Motu Working Paper 21-01

Living on the edge: An anatomy of New Zealand's most productive firms

Motu economic & public policy research

Richard Fabling

February 2021

Document information

Author contact details

Richard Fabling
Independent Researcher
richard.fabling@xtra.co.nz

About the frontier firms inquiry

The New Zealand Productivity Commission (NZPC) is undertaking an inquiry into a central aspect of New Zealand's productivity performance – the economic contribution of its frontier firms. The Government has asked the Commission to investigate how the economic contribution of New Zealand's frontier firms can be maximised through policies aimed at 1) improving the performance of frontier firms themselves; and 2) helping innovations diffuse more effectively from frontier firms to other New Zealand firms. This paper has been commissioned to understand the key characteristics of firms at New Zealand's productivity frontier through the use of Stats NZ's Longitudinal Business Database. It is one of a number of research inputs into the Commission's inquiry (see www.productivity.govt.nz/inquiries/frontier-firms/ for more information).

Acknowledgements

We gratefully acknowledge NZPC funding, and valuable feedback from Dave Maré (Motu), John Janssen (Treasury), and from Geoff Lewis, Gail Pacheco, Bill Rosenberg, and Penny Mok at the NZPC.

Disclaimer

The results in this paper are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI), managed by Stats NZ. The opinions, findings, recommendations, and conclusions expressed in this paper are those of the author, not Stats NZ, NZPC, Motu or individual data suppliers.

Access to the anonymised data used in this study was provided by Stats NZ under the security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business, or organisation, and the results in this paper have been confidentialised to protect these groups from identification and to keep their data safe. Careful consideration has been given to the privacy, security, and confidentiality issues associated with using administrative and survey data in the IDI. Further detail can be found in the Privacy Impact Assessment for the Integrated Data Infrastructure available from www.stats.govt.nz.

The results are based in part on tax data supplied by Inland Revenue to Stats NZ under the Tax Administration Act 1994. This tax data must be used only for statistical purposes, and no individual information may be published or disclosed in any other form, or provided to Inland Revenue for administrative or regulatory purposes. Any person who has had access to the unit record data has certified that they have been shown, have read, and have understood section 81 of the Tax Administration Act 1994, which relates to secrecy. Any discussion of data limitations or weaknesses is in the context of using the IDI for statistical purposes, and is not related to the data's ability to support Inland Revenue's core operational requirements.

Abstract

Theory and international evidence suggest that firms at the New Zealand productivity frontier may be especially important for the diffusion of knowledge from the global productivity frontier, acting as a conduit for new technologies and ideas to flow into the domestic economy. We identify the NZ productivity frontier in a novel way that is robust to some sources of measurement error, and to criticism that the frontier label is dependent on arbitrary assumptions. We show that economic activity is concentrated in the upper deciles of the productivity distribution, and that frontier firms are disproportionately important to aggregate output, even relative to firms just outside the frontier. Compared to laggard firms, frontier firms: employ a more skilled workforce concentrated in major Urban Areas (particularly Auckland); have superior human resource management practices; are more export intensive; are more likely to have up-to-date technology (including UFB use); and to be in markets with no competitors.

JEL codes

D20; L20; M21; O31

Keywords

Multifactor productivity; productivity frontier; productivity growth; management practices; innovation; exporting; foreign direct investment; competition

Summary haiku

Exceptional firms

drive the production frontier

out on the world's edge

Motu Economic and Public Policy Research

PO Box 24390	info@motu.org.nz	+64 4 9394250
Wellington	www.motu.org.nz	
New Zealand		

© 2021 Motu Economic and Public Policy Research Trust and the authors. Short extracts, not exceeding two paragraphs, may be quoted provided clear attribution is given. Motu Working Papers are research materials circulated by their authors for purposes of information and discussion. They have not necessarily undergone formal peer review or editorial treatment. ISSN 1176-2667 (Print), ISSN 1177-9047 (Online).

1 Motivation

As the New Zealand Productivity Commission (2020) note:

> "New Zealand is facing the prospect of a significant economic shock from the spread of COVID-19. Helping more Kiwi firms reach the productivity frontier would be a valuable step towards the economy reaching its full potential once the immediate effects of COVID-19 have passed.... Even a small improvement in productivity growth can have large cumulative effects in the form of improved jobs and earnings, more housing, better care of the environment and provision of social services. Lifting productivity is critical if New Zealand is to achieve higher incomes and living standards."

This focus on New Zealand's most productive firms is centered on an expectation that those firms are likely to make a disproportionately large contribution to the aggregate economy because they:

- produce more output for a given level of inputs (by definition);
- control a larger share of those inputs (since resources flow towards higher productivity firms); and
- may act as exemplars whose technologies can be replicated by lower productivity firms, thus raising average productivity level.

Bartelsman et al. (2008) formalise the last of these ideas, with laggard (non-frontier) firms potentially having the ability to learn from (converge in productivity towards) domestic productivity frontier firms and/or global productivity frontier firms.[1] Their microeconomic analysis focuses on United Kingdom (UK) manufacturing firms, finding that both the national and global productivity frontiers exert a "pull" on laggard firm productivity growth, but that the UK national frontier is more influential than the global frontier. Andrews et al. (2015) replicate and extend this analysis for a broader set of countries and sectors, also concluding that national productivity frontiers exert a stronger influence on non-frontier firm productivity growth than the global productivity frontier.[2] Since the domestic frontier is more relevant to

[1] In their empirical analysis, the global productivity frontier is generally determined by firms based in the United States of America.

[2] Conway et al. (2015) and Zheng (2016) explore the influence of frontier firm on laggard firms in NZ following the estimation approach of Griffith et al. (2009), concluding that faster frontier productivity growth increases the rate at which laggard firms converge to their own long-run average productivity. These NZ studies are not directly comparable

most firms, Bartelsman et al. (2008) hypothesise that national frontier firms may act as a technology conduit from the global productivity frontier, since their closer proximity to that frontier makes them the best able to learn from global technology leaders – a sort of global knowledge "trickle down" from the world frontier to the national frontier and then other domestic firms.

These results are particularly pertinent for New Zealand which, as a country, is a productivity laggard compared to its OECD peers implying that NZ firms are, on average, likely to be further behind the global productivity frontier than the European firms previously studied. Coupled with the negative impact of geographic distance on the ability of NZ firms to interact with firms at the global frontier (and other firms above the New Zealand frontier), NZ national frontier productivity firms may be especially important to the diffusion of knowledge from the global productivity frontier.

Aside from their closer proximity to the global production frontier, one key aspect of why frontier firms may be better equipped to absorb knowledge is the activities they undertake (eg, research and development), and the capabilities that they hold (eg, superior management and highly skilled workers). The aim of this paper is to identify the firms in the NZ productivity frontier, and to describe them. Syverson (2011) succinctly summarises the international literature on the sources of productivity variation across firms. Within the themes covered by Syverson, a number of New Zealand studies provide insights into the potential mechanisms that give frontier firms their productivity advantage, including:

- Fabling and Sanderson (2013), who find that the superior productivity of exporters is a selection effect (ie, firms that are already relatively high productivity become exporters), and that firms expand on entry into new markets, which is likely to increase aggregate productivity through resource reallocation;
- Fabling and Sanderson (2014a), who show that foreign-owned firms have higher productivity than domestically-owned firms, but that the superior productivity of foreign-owned firms is a selection effect – that is, new foreign direct investment (FDI) does not raise the productivity of firms;

to Bartelsman et al. (2008) and Andrews et al. (2015) due to the inclusion of firm fixed effects in the NZ results, which imposes the "conditional convergence" interpretation on their results. It is not clear that conditional convergence is the interesting economic phenomena, compared to absolute convergence to the frontier, and the inclusion of fixed effects exacerbates concerns that observed "convergence" is driven by mean-reversion in estimated productivity (caused by measurement error, one-off orders, etc).

- Fabling and Grimes (2014), who establish a link from adopting good human resource management (HRM) practices to improved productivity;
- Fabling and Grimes (2016), who find that ultrafast broadband (UFB) adoption raises productivity, but only when firms also make complementary ICT-enhancing investments;
- Maré (2016), who demonstrates that productivity is higher in major urban centres, particularly the Auckland Urban Area (UA), and that these UA premia are underestimated in the absence of controls for spatial price differences, and overestimated in the absence of controls for worker quality (skill);
- Maré et al. (2017) demonstrate the importance of controlling for worker quality (skill) in production functions, showing that skill-adjusted productivity growth was stronger than traditionally measured productivity growth prior to the Global Financial Crisis (GFC) due to relatively strong growth in the low-skilled workforce over that period;
- Chappell and Jaffe (2018) explore the relationship between intangibles investment and productivity, finding no clear link between the two;[3]
- Maré and Fabling (2019), who reach the tentative conclusion that more competition in an industry may improve average industry productivity by forcing low productivity firms to exit the market.

We begin our analysis by using the Longitudinal Business Database (LBD) to implement a novel definition of the NZ productivity frontier that is robust to the choice of production function and industry reference group. We then use this definition, the richness of the data in the LBD, and guidance from the prior NZ literature to describe the characteristics of firms at the frontier, and to quantify the frontier direct contribution (ie, excluding knowledge spillovers) to aggregate labour productivity growth in NZ. Finally, we use the changing composition of firms in the frontier over time, to identify characteristics associated with transitions into the frontier. In this last step we estimate relationships between frontier firm characteristics and productivity, by way of a series of ordinary least squares (OLS) regressions, with and without firm fixed effects. Our goal is not to provide conclusive evidence of a causal relationship from firm practices to improved performance to pro-

[3] We focus on links between product development/R&D expenditure, innovation outcomes and productivity, rather than "intangibles investment," which Chappell and Jaffe (2018) define as a subset of the reported innovation-related investment types performed by firms.

ductivity leadership, but rather to update and extend the set of stylised facts about firm practices and productivity, utilising updated data, coupled with a common (simplified) estimation approach.

The paper is set out as follows. Section 2 summarises the LBD components that we use and the population restrictions we impose to limit the analysis to subpopulations that excludes firms with characteristics associated with relatively high error in productivity measurement. We do this in a way that minimises the loss of aggregate inputs or output so that our results represent a subpopulation of firms that produce the bulk of private sector output. Section 3 explains the rationale for our frontier definition, and then reports the main findings that quantify the importance of the frontier, and describes the characteristics of firms at the frontier. Together, these analyses show that frontier firms make a disproportionately large contribution to the economy – a contribution that is likely driven by their superior engagement in activities that raise firm-level productivity. Section 4 provides a brief summary of the findings.

2 Data

We use the labour and productivity datasets available in the Longitudinal Business Database (LBD). These data contain standard production function variables – output (Y), intermediate consumption (M), capital services (K) and labour (L).[4] The last of these is derived from the linking of monthly Pay-As-You-Earn (PAYE) tax filings for employees and annual tax returns for working proprietors (WPs), with some downward adjustment to labour input for workers and WPs who are unlikely to be full-time in their job(s) (eg, multiple job holders). Remaining production function variables are derived from a mix of Annual Enterprise Survey (AES) returns and annual firm IR10 tax filings (Fabling 2011; Fabling and Maré 2015a, 2015b, 2019).

Production function data is augmented by a number of LBD datasets, including the Business Register, which identifies industry and physical firm locations (weighted by employment share), and the Business Operations Survey (BOS), which provides information on: exporting; inward and outward foreign direct investment (FDI); firm technology (core equipment and ultra-fast broadband (UFB) usage), innovation and investment in product development (including R&D); perceived competition; and collective employment

[4] Throughout the paper, production function variables in upper case are real (2018) dollars, while lower case variables indicate natural logs have been taken.

agreement coverage.[5] For consistency with the BOS components of the analysis, the analysis period is restricted to 2005 (the first BOS year) to 2018 (the latest available BOS and productivity data year).[6]

Multifactor productivity (MFP), derived from an estimated production function, is our preferred measure for determining frontier firms since this controls for the use of both labour and capital inputs. Labour productivity (LP) is used in some parts of the analysis, because LP is easier to aggregate and make cross-industry comparisons than MFP (which is inherently a within-industry construct).

Two types of firms have high apparent MFP – firms with high productivity, and firms with measurement error that make them appear more productive than they are. Fabling and Sanderson (2014b) suggests that micro enterprises – particularly WP-only firms – are substantially overrepresented in both tails of the productivity distribution and that this is likely to be due to greater true productivity variance for this firm type, and due to greater measurement error in inputs for small firms (especially the labour input). To avoid the estimated productivity frontier being dominated by firms that are not actually exceptional performers, we restrict our population of interest to exclude firms that may have greater measurement error, including micro enterprises, and firms in their first and last year of operation. This latter restriction is based on the logic that underlying (true) firm productivity is harder to observe in entry/exit years because firms may undertake exceptional activities, and because the productivity data construction makes a number of additional assumptions about labour and capital inputs in these transition years.

Table 1 shows the impact of these restrictions on the coverage of the productivity population broken down into sequential, cumulative steps. We make use of productivity dataset weights based on industry-firm-size cells to account for missing productivity data, using full coverage employment data to infer missing productivity population observations (Fabling and Maré 2019). The top panel of table 1 shows the proportion of productivity components (and firm-year observations) that are captured by the population, and the bottom panel shows how well the productivity data covers the population of interest (ie, the proportion of data that is actually observed, rather than accounted for by weighting). Focusing on the top (bold) row of each panel, the weighted dataset covers the entire productivity population (by construction),

[5] Fabling and Sanderson (2016) provide an overview of the BOS and other LBD data.

[6] We use productivity data based on the 20200120 IDI and December 2019 LBD archive. BOS data also come from the December 2019 LBD.

with approximately two thirds of firm-year observations having productivity data available. In the full productivity dataset, coverage of productivity components is significantly higher than the coverage of observations – ranging from 79% to 82% – primarily due to AES targeting full coverage of the largest firms in the economy.

Restricting the population to firms with observed MFP has minimal impact on the population, removing less than 2% of observations and aggregate labour. Since the weighting already accounts for firm-year observations that do not have usable data to construct all productivity components, this restriction simply removes observations where one or more productivity component is zero. The next restriction removes entering and exiting (in the following year) firms,[7] which has a more substantial impact on population size, removing a further 12.7% of firm-year observations. Since entering and exiting firms are less likely to file an IR10 form, their removal raises the firm-year coverage rate by three percentage points (to 70%).

The next restriction excludes WP-only firms, which removes a further 43.5% of the productivity population. The final restriction removes micro enterprises, and ensures each firm has at least two employees (excluding firms that are almost WP-only, and ensuring firm-level wage metrics are based on multiple individuals). This is shown in two steps, where we first impose full-time equivalent (FTE) employee labour input is greater than one, and then require that firm level L (which sums employee and WP labour input) is greater than two. Combined these steps reduce the firm-year population to 24% of it's original size. While these restrictions have a substantial impact on population size, productivity component aggregates are affected to a much lesser extent reflecting the comparatively small size of WP-only firms. In total, we lose 22% of population L (primarily from the WP-only exclusion); 21% of population K; 14% of population Y; and 13% of population M (final row of top panel in table 1). The exclusion of micro enterprises also yields a substantial improvement in data coverage, increasing the firm-year coverage rate by 10 percentage points from 67% to 77% (bottom panel of table, top vs bottom row).

Removed firm-year observations have relatively low labour productivity, which can be seen in figure 1, which plots annual aggregate labour produc-

[7]Entering (exiting) firms are transitioning from (to) inactivity, where activity is measured using full coverage employment (L) and Goods & Services Tax (GST) data, and partial coverage AES/IR10 productivity component data. These firms are distinct from those which join (leave) the productivity dataset due to incomplete data coverage, which remain in the analysis in the years where they have productivity data.

tivity – that is, aggregate value-added (Y-M) over aggregate L – at each restriction step. The exclusion of WP-only firms has the largest impact on observed aggregate labour productivity, since the loss of L input is much larger than the loss of output (table 1).[8] In terms of aggregate trends, all series display similar properties with largely static labour productivity up to the Global Financial Crisis (GFC), followed by a dip consistent with reduced output and labour hoarding, then a steady recovery and increase in productivity through to the end period (2018).

In summary, therefore, the population restrictions remove firm-year observations that may be more likely to have measurement error that would (incorrectly) cause the firm to look exceptional in productivity terms. At the same time, we have preserved the aggregate productivity dynamics of the population, while focussing on a relatively high productivity subset of firms with materially better data coverage.[9]

Table 2 reports the broad sectoral coverage of the productivity subpopulation, and shows that the majority of firms are in the services sector, despite the fact that productivity data coverage excludes large industries in that sector (notably, property operators, health care, education). In aggregate employment terms, the manufacturing sector make up a more substantial share of the total, reflecting the larger average firm size in that sector. In contrast, primary sector firms tend to be relatively small and their contribution to total employment is the lowest. The construction sector experienced the largest growth in both firms and employment, in part due to the Christchurch earthquake rebuild, while the manufacturing sector has been declining over most of the period.

Table 3 provides summary statistics of productivity and labour dataset variables for our productivity subpopulation. Aside from (logged) productivity components, these variables include the capital-labour ratio (ie, in logs $k - l$), estimated MFP, a proxy for average worker skill, firm age, and Urban Area (UA) employment shares.[10] MFP is estimated separately for each "production function" industry from the unweighted full productivity dataset using OLS with firm fixed effects, and assuming either a Cobb-Douglas (CD)

[8] In part, this likely reflects a particular measurement issue with WP labour inputs, since we rely on annual filing and attribute most WPs an FTE labour input of one, when many may be working part-time.

[9] The correlation between data quality and data coverage is likely to be positive, in part because productivity data cleaning steps remove internally inconsistent data, which will be more prevalent in subpopulations with more measurement error.

[10] A small number of firms have no plant location information, resulting in a slight decline in the total (unweighted and weighted) observations for statistics involving UAs.

or translog (TL) production function (Fabling and Maré 2015b, 2019). Formally, for a firm i in industry j, the estimated CD and TL production functions (including firm fixed effects) are, respectively:

$$\begin{aligned} y_{it} = \delta_{jt} &+ \beta_j^m m_{it} + \beta_j^l l_{it} + \beta_j^k k_{it} + \eta_i + \epsilon_{it} \\ y_{it} = \delta_{jt} &+ \beta_j^m m_{it} + \beta_j^l l_{it} + \beta_j^k k_{it} + \\ &+ \beta_j^{mm} m_{it}^2 + \beta_j^{ll} l_{it}^2 + \beta_j^{kk} k_{it}^2 \\ &+ \beta_j^{ml} m_{it} l_{it} + \beta_j^{mk} m_{it} k_{it} + \beta_j^{lk} l_{it} k_{it} + \eta_i + \epsilon_{it} \end{aligned}$$

In all cases, MFP is actual output less the estimated contribution of observed inputs (eg, for Cobb-Douglas, $y_{it} - [\beta_j^m m_{it} + \beta_j^l l_{it} + \beta_j^k k_{it}]$), which is equivalent to the sum of the industry-year effect (δ_{jt}), residual (ϵ_{it}) and firm fixed effect (η_i, where included).

Since the IDI does not include full coverage worker qualifications data, average worker skill is estimated from a two-way wage fixed effect model as in Maré et al. (2017), where the worker fixed effects capture the portable wage premium associated with each worker. Average worker skill is FTE-weighted at each firm, and includes the sum of observables (sex-age) and worker fixed effects components of the worker wage. We also make use of the firm wage fixed effect (FFE) from this model, which captures the wage premium paid by a firm to every worker at the firm. Given the arbitrary scaling of average worker skill and FFE, we have renormalised them to be mean zero, standard deviation one at the firm level. Firm age is based on Business Register birth year, first employment year or first GST year, whichever is the earliest.[11]

The BOS population only covers firms with six or more average monthly employment, creating a second population of interest when we link to the productivity data. Table 4 shows the impact of this further restriction on productivity and labour dataset variables. BOS-related statistics are weighted using BOS survey weights adjusted to include panel and Māori business top-ups in the weighted sample, and then further adjusted for the partial coverage of productivity data. The BOS reweighting to include panel observations occurs prior to applying the productivity population restriction (second weight adjustment) so that the final summed weights approximately reflect the restricted BOS-productivity subpopulation size. Since BOS is a sample survey, the data coverage rate of BOS statistics (0.132, bottom of table 4) is much

[11] We take the earliest of the three metrics to impose consistency between the (admin-based) productivity data view of firm activity and birth year, while acknowledging that the administrative data is left-censored. Because the productivity subpopulation restrictions exclude entrants, firm age is always greater than one (ln(firm age)>0).

lower than that of the productivity dataset (0.774, bottom of table 3).[12]

As expected, the BOS employment criteria means that the average BOS-productivity firm is larger – in terms of all productivity components – than the average productivity subpopulation firm. BOS-productivity firms also have higher average MFP, worker skill and firm wage fixed effect, and are less likely to operate in rural areas, where this last difference is partly due to the disproportionate removal of primary sector firms by the employment restriction. The impact of these restrictions on the observed industry composition of frontier firms is discussed later in the paper, but first we must define what we mean by the productivity frontier.

3 Results

3.1 Definition of the productivity frontier

It is common practice to define the productivity frontier as the 90th percentile of productivity within an industry, and any firm within the top decile of productivity as a frontier firm (Andrews et al. 2015).[13] However, for New Zealand at least, the outcomes of this classification is very dependent on the assumed production function and on the industry comparator group.[14] Tables 5 and 6 provide a first indication of the scale of this issue by comparing productivity deciles of firms under different production functions (Cobb-Douglas vs translog, table 5), and using a consistent production function (Cobb-Douglas), but ranking firms at different industry aggregations (production function industry vs three-digit ANZSIC, table 6).[15]

[12] Summary statistics in this paper will, therefore, differ from official Stats NZ BOS statistics because of this dual reweighting; because we exclude non-response and "don't know" responses; and because of productivity population criteria (ie, in a productivity industry and always private-for-profit).

[13] While Andrews et al. (2015) note that a percentile approach is common practice, they adopt a "top X firms" frontier measure because the coverage rate of the data they use (OECD-ORBIS) varies substantially over time.

[14] While not tested in this paper, the determination of frontier firms is also likely to be affected by the industry grouping over which the production function is estimated.

[15] There are 39 production function industries (listed in table A.1), and 191 three-digit industries in the 2006 Australia-New Zealand Standard Industry Classification (ANZSIC) that are in production function industries. For the detailed industry measure,we exclude industries with, on average, less than 50 firms in the population (ie, where less than five firms would be included in the top decile in any given year).

In each table, the main diagonal where productivity deciles match is shown in bold. With complete agreement between any two decile allocations, the main diagonal components would each be 0.1, since each decile contains 10% of firms, and the off-diagonal components would be zero. In practice, while agreement between Cobb-Douglas (CD) and translog (TL), and between Cobb-Douglas at the production function industry level and at the three-digit (detailed) industry level, is the most likely outcome, it is far from a certain outcome, particularly for deciles close to the median productivity. In terms of defining a productivity frontier, the value in the bottom right-hand corner of the tables is the most pertinent, indicating that the agreement between metrics on the top decile of firms is around 80%. When these measures disagree on the frontier, it is generally true that the firm is in the 9th decile on the other metric though, in the case of the CD-TL comparison, there is a non-zero probability of being a top decile firm on one measure and a bottom decile firm on the other.

Table 7 extends this analysis to all eight permutations of CD and TL production functions combined with production function industry or detailed industry (subscript "*3d*") comparison. To make this extension manageable, each column conditions on being in the top decile under a particular metric, and then each row reports the conditional probability of being in a given decile of the other metric. For example, the first column of table 7 corresponds to the last column of table 5 (multiplied by a factor of ten), since it reports the CD decile conditional on being in TL decile ten. Focussing on the bottom row of the table, we can see a common pattern across all pair-wise comparison. In the best case scenario – comparing translog productivity industry and detailed industry – we have agreement on the top decile firms at most 84% of the time.

Table 8 tests whether this disagreement could be driven by transitory membership of the top decile, by conditioning on presence in one frontier over consecutive years. To do this, the sample is restricted to firms where MFP is observed in consecutive years, and we restrict the comparison to CD and TL at the productivity industry level. Given this temporal restriction of the population, the first two columns repeat the first two columns of table 7 on the reduced sample, confirming that the basic pattern in the data is not affected by the restriction. The next two columns report productivity deciles on one metric at time t, conditional on being in the top decile on the other metric two years running (t and $t-1$). While consistency increases slightly from the base case – 83% compared to 80% – the general conclusion remains the same, that top decile performance metrics are substantially different depending on somewhat arbitrary choices regarding production function or industry granularity.

The last four columns of table 8 condition solely on being in a top decile on one metric in the prior year ($t - 1$). In this case, not only can we compare the consistency of two difference productivity metrics, but we can compare the consistency of a single metric in consecutive years (last two columns). Probabilities of agreement fall quite substantially in both cases, reflecting the fact that some firms do not have persistently high productivity over time.

To sidestep the issue of picking a particular productivity metric to identify the industry-specific frontier, knowing that an alternative choice would yield only an eighty percent overlap, we construct a composite measure using all four top decile definitions. The composite measure requires a firm to be in at least three of the four top deciles in a particular year to qualify as being a frontier firm.[16] By requiring at least three out of four top deciles, we ensure that a frontier firm must be considered high productivity for each production function choice (CD or TL) and for each industry granularity (productivity industry or 3-digit ANZSIC).

Table 9 disaggregates the population by the number of CD and TL top deciles that each firm is a member of, with combinations satisfying the frontier definition italicised. Overall 7.9% of firm-year observations are in this productivity frontier, while 86% of the population are in no frontier, and the remaining 6% of observations satisfy one or two top decile metrics and are excluded from the frontier. Most frontier observations (83%) are top decile firms under all four metrics.[17] The remainder of frontier observations having a tendency towards being in both TL top deciles (and one CD top decile), over being in both CD top deciles (and one TL top decile), which follows from the greater top decile consistency for TL over CD when comparing productivity industry and detailed industry (table 7, middle four columns).

A key feature of the frontier is that, unlike a top decile-based measure, this frontier does not contain 10% of firms, nor does it have to contain the same proportion of firms over time or across industries.[18] In practice, time

[16]Rather than exclude small industries where we do not calculate detailed industry measures, we require a firm to be in both the CD and TL productivity industry top decile to be counted as a frontier firm.

[17]This percentage is derived from the total proportion in all four top deciles (bottom right italicised cell, 0.066) divided by the total proportion of observations in the frontier (sum of all italicised cells, 0.079).

[18]The initial choice of focussing on the top 10% of firms is arbitrary, so that deviating from that percentage does not matter.

variation in the proportion of frontier firms is minimal, while variation across productivity industries is more substantial. Figure 2 demonstrates the productivity industry variation in the proportion of firm-year observations in the frontier (grey bars) relative to the overall mean (dashed line).[19] Industries are ordered based on their contribution to total frontier observations (black bars), which is a function of industry size, and the proportion of observation in an industry that are in the frontier (the industry frontier rate). The industry frontier rate varies from a low of 6% for Food, Beverage & Tobacco Manufacturing (industry CC1), to 9.3% for Non-auxiliary Finance & Insurance Services (industry KK1_).

The final column of table 10 shows the proportion of firms in the frontier by year, which is stable over time. To enable comparison of firm productivity across something approximating productivity deciles, this table also demonstrates a classification of firms into approximate "deciles" where firms in the top "decile" are frontier firms. Specifically, firms-year observations are allocated to the median (rounded down) of the four decile metrics, which we label the "composite" productivity decile.[20] Since our frontier has less than 10% of the observations, other "deciles" must have more than 10%, resulting in "deciles" one through seven being overrepresented in the composite measure.

