) \tag{69}$$

The capacity effect will be positive if the aggregate productivity of new firms is larger than the aggregate productivity of disappearing firms. This underlines the interpretation that new firms take the place of disappearing firms and make productivity improvement. Given the capacity effect, the portion effect can be specified in a similar way. For simplicity, I skip the DID demonstration and construct the portion effect straightforwardly by adding and subtracting $s_N^t\Psi_D^{t-1}$. Meanwhile, the term $s_C^t\Psi_C^{t-1}$ is added and subtracted to separate the within effect and the between effect:

$$\begin{aligned}\Delta\Psi^{t,t-1} &= s_C^t(\Psi_C^t - \Psi_C^{t-1}) + (s_C^t - s_C^{t-1})\Psi_C^{t-1} \\ &+ s_N^t(\Psi_N^t - \Psi_D^{t-1}) + (s_N^t - s_D^{t-1})\Psi_D^{t-1}\end{aligned} \tag{70}$$

These effects are the same as the components in the DID specification. To redress the balance of different periods, the term $s_D^{t-1}\Psi_N^t$ and the term $s_C^{t-1}\Psi_C^t$ are added and subtracted in a similar way for continuing firms:

$$\begin{aligned}\Delta\Psi^{t,t-1} &= s_C^{t-1}(\Psi_C^t - \Psi_C^{t-1}) + (s_C^t - s_C^{t-1})\Psi_C^t \\ &+ s_D^{t-1}(\Psi_N^t - \Psi_D^{t-1}) + (s_N^t - s_D^{t-1})\Psi_N^t\end{aligned} \tag{71}$$

An average of these two decomposition methods returns the productivity change that shows the within, between, capacity and portion effects:

$$\begin{aligned}\Psi^t - \Psi^{t-1} = &\frac{1}{2}(s_C^t + s_C^{t-1})(\Psi_C^t - \Psi_C^{t-1}) \\ &+ \frac{1}{2}(s_C^t - s_C^{t-1})(\Psi_C^t + \Psi_C^{t-1}) \\ &+ \frac{1}{2}(s_N^t + s_D^{t-1})(\Psi_N^t - \Psi_D^{t-1}) \\ &+ \frac{1}{2}(s_N^t - s_D^{t-1})(\Psi_N^t + \Psi_D^{t-1})\end{aligned} \tag{72}$$

More specifically, the decomposition highlighting firm replacement dynamics is expressed as:

$$\begin{aligned}\Delta\Psi^{t,t-1} &= \sum_{i\in U} w_i^t \Psi_i^t - \sum_{i\in U} w_i^{t-1}\Psi_i^{t-1} \\ &= \underbrace{\frac{1}{2}(s_C^t + s_C^{t-1})\left(\sum_{i\in U_C} s_{Ci}^t \Psi_i^t - \sum_{i\in U_C} s_{Ci}^{t-1}\Psi_i^{t-1}\right)}_{within} \\ &+ \underbrace{\frac{1}{2}(s_C^t - s_C^{t-1})\left(\sum_{i\in U_C} s_{Ci}^t \Psi_i^t + \sum_{i\in U_C} s_{Ci}^{t-1}\Psi_i^{t-1}\right)}_{between} \\ &+ \underbrace{\frac{1}{2}(s_N^t + s_D^{t-1})\left(\sum_{i\in U_N} s_{Ni}^t \Psi_i^t - \sum_{i\in U_D} s_{Di}^{t-1}\Psi_i^{t-1}\right)}_{capacity} \\ &+ \underbrace{\frac{1}{2}(s_N^t - s_D^{t-1})\left(\sum_{i\in U_N} s_{Ni}^t \Psi_i^t + \sum_{i\in U_D} s_{Di}^{t-1}\Psi_i^{t-1}\right)}_{portion}\end{aligned} \tag{73}$$

This new decomposition features the comparison of aggregate values. The within effect measures the change of aggregate productivity for continuing firms weighted by an averaged market share. The between effect measures the change of aggregate market shares for continuing firms weighted by average productivity. They are distinct from the components in all preceding decomposition methods where the difference is computed for each unit. In addition, the capacity and portion effects emphasise the consequence of firm replacement. The capacity effect will be positive if new firms in period t present larger aggregate productivity than disappearing firms in period $t-1$. It could indicate new firms entering the market to replace less productive firms and thereby improving the industry productivity. The portion effect will be positive if new firms in period t occupy larger market shares than disappearing firms in period $t-1$. It shows that new firms enter the industry to expand market shares and the industry productivity is enhanced. It may seem surprising that increasing shares from new firms will improve the industry productivity even without higher productivity than continuing firms. But the portion effect is analogous to the between effect; given constant firm productivity, the between effect will contribute to industry productivity with larger market shares, and so will the portion effect.

