Noname manuscript No.
(will be inserted by the editor)

Firm-level dynamics and aggregate productivity:
Australian evidence using linked employer-employee data

**Chien-Hung Chien · Robert Breunig ·
A.H. Welsh**

Received: date / Accepted: date

Abstract We estimate the contributions of entry, exit, within-firm growth and across firm re-allocation to productivity growth in Australia in 2002 - 2013. We use a novel approach for estimating labor inputs into the production function using a large, linked employer-employee dataset with over 43 million observations on firms and 130 million observations on employees. We estimate worker- and firm-specific effects using a grouping algorithm appropriate for sparse matrices. Entry and exit are the largest contributors to productivity movements across all industries. Firm exit contributes positively to productivity growth whereas firm entry contributes negatively.

Keywords productivity · firm dynamics · linked employer-employee data · production function estimation
JEL codes: C14, C55, D22, D24, J24, L20

Chien-Hung Chien
Australian Bureau of Statistics ANU Mathematical Science Institute
Tel.: +61262525917
E-mail: joseph.chien@abs.gov.au

Robert Breunig
Tel.: +61261252148
E-mail: robert.breunig@anu.edu.au A.H. Welsh
Tel.: +61261257313 E-mail: alan.welsh@anu.edu.au

1 Introduction

As Lentz and Mortensen (2010, p.2) point out, an efficient market allocates resources from less productive firms to more productive ones: *Firms are living beings; they are born, some grow to maturity, and all eventually die. Child mortality is high, the few who survive grow rapidly, but only a handful enjoy old age. New entrants are smaller and less productive on average but more diverse than continuing firms. Although size and productivity diversity diminish over time due to the selection process responsible for early death, differences in factor productivity of those that continue are still large and persistent.*

Firm dynamics—that is, how contributions from established, entering and exiting firms affect productivity—is one of the critical factors that influence aggregate productivity (Foster et al., 2001). The seminal surveys by Bartelsman and Doms (2000) and Syverson (2011) discuss the advantages of using micro-data to better understand the determinants of aggregate productivity. Aggregate statistics, which give a good overview of trends in productivity growth, do not show the variability that occurs at the firm-level. This variability is itself an important determinant of productivity growth. It is essential to develop a good understanding of the degree to which different aspects of productivity growth within and across firms contribute to differences in productivity growth across industries. Understanding these differences can help the government to create an environment for firms to innovate and improve goods and services.

In this paper, we assess the role of firm dynamics in aggregate and industry-level multi-factor productivity growth in Australia in the 2002-2013 period. To do this, we build a linked employer-employee dataset from over 43 million firm-year records and over 130 million employee-year records.[1] After linking workers to firms, we use the data on employees to create an instrument for the labor input into the production function. Many previous studies (see discussion and references below) have shown that productivity function estimates, and hence estimates of multi-factor productivity, are biased from correlation between labor inputs, which can be adjusted in real time, and unobserved determinants of productivity.

The first contribution of our paper is to use a novel approach to solving this problem. We estimate a wage equation with standard covariates and a fixed effect for each worker and each firm. Using the method of Abowd et al. (2002), we solve for the sparse matrix created by these extensive fixed effects. We then use the predicted values from the wage equation to create an estimate for the labor input into production that eliminates worker and firm fixed effects which could be correlated with unobserved determinants of productivity. To

[1] The dataset we use is different than the official linked employer-employee dataset of the Australian Bureau of Statistics, which can be found here: `https://www.abs.gov.au/about/data-services/data-integration/integrated-data/linked-employer-employee-database-leed`. Our dataset was created earlier for the purpose of this study and goes back further in time.

our knowledge, this approach has only been used once before by Maré et al. (2017).

The second contribution of our paper is to provide country-specific evidence on the dynamics of productivity growth at the national and industry level for Australia for the period 2002-2013. Our results are more up-to-date than other studies and provide information on the dynamics of productivity growth that are not provided by the Australian Bureau of Statistics. We examine the components of productivity growth by looking at firm entry and exit, reallocation across continuing firms and productivity growth within firms. We also examine whether these patterns differ across industries. We compare two different approaches to decomposing productivity changes over time. Their different treatment of entering and exiting firms does not matter much for national level productivity but does matter for some industries.

Our key finding is that entry and exit, combined, is the largest contributor to productivity movements. Combined, and on average across all industries, they explain about 61 per cent of productivity movements. In most industries, they often cancel each other out with exit contributing to positive productivity growth and entry often detracting from productivity growth. Breunig and Wong (2008) found a similar pattern when examining the Australian productivity growth experience in the 1990s. Those firms that survive, however, contribute strongly to productivity growth. Within-firm productivity growth is the largest single contributor to productivity changes, accounting for 36 per cent of productivity movements, on average across all industries. Within-firm productivity growth is generally negative. Our results suggest that government policies which create barriers to entry (such as regulation) or which prop up failing firms are likely to deter productivity growth.

This paper is structured as follows: Section 2 provides a brief literature review, Section 3 describes the scope of the data and Section 4 presents our estimation approach. We discuss our approach to identifying the wage equation in this section and outlines how we ensure the unique identification of the fixed effects in the wage equation. Section 5 summarises our empirical results. The final section provides some discussion and conclusions.

2 Literature review

The last three decades have seen an explosion of work on the impact of productivity movements at the firm level on overall, aggregate productivity. Productivity can be positively influenced in many ways: firms can become more productive, more productive firms can garner a greater share of output, low productivity firms can exit and new firms can enter who might bring higher productivity and innovation (if not immediately, perhaps in the future). A host of government policies from regulation to research and development subsidies may affect productivity and its reallocation amongst firms. Understanding these firm-level dynamics of productivity are thus key to evaluating and understanding the impact of policy on productivity.

Influential work by Olley and Pakes (1996) and Bartelsman and Dhrymes (1998) developed the principal methods that most economists use to measure the impact of firm dynamics on aggregate productivity. These methods are often used in analyses to better understand the process of creative destruction that can occur within and between sectors of the economy (Foster et al., 2001). Bartelsman and Doms (2000), in a a review of the early literature, highlighted the importance of the development of firm-level datasets across a range of countries which created the preconditions for this research.

One contribution of our paper is adding to the country-specific evidence about the role of firm dynamics in productivity growth. Lafrance and Baldwin (2011) explore the contribution that firm turnover has on productivity growth in the Canadian services industries. They find that the market naturally allocates resources from uncompetitive firms to new entrants. Brandt et al. (2009) examines Chinese manufacturing and finds that new entrants are particularly dynamic contributors to overall productivity and efficiency. Net entry contributes roughly half to TFP growth. However, aggregate productivity growth is tempered by a relatively less vigorous reallocation of inputs towards higher productivity firms compared to the US. Cincera and Galgau (2005) provide evidence for a range of European countries on the relationship between firm entry and exit and productivity growth. Pavcnik (2002) examines productivity in Chilean manufacturing and its relationship to trade liberalization. Brown et. al (2018) examine firm-level data from Chile, Columbia, Mexico and Peru. After decomposing productivity growth into within-plant growth and market allocation forces, average productivity of survivors is higher than the overall contribution of reallocation forces. During recessions the inverse is true and reallocation gives a positive, albeit small, contribution to aggregate productivity growth. Bartelsman et. al (2013) link the country-specific literature on firm-level heterogeneity in productivity to differences in aggregate performance across countries in a model that uses heterogeneous adjustment frictions and distortions to match observed data.

Previous Australian evidence is relatively scant. Breunig and Wong (2008) provided evidence for Australian productivity growth for the 1990s, showing the important role of entry and exit. Similar to this paper, they estimated productivity functions, but corrected for the endogeneity of labor inputs using the methods of Olley and Pakes (1996) and Levinsohn and Petrin (2003). Nguyen and Hansell (2014) explore the firm dynamic effects on productivity growth for Australian manufacturing and business services industries using an index number theory approach. Unlike our paper, they did not derive multifactor productivity from an estimated production function. They find that entering and exiting firms make larger contributions to overall productivity than established firms.

Economists have observed strong correlation between firm productivity and wage costs per worker (Lentz and Mortensen, 2010). However, this strong correlation can potentially cause endogeneity (Fox and Smeets, 2011). The instrumental variable approach is one of the most common ways to address endogeneity in productivity analysis (Van Biesebroeck, 2007; Eberhardt and Helmers, 2010; Del Gatto et al., 2011).

A unique aspect of our paper is in our construction of the instrumental variable for labor inputs in the firm production function. The issue of endogeneity when estimating production functions is well-documented in the literature. It refers to the problem caused by firms increasing inputs used for production in response to temporary increases in demand for their products and services (Eberhardt and Helmers, 2010). The temporary positive correlation between input and output makes it difficult to distinguish whether an increase in firm production is caused by increased firm productivity or a temporary firm response to increased demand. We discuss other solutions to this problem further in the methodology section below.

We instrument for labor inputs using predicted wages from wage equation estimates which are constructed to be uncorrelated with unobserved productivity shocks in the production function. Our method exploits a very large linked employer-employee dataset with over 130 million observations. Using a method proposed by Abowd et al. (2002) to cope with large, sparse matrices, we are able to identify worker- and firm-specific effects that determine wages (productivity). We then create an instrument using predicted wages but removing these worker- and firm-specific effects. By construction, this instrument for labor inputs is uncorrelated with firm-specific, unobserved effects on productivity such as managerial quality. To our knowledge, this approach has only been used before by Maré et al. (2017) in the context of productivity estimation. This departs from much of the previous literature which creates instruments using lagged values of capital or material inputs (Gandhi et al., 2011; Olley and Pakes, 1996; Breunig and Wong, 2008; Levinsohn and Petrin, 2003; Bakhtiari, 2015; Fox and Smeets, 2011).

The decomposition of firm-level contributions to aggregate industry productivity is derived using the approaches of Griliches and Regev (1995) and Melitz and Polanec (2015) to take into account firm dynamics. As we discuss below, these two approaches treat entering and exiting firms quite differently and thus can produce different results.

We turn now to describing the data.

3 Data sources

We develop a firm-level panel dataset for which we create linkages to a worker-level panel dataset which has been developed by the Australian Bureau of Statistics (ABS). We use both panel datasets, and the linkages, in our analysis.

We use firm-level data from the Australian Taxation Office (ATO), the Australian Business Register (ABR) and the ABS.[2] The sample period is from

[2] This study uses a strict data access protocol. Access to the datasets includes audit trails and is limited to a need-to-know basis. All ABS officers are legally bound to secrecy under the *Census and Statistics Act 1905*. Officers sign an undertaking of fidelity and secrecy to ensure that they are aware of their responsibilities. The ABS policies and guidelines govern

2001–02 to 2012–13. The data contain detailed firm characteristics from Business Income Tax filings, Business Activity Statements and the ABR. The resulting panel data set has $43,191,403$ firm-year observations on $6,846,067$ Australian Business Numbers (ABNs) for the financial years 2001–02 and 2012–13.[3] The vast majority of these firms are micro-businesses with one or no employees. We will drop these below, when we merge worker data with the firm-level data. Most firm-level variables, such as firm sales and materials costs, come from Business Income Tax or Business Activities Statements.

Information on firm industry classification comes from both the ATO and the ABS Business Register (ABSBR). Most firms have a non-missing industry classification. For a very small number of firms ($45,961$) with missing industry classification, we impute the classification using the method discussed in Chien et al. (2019a).

335 firms in 2001–02 have a missing year of incorporation. In order to be able to determine an age for these firms, which we use in our modelling below, we set the year of incorporation equal to 2000–01. This was the year that the ABN was first introduced as part of the introduction of Australia's Goods and Service Tax. For 2001–02 only, we use a variable created by the ABS available in the ABR which indicates whether a firm is continuing or a new entry. For all other years, we determine a firm's status (continuing or new entry) from our data.

When there are gaps in the data, that is when firms appear in some years and then are missing and then subsequently re-appear, we classify such firms as continuing.[4] For example, firm A has observations for 2002–03, 2004–05 and 2005–06. Firm A is classified as an entry firm in 2002–03, a continuing firm in 2004–05 and an exiting firm in 2005–06.[5] We do not include these firms in estimation during the years in which they are missing, but when we calculate and decompose industry-level productivity, we interpolate multi-factor productivity for these firms to allow them to contribute to the 'continuing firm' component.[6] We similarly interpolate the industry weights which we estimate in equation (7) below.

the disclosure of information to maintain the confidentiality of individuals and organisations. This study presents only aggregate results to ensure that they are not likely to enable the identification of a worker or a firm to comply with provisions that govern the use and release of this information, including the *Privacy Act 1988* (ABS, 2015).

[3] Australia operates a July 1 to June 30 financial year for tax purposes. ABN is not quite equivalent to a firm. Some business groups operate businesses using more than one ABN and multiple businesses can be listed under one ABN provided that they all operate under the same business structure. Our analysis is conducted at the ABN level.

[4] There are 588,177 such firms out of our final sample of just over 2 million firms.

[5] The ABS identifier for firm continuity in other years is based upon an 'Economic Statistics Units Model' (see `https://www.abs.gov.au/ausstats/abs@.nsf/dossbytitle/AC79D33ED6045E88CA25706E0074E77A?OpenDocument`). We prefer our measure based upon ABN numbers which is more narrow than the 'Enterprise Group' used by the ABS.

[6] We do not include these firm-year observations in estimation because we have no information on the firms in the years in which they are missing. An alternative would be to impute firm characteristics and include them in estimation.

Our firm panel is constructed similarly to the existing Business Longitudinal Analysis Data Environment (BLADE), see Hansell and Rafi (2018). Descriptive statistics may not match exactly to the public use version of BLADE.

Our worker panel uses data from the ABS prototype Graphically Linked Information Discovery Environment (GLIDE) (Chien and Mayer, 2015a). The worker panel contains $130,281,096$ worker-year observations from $2,028,564$ ABNs for firms. We do not include any workers from ABNs that have less than two employees, as we did for the firm panel. All of these ABNs have a matching firm record in the firm-level data. There are $13,131,074$ de-identified and encoded Tax File Numbers (DETFNs) which uniquely identify workers. We only include workers whose age is between 17 and 64, inclusive, in the years between 2001–02 and 2012–13. For example, a worker who is 64 in 2004–05 will appear in the data in that year but will not appear in 2005–06

Worker characteristics such as age, sex and occupation are taken from Personal Income Tax (PIT) filings and wage information is drawn from Pay-as-You-Go (PAYG) summaries. PAYG contains a longer time series than PIT, so this study backcasts sex, age and skills in the PIT data so that it is the same length as the PAYG data. The earliest available PIT information is used to backcast sex (holding it constant) and age (by subtracting 1 year). The ABS's Australian and New Zealand Standard Classification of Occupations is used to convert occupations into a 5-point skills categorical variable for the analysis (ABS, 2009). Two methods to backcast the skill categories for workers were explored: using the average over all available observations or using the earliest available information for each worker. Workers become more skilled over time so using the first approach, the average skill level held constant over time, inflates the worker's skill level over the backcast period. Hence, we used the earliest observed skill level for each worker and backcast that.

We add aggregate information about workers to the firm-level panel dataset. We do this by aggregating the worker panel to the firm-level to derive worker-level variables for each ABN. The employee counts come from PAYG records, which contain both ABNs and DETFNs. We count the total number of DETFNs for each ABN in each year. We generate a variable for this total number of employees (from the count of DETFNs) for each firm. Note that this is different than full-time equivalent (FTE) employees. PAYG records do not include hours worked, only remuneration, and we are thus unable to distinguish between full-time and part-time workers. Across all years, 6% of firms are available only in one of the two data sets (PAYG and PIT). This difference could be driven by different reporting periods for the two datasets or differences in the timing of payment of salaries and reporting of the income in the personal tax system.

