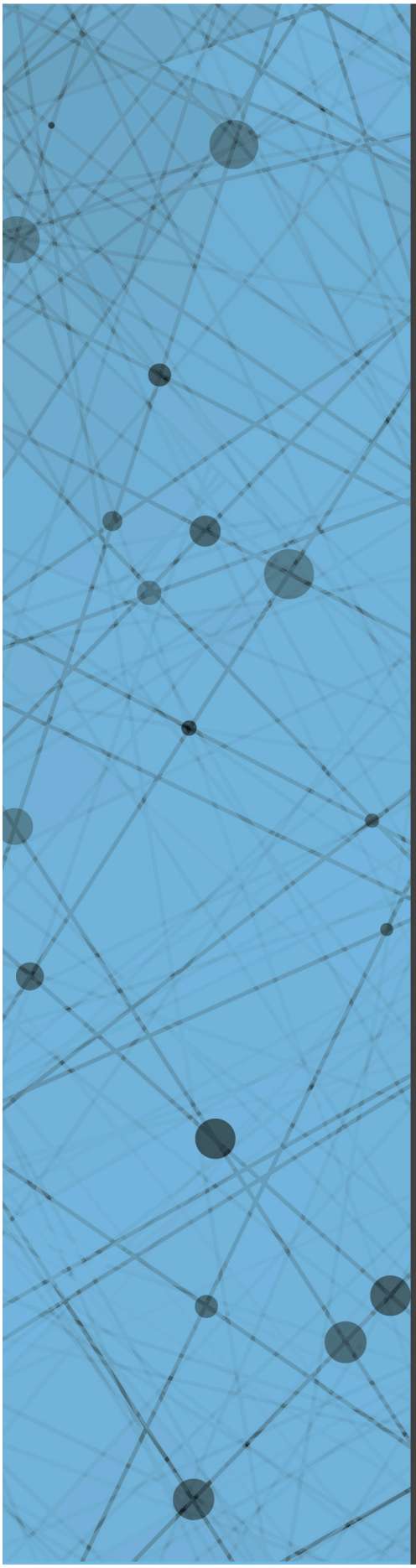
Motu



Entrepreneurial beginnings: Transitions to self-employment and the creation of jobs

Richard Fabling

Motu Working Paper 18-12

Motu Economic and Public Policy Research

October 2018

Document information

Author contact details
Richard Fabling
Independent researcher
richard.fabling@xtra.co.nz

Acknowledgements
I gratefully acknowledge funding and anonymous feedback from the Ministry of Business, Innovation & Employment (MBIE), and thank Stats NZ for supplying and enabling access to the data.

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, MBIE or Motu.
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.

Motu Economic and Public Policy Research
PO Box 24390 info@motu.org.nz +64 4 9394250
Wellington www.motu.org.nz
New Zealand

 ISSN 1176-2667 (Print), ISSN 1177-9047 (Online).

Abstract

Owner-operated firms are an important part of the New Zealand economy. They employ approximately 30% of the private-for-profit workforce, as well as providing jobs and income to the working proprietors themselves. This paper addresses two questions: what characteristics are associated with entrepreneurship (starting a self-employed business); and which sorts of entrepreneurs are more successful (create jobs)? We pay particular attention to differences in start-up and survival rates by business owner sex and ethnicity, but also consider whether other individual characteristics (including age and skill) and prior job characteristics also relate to the decision to start a business or to create jobs. We find substantial negative gaps in entrepreneurship for females and non-European-only ethnicity groups – gaps that arise in large part because of differential rates of entry into self-employment and, in the case of non-European-only ethnicities, higher attrition rates from self-employment after entry. These gaps persist in the presence of controls for skill, prior labour market experience and other individual characteristics.

JEL codes

J23; L26; M13

Keywords

Entrepreneurship; self-employment; job creation; survival; ethnicity; sex; Integrated Data Infrastructure (IDI)

Summary haiku

Entrepreneurship
creates jobs, but few firms grow
Who survives and thrives?

Executive summary

- The self-employed constitute a significant proportion of the labour force, and create a substantial number of jobs for their employees
- However, the majority of working proprietor (WP) firms never take on employees, which may reflect the preferences of the WPs, or that such transitions represent an important "step-change" in operations
- Taking into account the inherent risk in establishing and growing innovative start-up businesses, the data suggests the entrepreneurial spirit is alive and well in New Zealand
- There has, however, been an absolute decline in self-employment over the last decade with WP labour input falling from 28.6 percent to 21 percent of full-time equivalent labour input from 2005 to 2015
- Self-employment rates vary substantially by sex and ethnicity, with Pasifika-only and Māori-only ethnicity groups having a 9.4 percentage points (pp) and 8.1pp, respectively, lower probability of being self-employed than European-only. These differences are substantial when compared to the overall self-employment rate of 7.5%
- While partially explainable by differences in individual characteristics, such as age and migrant status, entrepreneurship gaps persist to some extent for all ethnicity groups relative to European-only, and for females relative to males
- For example, the entrepreneurship gap for females represents 48% of the average WP rate after controlling for individual characteristics
- These gaps arise in large part because of differential rates of entry into self-employment and, in the case of non-European-only ethnicities, higher attrition rates from self-employment after entry
- Controlling for both individual characteristics and prior labour market outcomes, the gap in the WP entry rate for Pasifika-only individuals is -75% of the mean entry rate, and the five-year survival rate gap after entry is -36% of the mean survival rate. For Māori-only ethnicity individuals, the corresponding entry and survival rate gaps are -54% and -18%

- The international literature suggests that differences in access to financial capital and specific business human capital (from, eg, parents or peers) may go some way towards explaining the residual ethnicity gaps in entrepreneurship
- Consistent with the observed ethnicity gaps, NZ-born individuals are more likely to be self-employed than immigrants (in contrast to US findings)
- Individuals with better prior labour market outcomes (higher earnings; better employers; no benefit receipt), and with formal qualifications, are also more likely to become self-employed
- However, individual skill and employee job creation are negatively correlated, consistent with high-skilled individuals electing self-employment simply as a preferred way of supplying their own labour services to the market
- Recent short-term negative labour market outcomes also appear to raise the probability of becoming a WP, suggesting the "necessity" channel to self-employment is relevant for some individuals. In turn, this may result in short self-employment spells as individuals return to jobs as employees

Contents

1 Motivation

Owner-operated businesses are an important part of the New Zealand economy. They employ a substantial proportion of the workforce, as well as providing jobs and income to the working proprietors (WPs) themselves. Table 1 shows this labour market contribution for both the whole economy, and restricted to economically significant private-for-profit firms.[1] Focussing on the private sector (bottom panel), WP firms account for between 29 and 33 percent of total full-time equivalent (FTE) employment of employees (column 4), while WPs themselves contribute an additional 21 to 29 percent to total labour input (column 5) when counted as equivalent to one FTE each.

While the aggregate contribution of WP firms to employment is substantial, the average WP firm is very small. The majority of WP firms never take on employees, which may partly reflect that such transitions represent an important "step-change" from management, risk and regulatory compliance perspectives. Additionally, some WPs may not wish to grow their business, instead seeing self-employment as an alternative mode of supplying labour to the market, which may yield non-pecuniary benefits such as flexibility over hours, autonomy, or the avoidance of management (eg, Blanchflower and Oswald 1998; Blanchflower 2004; Hurst and Pugsley 2011). Indeed, the self-employed receive lower financial returns than might be expected, consistent with material non-pecuniary benefits (Hamilton 2000).

Blanchflower et al. (2001) use International Social Survey Programme (ISSP) data to show that there is substantial unrealised demand for self-employment in many countries, including New Zealand. Table 2 extends their estimates for NZ to 2005 – the latest year where ISSP data is available on this topic. The top row of the table shows the self-employment rate preferred by individuals, which is roughly three times the actual self-employment rate, because a large proportion of employees state that they would rather be self-employed. While some self-employed would prefer to be employees, this desire for change is not as prevalent as it is for employees (bottom two rows of table 2).

Setting the evidence on preferences aside, the transitory nature of many self-employment spells suggests that a significant proportion of WPs: work in industries where the practical distinction between self-employment and employee is limited (see, eg, Hurst and Pugsley 2011); use self-employment as a

[1]The labour dataset from which these statistics are derived is described fully in Fabling and Maré (2015a) and summarised in section 2.

stop-gap between jobs; determine ex-post that having tried self-employment it is not for them; or establish that they are not well equipped to run a business despite a desire to do so. In some cases, also, exit from self-employment may reflect the on-selling of a successful business venture. Some WPs continue to be employees in other businesses, so that self-employment is potentially a source of supplemental income. Alternatively, the employee job may be maintained as a way of funding a new business idea, or retained as insurance against that idea failing (Garcia-Perez et al. 2013).

Taking into account the inherent risk in establishing and growing innovative start-up businesses, there is much in the data to suggest the entrepreneurial spirit is alive and well in New Zealand. The data shows, however, an absolute decline in self-employment over the last decade so that WP labour input falls from 28.6 percent to 21 percent of FTE labour input from 2005 to 2015 (table 1, column 5). While the ISSP data covers an earlier period, those statistics additionally imply a declining desire to be self-employed as well as a declining actual self-employment rate (table 2).

The trend decline in self-employment runs counter to the rise in "alternative work arrangements" documented in the US where independent contractors have been increasing as an employment group (Katz and Krueger 2016), though the US has also seen a decline in business dynamism (Decker et al. 2016). It is also counter to earlier trends in New Zealand where the rate of self-employment had been increasing over the three decades from 1966 to 1996, in contrast to most other OECD economies (Blanchflower 2000).[2] It is unclear, however, whether these trends are an issue for aggregate economic performance.[3] For example, from a labour market perspective, the total FTE jobs created by the self-employed has not declined – at least in the private-for-profit sector – despite the dip and recovery in total FTE, evident in table 1, following the Global Financial Crisis (GFC).[4]

[2]Recent OECD-based statistics for NZ show a decline in the self-employment rate over the period we analyse (see, eg, Blanchflower 2015).

[3]Stats NZ's Business Demography statistics (downloaded from NZ.Stat) suggest that a significant proportion of the decline in WP numbers is due to a decline in the total number of businesses in the agriculture, forestry & fishing (AFF) sector. From 2005 to 2015, the aggregate number of businesses in that sector fell by 10,488, but the total number of sole proprietor and partnership businesses fell by over 18,000 (offset by increases in other business types). Consistent with these aggregates, we find that 40% of the decline in the number of WPs over the study period is due to a loss of self-employed in the AFF sector.

[4]As column (4) of table 1 shows, the share of FTE employment in WP firms fell over the same period because of a substantial increase in total FTE employment. This increase in total FTE is, presumably, mechanically linked to the decline in WPs, since many individuals leaving self-employment will take jobs as employees.

Given the importance of self-employed business ownership in the NZ labour market, it is perhaps surprising that there is very little research on the characteristics and performance of the self-employed and their businesses.[5] We fill this gap by answering two descriptive questions: what characteristics are associated with entrepreneurship (starting a self-employed business);[6] and which sorts of entrepreneurs are more successful? The success metric we focus on is employing because of the potential long-run impact on aggregate jobs, and to avoid measurement issues with self-employed financial performance metrics (Fabling and Sanderson 2014). We also examine the dynamics of WP firm survival, since attrition from self-employment can exert a tangible effect on the composition and size of the aggregate WP stock.

There is a substantial international literature that canvases these issues. In the US, for example, one strand of research has focussed on differences in business start-up rates by ethnicity and migrant status, finding that: immigrants have relatively high self-employment rates (eg, Borjas 1986; Lofstrom 2002; Fairlie and Lofstrom 2015; Kerr and Kerr 2016); and African-American (Asian) men are less (more) likely to operate self-employed businesses than white Americans (Fairlie and Meyer 1996; Fairlie 1999; Fairlie 2007). Higher exit rates from self-employment for African-Americans are part of the explanations for gaps in the overall business ownership rate (Fairlie 1999; Ahn 2011).

While systematic differences in characteristics such as age and education partially account for these gaps, other factors such as access to financial capital and specific business human capital (perhaps attained from peers, siblings or parents) also seem to be important (Blanchflower and Oswald 1998; Blanchflower et al. 2003; Fairlie and Robb 2007; Bates 2011). Indeed for some ethnic groups in the US, such as Mexican-Americans, lower educational attainment and wealth completely explain the lower business formation rate, relative to non-Latino white Americans (Fairlie and Woodruff 2010).

More generally, across OECD countries self-employment is consistently more prevalent among men and older individual (Blanchflower 2004), and is more likely to be successful for higher skilled individuals, including those

[5] See, for example, Yuan et al. (2013) which summarises the NZ empirical literature on migrant entrepreneurship.

[6] We use the terms self-employed (or WP) and entrepreneur interchangeably. The data does not enable a clear distinction between the two concepts except in an unsatisfactorally biased ex-post (ie, success-based) manner. Much of the literature, particularly that relying on administrative data, uses the minimal standard of business ownership as the operational definition of entrepreneurship and we follow that approach.

with broader experience both academically and in the workforce (Lazear 2004; Lazear 2005; van der Sluis et al. 2008). For individuals close to retirement age, joining the ranks of the self-employed may enable a smoother transition out of the workforce (eg, Ramnath et al. 2017). While the Great Recession had a substantial impact across the OECD on the closure of many self-employed businesses, the number of transitions to self-employment increased in some countries, potentially linked to reduced opportunities in the (employee) labour market (Fairlie 2013; Blanchflower 2015).

We follow the broad themes of this entrepreneurship literature, paying particular attention to differences in start-up and success rates by business owner sex and ethnicity, but also consider whether other individual characteristics (including age and skill) and prior job characteristics (including earnings and industry) also influence the decision to start a business or to create jobs.[7] Our study is closest in nature to Fairlie and Miranda (2016) who use a comprehensive integrated self-employed/employer business register in the US (the ILBD, Davis et al. 2009) together with survey data on WP characteristics. We make use of similar data in New Zealand (the Longitudinal Business Database, LBD), but have the advantage of access to business owner characteristics and linked employer-employee data (LEED) from within the Integrated Data Infrastructure (IDI), enabling an entirely population-based analysis of self-employment.