4 Firm-level evidence

The Business Longitudinal Analysis Data Environment (BLADE) is a collection of data sets that provide firm-level information with data sets linked through an Australian Business Number (ABN). The data sets collected from the Australia Bureau of Statistics (ABS) are integrated with data reported to the Australian Taxation Office (ATO). ABS survey databases include the Economic Activity Survey, the Business Characteristics Survey and the Survey of the Business R&D while ATO databases include Business Activity Statements and Business Income Tax. The partnership between the ATO and the ABS contributes to data products enclosed in the BLADE.

I focus on the two main administrative data sets from the ATO in BLADE: Business Activity Statements (BAS) and Business Income Tax (BIT). Key indicators from BAS are total sales, non-capital purchases, and wages, salaries and other payments.

These variables can be used to measure gross output, intermediate input and labour input. From the BIT data, I collect detailed items on capital depreciation and capital work deduction that help to measure capital services (see the calculation of capital cost in Chien et al. (2019)). An alternative approach is to estimate capital services by subtracting labour value and intermediate input value from gross output. With these variables, a clear understanding is informed about the value measures of gross output, labour input, capital input and intermediate input. The research period covered by these variables is from 2002 to 2016.

Apart from BLADE, external data sources are also introduced from the ABS national accounts and productivity estimate tables so that price measures can be derived for gross output, labour input, capital input and intermediate input. The Australian Systems of National Accounts have published current prices of gross value added at the industry level. I take value added for 12 selected industries, specifically Divisions A–K and R (see Table 2). Productivity estimate tables from the ABS provide cost shares of labour, capital and intermediate input by which I estimate gross output value. The ABS (2015) has specified cost shares of capital (Z_L), labour (Z_K) and intermediate input (Z_X) that are expressed as:

$$Z_L = \frac{Lw_L}{Gp_G} \tag{74}$$

$$Z_K = \frac{Kr_K}{Gp_G} \tag{75}$$

$$Z_X = \frac{Xp_X}{Gp_G} \tag{76}$$

where w_L is the price of labour, r_K is the price of capital services, p_X is the price of intermediate input, and p_G is the price of gross output. L, K, X and G are quantity measures of labour, capital services, intermediate input and gross output respectively. Taking the current prices of value added for a certain industry, I model the current prices of gross output as value added divided by the sum of cost shares of labour and capital services. Gross output values multiplied by cost shares Z_L, Z_K and Z_X produce the current prices of labour, capital services and intermediate input. Following the data set construction for OECD countries by Fox (1997), I take value measures and divide them by quantity indexes, which leads to implicit price indexes. The quantity indexes are labour input indexes (quality adjusted hours worked basis), capital services indexes, intermediate input indexes and gross output indexes that are all obtained from the ABS multifactor productivity data cube.[5]

Table 2: Selected industries from Australian market sector industry classification

Division	Industry
A	Agriculture, Forestry and Fishing
B	Mining
C	Manufacturing
D	Electricity, Gas, Water and Waste Services
E	Construction
F	Wholesale Trade
G	Retail Trade
H	Accommodation and Food Services
I	Transport, Postal and Warehousing
J	Information, Media and Telecommunications
K	Financial and Insurance Services
R	Arts and Recreation Services

A set of filters is applied to the raw data from the BLADE before the productivity decomposition and productivity dynamics analyses are conducted. These filters address the following issues:

- ANZSIC code mismatch. If firm business activities switch from their original division to a new division, the Australian and New Zealand Standard Industrial Classification (ANZSIC) code is updated retrospectively, that is, the ANZSIC code in previous years are updated to be consistent with the new division. However, the record of division names in previous years remains unchanged,

[5]Estimates of industry multifactor productivity, cost shares and quantity indexes are available from https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/5260.0.55.0022018-19?OpenDocument (accessed December 4, 2019).

causing a mismatch between the ANZSIC code and division names. Firms with such mismatch are deleted because these observations act as if firms kept contributing to the initial industry while they have already moved to a different industry.