We follow ABS (2015) and Chien and Mayer (2015b) and use a similar strategy to create links between the firm-level and worker-level panel datasets. We use ABNs to deterministically link firm-level records to worker-level variables, identified by DETFNs, through the PAYG records, which contain both ABNs and DETFNs. The linking process is high quality because ABNs and DETFNs are crucial to tax compliance and are thus well-managed by the ATO, ABR and ABS. The final panel dataset that we use for analysis includes worker and employer information and has $10,039,638$ firm-year observations containing $2,028,564$ ABNs between 2001–02 and 2012–13. We impute missing values for observations which have missing data in any of the variables that we use in our analysis such as materials or capital costs (see Chien et al. (2019a) for details). Results from the dataset which includes imputed values, which we refer to below as the 'imputed dataset', match ABS official productivity statistics much better than the dataset which results from dropping all firms with missing items. The analysis of the complete case data involves dropping 80 per cent of the data which produces highly volatile year-on-year changes in aggregate productivity and estimates that are inconsistent with ABS aggregate productivity results.[7] Therefore, we prefer the dataset that includes the imputed values. However, we conduct all our analysis on both the imputed dataset and the complete case dataset and present those results below.

Table 1 shows the percentage of firms who fall into different categories of firm size and years in the sample. Large firms, that is, firms with more than 200 employees size, are more likely to be in the sample for a longer period.

Table 1: Firm size and years in sample

Firm size	**years in sample**											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**>10**	**Total**
1 to 4	3.9	4.9	5.7	7.1	5.2	5.3	5.1	4.7	4.8	5.0	14.7	66.4
5 to 19	0.2	0.5	0.9	1.8	1.4	1.6	1.7	1.7	1.8	2.1	11	24.7
20 to 199	0.0	0.1	0.1	0.5	0.4	0.4	0.5	0.5	0.5	0.7	4.6	8.3
200 plus	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.6
Total	4.1	5.5	6.8	9.5	7.0	7.3	7.3	6.8	7.1	7.8	30.9	100.0

Table entries are percentage of all firms in the size/years in sample combination

Tables $A1$ and $A2$ in Appendix A show summary statistics for the firm- and worker-level panels. The data for labor inputs ($\ln L_{jt}$), the cost of capital ($\ln K_{jt}$), material costs ($\ln M_{jt}$), firm age, and the natural log of average wages ($\ln w_{jt}$) come from the firm-level information. Their construction is described in more detail below and in the Appendix. $\widehat{z}_{jt}$ is estimated labor inputs derived from the worker level data and is described in subsection 4.1.

The summary statistics for the firm panel show broad consistency between the complete case and the imputed datasets. This is also the case for the correlation analysis presented in Tables 2 and 3 which compares the correlation coefficients for the key variables of our study in the complete case and imputed datasets. These tables show that the correlation coefficients are con-

[7] For the vast majority of the firms with missing values, imputation usually involves imputing only one or two data items. Materials and capital are the variables most often missing, about 25 per cent of the time. Output and wages are missing in about 10 per cent of cases. Industry, year, number of employees and firm age are never missing. The length and breadth of our dataset provides a very strong base from which to do imputation.

sistent when we compare the complete case and imputed datasets. The only remarkable difference is in the correlation between the natural log of wages lnw_{jt} and the estimated labor component $\widehat{z}_{jt}$ which is small but negative in the complete case data and small but positive in the data which uses the imputed values.

Table 2: Pearson Correlation Coefficients – complete case dataset

	lny_{jt}	lnK_{jt}	lnM_{jt}	lnAge_{jt}	ln$\widehat{z}_{jt}$	lnw_{jt}
lny_{jt}	1	0.5206***	0.58389***	0.13821***	0.01096***	0.72011***
lnK_{jt}		1	0.4829***	0.10082***	-0.12324***	0.50686***
lnM_{jt}			1	0.12111***	0.02614***	0.62887***
lnAge_{jt}				1	-0.42234***	0.15778***
ln$\widehat{z}_{jt}$					1	-0.0151***
lnw_{jt}						1

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3: Pearson Correlation Coefficients – imputed dataset

	lny_{jt}	lnK_{jt}	lnM_{jt}	lnAge_{jt}	ln$\widehat{z}_{jt}$	lnw_{jt}
lny_{jt}	1	0.53644***	0.55506***	0.13707***	0.0238***	0.7391***
lnK_{jt}		1	0.4823***	0.10351***	-0.11546***	0.5301***
lnM_{jt}			1	0.07288***	0.0215***	0.5616***
lnAge_{jt}				1	-0.42192***	0.137***
ln$\widehat{z}_{jt}$					1	0.0176***
lnw_{jt}						1

Note: *p<0.1; **p<0.05; ***p<0.01

Figure $B1$ in Appendix B shows yearly firm entry and exit rates over the sample period. Exit rates are higher after 2010 were except for the finance industry in 2007–08. This is most likely due to the Great Recession causing higher firm exit rates in the finance industry.

4 Methodology

We begin by estimating a wage equation for workers which controls for observable worker characteristics and unobservable characteristics of firms and workers. We use predicted values from this equation to instrument for the labor input in a firm-level production function. We use the predictions from the worker-level wage equation without the firm- and worker-specific effects. The predicted value thus captures what a 'typical' worker with a given set of characteristics averaged across all workers and firms would contribute to firm productivity. This removes any endogeneity in the quality of labor inputs which might be correlated with unobservable characteristics of the firm, such as managerial quality. It also removes endogeneity which might arise from positive spillovers between high-quality workers who are clustered in the same firm. It removes a further source of endogeneity if high quality workers are attracted to higher productivity firms.

We estimate production functions separately by industry and derive firm-, industry- and economy-level productivity from these industry-specific estimates. We decompose productivity changes to calculate the contributions from entering, exiting and continuing firms to aggregate productivity.

4.1 Wage equation

This analysis uses a modified wage equation adapted from Abowd et al. (2002). The worker panel is unbalanced, meaning that the available observations for each worker i, $i = 1, \ldots, N$, can be different. Suppose that the observations for worker i are available at time $t = 1, \ldots, T_i$ such that $t = 1$ is the first time period and $t = T_i$ is the last time period for the available observations for worker i. Note that there can be gaps. A worker might appear in periods 1 and 3 but not in period 2, for example.

We model w_{it}, the wages for worker i at time t, as:

$$\ln w_{it} = \boldsymbol{x}_{it}^{\intercal}\boldsymbol{\alpha} + \theta_i + \boldsymbol{f}_{it}^{\intercal}\boldsymbol{\psi} + \epsilon_{it}, \tag{1}$$

where $\boldsymbol{x}_{it}$ is a p-vector of the characteristics of worker i at time t, $\boldsymbol{\alpha}$ is a p-vector of unknown coefficients of worker characteristics, θ_i represents unobserved (time-invariant) worker effects and the components of the J-vector $\boldsymbol{\psi} = (\boldsymbol{\psi}_1, \cdots, \boldsymbol{\psi}_J)^{\intercal}$ represent firm effects (e.g., specific factors such as pay structure that affects workers' wages and firm-specific characteristics which influence worker productivity). Note that firms are indexed by j from $j = 1, \ldots, J$. $\boldsymbol{f}_{it}^{\intercal} = (f_{i1t}, \cdots, f_{iJt})^{\intercal}$ is a firm indicator vector with components

$$f_{ijt} = \begin{cases} 1, & \text{if worker } i \text{ works for firm } j \text{ at time } t \\ 0, & \text{otherwise,} \end{cases}$$

and the unobservable components ϵ_{it} are assumed to satisfy $\epsilon_{it} \overset{iid}{\sim} D(0, \sigma^2)$.

Worker characteristics include: $Sex = 1$ indicates if the worker i is male and 0 otherwise; $HighSkill = 1$ if worker i has a tertiary qualification and 0 otherwise; $MediumSkill = 1$ if worker i has at most a diploma-level qualification and 0 otherwise; $WorkingSkill = 1$ if worker i has at most a Certificate III qualification and 0 otherwise. Workers with qualifications lower than a Certificate III qualification are treated as the baseline and included in the intercept. $Time$ is represented by 11 time indicator variables: one for each year with 2001–02 as a baseline. The variable Age, the age of worker i at time t, is fitted by a quartic polynomial including linear, quadratic, cubic and quartic terms. We found that a quartic function was superior to a quadratic function in describing the decline in workers' wages as they age. We also include interaction terms $Sex : Poly(Age, 4)$ between Sex and Age and $Sex : Time$ between Sex and $Time$. $\boldsymbol{x}_{it}^{\intercal}$ thus contains $p = 34$ terms.

This paper uses the methodology proposed by Abowd et al. (2002) to estimate a wage equation with high dimensionality. Abowd et al. (1999) highlight the challenges of fitting model (1) due to the large number of workers and firms. Their US study uses data on over one million workers and more than $50,000$ firms. As indicated above, our dataset contains over two million firms and over 130 million worker-year observations.

The method of Abowd et al. (2002) uses a conjugate gradient algorithm to find a solution to equation (1)–specifically for the coefficients on worker and firm fixed effects which suffer from being very sparse. This solution may not be a global maximum, however. We then use the grouping methodology of Abowd et al. (2002) which constructs groups of firms based upon worker movements between firms. This additional source of information allows identification of the fixed effects for workers and firms. Please see Appendix D for a detailed discussion of our estimation approach.

We estimate equation (1) pooling across all workers at all time periods in all industries. We then derive an instrument for firm-specific labor inputs, which we use in equation (3) below. The instrument is constructed from the average fitted values from equation (1) summed for each firm j. Specifically, let $\widehat{\boldsymbol{\alpha}}$, $\widehat{\theta}_1, \ldots \widehat{\theta}_N$ and $\widehat{\boldsymbol{\psi}}$ denote estimates of parameters in (1). Note that the estimated person and firm effects from (1) are correlated with firm productivity because firms with higher quality workers or better management practices are likely to be more productive. The instrument removes these potential sources of endogeneity. The instrumental variable at the firm level is:

$$\widehat{z}_{jt} = \boldsymbol{x}_{jt}\widehat{\boldsymbol{\alpha}} \quad \text{where} \tag{2}$$

$$\boldsymbol{x}_{jt} = \sum_{i=1}^{N} f_{ijt}\boldsymbol{x}_{it}$$

Note that the variables in equation (2) now have a firm subscript, j, to reflect the summing of worker effects within each firm j. Below, when we want to emphasise that firm j belongs to industry k, we also include the industry subscript k so that the estimated firm-average worker effect for a firm j in industry k is denoted as $\widehat{z}_{jkt}$.

The instrument will be uncorrelated with unobserved factors in the firm production function (ε_{jkt} in equation (3) below) but correlated with the labor inputs (measured through wages, w) of the firm.

$$\mathrm{Cov}(\widehat{z}_{jkt}, \varepsilon_{jkt}) = 0 \text{ and } \mathrm{Cov}(\widehat{z}_{jkt}, w_{jkt}) \neq 0.$$

At the firm level, there is no instrument for firm-level labor inputs which would allow for standard two-stage least squares estimation. The detailed information about workers combined with their movements across firms over time produces an instrument which meets the two required conditions for a valid instrument–it is correlated with the firm's labor input but uncorrelated with the unobserved productivity movements which are correlated with a firm's labor input.

4.2 Firm-level productivity model

The volume of firm outputs can be modelled as functions of observed inputs such as capital, materials and labor in volume terms, and unobserved components in the production process (Fox and Smeets, 2011). Firm outputs, y_{jkt}, are defined as sales adjusted for repurchase of stock and are deflated by industry gross value added implicit price deflators for industry k at time t (ABS, 2018b). We use a Cobb–Douglas production function, similar to Breunig and Wong (2008) and Maré et al. (2017):

$$\begin{aligned}\ln y_{jkt} = \beta_{0k} + \beta_{1k}\ln L_{jkt} + \beta_{2k}\ln K_{jkt} + \beta_{3k}\ln M_{jkt} + \\ \beta_{4k}\ln Age_{jkt} + \boldsymbol{\tau}_k + \varepsilon_{jkt},\end{aligned} \tag{3}$$

where $\ln L_{jkt}$ is the logarithm of labor inputs (wages paid by the firm) deflated by *Wage Price Index* and $\ln K_{jkt}$ is the logarithm of the firm's cost of capital, which includes depreciation, capital rental expenses and capital work deductions, deflated by implicit price deflators based upon industry consumption of fixed capital (ABS, 2018b). The logarithm of material costs $\ln M_{jkt}$ include the inputs used in the production, deflated by *Producer Price Indexes: Intermediate Goods* (ABS, 2018d). The logarithm of firm age is $\ln Age_{jkt}$. We also include different intercepts β_{0k} for each industry and a vector of time-fixed effects which differ by industry, $\boldsymbol{\tau}_k$. ε_{jkt} captures unobserved, firm-level multifactor productivity. If the ε_{jkt} are uncorrelated with all of the right-hand side variables in equation ((3)), then unbiased coefficients for the Cobb–Douglas production function can be estimated by an ordinary least squares regression. This assumption is unlikely to hold in reality, particularly for labor inputs which can be adjusted in real time to adjust to firm needs.

The endogeneity caused by simultaneous determination of firm productivity and the amount of labor inputs used by a firm causes bias in estimating the production function equation (3). To mitigate the bias, many studies use predicted values from instrumental variables equations—that is, using lagged inputs as instruments for the current labor inputs (Gandhi et al., 2011). For example, Olley and Pakes (1996) and Breunig and Wong (2008) use lagged capital investment, Levinsohn and Petrin (2003) and Bakhtiari (2015) use lagged material inputs and Fox and Smeets (2011) uses lagged wage costs as instrumental variables. Instead of using lagged values, we construct an instrument for labor inputs, $\widehat{z}_{jkt}$, from equation (2). By construction, this removes components in equation (1) that can correlate with firm multi-factor productivity ε_{jkt} in (3). We follow the convention in the literature which is to assume that the non-labor components are pre-deterined at time t so all endogeneity comes through the correlation between the labor input and the unobserved multi-factor productivity term.

We fit separate models for each industry. We include k to emphasise the nesting of firm j in industry k. This specification restricts the same production technology within industries but varies across industries. Our estimation model, with the instrument included, is:

$$\begin{aligned}\ln y_{jkt} = \beta_{0k} + \beta_{1k}\widehat{z}_{jkt} + \beta_{2k}\ln K_{jkt} + \beta_{3k}\ln M_{jkt} + \\ \beta_{4k}\ln Age_{jkt} + \boldsymbol{\tau}_k + \varepsilon_{jkt}.\end{aligned} \tag{4}$$

The estimated parameters from (4) include the industry intercepts $\widehat{\beta}_k$, labor inputs $\widehat{\beta}_{1k}$, cost of capital $\widehat{\beta}_{2k}$, materials costs $\widehat{\beta}_{3k}$, firm age $\widehat{\beta}_{4k}$, industry-specific time-fixed effects $\widehat{\boldsymbol{\tau}}_k$ and the residual from the regression, $\widehat{\varepsilon}_{jkt}$.

Firm multi-factor productivity is the ratio of output to measured inputs normalised relative to the mean of industry k. Estimated firm multi-factor productivity from equation (4) is given by:

$$\widehat{mfp}_{jkt} = \widehat{\boldsymbol{\tau}}_k + \widehat{\varepsilon}_{jkt}. \tag{5}$$

In order to generate firm-level weights that will allow us to sum our productivity estimates to the industry- and economy-wide levels, we first estimate a model where we pool all firms and ignore any differences across industry. This method of weighting is sometimes called the composite-input shares approach and has been used by others including: Bartelsman and Dhrymes (1998); Van Biesebroeck (2008); Balk (2018); Dias and Marques (2021); Dosi et al. (2021).

$$\begin{aligned}\ln y_{jt} = \beta_0 + \beta_1\widehat{z}_{jt} + \beta_2\ln K_{jt} + \beta_3\ln M_{jt} + \\ \beta_4\ln Age_{jt} + \boldsymbol{\tau} + \varepsilon_{jt},\end{aligned} \tag{6}$$

We use (6) to define firm-level weights $\widehat{\omega}_{jkt}$ as:

$$\widehat{\omega}_{jkt} = \widehat{\beta}_1\widehat{z}_{jkt} + \widehat{\beta}_2\ln K_{jkt} + \widehat{\beta}_3\ln M_{jkt} + \widehat{\beta}_4\ln Age_{jkt}. \tag{7}$$

We normalise $\widehat{\omega}_{jkt}$ to sum to one within each industry and refer to these normalised weights as $\widehat{\omega}^*_{jkt}$ in what follows.