We find substantial negative entrepreneurship gaps for both females (relative to males) and non-European-only ethnicities (relative to European-only).[8] These gaps arise in large part because of differential rates of entry into self-employment and, in the case of non-European-only ethnicities, higher attrition rates from self-employment after entry. While observable characteristics go a significant way towards explaining the size of some entrepreneurship gaps, large (relative to average participation rates) unexplained differences remain.

In relation to observables characteristics, we find that: NZ-born individuals are more likely to be self-employed than immigrants (in contrast to US findings, but consistent with the observed ethnicity gaps); individu-

[7] We look, eg, at the prevalence of Māori business owners, as opposed to the prevalence of "Māori businesses." Partly this is because we focus on business owners, rather than the businesses themselves, as discussed in section 2. More fundamentally, though, there are important conceptual differences between Māori business ownership and Māori business, and we generally only have data sufficient to identify the former.

[8] In this context, an "entrepreneurship gap" is the difference – between two groups of individuals – in the probability of being self-employed (or entering self-employment or employing).

als with formal qualifications are more likely to be entrepreneurial, though bachelor and higher-qualified WPs are significantly less likely to be employers than less qualified WPs; individuals with young dependent children are more likely to be self-employed and employ others, controlling for the fact that self-employment is positively correlated with age. In relation to labour market variables, we find consistent evidence that individuals with better prior labour market outcomes (higher earnings; better employers; no benefit receipt) are also more likely to become self-employed. In addition, recent short-term negative labour market outcomes also appear to raise the probability of becoming a WP, suggesting the "necessity" channel to self-employment is relevant for some individuals.

Section 2 outlines the empirical method and the data used. Findings are discussed in detail in section 3, and summarised in section 4. Finally, section 5 outlines the potential of the dataset to enable future research.

2 Data and method

The empirical analysis is descriptive in nature, seeking to establish correlations between individual characteristics, self-employment, and employing. We explore these relationships graphically and using ordinary least squares (OLS) regressions. We start by looking at the characteristics of WPs in the *stock* of non-employer and employing WP firms (ie, the correlates of the overall WP rate). These simple cross-sectional statistics set the scene for investigating the dynamics of entry into self-employment, survival in business and the creation of jobs. In this latter analysis, we restrict attention to individuals who haven't been working proprietors for at least five years and look at the relationship between entry decisions, individual characteristics and prior labour market outcomes.

This two-part analysis necessitates four samples/populations as shown in table 3. For self-employment participation decisions (columns 1 and 3) we analyse a 10% random sample of the full population, selected at the individual level and weighted to account for sampling. For analysis of working proprietors (columns 2 and 4) we use data on the full population of interest. The first two datasets are used to analyse the overall WP rate and, therefore, include all individuals (column 1) or all working proprietors (column 2) in the Estimated Resident Population (ERP), as defined below. In contrast, the remaining datasets (columns 3 and 4) are used to analyse entry into self-employment and are restricted to individuals who haven't been working

proprietors for at least five years.

The analysis relies on Statistics New Zealand's Integrated Data Infrastructure (IDI) and exploits the linking of full-coverage firm- and worker-level administrative datasets with data from the 2013 Population Census. Individual-level characteristics (age, sex, and ethnicity), and monthly jobs, benefit receipt, and border movement data come from the October 2016 instance of the IDI and are discussed below by topic. These data are linked through to firms on the Longitudinal Business Database (LBD) using linked employer-employee data (LEED) and the permanent enterprise number.

In most cases, we include indicator variables for missing data rather than imputing data or dropping individuals. The exceptions to this rule are sex and birth date, where we drop individuals without these characteristics.[9] Table 3 summarises the "permanent" individual characteristics that we use, and the associated rate of missing data for each variable and sample.[10] We include information on ethnicity, NZ-born/migrant, English language skills, highest qualification, and an earnings-based measure of skill. In subsequent regressions, these permanent characteristics are supplemented with time-varying information on the number of young dependent children (also in table 3),[11] absences from NZ, and prior job and benefit histories.

The period of analysis is the eleven years from 2005-2015, so that we have at least five years of prior job and benefit history available in each analysis year.[12] When we analyse future outcomes for WP entrants (population in column 4 of table 3) we restrict the time period for entry to the six years from 2005 to 2010 so that each entering WP cohort has a minimum of five years of post-entry data.

[9] We do this because missingness may indicate that the entity associated with the id may not actually be a person. Further, for actual individuals, the absence of these variables may be indicative of lower quality data linking across datasets in the IDI. We also impose an age cut-off for the population, which is less clearly defined if we include "individuals" with missing birth date. Age (birth date) and sex come from the personal details table.

[10] Some of these characteristics are not truly permanent, eg highest qualification, but are treated as such since we only observe them once in the data, and we expect them to be unchanging for a significant proportion of the population.

[11] Because the data on dependent children come exclusively from Census 2013, consistent backcasting to 2005 restricts the analysis to children aged 8 or less (see section 2.6).

[12] Starting the analysis in 2005 avoids issues prior to this with higher rates of missing data for birth date, sex and ethnicity. The method for deriving the Estimated Resident Population also deteriorates prior to 2005 – compared to official statistics and the more consistent patterns observed in latter years – which is an artefact of the method's reliance on data that is left-censored.

2.1 Estimated Resident Population (ERP)

An Estimated Resident Population (ERP) is necessary to provide a population from which potential working proprietors might be drawn. We derive the ERP by starting with the IDI "spine," which attempts to replicate "an 'ever-resident' population" (Black 2016, pg9). To translate this ever-resident measure to actual residence in any particular March year, we first use education, tax and border movement data to identify the subset of "spine" individuals that are ever present in New Zealand during the reference period. Then, given that we observe each individual in NZ at some point and we observe all border movements over this period, we can use border and date of death information to identify the subset of individuals that are present (and alive) in NZ in any particular year.

The population of interest, in a given year, is restricted to individuals who were 17-74 years of age in the prior March (ie, turning 18-75 during the year of interest), reflecting a desire to exclude the school-aged, and to include potential transitions to self-employment up to ten years after the age of eligibility for national superannuation (65). Aside from focussing on the subset of ages where labour market activity and self-employment are most likely, the imposed age limits remove observations with dubious birth date information (eg, "individuals" employed before birth).

Appendix figure A.1 compares the study ERP to Stats NZ's official ERP in 2005, 2010 and 2015. Because the official ERP is measured as at March, rather than over the full year, the study ERP is higher than the official ERP at all ages. This overestimation is particularly pronounced for 20-35 year olds and is a feature of Stats NZ's own attempts at an admin-data ERP, which involve more identification rules (datasets) and which have a primary goal of replicating the official point-in-time ERP (Gibb et al. 2016; Statistics New Zealand 2016).[13]

Table 4 (column 2) shows the annual effect on population size of requiring presence in the ERP (ie, alive and in NZ during the year), and imposing age restrictions and non-missing age and sex. Relative to the private-for-profit population of table 1 (bottom panel), we lose 2.1-2.7 percent of working proprietors, and 3.1-4.7 percent of total FTE employment in WP firms. On average, the loss of working proprietors comes predominantly from the

[13]We deviate from these previous methods because they rely on additional data to derive their ERPs (eg, health data), which we do not use in this research. They also produce point-in-time ERP measures – which isn't suitable for the current analysis – relying on observed *activity* proximate to that point-in-time reference date.

upper-bound restriction on age (1.2pp, percentage points), followed by missing age/sex (0.5pp) and individuals overseas for the entire year (0.4pp). In contrast, the largest loss of associated FTE employment comes from missing age/sex (2.3pp of the average 3.9pp). The larger average employment size of dropped firms is consistent with the hypothesis that missing basic demographic data is associated with non-person business ownership.

2.2 Self-employment and business characteristics

As Acs et al. (2008) demonstrate, the choice of data source can exert quite a substantial influence over estimated entrepreneurship rates. We rely on self-employment information drawn directly from individual, partnership and company tax filings, and impose conditions designed to limit the population of interest to individuals who have an ownership stake in a business and provide labour input to that business (ie, *working* proprietors). In addition, we require associated business(es) to meet an economic significance test that should ensure presence on Stats NZ's Business Register (BR). Given the low thresholds for mandatory GST registration and tax filing, and the low cost of business registration in New Zealand, undercoverage of within-scope working proprietors is likely to be limited.

These rules will, however, exclude individuals who are "entrepreneurial" but are – at least temporarily – not involved in an economically significant private-for-profit business (eg, social entrepreneurs in the not-for-profit sector).[14] Conversely, the analysis will include individuals who are not "entrepreneurial" under some definitions of the term – eg, those self-employed who have no desire to grow their business. We account for this latter definitional issue through the interpretation of the findings.

Both the working proprietor and PAYE-based earnings (jobs) data are discussed in detail in Fabling and Maré (2015a), and we make use of their derived FTE measure of labour input and their method for identifying working proprietors and for removing them from the PAYE data. Briefly, WPs are identified from four tax sources:

1. *Sole proprietors paying themselves PAYE income,* defined as wage and salary earnings where the payer and payee IR numbers are the same on the EMS (Employer Monthly Schedule)

[14] Passive investors – business owners who do not supply labour inputs – are deliberately excluded from the analysis, where they can be identified.

2. *Sole proprietors receiving non-zero self-employment income* as reported in box 23 of the IR3 (Individual Tax Return)

3. *Partners receiving a share of total partnership income* as reported in box 25J of the IR7P (Partnership Income/Loss Distribution)

4. *Company owners receiving remuneration with no PAYE deducted* as reported in box 41C of the IR4S (Company Shareholders' Details)

Reported income levels are not a determining factor in the identification of WPs because this income is likely to include profit distribution (including any return on capital) and, therefore, may bear little relationship to labour input. An exception to this rule is applied in the case of companies, where a threshold of ever receiving $15,000 (real, 2000 dollars) in remuneration is applied to eliminate potential non-owners from being counted as WPs. Partners in partnerships are also excluded in years where their own individual tax return (IR3) indicates that they were passive investors in the partnership, rather than working proprietors. WPs who switch between receiving PAYE income and profit distributions are treated as WPs in all years that income is received. Fabling and Maré (2015a) explains the logic for these rules and the impact that they have on estimated WP counts.

The unit of observation is individual working proprietors, rather than the firms they operate, to avoid issues with business identifier continuity for micro enterprises. Specifically, for non-employing businesses, a change in business type from sole proprietor or partnership to limited liability company is likely to result in a change in enterprise number on the BR for the real-world business (Fabling 2011). Changes in legal form may be triggered by important milestones in the evolution of a business, such as the transition to employing or the addition of business owners, making it important that the analysis tracks activity across business type changes.

Where firm-level identifiers are necessary, eg to establish job starts and ends for employees, we use Fabling's (2011) method for repairing enterprise identifier breaks using employee-tracking. This method successfully repairs breaks in identifiers due to business type transitions, which are largely unobserved in the BR data in the absence of repaired enterprise links (Fabling 2011). However, the employee-tracking method does not cover firms with fewer than three employees because of the inherent difficulty in defining continuing business locations from small numbers of individuals. Instead, tracking the owner largely solves these continuity issues since the IDI captures mandatory WP filing of self-employment income across the relevant business types. The downside of following individuals is that we may char-

acterise some transitions out of self-employment as business "failure" when these are, in fact, the sale of a successful (ongoing) business by the owner.

The LBD provides information on businesses – namely industry, private-for-profit status, and annual Goods and Services Tax (GST) sales and purchases. GST data are deflated to 2000 real dollar values in order to impose a constant cut-off for economic significance at $30,000 in sales or purchases. Additionally, all firms with employees are classified as economically significant. As table 1 demonstrates (top panel vs bottom panel), the application of the economic significance threshold substantially affects the number of working proprietors in the population of interest – with 24-26 percent of all WPs being non-employers and having total income and expenditure below the $30,000 threshold.[15] The threshold is applied to achieve consistency with BR maintenance rules, which generally require businesses to reach the $30,000 threshold before an enterprise number and characteristics are added to the register. Applying the threshold as a constant real figure improves internal consistency of the analysis, but also allows for the fact that the nominal threshold for mandatory GST filing has increased over the observation period so that some new firms above the BR's $30,000 threshold may no longer register for GST and, therefore, not appear on the BR.[16]

2.3 Job and benefit history

The working proprietor outcomes of interest are business survival (ie, continuing economic significance) and employing staff. The latter information is sourced directly from PAYE (LEED) records within the IDI after being transformed using the methodology outlined in Fabling and Maré (2015a) to remove any self-employed who receive pay through the PAYE system from a firm that they themselves own.[17]

[15]The reported population loss includes the effect of the private-for-profit (with observed industry) criteria, though this only causes the loss of 0.4 percent of WP-years, and 1.6 percent of total FTE. By construction, the economic significance threshold does not cause any loss of WP firm FTE, nor the elimination of any firms from the LEED data.

[16]Fabling and Sanderson (2016) provide detail on the BR maintenance rules and the GST mandatory filing thresholds, as well as an overview of the other LBD data used.

[17]Some self-employed continue to have jobs in businesses they do not own. These jobs are retained in the LEED data when we estimate firm-level employment, and when we calculate individual job histories. When analysing WP outcomes, we do not include variables associated with concurrent wage and salary employment for two main reasons. Firstly, self-employment data is annual rather than monthly which can make it hard to determine whether self-employment is concurrent to employment or (within-year) sequential. Secondly, and related to this timing issue, individual WPs may gain or lose jobs because

This same data is used to create job histories for all individual in the ERP, where we construct an indicator variable for having a job in each of the five (March) years prior to the analysis year. In the same manner, we construct lagged indicator variables for receiving a main Government benefit, which is also identified through the PAYE data.