- Missing values. Firms are removed if they contain missing values for firm gross output (total sales), labour input (total salary, wages and other payments) and intermediate input (non-capital purchases). For internally computed variables such as price measures and quantity measures of output and input, observations with missing values are dropped.

- Zero values. Firms are removed if they contain zeros values for firm gross output (total sales) and labour input (total salary, wages and other payments), that is, they are treated as being inactive. For internally computed quantity measures of output and input, observations with zero values as well as negative values are dropped.

- GST adjustment. The value measures of output and input need to be adjusted to exclude the goods and services tax (GST) that is a tax rate of 10%. A tax rate of firm gross output can be derived from the total sales and the total taxes on sales from the raw data. If the total taxes are greater than what is supposed to be for a tax rate of 10%, these records seem spurious and the tax rate is set to 10% of total sales before taxes are excluded from firm gross output.[6] Similarly, value measures of labour input and intermediate input are adjusted to rule out the GST.

- Duplicates. A check of duplicates is conducted to examine if there are observations with the same firm code showing repeatedly in any period. These

[6]Information on the inclusion (or not) of the GST will help to produce total sales records with the GST excluded in future releases (since the release of 2017/2018 records). The GST adjustment filter will not be in use then.

items will be merged into a single observation if duplicates are detected after applying the filters above.

- Outliers. For each period, observations with finally computed productivity outside three standard deviations from mean values are labelled as outliers and are dropped from the data set. This outlier filter examines productivity measures rather than output or input measures. A wide variety of firms exist in each industry. If outliers are identified according to the output and input measures like total sales, leading firms with large sales will be wrongly removed and the conclusions on firm dynamics may not reveal the true industry performance.

I explore 12 industries from the Australian market sector, as listed above in Table 2. Figure 1 plots industry productivity estimates. Red lines indicate productivity measures from the ABS, green dashed lines indicate productivity estimates with firm-level data from the BLADE, and blue dashed lines indicate productivity estimates with industry-level data from the BLADE. Productivity estimates with the new firm-level productivity method match the ABS results well in industries such manufacturing, information (plus media and telecommunications), and finance (plus insurance services), which validates the reliability of the firm-level productivity method. The estimates in some industries do not perfectly match the ABS outcome but they essentially share the same trend over years. These industries include agriculture (plus forestry and fishing), mining, and electricity (plus gas, water and waste services). Large disparities are observed between productivity estimates in the remaining industries. Since productivity measures from the ABS are not based on the BLADE, it is reasonable to expect such for such disparities. Another disparity lies in the productivity estimates from firm-level data and those from industry-level data. The firm-level estimates are conducted by decomposing firm output ratios into explanatory factors and then aggregating them while the industry-level estimates are conducted by aggregating input and output values before decomposing the aggregate output growth. The differences between firm-level estimates and industry-level estimates highlight the importance of firm-specific details from which different conclusions may be drawn.