4.3 Industry productivity

We follow Maré et al. (2017) and define the aggregate productivity index A_{kt} for industry k at time t as:

$$A_{kt} = \sum_{j=1}^{J_{kt}} \widehat{\omega}^{*}_{jkt} \widehat{mfp}_{jkt}, \tag{8}$$

where J_{kt} is the number of firms in industry k at time t.
For decomposing productivity at the economy-wide level, let $\widehat{\omega}_{kt} = \sum_{j=1}^{J_{kt}} \widehat{\omega}_{jkt}$ and $\widehat{mfp}_{kt} = \sum_{j=1}^{J_{kt}} \widehat{mfp}_{jkt}$. Then the aggregate productivity index A_t, for all industries at time t, is:

$$A_t = \sum_{k=1}^{K_t} \widehat{\omega}^{**}_{kt} \widehat{mfp}_{kt}, \tag{9}$$

where we use the normalised weights $\widehat{\omega}^{**}_{kt} = \dfrac{\widehat{\omega}_{kt}}{\sum_{k=1}^{K} \widehat{\omega}_{kt}}$,

and K is the number of industries (which does not vary over the time period of our sample). Note the weights $\widehat{w}^{**}_{kt}$ sum to one in each time period t.
Griliches and Regev (1995) propose decomposing the changes in aggregate industry-level productivity from time $t-1$ to t into contributions from surviving (S), entering (EN) and exiting (EX) firms as:

$$\Delta A_{kt} = W_{kt} + B_{kt} + EN_{kt} + EX_{kt}, \tag{10}$$

$$\begin{aligned}
\text{where}\quad W_{kt} &= \sum_{j \in S_{kt}} \overline{\widehat{\omega}}^{*}_{jk} \Delta \widehat{mfp}_{jkt},\\
B_{kt} &= \sum_{j \in S_{kt}} \Delta \widehat{\omega}^{*}_{jkt} (\overline{\widehat{mfp}}_{jk} - \overline{A}_k),\\
EN_{kt} &= \sum_{j \in EN_{kt}} \widehat{\omega}^{*}_{jkt} (\widehat{mfp}_{jk} - \overline{A}_k) \text{ and}\\
EX_{kt} &= -\sum_{j \in EX_{kt}} \widehat{\omega}^{*}_{jkt-1} (\widehat{mfp}_{jkt-1} - \overline{A}_k).
\end{aligned}$$

Letting Δ represent change, $\Delta A_{kt} = A_{kt} - A_{kt-1}$ is the change in aggregate productivity for industry k from time $t-1$ to time t. Bars represent averages between t and $t-1$, so $\overline{\widehat{\omega}}^{*}_{jk} = \frac{(\widehat{\omega}^{*}_{jkt} + \widehat{\omega}^{*}_{jkt-1})}{2}$ and $\overline{A}_k = \frac{(A_{kt} + A_{kt-1})}{2}$. The definitions of surviving $j \in S_{kt}$, entering $j \in EN_{kt}$ and exiting firms $j \in EX_{kt}$

are based on firm transitions on an annual basis over the observed sample period. Survivors are firms operating in t and $t-1$, exiting firms are firms that exist at time $t-1$ but not at time t and entering firms are firms that did not exist at time $t-1$ but did at time t. The contribution of the surviving firms is decomposed into two components: the within-industry reallocation W_{kt}, which measures the change in firm productivity weighted by the average of the weights at t and $t-1$ (i.e., $\overline{\widehat{\omega}}^*_{jk}$) and the between-industry reallocation B_{kt}, which measures deviations from the average productivity (i.e., $\overline{A}_k$) including the impact of firm entry and exit (Foster et al., 2001).

Fox and Smeets (2011) and Nguyen and Hansell (2014) discuss the importance of using appropriate benchmarks to calculate the contributions of surviving, entering and exiting firms to aggregate productivity. The simple average of industry-level productivity at times t and $t-1$, used by Griliches and Regev (1995), has come under some criticism as a benchmark.

To check the robustness of our productivity decomposition to alternative benchmarks, our study also uses an alternative method proposed by Melitz and Polanec (2015). Their dynamic decomposition extends a decomposition proposed by Olley and Pakes (1996) to account for firm entry and exit. They define the within contribution as the unweighted mean change in the productivity of surviving firms and between contributions as the change in the covariance between market share and firm productivity for surviving firms. Their method uses surviving firms at time t as a benchmark for entering firms and surviving firms at time $t-1$ as a benchmark for exiting firms.[8],

We can re-write industry-level productivity as:

$$\begin{aligned} A_{kt} &= \overline{\widehat{mfp}}_{kt} + \frac{1}{(J_t-1)}\sum_{j=1}^{J_t}(\widehat{\omega}^*_{jkt} - \overline{\widehat{\omega}^*}_{kt})(\widehat{mfp}_{jkt} - \overline{\widehat{mfp}}_{kt}) \qquad (11) \\ &= \overline{\widehat{mfp}}_{kt} + \mathrm{Cov}(\widehat{\omega}^*_{jkt}, \widehat{mfp}_{jkt}), \end{aligned}$$

where $\overline{\widehat{mfp}}_{kt}$ and $\overline{\widehat{\omega}^*}_{kt}$ are the weighted averages of these variables for the J_t firms in industry k at time t using weights $\widehat{\omega}^*_{jkt}$.

The dynamic Olley–Pakes approach decomposes aggregate productivity into contributions from surviving, entering and exiting firms as:

$$\Delta A_{kt} = W^*_{kt} + B^*_{kt} + EN^*_{kt} + EX^*_{kt}, \qquad (12)$$

[8] We also explored the productivity decomposition method proposed by Foster et al. (2001). We do not report them as the results are similar to those for the approach of Griliches and Regev (1995).

$$\begin{aligned}\text{where}\quad W^*_{kt} &= \Delta\overline{P}_{kt}, \qquad B^*_{kt} = \Delta\mathrm{Cov}_{kt},\\ EN^*_{kt} &= \sum_{j\in EN} \widehat{\omega}^*_{jkt}(A_{jkt\in EN} - \overline{A}^{(t\in S)}_{jk}) \text{ and}\\ EX^*_{kt} &= \sum_{j\in EX} \widehat{\omega}^*_{jkt}(A_{jkt\in EX} - \overline{A}^{(t-1\in S)}_{jk}).\end{aligned}$$

and where $\Delta\overline{P}_{kt}$ is the unweighted change in mfp of surviving firms and $\overline{A}^{(t\in S)}_{jk})$ and $\overline{A}^{(t-1\in S)}_{jk})$ is the weighted average productivity of surviving firms at time t and $t-1$, respectively.[9]

This decomposition approach uses only surviving firms at time t and surviving firms at time $t-1$ as benchmarks for entering and exiting firms which is arguably more appropriate. We discuss the impact that this has on the results in Section 5. One advantage of using these benchmarks is that entering firms will only generate positive productivity growth when they have higher productivity than surviving firms at time t. Similarly, exiting firms can only generate a positive contribution if they have lower productivity than surviving firms at time $t-1$.

5 Empirical results

5.1 Firm dynamics and aggregate productivity

Figure 1 and Table 4 show the estimated contributions from surviving, entering and exiting firms to aggregate annual productivity using the methods of Griliches and Regev (1995) and Melitz and Polanec (2015) as discussed above. These two methods use different counterfactuals in the productivity decomposition. These different counterfactuals mostly matter for the exiting and entering firms, and in Figure 1 we can see that there are some small differences in the contribution of exiting and entering firms to productivity in the different years. However, the overall impression from the two methods is very much the same when we aggregate all industries. As we will see below, this decomposition matters more when we consider industry-level multi-factor productivity.

Within-firm productivity changes are the most important contributor to overall productivity movements accounting for 35 per cent when we use the method of Griliches and Regev (1995) and 36 per cent when we use that of Melitz and Polanec (2015).[10] This is followed by exit (about 33 per cent in both methods) and then entry (28 to 30 per cent). Exiting firms generally contribute

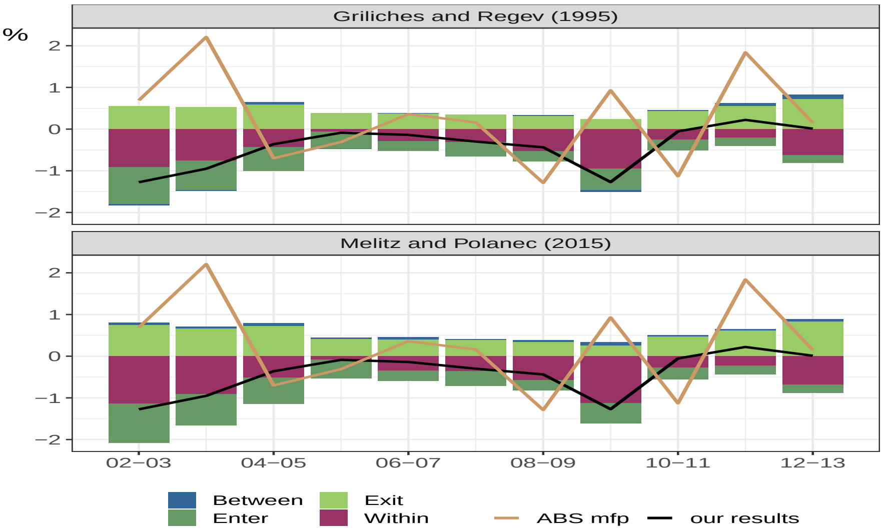
[9] The authors are happy to share their R code for producing these two decompositions. Please contact us via email.

[10] These percentages are calculated taking the absolute value of the numbers in the last row of Table 4 and looking at the percentage of the total of the absolute values.

positively to productivity, consistent with less productive firms ceasing operation. Entering firms generally contribute negatively to productivity, consistent with there being an advantage for incumbents in industry-specific knowledge about regulation, the marketplace and other factors. Between-firm changes in productivity only make very small contributions (1 to 3 per cent) to overall productivity changes. Breunig and Wong (2008) found the same ordering and similar magnitudes for productivity changes in Australia in the 1990s.

Figure 1 shows that our results are different than the published ABS annual productivity measures in about half the years of our sample. We find larger productivity growth in 2002-03, 2003-04, 2009-10 and 2011-12 than the ABS statistics. We find smaller growth in 2008-09 and 2010-11. What is most notable about our estimates is that that are much more volatile on a year-on-year basis than the official ABS productivity estimates. There are a variety of reasons why our estimates might be different than the official ABS estimates. First, ABS productivity measures are based off of a survey of a small number of firms whereas our results are close to the full population. The ABS also include sole traders and small businesses, which we exclude. The differences may reflect that we use different prices to derive the volume measures of inputs and outputs, which introduces differences in relative prices when estimating firm productivity. These differences can lead to different productivity estimates (see Dumagan and Balk, 2016 and Duarte and Restuccia, 2017 on the role of relative prices in estimating productivity.) Also, we use firm-level capital costs as a proxy for firm-level capital stock measures for our analysis because there is no information on firm-level asset prices. This information would be required to derive capital stock measures using the perpetual inventory method (Walters and Dippelsman, 1986).

Fig. 1: All industry decomposition



Note. Between and Within are the contributions from surviving firms, and Enter and Exit are the contributions from entering and exiting firms to the aggregate productivity derived from our estimates. The derivation of these measures can be found in (10) and (12) Griliches and Regev (1995); Melitz and Polanec (2015). ABS mfp are the published ABS Estimates of Industry Multi-factor Productivity (ABS, 2013).

Table 4: All industry decomposition results

Griliches and Regev (1995)					Melitz and Polanec (2015)				
year	**between**	**enter**	**exit**	**within**	**year**	**between**	**enter**	**exit**	**within**
02-03	-0.0317	-0.8856	0.5525	-0.9094	02-03	0.0558	-0.9423	0.7454	-1.1333
03-04	-0.0213	-0.7071	0.5273	-0.7497	03-04	0.0469	-0.7528	0.6678	-0.9127
04-05	0.0610	-0.5725	0.5873	-0.4365	04-05	0.0600	-0.6277	0.7254	-0.5183
05-06	-0.0097	-0.4052	0.3833	-0.0562	05-06	0.0386	-0.4478	0.4100	-0.0886
06-07	0.0041	-0.2298	0.3741	-0.2867	06-07	0.0484	-0.2467	0.4018	-0.3418
07-08	-0.0010	-0.3388	0.3520	-0.3116	07-08	0.0236	-0.3587	0.3905	-0.3547
08-09	0.0176	-0.2401	0.3130	-0.5269	08-09	0.0444	-0.2419	0.3368	-0.5757
09-10	-0.0455	-0.5150	0.2329	-0.9427	09-10	0.0762	-0.4861	0.2612	-1.1215
10-11	0.0330	-0.2598	0.4264	-0.2542	10-11	0.0261	-0.2800	0.4740	-0.2748
11-12	0.0763	-0.1961	0.5481	-0.2070	11-12	0.0400	-0.2107	0.6142	-0.2222
12-13	0.0981	-0.1892	0.7237	-0.6213	12-13	0.0609	-0.1957	0.8292	-0.6832
Average	0.0164	-0.4127	0.4564	-0.4820	Average	0.0474	-0.4355	0.5324	-0.5661

Appendix $C.2$ presents figures for the industry-by-industry decompositions across the 13 years of our data. Table $C11$ shows the averages from those figures across all years for each industry.[11] The importance of having an ap-

[11] Numbers for each industry year-by-year are available from the authors upon request.

propriate benchmark for exiting and entering firms becomes more apparent when we look at individual industries. Particularly for the smaller industries, we see larger differences in the results using the methods of Griliches and Regev (1995) and Melitz and Polanec (2015) than we did in the decompositions of all firms. For example, for Mining (B), Telecommunications (J), Arts and Recreation (R) and Other Services (S), the contribution of exiting and entering firms is quite different using the two methods.

Unlike Nguyen and Hansell (2014), we find that the net contribution from entering and exiting firms is similar in manufacturing and service industries. The within-industry contribution component is also broadly similar, at least on average across all service industries. At the industry level, our results show similar patterns with the published ABS industry-level productivity results, particularly for the Agriculture, Forestry and Fishing (A), Construction (E), Financial and Insurance Services (K) and Administrative Service industries. The industries with notable differences are Mining (B) and Electricity, Gas and Water (D), especially in 2012–13. The difference may be caused by our use of different price deflators and the methods we use to derive capital measures.

Overall, and across all industries, the combination of firm entry and exit is the dominant contribution to year-on-year productivity changes. Our study highlights this important aspect of the dynamics of productivity change that is not apparent from the national statistics.

5.2 Firm-level model results

Table C1 presents the estimation results for the firm-level production function estimated by ordinary least squares (OLS) and using our instrumental variable approach as discussed in subsection 4.1. These correspond to equations (3) and (4), respectively. In Table C1, we show results for all industries pooled together. We present results for both the *Complete case* and *Imputed* datasets. Recall that using only the complete case data results in a very large drop in sample size as reflected in Table C1. As discussed above in Section 3, we prefer the full dataset that uses the imputed values. Interestingly, despite the nearly 500 per cent increase in sample size, the results do not change by much.