Finally the Fabling and Maré labour dataset is used to directly estimate average job churn, which is the excess turnover in jobs an individual has over and above that necessary to give effect to the observed net job change.[18] Job churn is treated as a measure of success in the labour market since individuals with lower values, conditional on having a job,[19] transition between jobs less often – perhaps because they find better average job matches when they switch jobs, or because they are not as exposed to the temporary job market.[20]

2.4 Skills – wage fixed effects and formal education

Also derived from the Fabling and Maré labour dataset, we use the two-way wage fixed effects estimates of Maré et al. (2017), interpreting the estimated worker fixed effects (WFE) from that model as a proxy for worker skill.[21] WPs and other individuals who never hold an employee job over the full 17 years of data (April 1999-March 2016) do not have an estimated worker fixed effect. As table 3 (bottom row) shows, worker fixed effects are absent for 33.8 percent of all self-employed. For this reason, the regression analysis

of their self-employment outcomes which would make it difficult to interpret regression coefficients. See, eg, Garcia-Perez et al. (2013) for discussion and analysis of the potential role concurrent job earnings have in the entrepreneurial process.

[18] Job churn is numerically equal to the minimum of the number of job starts and job ends in a year. For example, if an individual starts three jobs in a year, and ends one job in the same year, job churn is one. That is, there was one "unnecessary" job start-end pair from the perspective of achieving a net job change of two over the year. We average this measure over the five prior years and cap the resulting average at twelve, adding an indicator variable for the small number of capped observation.

[19] In regressions, the indicator variables for having a job provide the necessary controls for this conditional interpretation to be valid.

[20] Job spell length would be a good additional covariate, but is subject to left-censoring for a large proportion of individuals in earlier years.

[21] The worker fixed effect can be thought of as a portable (permanent) wage premium that a worker gains, regardless of who their current employer is. The fixed effects model includes controls for year, and a quartic in worker age by sex. By construction, the WFE is mean zero by sex on an FTE-weighted basis. In this paper, we renormalise the worker fixed effect so that it is mean zero in the 10% ERP sample by sex.

relating to all self-employed does not include the WFE, since persistent non-employment may reflect reverse causation from always being a WP to never being an employee.

This issue is less severe when we consider entry into self-employment, because the five year exclusion period substantially lowers the probability of never having had a job (table 3, column 4).[22] In addition, the restriction to potential entrants results in a greater proportion of job histories pre-dating any WP experience. Given that WFEs are estimated over the full 17 years of data, and WP experience may affect subsequent job wages (positively or negatively), having the majority of jobs data preceding (rather than post-dating) WP entry makes it easier to interpret coefficients on WFE variables in regressions.

The two-way fixed effect estimates also allow us to establish whether the employer wage premium (the firm fixed effect) in prior jobs is related to becoming a WP. At the individual level, we FTE-weight these data over all jobs held during a year, and we include a separate variable for each of the five years leading up to the analysis year. The firm fixed effect is set to zero in years where the individual does not have a job, which is controlled for by the job indicator variables in the case of transitory absence from the jobs data, and the missing WFE indicator variable for individuals who never appear in the employee data.

The skill (WFE) data is supplemented with highest qualification and conversational English language ability indicator variables, derived from Census 2013 data. Because there is only a single Census integrated into the IDI, over a third of individuals in the ERP have missing data and the absence of this data is correlated with other variables that relate to the likelihood that the individual was in NZ in 2013, particularly migration status. This is a key reason for including both qualification and WFE measures in the analysis.[23]

[22] The likelihood of never having a job in the full sample (column 1) and the potential entrant sample (column 3) is similar, with the latter being lower because of the exclusion of WPs with a higher probability of never being employees.

[23] Administrative education data in the IDI could be used to populate or update the formal qualification variable for some individuals. We have not attempted this since it is unlikely to generate a convincing *highest* qualification measure for people with missing data, particularly adult migrants to New Zealand.

2.5 Ethnicity, overseas-born and absences from NZ

Ethnicity data come from the IDI's source-ranked ethnicity table, which provides level 1 ethnicities. Stats NZ's preferred source for populating this table is Census 2013. However, they improve the coverage of the table by adding administrative data on ethnicity, which is prioritised (source-ranked) based on the consistency of the administrative source with Census ethnicity.[24] Adding administrative ethnicity data provides coverage for almost 94 percent of the ERP and over 98 percent of the WP population (table 3, columns 1 and 2). The higher missing rate in the former is primarily due to lower coverage of ethnicity data for young individuals, who are under-represented in the self-employed population, relative to the ERP.

Most agencies allow individuals to report multiple level one ethnicities and we retain permutations of multiple ethnicity that have more than 10,000 individuals in the ERP. We combine the level one "Other" ethnicity with European because most (97%) of Census responses in the "Other" category are individuals who have identified "New Zealander" as their ethnicity, and because administrative collections do not necessarily distinguish between "New Zealander" and "European." These choices yield ten mutually exclusive groupings as listed in table 3, consisting of five individual level one grouping (the "-only" groups), the four two-way interactions with European, and a "residual" category that includes all other multi-ethnicity groupings.

Appendix figure A.2 compares the representation of each grouping (in 2015) depending on whether the data source is Census 2013 or source-ranked administrative data. For presentation purposes, the figure excludes the European-only group (the largest sub-population). Because absence from the Census data is partly determined by migration, we might expect some non-European-only rates in Census to be lower than the equivalent administrative data rates. This is apparent for Pasifika, Asian, and Middle Eastern, Latin American and African (MELAA) ethnicity groups and the residual category. However, the overestimation – relative to Census – of Māori-only in administrative data, together with the scale of the differences for the other non-European ethnicities, suggests systematic differences between the administrative data sources and Census, which could be due to collection method or the context in which the data is supplied. In this paper, we assume that both the Census and source-ranked administrative data are adequate identi-

[24] After Census 2013, the prioritised source ranking is: Department of Internal Affairs; Ministry of Health; Ministry of Education; Accident Compensation Corporation; and Ministry of Social Development. These five agencies' plus Census data account for almost all of the ethnicity information in the ERP.

fiers of ethnicity, noting that the majority of these data come from Census. As with other data, we retain individuals with missing ethnicity to avoid any potential bias from, eg, excluding individuals who have not interacted with any of the agencies collecting ethnicity data.

Census 2013 also provides information on whether individuals are New Zealand-born or the year of their first arrival in New Zealand. We divide the ERP overseas-born into approximately equal groups based on first arrival year (groupings shown in table 3). Border movement data are also used to construct annual indicator variables for complete absence from New Zealand in each of the five prior years. As with other lagged five-year data, these latter variables are only included in WP entry regressions. They are defined as full year absence for two reasons. Firstly, defining them in this manner means that their inclusion fully removes the effect of absence from NZ on the relevant coefficients for job and benefit (non-)receipt (ie, non-presence in NZ determining job and benefit status). Secondly, we want to test whether offshore experience is important to the decision to become self-employed, over and above any effect of being born overseas. The accumulation of such experience may requires a substantial exposure to foreign ideas, markets or culture to become evident.

2.6 Dependent children

Finally, we make use of Census 2013 to identify the number of dependent children each individual has. Census provides data on dependants less than 18 years of age as at 5 March 2013. Consistent with how we link other time-varying characteristics such as age (ie, as at the prior March), these Census responses directly provide the dependant count for the 2014 analysis year. We then project these counts back to 2005, restricting to dependants up to age 8 (ie, age 17 less 9 analysis years) in order to have a consistent upper age limit over time.[25] We divide this data into pre-school (0-4yr) and school age (5-8yrs), creating create five distinct count groups as shown in table 3. Finally, projecting these groupings forward to 2015 is problematic (because of births) and is only possible for a subset of individuals, resulting in a higher missing data rate compared to other Census-based variables (except in the WP entrant population, which is restricted to entry up to 2010).

[25]Projecting back relies on assuming stable family and household structure.

3 Results

In this section, we summarise the characteristics of the self-employed and employers, firstly examining the overall WP rate and then, secondly, transitions into self-employment, survival and job creation. In each sub-section, we begin with univariate statistics for key characteristics before undertaking multivariate analysis using OLS regressions. A key question we wish to answer with the multivariate analysis is how much of the observed gaps in entrepreneurship rates by sex and ethnicity are due to underlying differences in other observable characteristics, such as education, age, labour market outcomes and migration.

3.1 Entrepreneur characteristics – descriptives

Table 3 provides a first hint at the scale of these entrepreneurship gaps – for example, while females and males each make up approximately half of the ERP and the potential WP entrant population (columns 1 and 3), only 37.5 percent of self-employed are female (column 2) and only 41.2 percent of individuals who transition to self-employment are female (column 4).

Figure 1 shows – using local polynomial regressions reported with 95% confidence intervals – how these propensities vary by sex, age and skill (WFE). Panel A shows that the male self-employment rate is proportionately much higher for older individuals, particularly in the ten years leading up to the official retirement age of 65. There is no part of the age distribution where females are significantly more likely to be entrepreneurial than males. In contrast, Panel C shows that low-skilled (WFE) females are more likely to be self-employed than low-skilled males, but that moderately-skilled and high-skilled (WFE) males are more likely to be self-employed than similarly-skilled females, with the distributions crossing at approximately the 25th percentile of the WFE.

These U-shaped distributions are suggestive of multiple motivations for entrepreneurial behaviour related to the ability of individuals to be well-paid in the labour market. Differences in self-employment probabilities are significant (relative to the mean WP rate) between low-, moderately- and high-skilled individuals. Specifically, going from the 10th percentile to the 50th percentile of the WFE distribution is associated with a 4pp (1pp) decline in the self-employment rate for females (males), while going from the 50th to the 90th percentile is associated with a 2pp (5pp) increase in the self-

employment rate for females (males).[26]

Panels A and C of figure 2 show the (smoothed) distribution of individuals in the ERP which, for age, is very similar by sex, and which is somewhat more left-skewed in skill for females. Panels B and D of figure 2 show the effect of the differential rates of entrepreneurship by age and skill, respectively, on the distribution of age and skill in the working proprietor population (ie, the column 2 population in table 3).[27]

In addition to the gap in self-employment, we are also concerned with identifying gaps in the likelihood of employing, conditional on being self-employed.[28] Being an employer of substance is a relatively rare event, with the majority of WPs employing no or very few staff. Table 4 summarises the size distribution of owner-operated firms, grouping these firms into six annual average full-time equivalent employment size groups (columns 3-8), including those who are non-employers.[29] The top panel of the table shows proportions of WPs by size category, while the bottom panel shows the share of total FTE employment in WP firms. Approximately half of all WPs operate firms with no employees, and this proportion has been increasing over time (column 3, top panel). The majority of remaining WPs (33-38%) have at most two FTE employees, with a further 9-10% having up to five FTE employees. Larger employment sizes account for the remaining 5-6% of WPs.

As expected though, larger employers are important contributors to total employment in WP firms, with the over-twenty FTE group accounting for an increasing share of total FTEs rising from less than 20 percent in 2005 to over 27 percent in 2015 (bottom panel, table 4). Over the eleven years, total FTE employment in owner-operated firms is static (column 1), so that the increasing share of large employers in total employment is largely at the expense of WPs with two or less employees, consistent with the overall decline

[26]The 10th, 50th and 90th percentiles of the female (male) WFE distribution are, respectively, -0.33 (-0.43), -0.06 (-0.03), and 0.41 (0.44).

[27]These densities are within-sex and not rescaled to reflect cross-group differences in average self-employment, so that the higher relative density for females in the 35-50 age group doesn't reflect a higher overall proportion of female employers in this age range, relative to males.

[28]Because FTE employment is derived from the jobs data after own-firm WP labour input has been removed, WPs are never counted as employees of businesses they own. Following from this, and the assumed permanent WP status of an individual with respect to a firm, the addition of a working proprietor to a business never affects the assessment of whether a firm is an employer of staff or not.

[29]Firms, eg, partnerships can have multiple WPs either with or without employees. To avoid double-counting of total employees in the table, the FTE in multi-WP firms is apportioned to the individual WPs.

in the number of WPs in this employment size category. In the analysis that follows, we examine the characteristics of all employers regardless of WP firm size, before turning to the employment behaviour of new WPs, which usually involves small numbers of employees. We leave understanding the dynamics of how WP firms reach larger employment milestones to future work.

Panels B and D of figure 1 show the propensity for all WPs to employ (at any FTE size) by age and skill, respectively. For females, the likelihood of being an employer, conditional on being self-employed, is strictly declining in both age and skill. The decline from the 10th to 90th percentile of WP age is approximately 3pp, while the decline from the 10th to 90th percentile of WFE is 14pp. In contrast, for males the profile is flatter (and at some points increasing) over the 10th-90th percentile range for both age (declining less than 1pp) and WFE (declining 6pp).[30] For both sexes, the likelihood of employing drops steeply beyond aged 55, perhaps associated with retirement decisions.

Figures 3-5 repeat this graphical overview of entrepreneurship and employer rates by ethnicity, NZ/overseas-born status, and highest formal qualification respectively. These additional descriptive statistics are motivated by the average propensities reported in table 3 – in particular, the much higher proportion of self-employed of European-only ethnicity (83.4%) relative to their share of the ERP (61.8%), which is mirrored in higher self-employment for NZ-born. Highest qualification is included in this descriptive analysis to test the power of the wage-based skills measure to distinguish something useful over and above the standard skills metric of formal qualifications. For ethnicity, we focus on the five largest ethnicity groups to make the figure legible. For the same reason, we combine some formal qualifications to reduce the analysis to five groups.