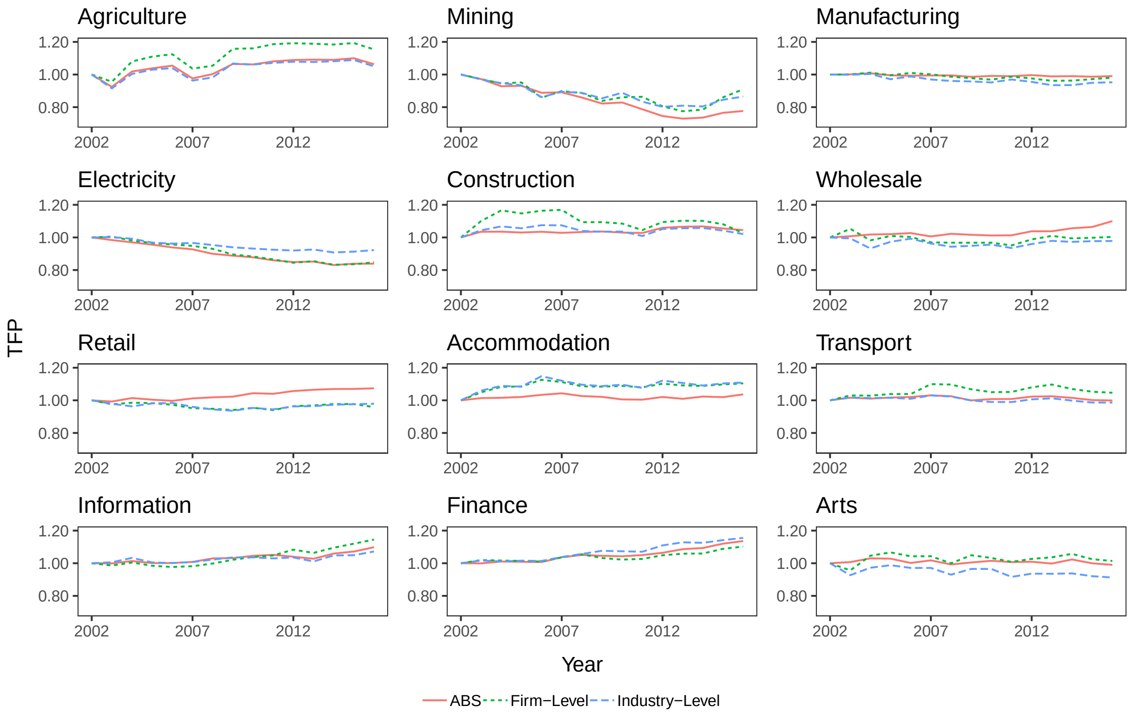

Figure 1: Productivity estimates for 12 selected industries

A series of methods about productivity dynamics has been introduced, among which the DF decomposition (Diewert and Fox, 2010) is supported by the difference-in-differences specification. Figure 2 follows the DF decomposition and decomposes industry productivity change into the between, with, entry and exit effects. These effects are averaged over years for each industry in the plot. It can be seen that the within effect contributes most to industry productivity change in all 12 industries. It means the productivity growth of continuing firms substantially determine aggregate productivity growth. Positive entry effects are observed in utilities, manufacturing, transport and agriculture industries, but their influences are fairly limited. Negative exit effects are observed in utilities, construction, transport, finance and agriculture industries. Positive signs show that firms enter the market with larger productivity than continuing firms while negative signs show that firms exit the market with larger productivity than continuing firms. But the productivity estimates of new and disappearing firms are not greatly different from those of continuing firms, which accounts for the limited entry and exit terms.