For all industries, the estimated coefficient on the firm's wage bill is 0.723 in the complete case data and 0.706 when using the full data with imputed values (in ALL.OLS columns). When we use the instrument, $\widehat{z}_{jkt}$, we find smaller effects, 0.691 in the complete case data and 0.648 in the full data set with the imputed values (in ALL.2SLS columns). As previously discussed, we expect the OLS estimates to be biased because of the correlation between labor inputs and unobserved factors in firm productivity. The fact that we get smaller estimates of the impact of labor after using the instrumental variable is consistent with similar studies using instrumental variables to correct for endogeneity (see Breunig and Wong, 2008, Olley and Pakes, 1996, Pavcnik, 2002 and Levinsohn and Petrin, 2003).

We find slightly smaller coefficients on the labor imput when using the imputed data. Recall that the complete case data are quite selected. The imputation of missing data for capital and material and inclusion of all of these observations should reduce bias in the coefficient on labor. In this case, it appears that the selected sample produces an upwards bias in the labor coefficient.

Appendix C presents results estimated by individual industry. In all industries, we find a similar pattern. Instrumenting for labor produces smaller coefficients on the labor inputs to firm output, consistent with bias in that coefficient caused by simultaneity between the choice of labor inputs and other determinants of firm productivity. In the industry-by-industry results we find greater variability between the complete case and imputed datasets. We prefer the full data with imputed values. These results demonstrates the importance of correcting for endogeneity in estimating firm-level production functions. The differences in the estimated labor input can be quite large.

Table 5: Results for all industries

	Complete cases		*Imputed*	
	ALL.2SLS	ALL.OLS	ALL.2SLS	ALL.OLS
$\ln\widehat{z}_{jkt}$	0.691^{***}		0.648^{***}	
	(0.001)		(0.0003)	
$\ln w$		0.723^{***}		0.706^{***}
		(0.001)		(0.0003)
$\ln K$	0.223^{***}	0.251^{***}	0.246^{***}	0.239^{***}
	(0.001)	(0.001)	(0.0002)	(0.0002)
$\ln M$	0.242^{***}	0.183^{***}	0.238^{***}	0.170^{***}
	(0.0004)	(0.0005)	(0.0002)	(0.0002)
$\ln Age$	0.120^{***}	0.054^{***}	0.159^{***}	0.060^{***}
	(0.001)	(0.001)	(0.0005)	(0.0005)
Observations	2,296,984	2,296,984	10,039,638	10,039,638
Adjusted R^2	0.992	0.992	0.990	0.990

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

5.3 Worker-level model results

In this subsection we discuss the results from the worker-level wage equation (1) which we estimated in the first stage. As discussed above, it is essential to include firms connected by workers to uniquely identify worker and firm effects (Abowd et al., 1999). Table 6 shows the pattern of workers who have different employers in the sample. The columns indicate the number of years that a worker stays in the sample, and the rows correspond with the number of employers that workers have over the 11 years of data. It is increasingly likely for workers to work for more employers when they stay in the sample for longer. There are significant worker movements between firms in the sample. Only 23.27% of workers have one employer over the 11-year period.

Table 6: Number of job changes and number of years in the sample

number of employers	**number of years in sample**											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**>10**	**Total**
1	3.54	3	2.12	5.63	0.83	0.74	0.66	0.62	0.72	0.87	4.53	23.27
2	1	2.17	1.79	3.69	1.07	0.97	0.86	0.8	0.88	1.03	4.57	18.82
3	0.33	1	1.16	2.22	1.04	0.99	0.91	0.85	0.92	1.02	4.13	14.56
4	0.12	0.46	0.61	1.26	0.81	0.87	0.84	0.81	0.88	0.95	3.5	11.11
5	0.05	0.21	0.32	0.7	0.54	0.66	0.7	0.72	0.78	0.83	2.87	8.37
6	0.02	0.1	0.16	0.39	0.34	0.46	0.54	0.59	0.66	0.69	2.29	6.24
7	0.01	0.05	0.09	0.22	0.2	0.31	0.39	0.45	0.53	0.57	1.8	4.63
8	·	0.02	0.05	0.12	0.12	0.2	0.28	0.34	0.41	0.44	1.4	3.4
9	·	0.01	0.02	0.07	0.07	0.13	0.19	0.25	0.32	0.35	1.07	2.48
10	·	0.01	0.01	0.04	0.04	0.08	0.13	0.18	0.24	0.26	0.82	1.81
>10	·	0.01	0.02	0.07	0.07	0.15	0.28	0.44	0.66	0.81	2.81	5.31
Total	5.07	7.04	6.36	14.39	5.14	5.57	5.78	6.03	6.99	7.83	29.81	100

Note. Number of employers measures how many unique ABN a worker i has over the sample period and number of years in sample measures how many unique year counts a worker i has in the sample.

Table 7 shows the correlation structure of the estimated components from the wage equation (1). We find a positive correlation between worker and firm effects, in line with the finding of Iranzo et al. (2008), but different from Abowd et al. (2002). Andrews et al. (2008) suggest that the negative correlation in previous studies may arise from a lack of worker mobility, which is not the case in Australia (see Table 6).

Table 7: Pearson correlation coefficients of estimated components

	$logL$	θ	ψ	$\boldsymbol{X\alpha}$	ϵ
$logL$	1	0.3063***	0.5490***	–0.2115***	0.5923***
θ		1	0.1058***	–0.9793***	–0.0085***
ψ			1	–0.0966***	–0.0021***
$\boldsymbol{X\alpha}$				1	–0.0267***
ϵ					1
Note *p¡0.1; **p¡0.05; ***p¡0.01					
Note:					*p<0.1; **p<0.05; ***p<0.01

6 Discussion and conclusions

This study shows the usefulness of large, longitudinal linked employer-employee datasets in deriving estimates of firm-level contributions to aggregate productivity. Rather than relying on lagged measures of capital or material costs to instrument for a firm's labour input, we use contemporaneous wage equation estimates that are purged of any firm- or worker-level effects which might be correlated with unobserved elements of firm productivity. The exclusion restriction required in this case seems more intuitively plausible than the functional form assumptions used by studies which instrument using lagged variables of other inputs.

Using the instrument from the wage equation in the production function, we estimate productivity at the national and industry level for the period 2002-2013. We then use two different decomposition approaches, which differ primarily in how they treat entering and exiting firms, to examine the contributions that entering, exiting and surviving firms make to aggregate productivity.

Across all industries, we find that firm exit and within-firm changes are the most important contributor to productivity growth and that each accounts for about one-third of productivity movements. Firm entry generally negatively affects industry–level productivity growth, accounting for 28 to 30 per cent of productivity movements. This result is similar to what was found by Breunig and Wong (2008) for the 1990s in Australia. It is not surprising, as many new firms end up not surviving: they may lack access to industry-specific knowledge and skills or necessary capital (Mata and Portugal, 1994; Honjo, 2000).

Within-firm productivity increases are generally a negative contributor to industry–level productivity. Overall, they also account for about one third of productivity movements. Reallocation effects for continuing firms are virtually non-existent. Almost all of the reallocation happens through entry and exit, suggesting that policies that facilitate firm entry and exit are likely to help increase productivity gains.

We find substantial heterogeneity at the industry level. We also find substantial differences between our estimates and the official statistics provided by the Australian Bureau of Statistics (ABS). These differences are driven by our sample inclusion rules (we eliminate single-employee firms and sole traders, which are included by the ABS) and our approach to estimating capital and material inputs of firms which differ from the approach of the ABS.

Our paper provides further evidence, in a growing literature, of the value of using micro-data to understand the components of industry-level productivity growth. It confirms the usefulness of firm-level data in understanding the contributions that entering, exiting and surviving firms make to productivity. What is clear from our analysis is that policies that provide considerable advantages to incumbent firms (such as cumbersome regulation, which is difficult for new entrants to comply with or assistance to existing firms which keep otherwise failing firms in business) are likely to detract from productivity growth. One potential area for future work would be to explore the link between younger firms' contribution and overall growth to inform policies and encourage economic growth (see Andrews et al., 2015).

Acknowledgements We would like to thank Benjamin Jefferson for research assistance. We benefited from the comments of the Tax and Transfer Policy Institute's Friday seminar gang and seminar participants at the ANU Mathematical Science Institute. We are grateful to colleagues from the Australian Bureau of Statistics–Lisette Aaron, Michael Beahan, Diane Braskic, Bruce Hockman, Dr Anders Holmberg, Grace Kim Sybille McKeown, Katrina Richardson, Paul Roberts, Dr Siu-Ming Tam, David Taylor, Liza Tiy and Carter Wong–for their helpful comments and support for this research. We are also grateful to Dr David Maré and Cornell University Labor Dynamics Institute for making Fortran and SAS codes available.

A Summary Statistics

Table A1: Summary statistics: firm-level dataset

Statistic	$P_{1^{st}}$	$P_{50^{th}}$	$P_{99^{th}}$	St. Dev.
Complete case data				
lny_{jt}	7.28	10.57	13.34	1.14
ln$\widehat{z}_{jt}$	6.06	7.49	9.70	1.06
lnK_{jt}	5.01	8.88	11.60	1.27
lnM_{jt}	5.40	10.59	14.11	1.74
lnAge_{jt}	0.00	1.79	2.94	0.79
lnw_{jt}	6.53	9.82	11.90	0.99
Sample size		2,296,984		
Imputed data				
lny_{jt}	7.03	10.55	13.33	1.24
ln$\widehat{z}_{jt}$	6.03	7.38	9.77	1.08
lnK_{jt}	4.41	8.69	11.76	1.48
lnM_{jt}	5.06	10.18	14.41	1.94
lnAge_{jt}	0.00	1.79	2.94	0.83
lnw_{jt}	6.30	9.75	12.03	1.13
Sample size		10,039,638		

lny_{jt} is logarithm of output (i.e., sales adjusted for repurchase of stock) deflated by industry gross value added implicit price deflators.
ln$\widehat{z}_t^{(j)}$ the logarithm of estimated labor inputs.
lnK_{jt} is the logarithm of capital that includes depreciation, capital rental expenses and capital work deductions deflated by the industry consumption of fixed capital implicit price deflators.
lnM_{jt} is the logarithm of material costs deflated by *Producer Price Indexes: Intermediate Goods* (ABS, 2018d).
lnAge_{jt} is the logarithm of firm age. Firm age is derived as the current year minus the year of incorporation.
ln$Wages_{jt}$ is the logarithm of wage costs (reported in Business Activity Statements) deflated by *Wage Price Index: All Industries*.

Table A2: Summary statistics: worker-level dataset

Statistic	Mean	St. Dev.	Min	Max
$SKILLH$	0.31			
$SKILLHM$	0.11			
$SKILLM$	0.12			
2003	0.07			
2004	0.07			
2005	0.07			
2006	0.07			
2007	0.08			
2008	0.08			
2009	0.08			
2010	0.12			
2011	0.11			
2012	0.10			
2013	0.09			
AGE	37	37	17	64
AGE^2	1549	1369	289	4096
AGE^3	70029	50653	4913	262144
AGE^4	3370501	1874161	83521	16777216
$SEX : AGE$	19	18	0	64
$SEX : AGE^2$	792	324	0	4096
$SEX : AGE^3$	35825	5832	0	262144
$SEX : AGE^4$	1726528	104976	0	16777216
SEX : 2003	0.03			
SEX : 2004	0.04			
SEX : 2005	0.04			
SEX : 2006	0.04			
SEX : 2007	0.04			
SEX : 2008	0.04			
SEX : 2009	0.04			
SEX : 2010	0.06			
SEX : 2011	0.06			
SEX : 2012	0.05			
SEX : 2013	0.05			
	Sample size: 130,281,096			

High Skill ($SKILLH$) equals 1 if a worker has at least a tertiary qualification.
Medium Skill ($SKILLHM$) equals 1 if a worker has at most a diploma qualification.
Working Skill ($SKILLM$) equals 1 if a worker has at most a Certificate III qualification.
2003 represents financial year 2002–03.
AGE is the logarithm of worker age, abbreviated without the ‘ln’.
$SEX : AGE$, $SEX : AGE^2$, $SEX : AGE^3$ and $SEX : AGE^4$ are the interaction terms between worker sex (SEX) and polynomial AGE.
SEX : 2003, $\cdots$, SEX : 2013 are interaction terms between SEX and time dummies.

B Firm entry and exit rates

We follow Nguyen and Hansell (2014) and define firm entry rate as the number of new firms divided by the total number of incumbent and entering firms in a given year. Exit rate is defined as the number of firms exiting the market in a given year divided by the incumbents (including the exiting firms) in the previous year.

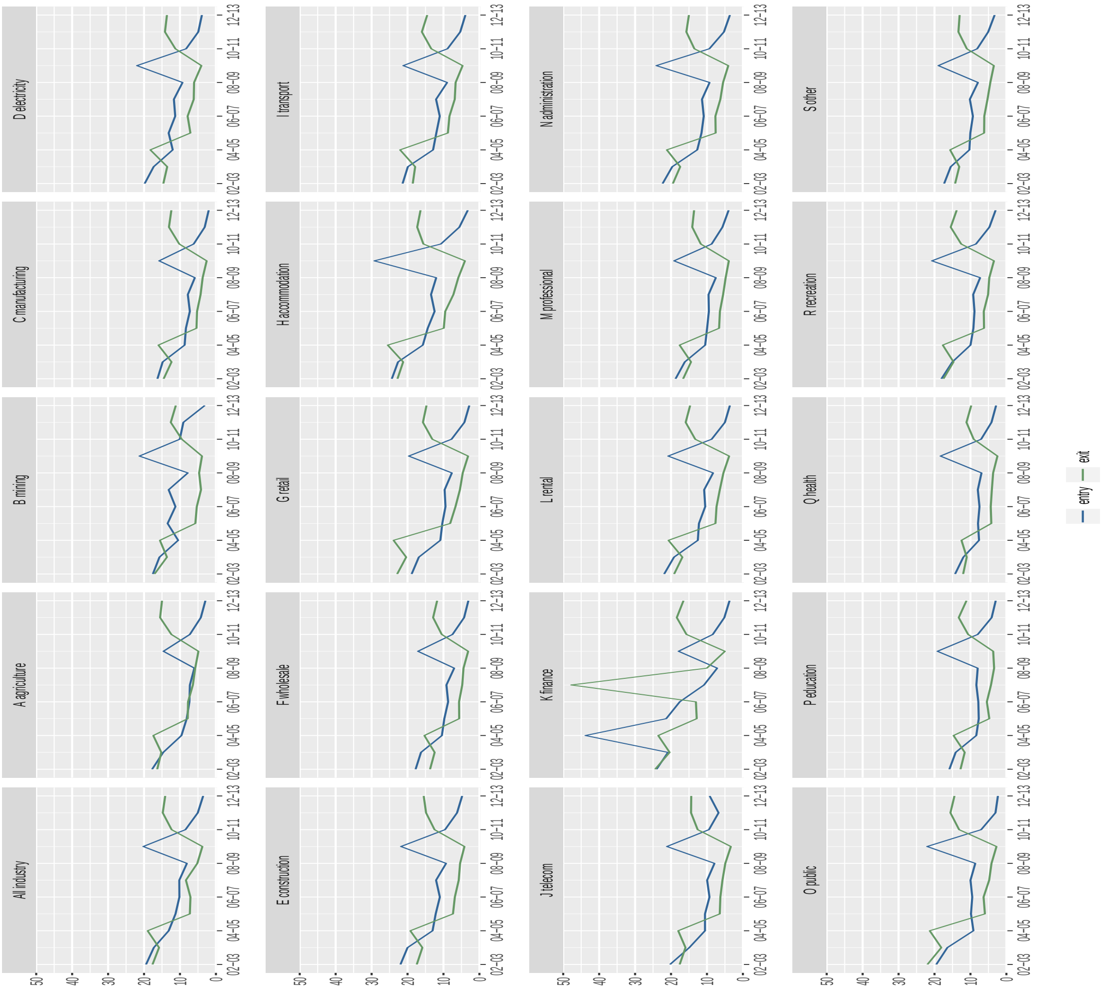


Fig. B1: Firm entry and exit

C Empirical results

C.1 Firm model results

Table C1: All industries (ALL) and Agriculture, Forestry and Fishing (A) industry results