Figure 3 illustrates how the entrepreneurship gap varies by ethnicity, age (panel A) and skill (panel C). Across all ethnicities, the likelihood of being a WP increases with age though there is substantial variation in the peak age for self-employment ranging from approximately 45 (for Pasifika-only) to 55 (for European-only, Māori-only and European×Māori). Consequently, the rate at which self-employment rates decrease around the retirement threshold also varies substantially across groups, with the steepest rate of decrease for European-only. All ethnicities display a U-shaped pattern in entrepreneurship by skill, though this is far less pronounced for Pasifika-only

[30]The 10th and 90th percentiles of the female (male) WP age distribution are, respectively, 34 (34) and 55 (58). The 10th and 90th percentiles of the female (male) WP WFE distribution are, respectively, -0.41 (-0.54) and 0.55 (0.71).

and Māori-only, where the self-employment rate for low-skilled individuals is barely higher than that for moderately-skilled individuals of the same ethnicity. In contrast, and aside from Pasifika-only, the self-employment rate for high-skilled individuals is often insignificantly different (at the 95% confidence interval) across ethnicity groups, though standard errors become large for non-European-only groups.

Conditional on being a WP, the likelihood of being an employer is similar across ethnicity groups, with the exception of Pasifika-only who are less likely to be employers than other individuals at most ages and skill (WFE) levels. With the exception of Pasifika-only, the likelihood of employing – conditional on being a WP – is declining in age, particularly from age 50 onwards. The Asian-only group show a distinctive pattern by skill, with moderately-skilled individuals showing a higher likelihood of employing than low-skilled. This is in contrast to other ethnicities, which show declining employer rates by skill.

This pattern is replicated when we group individuals by country of birth, where NZ-born have a decreasing likelihood to employ by skill, but overseas-born show a peak employer probability for moderately-skilled (figure 4D). Other patterns for NZ-born and overseas-born are similar to each other, with NZ-born having higher probabilities of being a WP and for being an employer, conditional on self-employment. In the latter instance, the profile by age (panel B) is flat up to age 50 for NZ-born before dropping, where overseas-born show a more steady decline in employer propensity with age.

Figure 5 illustrates the benefit of including both the skill (WFE) and formal qualifications variables in the analysis. Specifically, the skill (WFE) variable appears to capture a different aspect of individual ability as evidenced by all qualification groups displaying the same U-shaped propensity to be self-employed by skill (WFE). There is clearly a relationship between an individuals' ability to generate a wage premium in the labour market and becoming self-employed, that is over and above the knowledge gained by the individual in the formal education system. Conversely, the qualifications data shows that individuals without formal qualifications are less likely to be self-employed than other qualification groups over the 35-65 age range, providing a different perspective from the skill data, which generally shows that low-skilled individuals have higher self-employment rates than moderately-skilled individuals. Independent of skill (WFE), it appears that self-employed individuals with higher degrees (masters and doctorates) are less likely to be employers (figure 5D).

Table 5 (columns 1-3) summarise the gaps in self-employment by eth-

nicity (relative to European-only), as well as the gap for females (relative to males). Column (1) reports the mean proportion of WPs by sex or ethnicity group, while column (2) is the difference between these rates and the reference group rate. Column (3) expresses the gaps as a percentage of the mean self-employment rate in the population, which is 7.5% (top row of table). As intimated by figures 1A and 3A, these gaps are substantial with the raw entrepreneurship gap for Māori-only and Pasifika-only over 100 percent of the average WP rate (-108% and -126% gaps respectively). For females, the gap is almost half the mean entry rate. Gaps in the likelihood of self-employment are closely related to gaps in the entry rate into self-employment (columns 4-6), which are are discussed in more detail in section 3.3. Before that analysis, we consider how the characteristics examined in this section relate to the estimated gaps in the stock of self-employed and employers.

3.2 Entrepreneur characteristics – regressions

Tables 6 and 7 report results from multivariate (OLS) tests of the power of these individual characteristics to explain self-employment in the ERP (table 6), and employing in the population of working proprietors (table 7). We exclude skill (WFE) from these regressions because persistent self-employment creates an interpretation (reverse causality) issue for the missing WFE indicator variable, which affects over a third of the population of WPs (bottom row of table 3, column 2).[31]

These regressions are sequenced to enable us to examine the effect that adding additional individual characteristics has on the estimated sex and ethnicity entrepreneurship gaps. We initially (column 1) report the regression results simultaneously including sex and ethnicity group indicator variables,[32] where the reference group is male and European-only. As might be expected, estimated coefficients are very close to the raw gaps in the data, which can be seen – for table 6 – by comparing column (1) to table 5 (column 2).

Columns (2)-(5) introduce additional covariates in a cumulative fashion: column (2) adds a quartic functions of age;[33] column (3) adds indicator variables for (range-grouped) time since first arrival in New Zealand (NZ-

[31] This issue is largely negated in the entry into self-employment regressions because the five year exclusion period means that most individuals have (preceding) jobs as employees.

[32] All regressions include controls for year.

[33] More precisely, age is included as a quartic of $a = (\text{age} - 41)$, where 41 is the median age, which is consistent with the method used to estimate worker and firm wage fixed effects, and provides sufficient flexibility to capture the raw relationships observed in figure 1.

born is the reference group), and an indicator for conversational English language skills; column (4) adds highest qualification indicator variables (no formal qualification is the reference group); and column (5) adds dependants.

Coefficients for age are not reported in the table for brevity. Instead, figure 6 (panels A and B) approximates the estimated age profiles using local polynomial regressions, where the dependent and independent variables are estimated residuals from regressions on all the other covariates from the final (column 5) specifications in tables 6 and 7. This approach has the advantage of not imposing a quartic functional form and, therefore, enabling a better comparison to the raw profiles presented in figure 1 (panels A and B). While the raw profiles imply potentially important differences in age/WFE profiles by sex, sex-specific coefficients (aside from the female indicator) are not included in the main estimates, which enables a consistent interpretation of the female entrepreneurship gap across specifications. Figure 6 (panels C and D) re-estimates the smoothed propensities by sex – mimicking the inclusion of sex-specific age controls – to enable more direct comparison with figure 1, panels A and B respectively.[34]

The initial estimated self-employment gap between females and males is −3.7pp (table 6, column 1). In contrast, conditional on self-employment, the female-male employing gap is positive at 1.7pp (table 7, column 1). Controlling for additional covariates has a relatively minor effect on the estimated self-employment gap for females – increasing from −3.7pp to −3.8pp – and the employer gap – falling from 1.7pp to 1.6pp. Examining the estimated variation in self-employment by age suggests that the gaps in self-employment rates between young and old individuals within each sex are slightly larger after controlling for covariates (figure 6, panels A and C), compared to the raw estimates (figure 1A).

Relative to the reference group of European-only, the nine remaining ethnicity groupings all have negative self-employment gaps, ranging from −4.6pp for Asian-only to −9.4pp for Pasifika-only when additional controls are excluded (table 6, column 1). All nine gaps close somewhat when we control for age (column 2), reflecting the younger average age of non-European-only groups and the negative correlation between age and self-employment.

Being a recent arrival to New Zealand is associated with lower self-employment relative to NZ-born, with this negative relationship generally reducing the longer an individual has been in New Zealand (column 3). In

[34]In panels C and D of figure 6, the female indicator is excluded from the covariates so that mean difference between sexes are included in the age profiles, consistent with figure 1.

addition, having conversational English skills – which is a characteristics more commonly associated with long-term migrants and NZ-born – is associated with a 1.0pp higher likelihood of being a WP (at least in the absence of controls for qualifications). Taken together, the penalty for recency of arrival is a substantial component of the overall ethnicity gap for Asian and MELAA ethnicity groups (both -only and ×European). Column (4) of table 6 shows that individuals with formal qualifications have a higher likelihood of being self-employed, relative to the unqualified. This likelihood is highest for individuals with level 4 certificates, which includes trade certificates and, therefore, occupational groups often associated with self-employment. The inclusion of highest qualification reduces most ethnicity gaps, but the magnitude of the changes are small relative to the overall gaps. Finally, column (5) shows that the self-employed are more likely to have dependants, with the probability of being self-employed increasing in the number of children. The inclusion of dependant children variables has a negligible effect on other estimated coefficients, including the female gap.[35]

In summary, after adding controls for individual characteristics, substantial entrepreneurship gaps remain for Māori-only (−6.8pp) and Pasifika-only (−6.7pp). As a percentage of the raw (table 5) gap, Māori-only are the only ethnicity group where the gap shrinks by less than 29% with the introduction of controls – falling a relatively small 16% – partly because overseas-born cannot explain this particular entrepreneurship gap. The Pasifika-only gap closes by 29% due primarily to the relatively young age structure of this group, and due to a relatively high proportion of overseas-born. Asian and MELAA gaps (for both -only and ×European) close by at least 61% of the raw gap, largely explained by overseas-born status, though age structure is the primary factor for the European×Asian gap.

Conditional on self-employment, employing is also less likely for most ethnicity groups relative to European-only, with exceptions being Māori-only and Asian-only who are more likely to be employers (by approximately 1pp), and European×Māori who are insignificantly different from European-only in their likelihood to be employers in the absence of additional controls (table 7, column 1). With all controls added (column 4), some employer gaps increase

[35]Partly, the estimate of the coefficient on the female indicator variable is unaffected because both females and males appear more likely to become self-employed when they have dependent children. To see this, table B.1 (column 1) shows reestimated coefficients on dependants, where these coefficients are allowed to differ by sex. The coefficient on the female indicator variables is reported for completeness, but it is not comparable to the main estimates in table 6 because of the inclusion of other sex-specific coefficients in the regression.

and others decrease. The largest negative changes in the gap occur for Māori, with the apparent raw positive gap for Māori-only becoming insignificantly different from zero, and a significant negative gap (−1.8pp) appearing for European×Māori.

As with self-employment, the raw gap for Asian-only and MELAA-only appears to be affected by overseas-born status, with the largest positive changes in gap occurring for these groups following the introduction of overseas-born control variables – changes of 4.8pp and 6.3pp respectively comparing columns (1) and (5). The relationship between first arrival date and employing is substantial with a 14.4pp lower probability of employing if first arrival in NZ was within the last five years. While the relationship weakens the longer an individual has been in NZ, it is still −6.0pp at 31 years or more since first arrival. Conditional on being a WP, the probability of employing is declining in highest qualification (and conversational English ability), with post-graduate and higher-qualified individuals between 8.4pp and 12.5pp less likely to be employers than individuals with no formal qualifications. Self-employed individuals with dependants are more likely to employ and the probability of employing is approximately doubled going from having one child to having two children.[36]

Finally, figure 6 (panel B) shows the estimated relationship between employing and age based on the column (5) specification of table 7 and, additionally, allowing the age profile to be sex-specific (panel D). The estimated age profiles controlling for other covariates show an inverted U-shape over the 10th to 90th percentile of (residual) age (ie, from -12.4 to 11.8). In the specification with common age coefficients (figure 6B) the probability of employing increases with age up to around the 25th percentile (ie, residual age 5.6) before declining. Inspection of the quartic age coefficients across specifications (not reported) suggests that the more pronounced inverted U-shape – compared to the raw relationship in figure 1B – is largely driven by the inclusion of controls for the number of dependent children. This result is consistent with the higher employer probabilities for individuals below age 40 in the raw distribution (figure 1B) – relative to the multivariate estimates (figure 6D) – being driven by individuals with dependants.

[36]Table B.1 (column 2) confirms that the relationship between employing (conditional on self-employment) and dependants is similar for females and males.

3.3 Entry into entrepreneurship

We now estimate entrepreneurship gaps in transitions to self-employment. This second analysis is important for understanding how the stock of WPs evolves over time, particularly since we also examine the survival of new entrants in subsequent years. The analysis of entry into self-employment is also an opportunity to consider – in a setting less prone to reverse causality – the relationship between skill (labour market wage premia), labour market outcomes and self-employment, which may provide a deeper understanding of differential WP outcomes by ethnicity and migrant status. In addition, we consider the relationship between absences from NZ and entrepreneurship, which includes NZ-born who have lived overseas.

Table 8 summarises the WP entry rate by year. The "potential entrant" population of interest is those individuals in the ERP who have not been a WP in any of the five previous years. Column (1) shows the size of this population, which corresponds to the 10% sample in column (3) of table 3 (scaled by a factor of ten to represent the full population). The number of potential entrants increases over time because of general increases in the ERP and because a declining entrepreneurship rate over time results in more individuals satisfying the (five-year) non-WP population criteria.

Columns (2) and (3), respectively, report the number of entrants into self-employment and the number of entrants who have employees in their first year of business. There is a distinct drop-off in the absolute number WP entrants following the Global Financial Crisis in 2008 – a decline that hasn't subsequently recovered. Column (4) reports WP entry as a proportion of potential entrants, while columns (5) and (6) show the difference in the entry rate between 2005 and each subsequent year in raw terms (column 5) and as estimated in a model with a full set of individual-level covariates that might help explain changing trends in entry rates (from table 9, column 2). Column (6) includes stars signifying that all changes from 2005 are significantly different from zero, as are differences between 2006-2008 and subsequent years (not reported in table).

Overall, both raw and regression-based estimates suggest a substantial decline in entry dynamics over the last decade. Setting aside the 2015 year, where the entry rate may be subject to a late-filing undercount,[37] the entry rate had declined by 38% comparing the first three entry years (2005-2008)

[37] Additionally, the regression-based estimate for 2015 is affected by the higher missing rate for dependants in that year, and is much closer to the actual rate when dependent children variables (including the corresponding missing indicator) are excluded.

to the last three years (2012-2014). This declining dynamism is consistent with the observed decline in total WP firms (table 1, column 3), with exiting WPs not being replenished by a matching flow of new entrants. Relative to the total number of WPs reported in the top panel of table 4, the proportion of WPs who are new entrants has fallen from 9.0% to 7.1% using the same start/end three year comparison (column 7). Overall, though, almost 44% of all WPs we observe over the 2005-2015 period are entrants during that period, reflecting the importance of entrants in understanding the stock of WPs. Finally, column (8) reports the proportion of entering WPs who have employees in their first year of operation. The decline in overall entry has not been associated with an increase in early employment behaviour, suggesting the average quality of entering WPs has not increased – at least on one dimension – as a results of the reduced inflow rate.