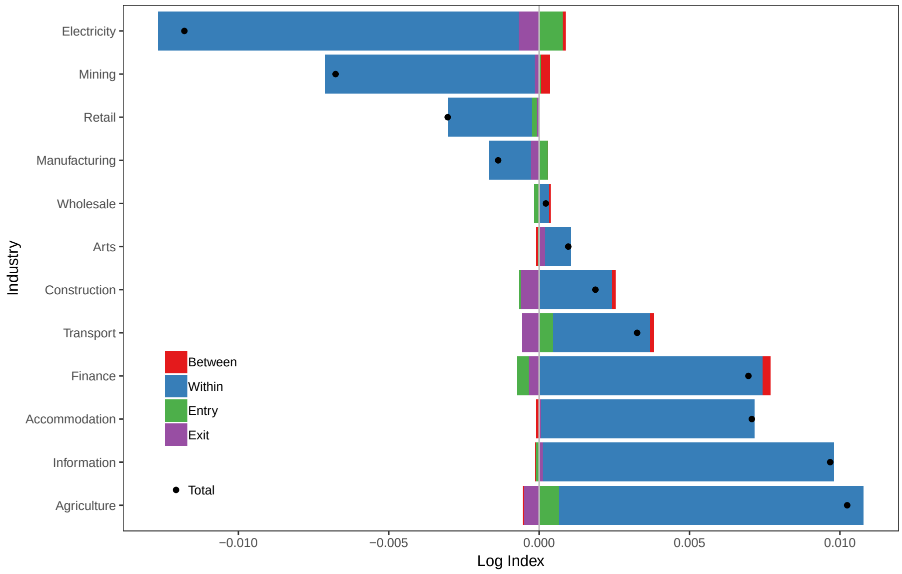

Figure 2: Productivity dynamics average from 2002 to 2016

Based on the firm-level productivity decomposition, explanatory factors of productivity can also be analysed for firm dynamics. Figure 3 demonstrates inefficiency dynamics that specifies the contribution of incumbents, entrants and exiters to aggregate production efficiency. The within effect is once again confirmed to have the largest impact on aggregate performance, indicating the contribution of incumbents. Recall that inefficiency measures the production gap between normal firms and frontier firms. Only a few firms are on the frontier so that the remaining firms seem inefficient compared with the leading firms. Since the market is mainly composed of incumbents, the inefficiency would be aggregated over these incumbents, leading to the large within effect. It is reasonable to interpret the within effect about inefficiency as the result of a large proportion of incumbents.

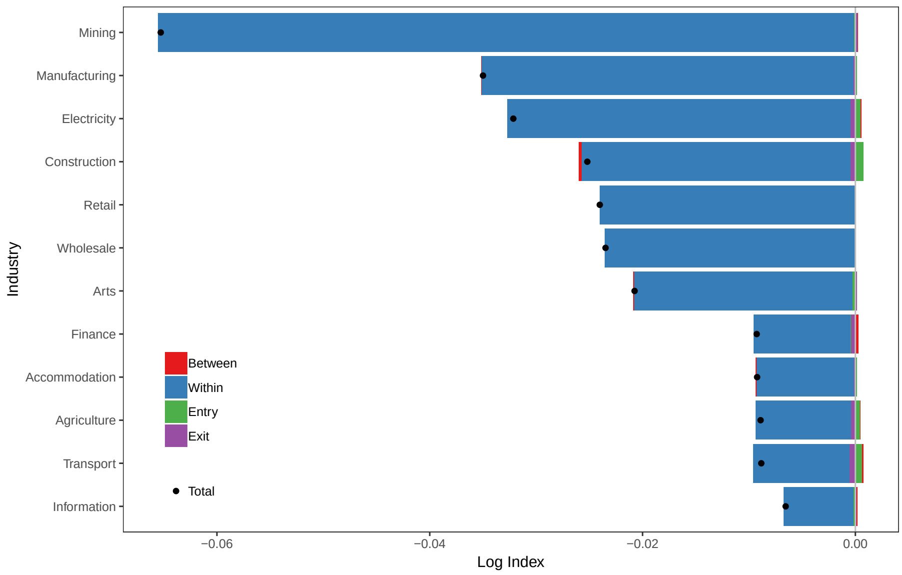

Figure 3: Inefficiency dynamics average from 2002 to 2016

Previous studies on firm dynamics support the firm-level evidence of this paper that there are remarkable contributions from incumbents to aggregate performance. The within-firm improvement is found to dominate the overall development of aggregate productivity in Balk (2003), Baldwin and Gu (2011), and Riley et al. (2015), though they draw conclusions for different economies. Baldwin and Gu (2011) compared the firm churning of retail trade and manufacturing sectors in Canada, Riley et al. (2015) analysed the productivity puzzle after the financial crisis in the United Kingdom, and Balk (2003) reviewed the empirical evidence of manufacturing business churn in Finland, France, Italy, Netherlands, Portugal and the United Kingdom. All of them found strong effects of inter-firm and intra-firm relations, that is, the within and between relations on the industry productivity fluctuation. For the case of Australia, the existing research about productivity dynamics using BLADE is at an early stage where limited papers are available. The only study, to my best knowledge, is done by He (2018) who conducted an empirical investigation of Australian business dynamism and also validated the contribution within firms to illustrating industry productivity change. In his research, aggregate productivity growth is decomposed into common productivity, internal dynamism and external dynamism. He (2018) finds that firm survivors have been the driving force behind industry productivity growth, which provides similar insights to this paper.