	Complete cases		*Imputed*		*Complete cases*		*Imputed*	
	ALL.2SLS	ALL.OLS	ALL.2SLS	ALL.OLS	A.2SLS	A.OLS	A.2SLS	A.OLS
$\ln\widehat{z}_{jkt}$	0.691^{***}		0.648^{***}		-0.036^{**}		-0.016^{***}	
	(0.001)		(0.0003)		(0.018)		(0.004)	
$\ln w$		0.723^{***}		0.706^{***}		0.212^{***}		0.248^{***}
		(0.001)		(0.0003)		(0.002)		(0.001)
$\ln K$	0.223^{***}	0.251^{***}	0.246^{***}	0.239^{***}	0.503^{***}	0.447^{***}	0.453^{***}	0.402^{***}
	(0.001)	(0.001)	(0.0002)	(0.0002)	(0.002)	(0.002)	(0.001)	(0.001)
$\ln M$	0.242^{***}	0.183^{***}	0.238^{***}	0.170^{***}	0.114^{***}	0.069^{***}	0.129^{***}	0.076^{***}
	(0.0004)	(0.0005)	(0.0002)	(0.0002)	(0.001)	(0.001)	(0.001)	(0.001)
$\ln Age$	0.120^{***}	0.054^{***}	0.159^{***}	0.060^{***}	0.008	-0.061^{***}	0.022^{***}	-0.033^{***}
	(0.001)	(0.001)	(0.0005)	(0.0005)	(0.005)	(0.005)	(0.003)	(0.002)
Observations	2,296,984	2,296,984	10,039,638	10,039,638	162,766	162,766	662,553	662,553
Adjusted R^2	0.992	0.992	0.990	0.990	0.363	0.399	0.357	0.405
Note:								$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table C2: Mining (B) and Manufacturing (C) industries results

	Complete cases		*Imputed*		*Complete cases*		*Imputed*	
	B.2SLS	B.OLS	B.2SLS	B.OLS	C.2SLS	C.OLS	C.2SLS	C.OLS
$\ln\widehat{z}_t^{(jk)}$	0.584^{***}		0.235^{***}		0.565^{***}		0.090^{***}	
	(0.115)		(0.014)		(0.012)		(0.003)	
WAGES		0.563^{***}		0.546^{***}		0.527^{***}		0.469^{***}
		(0.017)		(0.005)		(0.002)		(0.001)
LnK	0.313^{***}	0.218^{***}	0.226^{***}	0.162^{***}	0.200^{***}	0.127^{***}	0.179^{***}	0.126^{***}
	(0.011)	(0.011)	(0.004)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)
LnM	0.159^{***}	0.090^{***}	0.203^{***}	0.108^{***}	0.317^{***}	0.189^{***}	0.352^{***}	0.216^{***}
	(0.007)	(0.007)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
	-0.012	0.016	0.002	-0.008	0.111^{***}	0.041^{***}	0.135^{***}	0.049^{***}
	(0.022)	(0.019)	(0.007)	(0.006)	(0.002)	(0.002)	(0.002)	(0.001)
Observations	4,902	4,902	36,559	36,559	288,335	288,335	645,869	645,869
R^2	0.289	0.413	0.261	0.431	0.303	0.414	0.320	0.430
Adjusted R^2	0.287	0.411	0.261	0.431	0.303	0.414	0.320	0.430
Note:								*p<0.1; **p<0.05; ***p<0.01

Table C3: Electricity, Gas, Water and Waste Services (D) and Construction (E) industries results

	Complete cases		*Imputed*		*Complete cases*		*Imputed*	
	D.2SLS	D.OLS	D.2SLS	D.OLS	E.2SLS	E.OLS	E.2SLS	E.OLS
$\ln\widehat{z}_t^{(jk)}$	0.409^{***}		0.114^{***}		0.123^{***}		0.023^{***}	
	(0.101)		(0.015)		(0.010)		(0.002)	
WAGES		0.541^{***}		0.444^{***}		0.403^{***}		0.355^{***}
		(0.016)		(0.006)		(0.002)		(0.001)
LnK	0.294^{***}	0.180^{***}	0.313^{***}	0.232^{***}	0.170^{***}	0.129^{***}	0.164^{***}	0.134^{***}
	(0.011)	(0.010)	(0.004)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)
LnM	0.134^{***}	0.069^{***}	0.145^{***}	0.084^{***}	0.229^{***}	0.169^{***}	0.245^{***}	0.183^{***}
	(0.007)	(0.007)	(0.003)	(0.003)	(0.001)	(0.001)	(0.0005)	(0.0005)
$\ln Age_{jt}$	0.079^{***}	0.016	0.122^{***}	0.060^{***}	0.030^{***}	-0.012^{***}	0.063^{***}	0.018^{***}
	(0.019)	(0.016)	(0.008)	(0.007)	(0.002)	(0.002)	(0.001)	(0.001)
Observations	5,022	5,022	28,837	28,837	373,859	373,859	1,477,460	1,477,460
Adjusted R^2	0.234	0.382	0.269	0.387	0.222	0.313	0.248	0.336
Note:								*p<0.1; **p<0.05; ***p<0.01

Table C4: Wholesale Trade (F) and Retail Trade (G) industries results

	Complete cases		*Imputed*		*Complete cases*		*Imputed*	
	F.2SLS	F.OLS	F.2SLS	F.OLS	G.2SLS	G.OLS	G.2SLS	G.OLS
$\ln\widehat{z}_t^{(jk)}$	0.570^{***} (0.017)		0.110^{***} (0.004)		0.622^{***} (0.010)		0.162^{***} (0.003)	
WAGES		0.493^{***} (0.003)		0.442^{***} (0.002)		0.450^{***} (0.002)		0.412^{***} (0.001)
LnK	0.167^{***} (0.002)	0.085^{***} (0.002)	0.160^{***} (0.001)	0.096^{***} (0.001)	0.139^{***} (0.001)	0.088^{***} (0.001)	0.132^{***} (0.001)	0.090^{***} (0.001)
LnM	0.336^{***} (0.002)	0.235^{***} (0.002)	0.363^{***} (0.001)	0.252^{***} (0.001)	0.362^{***} (0.001)	0.246^{***} (0.001)	0.395^{***} (0.001)	0.271^{***} (0.001)
lnAge_{jt}	0.112^{***} (0.003)	0.052^{***} (0.003)	0.138^{***} (0.002)	0.059^{***} (0.002)	0.187^{***} (0.002)	0.111^{***} (0.002)	0.221^{***} (0.001)	0.126^{***} (0.001)
Observations	213,389	213,389	480,515	480,515	434,058	434,058	1,072,727	1,072,727
Adjusted R^2	0.254	0.330	0.292	0.370	0.241	0.310	0.273	0.345
Note:								*p<0.1; **p<0.05; ***p<0.01

Table C5: Accommodation and Food Services (H) and Transport, Postal and Warehousing (I) industries results

	Complete cases		*Imputed*		*Complete cases*		*Imputed*	
	H.2SLS	H.OLS	H.2SLS	H.OLS	I.2SLS	I.OLS	I.2SLS	I.OLS
$\ln\widehat{z}_t^{(jk)}$	0.721^{***} (0.013)		0.217^{***} (0.003)		0.604^{***} (0.032)		0.086^{***} (0.003)	
WAGES		0.447^{***} (0.002)		0.420^{***} (0.001)		0.417^{***} (0.005)		0.424^{***} (0.001)
LnK	0.189^{***} (0.002)	0.132^{***} (0.002)	0.174^{***} (0.001)	0.122^{***} (0.001)	0.291^{***} (0.004)	0.208^{***} (0.003)	0.278^{***} (0.001)	0.213^{***} (0.001)
LnM	0.389^{***} (0.002)	0.268^{***} (0.002)	0.452^{***} (0.001)	0.315^{***} (0.001)	0.132^{***} (0.002)	0.091^{***} (0.002)	0.125^{***} (0.001)	0.080^{***} (0.001)
lnAge_{jt}	0.279^{***} (0.002)	0.213^{***} (0.002)	0.299^{***} (0.001)	0.224^{***} (0.001)	0.103^{***} (0.007)	0.029^{***} (0.006)	0.170^{***} (0.002)	0.097^{***} (0.002)
Observations	258,373	258,373	721,244	721,244	45,677	45,677	463,843	463,843
Adjusted R^2	0.360	0.433	0.383	0.461	0.229	0.327	0.257	0.380
Note:								*p<0.1; **p<0.05; ***p<0.01

Table C6: Telecommunications (J) and Financial and Insurance Services (K) industries results

	Complete cases		*Imputed*		*Complete cases*		*Imputed*	
	J.2SLS	J.OLS	J.2SLS	J.OLS	K.2SLS	K.OLS	K.2SLS	K.OLS
$\ln\widehat{z}_t^{(jk)}$	1.202***		0.095***		1.366***		0.081***	
	(0.061)		(0.009)		(0.053)		(0.003)	
WAGES		0.593***		0.566***		0.592***		0.529***
		(0.009)		(0.003)		(0.008)		(0.001)
LnK	0.277***	0.138***	0.239***	0.142***	0.224***	0.092***	0.248***	0.139***
	(0.007)	(0.006)	(0.003)	(0.002)	(0.006)	(0.006)	(0.001)	(0.001)
LnM	0.207***	0.116***	0.246***	0.138***	0.202***	0.137***	0.188***	0.132***
	(0.005)	(0.005)	(0.002)	(0.002)	(0.004)	(0.004)	(0.001)	(0.001)
lnAge_{jt}	0.056***	0.047***	0.144***	0.050***	0.090***	0.092***	0.124***	0.072***
	(0.013)	(0.011)	(0.005)	(0.004)	(0.012)	(0.011)	(0.002)	(0.002)
Observations	13,619	13,619	89,794	89,794	21,013	21,013	471,502	471,502
Adjusted R^2	0.267	0.430	0.266	0.461	0.237	0.384	0.242	0.435

Note: *p<0.1; **p<0.05; ***p<0.01

Table C7: Rental, Hiring and Real Estate Services (L) and Professional Services (M) industries results

	Complete cases		*Imputed*		*Complete cases*		*Imputed*	
	L.2SLS	L.OLS	L.2SLS	L.OLS	M.2SLS	M.OLS	M.2SLS	M.OLS
$\ln\widehat{z}_t^{(jk)}$	1.663***		0.225***		0.757***		0.156***	
	(0.036)		(0.004)		(0.019)		(0.002)	
WAGES		0.546***		0.404***		0.607***		0.577***
		(0.006)		(0.002)		(0.003)		(0.001)
LnK	0.278***	0.176***	0.262***	0.209***	0.215***	0.104***	0.152***	0.098***
	(0.004)	(0.004)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
LnM	0.200***	0.123***	0.204***	0.141***	0.144***	0.085***	0.165***	0.089***
	(0.003)	(0.003)	(0.001)	(0.001)	(0.002)	(0.001)	(0.0005)	(0.0004)
lnAge_{jt}	0.106***	0.076***	0.181***	0.117***	0.027***	0.008**	0.064***	0.009***
	(0.008)	(0.007)	(0.002)	(0.002)	(0.004)	(0.003)	(0.001)	(0.001)
Observations	40,665	40,665	386,405	386,405	124,096	124,096	1,298,560	1,298,560
R^2	0.288	0.391	0.260	0.356	0.178	0.400	0.161	0.427
Adjusted R^2	0.288	0.391	0.260	0.355	0.178	0.400	0.161	0.427

Note: *p<0.1; **p<0.05; ***p<0.01

Table C8: Administrative and Support Services (N) and Public Administration and Safety (O) industries results

	Complete cases		*Imputed*		*Complete cases*		*Imputed*	
	N.2SLS	N.OLS	N.2SLS	N.OLS	O.2SLS	O.OLS	O.2SLS	O.OLS
$\ln\widehat{z}_t^{(jk)}$	0.876***		0.174***		0.241***		0.118***	
	(0.036)		(0.004)		(0.079)		(0.012)	
WAGES		0.576***		0.549***		0.581***		0.563***
		(0.005)		(0.001)		(0.013)		(0.004)
LnK	0.246***	0.141***	0.222***	0.136***	0.227***	0.132***	0.224***	0.145***
	(0.004)	(0.003)	(0.001)	(0.001)	(0.009)	(0.008)	(0.003)	(0.003)
LnM	0.130***	0.069***	0.143***	0.073***	0.179***	0.105***	0.255***	0.148***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.007)	(0.006)	(0.003)	(0.002)
lnAge_{jt}	0.044***	0.003	0.116***	0.043***	0.043***	$-$0.018	0.093***	0.025***
	(0.007)	(0.006)	(0.002)	(0.002)	(0.015)	(0.013)	(0.006)	(0.005)
Observations	43,106	43,106	441,659	441,659	6,653	6,653	56,044	56,044
Adjusted R^2	0.246	0.410	0.261	0.451	0.276	0.448	0.375	0.550
Note:								*p<0.1; **p<0.05; ***p<0.01

Table C9: Education and Training (P) and Public Administration and Safety (Q) industries results

	Complete cases		*Imputed*		*Complete cases*		*Imputed*	
	P.2SLS	P.OLS	P.2SLS	P.OLS	Q.2SLS	Q.OLS	Q.2SLS	Q.OLS
$\ln\widehat{z}_t^{(jk)}$	1.227***		0.184***		0.756***		0.228***	
	(0.067)		(0.008)		(0.038)		(0.004)	
WAGES		0.592***		0.555***		0.593***		0.583***
		(0.010)		(0.002)		(0.006)		(0.001)
LnK	0.218***	0.096***	0.195***	0.110***	0.237***	0.127***	0.120***	0.069***
	(0.008)	(0.007)	(0.002)	(0.002)	(0.005)	(0.004)	(0.001)	(0.001)
LnM	0.203***	0.108***	0.265***	0.155***	0.125***	0.078***	0.186***	0.111***
	(0.005)	(0.005)	(0.001)	(0.001)	(0.003)	(0.003)	(0.001)	(0.001)
lnAge_{jt}	0.087***	0.037***	0.168***	0.041***	0.114***	0.025***	0.202***	0.012***
	(0.014)	(0.012)	(0.004)	(0.004)	(0.007)	(0.006)	(0.002)	(0.002)
Observations	10,478	10,478	181,655	181,655	35,078	35,078	629,550	629,550
Adjusted R^2	0.311	0.481	0.297	0.485	0.181	0.351	0.153	0.358
Note:								*p<0.1; **p<0.05; ***p<0.01