For the BOS-productivity subpopulation, we expect the frontier firm rate to be higher than in the productivity population, since average firm productivity is higher in the former (tables 3 and 4). Figure 3 confirms this expectation, with the average BOS-productivity frontier rate being 8.6% (dashed grey line). Additionally, the sample survey (lower coverage) nature of BOS introduces substantial time variation into the BOS sample, with the frontier rate (grey bars) exceeding 9% in 2007, 2008 and 2016. Figure 4 compares the productivity industry frontier rate across the two populations, with industry bubbles scaled to industry size in the BOS-productivity subpopulation. As with the temporal variation, BOS has more variability in the industry frontier than the productivity subpopulation, due to sampling and because the six employee cut-off affects the firm productivity distribution differently across industries. In particular, Professional, Scientific & Technical Services (industry MN11), which is the second-largest BOS industry accounting for over ten percent of observations, has a frontier rate of 12.3% compared to an 8.4% frontier rate in the productivity subpopulation.

[19] The productivity industry classification is explained in appendix table A.1.

[20] For example a firm that is in the seventh decile of each CD metrics (productivity and detailed industry), but in the eighth decile of each translog metrics, would be in the seventh composite decile (median of the four metrics is 7.5, rounded down to 7).

Given this substantial variation, subsequent regression-based estimates of the relationship between firm characteristics and presence in the frontier always include productivity industry by year dummy variables to control for cross-industry variation – and, in the case of BOS, cross-time variation – in the frontier firm rate.

3.2 Persistence in the productivity frontier

The final two columns of table 8 show that firms often drop out of the top productivity decile over time. Similarly, there will be firms whose high productivity potential is only realised over time, leading to entry into the productivity frontier. In this subsection, we expand this transition analysis for the four productivity decile metrics and, for the firm frontier.

Table 11 shows the year-on-year transition matrix between Cobb-Douglas productivity industry deciles. For a given row, the top panel shows the probability of being in each CD productivity decile in the current year, conditional on being in a given CD productivity decile in the previous year. Because the population excludes exiting and entering firms, firms observed at either time $t-1$ or t must be active in both time periods, however these firms may be unobserved in one of those years because they are outside the population in that year (eg, transition to/from WP-only) or have missing productivity data. These firms are dubbed "joiners" and "leavers" to distinguish them from entrants/exiters and to highlight the fact that they join or leave our (observed MFP) subpopulation. The rightmost column of the table reports the probability of being in the top decile at t conditional on not being a leaver.

The main diagonal of the top panel (bold), where firms stay in the same decile between years, is the most likely state for non-leavers. Persistent decile ranking is strongest for the poorest performing firms at $t-1$ – with decile one persistence 34% unconditional, and 50% conditional on not leaving[21] – and the best performing firms at $t-1$ – with decile nine 31% unconditional and 61% conditional, and decile ten 50% unconditional and 63% conditional on not leaving. The leaving rate is particularly elevated for decile one firms, which likely reflects their poorer performance leading to shrinking employment (below the population threshold) and/or future exit (at $t+2$).

[21] The probability of staying in decile one, conditional on not leaving, is not reported in the table and is given by 0.340/[1-0.314] (ie, the probability of staying in decile one divided by the probability of all outcomes excluding leaving).

The bottom panel of the table reports the $t-1$ top decile row for each of the metrics (repeating the CD from the panel above), plus the composite metric top decile (which matches the frontier, as in table 10). Probabilities of remaining in the top decile are fairly consistent across each metric –in the range of 47-50% unconditional, and 60-63% conditional on not leaving – but lowest for the productivity frontier (composite metric).[22] Overall, therefore, while the frontier is the productivity decile with the greatest persistence, a significant proportion of firms in the frontier will drop out of it from one year to the next, with non-trivial probability of falling into the bottom half of the productivity distribution. Conversely, a significant proportion of firms from decile nine move to the top decile over consecutive years, reflecting the somewhat arbitrary nature of the boundary.

Figure 5 plots the probability of persistence in the frontier for five years following presence in the frontier (at time t), together with the probability of being in the frontier in the five years prior to t. Three probabilities are plotted: the unconditional probability of being in the frontier in a given year (dashed line); the same probability conditional on not leaving (solid line) and the probability of being in at least one top decile across all four metrics, though not necessarily in the frontier, conditional on not leaving. Focussing on the solid line (which is directly comparable to the bottom right figure of table 11), the conditional probability of survival on the frontier falls most rapidly one year out from being on the frontier (to 59%), and then steadily declines to 43% five years on. Figure 6 reinforces this point by plotting the associated hazard rate, ie, the firm exit rate from the frontier as a proportion of those remaining in the frontier at the prior year. The dotted line in figure 5 shows that, while attrition from the frontier is quite severe over time, a significant proportion of firms that exit the frontier are still top decile firms on one or two metrics.

Despite attrition from the frontier, a significant proportion of frontier observations are associated with firms that have been in the frontier for a substantial period of time. Table 12 reports frontier statistics by the number of years that a firm is observed in the productivity subpopulation. The table also reports the number of observations and firms showing that, while the majority of firms have been observed for four or less years, the majority of observations are from firms that are observed for at least nine years (columns 1 and 2). The frontier rate (column 3), however, is largely uncorrelated with how long the firm has been observed. The probability that a firm is ever

[22]The conditional persistence in decile ten for CD and TL are the same result reported earlier in bottom right-hand corner of table 8.

observed in the frontier is positively correlated with the number of years observed (column 4), rising from 9% for firms observed once to 29% for firms observed in all 14 years. The final column of table 12 shows the proportion of firms that spend the majority of their observed time in the frontier. While a noisy measure of persistence for firms that are observed a small number of times, this statistics doesn't systematically vary with the number of years observed, meaning that a large share of frontier firm observations are associated with firms that have been in the frontier for many years.

Figure 7 shows how these patterns affect the aggregate distribution of frontier firm observations by number of years observed in the frontier. For example, the figure shows that firms that have been observed for 14 years account for almost 16% of frontier observations and that, of those frontier observations, the majority are associated with firms that have been in the frontier for at least six years. In total, half of frontier observations are associated with firms that are in the frontier for at least four years.

3.3 Contribution of frontier to aggregate

Figure 8 plots the aggregate share of firms, output, inputs, and value-added by composite productivity decile, showing an increasing share of each by decile, except for the number of firms (see discussion of table 10). The fact that economic activity is concentrated in the upper deciles of the productivity distribution, is consistent with a properly functioning economy that allows resources to shift to more productive uses. While frontier firms constitute only 8% of firm-years, they account for 13% of total labour input, 27% of aggregate gross output, 25% of intermediate consumption, 29% of value-added (Y−M) and 22% of aggregate capital services.

Even relative to decile nine firms, frontier firms are substantially more important to aggregate output. Controlling for the difference in firm-year observations, an average decile nine and frontier firm use the same average labour input, but frontier firms produce 87% more value-added. Figure 9 shows this difference in terms of aggregate labour productivity, with each individual year shown in grey lines, and the average level over the first three and last three years shown in black lines (dashed and solid, respectively).[23] We focus on labour productivity, in part because the frontier definition abstracts away from a particular preferred metric of MFP, and because aggregation

[23] We average over three years to minimise the risk that the choice of endpoints unduly affects the measured productivity growth over the entire period.

and decomposition are easier to explain with a single input (labour).

Several features of figure 9 are worth noting: aggregate labour productivity is increasing in composite productivity decile; substantial dispersion in productivity across firms is evident, with the frontier being nine times as productive as the bottom decile of firms; frontier firms and bottom decile firms are substantially different from other firms; and aggregate labour productivity has improved over time in every decile, consistent with the aggregate trend (figure 1).

Some of the labour productivity advantage of frontier firms is explained by the capital-labour ratio, as shown in figure 10. Pooling all years by decile, the figure shows aggregate labour productivity (solid line), the aggregate capital-labour ratio (dashed line), and aggregate capital productivity (dotted line), which (in logs) is the difference between the other two series. The aggregate capital-labour ratio is quite stable over deciles one to eight, but rises for decile nine and for the frontier (decile ten), explaining some of the labour productivity advantage of frontier firms. Of course, when we control for both labour and capital inputs, it must also be the case that frontier firms are more productive, since they are the top firms ranked by MFP. Interestingly, the relatively poor aggregate labour productivity performance of bottom decile firms is not explained by relative capital shallowness, rather, the worst productivity firms have very low aggregate capital productivity and labour productivity. As with labour productivity, capital productivity is generally increasing in composite productivity decile, though this is not the case for decile nine firms, which exacerbates the labour productivity gap between decile nine and the frontier.

The top panel of table 13 shows the annualised growth rate in aggregate components (and total number of firms) by composite decile, where growth is measured across the two three-year time periods shown in figure 9. In percentage terms, aggregate labour growth and labour productivity growth was strongest in the bottom productivity decile (3.3% and 2.8% per annum respectively), while aggregate labour productivity growth was slowest in the frontier (0.4% per annum). Decile nine was the only decile to experience an aggregate decline in gross output, offset by an even larger decline in intermediate consumption, so that value-added growth was still positive. For the entire subpopulation, the annual growth rate in aggregate labour productivity was 0.83% per annum (rightmost column).[24]

[24] For the interested reader, figure A.1 in the appendix plots changes over time in the (translog) MFP distribution on an industry-by-industry basis, showing that some gaps between industry frontier and non-frontier firms closed over time while other gaps in-

While the relatively low growth rate in frontier labour productivity gives the impression that the frontier does not make a substantial contribution to aggregate labour productivity growth, that ignores the fact that the frontier initially has much higher productivity (figure 9) and greater share of inputs (figure 8). To understand the impact of this dominance on aggregate labour productivity growth we do two things.

Firstly, in the bottom panel of table 13, we report the contribution of each decile to aggregate growth in output and inputs (ie, each row sums to 100%, reflecting the total aggregate change in the productivity component). When weighted by contribution to the aggregate, the relatively rapid growth of bottom decile value-added is less impressive, contributing only 4.3% of the aggregate change. In contrast, the frontier accounts for 21% of aggregate growth in value-added. In aggregate labour productivity terms, this disparity in contribution is exacerbated by the fact that the bottom decile grabbed a disproportionate share of aggregate labour input growth (15.8%) compared to the frontier (11.4%), which is not commensurate with the relative contribution to output growth.

Secondly, we quantify the importance of the frontier to aggregate labour productivity growth by performing counterfactual exercises where we suppress labour productivity growth and/or labour input growth in the frontier, and calculate the implied aggregate labour productivity under these scenarios. The leftmost column of table 14 reports observed aggregate labour productivity levels in the (three-year averaged) start and end periods, the dollar value change in productivity, and the annualised growth rate associated with that change (0.83% per annum, as in the top panel of 13). Subsequent columns repeat these statistics, but turning off frontier firm labour productivity growth (column 2), employment growth (column 3-4), or growth in both (column 5-6). For the counterfactual exercise, we assume that other deciles are unaffected by the absence of change in the frontier, except that we assume aggregate labour input growth remains the same under all scenarios, allocating frontier labour growth to either decile nine, or to all deciles one through nine. The former allocation imposes positive sorting of high skill workers into high productivity firms (which we confirm later in the paper), while the latter ignores matching between workers and firms.

creased. Figure A.2 summarises these trends by plotting the initial gap between the 90th and 50th percentile of the MFP distribution against the change in the size of this gap. Overall, the relationship between the initial gap and the change in the gap is mildly positive, implying that the gap between frontier firms and laggard firms tended to increase more in industries where the gap was initially larger.

Assuming worker sorting applies, table 14 shows that the impact of no frontier labour productivity growth (column 2) and no labour input growth (column 3) are roughly equal, reducing aggregate annual labour productivity growth by 0.11% and 0.14% respectively. When combined (column 5), these impacts are almost additive, reducing the annual growth rate from 0.83% to 0.59% or, equivalently, shaving over $3,400 off average 2016/18 labour productivity.

Obviously, the simplicity of the counterfactual assumptions raise questions about the precision of these estimates, however it is unclear whether they represent an under- or over-estimate of the impact of the frontier. For example, we could argue that additional high skill workers at decile nine firms should raise decile nine firm productivity measurably, since the added labour is a significant proportion (9.7%) of aggregate L at those firms. Alternatively, we could argue that the absence of labour productivity growth at the frontier should stunts productivity growth in other deciles because fewer new technologies become available to be adopted by firms following the industry leader (as in Bartelsman et al. 2008).

We conclude this section by noting that the estimated contribution of the frontier is not the same as estimating the contribution of incumbent frontier firms to subsequent productivity growth, since firms move in and out of the frontier (see, eg, figure 5). Instead we are examining how the frontier itself – with constantly evolving membership – pushes forward productivity growth, not the firms in the frontier at any particular point in time.

3.4 Characteristics of frontier firms

3.4.1 Mean differences between frontier and non-frontier firms

Having identified the importance of the frontier to aggregate labour productivity growth, we now set out to establish the non-productivity characteristics of firms in the frontier. Table 15 reports means of full coverage variables based on whether a firm is in the frontier or not. Stars indicate a significant difference between the frontier and non-frontier mean, with most differences being significant at the 1% level (represented by three stars).

Frontier firms tend to be larger in employment terms, but less capital-intensive, than non-frontier firms (top two rows of table 15). Digging into these differences further, figure 11 shows how the probability of being in the frontier varies with each of these characteristics, using a smoothed propensity from a local polynomial regression (solid line with 95% confidence interval). For reference, the figure also plots the linear estimate of the relationship (dashed black line with 95% confidence interval), the mean proportion of firms in the frontier (horizontal grey dashed line), and the 25th and 75th percentile of the independent variable (vertical grey dashed lines). These latter lines are a useful guide to focus attention on the bulk of the distribution of observations.[25] For total labour input, frontier firms are significantly more likely to appear above twelve employment (l>2.5) with the probability of being a frontier firm rising by over one percentage point (from 7.8% to over 9%) for firms over 20 employment (l>3). Frontier firms being larger, on average, than non-frontier firms is consistent with more productive firms attracting a larger share of aggregate inputs.

There is much greater variation in the frontier probability over the capital-labour ratio. Frontier firms are more concentrated in the bottom quartiles of the k-l ratio with the probability of being a frontier firm rising by over two percentage point for firms with a capital-labour ratio below 8.5 (around $4,900 per unit of labour). While the average frontier firm has a lower k-l ratio than the average non-frontier firm, the aggregate capital-labour ratio is highest in the frontier (figure 10) and the frontier makes the largest contribution to aggregate capital (figure 8) implying that some frontier firms must be very capital intensive. Subsequent regressions of frontier membership on firm characteristics control for industry, but it is worth remembering that these mean characteristics include differences derived from cross-industry variation. In particular, the difference between the firm average and aggregate k-l ratio picture are, in part, driven by the composition of the frontier being dominated by relatively low k-l ratio service sector firms (figure 2).

As well as having more labour input, frontier firms attract more skilled workers and pay a higher firm wage premium than non-frontier firms (third and fourth rows of table 15). Since worker skill and FFE have been normalised, the table shows that average worker skill in the average frontier firm worker is 31% of one standard deviation above that of the equivalent non-frontier worker skill level, and the average frontier firm pays 41% of one standard deviation higher firm wage premium than the average non-frontier firm. Figure 12 shows how the probability of being in the frontier varies over average worker skill and the firm wage fixed effect. Higher average worker skill in the frontier is a potential explanation for higher frontier firm pro-

[25] The top and bottom 5% of the independent variable are also trimmed for presentation purposes and for confidentiality.

ductivity (see, eg, Maré et al. 2017), while a high firm wage premium is consistent with rent-sharing associated with high firm productivity (see, eg, Card et al. 2018).

Figures 13 and 14 show how mean differences (dotted line) between the frontier (solid line) and non-frontier (dashed line) have varied over time for labour (l), the k-l ratio, average worker skill, and the firm wage premium (FFE). In the case of labour input, variation in the gap seems to derive mainly from employment responses following the GFC, where non-frontier firms shrank in size and remained smaller through to 2012, while frontier firm average firm size grew steadily over that period. Subsequent non-frontier employment growth has been more rapid than frontier growth, though, overall, these changes are small relative to average firm size. Changes in the k-l ratio gap have been more substantial with the frontier and non-frontier firm averages trending in different directions, resulting in the gap increasing from \$4,200 per FTE to \$7,100 per FTE.

The firm-level average worker skill measure increased by similar amounts in both frontier and non-frontier firms following the GFC, consistent with lost jobs being concentrated in lower skill occupations (top panel of figure 14). Post-GFC, frontier firm skill has declined slightly while non-frontier firm average skill has been static. The firm wage premium is steadily declining for both types of firm, though at a slightly higher rate for non-frontier firms leading to a slight increase in the gap between average frontier and non-frontier firm wage premia.

Remaining significant differences between frontier and non-frontier firms (table 15) indicate that frontier firms are marginally younger than non-frontier firms, and that they are more likely to be in most major Urban Areas (presence in UA variables). The same spatial pattern is apparent in UA employment shares, though is muted by the fact that frontier firms are one percentage point more likely to be located in multiple regions (bottom row of table 15), implying that their employment share in any particular region will be lower than a single region firm.

Table 16 makes use of the richness of the BOS dataset to explore other differences between frontier and non-frontier firms. This table follows the same format as table 15, but also includes unweighted and weighted observation counts by variable, since we exclude unusable observations on a case-by-case basis, and because some variables are not collected in each year. Many of the BOS variables are binary characteristics and, for those variables, we also report the proportion of firms with that characteristic that are frontier firms, together with the number of firm-year observations with that characteristic averaged over the years in which the characteristic is observed (table 17). Other characteristics can be measured conditional on doing an activity, eg, R&D intensity conditional on doing R&D, and table 18 reports mean differences in these intensity variables conditional on activity.

Export-related variables appear at the top of each of the three tables and provides a good example of how the tables work together. Export intensity is recorded in the annual component of BOS, but separate from new export market entry, which is a binary yes/no question. Because firms leave the export question unanswered this variable has lower coverage than the export market entry question (approximately 5,000 fewer responses, table 16).[26] Frontier firms are more likely to be exporters and have higher export intensity than non-frontier firms, though the former difference is not statistically significant at the 10% level (table 16). As a consequence, frontier firms make up 9.2% of the average 4,800 exporting firms in the BOS-productivity population each year, compared to the base frontier firm rate of 8.6% (table 17). Conditional on being an exporter, the gap in export share is nine percentage points (pp) and non-frontier firms are significantly more likely to enter new export markets in a year (table 18).

Focussing on other significant relationships across these tables, we see that frontier firms are significantly more likely than non-frontier firms to be foreign-owned and to have foreign-ownership stakes in overseas ventures (outward direct investment, ODI). They make up almost 17% of all foreign-owned firms and 12% of all ODI firms in the BOS-productivity population. The FDI gap is very stable over time, with both frontier and non-frontier FDI rising through to 2011 and then dropping off (figure 15).

Frontier firms are 7pp more likely to report that their core equipment is fully up-to-date compared to the best commonly available technology (table 16), though the gap between frontier and non-frontier firms has been closing over time (figure 16). Frontier firms are also more likely to believe that they have no competitors. Overall, only 4% of firms in the population report having no competitors (corresponding to 1,165 firms per year on average), and 14% of those firms are in the frontier which could reflect frontier firms ability to produce differentiated products, or could reflect the shortcomings of revenue-based productivity metrics which may confound high mark-ups

[26] The "has exports" indicator variable is constructed from responses to the export intensity question. Changes to the survey in 2012 caused some firms to switch from reporting 0% to 1% of sales as exports, after being instructed to round up to 1% if sales were non-zero but <1%. To ensure consistency over time, we set responses of 1% (in any year) to zero, and then count firms with non-zero adjusted exports as exporters.

(due to lack of competition) with high productivity.

Human resource management questions are asked every four years, substantially reducing the number of firm-year observations available (bottom of first page of table 16). The HRM practices indexes is compiled from nine management questions that are reduced to binary high-low responses, added and then converted into a z-score (ie, mean zero, standard deviation one).[27] Frontier firms have HRM practices on average almost one quarter of a standard deviation higher than non-frontier firms. Frontier firms are also more likely to have ultrafast broadband (UFB). The complementary information and communication technology (ICT) investments variable measures how many different investment types firms make to improve the benefit they get from their ICTs.[28] As with HRM practices, complementary ICT investments is a z-score based on binary responses. Differences in complementary ICT investments are insignificantly different between frontier and non-frontier firms, despite the fact that frontier firms are more likely to have UFB and fully up-to-date core equipment.

The second page of table 16 focuses on R&D, broader measures of expenditure on product development, and innovation outcomes. Unlike the exporting question, these expenditure questions are dollar value responses which we normalise by intermediate consumption (M).[29] Questions on product development and two-yearly innovation are asked in BOS every second year, with the former not being in the survey in the first year of running.

[27] Respondents must answer at least eight of the nine questions, with the HRM questions covering (BOS 2005 question number in brackets): job satisfaction (Q25); performance reviews (Q26); performance pay (Q27); training (Q28); assessing skills gaps (Q30); health & safety (Q31); incorporating employee requirements into business goals (Q6); measuring firm HRM performance (Q20); and identifying skills-related risks and opportunities (Q23). Questions with scale responses are converted to binary high/low indicators where the grouping of responses divides the population as evenly as possibly (following Fabling and Grimes 2014).

[28] Investment categories are: Changing staff level/mix; training; new work practices; restructuring; new business strategies/management techniques; relocation; investment in other forms of capital; R&D; redesigned production/distribution processes; and shifting output mix towards ICT intensive products.

[29] To be consistent with this normalisation, we also deflate values using the productivity industry-specific input price deflator. In rare cases, reported investments exceed M, in which case we set the relevant ratio to one to avoid the impact outliers could have on means or estimated regression coefficients. A ratio of more than one does not imply that reported expenditure or productivity data (or both) are incorrect since M does not capture all components that might be included in the reported expenditure. For example, a large component of R&D expenditure is employee wages, which is not a component of intermediate consumption.

Research and development questions, and the one-yearly innovation question are asked in every year of the survey, with the latter asking a single question covering all four two-yearly innovation categories.[30]

The propensity to do R&D is not significantly different for frontier firms compared to non-frontier firms, though the R&D share of expenditure is higher. Frontier firms are less likely than non-frontier firms to have expenditure on product design, but spend more (on average) on product design and on marketing and market research (table 16). Conditional on investing in R&D, the average share of expenditure spent is two thirds higher in frontier firms compared to non-frontier firms (15.7% compared to 9.4% of M, table 18). Conditioning on any product development expenditure more generally, frontier firms spend significantly more on all categories of expenditure except "other." Average total product development costs as a proportion of intermediate consumption in frontier firms, therefore, is almost double what it is in non-frontier firms. Since total product development costs are only observed every second year, figure 17 plots the frontier firm gap in R&D intensity over time. While the measure for frontier firms is noisy, there is tentative evidence that both frontier and non-frontier intensity are growing, with the latter growing at a somewhat slower rate resulting in a slowly increasing gap over time.

In contrast to the product development statistics, reported innovation outcomes are, on average, weaker for frontier firms across all metrics (bottom of table 16), with the innovation rate gap ranging from 2.2pp for new product innovation to 5.6pp for new marketing methods. For any reported successful innovation at all, frontier firms trail non-frontier firms by 5% on the one-year measure and 6% on the two-year measure. Figure 18 shows the one-year innovation gap over time, which shows that the significant difference between frontier and non-frontier firms seen on average across all years is mainly a pre-2012 phenomenon where innovation rates were relatively high for non-frontier firms.

One way to reconcile the overall innovation investment-outcome deficit of frontier firms – that also gels with their superior MFP – is that being at the frontier requires greater investment in order to innovate successfully (push back the frontier). Non-frontier firms, in contrast, have the ability to follow behind adopting at lower cost than the original investment by frontier firms.[31]

[30] The "any of the above" two-yearly innovation question is not an asked question and is an indicator variable set to one if any of the four two-year innovation questions were answered positively, and zero otherwise.

[31] This positive externality argument is often cited in support of R&D tax credits and other

Following from this, innovations must vary in their value – both to the innovating firm, and to the wider economy – meaning that binary measures of success are unlikely to adequately capture the (quality-adjusted) volume of innovation output, which may be higher for frontier firms. Firms themselves must make this quality assessment and it is plausible that frontier firms set a higher standard for responding positively to the BOS innovation questions, which then manifests as lower measured innovation at the frontier. Alternatively, non-frontier firm innovation expenditure may be underestimated because the investment is more ad hoc in nature or not captured by the BOS questions (which has an R&D/product development focus). A further possibility is that frontier firms actually have relatively low productivity at producing (quality-adjusted) innovation, though this seems inconsistent with their superior MFP performance over non-frontier firms.

3.4.2 Regression-based estimates of differences between frontier and non-frontier firms

Taken together, the exceptional characteristics of frontier firms support the hypothesis that these firms are productivity frontier firms, rather than a set of firms grouped by the relative size of measurement error. Tables 19-26 extend the analysis of frontier firm characteristics by regressing an indicator variable for frontier membership on one or more firm characteristics plus a full set of (production function) industry by year dummy variables. The purpose of these regressions is not to establish causal relationships from activity to frontier membership, but rather to disentangle whether the observed univariate relationships are robust to the addition of controls, such as industry or firm size. For example, firms in the frontier may simply have better management practices than non-frontier firms because they are larger and, therefore, have more need for formal management systems.

Most tables present estimated regression coefficients (and standard errors) for all sectors pooled, following an identical format where estimates are based on ordinary least squares (OLS) in the first two columns and firm fixed effects (FE) regressions in the next two columns. For each regression method the leftmost column estimates the relationship between frontier firm presence and the reported variables in the table, broken into thematic subgroups. The adjusted R^2 for each regression is reported at the bottom of the table, which provides an indication of the variable groupings.[32]

innovation policies.

[32] The adjusted R^2 penalises the R^2 for the number of explanatory variables included in

For each regression method (OLS, FE), the rightmost column then conforms to a more standard regression table format, reporting coefficients estimated from a single regression including all variables in the table and, potentially, additional unreported coefficients as described in the main text and table notes. We will focus mainly on columns 1, 2 and 4 which represent a natural progression of an increasing number of simultaneous independent variables and controls. Tables 21 and 23 repeat multivariate specifications separately estimated for each sector.

Table 19 reports coefficients for productivity dataset variables, with Urban Area (UA) employment share coefficients reported separately in table 20 for space reasons. Column 1 of table 19 reports estimated coefficients for labour, the *k*-*l* ratio, worker skill, firm wage premium and log firm age where each variable is included in a separate OLS regression (including industry-year dummies). These coefficients are consistent with the mean characteristic gaps (frontier vs non-frontier) reported at the top of table 15 and with the estimated linear relationships in figures 11 and 12, but controlling for industry composition. These coefficients allow us to calculate the change in probability of being in the frontier based on a changed characteristic (where we have assumed a linear relationship). Using a transition from the 25th to the 75th percentile of the independent variable (reported in table 3) as an appropriate metric for a substantial change (holding other variables constant), column 1 coefficients imply that such a move is associated with an increase in the probability of being in the frontier of 0.9pp for labour, 3.9pp for average worker skill, and 4.2pp for the firm wage premium.[33] Similarly, a 25th to 75th percentile move is associated with a decrease in the probability of being in the frontier of 5.4pp for the *k*-*l* ratio and 0.7pp for firm age. These estimated changes are substantial when compared to the average proportion of firms in the frontier (7.9%).

---

the regression. We include a large number of industry-year dummies (39 industries by 14 years) in the regressions, and our earlier analysis (figures 2 and 3) suggest these could be replaced by industry dummies for productivity variable related OLS regressions – or excluded entirely for FE regressions – with little loss of explanatory power, because of the absence of time variation in the frontier. Industry-time dummies are potentially necessary for the BOS analysis because sampling introduces substantial variation in the frontier rate. For consistency we include the full set of industry-year dummies in all regressions, noting that this will impact on the apparent explanatory power (adjusted R^2) of productivity variable based regressions.