Figure 7A sets the scene for regressions including labour market variables by plotting pre- and post-entry dynamics for WP entrants across three dimensions: having a job; receiving a main government benefit; and being absent from NZ for the entire (March) year. Individuals in this analysis become self-employed at time t (ie, entering WP cohorts are pooled). Mechanically, nobody is absent from NZ (dotted line, right-hand scale) at t because of the ERP rules, so that the rate of absence increases on either side of t. Aside from this mechanical decline in absence, we observe distinct changes in behaviour before and after individuals become self-employed. Both the likelihood of having a job (solid line) and receiving a benefit (dashed line, right-hand scale) decline prior to becoming self-employed. For employment, the decline occurs almost exclusively in the year prior to entry, with a 5.1pp drop in the probability of having a job in $t-1$ relative to $t-2$. For benefit receipt there is a consistent decline over each year leading to entry, aggregating to a 4.6pp decline in benefit receipt over the five years prior to becoming self-employed.

Between 65 and 67 percent of individuals hold a job in the five to two years prior to becoming a WP, explaining why coverage for the skill (WFE) variable is over 90% for the WP entrant population. The job-holding rate drops to 62% in the year prior to entry, before falling substantially on entry and the following year. Benefit receipt rates fall more consistently year-on-year leading up to entry, with the rate dropping by at least 1pp every year to $t-1$. Both job and benefit rates rise again after entry, partly reflecting the fact that a significant proportion of entrants do not continue with self-employment in subsequent years, as discussed in the next subsection.

While these descriptive statistics provide some insight into the dynamics of pre-entry labour market outcomes, understanding the relationship be-

tween these variables and the WP entry decision also requires an understanding of the labour market histories of non-entrants, and acknowledgement of the relationship between the labour market variables – job and benefit receipt – and absence from NZ. To do this, we turn to multivariate (OLS) regressions where the inclusion of these variable simultaneously allows us to understand the role of each, conditional on other covariates.

Table 9 shows the estimated relationship between individual characteristics and entry into self-employment, excluding and including controls for labour market experience and absences from NZ (columns 1 and 2-4 respectively). Labour market variables are added sequentially starting with a quartic in WFE (column 2) and prior job industry indicator variables (column 3). Coefficients are reported to five decimal places because the mean entry rate is 0.65% of the potential entrant population. As noted above, while this is a relatively rare event, entrants form a material proportion of the total WP stock, and 4.3% of potential entrants (228,870 individuals) will become WPs at some point over the 11 year analysis period (table 8, bottom row).

Table 9 (column 1) and table 6 (column 4) enable a comparison of observed sex and ethnicity entrepreneurship gaps in WP entry and being a WP, respectively, controlling for the same additional individual characteristics. Comparison of the incumbent and entry regression coefficients is hampered by the mean difference in the dependent variable. Table 10 simplifies this comparison by reporting each gap as a percentage of the relevant dependent variable mean, which is either the proportion of the ERP who are WPs (column 1) or the proportion of potential entrants who become self-employed (column 2) – in each case controlling for individual characteristics, but not labour market experience. Column (3) reports estimates from the entry regression (table 9, column 5) including complete controls, including those for labour market experience. Finally, columns (4)-(6) report gaps in WP outcomes five years after entry, which are discussed in the next subsection.

For example, the female entrepreneurship gap for entry, controlling for individual characteristics, is estimated at -47.5% of the entry rate (ie, $-0.0031/0.0065$). With additional controls for labour market history, this gap decreases substantially to -37.4% of the entry rate. The reduction in the female entry gap comes almost exclusively from the introduction of prior job industry controls (column 3 of table 9), since the industries that tend to have high self-employment rates are also male-dominated industries. As table B.2 show, 44-47% of WPs own firms in agriculture, manufacturing or

construction.[38]

For ethnicity, the entrepreneurship gaps in entry are highly correlated with gaps in the WP population (correlation of 0.96 comparing columns 1 and 2), as might be expected given the importance of entrants in the overall stock. The gap in WP entry rate, relative to European-only (excluding labour market controls), varies from insignificantly different from zero for European×Asian and European×MELAA to −78.3% for Māori-only and −97.7% for Pasifika-only. The introduction of labour market experience variables materially reduces the estimated ethnicity gaps (table 10, column 3), though Māori-only and Pasifika-only gaps remain substantial at 54.4% and 75.2% of the mean entry rate, respectively.

Returning to table 9 (column 4), we can see the relationship between the control variables and WP entry. There is a negative, but weakening over time, relationship between entry into self-employment, benefit receipt and full-year absences from NZ. The inclusion of absence variables has almost no impact on the estimated relationships between entry and first arrival, though we would expect the greatest risk of collinearity between the lagged absence variables and the first arrival in the last 5 years indicator variable, the latter of which has an insignificantly different from zero coefficient in both specifications including and excluding absence controls. The coefficients on some first arrival variables have a positive sign (significantly different from zero) in the entry regression. Taken together with the absence variables, these estimates suggest that very recent arrivals to NZ – regardless of migrant status – are less likely to start their own business, but that overseas-born who have been in NZ for a significant length of time (6-20 years) are more likely to become self-employed.

Aside from the immediately preceding year (ie, $t-1$), having a job and working for a good (high-paying) firm are positively associated with becoming a WP. In the year prior to potential entry, having a job is negatively associated with a transition to self-employment. These results match the pattern shown in figure 7A, where employment drops prior to entry into self-employment. This dip could be interpreted as a deliberate transition – quitting a job to start a business – or indicative of an unexpected negative shock to employment.

[38] In some sense this explanation is tautological – ie, if males are more likely to be self-employed then more male-dominated industries should be over-represented in the WP data. On the other hand, if there is something about those industries that make them more amenable to self-employment, then causality could run from industry to higher self-employment rates for males.

Several factors support the unexpected shock interpretation. Firstly, the job variable represents non-employment for the entire prior year, and we might expect a timed exit from the labour market to go hand-in-hand with a quicker transition to self-employment. Having said that, it could be that the transition is immediate, but because it takes time for a WP firm to become economically significant, there is an apparent lag between job cessation and WP starting. Secondly, for individuals who do have jobs at $t-1$, the quality of their employer (firm fixed effect) is negatively associated with entry in the following year, so that individuals that transition are more likely to have recently been employees of low-paying firms, which are potentially more likely to shed workers or have workers leave for better jobs.[39] Finally, individuals who generally experience greater job churn are also more likely to become WPs, as are individuals with weak English language skills. Conversely, the negative shock interpretation is less consistent with the strong negative coefficient on benefit receipt in the immediately preceding year, and the steady decline in benefit receipt among individuals who will become WPs (shown in figure 7A).

Overall, results suggest that individuals who tend to have jobs (and not receive benefits) – particularly good jobs – are more likely to become WPs. In addition, workers who have left their jobs recently (and aren't on benefits); who recently worked in low-paying firms; or who are subject to excess job churn are all more likely to transition to self-employment in the following year.

Estimated age profiles are consistent with earlier results, showing an initially increasing association between age and entry into self-employment for both females and males, with entry rates decreasing before and beyond retirement age (figure 8, panels A and C). After controlling for age, individuals with young children are more likely to become self-employed than individuals without young children. Several aspects of these results are informative. Firstly, both females and males are more likely to become self-employed (table B.1, column 3) if they have young children – with the estimated relationship stronger for males. Secondly, relationships are stronger for pre-school children than older (5-8yr old) children, particularly for males. Finally, within each dependent child age groups, the likelihood of becoming self-employed increases with the number of children. Overall, the observed relationships are consistent with the transition to self-employment being at

[39] We control for job industry, to avoid confounding industry and firm wage premia. Industry controls are indicator variables for being employed in each of 19 ANZSIC'06 divisions over the five prior years.

least partly related to gaining flexibility to meet childcare responsibilities, managing work-life balance, adapting to changing financial demands (perhaps particularly where self-employment supplements job wages), and/or changing aspirations for future income.

Remaining coefficients support the hypothesis that skilled individuals are more likely to become entrepreneurs, where skill is measured either as highest qualification or WFE. In the former case, the inclusion of WFE mutes the apparent role of formal qualifications (compare columns 1 and 2), reflecting the correlation between the skill metrics. The relationship between WFE and entry is shown in figure 8 (panels B and D), which shows an almost strictly increasing relationship between skill and entry into self-employment. This is in contrast to the raw estimated relationship between skill and being a WP, which showed a U-shaped relationship (figure 1C). The entry regression results control for time-varying characteristics associated with low skill, such as job and benefit history, which may explain the non-U-shaped relationship between entry and skill. In addition, as the next subsection demonstrates, differences in WP survival rates by skill may contribute to a U-shaped skill profile in the stock of entrepreneurs that is not present in the initial inflow of WP entrants.

3.4 Continued entrepreneurship and job creation

In this subsection we track two outcomes over the five years following WP entry – continued self-employment and employing. We treat employment in the year of entry as a "post-entry" outcome, since we cannot convincingly establish the sequencing of entry into self-employment and entry into employing if both events occur in the same year. In addition, we consider a third outcome – employing conditional on continued self-employment – so that we can separately distinguish gaps in employing due to attrition from self-employment, from gaps in employing due to other causes such as variation in the desire to grow a business. Figure 7B shows the post-entry probability of these three outcomes for the population of individuals entering self-employment at time t. By construction, the probability of being a WP (solid line) at time t is one (and zero in the previous five years). In the year following entry, 28% of WPs have exited and this attrition process continues at a slower pace until year $t+5$ where 46% of WPs are still self-employed. The probability of being an employer at time t is 27% (dashed line), and this rate declines less rapidly than the overall WP participation rate, so that the employer rate conditional on survival in self-employment actually rises over time (dotted line). By the

fifth year after entry, 20% of entering WPs are employers, corresponding to 43% of surviving WPs.

Since over half of all entering WPs exit by year five, understanding gaps in survival rates is an important step towards understanding entrepreneurial gaps in the WP population. Tables 11-13 relate the full set of WP characteristics, including labour market histories, to continued self-employment and, potentially, becoming an employer over the following five years. Figures 9-14 show the estimated age and skill profiles that accompany these regression estimates. This analysis follows the same specification as table 9 (column 4) and associated figure 8 (panels A and B). All individual characteristics are measured as at the year prior to WP entry (ie, are held at their initial $t-1$ values).[40] Time-varying labour market and absence from NZ variables are held fixed to avoid interpretation issues arising from WP outcomes affecting subsequent labour market outcomes.

Due to the volume of results, we concentrate on describing WP survival and employer gaps by sex and ethnicity. Conditional on becoming a WP, females are 0.8pp less likely than males to still be self-employed five year after entry (table 11, column 5), despite the fact that females are 0.6pp more likely than males to survive into the second year of self-employment (column 1). Conditional on continued self-employment, females are more likely to be employers than males (table 12). This difference exists at entry with a positive 5.3pp gap in employer likelihood at t, relative to men, decreasing to a 4.6pp positive gap five years after entry. Considering employer status without conditioning on WP survival, the gap is smaller – due to the higher attrition rate for females – but still positive (1.8pp) and significant (table 13, column 6).[41]

Table 10 (columns 4-6) summarise these fifth year post-entry results, reporting related coefficients as a percentage of the dependent variable mean to aid comparison with other gaps. For ethnicity, this table shows a negative and significant gap in the WP survival rate relative to European-only, except in the case of European×Asian where point estimates are negative but insignificantly different from zero at the 5% level, and Asian-only where the gap is positive (column 4). Ethnicity-related gaps in the survival rate

[40] As a robustness test (table B.3), we allow the dependent children indicator variables to be time-varying (measured at the prior March) since they may change substantially over five years. These additional results support the main estimate findings – in particular, the alternative estimates has no effect on estimated female entrepreneurship gaps.

[41] The first column of tables 12 and 13 are identical, since all entrants are WP at time t. These results are duplicated to aid comparison of coefficients over time.

are positively correlated with WP entry gaps (correlation of 0.85).[42] Thus, if we think of continued self-employment as a success metric, a lower entry rates for a particular ethnicity group does not, in turn, imply a higher average quality of WPs who do enter. Pasifika-only are 16.5pp less likely than European-only to still be self-employed five year after entry (table 11), with Māori-only and the residual ethnicity group having −8.5pp and −8.4pp gaps respectively.

In the case of both Pasifika-only and Māori-only, these survival gaps develop immediately after entry and grow over time. These results are consistent with the interpretation of the entry regression results, which concluded that some transitions to self-employment are likely driven by poor labour market outcomes. In that sense, rapid departure from self-employment – if it signals a return to employment as an employee – is a positive outcome for the individual. Conversely, higher exit rate gaps then imply a reduced pool of entrants who desired to create their own business – ie, true entrepreneurs.

As a consequence of these survival rate gaps, six of the nine ethnicity groups also have negative employer gaps, relative to European-only, five years after entry (table 10, column 6). However, conditioning on survival most gaps in the probability of employing are insignificantly different from zero (column 5). Māori-only (2.8pp) and Asian-only (7.2pp) each have positive and significant employer gaps in the last year of analysis, with this advantage appearing in the year of entry and persisting, though somewhat weakening over time in the case of Māori-only (table 12). Overall, ethnicity-related gaps in new entrepreneur employment appear to emerge because of differential WP survival rates, except in these two instances where WPs have a higher tendency to employ on entry.