The importance of incumbents to firm dynamics may seem surprising because a prima facie statement is that start-ups and young firms are expected to perform more productively and more efficiently than incumbents. However, the concept of new firms should not be confused with young firms. Young firms are defined as entrants for productivity dynamics only in the year they enter the market. They will be specified as incumbents if they survive in consecutive years. In addition to the concept clarification, a noticeable fact is that the evolution of Australian firms begins to move at a lower speed as a decreasing entry rate is observed according to Bakhtiari (2019). The impact of firm entries is limited due to the decline in entry rates and therefore the aggregate productivity does not benefit much from these firms. A symmetric case to the decreasing entry rate is that the possibility of new firms exiting the market is increasing, with an exit rate at approximately 24% in their first three years and beyond 60% in their first ten years (Bakhtiari, 2019). The inactive performance of new firms is attributed to factors that discourage potential firms from joining the existing market. A generic identification of such factors comes from Orr (1974) that modelled the firm entry as a function of incentives and barriers. Key incentives to enter include profit rates and industry output growth, while market barriers are mainly reflected by economies of scale, capital requirements and industry concentration. To explain the entry deterioration in Australia, Bakhtiari (2019) proposed some reasons that may be relevant: a price boom of resources, an increasing cost of loans, competitive pressure caused by globalisation, demographic features like aged population, and market monopolies. These reasons basically fall into the category of incentives or barriers.

The analytical framework of incentives and barriers to entry also applies to the virus outbreak that seriously strikes the economy. Due to the global spread of the coronavirus discovered in 2019, a large number of countries have placed strict restrictions on social distancing to slow down and stop the transmission of the virus. Australia's coronavirus restrictions advise the avoidance of non-essential travel within Australia to protect citizens and visitors from potential infection risks. Economic sectors are heavily hit by a dramatic decline in the demand of products and services. Business profits drop rapidly under these circumstances and discourage the entry of potential firms. Numerous firms have to close down and exit the market in a period of financial hardship. Incumbents decide to operate below the full capacity level in response to the decreasing demand. A typical example is that only delivery orders are available in some restaurants and shops that are still running. The reasons that firms choose to enter, continue or exit can be well explained by incentives and barriers amid the virus shock to industries.

A large amount of inefficiency is produced and it will validate the use of the productivity decomposition at the firm level once data covering 2020 become available. Clearly, the deterioration of production output after the virus outbreak is not originated by a decreasing level of technology because the existing expertise of industry production will not become less immediately. The economic deterioration is more correctly interpreted as inefficiency that occurs widely in firms where product orders fall down and firm resources are left idle. The firm-level productivity decomposition built upon the DF decomposition (Diewert and Fox, 2018) distinguishes between inefficiency and technical regress. Once statistical data sets required for the model implementation are ready in 2021, researchers will be able to run the productivity decomposition method on the data sets accordingly and examine the decomposition results to quantify declines in production efficiency that is anticipated due to the economy facts in 2020. This serves as a validation approach to judge the merit of the productivity decomposition which clarifies the contributions of efficiency change, and avoids productivity decline from being incorrectly interpreted as technical regress.

5 Conclusion

This paper proposes a new framework where the productivity decomposition can be integrated with productivity dynamics. It is done by establishing a firm-level productivity decomposition and then attributing the industry performance to the firm performance. The productivity decomposition specifies explanatory factors of firm productivity such as efficiency and technical progress, while productivity dynamics allow firm contributions to be allocated as the entry, exit, between and within effects. With the firm-level productivity decomposition, not only productivity can be analysed by the entry and exit effects but also the explanatory factors can be treated in the same way.

The firm-level productivity decomposition is built upon the DF decomposition (Diewert and Fox, 2018), but it has been adapted to be compatible with firm-level data. Key assumptions have been made to facilitate productivity decomposition at the firm level, providing new perspectives about how productivity can be explained: a sequential approach rules out technical regress with a period to period examination of firms; the concept of firm gross output is taken as the appropriate output measure; a rolling window ensures that the firm-level decomposition is equivalent to the DF decomposition when firm-level data collapse into industry-level data; and industry-level prices improve the availability of the new framework given that firm-level prices are difficult to collect.

For productivity dynamics, a difference-in-differences approach follows the review of different methods for quantifying firm contributions to aggregate performance such as entry and exit effects. The counterfactual exercise considers, for example the entry effect by comparing the outcome change in a real market and a pseudo market. It produces the same components as those in the DF decomposition of productivity dynamics (Diewert and Fox, 2010), validating the DF decomposition with the causal specification.

Empirical results from the BLADE are produced with the firm-level productivity decomposition. Compared with the productivity release from the ABS, firm-level and industry-level productivity estimates match the official records in industries such as manufacturing, information (plus media and telecommunications) and finance (plus insurance services), and they share the similar pattern with the ABS productivity estimates in industries such as agriculture (plus forestry and fishing), mining and utilities. The disparity between firm-level and industry-level results reminds us of the importance of firm-level information, which justifies the use of firm-level productivity decomposition.