[33] For example, for average worker skill, moving from the 25th to 75th percentile of the distribution is a change in skill of 1.329 (= 0.625 – [−0.704], eighth row of table 3). Multiplying this change by the column one (table 19) coefficient on skill yields an implied 3.9pp ($= 0.029 \times 1.329$) change in the probability of being in the frontier.

Column 2 of table 19 reports OLS estimated coefficients when all productivity covariates are included simultaneously, including the UA employment share variables in table 20. Patterns are generally similar to column 1, though the sign on the labour coefficient reverses, due to the simultaneous inclusion of the firm wage fixed effect (FFE), which is postively correlated with firm size. Fixed effect estimates where labour is the only independent variables (top row, column 3) also support the hypothesis that smaller firms are more likely to be in the frontier, once permanent firm characteristics are controlled for. Fixed effect estimates with all variables included (column 4) are similar to OLS estimates, noting that the firm wage premium variable must be dropped as it is a permanent characteristics of a firm (by construction). Firm age is no longer a significant correlate of frontier presence once other observed and permanent (unobserved) firm characteristics are controlled for. Overall, these estimates support the view that firms in the frontier are less capital intensive than non-frontier firms, controlling for industry composition, and over time are more likely to be in the frontier in periods when they are less capital intensive. Average worker skill estimates support the notion of that skilled workers are a key contributor to higher productivity in firms, supporting the inclusion of worker quality as a missing input in the standard New Zealand industry production function (Maré, Hyslop, and Fabling 2017).

Table 20 reports related results for UA employment shares, where Auckland Urban Area is the omitted category. The first two columns of table are consistent with results reported by Maré (2016), who finds that Auckland firms have a productivity premium over firms in other regions but that this performance premium is eroded when the quality of labour inputs are accounted for.[34] In our results (column 1), the Auckland productivity premium converts into Auckland-only firms having a 0.9-2.7pp higher probability of being in the frontier than firms only located in other main Urban Areas. A substantial proportion of firms are not located in main Urban Areas (bottom of table 3), and for these firms the probability of being in the frontier is 3.1-4.2pp lower than an Auckland-only firm, after controlling for differences in industry composition including the increased likelihood that non-urban firms are in the primary sector. Once we control for other productivity characteristics (column 2 of table 20) the Auckland frontier "advantage" is eroded substantially, consistent with the observed productivity advantage of

[34] The most directly comparable results in Maré (2016) are table 3, columns 2 and 4, which are comparable to our table 20, columns 1 and 2 respectively. The main points of difference between the two analyses are the different production functions, differences in the productivity population restrictions, and weighting.

Auckland firms coming, in part, from access to higher quality labour inputs.

Adding firm fixed effects as well as productivity variables (column 4), we still find an economic and statistically significant lower probability of non-main UAs in the frontier compared to Auckland firms (with similar coefficients to column 2). This final specification is quite stringent, given the persistence of most firm location choices over time, which is reflected in the larger standard errors compared to the OLS specifications.[35]

Table 21 breaks down the productivity covariate results by sector, where the UA employment share variables are now relative to all main UAs for parsimony in the table, and because this appears to be the main common feature of the pooled OLS and FE results.[36] Most coefficients are consistent across sectors, except (log) firm age, for which the near-zero average coefficient conceals substantial sector-level variation. Focussing on the fixed effects coefficients (four rightmost columns), these results suggest that, in construction, older firms are more likely to be in the frontier than younger firms, while the opposite is true in the manufacturing and services sectors. Urban Area employment share results also vary by sector, with the primary sector – perhaps unsurprisingly – not experiencing a frontier penalty for being outside main UAs. None of the fixed effect coefficients on UA shares are significantly different from zero at the sector level, because these are now relative to all main UAs, not just Auckland, which reduces the frontier gap.

The modular structure of the BOS means that we cannot estimate models with all BOS variables at the same time, since some questions never appear in the same year as other questions. Instead, we break BOS-related results down into those that are surveyed annually (itemised in table 22) and having sufficient observations to enable sector-specific estimates (table 23), and the remaining BOS topics, which are covered less frequently and which we restrict to whole economy estimates (tables 24-26). In all BOS-related regressions, multivariate models include productivity variables and, in the case of non-annual variables, also include annual BOS variables as additional controls.

Table 22 reports estimated coefficients for annually measured BOS characteristics. We discuss these results sequentially in the column 1 groupings indicated by the adjusted R2 listing, starting with the three exporting variables. Incumbent exporters entering new export markets appear less likely

[35] According to Maré's (2016) analysis, these estimates may be biased downwards because the MFP estimates do not account for spatial price differences across regions.

[36] Figure A.3 in the appendix graphically summarises coefficients on table 21 productivity covariates on an industry by industry basis.

to be in the productivity frontier by 2.0-2.7pp depending on other controls (and assuming the overall export share remains constant), which may reflect a transitional issue consistent with Fabling and Sanderson (2013) who observe significant gearing up of capital for incumbent manufacturing exporters entering new markets. In specifications without firm fixed effects (columns 1 and 2), export intensity is positively related to frontier presence, with an offsetting negative coefficient on export participation.

Combining these estimates, an incumbent exporter moving from the non-frontier to frontier mean export share (an increase of 8.8pp, table 18) has a 0.7-0.8pp increased probability of being in the frontier (columns 1 and 2 of table 22). Using column 2 estimated coefficients, a firm transitioning from non-exporting to the frontier firm average export share (48%) has a reduced probability of being in the frontier of 0.7pp in the entry year, and an elevated probability of being in the frontier of 1.4pp in the following year (all else the same).[37] The export intensity relationship disappears in the presence of firm fixed effects, which may be due to the persistence of export behaviour over time amongst firms in the BOS population.

Despite over-representation of foreign-owned firms in the frontier (table 17), the relationship between FDI and frontier presence is insignificantly different from zero when industry is controlled for (table 22, column 1). Similarly, ODI is not significantly related to frontier presence once other firm characteristics and/or firm fixed effects are included (columns 2-4).

Firms with higher research and development intensity are more likely to appear in the frontier, even controlling for firm fixed effects and other time-varying characteristics (column 4). This relationship is offset by a negative coefficient on R&D participation, so that a firm transitioning from not doing R&D to doing R&D at the average frontier firm intensity of R&D performers (15.7%, table 18) has a 0.9pp increase in the probability of being in the frontier, while an incumbent R&D performing firm moving from the non-frontier average intensity (9.4%) to the frontier average intensity (15.7%) has a 0.8pp increase in the probability of being in the frontier. Despite more R&D intensive firms being more likely to be in the frontier, the relationship between one-year innovation and frontier presence is negative (column 1), though the magnitude and significance of the relationship dissipate as additional controls and firm fixed effects are added.

[37] In this case, on the entry year we must sum both indicator coefficients since the firm must also be entering new export markets (ie, $-3.0 - 2.1 + 0.092 \times 0.48$). The "following year" calculation assumes stable export partners so that the new market coefficient is no longer applicable.

For parsimony, only the no collective employment agreement category is included in the regression, reflecting the key mean difference between frontier and non-frontier firms (table 16). As a consequence, the reported coefficient reflects the difference in frontier firm probability between having and not having any collective employment agreements. While frontier firms are less likely, on average, to have one or more collective employment agreements than non-frontier firms (table 16), there is no significant difference in this characteristic once industry-year is controlled for (column 1).

Similarly for core equipment and competition, we include a single category reflecting the group most over-represented by frontier firms – core equipment fully-up-to-date, and no competitors. In each case, OLS estimates suggest a positive relationship between having the characteristic and frontier presence, and that the relationship is robust to the inclusion of other BOS controls and productivity variables (column 2). The inclusion of firm fixed effects results in smaller and insignificant coefficients, which is likely due to technology and competition being slowly evolving firm characteristics.

Table 23 reports coefficients on annual BOS characteristics estimated by sector, and including all covariates simultaneously (including productivity variables). Focussing on coefficients that are significantly different from zero in column 4 of table 22 (ie, when all industries are pooled), table 23 shows that the relationship between entering new markets and frontier presence is strongest in the services sector, but is also negative (and insignificant) for other sectors. Similarly, the pooled R&D intensity coefficients are mainly derived from service sector firms, though the point estimate on the R&D share is also large for manufacturing.

Looking at OLS estimates, the zero overall relationship between FDI and frontier presence conceals underlying heterogeneity with the manufacturing and primary sectors having significant positive coefficients on the FDI share, which is the opposite of the services sector (coefficient negative and insignificant). Taking into account the contrary sign on the FDI indicator variable, compared to a wholly domestically-owned firm, the probability of being in the frontier is 0.056pp (0.016pp; 0.022pp) higher for a manufacturing (primary; services) sector firm that is foreign-owned at the mean conditional FDI share (80%, table 18).

Core equipment and competition coefficients are strongest in the manufacturing sector (table 23, column 1), followed by the services sector for core equipment ,and construction for competition, though the latter have large standard errors. Overall, the relatively small number of observations in the construction and primary sectors make it hard to pin down coefficients with any precision – a problem which becomes more pronounced if we attempt to estimate relationships at the sector level for variables that appear in only a subset of years. Because of this issue, remaining estimates are only estimated for all industries pooled.

Table 24 reports coefficients related to product development and successful two-year innovation. With the most complete control set (column 4), only the indicator variable for having product design is significantly different from zero (at the 10% level). The coefficient is negative, though the coefficient on product design intensity is positive (and significantly different from zero in OLS specifications). Taking the latter coefficient as plausible, but imprecisely estimated, a firm undertaking product design would need an expenditure share of 22% (ie, 0.028/0.128) or higher for the probability of being in the frontier to be higher than when the firm were not undertaking product design (all else the same). However, the average product design intensity, conditional on any product development, is much lower than this averaging 3.9% for frontier firms and 1.6% for non-frontier firms (table 18), implying that the average firm is less likely to be in the frontier when it is doing product design compared to when it isn't.

Product marketing/market research expenditure is generally positively (significantly) associated with being in the frontier, though insignificantly different from zero in the most stringent specification (column 4). Using the most conservative significant estimates (column 3), a firm at the mean frontier firm expenditure share, conditional on product development (5.6%), is 1.3pp more likely to be at the frontier than a firm not doing product marketing or market research.

Where statistically significant, innovation outcomes are negatively related to frontier presence. Only on the case of new operational processes is the coefficient significantly different from zero in the presence of non-innovation covariates, implying a 2pp increased likelihood of being in the frontier for firms that are not innovating operationally, compared to firms that are.[38] Looking at the pattern of coefficients across all four specifications, there is reasonably consistent evidence that both operational and new marketing method innovations are less likely in frontier firms, while new product and

[38] Columns 2 and 4 of table include annual BOS controls, which include the one-year innovation measure. Coefficients (standard errors) on this variable are: 0.000 (0.009) for OLS; and 0.012 (0.009) for FE. The one-year innovation variable is included because responses to this question in BOS are sometimes not internally consistent with two-year innovation measures, suggesting that the two sets of variables – one-year vs two-year – potentially capture different information about the scale or scope of innovations at the firm.

organisational process innovation appears unrelated to frontier presence.

Remaining estimates for HRM practices (table 25) and UFB usage (table 26) show a positive relationship between each activity and being in the frontier, in the presence of industry-year controls (column 1). The OLS-estimated relationship with HRM practices is robust to the inclusion of other variables (column 2) – which show that a one standard deviation increase in HRM practices is associated with a 1.1pp higher probability of being in the frontier – but not the inclusion of firm fixed effects. For UFB, the inclusion of other controls reduces the estimated coefficient to zero. Like product design expenditure, the relationship between complementary ICT investments and frontier presence is negative for OLS estimates (including other controls), and zero for FE estimates. The interaction between UFB and complementary investments is never significantly different from zero, though OLS point estimates suggest that firms that have UFB do not suffer the same frontier penalty from their complementary investments that non-UFB firms do (ie, the interaction term is positive and similar in magnitude to the investments main effect coefficient).

The HRM and UFB results differ substantially from the findings of Fabling and Grimes (2014) and Fabling and Grimes (2016) respectively, where those authors found a positive impact of HRM and UFB (interacted with complementary investments) on productivity, after controlling for firm fixed effects (or, similarly, estimating in first differences). There are a number of reasons why our estimates are less definitive than those results, some of which relate to specific differences between the hypotheses in those papers and those being tested here. Put another way, the purpose of this paper is to estimate comparable coefficients across the potential range of frontier correlates, rather than replicating prior studies (with additional years of data).

In the case of both the HRM and UFB analysis, key independent variables differ from those investigated by the original authors. Specifically, the individual HRM practices in Fabling and Grimes (2014) overlap, but do not exactly match, those used in this study, and Fabling and Grimes (2014) construct weighted practice indices using principal components, rather than z-scoring a simple count of practices. In the case of the UFB analysis, Fabling and Grimes (2016) focussed on the subset of complementary investments for which estimated interaction effects with UFB adoption are the strongest, rather than compiling the investment metric from all investment types as we have. More importantly, they also estimated the effect of UFB on productivity in first differences, treating investment as an already first differenced variable, whereas our FE specification relies on changes in the complementary investment index to identify the relevant coefficients (for consistency with the treatment of other firm practices in this paper). While both approaches are legitimate, they should be expected to yield different results.

More generally, there are good reasons to believe that the methodology in this paper limits the ability of firm fixed effects regressions to identify causal relationships between firm activities and productivity, relative to other New Zealand studies. This shortcoming relates to the lack of variation in our dependent variable,[39] and due to the potential for firms that actually switch in and out of the frontier to have consistent high-performance practices over time. Specifically, New Zealand productivity studies usually have an MFP variant as the dependent variable which is likely to be more responsive to a firm practice that has a causal impact on productivity than the frontier indicator variable, in part because only a small proportion of firms are close enough to the frontier that a moderate increase in MFP could lift them into the frontier. Indeed, most transitions into and out of the frontier are firms coming from or going to decile 9 of the productivity distribution (table 11), and firms close to the frontier are more likely to already have the characteristics of high-performing firms than firms in lower MFP deciles.

Additionally, any positive impact on MFP from adopting productivity-enhancing practices for firms already in the frontier does not affect frontier presence, meaning that average productivity in the frontier can grow due to specific firm activities, without any observed difference in the dependent variable.

Since we are interested in the characteristics of firms at the frontier, we have included variables such as firm size (l) as independent variables. However, these may also be outcome variables impacted by firm activities or investments, either directly or via an effect on productivity. If, for example, adopting good HRM practices raises MFP (as Fabling and Grimes, 2014, and others have found), then adopting firms may expand their market share and raise total employment as a consequence of their superior relative productivity. In that sense, the estimated modelled relationships in this paper do not reflect a causal model (and have deliberately not been interpreted causally). Direct labour input controls compete with HRM practices as an explanation of higher MFP when, in fact, adopting HRM practices caused the change in both MFP (frontier presence) and labour input.

Finally, and on a more technical note, while the full set of controls we

[39] As the bottom of table 12 indicates, only 17% of firms are ever in frontier, some of which are always in the frontier.

adopt add to the robustness of the analysis, they also eat into the sample size, particularly the inclusion of the exporting and core equipment questions in the BOS analysis, affecting panel size and, therefore, estimated standard errors.[40] Prior studies using BOS and productivity variables have also tended to estimate on the unweighted sample and are, therefore, focussed on the experience of the observed firms. Since this paper is interested in estimates that reflect the population of (BOS-productivity) firms it has made more sense to use the (recently developed) firm productivity weights together with BOS sampling weights. The consequence of this decision is to provide more weight to small firm observations in the analysis, since these firms are less likely to file the necessary tax returns to be included in the productivity dataset, and are sampled at a lower frequency in the BOS. It is plausible that the impact of some firm activities, eg improved formal management practices, have a weaker effect on small firm performance, which would lead to smaller estimated coefficients in the weighted analysis here, compared to unweighted analysis. Conditional on the weighting scheme being sensible, both approaches are defensible and differences in estimates reflect differences in the sample/population of interest.

4 Conclusions

New Zealand's position as an OECD productivity laggard – coupled with the small, remote nature of our economy – is likely to inhibit learning from the global productivity frontier. Theory and international evidence suggest that firms at the NZ national productivity frontier may be important for the diffusion of knowledge from the global frontier, providing an important conduit for new technologies that will allow average productivity to grow at a more rapid rate than otherwise. Frontier firms have the advantage of being closer to the global production frontier (in technology terms), and may be better equipped to absorb knowledge due to the activities they undertake, and the capabilities that they hold. Frontier firms are also major producers in the New Zealand economy, increasing laggard (non-frontier) firm exposure to the potential of learning from the (local) best.

In this paper, we identify the NZ productivity frontier in a novel way. Population restrictions reduce the issue of MFP measurement error leading

[40] An alternative estimation approach would be to include observations with missing variables and use dummy variable controls to account for this missingness, to improve sample size. The potential downside of this approach is that it adds firms to the regression for whom we are not controlling for the unobserved variable(s).

to misidentification of the frontier. At the same time, we have preserved the aggregate productivity dynamics of the population, while focussing on a relatively high productivity subset of firms with materially better data coverage. We then triangulate four firm-level top MFP measures to compose a frontier that is robust to both production function and industry benchmark choice, mitigating the criticism that the frontier classification is dependent on arbitrary assumptions.

We then show that economic activity is concentrated in the upper deciles of the productivity distribution, which is consistent with a properly functioning economy that allows resources to shift to more productive uses. While frontier firms constitute only 8% of firm-years, they account for 29% of value-added. Even relative to firms just outside the frontier, frontier firms are substantially more important to aggregate output with frontier firms producing 87% more value-added than the average decile nine firm, using the same average labour input. The dispersion in firm performance is even more stark when we compare aggregate labour productivity of frontier firms to the bottom decile of firms, finding that the frontier is nine times as productive as the bottom decile of firms.

The disproportionately large contribution that frontier firms make to the economy is likely supported by their superior engagement in activities that contribute to firm-level productivity growth. Frontier firms employ a more skilled workforce that they draw on by being concentrated in major Urban Areas (particularly Auckland) and – perhaps – through their adoption of superior human resource management practices. They are more export intensive than non-frontier firms, likely leading to greater exposure to the global productivity frontier. Frontier firms are also more likely to have up-to-date technology (including UFB use) and to be in markets where they have no competitors, which may reflect their ability to produce differentiated products that are in demand in the marketplace – a notion supported by their greater international presence.

From a data perspective, these exceptional characteristics of frontier firms support the hypothesis that these firms are productivity frontier firms, rather than a set of firms grouped by MFP measurement error. From an economics perspective, the results suggest it may be right to expect much from this constantly evolving groups of firms as they push out the New Zealand productivity frontier.

References

Andrews, D., C. Criscuolo, and P. N. Gal (2015). Frontier firms, technology diffusion and public policy: Micro evidence from OECD countries. OECD Productivity Working Papers 2, OECD Publishing.

Bartelsman, E. J., J. Haskel, and R. Martin (2008). Distance to which frontier? Evidence on productivity convergence from international firm-level data. CEPR Discussion Papers 7032, Centre for Economic Policy Research.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics 36*(S1), 13–70.

Chappell, N. and A. Jaffe (2018). Intangible investment and firm performance. *Review of Industrial Organization 52*(4), 509–559.

Conway, P., L. Meehan, and G. Zheng (2015). Do New Zealand firms catch up to the domestic productivity frontier? Working Paper 2015/3, New Zealand Productivity Commission.

Fabling, R. (2011). Keeping it together: Tracking firms in New Zealand's Longitudinal Business Database. Working Paper 11-01, Motu Economic and Public Policy Research.

Fabling, R. and A. Grimes (2014). The "suite" smell of success: Personnel practices and firm performance. *Industrial and Labor Relations Review 67*(4), 1095–1126.

Fabling, R. and A. Grimes (2016). Picking up speed: Does ultrafast broadband increase firm productivity? Working Papers 16-22, Motu Economic and Public Policy Research.

Fabling, R. and D. C. Maré (2015a). Addressing the absence of hours information in linked employer-employee data. Working Paper 15-17, Motu Economic and Public Policy Research.

Fabling, R. and D. C. Maré (2015b). Production function estimation using New Zealand's Longitudinal Business Database. Working Paper 15-15, Motu Economic and Public Policy Research.

Fabling, R. and D. C. Maré (2019). Improved productivity measurement in New Zealand's Longitudinal Business Database. Working Paper 19-03, Motu Economic and Public Policy Research.

Fabling, R. and L. Sanderson (2013). Exporting and firm performance: Market entry, investment and expansion. *Journal of International Economics 89*(2), 422–431.

Fabling, R. and L. Sanderson (2014a). Foreign acquisition and the performance of New Zealand firms. *New Zealand Economic Papers 48*(1), 1–20.

Fabling, R. and L. Sanderson (2014b). Productivity distributions in New Zealand: The dangers of international comparisons. Working Paper 14-16, Motu Economic and Public Policy Research.

Fabling, R. and L. Sanderson (2016). A rough guide to New Zealand's Longitudinal Business Database (2nd edition). Working Paper 16-03, Motu Economic and Public Policy Research.

Griffith, R., S. Redding, and H. Simpson (2009). Technological catch-up and geographic proximity. *Journal of Regional Science 49*(4), 689–720.

Maré, D. C. (2016). Urban productivity estimation with heterogeneous prices and labour. Working Paper 16-21, Motu Economic and Public Policy Research.

Maré, D. C. and R. Fabling (2019). Competition and productivity: Do commonly used metrics suggest a relationship? Working Paper 19-16, Motu Economic and Public Policy Research.

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

New Zealand Productivity Commission (2020). New Zealand firms: Reaching for the frontier. Issues paper, New Zealand Productivity Commission.

Syverson, C. (2011). What determines productivity? *Journal of Economic Literature 49*(2), 326–365.

Zheng, G. (2016). Geographic proximity and productivity convergence across New Zealand firms. Working Paper 2016/04, New Zealand Productivity Commission.

Tables

Table 1: Effect of productivity population restrictions on coverage

	Proportion of population total included				
	N(obs)	Y	M	K	L
Unrestricted	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Observed MFP ($Y, K, M > 0$)	0.982	0.999	0.999	0.997	0.984
+ non-entry/exit year	0.855	0.977	0.977	0.975	0.955
+ has employees	0.419	0.919	0.926	0.878	0.836
+ FTE employment>1	0.269	0.872	0.879	0.806	0.787
+ $\boldsymbol{L > 2}$	**0.239**	**0.860**	**0.865**	**0.786**	**0.777**
	Productivity dataset coverage rate				
	N(obs)	Y	M	K	L
Unrestricted	**0.668**	**0.820**	**0.823**	**0.816**	**0.793**
Observed MFP ($Y, K, M > 0$)	0.670	0.820	0.823	0.817	0.793
+ non-entry/exit year	0.702	0.824	0.827	0.821	0.799
+ has employees	0.746	0.834	0.836	0.838	0.819
+ FTE employment>1	0.765	0.841	0.843	0.849	0.825
+ $\boldsymbol{L > 2}$	**0.774**	**0.844**	**0.845**	**0.853**	**0.827**

The top panel uses productivity population weights to estimate the proportion of the population total (pooled 2005-2018) that is included in the restricted population. The data coverage rate (bottom panel) is the ratio of the unweighted total to the weighted total. The data coverage rate increases as we add population restrictions because small and WP-only firms have relatively lower coverage rates than large firms in the productivity dataset. The bottom row of each panel (bolded) reflects the final analysis sample. Y and M are deflated using the relevant official Stats NZ producer price index, while K is deflated using an industry-specific weighting of official asset-type deflators as described in Fabling and Maré (2019).

Table 2: Sectoral composition of productivity subpopulation

	N(firms)				L			
	Construction	Manufacturing	Primary	Services	Construction	Manufacturing	Primary	Services
2005	9,927	9,552	10,311	41,445	99,500	230,100	65,400	603,100
2006	10,656	9,567	10,026	42,333	111,300	226,200	64,000	626,900
2007	11,256	9,489	9,885	42,906	114,900	229,900	64,300	634,700
2008	11,721	9,369	9,789	43,302	123,000	225,300	64,900	656,600
2009	11,259	9,141	9,891	42,603	121,200	216,800	66,100	647,200
2010	10,140	8,709	9,771	41,652	112,000	206,000	65,500	626,300
2011	9,921	8,571	10,038	41,772	109,300	204,500	67,800	632,200
2012	9,789	8,379	10,215	42,015	111,500	199,500	69,200	641,200
2013	10,179	8,268	10,221	42,141	118,000	201,800	70,800	648,300
2014	10,980	8,190	10,677	42,918	126,200	203,300	74,500	665,500
2015	11,589	8,235	10,629	43,545	134,600	205,700	74,400	684,500
2016	12,129	8,211	10,407	44,400	141,200	206,400	73,600	709,900
2017	12,978	8,319	10,317	45,711	150,700	209,700	73,600	736,600
2018	13,821	8,340	10,326	46,620	162,600	212,500	76,500	761,800
Total	**156,345**	**122,340**	**142,503**	**603,363**	**1,736,000**	**2,977,700**	**970,600**	**9,274,800**
Share in 2005	0.139	0.134	0.145	0.582	0.100	0.231	0.066	0.604
Share in 2018	0.175	0.105	0.131	0.589	0.134	0.175	0.063	0.628
Growth rate	0.026	-0.010	0.000	0.009	0.039	-0.006	0.012	0.018

Utilities (electricity, gas, water and waste) are included in the construction sector. Mining is included in the primary sector, along with agriculture, forestry and fishing. The productivity population is restricted to industries Stats NZ's "measured sector" with the exclusions – property operators, health care, education and public administration – reducing the apparent size of the services sector. The "Growth rate" is the annual growth rate required to achieve the observed aggregate growth over the entire period (ie, $(x_{2018}/x_{2005})^{1/13} - 1$.

Table 3: Summary statistics for productivity subpopulation

	Mean	Standard deviation	Percentiles 25th	50th	75th
Gross output (y)	13.767	1.208	12.937	13.587	14.384
Intermediate consumption (m)	12.793	1.445	11.847	12.662	13.585
Labour (l)	1.824	0.902	1.193	1.599	2.212
Capital services (k)	11.431	1.387	10.568	11.319	12.201
Capital-labour ratio ($k-l$)	9.607	1.113	8.984	9.550	10.175
MFP (Cobb-Douglas)	0.014	0.362	-0.157	-0.010	0.157
MFP (translog)	0.011	0.326	-0.141	-0.004	0.150
Average worker skill	0.000	1.000	-0.704	-0.024	0.625
Firm wage fixed effect (FFE)	0.000	1.000	-0.645	-0.069	0.596
ln(firm age)	2.504	0.736	1.946	2.565	3.045
Urban Area employment share					
Auckland	0.272	0.439	0.000	0.000	1.000
Hamilton	0.044	0.203	0.000	0.000	0.000
Tauranga	0.029	0.166	0.000	0.000	0.000
Napier	0.029	0.166	0.000	0.000	0.000
Wellington	0.067	0.246	0.000	0.000	0.000
Christchurch	0.092	0.284	0.000	0.000	0.000
Dunedin	0.021	0.140	0.000	0.000	0.000
Other main	0.114	0.313	0.000	0.000	0.000
Secondary	0.066	0.244	0.000	0.000	0.000
Minor	0.087	0.278	0.000	0.000	0.000
Rural	0.180	0.381	0.000	0.000	0.000

	N(observations) Unweighted	Weighted	Coverage rate
Gross output – ln(firm age)	792,633	1,024,542	0.774
Urban Area employment share	792,492	1,024,356	0.774

Lower case indicates production function variable is logged. Multifactor productivity (MFP) estimated separately for each "production function" industry from the unweighted full productivity dataset using OLS with firm fixed effects, and assuming either a Cobb-Douglas (CD) or translog (TL) production function (Fabling and Maré 2015b, 2019). In either case, MFP is the sum of the industry-year effect, firm fixed effect and residual. Average worker skill and firm wage fixed effect (FFE) are estimated from a two-way wage fixed effect model as in Maré et al. (2017). Average worker skill is FTE-weighted at each firm, and includes the sum of observables (sex-age) and worker fixed effects components of the worker wage. Given the arbitrary scaling of average worker skill and FFE, we have renormalised them to be mean zero, standard deviation one at the firm level. Hamilton Urban Area (UA) includes Cambridge & Te Awamutu; Napier UA includes Hastings; and Wellington UA includes Porirua, Lower Hutt & Upper Hutt. A small number of firms have no plant location information.