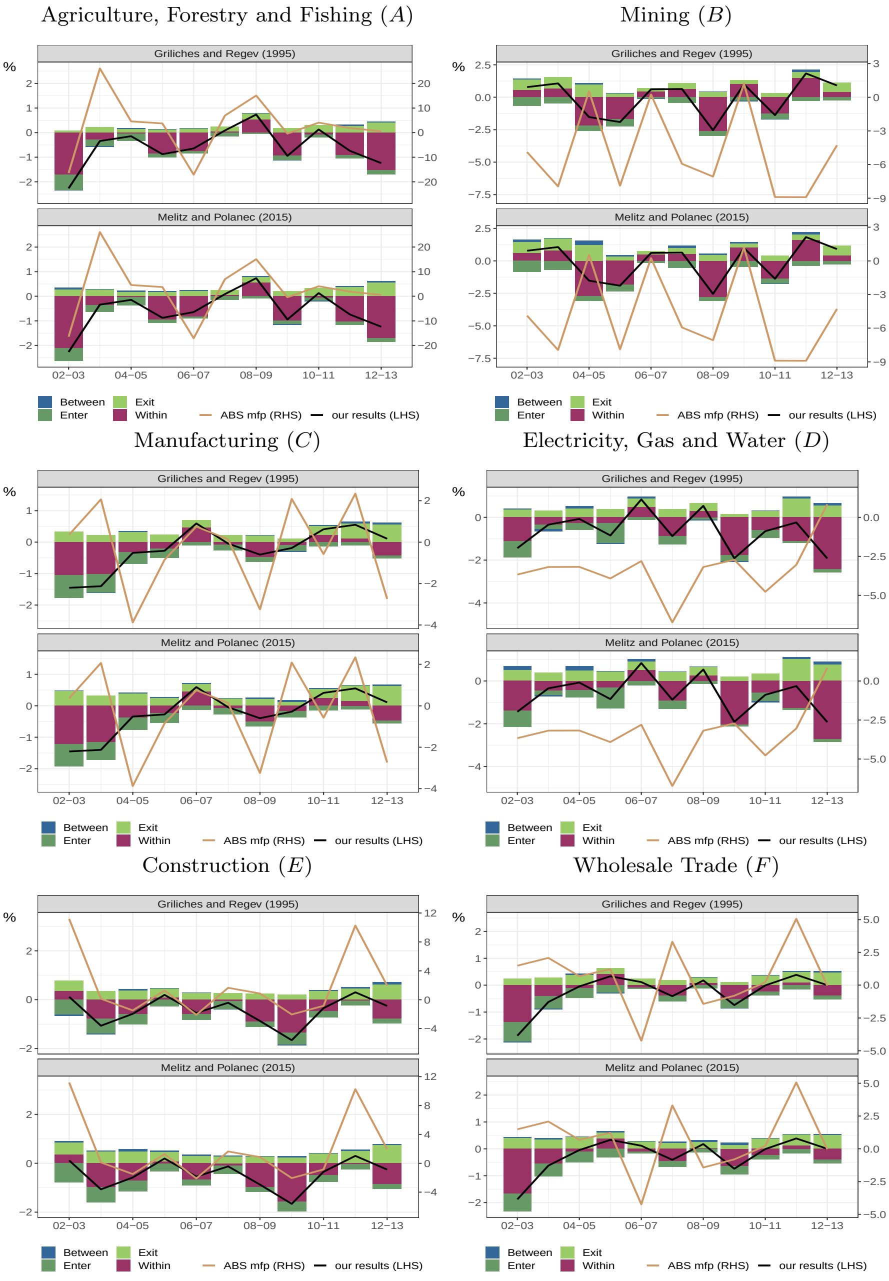
Table C10: Arts and Recreation Services (R) and Other Services (S) industries results

	Complete cases		*Imputed*		*Complete cases*		*Imputed*	
	R.2SLS	R.OLS	R.2SLS	R.OLS	S.2SLS	S.OLS	S.2SLS	S.OLS
$\ln\widehat{z}_t^{(jk)}$	1.093^{***}		0.154^{***}		0.247^{***}		0.049^{***}	
	(0.052)		(0.008)		(0.013)		(0.003)	
WAGES		0.589^{***}		0.558^{***}		0.498^{***}		0.459^{***}
		(0.008)		(0.002)		(0.003)		(0.001)
LnK	0.208^{***}	0.113^{***}	0.148^{***}	0.065^{***}	0.158^{***}	0.101^{***}	0.144^{***}	0.106^{***}
	(0.005)	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
LnM	0.234^{***}	0.152^{***}	0.288^{***}	0.192^{***}	0.279^{***}	0.179^{***}	0.348^{***}	0.229^{***}
	(0.004)	(0.004)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
lnAge_{jt}	0.139^{***}	0.073^{***}	0.182^{***}	0.062^{***}	0.100^{***}	0.035^{***}	0.132^{***}	0.054^{***}
	(0.010)	(0.009)	(0.004)	(0.004)	(0.003)	(0.002)	(0.002)	(0.001)
Observations	19,502	19,502	146,098	146,098	196,393	196,393	748,764	748,764
Adjusted R^2	0.316	0.462	0.290	0.478	0.267	0.383	0.315	0.427

Note: *p<0.1; **p<0.05; ***p<0.01

C.2 Industry decomposition

Fig. C1: industry decomposition for industries (A) through (F)



Agriculture, Forestry and Fishing (A)

Mining (B)

Manufacturing (C)

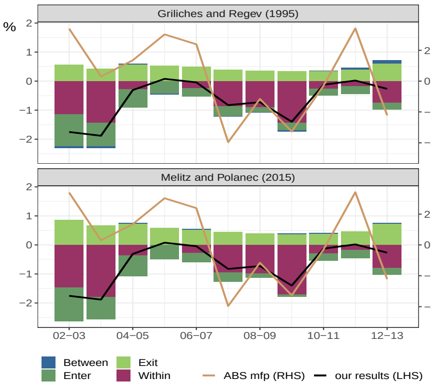
Electricity, Gas and Water (D)

Construction (E)

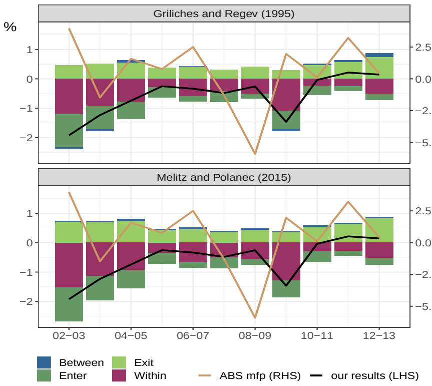
Wholesale Trade (F)

Fig. C2: industry decomposition for industries (H) through (M)

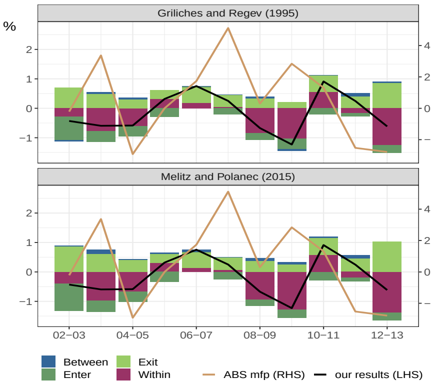
Accommodation and Food Services (H)



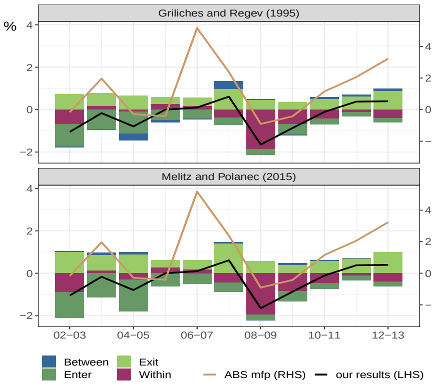
Transport, Postal and Warehousing (I)



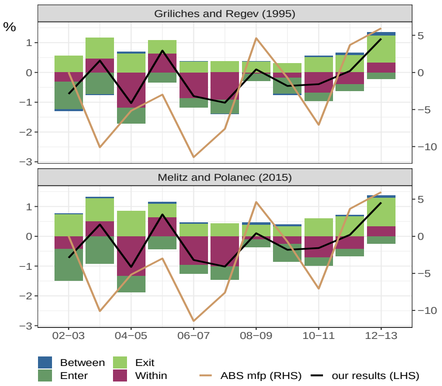
Telecommunications (J)



Financial and Insurance Services (K)



Rental and Real Estate Services (L)



Professional and Technical Services (M)

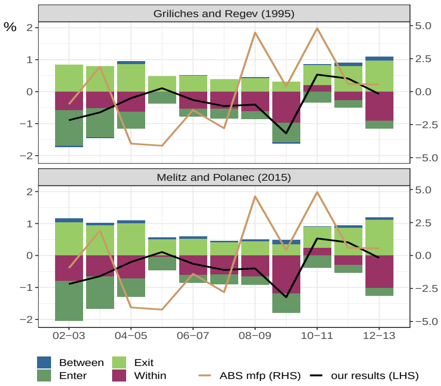
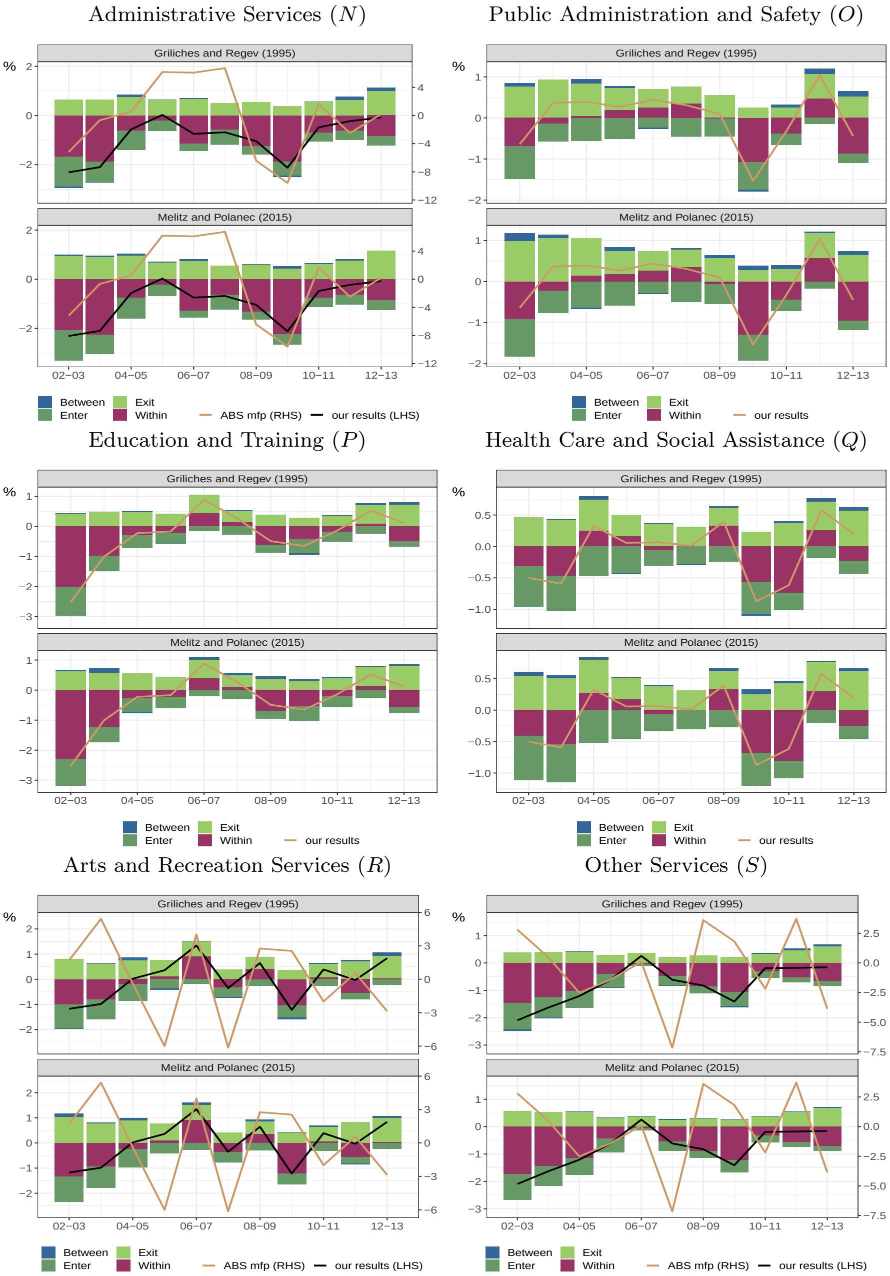


Fig. C3: industry decomposition for industries (N)through (S)



Administrative Services (N)

Public Administration and Safety (O)

Education and Training (P)

Health Care and Social Assistance (Q)

Arts and Recreation Services (R)

Other Services (S)

Table C11: Average by industry across all years

Griliches and Regev (1995)					Melitz and Polanec (2015)				
industry	**between**	**enter**	**exit**	**within**	**industry**	**between**	**enter**	**exit**	**within**
A	0.0136	-0.2033	0.2096	-0.5881	A	0.0302	-0.1916	0.2722	-0.6790
B	0.0382	-0.3807	0.5239	-0.2294	B	0.1079	-0.4400	0.5711	-0.2870
C	0.0130	-0.2653	0.2880	-0.2569	C	0.0314	-0.2742	0.3285	-0.3069
D	0.0194	-0.3316	0.4073	-0.7249	D	0.0619	-0.3310	0.4819	-0.8426
E	0.0131	-0.3460	0.3557	-0.4626	E	0.0487	-0.3648	0.4144	-0.5380
F	0.0064	-0.2687	0.2814	-0.2642	F	0.0413	-0.2781	0.3198	-0.3282
G	0.0077	-0.3580	0.3186	-0.4866	G	0.0176	-0.3608	0.4003	-0.5754
H	-0.0005	-0.4288	0.4513	-0.6799	H	0.0155	-0.4327	0.5576	-0.7984
I	0.0122	-0.4381	0.4665	-0.6192	I	0.0482	-0.4537	0.5596	-0.7328
J	0.0211	-0.2841	0.4627	-0.3493	J	0.0683	-0.3088	0.5231	-0.4322
K	0.0094	-0.5157	0.5889	-0.3682	K	0.0444	-0.6208	0.7298	-0.4390
L	0.0171	-0.4321	0.5201	-0.2861	L	0.0490	-0.4846	0.5940	-0.3394
M	0.0327	-0.4638	0.6313	-0.4899	M	0.0772	-0.5033	0.7134	-0.5771
N	0.0291	-0.5567	0.6230	-1.0349	N	0.0532	-0.5540	0.7473	-1.1859
O	0.0439	-0.4280	0.5503	-0.1684	O	0.0693	-0.4730	0.6152	-0.2136
P	0.0194	-0.3768	0.4623	-0.4161	P	0.0516	-0.3904	0.5180	-0.4903
Q	0.0122	-0.3664	0.3886	-0.1208	Q	0.0382	-0.3931	0.4215	-0.1529
R	0.0160	-0.4310	0.6151	-0.2149	R	0.0580	-0.4691	0.6877	-0.2914
S	0.0028	-0.4111	0.3548	-0.7349	S	0.0182	-0.4084	0.4276	-0.8258

D Estimation methods

This section explains the approach of Abowd et al. (2002), in particular, how the method identifies unique solutions.

D.1 Data structure

The model in (1) can be written as a model for each worker i by stacking the observations over time. We obtain:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\alpha} + \mathbf{1}_i\theta_i + \boldsymbol{F}_i\boldsymbol{\psi} + \boldsymbol{\epsilon}_i, \tag{13}$$

where $\boldsymbol{y}_i = \overset{T_i\times 1}{\begin{bmatrix} y_{it_{i1}} \\ \vdots \\ y_{it_{iT_i}} \end{bmatrix}}$, $\boldsymbol{X}_i = \overset{T_i\times 34}{\begin{bmatrix} \boldsymbol{x}^{\mathsf{T}}_{it_{i1}} \\ \vdots \\ \boldsymbol{x}^{\mathsf{T}}_{it_{iT_i}} \end{bmatrix}}$, $\mathbf{1}_i = \overset{T_i\times 1}{\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}}$, $\boldsymbol{F}_i \quad = \overset{T_i\times J}{\begin{bmatrix} \boldsymbol{f}^{\mathsf{T}}_{it_{i1}} \\ \vdots \\ \boldsymbol{f}^{\mathsf{T}}_{it_{iT_i}} \end{bmatrix}}$ and $\boldsymbol{\epsilon}_i = \overset{T_i\times 1}{\begin{bmatrix} \epsilon_{it_{i1}} \\ \vdots \\ \epsilon_{it_{iT_i}} \end{bmatrix}}$.