Briefly, other characteristics have the following (significant) relationship with five year outcomes. Continued self-employment in the long-term is positively associated with: recently-arrived overseas-born (up to 20 years since arrival); qualifications at or below bachelors level, and at doctoral level; having young children; and having had a job prior to entry into self-employment (table 11). Conversely, prior beneficiaries are less likely to still be self-employed in the long-term, as are individuals who worked in high-paying firms prior to entry. Continuing self-employment is also negatively associated with age, and this negative relationship is stronger for older individuals (figure 9). Absence from NZ is positively related to survival if it was recent and not long-term, but of ambiguous sign for individuals whose absence stretches back further

[42]The correlation is 0.69 excluding Pasifika-only, which is an outlier in both the WP entry and survival gaps.

than one year. Other covariates (conversational English and job churn) have coefficients insignificantly different from zero. Figure 10 shows how the pattern of selection based on skill (WFE) develops over time. By $t + 5$ there is a substantial loss of moderately-skilled individuals from the entry cohorts, contributing to the U-shaped distribution in the skill distribution of the stock of entrepreneurs.

Conditional on still being self-employed at $t+5$, employing is negatively associated with: overseas-born (except for most recent arrivals); all qualifications above level 4 certificate, particularly post-graduate qualifications; WPs who previously worked in high quality firms; and age (table 12 and figure 11). Recent (in the year immediately preceding entry) prior job experience is strongly positively associated with employing, as is recent absence from NZ. Job churn and conversational English ability are unrelated to $t + 5$ employment, though the latter has a consistently negative sign over time and is significantly different from zero in three of the six years. Finally, skill appears to be negatively associated with employing at time t, but becomes positively associated with employing at $t + 5$, at least over the 10th to 90th percentile range of -0.41 to 0.43 (figure 12). For high-skilled individuals beyond the 90th percentile of the WFE, skill is consistently negatively related to employing with the relationship strengthening over time.

4 Conclusions

The self-employed constitute a significant proportion of the labour force, and create a substantial number of jobs for their employees. Raw differences in the overall self-employment rate vary substantially by sex and ethnicity, with Pasifika-only and Māori-only ethnicity groups having a 9.4pp and 8.1pp, respectively, lower probability of being self-employed than European-only. These entrepreneurship gaps are substantial when compared to the overall self-employment rate of 7.5% of the ERP. While partially explainable by differences in other individual characteristics, such as age and migrant status, gaps persist to some extent for all ethnicity groups, relative to European-only individuals. A similar entrepreneurship gap exists for females, representing 48% of the average WP participation rate after controlling for individual characteristics.

In the case of ethnicity, overall entrepreneurship rates are strongly related to the dynamics of entry into self-employment and survival rates for entrants. Controlling for individual characteristics and labour market out-

comes the WP entry rate gap for Pasifika-only ethnicity individuals (relative to European-only) is −75% of the mean entry rate, and the five-year survival rate gap after entry is −36% of the mean survival rate. For Māori-only ethnicity individuals, the corresponding entry and survival rate gaps are −54% and −18%.

The majority of ethnicity groups have negative gaps in employing five years after entry relative to the European-only group. However, these gaps are due entirely to differences in WP survival rates. After controlling for attrition from self-employment, differences in long-term employment outcomes are often unrelated to ethnicity, and the employer entrepreneurship gap for females is positive.

Results relating to control variables, particularly the U-shaped relationship between skill and entrepreneurship and the coefficients on labour market, are suggestive of additional effects that may vary by ethnicity, or other characteristics. Specifically, the results are consistent with involuntary self-employment transitions, to some extent. This type of transition is, perhaps, more likely to be associated with short spells of self-employment. While we include labour market controls aimed at capturing the relationship between, eg, involuntary job loss, and transitions to self-employment, these relationships are estimated across all individuals, rather than within groups.

For Pasifika-only and Māori-only, rapid exit from self-employment – if it signals a return to employment as an employee – is a positive outcome if job-holding is the preferred state (ie, self-employment is a stop-gap measure). Conversely, under this interpretation, higher exit rate gaps then imply a reduced pool of entrants who desired to create their own business – ie, true entrepreneurs. In addition, in many occupations, transitions between self-employment and job-holding may be costly.

Higher WP rates in the upper end of the skills distribution and for those with persistently positive prior job outcomes, indicate that self-employment is also closely associated with individuals with high earnings potential in the labour market. Such individuals are unlikely to be being forced into self-employment. The lack of a strong positive relationship between skill and job creation, is consistent with high-skilled individuals electing self-employment as a preferred way to supply their labour services to the market. These individuals too may have high personal net worth (which is not captured in the dataset), which may affect the ability of individuals to start businesses in NZ. The international literature suggests that differences in access to capital and specific business human capital (from, eg, parents or peers) may go some way to explaining the residual ethnicity gaps in entrepreneurship. In

the latter instance, high-skilled individuals are more likely to reach management positions in firms that would allow them to, perhaps, develop the skills necessary to operate their own business successfully.

5 Potential extensions

We have created a general purpose dataset to begin to understand the dynamics of entrepreneurship in New Zealand. A number of potential avenues for further work stand out. Firstly, the declining dynamism of WP entry is perplexing and, perhaps, an issue for future job growth in NZ. It would be interesting to investigate changes in these dynamics pre- and post-GFC to understand why the entry rate for entrepreneurs has dropped so substantially. Secondly, and along similar lines, estimating the potentially sizeable effect of changing demographics on future entrepreneurship rates would be interesting, particularly given the observed relationships between age and WP entry, particularly around the age of retirement (see, eg, Liang et al. 2014). Thirdly, and relating to the possibility that labour market opportunities may affect the goals of entrepreneurs, understanding the possible heterogeneity in the importance of different factors for ethnicity groups would be useful (eg, separate labour market coefficients by ethnicity). Fourthly, given the substantial gap in entrepreneurship by sex, it would be sensible to investigate how the sex gap varies by ethnicity (see, eg, Mora and Dávila 2014). Fifthly, while we have considered the probability of WPs creating jobs, we have not investigated the quality of those jobs and, in particular, the characteristics of first hires, which may give further insight into the entrepreneurial potential of new ventures. Sixthly, the impact of short-term WP spells could be examined in more detail – specifically the nature of jobs before and after these spells and whether (transitory) self-employment has a positive or negative impact on future job earnings. Finally, we have made several important data decisions to simplify this analysis. It would be good to revisit these, given that they may deepen our understanding of the entrepreneurial process. These issues include: the longer-term growth of WP firms and the characteristics of WPs who manage to grow large employment firms; the role concurrent jobs have in affecting longer-term WP outcomes; and whether WP experience – both success and failure – leads to improved outcomes in future entrepreneurial endeavours (see, eg, Lafontaine and Shaw 2016; Shaw and Sørensen 2017).

References

Acs, Z., S. Desai, and L. Klapper (2008). What does "entrepreneurship" data really show? *Small Business Economics 31*(3), 265–281.

Ahn, T. (2011). Racial differences in self-employment exits. *Small Business Economics 36*(2), 169–186.

Bates, T. (2011). Minority entrepreneurship. *Foundations and Trends(R) in Entrepreneurship 7*(3-4), 151–311.

Black, A. (2016). The IDI prototype spine's creation and coverage. Working Papers 16-03, Statistics New Zealand.

Blanchflower, D. G. (2000). Self-employment in OECD countries. *Labour Economics 7*(5), 471–505.

Blanchflower, D. G. (2004). Self-employment: More may not be better. *Swedish Economic Policy Review 11*, 15–73.

Blanchflower, D. G. (2015). Self-employment across countries in the Great Recession of 2008-2014. In *Yearly report on flexible labor and employment*, flexibility@work, pp. 7–128. randstad: Amsterdam.

Blanchflower, D. G., P. B. Levine, and D. J. Zimmerman (2003). Discrimination in the small-business credit market. *The Review of Economics and Statistics 85*(4), 930–943.

Blanchflower, D. G. and A. J. Oswald (1998). What makes an entrepreneur? *Journal of Labor Economics 16*(1), 26–60.

Blanchflower, D. G., A. J. Oswald, and A. Stutzer (2001). Latent entrepreneurship across nations. *European Economic Review 45*(4-6), 680–691.

Borjas, G. J. (1986). The self-employment experience of immigrants. *Journal of Human Resources 21*(4), 485–506.

Davis, S. J., J. Haltiwanger, R. S. Jarmin, C. J. Krizan, J. Miranda, A. Nucci, and K. Sandusky (2009). Measuring the dynamics of young and small businesses: Integrating the employer and nonemployer universes. In T. Dunne, J. B. Jensen, and M. J. Roberts (Eds.), *Producer Dynamics: New Evidence from Micro Data*, NBER Studies in Income and Wealth, Chapter 9, pp. 329–366. University of Chicago Press: Chicago.

Decker, R. A., J. Haltiwanger, R. S. Jarmin, and J. Miranda (2016). Declining business dynamism: What we know and the way forward. *American Economic Review 106*(5), 203–07.

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 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 L. Sanderson (2014). 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.

Fairlie, R. W. (1999). The absence of the African-American owned business: An analysis of the dynamics of self-employment. *Journal of Labor Economics 17*(1), 80–108.

Fairlie, R. W. (2007). Entrepreneurship among disadvantaged groups: Women, minorities and the less educated. In S. Parker (Ed.), *The Life Cycle of Entrepreneurial Ventures*, Volume 3 of *International Handbook Series on Entrepreneurship*, pp. 437–475. Springer: Boston, MA.

Fairlie, R. W. (2013). Entrepreneurship, economic conditions, and the great recession. *Journal of Economics & Management Strategy 22*(2), 207–231.

Fairlie, R. W. and M. Lofstrom (2015). Immigration and entrepreneurship. Volume 1 of *Handbook of the Economics of International Migration*, Chapter 17, pp. 877–911. Elsevier: Oxford.

Fairlie, R. W. and B. D. Meyer (1996). Ethnic and racial self-employment differences and possible explanations. *Journal of Human Resources 31*(4), 757–793.

Fairlie, R. W. and J. Miranda (2016). Taking the leap: The determinants of entrepreneurs hiring their first employee. *Journal of Economics & Management Strategy* (forthcoming, http://dx.doi.org/10.1111/jems.12176).

Fairlie, R. W. and A. M. Robb (2007). Why are black-owned businesses less successful than white-owned businesses? The role of families, inheritances, and business human capital. *Journal of Labor Economics 25*(2), 289–323.

Fairlie, R. W. and C. M. Woodruff (2010). Mexican-American entrepreneurship. *The B.E. Journal of Economic Analysis & Policy 10*(1), 1–44.

Garcia-Perez, M., C. Goetz, J. Haltiwanger, and K. Sandusky (2013). Don't quit your day job: Using wage and salary earnings to support a new business. Working Papers 13-45, Center for Economic Studies,

U.S. Census Bureau.

Gibb, S., C. Bycroft, and N. Matheson-Dunning (2016). Identifying the New Zealand resident population in the Integrated Data Infrastructure (IDI). Census Transformation Papers, Statistics New Zealand.

Hamilton, B. H. (2000). Does entrepreneurship pay? An empirical analysis of the returns to self-employment. *Journal of Political Economy 108*(3), 604–631.

Hurst, E. and B. W. Pugsley (2011). What do small businesses do? NBER Working Papers 17041, National Bureau of Economic Research.

Katz, L. F. and A. B. Krueger (2016). The rise and nature of alternative work arrangements in the United States, 1995-2015. NBER Working Papers 22667, National Bureau of Economic Research.

Kerr, S. P. and W. R. Kerr (2016). Immigrant entrepreneurship. In J. Haltiwanger, E. Hurst, J. Miranda, and A. Schoar (Eds.), *Measuring Entrepreneurial Businesses: Current Knowledge and Challenges*, NBER Studies in Income and Wealth, Chapter 5, forthcoming. University of Chicago Press: Chicago.

Lafontaine, F. and K. Shaw (2016). Serial entrepreneurship: Learning by doing? *Journal of Labor Economics 34*(S2), 217–254.

Lazear, E. P. (2004). Balanced skills and entrepreneurship. *American Economic Review 94*(2), 208–211.

Lazear, E. P. (2005). Entrepreneurship. *Journal of Labor Economics 23*(4), 649–680.

Liang, J., H. Wang, and E. P. Lazear (2014). Demographics and entrepreneurship. NBER Working Papers 20506, National Bureau of Economic Research.

Lofstrom, M. (2002). Labor market assimilation and the self-employment decision of immigrant entrepreneurs. *Journal of Population Economics 15*(1), 83–114.

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

Mora, M. T. and A. Dávila (2014). Gender and business outcomes of black and hispanic new entrepreneurs in the United States. *American Economic Review 104*(5), 245–249.

Ramnath, S., J. B. Shoven, and S. N. Slavov (2017). Pathways to retirement through self-employment. NBER Working Papers 23551, National Bureau of Economic Research.

Shaw, K. L. and A. Sørensen (2017). The productivity advantage of serial entrepreneurs. NBER Working Papers 23320, National Bureau of Economic Research.

Statistics New Zealand (2016). Experimental population estimates from

linked administrative data: Methods and results. Census Transformation Papers, Statistics New Zealand: Wellington.

van der Sluis, J., M. van Praag, and W. Vijverberg (2008). Education and entrepreneurship selection and performance: A review of the empirical literature. *Journal of Economic Surveys 22*(5), 795–841.

Yuan, S., T. Cain, and P. Spoonley (2013). Immigrant entrepreneurship and tax compliance. Working Papers, Inland Revenue Department: Wellington.