Table 4: Summary statistics for BOS-productivity subpopulation

	Mean	Standard deviation	25th	Percentiles 50th	75th
Gross output (y)	14.519	1.221	13.688	14.329	15.144
Intermediate consumption (m)	13.531	1.479	12.572	13.376	14.343
Labour (l)	2.514	0.906	1.930	2.300	2.870
Capital services (k)	12.073	1.373	11.255	11.920	12.796
Capital-labour ratio ($k - l$)	9.559	1.061	9.013	9.546	10.105
MFP (Cobb-Douglas)	0.069	0.365	-0.103	0.029	0.189
MFP (translog)	0.048	0.316	-0.103	0.025	0.175
Average worker skill	0.088	0.961	-0.621	0.068	0.694
Firm wage fixed effect (FFE)	0.240	0.953	-0.441	0.165	0.846
ln(firm age)	2.681	0.730	2.197	2.773	3.178
Urban Area employment share					
Auckland	0.301	0.447	0.000	0.000	1.000
Hamilton	0.048	0.209	0.000	0.000	0.000
Tauranga	0.030	0.166	0.000	0.000	0.000
Napier	0.033	0.175	0.000	0.000	0.000
Wellington	0.070	0.245	0.000	0.000	0.000
Christchurch	0.102	0.293	0.000	0.000	0.000
Dunedin	0.025	0.153	0.000	0.000	0.000
Other main	0.120	0.318	0.000	0.000	0.000
Secondary	0.068	0.245	0.000	0.000	0.000
Minor	0.080	0.265	0.000	0.000	0.000
Rural	0.121	0.322	0.000	0.000	0.000

	N(observations) Unweighted	Weighted	Coverage rate
Gross output – ln(firm age)	60,936	461,625	0.132
Urban Area employment share	60,924	461,568	0.132

All statistics weighted using BOS survey weights adjusted to include panel and Māori business top-ups in the weighted sample, and then further adjusted for the partial coverage of productivity data. The BOS reweighting to include panel observations occurs prior to applying the productivity population restriction (second weight adjustment) so that the final summed weights approximately reflect the restricted BOS-productivity subpopulation size. Summary statistics differ from official Stats NZ BOS statistics because of this dual reweighting; because we exclude non-response and "don't know" responses; and because of productivity population criteria (ie, in a productivity industry and always private-for-profit). See table 3 for further notes.

Table 5: Proportion of firm-years by MFP decile – Cobb-Douglas vs translog

CD decile	TL decile 1	2	3	4	5	6	7	8	9	10
1	**0.085**	0.014	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.011	**0.060**	0.023	0.004	0.001	0.000	0.000	0.000	0.000	0.000
3	0.002	0.017	**0.046**	0.026	0.006	0.002	0.001	0.000	0.000	0.000
4	0.001	0.005	0.019	**0.039**	0.027	0.007	0.002	0.001	0.000	0.000
5	0.001	0.002	0.006	0.019	**0.036**	0.027	0.007	0.002	0.001	0.000
6	0.000	0.001	0.003	0.007	0.019	**0.035**	0.027	0.006	0.001	0.000
7	0.000	0.001	0.001	0.003	0.008	0.019	**0.037**	0.026	0.004	0.001
8	0.000	0.000	0.001	0.001	0.003	0.007	0.020	**0.042**	0.024	0.002
9	0.000	0.000	0.000	0.000	0.001	0.002	0.005	0.020	**0.054**	0.016
10	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.015	**0.080**

Decile one includes the lowest MFP firms, whereas decile ten includes the highest MFP firms. By definition, each table row or column adds to 0.1 (10%). Cells where the decile is the same across the two measures are highlighted in bold.

Table 6: Proportion of firm-years by Cobb-Douglas MFP decile – production function industry vs three-digit industry

CD decile	CD_{3d} decile 1	2	3	4	5	6	7	8	9	10
1	**0.091**	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.009	**0.076**	0.013	0.001	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.014	**0.066**	0.017	0.002	0.001	0.000	0.000	0.000	0.000
4	0.000	0.002	0.017	**0.058**	0.021	0.002	0.001	0.000	0.000	0.000
5	0.000	0.000	0.003	0.017	**0.053**	0.023	0.003	0.000	0.000	0.000
6	0.000	0.000	0.001	0.004	0.017	**0.051**	0.023	0.003	0.000	0.000
7	0.000	0.000	0.000	0.001	0.005	0.017	**0.051**	0.023	0.003	0.000
8	0.000	0.000	0.000	0.000	0.001	0.005	0.017	**0.052**	0.023	0.001
9	0.000	0.000	0.000	0.000	0.000	0.001	0.005	0.018	**0.059**	0.017
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.015	**0.081**

CD_{3d} decile is not calculated for any three-digit industry with, on average, less than 50 firms in the population (ie, where less than five firms would be included in the top decile in any given year). This exclusion reduces the productivity subpopulation from 1,024,542 to 1,007,244 firm-year observations. Excluded industry observations are overrepresented in CD MFP deciles one (12.5% of excluded firm-year observations), nine (11.4% of obs) and ten (13.4% of obs). Table proportions are calculated excluding those observations where CD_{3d} decile is missing. See table 5 for additional notes.

Table 7: Productivity decile distribution conditional on being in a top decile

	Probability of being in MFP decile of A | in top decile B							
Decile of A	A=CD B=TL	A=TL B=CD	A=CD B=CD$_{3d}$	A=CD$_{3d}$ B=CD	A=TL B=TL$_{3d}$	A=TL$_{3d}$ B=TL	A=CD$_{3d}$ B=TL$_{3d}$	B=TL$_{3d}$ A=CD$_{3d}$
1	0.001	0.004	0.000	0.000	0.000	0.000	0.001	0.004
2	0.002	0.002	0.000	0.000	0.000	0.000	0.002	0.002
3	0.001	0.002	0.000	0.000	0.000	0.000	0.001	0.002
4	0.002	0.002	0.000	0.000	0.000	0.000	0.001	0.002
5	0.002	0.003	0.000	0.000	0.000	0.000	0.003	0.003
6	0.003	0.004	0.000	0.002	0.000	0.001	0.004	0.005
7	0.006	0.009	0.001	0.003	0.001	0.002	0.007	0.010
8	0.017	0.026	0.010	0.024	0.010	0.017	0.018	0.028
9	0.165	0.146	0.170	0.152	0.154	0.142	0.167	0.146
10	**0.802**	**0.802**	**0.820**	**0.819**	**0.836**	**0.837**	**0.797**	**0.797**

Each column reflects a different pairing of two productivity decile metrics, where the decile of the metric A is reported in the leftmost column of the table, and the productivity decile of metric B is ten. The final row of the table (in bold), is the proportion of observations where metric A agrees with metric B that this is a top MFP decile firm in a particular year. As in table 6, comparisons that include three-digit industry metrics have proportions calculated excluding missing (small industry) observations. For brevity, four permutations are not reported – those pairing CD with TL$_{3d}$, and CD$_{3d}$ with TL.

Table 8: Productivity decile distribution conditional on being in prior top decile

	Probability of being in MFP decile of A | in top decile B at time							
	t		t & $t-1$		$t-1$			
Decile of A	A=CD B=TL	A=TL B=CD	A=CD B=TL	A=TL B=CD	A=CD B=TL	A=TL B=CD	A=CD B=CD	A=TL B=TL
1	0.001	0.004	0.002	0.005	0.013	0.017	0.012	0.011
2	0.002	0.002	0.003	0.002	0.012	0.013	0.010	0.009
3	0.002	0.002	0.002	0.001	0.014	0.014	0.011	0.011
4	0.002	0.002	0.003	0.002	0.017	0.018	0.013	0.013
5	0.003	0.003	0.003	0.003	0.021	0.021	0.016	0.016
6	0.004	0.005	0.005	0.003	0.029	0.029	0.022	0.022
7	0.007	0.009	0.008	0.007	0.045	0.043	0.034	0.033
8	0.019	0.027	0.018	0.022	0.082	0.082	0.062	0.064
9	0.164	0.155	0.127	0.128	0.204	0.201	0.188	0.188
10	**0.797**	**0.792**	**0.829**	**0.827**	**0.564**	**0.563**	**0.632**	**0.633**

Each column reflects a different pairing of two productivity decile metrics, where the decile of the metric A is reported in the leftmost column of the table, and the productivity decile of metric B is ten in the contemporaneous (t) and/or prior ($t-1$) year. The first two columns of the table differ from the first two columns of table 7 because the former excludes firms not observed at $t-1$ for consistency with the remainder of the table. The final row of the table (in bold), is the proportion of observations where metric A agrees with metric B that this is a top MFP decile firm in a particular year or pair of years.

Table 9: Definition of frontier firms using number of top decile appearances

Top decile Cobb-Douglas	Top decile translog Neither TL	TL only	TL_{3d} only	Both TL	**Total**
Neither CD	0.861	0.005	0.007	0.010	**0.883**
CD only	0.006	0.008	0.000	*0.004*	**0.018**
CD_{3d} only	0.006	0.000	0.007	*0.005*	**0.018**
Both CD	0.011	*0.003*	*0.003*	*0.066*	**0.082**
Total	**0.884**	**0.016**	**0.016**	**0.084**	**1.000**

Cells in italics satisfy the criterion for the frontier, constituting a total of 7.9% of firm-year observations. Analysis includes industries with insufficient firms in the detailed (3-digit) industry to justify identifying the CD_{3d} and TL_{3d} MFP deciles, placing these observations in relevant neither/both cells for consistency with the definition of the frontier. The table is almost identical if these industries are excluded, since small observation industries account for only 1.7% of total observations.

Table 10: Composite productivity decile by year

	Median decile (rounded down)									frontier
t	1	2	3	4	5	6	7	8	9	10
2005	0.108	0.103	0.102	0.106	0.104	0.104	0.102	0.101	0.091	0.079
2006	0.108	0.102	0.104	0.104	0.103	0.105	0.101	0.101	0.092	0.080
2007	0.109	0.103	0.103	0.104	0.104	0.105	0.103	0.099	0.093	0.078
2008	0.108	0.103	0.103	0.103	0.106	0.102	0.105	0.099	0.092	0.079
2009	0.108	0.103	0.103	0.104	0.105	0.104	0.101	0.101	0.092	0.080
2010	0.107	0.103	0.104	0.104	0.105	0.103	0.103	0.100	0.091	0.079
2011	0.107	0.104	0.103	0.104	0.105	0.105	0.101	0.101	0.091	0.080
2012	0.107	0.104	0.103	0.105	0.105	0.103	0.102	0.100	0.092	0.079
2013	0.108	0.103	0.104	0.104	0.106	0.102	0.104	0.098	0.092	0.079
2014	0.107	0.103	0.105	0.104	0.105	0.104	0.101	0.099	0.092	0.079
2015	0.108	0.103	0.102	0.105	0.104	0.105	0.103	0.099	0.091	0.080
2016	0.108	0.104	0.103	0.103	0.104	0.104	0.103	0.100	0.092	0.080
2017	0.108	0.102	0.104	0.104	0.105	0.105	0.102	0.100	0.091	0.080
2018	0.109	0.102	0.104	0.102	0.105	0.106	0.102	0.099	0.092	0.079
Total	**0.108**	**0.103**	**0.103**	**0.104**	**0.105**	**0.104**	**0.102**	**0.100**	**0.092**	**0.079**

The "composite productivity decile" is the median of the four MFP deciles (CD, TL, CD_{3d}, TL_{3d}) rounded down. Rounding down has the effect of making composite decile ten match our definition of the productivity frontier (ie, three out of four tenth deciles, or both CD and TL tenth deciles for small observation industries).

Table 11: Probability of transition between productivity deciles over consecutive years

CD decile at $t-1$	CD decile at t: 1	2	3	4	5	6	7	8	9	10	Leaver	Top decile at t | non-leaver
1	**0.340**	0.138	0.068	0.041	0.029	0.022	0.017	0.013	0.010	0.010	0.314	0.014
2	0.149	**0.236**	0.151	0.090	0.056	0.036	0.024	0.015	0.010	0.008	0.225	0.010
3	0.072	0.159	**0.199**	0.145	0.093	0.057	0.035	0.023	0.014	0.009	0.195	0.011
4	0.045	0.091	0.150	**0.180**	0.145	0.092	0.056	0.031	0.017	0.010	0.183	0.012
5	0.031	0.055	0.094	0.146	**0.176**	0.145	0.089	0.050	0.026	0.013	0.175	0.016
6	0.022	0.036	0.058	0.094	0.143	**0.179**	0.151	0.088	0.042	0.017	0.169	0.021
7	0.017	0.023	0.037	0.057	0.090	0.152	**0.200**	0.157	0.075	0.027	0.165	0.032
8	0.013	0.016	0.023	0.033	0.051	0.087	0.157	**0.239**	0.165	0.049	0.169	0.059
9	0.010	0.011	0.013	0.017	0.026	0.042	0.077	0.167	**0.315**	0.145	0.175	0.176
10	0.009	0.008	0.009	0.011	0.013	0.018	0.027	0.049	0.150	**0.503**	0.204	0.632
Joiner	0.146	0.114	0.100	0.095	0.090	0.087	0.087	0.086	0.090	0.105	–	0.105

Top decile at $t-1$	Decile of related measure at t: 1	2	3	4	5	6	7	8	9	10	Leaver	Top decile at t | non-leaver
CD	0.009	0.008	0.009	0.011	0.013	0.018	0.027	0.049	0.150	0.503	0.204	0.632
TL	0.009	0.007	0.009	0.010	0.013	0.017	0.026	0.050	0.149	0.502	0.206	0.633
CD_{3d}	0.010	0.009	0.009	0.012	0.014	0.020	0.029	0.055	0.156	0.486	0.200	0.608
TL_{3d}	0.009	0.008	0.009	0.011	0.015	0.019	0.029	0.055	0.155	0.486	0.205	0.611
Composite	0.010	0.008	0.009	0.011	0.015	0.020	0.030	0.057	0.159	0.472	0.209	0.597

For a given row, the top panel shows the probability of being in each CD productivity decile in the current year (t), conditional on being in a given CD productivity decile in the previous year ($t-1$). Because we have excluded entrants and exiters, firms that are active in both years but not observed in one of the years are outside the population in that year or have missing productivity data. These firms are dubbed "joiners" and "leavers" to distinguish them from entrants/exiters and to highlight the fact that they join or leave our (observed MFP) subpopulation. The rightmost column of the table reports the probability of being in the top decile at t conditional on not being a leaver. The main diagonal of the top panel, where firms stay in the same decile between years, is highlighted in bold. The bottom panel of the table reports the top ($t-1$) decile row of this table for the other metrics. The composite decile is the median of the four productivity deciles (CD, TL, CD_{3d}, TL_{3d}) rounded down, as in table 10, so that the top decile matches our definition of the productivity frontier.

Table 12: Proportion of frontier observations by number of years observed

N(years observed)	Number of observations	Number of firms	Proportion of observations in frontier	Proportion of firms ever in frontier	Proportion of firms mainly in frontier
1	42,372	42,372	0.090	0.090	0.090
2	58,824	29,412	0.085	0.125	0.125
3	64,239	21,413	0.083	0.149	0.068
4	65,622	16,406	0.081	0.169	0.085
5	64,263	12,853	0.077	0.181	0.058
6	63,837	10,640	0.078	0.195	0.069
7	63,849	9,121	0.081	0.221	0.058
8	63,804	7,976	0.082	0.237	0.066
9	63,636	7,071	0.078	0.241	0.053
10	67,956	6,796	0.075	0.245	0.057
11	68,334	6,212	0.077	0.259	0.052
12	77,985	6,499	0.076	0.271	0.056
13	100,257	7,712	0.074	0.276	0.049
14	159,570	11,398	0.081	0.288	0.060
Total	**1,024,542**	**195,879**	**0.079**	**0.173**	**0.078**

Firms are “mainly” in the frontier if they are frontier firms for at least half of the years that they are observed.

Table 13: Input, output and labour productivity growth from 2005-07 to 2016-18 by composite decile

	Annual growth rate (percentage) by composite decile										
	1	2	3	4	5	6	7	8	9	10	**Total**
N(firms)	0.59%	0.60%	0.66%	0.43%	0.64%	0.57%	0.58%	0.54%	0.57%	0.62%	**0.58%**
Labour (L)	3.30%	1.66%	1.32%	1.02%	1.94%	1.01%	1.21%	0.35%	0.97%	1.13%	**1.24%**
Gross output (Y)	3.88%	2.75%	2.11%	1.76%	3.47%	1.11%	1.25%	1.74%	-0.65%	2.06%	**1.41%**
Intermediate consumption (M)	3.07%	2.96%	1.64%	1.29%	3.33%	0.15%	0.47%	1.64%	-2.02%	2.60%	**0.88%**
Value-added ($Y - M$)	6.24%	2.45%	2.69%	2.32%	3.63%	2.18%	2.18%	1.87%	1.76%	1.54%	**2.08%**
Capital services (K)	4.18%	3.79%	3.51%	2.20%	4.34%	1.90%	1.37%	4.33%	1.87%	2.10%	**2.65%**
Labour productivity	2.84%	0.78%	1.35%	1.29%	1.66%	1.15%	0.95%	1.52%	0.78%	0.40%	**0.83%**

	Contribution of composite decile to aggregate change (percentage)										
	1	2	3	4	5	6	7	8	9	10	**Total**
N(firms)	11.0%	10.6%	11.7%	7.7%	11.6%	10.3%	10.3%	9.4%	9.0%	8.5%	**100%**
Labour (L)	15.8%	7.3%	7.1%	6.3%	15.3%	9.7%	11.5%	3.8%	11.9%	11.4%	**100%**
Gross output (Y)	6.7%	5.9%	5.7%	6.0%	17.7%	7.2%	8.4%	16.1%	-10.3%	36.5%	**100%**
Intermediate consumption (M)	11.2%	10.6%	7.0%	6.8%	26.4%	1.5%	4.9%	24.6%	-57.4%	64.6%	**100%**
Value-added ($Y - M$)	4.3%	3.3%	5.0%	5.6%	12.9%	10.3%	10.4%	11.5%	15.4%	21.2%	**100%**
Capital services (K)	7.8%	6.0%	6.4%	5.0%	13.4%	7.5%	4.7%	18.6%	14.3%	16.3%	**100%**

Initial (2005-2007) and final (2016-2018) periods pooled to mitigate the risk of endpoints overly affecting estimates. Compound annual growth rate in top panel calculated as if the analysis was from middle year to middle year (ie, the 11 years from 2006 to 2017). Bottom panel excludes labour productivity, because this cannot be simply aggregated across subgroups as can the other metrics.

Table 14: Impact of alternative frontier growth on aggregate labour productivity growth

	(1)	(2)	(3)	(4)	(5)	(6)
	Observed		Counterfactuals: No growth in frontier			
	labour	LP	Labour (L)		L & LP	
Labour productivity (LP)	productivity		$L\Rightarrow$decile 9	$L\Rightarrow$deciles 1-9	$L\Rightarrow$decile 9	$L\Rightarrow$deciles 1-9
Average (2005-2007 pooled)	\$122,871	\$122,871	\$122,871	\$122,871	\$122,871	\$122,871
Average (2016-2018 pooled)	\$134,566	\$132,969	\$132,536	\$131,800	\$131,125	\$130,389
Change (2016/18 - 2005/07)	**\$11,695**	**\$10,098**	**\$9,665**	**\$8,929**	**\$8,254**	**\$7,518**
Annual growth rate	0.83%	0.72%	0.69%	0.64%	0.59%	0.54%
Change in growth rate (actual to counterfactual)		-0.11%	-0.14%	-0.19%	-0.24%	-0.29%

Initial (2005-2007) and final (2016-2018) periods pooled to mitigate the risk of endpoints overly affecting estimates. Compound annual growth rate in top panel calculated as if the analysis was from middle year to middle year (ie, the 11 years from 2006 to 2017). Counterfactual where frontier firm L does not increase assumes that aggregate 2016/18 L remains unchanged and allocates the implied additional non-frontier employment to either composite decile nine, or across all non-frontier deciles (ie, has the same LP as the non-frontier aggregate). In all counterfactuals, the labour productivity of non-frontier deciles is assumed to be unchanged by changes to frontier LP and employment growth.

Table 15: Mean characteristics of frontier firms

	Frontier	Non-frontier	Overall
Labour (l)	1.907***	1.817	1.824
Capital-labour ratio ($k-l$)	9.127***	9.648	9.607
Average worker skill	0.284***	-0.025	0.000
Firm wage fixed effect (FFE)	0.376***	-0.032	0.000
ln(firm age)	2.469***	2.507	2.504
Urban Area employment share			
Auckland	0.324***	0.267	0.272
Hamilton	0.046**	0.044	0.044
Tauranga	0.031**	0.029	0.029
Napier	0.027***	0.029	0.029
Wellington	0.080***	0.066	0.067
Christchurch	0.101***	0.091	0.092
Dunedin	0.018***	0.021	0.021
Other main	0.100***	0.115	0.114
Secondary	0.055***	0.066	0.066
Minor	0.063***	0.089	0.087
Rural	0.155***	0.182	0.180
Presence in Urban Area			
Auckland	0.344***	0.281	0.286
Hamilton	0.059***	0.052	0.053
Tauranga	0.041***	0.035	0.036
Napier	0.035	0.034	0.034
Wellington	0.102***	0.080	0.081
Christchurch	0.126***	0.107	0.108
Dunedin	0.027	0.027	0.027
Other main	0.115***	0.127	0.126
Secondary	0.065***	0.075	0.074
Minor	0.072***	0.098	0.096
Rural	0.164***	0.190	0.188
Multi-UA group firm	0.064***	0.054	0.054

Stars (***;**;*) indicate frontier firm mean significantly different from non-frontier firm mean (at the 1%;5%;10% level respectively). Presence in Urban Area is an indicator variable set equal to one if the firm has employment in a specific UA, and zero otherwise. Multi-UA group firm is one for firms with multiple UA presences (zero otherwise), which in not equivalent to firms with a presence in multiple UAs because of the grouping of some UAs (ie, other main secondary and minor groupings).

Table 16: Mean BOS characteristics of frontier firms

	N(observations)		Mean		
	Unweighted	Weighted	Frontier	Non-frontier	Overall
Has exports	54,675	410,967	0.175	0.163	0.164
Export share of sales	54,675	410,967	0.084***	0.064	0.066
Entered new market	59,583	450,162	0.035	0.040	0.040
Has foreign direct investment (FDI)	59,940	454,761	0.137***	0.064	0.070
FDI share	59,748	454,209	0.109***	0.049	0.054
Has outward direct investment (ODI)	59,955	455,136	0.048***	0.033	0.035
Collective employment agreements: No employees	57,456	429,183	0.765*	0.744	0.745
1-10% of employees	57,456	429,183	0.023	0.021	0.021
11-50% of employees	57,456	429,183	0.029	0.027	0.027
51-90% of employees	57,456	429,183	0.030	0.032	0.032
91-100% of employees	57,456	429,183	0.153**	0.177	0.175
Core equipment: fully up-to-date	52,230	384,849	0.623***	0.549	0.555
up to 4yrs behind	52,230	384,849	0.298**	0.328	0.326
5-10yrs behind	52,230	384,849	0.064***	0.096	0.094
more than 10yrs behind	52,230	384,849	0.016***	0.027	0.026
Competition: no competition	57,696	430,839	0.063***	0.036	0.038
1-2 competitors	57,696	430,839	0.188	0.187	0.187
many competitors, several dominant	57,696	430,839	0.558	0.574	0.573
many competitors, none dominant	57,696	430,839	0.191	0.204	0.203
HRM practices (z-score)	16,659	118,683	0.212***	-0.019	0.000
Has ultrafast broadband (UFB)	19,683	154,449	0.339***	0.284	0.289
Complementary ICT investments (z-score)	19,683	154,449	-0.014	0.001	0.000

Table continued on next page...

...table continued from previous page.

	N(observations)		Mean		
	Unweighted	Weighted	Frontier	Non-frontier	Overall
Has R&D	59,841	454,254	0.085	0.079	0.079
R&D (share of M)	59,715	453,867	0.013***	0.007	0.008
Has expenditure on product design	24,954	188,979	0.079**	0.102	0.100
Has expenditure on product marketing	24,954	188,979	0.143	0.143	0.143
Has other expenditure on product development	24,954	188,979	0.088	0.100	0.099
Has any expenditure on product development	24,963	189,015	0.238	0.254	0.253
Product design (share of M)	24,954	188,979	0.009**	0.004	0.004
Product marketing & market research (share of M)	24,954	188,979	0.013***	0.005	0.006
Other product development (share of M)	24,954	188,979	0.007	0.005	0.005
Total product development incl. R&D (share of M)	24,963	189,015	0.039***	0.021	0.023
Successful innovation in the last two years					
New product	30,687	225,192	0.170*	0.192	0.190
New operational process	30,687	225,192	0.141***	0.181	0.178
New organisational process	30,687	225,192	0.197*	0.222	0.219
New marketing method	30,687	225,192	0.171***	0.227	0.222
Any of the above	30,687	225,192	0.366***	0.427	0.422
Any successful innovation in the last year	58,647	440,631	0.345***	0.395	0.391

Stars (***;**;*) indicate frontier firm mean significantly different from non-frontier firm mean (at the 1%;5%;10% level respectively). Observation counts primarily differ across characteristics because some questions are not surveyed in every year. Minor differences in counts between questions that are asked in the same number of years are due to dropping imputed and "don't know" responses on a question-by-question basis. Minor data cleaning has been performed to achieve consistency between questions that are related through questionnaire routing (eg, Has FDI and the FDI intensity question). Consistent with the remainder of the paper, reported proportions are based on weighted observations. Mean product development component shares do not add to the total product development share because each component (and total) is independently capped at a value of one to avoid outliers from unduly affecting the mean and/or estimated regression coefficients. A small proportion of firms, therefore, have summed components that exceed one, but a total capped at one.

Table 17: Mean proportion of firms with BOS characteristics that are frontier

	N(observations) Total	Avg/year	Proportion frontier
Has exports	67,305	4,808	0.092
Entered new market	17,856	1,275	0.076
Has foreign direct investment (FDI)	32,013	2,287	0.167
Has outward direct investment (ODI)	15,717	1,123	0.120
Collective employment agreements: No employees	319,926	22,852	0.087
1-10% of employees	8,910	636	0.095
11-50% of employees	11,607	829	0.091
51-90% of employees	13,557	968	0.080
91-100% of employees	75,174	5,370	0.074
Core equipment: fully up-to-date	213,555	15,254	0.096
up to 4yrs behind	125,343	8,953	0.078
5-10yrs behind	36,024	2,573	0.058
more than 10yrs behind	9,927	709	0.052
Competition: no competition	16,305	1,165	0.141
1-2 competitors	80,427	5,745	0.086
many competitors, several dominant	246,825	17,630	0.083
many competitors, none dominant	87,279	6,234	0.080
Has ultrafast broadband (UFB)	44,592	8,918	0.102
Has R&D	35,967	2,569	0.093
Has expenditure on product design	18,900	3,150	0.067
Has expenditure on product marketing	26,988	4,498	0.085
Has other expenditure on product development	18,624	3,104	0.076
Has any expenditure on product development	47,748	7,958	0.080
Successful innovation in the last two years			
New product	42,738	6,105	0.076
New operational process	39,984	5,712	0.067
New organisational process	49,425	7,061	0.076
New marketing method	50,043	7,149	0.065
Any of the above	95,088	13,584	0.073
Any successful innovation in the last year	172,119	12,294	0.076
Total	**461,625**	**32,973**	**0.086**

Observation counts of firms with different activities differ, in part, because some questions are not surveyed in every BOS year. The second column in the table reports the average observations per year that the question appears in the survey. Questions that are asked annually are available for 14 years (2005-2018). The remaining questions are available in seven years (two-year innovation questions), six years (product development costs, excluding R&D expenditure which is available in all 14 years), or five years (UFB usage). The final row of the table reports the proportion of frontier firms in the full BOS-productivity subpopulation, ignoring data availability for any particular question (consistent with figure 3).