The model for the whole sample can be written in matrix form as:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{P}\boldsymbol{\theta} + \boldsymbol{F}\boldsymbol{\psi} + \boldsymbol{\epsilon}, \tag{14}$$

where $\boldsymbol{y} = \overset{N^*\times 1}{\begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_N \end{bmatrix}}$, $\boldsymbol{X} = \overset{N^*\times p}{\begin{bmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_N \end{bmatrix}}$, $\boldsymbol{P} = \overset{N^*\times N}{\begin{bmatrix} \mathbf{1}_1 & & \Huge 0 \\ & \ddots & \\ \Huge 0 & & \mathbf{1}_N \end{bmatrix}}$, $\boldsymbol{\theta} = \overset{N\times 1}{\begin{bmatrix} \theta_1 \\ \vdots \\ \theta_N \end{bmatrix}}$, $\boldsymbol{F} \quad = \overset{N^*\times J}{\begin{bmatrix} \boldsymbol{F}_1 \\ \vdots \\ \boldsymbol{F}_N \end{bmatrix}}$,

$\boldsymbol{\epsilon} = \overset{N^*\times 1}{\begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_N \end{bmatrix}}$ and $N^* = \sum_{i=1}^{N} T_i$ is the total number of observations.

D.2 Preconditioned conjugate gradient algorithm

This study uses the direct estimation methodology proposed by Abowd et al. (2002), which involves first solving a large sparse linear system with a preconditioned conjugate gradient algorithm and then imposing constraints on the parameters to identify unique worker and firm effects. The conjugate gradient algorithm solves the sparse linear system $\boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{c}$, where $\boldsymbol{A}$ is a symmetric positive definite matrix, $\boldsymbol{\beta}$ is an unknown vector, and $\boldsymbol{c}$ is a known vector. For ordinary least squares estimation of parameters in (1), the system is defined with:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{X}^{\mathsf{T}}\boldsymbol{X} & \boldsymbol{X}^{\mathsf{T}}\boldsymbol{P} & \boldsymbol{X}^{\mathsf{T}}\boldsymbol{F} \\ \boldsymbol{P}^{\mathsf{T}}\boldsymbol{X} & \boldsymbol{P}^{\mathsf{T}}\boldsymbol{P} & \boldsymbol{P}^{\mathsf{T}}\boldsymbol{F} \\ \boldsymbol{F}^{\mathsf{T}}\boldsymbol{X} & \boldsymbol{F}^{\mathsf{T}}\boldsymbol{P} & \boldsymbol{F}^{\mathsf{T}}\boldsymbol{F} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\theta} \\ \boldsymbol{\psi} \end{bmatrix} \text{ and } \boldsymbol{c} = \begin{bmatrix} \boldsymbol{X}^{\mathsf{T}}\boldsymbol{y} \\ \boldsymbol{P}^{\mathsf{T}}\boldsymbol{y} \\ \boldsymbol{F}^{\mathsf{T}}\boldsymbol{y} \end{bmatrix}. \tag{15}$$

Since $\boldsymbol{A}$ is a large, sparse matrix, iterative methods like the conjugate gradient algorithm perform better if we transform $\boldsymbol{A}$ to improve its condition number (Shewchuk, 1994). There are many options for creating a preconditioning matrix, including incomplete Cholesky factorisation or diagonal preconditioning, which uses a diagonal matrix whose diagonal entries are identical to the diagonal elements of $\boldsymbol{A}$ (see Hestenes, 1952, for a review). The preconditioning matrix used in the algorithm is a variant of incomplete Cholesky factorisation. Let:

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{Z} & 0 & 0 \\ 0 & \boldsymbol{P}^{1/2} & 0 \\ 0 & 0 & \boldsymbol{F}^{1/2} \end{bmatrix},$$

where $\boldsymbol{Z}$ is the upper triangular matrix obtained from the Cholesky decomposition of $\boldsymbol{X}^\intercal \boldsymbol{X}$, $\boldsymbol{P}^{1/2}$ is the diagonal matrix with the square roots of the diagonal terms of $\boldsymbol{P}^\intercal \boldsymbol{P}$ on the diagonal and $\boldsymbol{F}^{1/2}$ is the diagonal matrix with the square roots of the diagonal terms of $\boldsymbol{F}^\intercal \boldsymbol{F}$ on the diagonal. Following Abowd et al. (2002) and Fasshauer (2007), rewrite the system as:

$$\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{\beta}} = \widetilde{\boldsymbol{c}},$$

$$\text{where } \widetilde{\boldsymbol{A}} = \boldsymbol{U}^{-\intercal}\boldsymbol{A}\boldsymbol{U}^\intercal = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{Z}^{-\intercal}\boldsymbol{X}^\intercal\boldsymbol{P}\boldsymbol{P}^{1/2} & \boldsymbol{Z}^{-\intercal}\boldsymbol{X}^\intercal\boldsymbol{F}\boldsymbol{F}^{1/2} \\ \boldsymbol{P}^{-1/2}\boldsymbol{P}^\intercal\boldsymbol{X}\boldsymbol{Z}^\intercal & \boldsymbol{I} & \boldsymbol{P}^{-1/2}\boldsymbol{P}^\intercal\boldsymbol{F}\boldsymbol{F}^{1/2} \\ \boldsymbol{F}^{-1/2}\boldsymbol{F}^\intercal\boldsymbol{X}\boldsymbol{Z}^\intercal & \boldsymbol{F}^{-1/2}\boldsymbol{F}^\intercal\boldsymbol{P}\boldsymbol{P}^{1/2} & \boldsymbol{I} \end{bmatrix},$$

$$\widetilde{\boldsymbol{\beta}} = \boldsymbol{U}^{-\intercal}\boldsymbol{\beta} \quad \text{and} \quad \widetilde{\boldsymbol{c}} = \boldsymbol{U}^{-\intercal}\boldsymbol{c}.$$

The preconditioned conjugate gradient algorithm used in this study was developed by Dongarra (1991) and implemented in Fortran (see Algorithm 1). Let (k) denote the current iteration and $(k+1)$ the next iteration. The conjugate gradient method computes $\widetilde{\boldsymbol{\beta}}^{(k+1)}$ by iterating:

$$\widetilde{\boldsymbol{\beta}}^{(k+1)} = \widetilde{\boldsymbol{\beta}}^{(k)} + \widetilde{\alpha}^{(k)}\widetilde{\boldsymbol{d}}^{(k)},$$

where $\widetilde{\alpha}^{(k)}$ is a scalar given by

$$\widetilde{\alpha}^{(k)} = \frac{\widetilde{\boldsymbol{r}}^{(k)\intercal}\boldsymbol{U}^{-1}\widetilde{\boldsymbol{r}}^{(k)}}{\widetilde{\boldsymbol{d}}^{(k)\intercal}\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{d}}^{(k)}}, \text{ with } \widetilde{\boldsymbol{r}} = \widetilde{\boldsymbol{c}} - \widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{\beta}}, \text{ and}$$

$$\widetilde{\boldsymbol{d}}^{(k+1)} = \widetilde{\boldsymbol{r}}^{(k+1)} + \widetilde{\delta}^{(k+1)}\widetilde{\boldsymbol{d}}^{(k)}, \text{ with } \quad \widetilde{\delta}^{(k+1)} = \frac{\widetilde{\boldsymbol{r}}^{(k+1)\intercal}\boldsymbol{U}^{-1}\widetilde{\boldsymbol{r}}^{(k+1)}}{\widetilde{\boldsymbol{r}}^{(k)\intercal}\boldsymbol{U}^{-1}\widetilde{\boldsymbol{r}}^{(k)}}.$$

The basic pseudocode is:

Algorithm 1 Preconditioned conjugate gradient algorithm

```
procedure
    compute the preconditioning matrix $\boldsymbol{U}$
    compute $\widetilde{\boldsymbol{A}}$ and $\widetilde{\boldsymbol{c}}$
    initial $\boldsymbol{r}^{(0)} = \widetilde{\boldsymbol{c}}$ and let $\boldsymbol{d}^{(0)} = \boldsymbol{U}^{-1}\boldsymbol{r}^{(0)}$
    for $k = 1, 2, 3, \cdots$ do
        $\widetilde{\alpha}^k = \frac{\widetilde{\boldsymbol{r}}^{(k)\intercal}\boldsymbol{U}^{-1}\widetilde{\boldsymbol{r}}^{(k)}}{\widetilde{\boldsymbol{d}}^{(k)\intercal}\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{d}}^{(k)}}$
        $\widetilde{\boldsymbol{\beta}}^{(k+1)} = \widetilde{\boldsymbol{\beta}}^{(k)} + \widetilde{\alpha}^{(k)}\widetilde{\boldsymbol{d}}^{(k)}$
        $\widetilde{\boldsymbol{r}}^{(k+1)} = \boldsymbol{r}^{(k)} - \widetilde{\alpha}^{k}\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{d}}^{(k)}$
        $\widetilde{\delta}^{(k)} = \frac{\widetilde{\boldsymbol{r}}^{(k+1)\intercal}\boldsymbol{U}^{-1}\widetilde{\boldsymbol{r}}^{(k+1)}}{\widetilde{\boldsymbol{r}}^{(k)\intercal}\boldsymbol{U}^{-1}\widetilde{\boldsymbol{r}}^{(k)}}$
        $\widetilde{\boldsymbol{d}}^{(k+1)} = \widetilde{\boldsymbol{r}}^{(k+1)} + \widetilde{\delta}^{(k+1)}\widetilde{\boldsymbol{d}}^{(k)}$
    until the difference between $\widetilde{\boldsymbol{\beta}}^{(k)}$ and $\widetilde{\boldsymbol{\beta}}^{(k+1)}$ is less than $10^{-7}$
```

The convergence criterion of $\frac{|\widetilde{\boldsymbol{r}}|}{|\widetilde{\boldsymbol{c}}|} < 10^{-7}$ that we use is similar to that used by others (e.g., Abowd et al., 2002; Hallez et al., 2007).

D.3 Identification using a grouping algorithm

The solutions provided by the preconditioned conjugate gradient algorithm for the firm and worker effects depend on the initial values, preconditioning matrices and convergence criteria. Koopmans (1949), Koopmans et al. (1950) and Fisher (1966) discuss the need to impose model constraints to identify the underlying economic relationship in the observed data. Since two parametric equations can have the same likelihood function, the desired equation is chosen by imposing some restrictions to uniquely identify parameters. There is an infinite number of possible constraints and solutions. Fujikoshi (1993) summarises several possible approaches for two-way cross-classified unbalanced data.

D.4 Issues in identification

We use a simplified version of (1) in this subsection to illustrate the issues faced in imposing appropriate restrictions on the model for workers' wages. For simplicity, we consider a single fixed t and replace the observable worker characteristics terms $\boldsymbol{x}_{it}^{\intercal}\boldsymbol{\alpha}$ with the fixed unknown constant μ. With these simplifications, the model (1) has expectation:

$$E\{\ln(y_{it})\} = \mu + \theta_i + \boldsymbol{f}_{it}^{\intercal}\boldsymbol{\psi} = \mu + \theta_i + \psi_j, \tag{16}$$

when worker i works for firm j at time t. With t fixed, it is convenient to make the dependence on j more explicit and, just for this subsection, replace y_{it} with y_{ij}. We consider a two-way table of five workers—labelled θ_i for $i = 1, \cdots, 5$—and four firms labelled ψ_j for $j = 1, \cdots, 4$. If we only have one observation in some cells; a simple example is shown in Figure $D1$(a). We describe the data in Figure $D1$a as unbalanced. The saturated model, the main effect without interaction model for the *balanced* data, is given by (16). The model matrix $(\boldsymbol{P}, \boldsymbol{F})$ is given in Figure $D1$a. The relationships between the columns in the model matrix in Figure $D1$b are:

$$\beta_0 = \sum_{i=1}^{5} \theta_i \qquad (17) \qquad\qquad \text{and} \qquad \beta_0 = \sum_{j=1}^{4} \psi_j, \qquad (18)$$

where the sums are interpreted as the sums of the vectors in the columns labelled by μ, θ_i and ψ_j. These relationships show that the model is over-parameterised and an infinite number of solutions satisfy the ordinary least squares normal equation (1).
As can be seen from Figure D1a, the observation pattern forms two groups. This model is also rank deficient. One way to identify unique solutions is by using the corner point constraint to set redundant parameters to zero— $\theta_5 = \psi_4 = 0$—then the model matrix is shown in Figure D1c (Holmes et al., 1997).

Fig. D1: a simple example

(a) unbalanced table

	ψ_1	ψ_2	ψ_3	ψ_4
θ_1	A	A	NA	NA
θ_2	A	A	NA	NA
θ_3	A	A	NA	NA
θ_4	NA	NA	A	A
θ_5	NA	NA	A	A

use 17 18 and write

(b) No constraints

β_0	θ_1	θ_2	θ_3	θ_4	θ_5	ψ_1	ψ_2	ψ_3	ψ_4
1	1	0	0	0	0	1	0	0	0
1	1	0	0	0	0	0	1	0	0
1	0	1	0	0	0	1	0	0	0
1	0	1	0	0	0	0	1	0	0
1	0	0	1	0	0	1	0	0	0
1	0	0	1	0	0	0	1	0	0
1	0	0	0	1	0	0	0	1	0
1	0	0	0	1	0	0	0	0	1
1	0	0	0	0	1	0	0	1	0
1	0	0	0	0	1	0	0	0	1

not full rank

(c) With corner point constraints

β_0	θ_1	θ_2	θ_3	θ_4	ψ_1	ψ_2	ψ_3
1	1	0	0	0	0	1	0
1	0	1	0	0	1	0	0
1	0	1	0	0	0	1	0
1	0	0	1	0	1	0	0
1	0	0	1	0	0	1	0
1	0	0	0	1	0	0	1
1	0	0	0	1	0	0	0
1	0	0	0	0	0	0	1
1	0	0	0	0	0	0	0

full rank

β_0	θ_1	θ_2	θ_4	ψ_1	ψ_3
1	1	0	0	1	0
1	1	0	0	0	0
1	0	1	0	1	0
1	0	1	0	0	0
1	0	0	0	1	0
1	0	0	0	0	0
1	0	0	1	0	1
1	0	0	1	0	0
1	0	0	0	0	1
1	0	0	0	0	0

(d) Corner point constraints with group structure

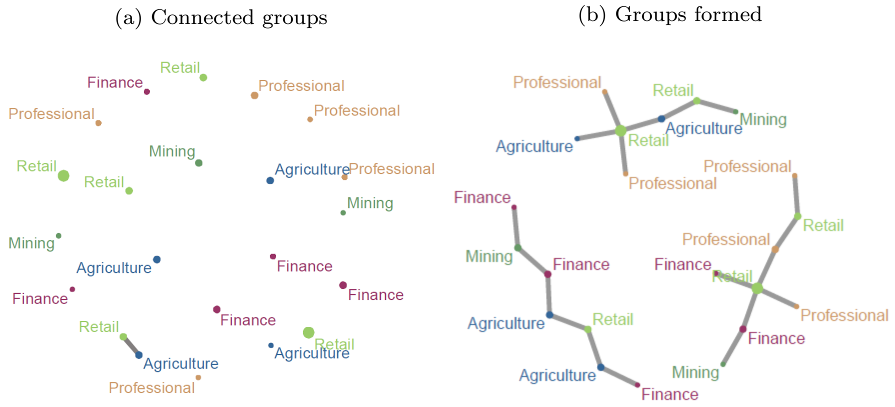
Note. A = available; NA = unavailable.