Tables & figures

Table 1: Employment in working proprietor-owned firms

	(1)	(2)	(3)	(4)	(5)
	FTE employees		WP	Ratios	
	Total	In WP firm	Headcount	(2)/(1)	(3)/(1)
All enterprises (including public sector)					
2005	1,394,700	354,200	392,844	0.254	0.282
2006	1,438,100	366,700	393,234	0.255	0.273
2007	1,464,200	364,100	390,486	0.249	0.267
2008	1,497,600	369,000	387,066	0.246	0.258
2009	1,504,600	358,400	377,106	0.238	0.251
2010	1,468,800	330,700	367,956	0.225	0.251
2011	1,478,600	333,300	366,606	0.225	0.248
2012	1,497,200	332,500	360,933	0.222	0.241
2013	1,518,300	337,700	355,509	0.222	0.234
2014	1,556,200	344,600	351,006	0.221	0.226
2015	1,603,700	351,800	336,474	0.219	0.210
Economically significant private-for-profit firms					
2005	1,049,600	349,900	299,775	0.333	0.286
2006	1,087,500	361,900	299,235	0.333	0.275
2007	1,106,800	360,800	297,993	0.326	0.269
2008	1,132,600	365,900	295,947	0.323	0.261
2009	1,128,700	354,900	286,005	0.314	0.253
2010	1,084,600	327,500	273,627	0.302	0.252
2011	1,089,900	329,600	272,391	0.302	0.250
2012	1,106,800	330,800	267,318	0.299	0.242
2013	1,126,100	336,000	263,967	0.298	0.234
2014	1,159,200	341,000	264,651	0.294	0.228
2015	1,202,100	350,500	252,921	0.292	0.210

FTE is derived from the Fabling-Maré labour dataset for each March year (eg, 2005 is the year ending on 31st March 2005). WP headcounts come from the same dataset and are assumed to relate to March years also. Firms are economically significant in a year if they have GST-exclusive sales or purchases of at least $30K (real, 2000 dollars) or if they have employees. Following Fabling and Maré (2015b), private-for-profit is a permanent characteristic assigned to a firm based on business type and institutional sector.

Table 2: Decomposition of gap between preferred and actual self-employment rates in New Zealand (using ISSP)

	Proportion	
Year	1997	2005
Preferred self-employment rate	**0.611**	**0.552**
LESS Employees preferring to be self-employed	-0.467	-0.394
PLUS Self-employed preferring to be employees	0.074	0.017
EQUALS Actual self-employment rate	**0.219**	**0.175**
Proportion of employees preferring to be self-employed	0.597	0.477
Proportion of self-employed preferring to be employees	0.339	0.096

Own calculation based on International Social Survey Programme (ISSP) data downloaded from http://www.gesis.org/issp/home/ on 1 December 2016. Preferred self-employment rate in 1997 differs from the 64.2% reported in Table 1 of Blanchflower et al. (2001) because we restrict the sample to employed individuals. Using the within-survey self-employment rate enables an internally consistent decomposition of the gap between "preferred" and "actual." The actual self-employment rate reported in Blanchflower et al. (2001) using OECD data is 22.7% for 1997.

Table 3: Sample and population characteristics

	(1)	(2)	(3)	(4)
	Regression sample/population			
Sample (10%) or population (WP)	10%	WP	10%	WP
Restricted to potential/actual entrants	N	N	Y	Y
Last year of observation	2015	2015	2015	2010
Observations	4,010,454	3,005,475	3,567,669	144,618
Individuals	554,838	521,268	533,052	144,618
	Proportion of total observations			
Male	0.506	0.625	0.493	0.588
Female	0.494	0.375	0.507	0.412
European-only	0.618	0.834	0.592	0.769
European×Māori	0.049	0.031	0.052	0.042
European×Pasifika	0.007	0.003	0.008	0.005
European×Asian	0.004	0.002	0.004	0.003
European×MELAA	0.010	0.006	0.010	0.009
Māori-only	0.070	0.018	0.076	0.029
Pasifika-only	0.049	0.005	0.055	0.010
Asian-only	0.105	0.076	0.109	0.109
MELAA-only	0.010	0.005	0.011	0.009
Residual	0.015	0.004	0.017	0.007
Missing	*0.062*	*0.017*	*0.067*	*0.009*
NZ-born	0.483	0.653	0.464	0.573
First arrival: 0-5yrs	0.040	0.013	0.044	0.067
6-10yrs	0.034	0.028	0.035	0.052
11-20yrs	0.042	0.051	0.041	0.045
21-30yrs	0.021	0.028	0.020	0.021
31+yrs	0.043	0.057	0.041	0.040
unknown	0.010	0.004	0.011	0.006
Missing	*0.327*	*0.166*	*0.344*	*0.197*
No conversational English	0.014	0.011	0.015	0.014
Has conversational English	0.653	0.824	0.634	0.788
Missing	*0.333*	*0.165*	*0.351*	*0.199*
No qualification	0.137	0.136	0.136	0.104
Highest qual: Level 1-3 cert	0.226	0.273	0.221	0.257
Level 4 cert	0.069	0.133	0.063	0.115
Level 5-6 dip	0.065	0.089	0.062	0.086
Bachelor/level 7	0.097	0.118	0.095	0.138
Post-grad/hons	0.021	0.025	0.020	0.030
Masters	0.020	0.023	0.019	0.033
Doctorate	0.005	0.008	0.005	0.009
Post-school unknown	0.016	0.024	0.015	0.023
Missing	*0.344*	*0.170*	*0.363*	*0.205*
No dependants	0.497	0.591	0.485	0.565
Dependants: One (0-4yr)	0.034	0.040	0.034	0.078
2+ (0-4yr)	0.017	0.027	0.016	0.041
One (5-8yr)	0.029	0.050	0.027	0.054
2+ (5-8yr)	0.009	0.018	0.008	0.017
2+ (mixed)	0.026	0.045	0.025	0.054
Missing	*0.388*	*0.230*	*0.405*	*0.191*
Has worker fixed effect (WFE)	0.814	0.662	0.829	0.903
Missing	*0.186*	*0.338*	*0.171*	*0.097*

The first year of observation for each sample/population is 2005. The working proprietor entrant population (column 4) is restricted to 2005-2010 so that five years of future outcomes are available for each entering cohort year. Because a potential entrant cannot have WP experience in the prior five years, individuals can only appear in the actual entrant population once over this time period.

Table 4: Distribution of working proprietors by number of FTE employees

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Loss from	Proportion of total					
		ERP & age	by employment size (FTE) group					
	Total	restrictions	0	$(0, 2]$	$(2, 5]$	$(5, 10]$	$(10, 20]$	$(20, \infty)$
WP headcount								
2005	293,484	0.021	0.484	0.375	0.091	0.032	0.012	0.005
2006	292,971	0.021	0.485	0.371	0.093	0.032	0.012	0.006
2007	291,837	0.021	0.493	0.364	0.092	0.032	0.013	0.006
2008	289,767	0.021	0.499	0.358	0.092	0.033	0.013	0.006
2009	279,981	0.021	0.510	0.348	0.090	0.033	0.012	0.006
2010	267,789	0.021	0.515	0.346	0.090	0.032	0.012	0.006
2011	266,430	0.022	0.517	0.342	0.090	0.033	0.012	0.006
2012	261,246	0.023	0.518	0.341	0.089	0.033	0.012	0.006
2013	257,760	0.024	0.519	0.337	0.090	0.033	0.013	0.007
2014	258,069	0.025	0.521	0.332	0.092	0.034	0.013	0.007
2015	246,147	0.027	0.515	0.330	0.096	0.037	0.014	0.008
FTE employees in WP firms								
2005	336,500	0.038	0.000	0.224	0.247	0.190	0.145	0.194
2006	346,000	0.044	0.000	0.216	0.246	0.189	0.142	0.208
2007	343,800	0.047	0.000	0.212	0.245	0.188	0.145	0.209
2008	350,900	0.041	0.000	0.203	0.239	0.188	0.141	0.228
2009	341,800	0.037	0.000	0.198	0.233	0.186	0.136	0.247
2010	317,400	0.031	0.000	0.203	0.237	0.188	0.135	0.237
2011	319,000	0.032	0.000	0.202	0.236	0.189	0.135	0.239
2012	318,200	0.038	0.000	0.199	0.231	0.186	0.136	0.248
2013	323,400	0.038	0.000	0.191	0.227	0.183	0.138	0.261
2014	327,900	0.038	0.000	0.189	0.227	0.185	0.143	0.256
2015	336,100	0.041	0.000	0.176	0.221	0.184	0.144	0.275

Column (2) reports the loss of sample, relative to the bottom panel of table 1, from imposing the age and estimated resident population (ERP) restrictions. The FTE contribution of a WP is calculated by apportioning the FTE in multi-WP firms, and by aggregating across firms for WPs with multiple businesses.

Table 5: Raw entrepreneurship gaps, as percentage of mean rate

	(1)	(2)	(3)	(4)	(5)	(6)
	Being a WP			Entry into self-employment		
	Proportion WP	Raw gap	Gap as % of mean	Proportion entering	Raw gap	Gap as % of mean
Mean	**0.075**			**0.0065**		
Male	0.092			0.0079		
Female	0.057	-0.036	-48.1%	0.0052	-0.0027	-41.1%
European-only	0.101			0.0084		
European×Māori	0.045	-0.056	-75.1%	0.0048	-0.0035	-54.0%
European×Pasifika	0.035	-0.066	-87.8%	0.0039	-0.0044	-68.2%
European×Asian	0.047	-0.054	-71.6%	0.0055	-0.0029	-44.3%
European×MELAA	0.049	-0.051	-68.6%	0.0055	-0.0029	-44.0%
Māori-only	0.020	-0.081	-108.0%	0.0025	-0.0059	-90.4%
Pasifika-only	0.007	-0.094	-125.6%	0.0012	-0.0071	-109.2%
Asian-only	0.054	-0.047	-63.1%	0.0072	-0.0011	-17.5%
MELAA-only	0.032	-0.069	-92.5%	0.0056	-0.0028	-42.9%
Residual	0.020	-0.080	-107.6%	0.0027	-0.0056	-86.3%

Raw gaps are calculated from 10% samples, and relative to male and European-only. Columns (3) and (6) express these gaps as a percentage of the mean of the corresponding dependent variable (reported in the top panel of the table), and are directly comparable to the estimated residual gaps in columns (1)-(3) of table 10.

Table 6: Correlates of being a working proprietor

	(1)	(2)	(3)	(4)	(5)
Female	-0.037**	-0.037**	-0.039**	-0.037**	-0.038**
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
European×Māori	-0.054**	-0.038**	-0.038**	-0.036**	-0.037**
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
European×Pasifika	-0.064**	-0.045**	-0.042**	-0.040**	-0.040**
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
European×Asian	-0.051**	-0.029**	-0.017**	-0.017**	-0.016**
	[0.005]	[0.004]	[0.004]	[0.004]	[0.004]
European×MELAA	-0.052**	-0.041**	-0.014**	-0.012**	-0.013**
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Māori-only	-0.081**	-0.076**	-0.073**	-0.067**	-0.068**
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Pasifika-only	-0.094**	-0.083**	-0.071**	-0.065**	-0.067**
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Asian-only	-0.046**	-0.032**	-0.012**	-0.010**	-0.009**
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
MELAA-only	-0.069**	-0.053**	-0.030**	-0.027**	-0.027**
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Residual	-0.080**	-0.059**	-0.044**	-0.041**	-0.042**
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
First arrival: 0-5yrs			-0.060**	-0.062**	-0.062**
			[0.001]	[0.001]	[0.001]
6-10yrs			-0.032**	-0.034**	-0.035**
			[0.001]	[0.001]	[0.001]
11-20yrs			-0.005*	-0.007**	-0.008**
			[0.002]	[0.002]	[0.002]
21-30yrs			0.000	-0.002	-0.002
			[0.002]	[0.002]	[0.002]
31+yrs			-0.013**	-0.014**	-0.014**
			[0.002]	[0.002]	[0.002]
unknown			-0.045**	-0.042**	-0.041**
			[0.003]	[0.003]	[0.003]
Has conversational English			0.010**	0.002	0.003
			[0.003]	[0.003]	[0.003]

Table continued on next page

Table continued from previous page

	(1)	(2)	(3)	(4)	(5)
Highest qual: Level 1-3 cert				0.028**	0.027**
				[0.001]	[0.001]
Level 4 cert				0.054**	0.053**
				[0.002]	[0.002]
Level 5-6 dip				0.026**	0.025**
				[0.002]	[0.002]
Bachelor/level 7				0.029**	0.028**
				[0.002]	[0.002]
Post-grad/hons				0.023**	0.022**
				[0.003]	[0.003]
Masters				0.016**	0.015**
				[0.003]	[0.003]
Doctorate				0.034**	0.032**
				[0.006]	[0.006]
Post-school unknown				0.039**	0.038**
				[0.004]	[0.004]
Dependants: One (0-4yr)					0.024**
					[0.001]
2+ (0-4yr)					0.049**
					[0.002]
One (5-8yr)					0.035**
					[0.002]
2+ (5-8yr)					0.051**
					[0.003]
2+ (mixed)					0.048**
					[0.002]
Observations	4,010,454	4,010,454	4,010,454	4,010,454	4,010,454
R^2	0.023	0.049	0.055	0.057	0.059
Mean of dependent variable	0.075	0.075	0.075	0.075	0.075
Quartic in age	N	Y	Y	Y	Y

Ordinary least squares regression for 10% random sample (weighted) of ERP, sampled at individual level. Dependent variable is an indicator variable set to one if the individual is a WP (zero otherwise). Robust (clustered on individual) standard errors in brackets (**/* implies coefficient significantly different from zero at the 1/5% level). Regressions include year dummies. Reference group is male; European-only; NZ-born; no conversational English; no qualification; no young dependants. Where relevant, unreported indicator variables also included for missing: ethnicity; NZ-born status; conversational English; highest qualification; number of dependants. First arrival unknown is overseas-born but year of first arrival is unknown. Highest qualification post-school unknown is post-school qualification of unknown level. Age is included as a quartic of $a = (\text{age}-41)$, where 41 is the median age. The relationship between being a working proprietor and age, conditional on other column (5) covariates, is reported in figure 6A.