Table 18: Mean BOS characteristics of frontier firms – conditional on activity

	Mean		
	Frontier	Non-frontier	Overall
Conditional on exporting			
Export share of sales	0.480***	0.392	0.400
Entered new market	0.173***	0.216	0.212
Conditional of foreign direct investment			
FDI share	0.818	0.802	0.805
Conditional on R&D (share of M)			
R&D	0.157***	0.094	0.100
Conditional on product development (share of M)			
R&D	0.056***	0.031	0.033
Product design	0.039**	0.016	0.017
Product marketing & market research	0.056***	0.021	0.024
Other product development	0.027	0.019	0.019
Total product development incl. R&D	0.163***	0.083	0.089

Stars (***;**;*) indicate frontier firm mean significantly different from non-frontier firm mean (at the 1%;5%;10% level respectively). The two conditional R&D intensities are substantially different because one is conditional on R&D activity and the other is conditional on any product development activity (including R&D). In the latter case, firms may have product development costs (eg, market research) but no R&D expenditure, which lowers the observed conditional R&D intensity. Mean product development component shares do not add to total product development share because each component (and total) is independently capped at a value of one to avoid outliers from unduly affecting the mean and/or estimated regression coefficients. A small proportion of firms, therefore, have summed components that exceed one, but a total capped at one.

Table 19: Estimated relationship between frontier productivity and full coverage firm characteristics

Dependent variable: δ(frontier)	(1) OLS	(2) OLS	(3) FE	(4) FE
Labour (l)	0.009***	-0.011***	-0.013***	-0.033***
	[0.001]	[0.001]	[0.001]	[0.002]
Capital-labour ratio ($k-l$)	-0.045***	-0.058***	-0.049***	-0.057***
	[0.001]	[0.001]	[0.001]	[0.001]
Average worker skill	0.029***	0.033***	0.016***	0.017***
	[0.001]	[0.001]	[0.001]	[0.001]
Firm wage fixed effect (FFE)	0.034***	0.041***		
	[0.001]	[0.001]		
ln(firm age)	-0.006***	-0.006***	-0.004*	0.004
	[0.001]	[0.001]	[0.002]	[0.002]
N(observations, unweighted)	792,633	792,492	792,633	792,492
Adjusted R^2		0.056		0.012
Labour (l)	0.001		0.001	
k-l ratio	0.023		0.009	
Average worker skill	0.008		0.001	
FFE	0.014			
ln(firm age)	0.000		0.001	
Separate univariate regressions	Y	N	Y	N
Full covariate set included	N	Y	N	Y

Dependent variable is an indicator variable set equal to one for firms on the frontier, and zero otherwise. All regressions include production function industry×year dummies. Robust standard errors (clustered on firm) reported in square brackets. Standard errors are not adjusted for estimation at first stage (ie, Cobb-Douglas and translog MFP, which determine frontier status). Stars (***;**;*) indicate a coefficient significantly different from zero (at the 1%;5%;10% level respectively). OLS regressions (columns 1 & 2) are weighted using productivity weights. Fixed effects (FE) regressions utilise firm-level average productivity weights. Weighted observation counts are reported at the bottom of table 3. Regressions excluding the full covariate set (columns 1 & 3) are separate univariate regressions for each independent variable, stacked to aid presentation of the results (with each related adjusted R^2 reported at the bottom of the table). Regressions including the full covariate set (columns 2 & 4) include all variables in this table in a single regression that also includes Urban Area employment share variables. UA employment share coefficients are reported separately in table 20 (columns 2 & 4). Fixed effects regressions exclude the firm wage fixed effect (FFE) as this is constant for each firm (by construction).

Table 20: Estimated relationship between frontier productivity and location

Dependent variable: δ(frontier)	(1) OLS	(2) OLS	(3) FE	(4) FE
Urban Area employment share				
Hamilton	-0.016***	-0.004	-0.010	-0.010
	[0.003]	[0.003]	[0.015]	[0.015]
Tauranga	-0.013***	-0.002	-0.027*	-0.023
	[0.004]	[0.004]	[0.016]	[0.015]
Napier	-0.024***	-0.012***	0.000	-0.002
	[0.004]	[0.004]	[0.020]	[0.019]
Wellington	-0.001	-0.001	0.002	0.006
	[0.003]	[0.003]	[0.014]	[0.014]
Christchurch	-0.009***	-0.005*	-0.026**	-0.027**
	[0.002]	[0.002]	[0.012]	[0.012]
Dunedin	-0.027***	-0.013***	-0.040*	-0.038*
	[0.005]	[0.004]	[0.021]	[0.020]
Other main	-0.027***	-0.014***	-0.009	-0.009
	[0.002]	[0.002]	[0.010]	[0.010]
Secondary	-0.031***	-0.018***	-0.024**	-0.026**
	[0.003]	[0.003]	[0.011]	[0.010]
Minor	-0.041***	-0.027***	-0.020**	-0.023**
	[0.002]	[0.002]	[0.010]	[0.010]
Rural	-0.042***	-0.019***	-0.013	-0.015*
	[0.002]	[0.002]	[0.008]	[0.008]
N(observations, unweighted)	792,492	792,492	792,492	792,492
Adjusted R^2	0.003	0.056	0.001	0.012
Full covariate set included	N	Y	N	Y

Dependent variable is an indicator variable set equal to one for firms on the frontier, and zero otherwise. All regressions include production function industry×year dummies. Robust standard errors (clustered on firm) reported in square brackets. Standard errors are not adjusted for estimation at first stage (ie, Cobb-Douglas and translog MFP, which determine frontier status). Stars (***;**;*) indicate a coefficient significantly different from zero (at the 1%;5%;10% level respectively). OLS regressions (columns 1 & 2) are weighted using productivity weights. Fixed effects (FE) regressions utilise firm-level average productivity weights. Weighted observation counts are reported at the bottom of table 3. Regressions including the full covariate set (columns 2 & 4) include labour (l), the k-l ratio, average worker skill, firm wage fixed effect, and ln(firm age). Coefficients on these variables are reported separately in table 19. Auckland is the omitted UA employment share category.

Table 21: Estimated relationship between frontier productivity and full coverage firm characteristics by sector

Dependent variable:	OLS				FE			
δ(frontier)	Construct.	Manu.	Primary	Services	Construct.	Manu.	Primary	Services
Labour (l)	-0.016***	-0.019***	-0.015***	-0.009***	0.000	-0.031***	-0.021***	-0.046***
	[0.002]	[0.002]	[0.002]	[0.001]	[0.003]	[0.004]	[0.004]	[0.002]
Capital-labour ratio ($k-l$)	-0.045***	-0.049***	-0.079***	-0.050***	-0.036***	-0.047***	-0.067***	-0.062***
	[0.002]	[0.004]	[0.002]	[0.001]	[0.003]	[0.004]	[0.003]	[0.002]
Average worker skill	0.012***	0.043***	0.026***	0.037***	0.010***	0.019***	0.010***	0.021***
	[0.002]	[0.003]	[0.002]	[0.001]	[0.002]	[0.003]	[0.002]	[0.001]
Firm wage fixed effect (FFE)	0.037***	0.055***	0.032***	0.041***				
	[0.002]	[0.003]	[0.002]	[0.001]				
ln(firm age)	-0.017***	-0.006**	-0.023***	0.002**	-0.026***	0.015**	-0.008	0.011***
	[0.002]	[0.002]	[0.002]	[0.001]	[0.006]	[0.007]	[0.006]	[0.003]
Urban Area employment share								
Secondary	-0.016***	-0.009	0.013	-0.014***	-0.018	-0.023	-0.005	-0.011
	[0.005]	[0.007]	[0.010]	[0.003]	[0.015]	[0.027]	[0.024]	[0.011]
Minor	-0.025***	-0.015***	0.027***	-0.025***	-0.010	-0.029	0.008	-0.014
	[0.004]	[0.006]	[0.009]	[0.002]	[0.016]	[0.020]	[0.022]	[0.009]
Rural	-0.010***	0.000	0.008	-0.021***	-0.007	-0.019	0.010	-0.004
	[0.004]	[0.006]	[0.005]	[0.003]	[0.011]	[0.017]	[0.014]	[0.007]
N(observations, unweighted)	792,492				792,492			
Adjusted R^2	0.061				0.013			

Dependent variable is an indicator variable set equal to one for firms on the frontier, and zero otherwise. All regressions include production function industry×year dummies. Robust standard errors (clustered on firm) reported in square brackets. Standard errors are not adjusted for estimation at first stage (ie, Cobb-Douglas and translog MFP, which determine frontier status). Stars (***;**;*) indicate a coefficient significantly different from zero (at the 1%;5%;10% level respectively). OLS regression weighted using productivity weights. Fixed effects (FE) regression utilise firm-level average productivity weights. Weighted observation counts are reported at the bottom of table 3. Coefficients by sector estimated in single regression (either OLS or FE), and stacked side-by-side to aid presentation of the results. Fixed effects regressions exclude the firm wage fixed effect (FFE) as this is constant for each firm (by construction). All main Urban Areas pooled is the omitted UA employment share category, reducing the number of coefficients to estimate/report compared to table 20 (where Auckland UA was the omitted category).

Table 22: Estimated relationship between frontier productivity and BOS characteristics available annually

Dependent variable: δ(frontier)	(1) OLS	(2)	(3) FE	(4)
Has exports	-0.014	-0.030***	-0.012	0.000
	[0.010]	[0.011]	[0.008]	[0.010]
Entered new market	-0.027**	-0.021*	-0.020**	-0.024**
	[0.011]	[0.013]	[0.010]	[0.011]
Export share of sales	0.078***	0.092***	0.029	0.009
	[0.022]	[0.029]	[0.020]	[0.026]
Has foreign direct investment (FDI)	0.063	0.070	0.039	0.031
	[0.059]	[0.067]	[0.027]	[0.027]
FDI share	0.029	-0.040	-0.021	0.001
	[0.064]	[0.073]	[0.034]	[0.041]
Has outward direct investment (ODI)	0.030*	-0.021	-0.006	0.008
	[0.018]	[0.015]	[0.009]	[0.012]
Has R&D	-0.003	-0.005	-0.013**	-0.012*
	[0.011]	[0.012]	[0.006]	[0.007]
R&D (share of M)	0.091**	0.004	0.085**	0.133***
	[0.045]	[0.050]	[0.042]	[0.050]
Successful innovation in the last year	-0.017***	-0.011*	-0.005	-0.003
	[0.006]	[0.006]	[0.003]	[0.004]
No employees on collective agreement	0.004	0.003	0.000	0.002
	[0.007]	[0.007]	[0.005]	[0.007]
Core equipment up-to-date	0.023***	0.013**	0.005	0.001
	[0.006]	[0.007]	[0.004]	[0.005]
No competitors	0.068***	0.052***	0.017	0.016
	[0.018]	[0.018]	[0.014]	[0.019]
N(observations, unweighted)	See T.16	41,847	See T.16	41,847
Adjusted R^2		0.082		0.029
Exporting	0.015		0.011	
FDI	0.017		0.011	
ODI	0.012		0.011	
R&D	0.012		0.011	
Innovation	0.012		0.011	
No collective agreement	0.013		0.011	
Core equipment up-to-date	0.015		0.012	
No competitors	0.014		0.011	
Separate regression for variable groups	Y	N	Y	N
Full covariate set included	N	Y	N	Y

Dependent variable is an indicator variable set equal to one for firms on the frontier, and zero otherwise. All regressions include production function industry×year dummies. Robust standard errors (clustered on firm) reported in square brackets. Standard errors are not adjusted for estimation at first stage (ie, Cobb-Douglas and translog MFP, which determine frontier status). Stars (***;**;*) indicate a coefficient significantly different from zero (at the 1%;5%;10% level respectively). OLS regressions (columns 1 & 2) are weighted using BOS productivity weights. Fixed effects (FE) regressions utilise firm-level average BOS productivity weights. Models estimated with probability weights to reflect sample survey design (but not accounting for survey stratification because of reweighting). Regressions excluding the full covariate set (columns 1 & 3) are separate regressions for three groupings of variables (exports; FDI; and R&D) or for individual variables, stacked to aid presentation of the results (with the range of adjusted R^2's reported at the bottom of the table). Unweighted and weighted observation counts for these regressions are reported in table 16. Regressions including the full covariate set (columns 2 & 4) include all variables in this table in a single regression that also includes Urban Area employment share variables and the variables listed in table 19. Coefficients on these variables are not reported and are available from the author on request. Fixed effects regressions exclude the firm wage fixed effect (FFE) as this is constant for each firm (by construction). Only a single response category is included for collective employment agreement, core equipment and competition variables. Alternative estimates with complete enumeration of categories show similar results.

Table 23: Estimated relationship between frontier productivity and BOS characteristics available annually by sector

Dependent variable:	OLS				FE			
δ(frontier)	Construct.	Manu.	Primary	Services	Construct.	Manu.	Primary	Services
Has exports	-0.036	-0.002	0.004	-0.054***	-0.030	0.005	0.002	0.001
	[0.024]	[0.012]	[0.030]	[0.016]	[0.034]	[0.012]	[0.032]	[0.014]
Entered new market	-0.008	-0.014	0.032	-0.042**	-0.022	-0.015	-0.005	-0.040*
	[0.068]	[0.015]	[0.040]	[0.021]	[0.049]	[0.011]	[0.038]	[0.021]
Export share of sales	0.036	0.027	0.041	0.149***	0.003	0.063	0.028	-0.020
	[0.041]	[0.033]	[0.053]	[0.051]	[0.043]	[0.048]	[0.040]	[0.047]
Has foreign direct investment (FDI)	0.248	-0.033	-0.135*	0.104	0.066	0.023	0.068	0.029
	[0.218]	[0.026]	[0.076]	[0.089]	[0.176]	[0.030]	[0.077]	[0.037]
FDI share	-0.027	0.110***	0.187*	-0.102	0.088	-0.010	0.166	-0.014
	[0.283]	[0.036]	[0.099]	[0.096]	[0.275]	[0.047]	[0.144]	[0.053]
Has outward direct investment (ODI)	0.045	-0.018	-0.036	-0.023	-0.026	0.016	-0.048*	0.013
	[0.059]	[0.017]	[0.035]	[0.021]	[0.029]	[0.021]	[0.025]	[0.017]
Has R&D	-0.036**	-0.005	-0.021	0.001	-0.022	0.003	-0.002	-0.022*
	[0.014]	[0.011]	[0.030]	[0.020]	[0.021]	[0.009]	[0.025]	[0.011]
R&D (share of M)	0.096	0.144	-0.033	-0.033	-0.095	0.200	0.064	0.136**
	[0.082]	[0.106]	[0.149]	[0.061]	[0.144]	[0.150]	[0.121]	[0.053]
Successful innovation in the last year	-0.008	-0.006	-0.008	-0.011	-0.011	-0.006	-0.010	0.000
	[0.011]	[0.008]	[0.016]	[0.009]	[0.008]	[0.006]	[0.013]	[0.006]
No employees on collective agreement	-0.013	-0.006	0.020	0.008	0.009	-0.006	0.014	-0.001
	[0.017]	[0.013]	[0.015]	[0.010]	[0.014]	[0.013]	[0.015]	[0.010]
Core equipment up-to-date	0.007	0.018*	0.003	0.017*	-0.013	0.001	-0.030*	0.007
	[0.013]	[0.010]	[0.021]	[0.009]	[0.012]	[0.007]	[0.017]	[0.007]
No competitors	0.079	0.129***	0.013	0.049*	-0.023	-0.043	0.021	0.038
	[0.072]	[0.043]	[0.025]	[0.025]	[0.031]	[0.042]	[0.021]	[0.029]
N(observations, unweighted)	3,264	11,367	3,600	23,613	3,264	11,367	3,600	23,613
Adjusted R^2	0.053	0.071	0.176	0.092	0.033	0.031	0.036	0.038

Dependent variable is an indicator variable set equal to one for firms on the frontier, and zero otherwise. All regressions include production function industry×year dummies. Robust standard errors (clustered on firm) reported in square brackets. Standard errors are not adjusted for estimation at first stage (ie, Cobb-Douglas and translog MFP, which determine frontier status). Stars (***;**;*) indicate a coefficient significantly different from zero (at the 1%;5%;10% level respectively). OLS regressions are weighted using BOS productivity weights. Fixed effects (FE) regressions utilise firm-level average BOS productivity weights. Models estimated with probability weights to reflect sample survey design (but not accounting for survey stratification because of reweighting). Coefficients are estimated in sector-specific regressions, which also include Urban Area employment share variables and the variables listed in table 19. Coefficients on these variables are not reported and are available from the author on request. Fixed effects regressions exclude the firm wage fixed effect (FFE) as this is constant for each firm (by construction).

Table 24: Estimated relationship between frontier productivity and innovation outcomes

Dependent variable:	(1)	(2)	(3)	(4)
δ(frontier)	OLS		FE	
Has R&D	0.012	0.001	-0.012	-0.016
	[0.012]	[0.013]	[0.011]	[0.012]
Has product design	-0.035***	-0.011	-0.025**	-0.028**
	[0.011]	[0.012]	[0.011]	[0.012]
Has product marketing/market research	-0.014	-0.005	-0.002	0.007
	[0.010]	[0.011]	[0.009]	[0.011]
Has other product development	-0.003	-0.008	0.001	-0.009
	[0.009]	[0.009]	[0.011]	[0.013]
Expenditure as share of M				
R&D	0.014	-0.052	0.013	0.031
	[0.051]	[0.062]	[0.070]	[0.111]
Product design	0.292**	0.295**	0.121	0.128
	[0.123]	[0.146]	[0.087]	[0.109]
Product marketing/market research	0.421***	0.348***	0.238*	0.185
	[0.119]	[0.127]	[0.126]	[0.157]
Other product development	-0.050	-0.038	-0.003	0.084
	[0.061]	[0.076]	[0.059]	[0.080]
Successful innovation in the last two years				
New product	-0.003	-0.009	-0.005	-0.002
	[0.007]	[0.009]	[0.007]	[0.010]
New operational process	-0.016**	-0.020**	-0.003	-0.012
	[0.007]	[0.009]	[0.007]	[0.010]
New organisational process	0.004	0.008	0.005	0.010
	[0.007]	[0.010]	[0.006]	[0.008]
New marketing method	-0.021***	-0.013	-0.021***	-0.015
	[0.008]	[0.010]	[0.007]	[0.011]
N(observations, unweighted)	See T.16	17,616	See T.16	17,616
Adjusted R^2		0.088		0.026
Product development	0.019		0.013	
Innovation	0.012		0.013	
Separate regression for variable groups	Y	N	Y	N
Full covariate set included	N	Y	N	Y

Regressions excluding the full covariate set (columns 1 & 3) are separate regressions for two groupings of variables (product development expenditure; and innovation outcomes), stacked to aid presentation of the results (with the respective adjusted R^2's reported at the bottom of the table). Unweighted and weighted observation counts for these regressions are reported in table 16. Regressions including the full covariate set (columns 2 & 4) include all variables in this table in a single regression that also includes Urban Area employment share variables, the variables listed in table 19, and the annually available BOS variables listed in table 22. Coefficients on these variables are not reported and are available from the author on request. Fixed effects regressions exclude the firm wage fixed effect (FFE) as this is constant for each firm (by construction). See table 22 for additional notes.

Table 25: Estimated relationship between frontier productivity and HRM practices

Dependent variable: δ(frontier)	(1) OLS	(2)	(3) FE	(4)
HRM practices (z-score)	0.014***	0.011**	0.000	0.000
	[0.004]	[0.005]	[0.005]	[0.007]
N(observations, unweighted)	16,659	12,009	16,659	12,009
Adjusted R^2	0.013	0.090	0.015	0.043
Full covariate set included	N	Y	N	Y

Regressions including the full covariate set (columns 2 & 4) include all variables in this table in a single regression that also includes Urban Area employment share variables, the variables listed in table 19, and the annually available BOS variables listed in table 22. Coefficients on these variables are not reported and are available from the author on request. Fixed effects regressions exclude the firm wage fixed effect (FFE) as this is constant for each firm (by construction). See table 22 for additional notes.

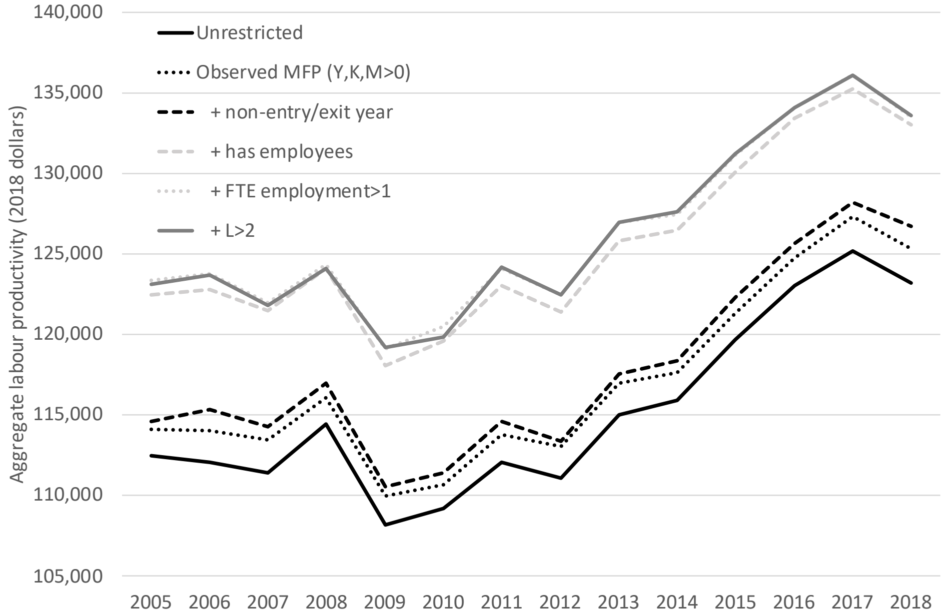
Table 26: Estimated relationship between frontier productivity and UFB

Dependent variable: δ(frontier)	(1) OLS	(2)	(3) FE	(4)
Has ultrafast broadband (UFB)	0.023**	0.001	0.005	0.009
	[0.010]	[0.013]	[0.007]	[0.007]
Complementary ICT investments	-0.009**	-0.013***	0.004	0.001
	[0.004]	[0.005]	[0.004]	[0.006]
UFB×complementary investments	0.007	0.007	-0.002	0.002
	[0.007]	[0.007]	[0.005]	[0.006]
N(observations, unweighted)	19,683	13,692	19,683	13,692
Adjusted R^2	0.018	0.095	0.019	0.061
Full covariate set included	N	Y	N	Y

Regressions including the full covariate set (columns 2 & 4) include all variables in this table in a single regression that also includes Urban Area employment share variables, the variables listed in table 19, and the annually available BOS variables listed in table 22. Coefficients on these variables are not reported and are available from the author on request. Fixed effects regressions exclude the firm wage fixed effect (FFE) as this is constant for each firm (by construction). See table 22 for additional notes.

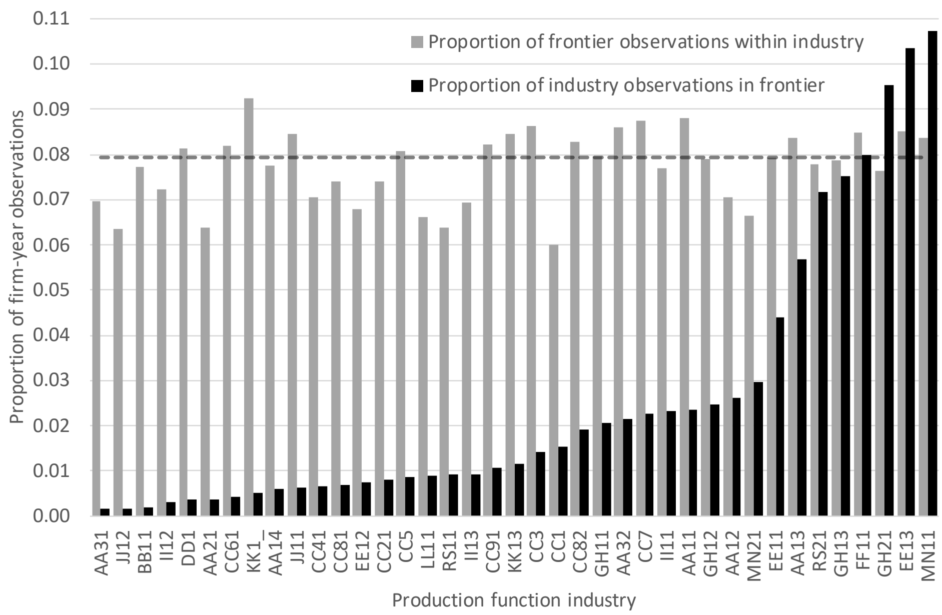
Figures

Figure 1: Effect of population restrictions on aggregate labour productivity



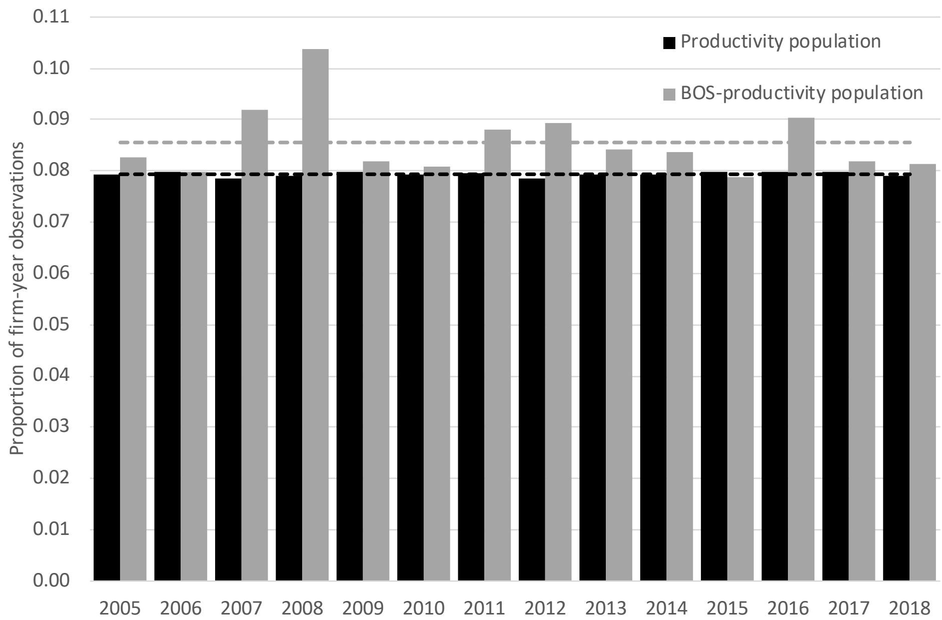
Aggregate labour productivity is (sub)population total value-added over total labour input (ie, $[\Sigma Y - \Sigma M]/\Sigma L$). Population restrictions are sequential and cumulative as indicated in table 1. Associated table footnote describes deflators.