The firm-level productivity decomposition also demonstrates micro drivers of industry inefficiency when enclosed with firm dynamics. The change of industry inefficiency is explained by entry, exit, between and within effects, in the same way as how the change of industry productivity is allocated. Incumbents are found to dominate both industry productivity and industry inefficiency, supported by remarkable within effects in the decomposition. The contributions of entries and exiters to the industry are limited due to their relatively small weights. Australian productivity has slowed, reflecting a decreasing entry rate and an increasing exiting rate of new firms.

The importance of distinguishing inefficiency from technical regress becomes clear when economic sectors are harshly hit by the outbreak of the coronavirus. To fight against the virus, governments have imposed strict restrictions on social distancing and non-essential businesses so that residents are protected from potential infection risks. The side effect is that the economy is suffering from a sudden decline in the demand of products and services. The following deterioration of industry productivity is obviously not caused by technical regress since the existing expertise and skills are unlikely to disappear immediately. Inefficiency appears as a more reasonable interpretation to account for the decreasing productivity as it can be noticed that numerous firms choose to operate below full capacity when product orders drop rapidly. A large amount of inefficiency is anticipated once the updated data sets are available in 2021, and this will justify the use of such a method.

Disclaimer

The results of these studies are based, in part, on Australian Business Registrar (ABR) data supplied by the Registrar to the ABS under A New Tax System (Australian Business Number) Act 1999 and tax data supplied by the ATO to the ABS under the Taxation Administration Act 1953. These require that such data is only used for the purpose of carrying out functions of the ABS. No individual information collected under the Census and Statistics Act 1905 is provided back to the Registrar or ATO for administrative or regulatory purposes. Any discussion of data limitations or weaknesses is in the context of using the data for statistical purposes, and is not related to the ability of the data to support the ABR or ATO's core operational requirements. Legislative requirements to ensure privacy and secrecy of this data have been followed. Source data are de-identified and so data about specific firms has not been viewed in conducting this analysis. In accordance with the Census and Statistics Act 1905, results have been confidentialised used to ensure that they are not likely to enable identification of a particular person or organisation.

References

Australian Bureau of Statistics (ABS) (2015). *Australian system of national accounts: concepts, sources and methods*. ABS Catalogue No. 5216.0.

Baily, M. N., Hulten, C., and Campbell, D. (1992). Productivity dynamics in manufacturing plants. *Brookings Papers on Economic Activity*, Microeconomics 1992:187–267.

Bakhtiari, S. (2019). Entrepreneurship dynamics in Australia: lessons from micro-data. *Economic Record*, 95(308):114–140.

Baldwin, J. R. and Gu, W. (2006). Plant turnover and productivity growth in Canadian manufacturing. *Industrial and Corporate Change*, 15(3):417–465.

Baldwin, J. R. and Gu, W. (2011). Firm dynamics and productivity growth: a comparison of the retail trade and manufacturing sectors. *Industrial and Corporate Change*, 20(2):367–395.

Balk, B. M. (2003). The residual: on monitoring and benchmarking firms, industries, and economies with respect to productivity. *Journal of Productivity Analysis*, 20(1):5–47.

Braguinsky, S., Ohyama, A., Okazaki, T., and Syverson, C. (2015). Acquisitions, productivity, and profitability: evidence from the Japanese cotton spinning industry. *American Economic Review*, 105(7):2086–2119.

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

Breunig, R. and Wong, M. (2008). A richer understanding of Australia's productivity performance in the 1990s: improved estimates based upon firm-level panel data. *Economic Record*, 84(265):157–176.

Brynjolfsson, E. and Hitt, L. M. (2003). Computing productivity: firm-level evidence. *The Review of Economics and Statistics*, 85(4):793–808.

Chien, C. H., Welsh, A., and Breunig, R. (2019). Approaches to analysing micro-drivers of aggregate productivity. Research paper, ABS catalogue No. 1351.0.55.164, Australian Bureau of Statistics.

de Haan, J. (2015). A framework for large scale use of scanner data in the Dutch CPI. Paper presented at the fourteenth Ottawa Group meeting, Tokyo, Japan.