The model matrix in Figure D1c is still singular because $\theta_1 + \theta_2 + \theta_3 = \psi_1 + \psi_2$:

$$\overset{\theta_1+\theta_2+\theta_3}{\begin{bmatrix} 1\ 0\ 0 \\ 0\ 1\ 0 \\ 0\ 1\ 0 \\ \hline 0\ 0\ 1 \\ 0\ 0\ 1 \\ \hline 0\ 0\ 0 \\ 0\ 0\ 0 \\ \hline 0\ 0\ 0 \\ 0\ 0\ 0 \end{bmatrix}} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \equiv \overset{\psi_1+\psi_2}{\begin{bmatrix} 0\ 1 \\ 1\ 0 \\ 0\ 1 \\ \hline 1\ 0 \\ 0\ 1 \\ \hline 0\ 0 \\ 0\ 0 \\ \hline 0\ 0 \\ 0\ 0 \end{bmatrix}} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Figure D1b shows that the unbalanced data separates into two groups called 'connected groups' (Searle, 1987). We need to take the grouping structure into account to identify unique firm and worker effects. There is an infinite number of possible constraints to make the model matrix of full rank; the particular choice from these is arbitrary. An example is to impose $\psi_2 = 0$; the model matrix for the resulting full rank model is shown in Figure D1d. Abowd et al. (2002) recognise the need to find connected groups of workers and firms to set model constraints for analysing linked employer and employee data. Firms and workers can be connected by a worker changing jobs or multiple-job holders who work for different firms. Abowd et al. (2002) propose a grouping algorithm that divides connected workers and firms into mutually exclusive groups (see Algorithm 2). A group is defined as all workers and firms connected through some migration of workers between firms in that group. The main result is that the imposition of constraints according to the grouping structure ensures unique solutions for the worker and firm effects in the ordinary least squares normal equations. Figure $D2$a and Figure $D2$b show how the algorithm connects firms and workers so that no one worker or firm can be included in more than one group. The circle's size represents the firms' size to show that connections can occur between firms of different sizes. An edge connects two firms through a worker changing jobs from one firm to another or holding jobs in both firms. These connected groups are mutually exclusive because there are no additional worker movements.

Fig. D2: Connected groups


(a) Connected groups

(b) Groups formed

Algorithm 2 Grouping algorithm

```
procedure
    order by firm id and then worker id
    for group $g = 1$
    assign first firm $j$ to group $g = 1$
    partitioning step
    repeat
        add all workers employed by a firm $j$ in group $g = 1$ to group $g = 1$
        add all firms that have employed a worker $i$ in group $g = 1$ to group $g = 1$
    until no more firms or workers can be added to group $g = 1$
    end partitioning step
    for $g = 2$: $\forall$ worker $i$ and $\forall$ firm $j \notin g = 1$ assign first firm $j$ to $g = 2$
    repeat partitioning step and add all workers and firms in group $g = 2$ to group $g = 2$
            ⋮
    for $g = G$: $\forall$ worker $i$ and $\forall$ firm $j \notin g = 1, \cdots, G-1$ assign first firm $j$ to $g = G$
    repeat partitioning step and add all workers and all firms in $g = G$ to $g = G$
    until all firms are assigned
```

Declarations

Robert Breunig received support from the Australian Research Council, Discovery Grant Program (DP160101914).

The authors have no relevant financial or non-financial interests to disclose.

Data Availability

The dataset generated during the current study is not publicly available as it contains proprietary information that the authors acquired through a license. Information on how to obtain it and reproduce the analysis is available from the corresponding author on request.

References

Abowd, J. M., Creecy, R. H. and Kramarz, F. (2002), *Computing person and firm effects using linked longitudinal employer-employee data*, technical paper no. TP-2002-06, US Census Bureau. Suitland, MD. `ftp://ftp2.census.gov/ces/tp/tp-2002-06.pdf`

Abowd, J. M., Kramarz, F. and Margolis, D. N. (1999), 'High wage workers and high wage firms', *Econometrica* **67**(2), 251–333. `http://www.jstor.org/stable/2999586`

ABS—*see* Australian Bureau of Statistics

Australian Bureau of Statistics (2009), *ANZSCO—Australian and New Zealand standard classification of occupations*, cat. no. 1220.0. `http://www.abs.gov.au/AUSSTATS/abs@.nsf/allprimarymainfeatures/4AF138F6DB4FFD4BCA2571E200096BAD?opendocument`

Australian Bureau of Statistics (2013), *Estimates of Industry Multifactor Productivity*, cat. no. 5260.0.55.002. `https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/5260.0.55.002Main+Features12012-13?OpenDocument`

Australian Bureau of Statistics (2015), *Construction of experimental statistics on employee earnings and jobs from administrative data, Australia, 2011-12*, information paper, cat. no. 6311.0. `https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/6311.0main+features12011-12`

Australian Bureau of Statistics (2018b), *Australian system of national accounts, 2017-18*, cat. no. 5204.0. `http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/5204.02017-18?OpenDocument`

Australian Bureau of Statistics (2018d), *Producer price indexes, Australia, Sep 2018*, cat. no. 6427.0. `http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/6427.0Sep%202018?OpenDocument`

Andrews, D. and Criscuolo, C. (2015), *Firm dynamics and productivity growth in Europe*, report, Mimeo.

Andrews, M. J., Gill, L., Schank, T. and Upward, R. (2008), 'High wage workers and low wage firms: negative assortative matching or limited mobility bias?', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **171**(3), 673–697. `http://dx.doi.org/10.1111/j.1467-985X.2007.00533.x`

Bakhtiari, S. (2015), 'Productivity, outsourcing and exit: the case of Australian manufacturing', *Small Business Economics* **44**(2), 425–447.

Balk, B. M. (2016), 'Aggregate productivity and productivity of the aggregate: Connecting the bottom-up and top-down approaches', in W. Greene, L. Khalar, P. Makdissi, R. Sickles, M. Veall and M. Voia (eds), *Productivity and Inequality*, Springer, 119–141.

Bartelsman, E. J. and Dhrymes, P. J. (1998), 'Productivity dynamics: U.S. manufacturing plants, 1972–1986', *Journal of Productivity Analysis* **9**(1), 5–34.

Bartelsman, E. J. and Doms, M. (2000), 'Understanding productivity lessons from longitudinal microdata', *Journal of Economic Literature* **38**(3), 569–594.

Bartelsman, E. J. and Haltiwanger, J. and Scarpetta, S. (2013), 'Cross-country differences in Productivity: The role of allocation and selection', *American Economic Review* **103**(1), 305–334.

Brandt, L., Van Biesebroeck, J. and Zhang, Y. (2009), *Creative Accounting or Creative Destruction: Firm-level Productivity Growth in Chinese Manufacturing*, NBER Working Paper 15152, National Bureau of Economic Research. `https://www.nber.org/system/files/working_papers/w15152/w15152.pdf`.

Breunig, R. and Wong, M.-H. (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.

Brown, J. D. and Crespi, G. A. and Lacovone, L. andMarcolin, L. (2018), 'Decomposing firm-level productivity growth and assessing its determinants: evidence from the Americas', *The Journal of Technology Transfer* **43**, 1571–1606.

Chien, C.-H. and Mayer, A. (2015*a*), *A new analytical platform to explore linked data*, research paper, cat. no. 1352.0.55.151, Australian Bureau of Statistics, Canberra. `http://www.abs.gov.au/ausstats/abs@.nsf/mf/1352.0.55.151`

Chien, C.-H. and Mayer, A. (2015b), *Use of a prototype linked employer–employee database to describe characteristics of productive firms*, research paper, cat. no. 1351.0.55.055, Australian Bureau of Statistics, Canberra. https://www.abs.gov.au/ausstats/abs@.nsf/mf/1351.0.55.055

Chien, C.-H., Welsh, A. H. and Breunig, R. V. (2019a), *Approaches to analysing micro-drivers of aggregate productivity*, research paper, cat. no. 1351.0.55.164, Australian Bureau of Statistics, Canberra. https://www.abs.gov.au/ausstats/abs@.nsf/mf/1351.0.55.164.

Cincera, M. and Galgau, O. (2005*a*), *Impact of Market Entry and Exit on EU Productivity and Growth Performance*, Economic Papers Number 222, European Commission. Directorate-General for Economic and Financial Affairs https://econwpa.ub.uni-muenchen.de/econ-wp/io/papers/0503/0503013.pdf

Del Gatto, M., Di Liberto, A. and Petraglia, C. (2011), 'Measuring productivity', *Journal of Economic Surveys* **25**(5), 952–1008.

Dias, D. A. and Marques, C. R. (2021), 'From micro to macro: a note on the analysis of aggregate productivity dynamics using firm-level data', *Journal of Productivity Analysis* **56**(1), 1–14.

Dongarra, J. J. (1991), *Solving linear systems on vector and shared memory computers*, Society for Industrial and Applied Mathematics, Philadelphia. **ISBN:** 9780898712704.

Dosi, G., Grazzi, M., Li, L., Marengo, L and Settepanella, S. (2021), 'Productivity Decomposition in Heterogeneous Industries', *The Journal of Industrial Economics***69**(3), 615–652.

Duarte, M. and Restuccia, D. (2017), *Relative prices and sectoral productivity*, working paper no. 23979, National Bureau of Economic Research, Cambridge, MA.

Dumagan, J. C. and Balk, B. M. (2016), 'Dissecting aggregate output and labor productivity change: a postscript on the role of relative prices', *Journal of Productivity Analysis* **45**(1), 117–119.

Eberhardt, M. and Helmers, C. (2010), *Untested assumptions and data slicing: a critical review of firm-level production function estimators*, discussion paper 513, University of Oxford, Department of Economics. https://www.economics.ox.ac.uk/materials/working_papers/paper513.pdf

Fasshauer, G. (2007), '477/577 numerical linear algebra/computational mathematics I', Illinois Institute of Technology, Chicago. http://www.math.iit.edu/~fass/477_577.html

Fisher, F. M. (1966), *The identification problem in econometrics*, McGraw-Hill, New York. **ISBN:** 0882753444.

Foster, L., Haltiwanger, J. C. and Krizan, C. J. (2001), 'Aggregate productivity growth: lessons from microeconomic evidence', in C. R. Hulten, E. R. Dean and M. J. Harper (eds), *New developments in productivity analysis*, University of Chicago Press, 303–372.

Fox, J. T. and Smeets, V. (2011), 'Does input quality drive measured differences in firm productivity?', *International Economic Review* **52**(4), 961–989. http://www.jstor.org/stable/41349184

Fujikoshi, Y. (1993), 'Two-way ANOVA models with unbalanced data', *Discrete Mathematics* **116**(1–3), 315–334. http://www.sciencedirect.com/science/article/pii/0012365X9390410U

Gandhi, A., Navarro, S. and Rivers, D. A. (2011), 'On the identification of production functions: how heterogeneous is productivity?', revised manuscript resubmitted to the *Journal of Political Economy*, University of Wisconsin–Madison. http://www.ssc.wisc.edu/~agandhi/homepage/Amit_Gandhi_files/production_9_25_13_FULL.pdf

Griliches, Z. and Regev, H. (1995), 'Firm productivity in Israeli industry 1979–1988', *Journal of Econometrics* **65**(1), 175–203. http://www.sciencedirect.com/science/article/pii/030440769401601U

Hansell, D. and Rafi, B. (2018), 'Firm-level analysis using the ABS' Business Longitudinal Analysis Data Environment (BLADE)', *Australian Economic Review* **51**(1), 132–138. https://doi.org/10.1111/1467-8462.12253

Hallez, H., Vanrumste, B., Grech, R., Muscat, J., De Clercq, W., Vergult, A., D'Asseler, Y., Camilleri, K. P., Fabri, S. G., Van Huffel and Lemahieu, I. (2007), 'Review on solving the forward problem in EEG source analysis', *Journal of NeuroEngineering and Rehabilitation* **4**(1), art. 46.

Hestenes, M. R. and Stiefel, E. (1952), *Methods of conjugate gradients for solving linear systems* **49**(6), 409–436. https://www.fing.edu.uy/inco/cursos/numerico/aln/hes_stief1952.pdf

Holmes, A., Poline, J. and Friston, K. (1997), 'Characterizing brain images with the general linear model', in R. Frackowiak, K. Friston, C. Frith, R. Dolan and J. Mazziotta, (eds), *Human brain function*, Academic Press, San Diego, 59–84.

Honjo, Y. (2000), 'Business failure of new firms: an empirical analysis using a multiplicative hazards model', *International Journal of Industrial Organization* **18**(4), 557–574.

Iranzo, S., Schivardi, F. and Tosetti, E. (2008), 'Skill dispersion and firm productivity: an analysis with matched employer–employee data', *Journal of Labor Economics* **26**(2), 247–285.

Koopmans, T. C. (1949), 'Identification problems in economic model construction', *Econometrica* **17**(2), 125–144.

Koopmans, T. C., Rubin, H. and Leipnik, R. B. (1950), 'Measuring the equation systems of dynamic economics', in T. C. Koopmans (ed.), *Statistical inference in dynamic economic models*, John Wiley & Sons, New York, 53–237.

Lafrance, A. and Baldwin, J. R. (2011), *Firm turnover and productivity growth in selected Canadian services industries, 2000 to 2007*, economic analysis research paper, cat. no. 11F0027 no. 72, Statistics Canada, Economic Analysis Division, Ottawa, Ontario. https://www150.statcan.gc.ca/n1/en/pub/11f0027m/11f0027m2011072-eng.pdf?st=QAqdVn12.

Lentz, R. and Mortensen, D. T. (2010), 'Labor market models of worker and firm heterogeneity', *Annual Review of Economics* **2**(1), 577–602.

Levinsohn, J. and Petrin, A. (2003), 'Estimating production functions using inputs to control for unobservables', *The Review of Economic Studies* **70**(2), 317–341. http://www.jstor.org/stable/3648636

Maré, D. C., Hyslop, D. R. and Fabling, R. (2017), 'Firm productivity growth and skill', *New Zealand Economic Papers* **51**(3), 302–326. http://doi.org/10.1080/00779954.2016.1203815

Mata, J. and Portugal, P. (1994), 'Life duration of new firms', *The Journal of Industrial Economics* **42**(3), 227–245.

McFadden, D. (1963), 'Constant elasticity of substitution production functions', *The Review of Economic Studies* **30**(2), 73–83.

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. http://dx.doi.org/10.1111/1756-2171.12088

Nguyen, T. and Hansell, D. (2014), *Firm dynamics and productivity growth in Australian manufacturing and business services Oct 2014*, research paper, cat. no. 1351.55.052, Australian Bureau of Statistics, Canberra. https://www.abs.gov.au/ausstats/abs@.nsf/mf/1351.0.55.052

Olley, G. S. and Pakes, A. (1996), 'The dynamics of productivity in the telecommunications equipment industry', *Econometrica* **64**(6), 1263–1297. http://www.jstor.org/stable/2171831

Pavcnik, N. (2002), 'Trade Liberalization, Exit, and Productivity Improvement: Evidence from Chilean Plants', *Review of Economic Studies* **69**(6), 245–276.

Searle, S. R. (1987), *Linear models for unbalanced data*, Wiley, New York. **ISBN:** 9780471840961.

Shewchuk, J. R. (1994), *An introduction to the conjugate gradient method without the agonizing pain*, 1 1/4 edn, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA. http://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf

Syverson, C. (2011), 'What determines productivity?', *Journal of Economic Literature* **49**(2), 326–365. http://home.uchicago.edu/syverson/productivitysurvey.pdf

Van Biesebroeck, J. (2007), 'Robustness of productivity estimates', *The Journal of Industrial Economics* **55**(3), 529–569.

Van Biesebroeck, J. (2008), 'Aggregating and decomposing productivity', *Review of Business and Economics* **53**(2), 122–146.

Walters, R. and Dippelsman, R. J. (1986), *Estimates of depreciation and capital stock, Australia*, occasional paper 1985/3, Australian Bureau of Statistics, Canberra.