Figure 2: Proportion of firms in composite productivity frontier by industry



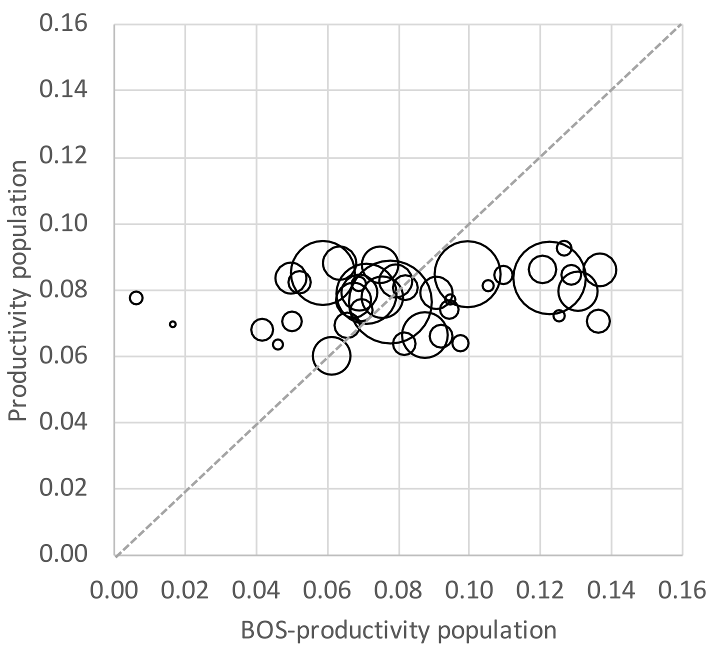
The dashed horizontal line shows the average proportion of frontier observations with all industries pooled. Descriptions of the production function industries are in appendix table A.1.

Figure 3: Proportion of firms in productivity frontier by year – productivity vs BOS-productivity



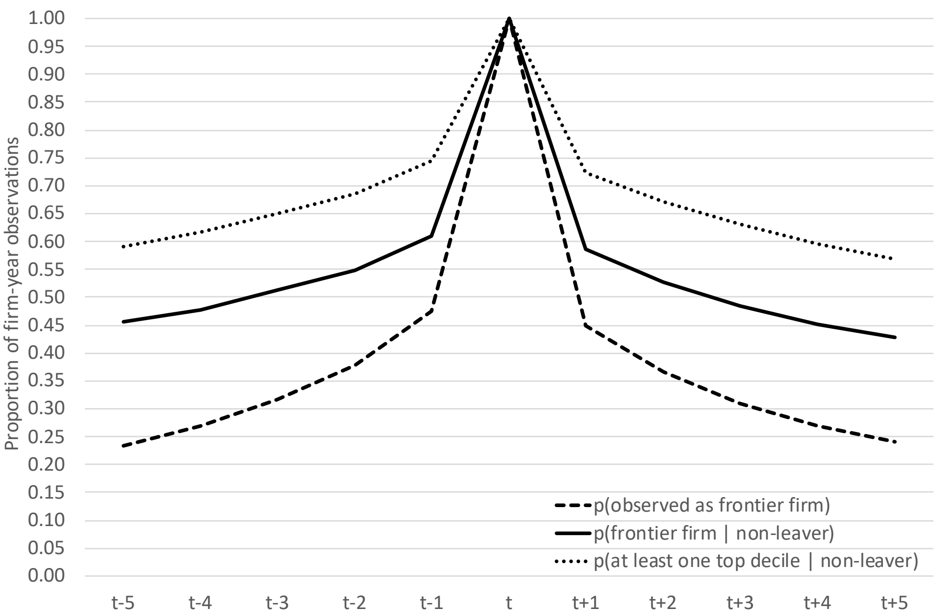
The dashed horizontal line shows the average proportion of frontier observations with all years pooled, which is 0.079 for the productivity subpopulation and 0.086 for the BOS-productivity subpopulation. See table 4 for notes on the weighting method used to estimate the BOS-productivity proportions.

Figure 4: Proportion of firms in productivity frontier by industry – productivity vs BOS-productivity



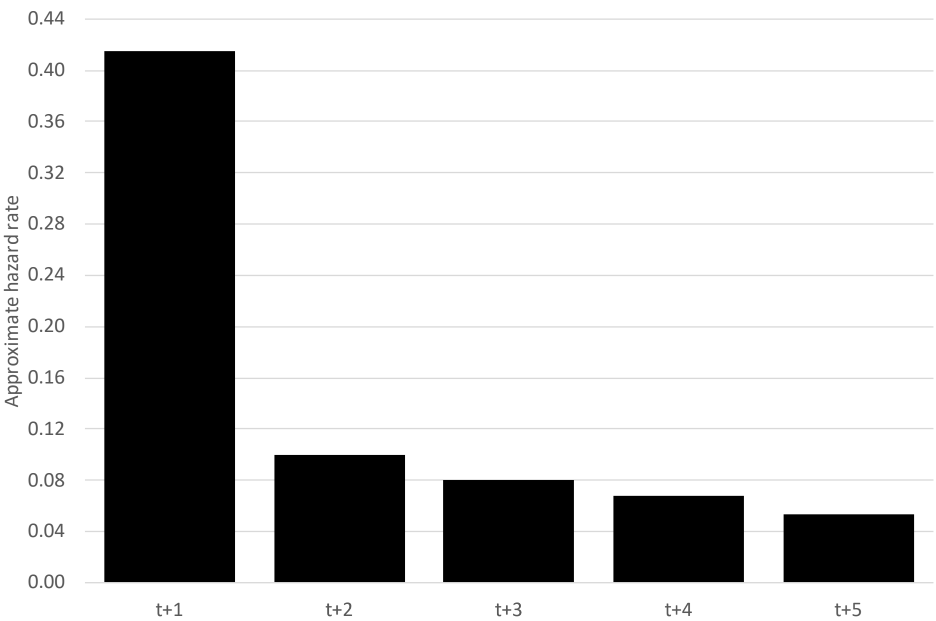
All years pooled. Bubble size scaled to total number of firms in each BOS-productivity industry. Dashed line indicates consistent proportion of frontier firms within industry between productivity and BOS-productivity subpopulations.

Figure 5: Persistence in the frontier up to five years before and after presence



Analysis excludes industries with insufficient firms in the detailed (3-digit) industry to justify identifying the CD_{3d} and TL_{3d} frontiers, so that the metric of "at least one top decile" is consistent across firms. The pre-t (post-t) pattern is constructed for $t \geqslant 2010$ ($t \leqslant 2013$) to maintain a consistent sample at each time period. While the dashed line doesn't adjust for population leavers, it is not affected by non-observation due to censoring because of the restrictions on t. Table 11 one-year persistence rates – reported in the last row of the table (0.472 unconditional; 0.597 conditional) – are higher than those in the figure (0.450 unconditional; 0.585 conditional) due to the different time periods covered.

Figure 6: Approximate hazard rate for frontier exit, conditional on non-leaver



Hazard rate derived from the year-on-year change in probability of being in the frontier, conditional on being a non-leaver (ie, the solid line in figure 5). The hazard rate is approximate in two senses: it does not account for the fact that firms can leave and re-enter the frontier and/or the population; and it makes no attempt to correct for the fact that leaving the population may be related to firm performance and, therefore, persistence in the frontier.

Figure 7: Proportion of frontier observations by number of years observed and number of years observed in frontier

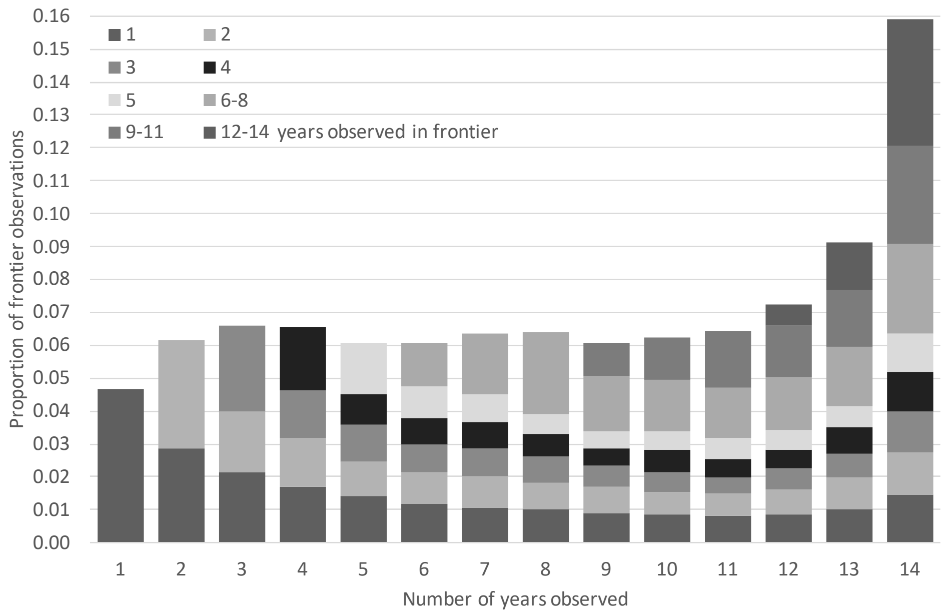
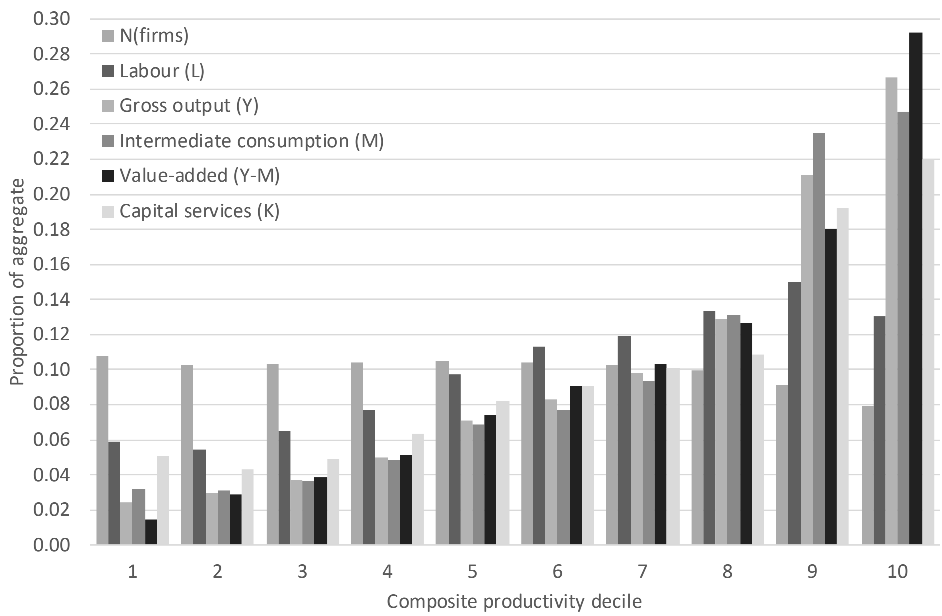
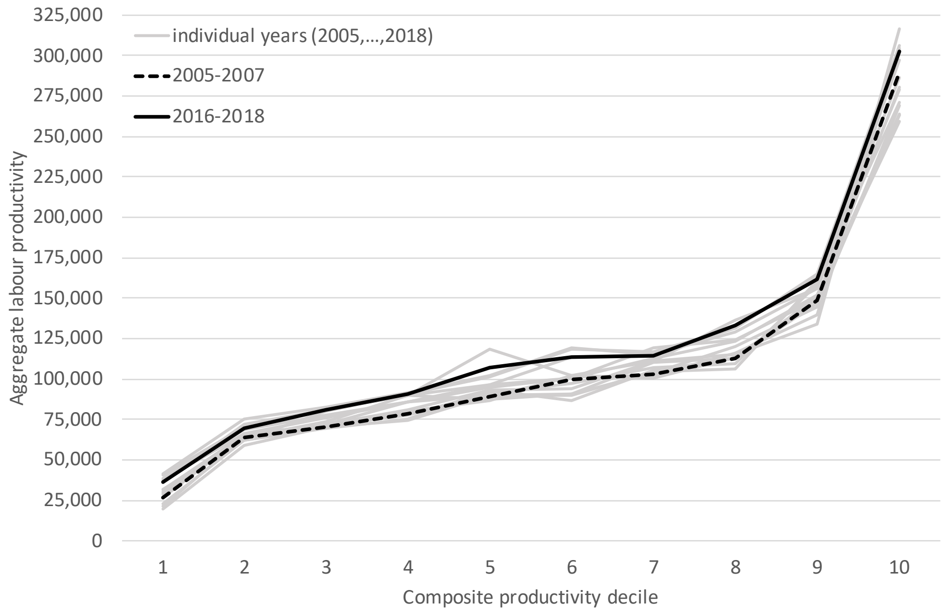


Figure 8: Share of aggregates by composite productivity decile



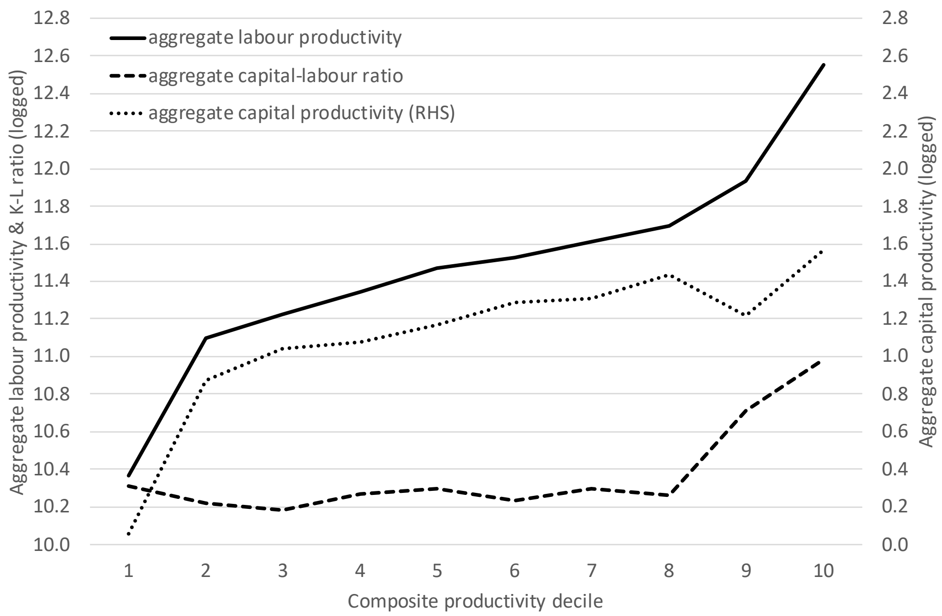
All years pooled using deflators described in the table 1 footnote. The composite productivity decile is the median of the four MFP deciles (CD, TL, CD_{3d}, TL_{3d}) rounded down, as in table 10.

Figure 9: Aggregate labour productivity by composite decile and year



The composite productivity decile is the median of the four MFP deciles (CD, TL, CD_{3d}, TL_{3d}) rounded down, as in table 10. Each individual year appears in grey, while the average of the two periods 2005-2007 and 2016-2018 is highlighted (in black dashed and solid lines respectively) to illustrate aggregate change over entire period. The change in aggregate labour productivity over these two time periods is decomposed in table 14.

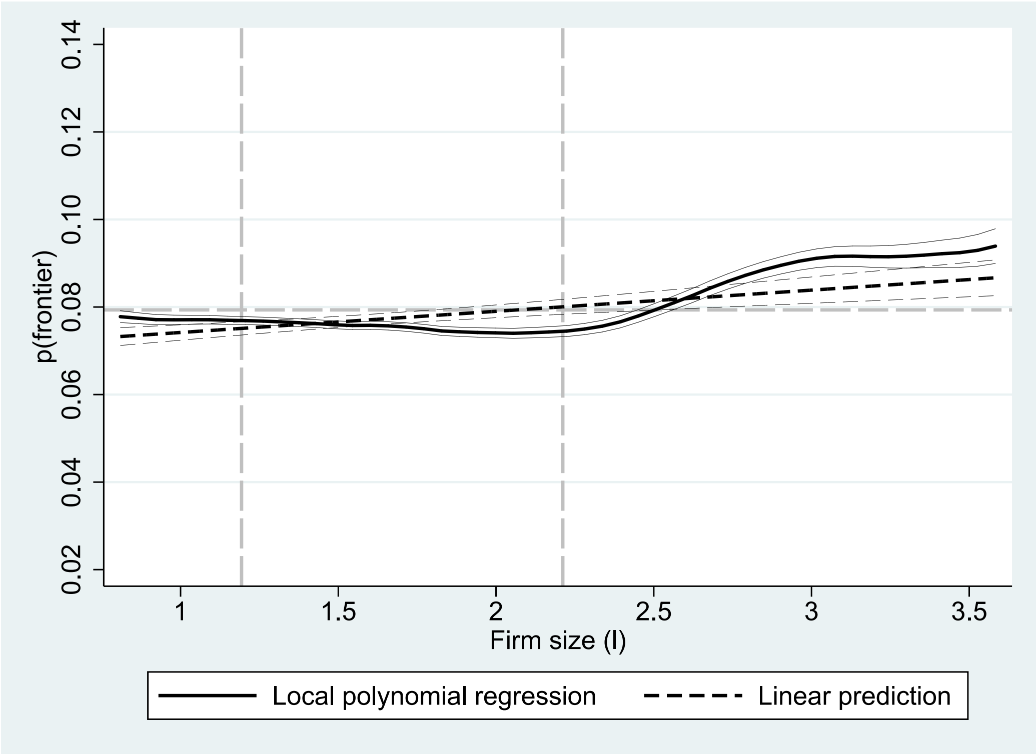
Figure 10: Aggregate K-L ratio and capital productivity by composite decile



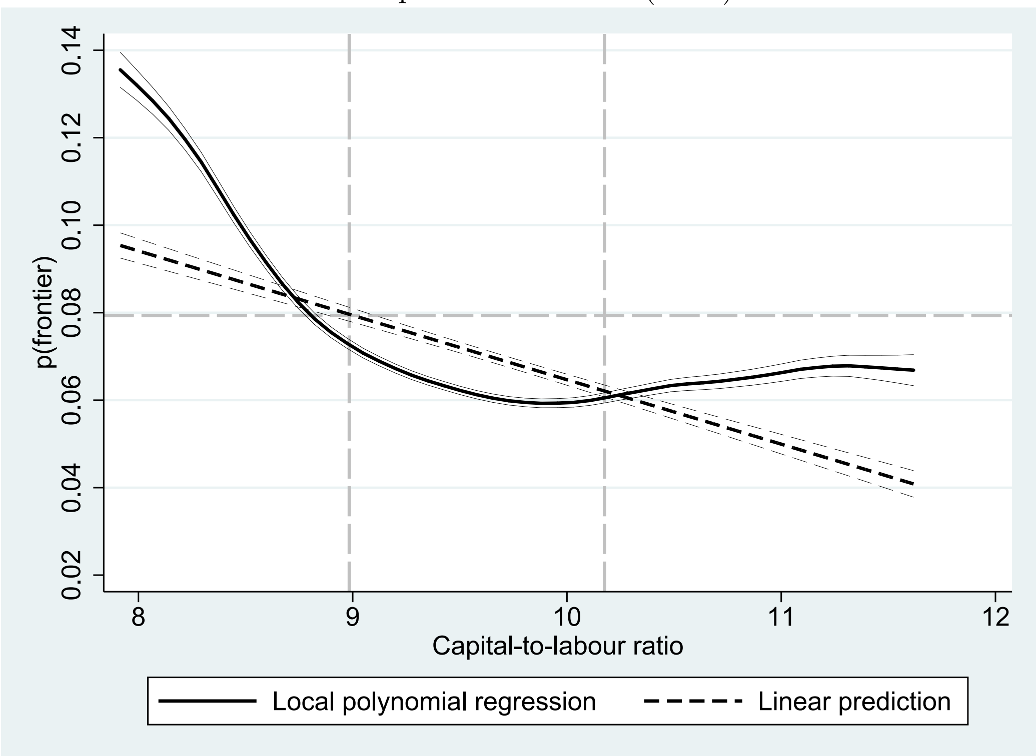
All years pooled using deflators described in the table 1 footnote. The composite productivity decile is the median of the four MFP deciles (CD, TL, CD_{3d}, TL_{3d}) rounded down, as in table 10. The aggregate capital-labour (K-L) ratio is total capital services over total labour input (ie, $\Sigma K/\Sigma L$). In natural logs, aggregate capital productivity (the dotted line) is the difference between aggregate labour productivity (the solid line) and the K-L ratio (the dashed line).

Figure 11: Estimated relationship between frontier firm and l or k-l ratio

Labour (l)



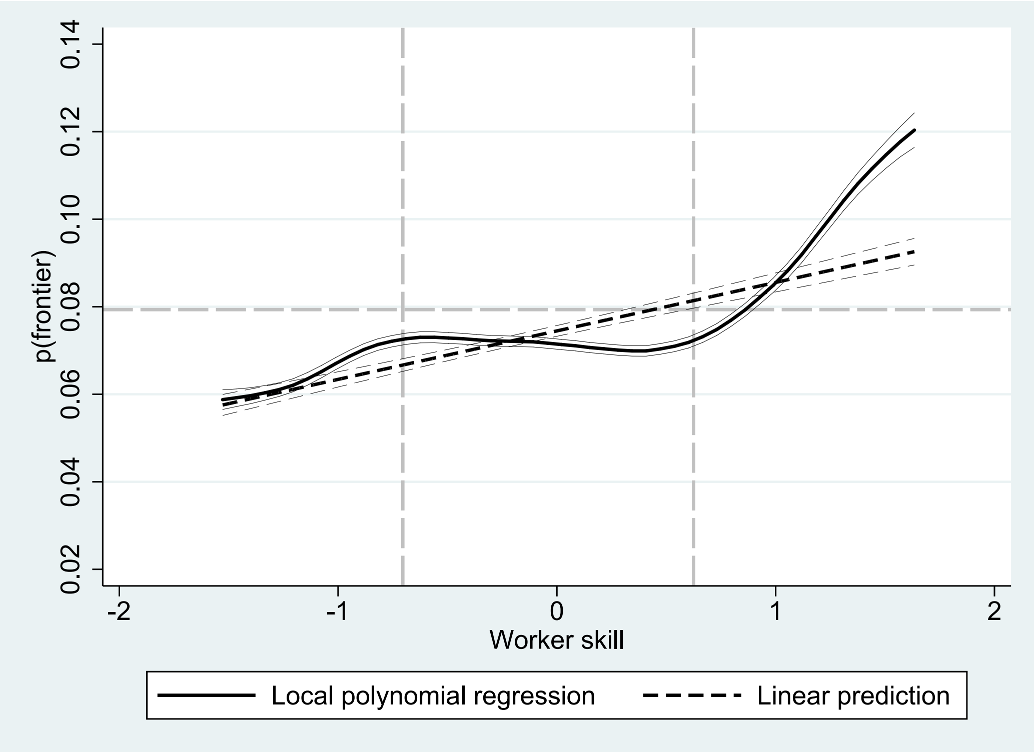
Capital-labour ratio ($k - l$)



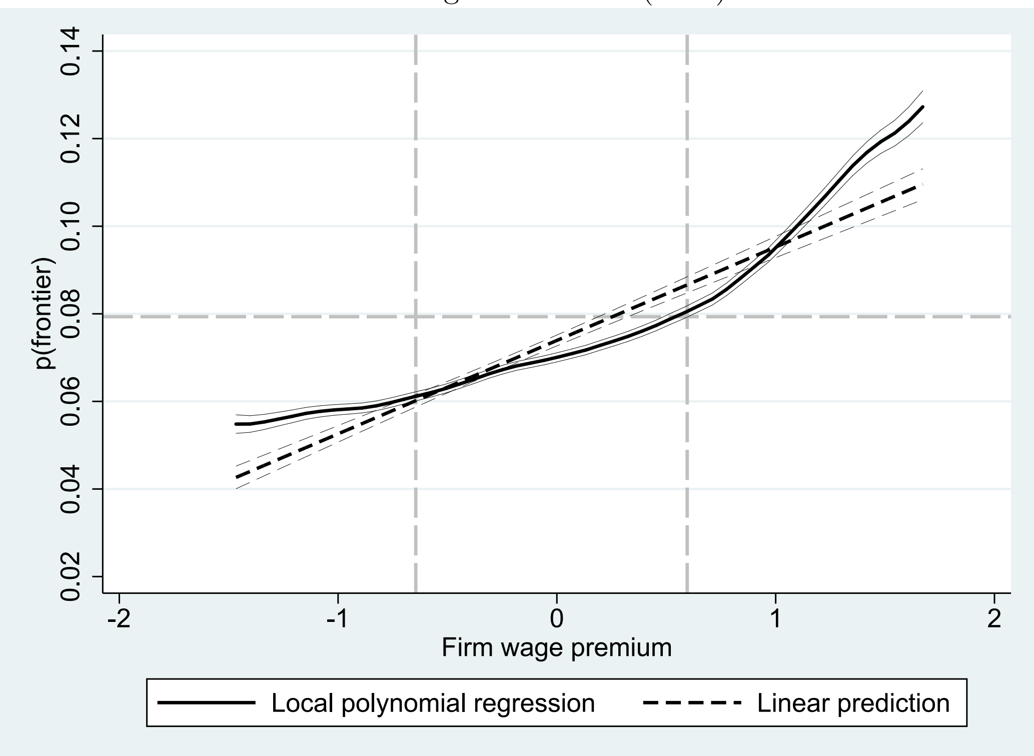
The dependent variable is the probability of being a frontier firm. Smoothed propensities, including 95 percent confidence intervals, are derived from (Epanechnikov) kernel-weighted local polynomial regressions (using Stata's default rule-of-thumb bandwidth). The estimated linear fit (dashed black line) is also plotted with 95% confidence interval (robust standard errors clustered on firm). The horizontal dashed grey line shows the mean probability of being a frontier firm (ie, 0.079 from table 10). Vertical dashed grey lines signify the 25th and 75th percentile of the independent variable. The top and bottom 5% of observations of the independent variables are trimmed for confidentiality and for presentation purposes.

Figure 12: Estimated relationship between frontier firm and skill or FFE

Average worker skill



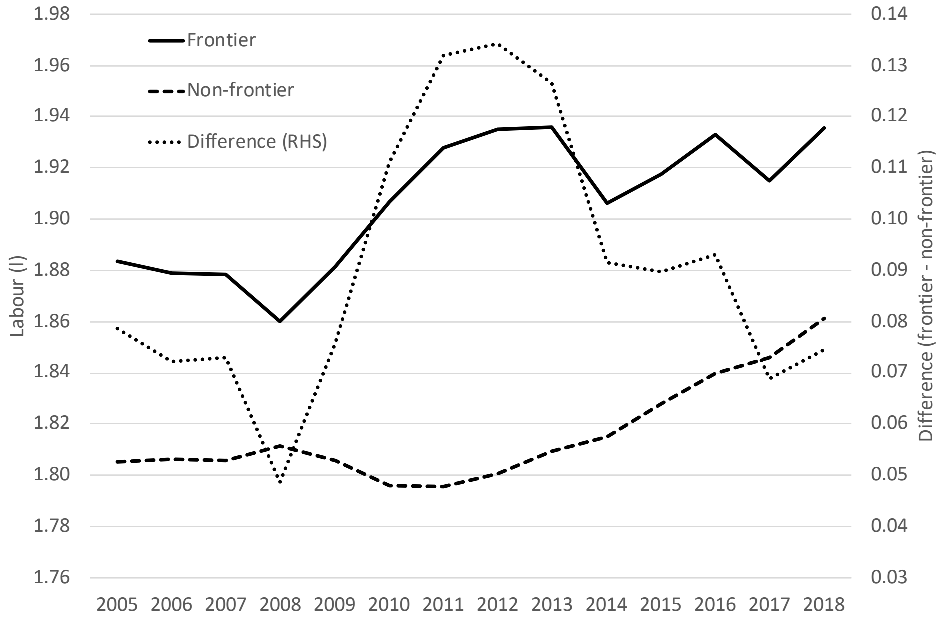
Firm wage fixed effect (FFE)



See figure 11 for notes.