Decker, R. A., Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2017). Declining dynamism, allocative efficiency, and the productivity slowdown. *American Economic Review*, 107(5):322–26.

Delgado, M. A., Fariñas, J. C., and Ruano, S. (2002). Firm productivity and export markets: a non-parametric approach. *Journal of International Economics*, 57(2):397–422.

Diewert, W. E. and Fox, K. J. (2010). On measuring the contribution of entering and exiting firms to aggregate productivity growth. In *Price and productivity measurement: volume 6 – index number theory*, chapter 3, pages 41–66. Trafford Press.

Diewert, W. E. and Fox, K. J. (2017). Substitution bias in multilateral methods for CPI construction using scanner data. Microeconomics working papers, Vancouver School of Economics.

Diewert, W. E. and Fox, K. J. (2018). Decomposing value added growth into explanatory factors. In *The Oxford Handbook of Productivity Analysis*, chapter 19, page 625–662. Oxford University Press: New York.

Eltetö, O. and Köves, P. (1964). On a problem of index number computation relating to international comparison. *Statisztikai Szemle*, 42:507–518.

Foster, L., Grim, C., Haltiwanger, J., and Wolf, Z. (2017). Innovation, productivity growth and productivity dispersion. Working paper, U.S. Census Bureau.

Foster, L., Haltiwanger, J., and Syverson, C. (2008). Reallocation, firm turnover, and efficiency: selection on productivity or profitability? *American Economic Review*, 98(1):394–425.

Foster, L., Haltiwanger, J. C., and Krizan, C. J. (2001). Aggregate productivity growth: lessons from microeconomic evidence. In *New Developments in Productivity Analysis*, chapter 8, page 303–372. University of Chicago Press.

Fox, K. J. (1997). Open economy data sets for 18 OECD countries, 1960–1992. Discussion paper, School of Economics, University of New South Wales.

Gini, C. (1931). On the circular test of index numbers. *Metron*, 9(9):3–24.

Griliches, Z. and Regev, H. (1995). Firm productivity in Israeli industry 1979–1988. *Journal of Econometrics*, 65(1):175–203.

He, C. (2018). Holding productivity accountable: business dynamism and productivity. Paper presented at the EMG workshop, Sydney, Australia.

Ivancic, L., Diewert, W. E., and Fox, K. J. (2011). Scanner data, time aggregation and the construction of price indexes. *Journal of Econometrics*, 161(1):24–35.

Konrad, A. M. and Mangel, R. (2000). The impact of work-life programs on firm productivity. *Strategic Management Journal*, 21(12):1225–1237.

Krsinich, F. (2016). The FEWS index: fixed effects with a window splice. *Journal of Official Statistics*, 32(2):375–404.

Maré, D. C., Hyslop, D. R., and Fabling, R. (2017). Firm productivity growth and skill. *New Zealand Economic Papers*, 51(3):302–326.

Melitz, M. J. and Polanec, S. (2015). Dynamic Olley-Pakes productivity decomposition with entry and exit. *The RAND Journal of Economics*, 46(2):362–375.

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

Orr, D. (1974). The determinants of entry: a study of the Canadian manufacturing industries. *The Review of Economics and Statistics*, 56(1):58–66.

Pierpaolo Parrotta, Dario Pozzoli, M. P. (2012). Does labor diversity affect firm productivity? Discussion paper No. 6973, Institute for the Study of Labor (IZA).

Riley, R., Rosazza-Bondibene, C., and Young, G. (2015). The UK productivity puzzle 2008–13: evidence from british businesses. Staff working paper No. 531, Bank of England.

Scarpetta, S., Hemmings, P., Tressel, T., and Woo, J. (2002). The role of policy and institutions for productivity and firm dynamics: evidence from micro and industry data. Working paper No. 329, OECD Economics Department.

Szulc, B. (1964). Indices for multiregional comparisons. *Przeglad statystyczny*, 3:239–254.

Tian, G. Y. and Twite, G. (2011). Corporate governance, external market discipline and firm productivity. *Journal of Corporate Finance*, 17(3):403–417.

Törnqvist, L. (1936). The bank of Finland's consumption price index. *The Bank of Finland Monthly Bulletin*, 10:1–8.