Figure 13: Mean l and k-l ratio of firms over time – frontier vs non-frontier

Labour (l)



Capital-labour ratio ($k - l$)

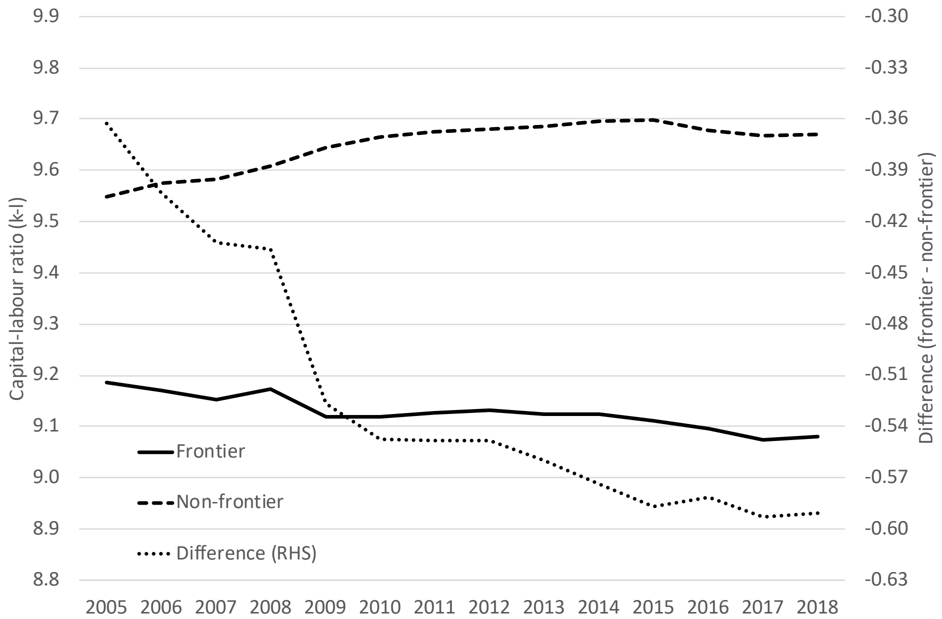
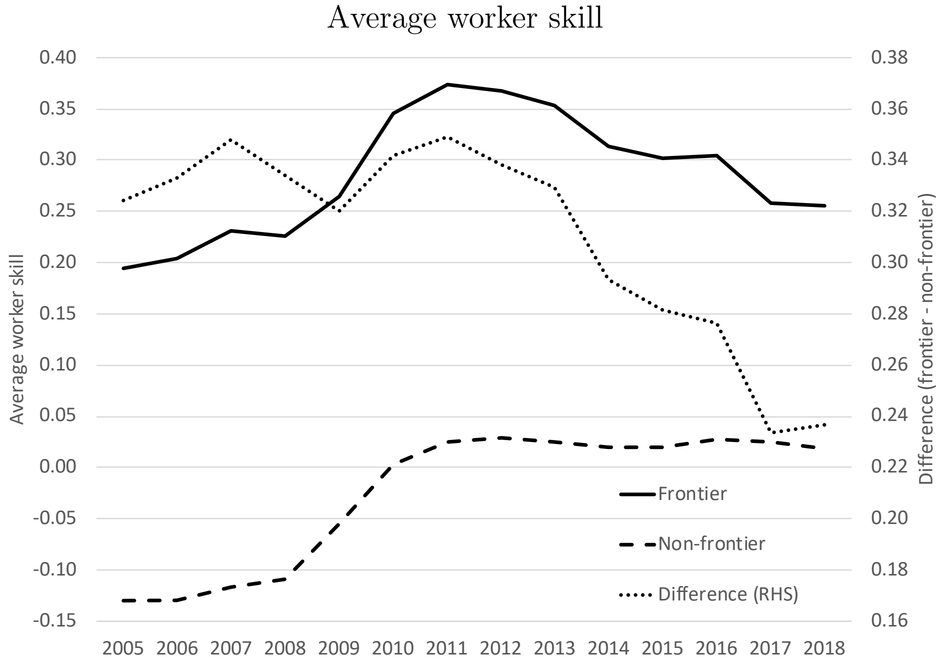


Figure 14: Mean skill and FFE of firms over time – frontier vs non-frontier



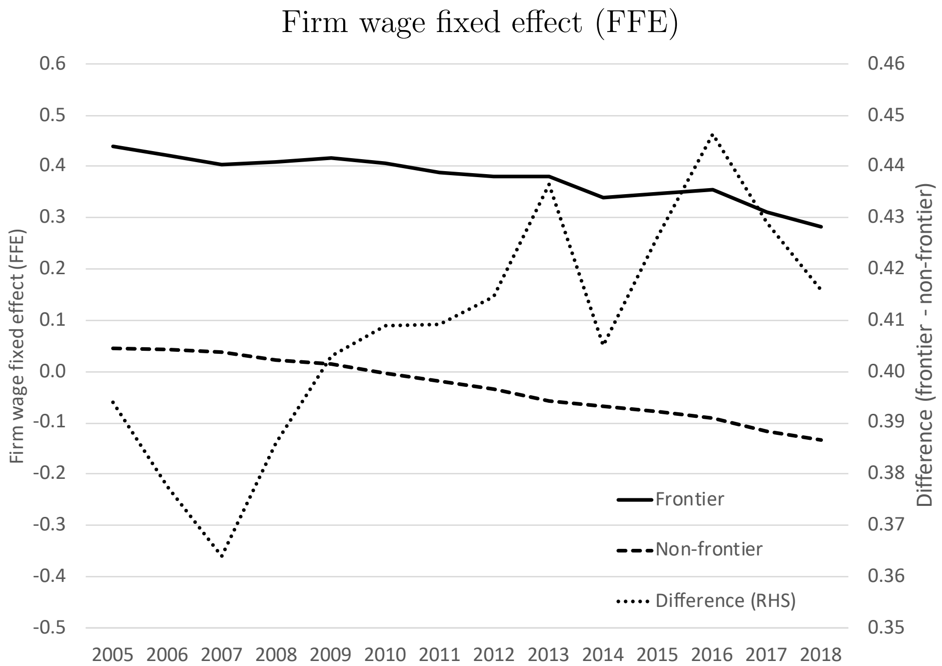


Figure 15: Proportion of firms with FDI over time – frontier vs non-frontier

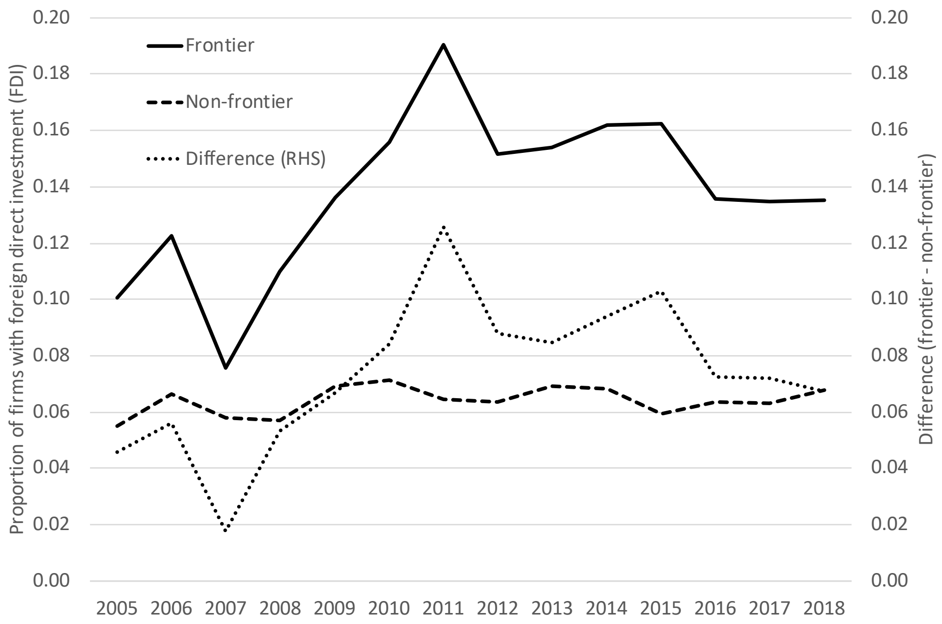


Figure 16: Proportion of firms with core equipment fully up-to-date over time – frontier vs non-frontier

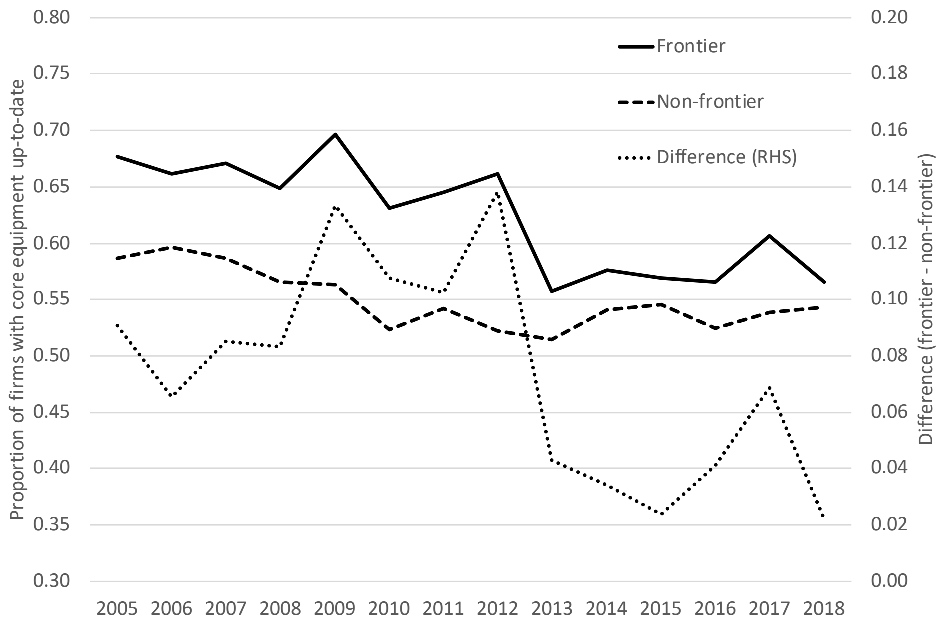


Figure 17: Mean R&D share of firms over time – frontier vs non-frontier

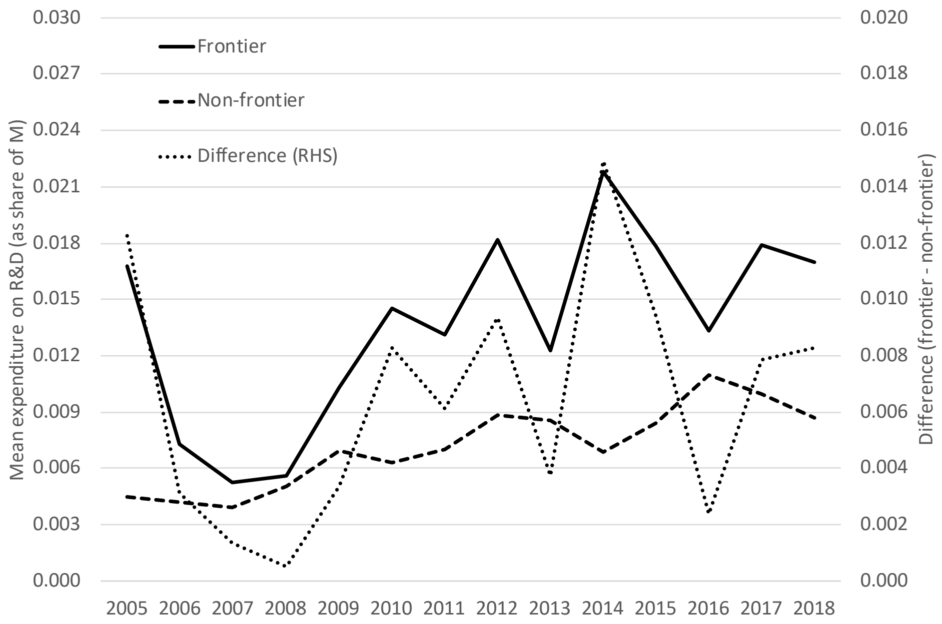
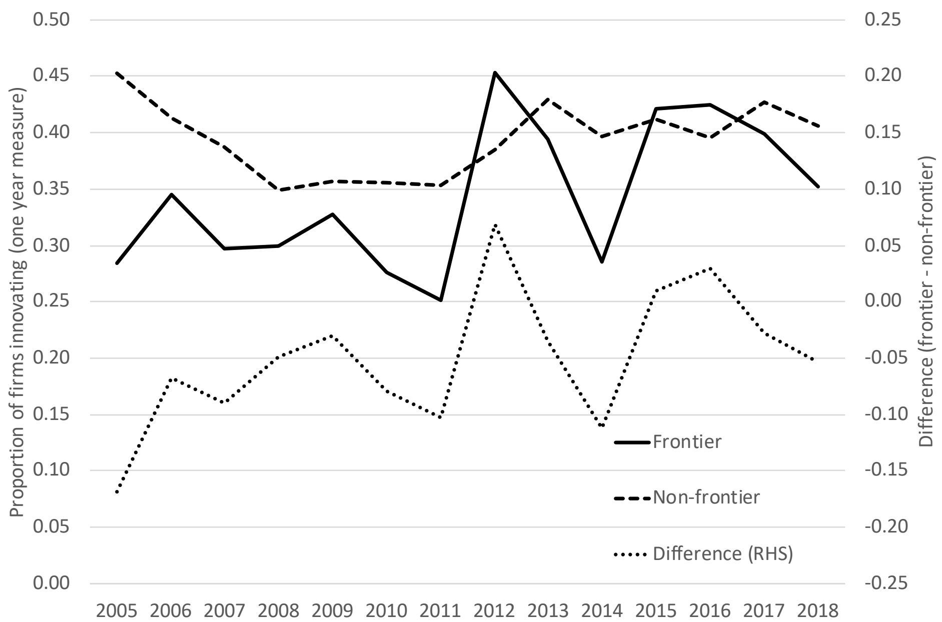


Figure 18: Proportion of firms innovating (one year) over time – frontier vs non-frontier



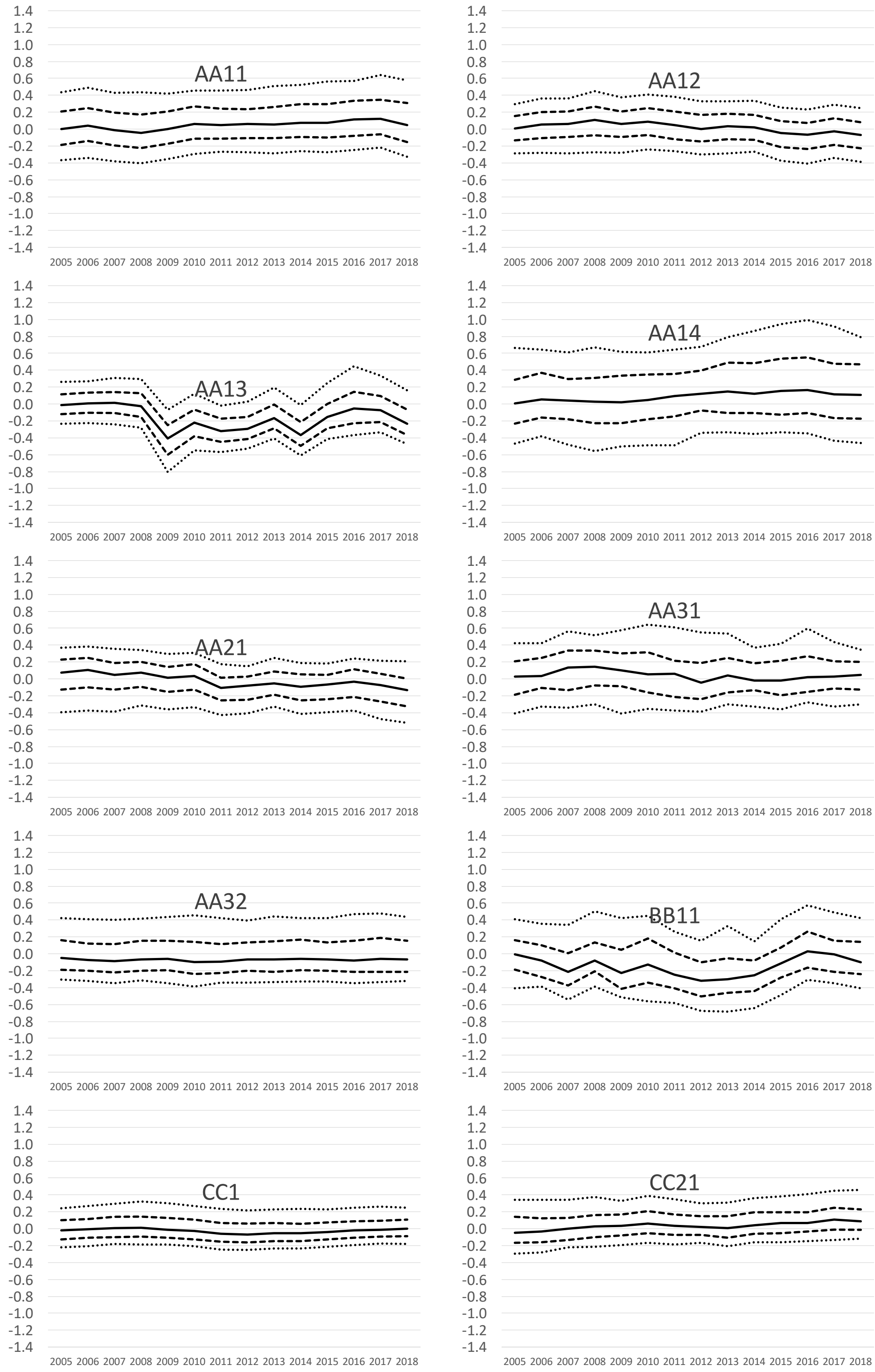
A. Appendix

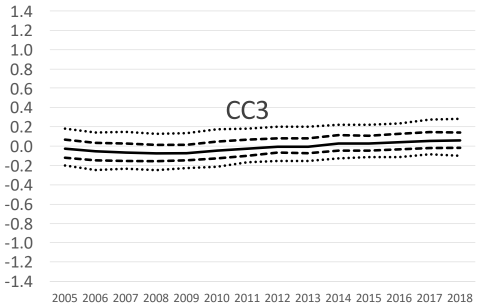
Table A.1: Production function industries

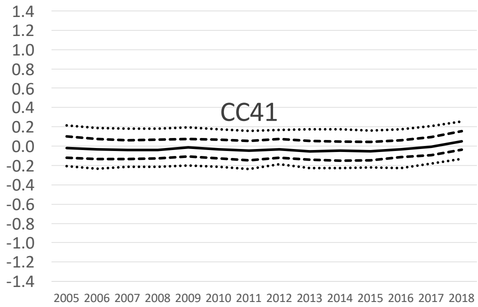
pf_ind	Industry (NZSIOC description)
AA11	Horticulture & Fruit Growing
AA12	Sheep, Beef Cattle & Grain Farming
AA13	Dairy Cattle Farming
AA14	Poultry, Deer & Other Livestock Farming
AA21	Forestry & Logging
AA31	Fishing & Aquaculture
AA32	Agriculture, Forestry & Fishing Support Services & Hunting
BB11	Mining
CC1	Food, Beverage & Tobacco Product Manufacturing
CC21	Textile, Leather, Clothing & Footwear Manufacturing
CC3	Wood & Paper Products Manufacturing
CC41	Printing
CC5	Petroleum, Chemical, Polymer & Rubber Product Manufacturing
CC61	Non-Metallic Mineral Product Manufacturing
CC7	Metal Product Manufacturing
CC81	Transport Equipment Manufacturing
CC82	Machinery & Other Equipment Manufacturing
CC91	Furniture & Other Manufacturing
DD1	Electricity, Gas, Water & Waste Services
EE11	Building Construction
EE12	Heavy & Civil Engineering Construction
EE13	Construction Services
FF11	Wholesale Trade
GH11	Motor Vehicle & Motor Vehicle Parts & Fuel Retailing
GH12	Supermarket, Grocery Stores & Specialised Food Retailing
GH13	Other Store-Based Retailing & Non Store Retailing
GH21	Accommodation & Food Services
II11	Road Transport
II12	Rail, Water, Air & Other Transport
II13	Postal, Courier Transport Support, & Warehousing Services.
JJ11	Information Media Services
JJ12	Telecommunications, Internet & Library Services
KK13	Auxiliary Finance & Insurance Services
KK1_	Non-auxiliary Finance & Insurance Services
LL11	Rental & Hiring Services (except Real Estate)
MN11	Professional, Scientific & Technical Services
MN21	Administrative & Support Services
RS11	Arts & Recreation Services
RS21	Other Services

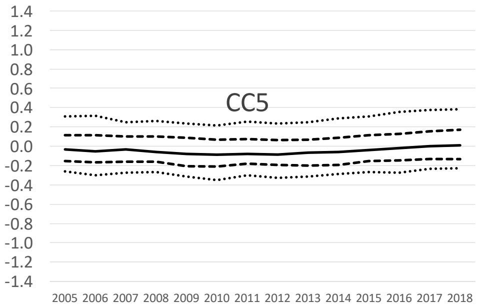
Production function industries are a mix of level two and level three New Zealand Standard Industry Output Categories (NZSIOC). KK1_ incorporates KK11 and KK12, which is the totality of KK1 (Finance & Insurance Services), excluding Auxiliary Finance & Insurance Services (KK13).

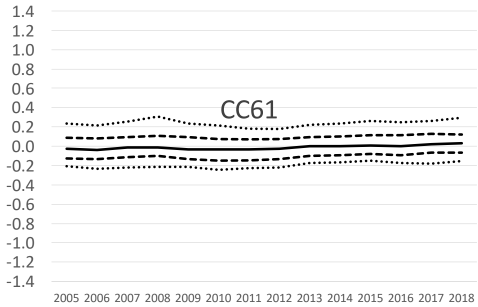
Figure A.1: 10th/25th/50th/75th/90th percentile of MFP (translog) by industry

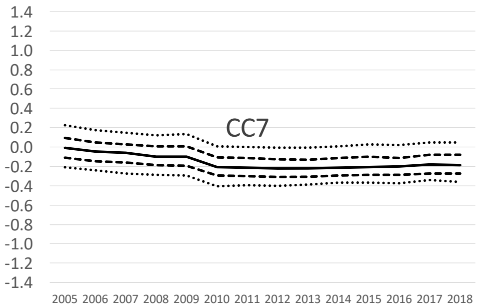


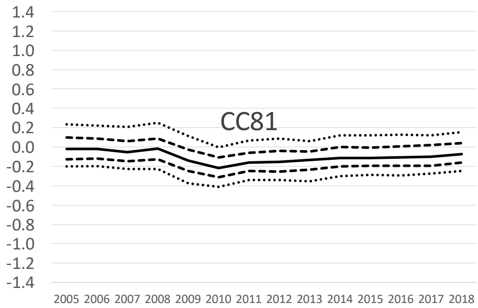

CC3

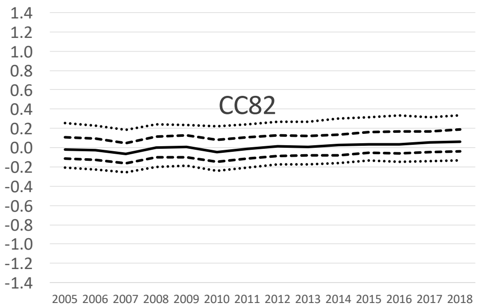

CC41

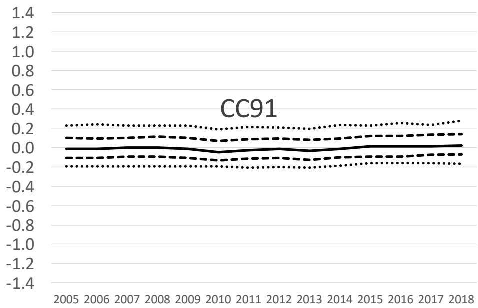

CC5

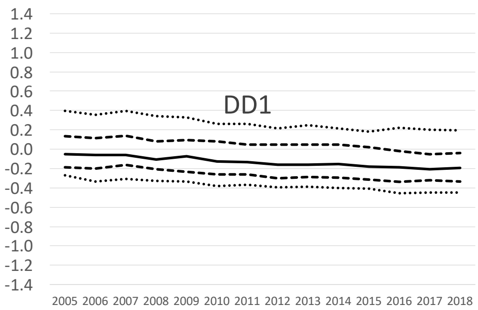

CC61

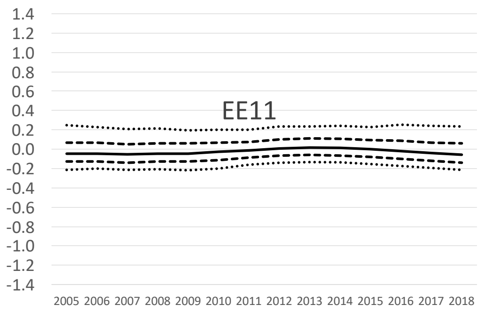

CC7

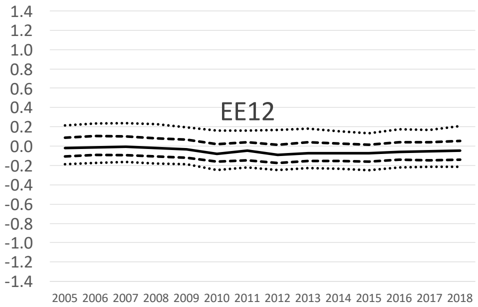

CC81

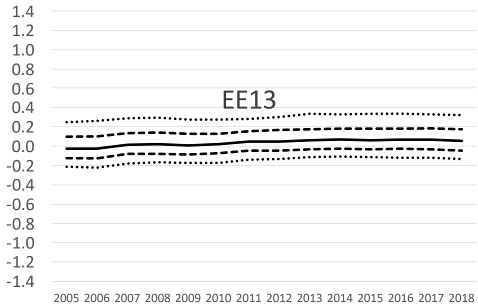

CC82

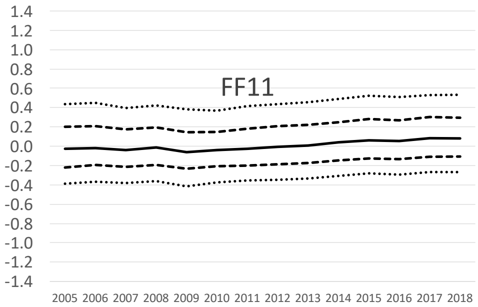

CC91

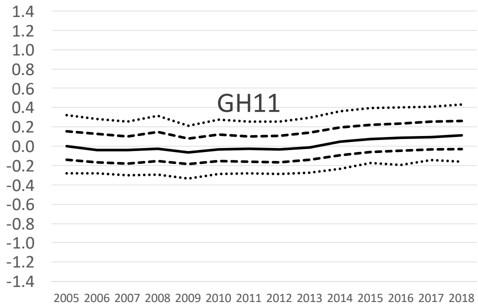

DD1

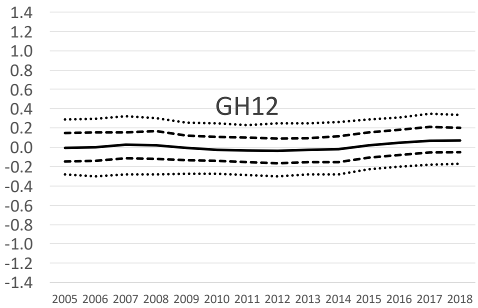

EE11

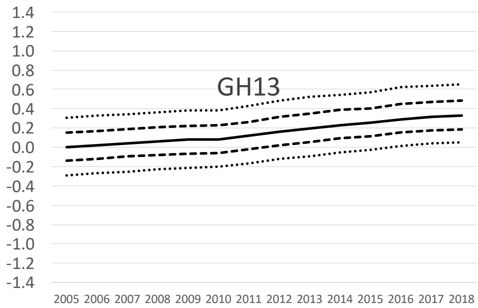

EE12
1.4 1.2 1.0 0.8 0.6 0.4 0.2 0.0 -0.2 -0.4 -0.6 -0.8 -1.0 -1.2 -1.4
2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018

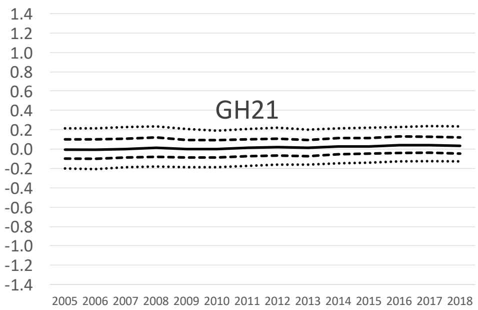

EE13
1.4 1.2 1.0 0.8 0.6 0.4 0.2 0.0 -0.2 -0.4 -0.6 -0.8 -1.0 -1.2 -1.4
2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018

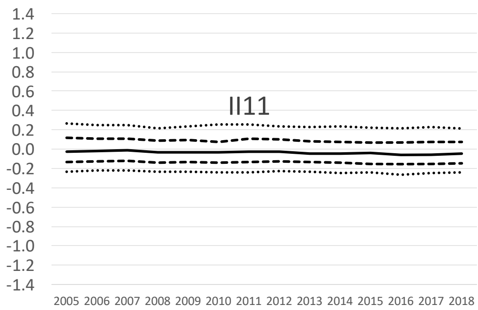

FF11
1.4 1.2 1.0 0.8 0.6 0.4 0.2 0.0 -0.2 -0.4 -0.6 -0.8 -1.0 -1.2 -1.4
2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018

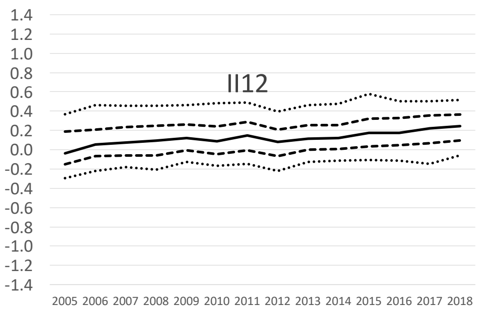

GH11
1.4 1.2 1.0 0.8 0.6 0.4 0.2 0.0 -0.2 -0.4 -0.6 -0.8 -1.0 -1.2 -1.4
2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018

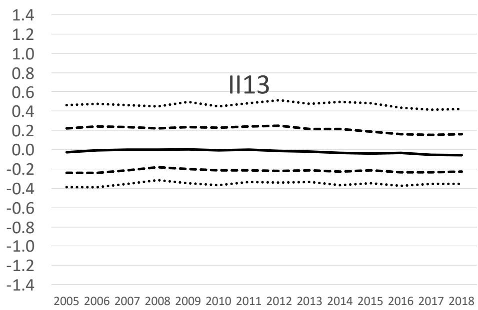

GH12
1.4 1.2 1.0 0.8 0.6 0.4 0.2 0.0 -0.2 -0.4 -0.6 -0.8 -1.0 -1.2 -1.4
2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018


GH13
1.4 1.2 1.0 0.8 0.6 0.4 0.2 0.0 -0.2 -0.4 -0.6 -0.8 -1.0 -1.2 -1.4
2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018


GH21
1.4 1.2 1.0 0.8 0.6 0.4 0.2 0.0 -0.2 -0.4 -0.6 -0.8 -1.0 -1.2 -1.4
2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018


II11
1.4 1.2 1.0 0.8 0.6 0.4 0.2 0.0 -0.2 -0.4 -0.6 -0.8 -1.0 -1.2 -1.4
2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018


II12
1.4 1.2 1.0 0.8 0.6 0.4 0.2 0.0 -0.2 -0.4 -0.6 -0.8 -1.0 -1.2 -1.4
2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018


II13
1.4 1.2 1.0 0.8 0.6 0.4 0.2 0.0 -0.2 -0.4 -0.6 -0.8 -1.0 -1.2 -1.4
2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018

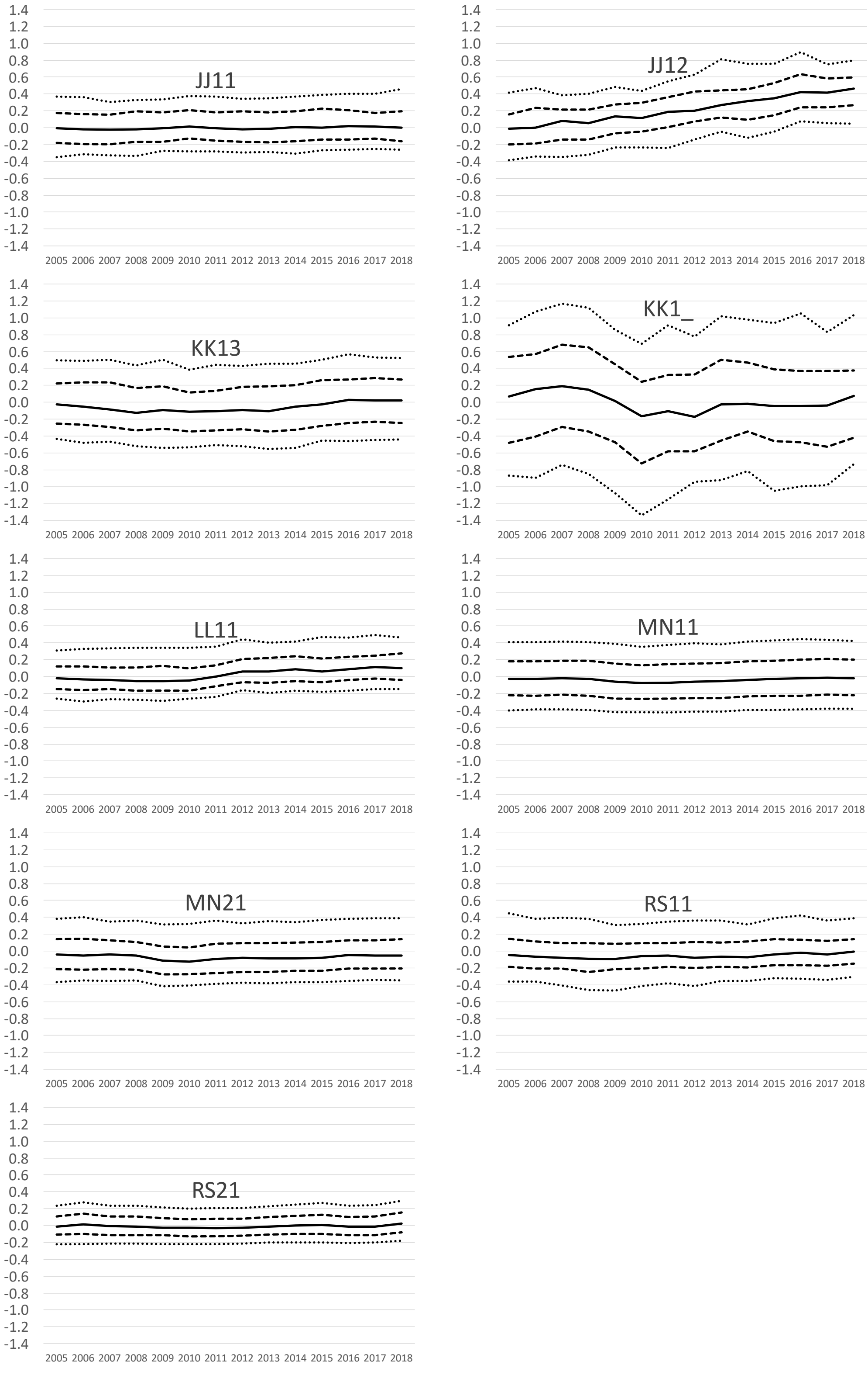
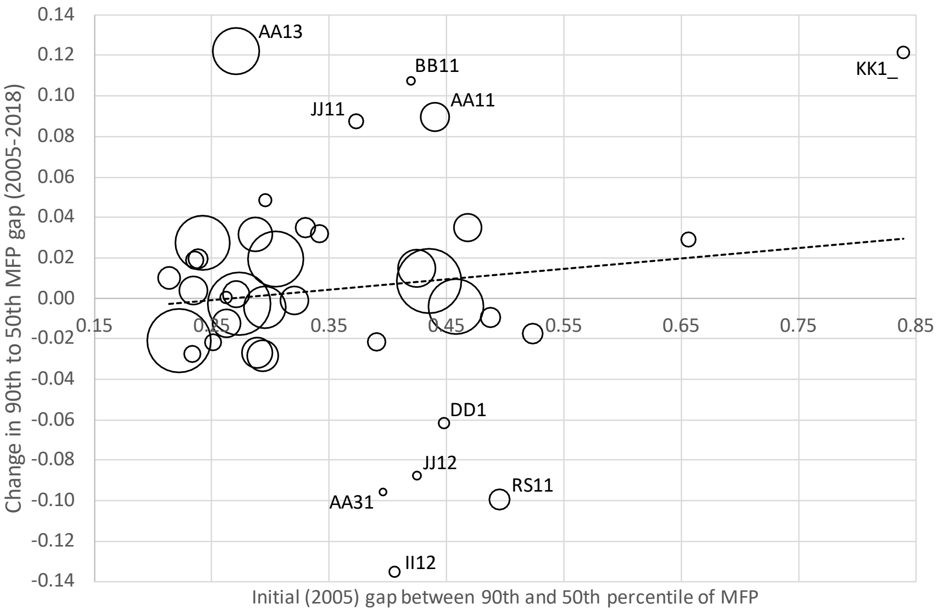
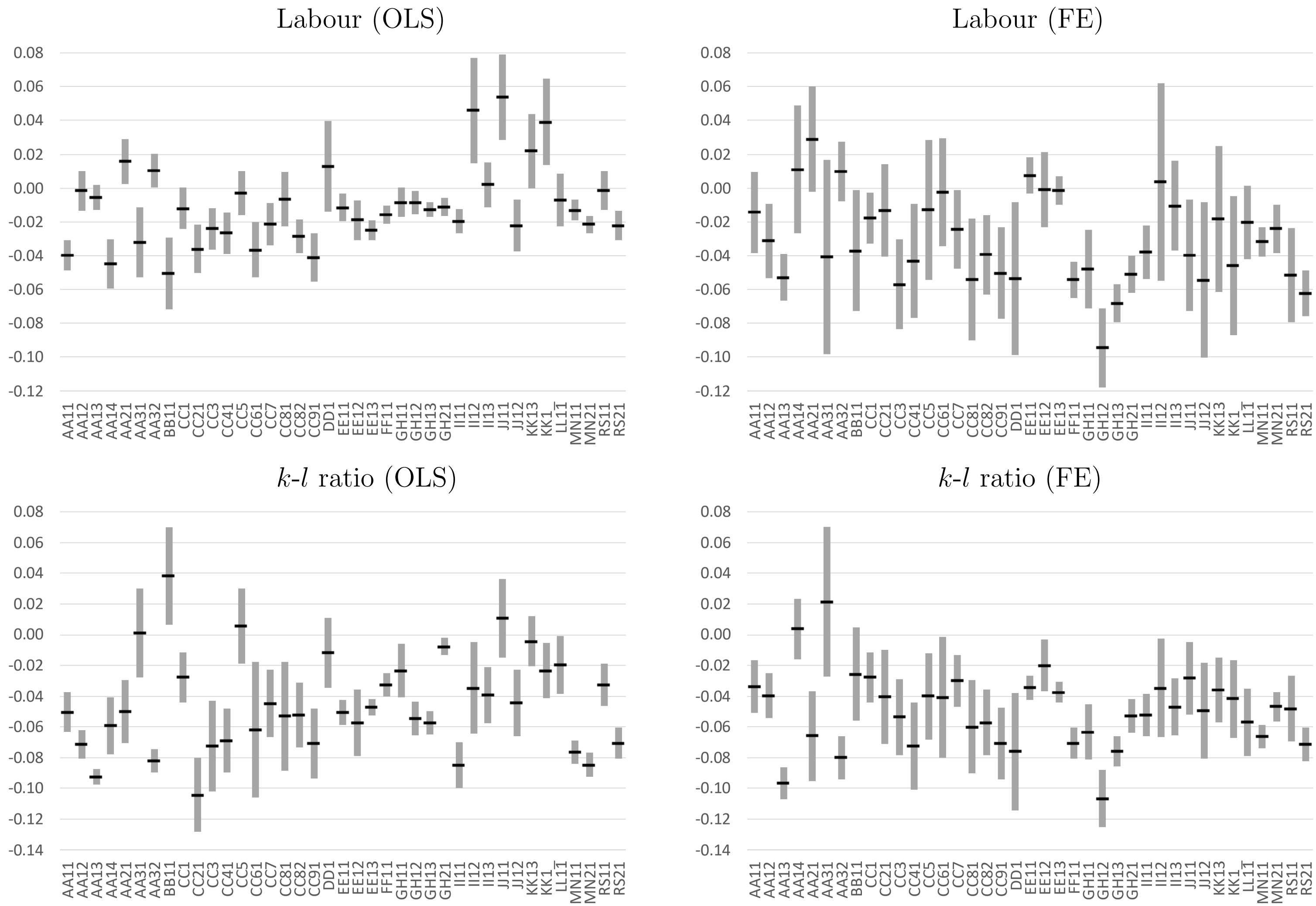

JJ11
JJ12
KK13
KK1_
LL11
MN11
MN21
RS11
RS21
1.4 1.2 1.0 0.8 0.6 0.4 0.2 0.0 -0.2 -0.4 -0.6 -0.8 -1.0 -1.2 -1.4
2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018

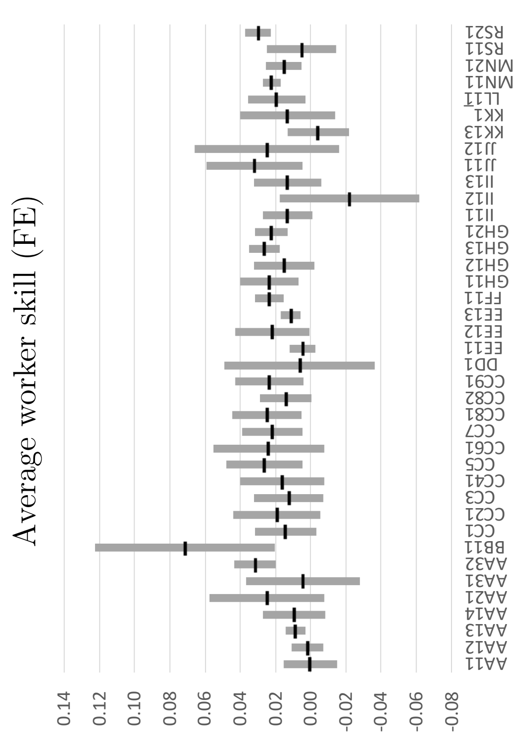
Figure A.2: Change in gap between 90th and 50th percentile of (translog) MFP (2005-2018) vs initial (2005) gap size

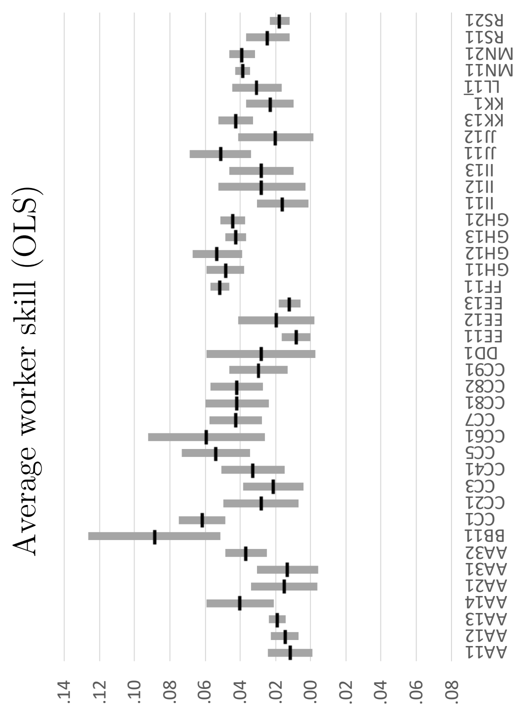


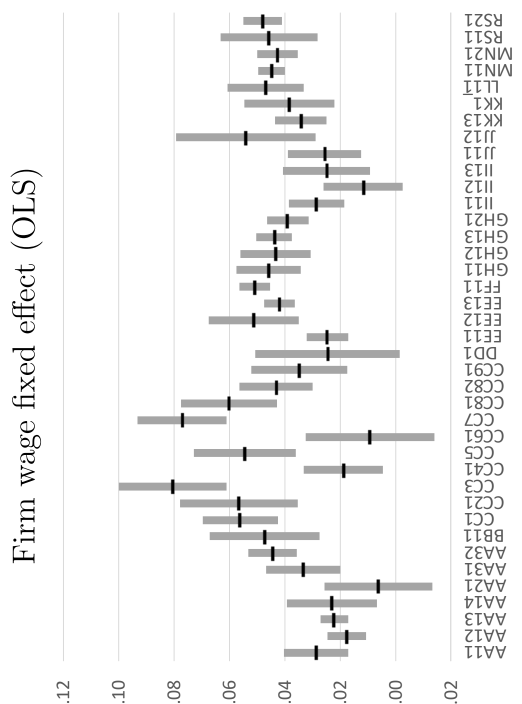
Based on percentiles reported in figure A.1. Top and bottom five industries based on change in gap are itemised (industry codes described in table A.1). Bubble size scaled to total number of firms in each productivity industry. Dashed line is (unweighted) linear fit between initial level and growth rate, with estimated slope of 0.052.

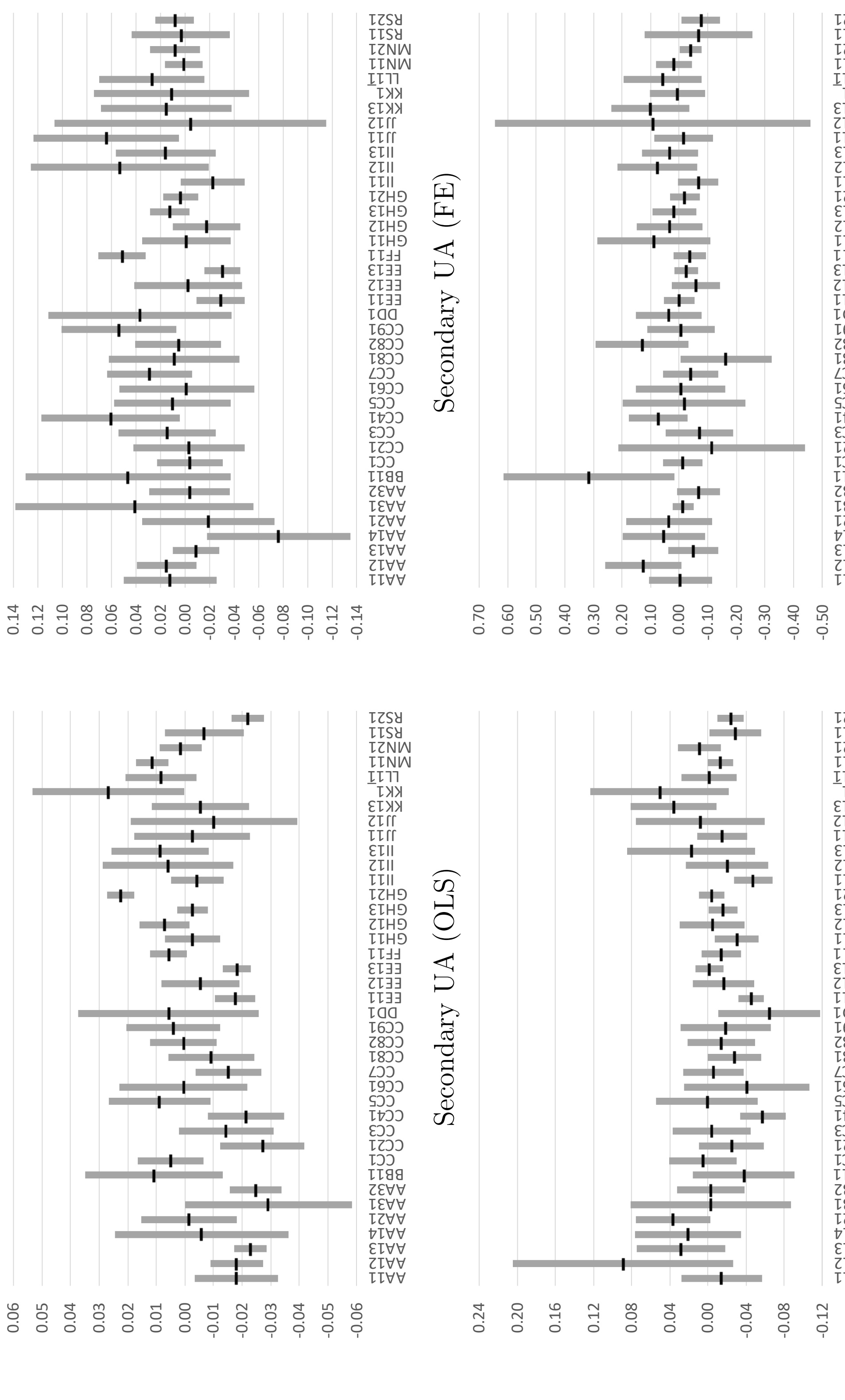
Figure A.3: Table 21 coefficients (and 95% confidence interval) by industry

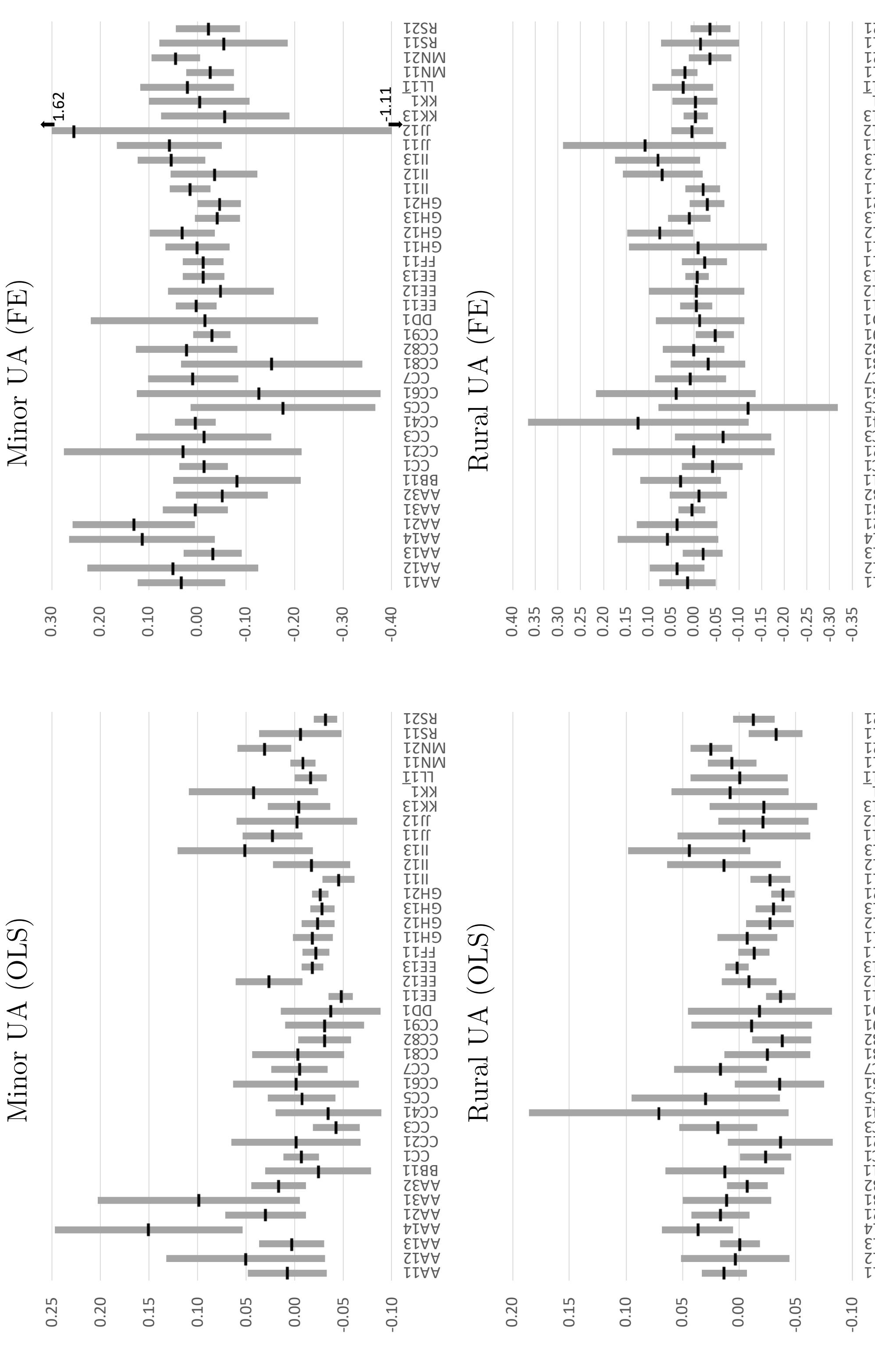



Average worker skill (FE)
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
-0.02
-0.04
-0.06
-0.08
AA11 AA12 AA13 AA14 AA21 AA31 AA32 BB11 CC1 CC21 CC3 CC41 CC5 CC61 CC7 CC81 CC82 CC91 DD1 EE11 EE12 EE13 FF11 GH11 GH12 GH13 GH21 II11 II12 II13 JJ11 JJ12 KK13 KK1 LL1 MN11 MN21 RS11 RS21


Average worker skill (OLS)
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
-0.02
-0.04
-0.06
-0.08
AA11 AA12 AA13 AA14 AA21 AA31 AA32 BB11 CC1 CC21 CC3 CC41 CC5 CC61 CC7 CC81 CC82 CC91 DD1 EE11 EE12 EE13 FF11 GH11 GH12 GH13 GH21 II11 II12 II13 JJ11 JJ12 KK13 KK1 LL1 MN11 MN21 RS11 RS21


Firm wage fixed effect (OLS)
0.12
0.10
0.08
0.06
0.04
0.02
0.00
-0.02
AA11 AA12 AA13 AA14 AA21 AA31 AA32 BB11 CC1 CC21 CC3 CC41 CC5 CC61 CC7 CC81 CC82 CC91 DD1 EE11 EE12 EE13 FF11 GH11 GH12 GH13 GH21 II11 II12 II13 JJ11 JJ12 KK13 KK1 LL1 MN11 MN21 RS11 RS21


ln(firm age) (OLS)
ln(firm age) (FE)
Secondary UA (OLS)
Secondary UA (FE)


Minor UA (OLS)
Minor UA (FE)
Rural UA (OLS)
Rural UA (FE)
1.62
-1.11

Motu
economic & public policy research
for other Motu working papers: www.motu.nz