Table 7: Correlates of employing, conditional on being a working proprietor

	(1)	(2)	(3)	(4)	(5)
Female	0.017**	0.013**	0.015**	0.016**	0.016**
	[0.001]	[0.001]	[0.001]	[0.002]	[0.002]
European×Māori	0.002	-0.005	-0.017**	-0.017**	-0.018**
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
European×Pasifika	-0.047**	-0.056**	-0.059**	-0.058**	-0.058**
	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]
European×Asian	-0.051**	-0.060**	-0.046**	-0.042**	-0.041**
	[0.015]	[0.015]	[0.015]	[0.015]	[0.015]
European×MELAA	-0.056**	-0.056**	-0.046**	-0.045**	-0.046**
	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]
Māori-only	0.012*	0.008	-0.003	-0.004	-0.004
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Pasifika-only	-0.077**	-0.086**	-0.051**	-0.058**	-0.060**
	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
Asian-only	0.011**	0.001	0.061**	0.058**	0.059**
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
MELAA-only	-0.030**	-0.041**	0.030**	0.032**	0.033**
	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
Residual	-0.034**	-0.041**	-0.037**	-0.037**	-0.036**
	[0.010]	[0.010]	[0.010]	[0.011]	[0.011]
First arrival: 0-5yrs			-0.160**	-0.145**	-0.144**
			[0.004]	[0.004]	[0.004]
6-10yrs			-0.130**	-0.116**	-0.117**
			[0.003]	[0.003]	[0.003]
11-20yrs			-0.095**	-0.081**	-0.083**
			[0.003]	[0.003]	[0.003]
21-30yrs			-0.071**	-0.062**	-0.061**
			[0.004]	[0.004]	[0.004]
31+yrs			-0.066**	-0.060**	-0.060**
			[0.003]	[0.003]	[0.003]
unknown			-0.053**	-0.049**	-0.051**
			[0.011]	[0.011]	[0.011]
Has conversational English			-0.047**	-0.033**	-0.032**
			[0.007]	[0.007]	[0.007]

Table continued on next page

Table continued from previous page

	(1)	(2)	(3)	(4)	(5)
Highest qual: Level 1-3 cert				0.011**	0.010**
				[0.002]	[0.002]
Level 4 cert				0.004	0.003
				[0.003]	[0.003]
Level 5-6 dip				-0.016**	-0.018**
				[0.003]	[0.003]
Bachelor/level 7				-0.017**	-0.019**
				[0.003]	[0.003]
Post-grad/hons				-0.082**	-0.084**
				[0.005]	[0.005]
Masters				-0.123**	-0.125**
				[0.005]	[0.005]
Doctorate				-0.111**	-0.114**
				[0.009]	[0.009]
Post-school unknown				0.003	0.002
				[0.005]	[0.005]
Dependants: One (0-4yr)					0.028**
					[0.002]
2+ (0-4yr)					0.061**
					[0.003]
One (5-8yr)					0.027**
					[0.002]
2+ (5-8yr)					0.052**
					[0.003]
2+ (mixed)					0.067**
					[0.003]
Observations	3,005,475	3,005,475	3,005,475	3,005,475	3,005,475
R^2	0.001	0.004	0.008	0.011	0.012
Mean of dependent variable	0.494	0.494	0.494	0.494	0.494
Quartic in age	N	Y	Y	Y	Y

Ordinary least squares regression for population of WPs. Dependent variable is an indicator variable set to one if the WP has employees (zero otherwise). Robust (clustered on individual) standard errors in brackets (**/* implies coefficient significantly different from zero at the 1/5% level). Regressions include year dummies. Reference group is male; European-only; NZ-born; no conversational English; no formal qualification; no young dependants. Where relevant, unreported indicator variables also included for missing: ethnicity; NZ-born status; conversational English; highest qualification; number of dependants. First arrival unknown is overseas-born but year of first arrival is unknown. Highest qualification post-school unknown is post-school qualification of unknown level. Age is included as a quartic of $a = (\text{age}-41)$, where 41 is the median age. The relationship between being an employer and age, conditional on other column (5) covariates, is reported in figure 6B.

Table 8: Working proprietor entry rate by year

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Potential	Entrants		WP	$\Delta_{2005,t}$[WP entry rate]		Ratios	
	entrants	WP	Employer	entry rate	Actual	Estimated	(2)/N(WP)	(3)/(2)
2005	2,954,370	27,660	7,890	0.0094	-	-	0.094	0.285
2006	2,997,600	25,650	7,530	0.0086	-0.00081	-0.00090**	0.088	0.294
2007	3,057,090	25,680	7,500	0.0084	-0.00096	-0.00110**	0.088	0.292
2008	3,118,770	25,890	7,140	0.0083	-0.00106	-0.00129**	0.089	0.276
2009	3,186,720	21,900	5,700	0.0069	-0.00249	-0.00280**	0.078	0.260
2010	3,253,410	18,030	4,500	0.0055	-0.00382	-0.00420**	0.067	0.250
2011	3,304,110	18,900	4,680	0.0057	-0.00364	-0.00418**	0.071	0.248
2012	3,368,880	19,050	5,160	0.0057	-0.00371	-0.00424**	0.073	0.271
2013	3,403,410	17,790	4,710	0.0052	-0.00414	-0.00468**	0.069	0.265
2014	3,473,190	18,540	5,190	0.0053	-0.00402	-0.00449**	0.072	0.280
2015	3,559,080	12,840	3,450	0.0036	-0.00575	-0.00458**	0.052	0.269
Observations	35,676,690	231,960	63,390	0.0065			0.077	0.273
Individuals	5,330,520	228,870	63,000	0.0429			0.439	0.275

Estimated counts derived from 10% sample (ie, multiplied by ten). The total observations row sums individual-year observations, while the total individuals row counts each potential entrant once across the eleven years. The former does not add to the sum of the individual years because confidentialisation is applied independently to the true total. Because of the 11 year timespan, multiple new entry in possible for an individual, though in practice it is not common (as can been seen by comparing the last two rows of column 3). The drop-off in entrants in 2015 may, in part, relate to delays in new enterprises appearing on the Business Register. Column (6) reports estimated coefficients on year dummies from specification (4) of table 9. ** indicates coefficient significantly different from zero at the 1% level, with associated (robust) standard errors of 0.00024 (for 2006-2008); 0.00023 (for 2009); and 0.00022 (for 2010-2015). Column (7) shows the proportion of entrants in the total WP population, where the latter is from table 4 (top panel) and table 3 (top of column 2). Column (8) shows the proportion of entrant working proprietors that employ in the year of entry.

Table 9: Correlates of entry into self-employment

	(1)	(2)	(3)	(4)
Female	-0.00309**	-0.00294**	-0.00229**	-0.00243**
	[0.00009]	[0.00009]	[0.00009]	[0.00010]
European×Maori	-0.00292**	-0.00253**	-0.00250**	-0.00192**
	[0.00018]	[0.00018]	[0.00018]	[0.00018]
European×Pasifika	-0.00371**	-0.00344**	-0.00308**	-0.00278**
	[0.00039]	[0.00039]	[0.00039]	[0.00039]
European×Asian	-0.00112	-0.00091	-0.00065	-0.00083
	[0.00064]	[0.00064]	[0.00064]	[0.00064]
European×MELAA	-0.00009	-0.00001	0.00017	0.00018
	[0.00041]	[0.00041]	[0.00041]	[0.00041]
Maori-only	-0.00509**	-0.00446**	-0.00448**	-0.00354**
	[0.00012]	[0.00012]	[0.00012]	[0.00013]
Pasifika-only	-0.00635**	-0.00564**	-0.00521**	-0.00489**
	[0.00012]	[0.00012]	[0.00012]	[0.00012]
Asian-only	-0.00105**	-0.00038*	-0.00024	-0.00058**
	[0.00017]	[0.00017]	[0.00017]	[0.00018]
MELAA-only	-0.00286**	-0.00232**	-0.00217**	-0.00176**
	[0.00040]	[0.00040]	[0.00040]	[0.00040]
Residual	-0.00350**	-0.00296**	-0.00259**	-0.00206**
	[0.00023]	[0.00023]	[0.00023]	[0.00023]
First arrival: 0-5yrs	-0.00004	0.00024	-0.00012	-0.00024
	[0.00028]	[0.00028]	[0.00028]	[0.00030]
6-10yrs	0.00164**	0.00191**	0.00190**	0.00164**
	[0.00032]	[0.00032]	[0.00032]	[0.00032]
11-20yrs	0.00077**	0.00088**	0.00088**	0.00086**
	[0.00027]	[0.00027]	[0.00027]	[0.00027]
21-30yrs	0.00036	0.00027	0.00036	0.00043
	[0.00036]	[0.00036]	[0.00036]	[0.00036]
31+yrs	-0.00056*	-0.00072**	-0.00066**	-0.00072**
	[0.00023]	[0.00023]	[0.00023]	[0.00023]
unknown	-0.00246**	-0.00242**	-0.00268**	-0.00263**
	[0.00036]	[0.00036]	[0.00036]	[0.00037]
Has conversational English	-0.00115**	-0.00112**	-0.00091*	-0.00133**
	[0.00039]	[0.00039]	[0.00039]	[0.00039]
Highest qual: Level 1-3 cert	0.00248**	0.00198**	0.00222**	0.00176**
	[0.00015]	[0.00015]	[0.00015]	[0.00015]
Level 4 cert	0.00538**	0.00480**	0.00477**	0.00437**
	[0.00026]	[0.00026]	[0.00026]	[0.00026]
Level 5-6 dip	0.00315**	0.00203**	0.00257**	0.00208**
	[0.00023]	[0.00024]	[0.00024]	[0.00024]
Bachelor/level 7	0.00369**	0.00195**	0.00265**	0.00210**
	[0.00021]	[0.00021]	[0.00022]	[0.00022]
Post-grad/hons	0.00293**	0.00062	0.00156**	0.00115**
	[0.00039]	[0.00039]	[0.00040]	[0.00040]
Masters	0.00516**	0.00240**	0.00335**	0.00294**
	[0.00045]	[0.00045]	[0.00046]	[0.00046]
Doctorate	0.00443**	0.00019	0.00131	0.00101
	[0.00083]	[0.00084]	[0.00084]	[0.00084]
Post-school unknown	0.00370**	0.00304**	0.00320**	0.00281**
	[0.00044]	[0.00044]	[0.00044]	[0.00043]
Dependants: One (0-4yr)	0.00508**	0.00497**	0.00490**	0.00471**
	[0.00035]	[0.00035]	[0.00035]	[0.00035]
2+ (0-4yr)	0.00776**	0.00747**	0.00727**	0.00651**
	[0.00055]	[0.00055]	[0.00055]	[0.00055]
One (5-8yr)	0.00225**	0.00229**	0.00228**	0.00253**
	[0.00036]	[0.00036]	[0.00036]	[0.00036]
2+ (5-8yr)	0.00399**	0.00389**	0.00379**	0.00380**
	[0.00070]	[0.00070]	[0.00070]	[0.00070]
2+ (mixed)	0.00377**	0.00370**	0.00353**	0.00319**
	[0.00039]	[0.00039]	[0.00039]	[0.00039]

Table continued on next page

Table continued from previous page

	(1)	(2)	(3)	(4)
Absent from NZ: $t-1$				-0.00546**
				[0.00013]
$t-2$				-0.00198**
				[0.00018]
$t-3$				-0.00102**
				[0.00021]
$t-4$				0.00023
				[0.00023]
$t-5$				0.00070**
				[0.00022]
Has benefit: $t-1$				-0.00352**
				[0.00016]
$t-2$				-0.00079**
				[0.00020]
$t-3$				-0.00095**
				[0.00020]
$t-4$				-0.00089**
				[0.00021]
$t-5$				-0.00122**
				[0.00018]
Has job: $t-1$				-0.00830**
				[0.00022]
$t-2$				0.00167**
				[0.00023]
$t-3$				0.00057**
				[0.00022]
$t-4$				0.00039
				[0.00021]
$t-5$				-0.00018
				[0.00018]
Average firm fixed effect: $t-1$				-0.01270**
				[0.00082]
$t-2$				0.00274**
				[0.00104]
$t-3$				0.00240*
				[0.00103]
$t-4$				0.00046
				[0.00100]
$t-5$				0.00092
				[0.00079]
Average job churn				0.00075**
				[0.00009]
Observations	3,567,669	3,567,669	3,567,669	3,567,669
R^2	0.005	0.005	0.006	0.007
Mean of dependent variable	0.007	0.007	0.007	0.007
Quartic in worker fixed effect	N	Y	Y	Y
Prior job industry	N	N	Y	Y

Ordinary least squares regression for 10% random sample (weighted) of potential entrants in ERP, sampled at individual level. Dependent variable is an indicator variable set to one if the individual becomes a WP (zero otherwise). Robust (clustered on individual) standard errors in brackets (**/* implies coefficient significantly different from zero at the 1/5% level). All regressions include year dummies and a quartic in age ($a = (\text{age}-41)$); columns (2)-(4) include a quartic in WFE; and columns (3)-(4) include prior job industry dummies (19 ANZSIC'06 divisions). The relationships between entry into self-employment and age/WFE, conditional on other column (4) covariates, are reported in figure 8. Reference group is male; European-only; NZ-born; no conversational English; no formal qualification; no dependent children. Where relevant, unreported indicator variables also included for missing: ethnicity; NZ-born status; conversational English; highest qualification; worker fixed effect; number of dependants. First arrival unknown is overseas-born but year of first arrival is unknown. Highest qualification post-school unknown is post-school qualification of unknown level.

Table 10: Summary of entrepreneurship gaps controlling for observables, as percentage of mean rate

	(1)	(2)	(3)	(4)	(5)	(6)
	Being a	Entry into		Five years after entry		
	WP	self-employment		WP	Employer | WP	Employer
Female	-50.8%**	-47.5%**	-37.4%**	-1.7%**	10.6%**	8.9%**
European×Māori	-49.5%**	-44.9%**	-29.5%**	-8.6%**	3.7%	-6.0%*
European×Pasifika	-53.5%**	-57.1%**	-42.8%**	-10.3%*	-2.3%	-10.4%
European×Asian	-21.4%**	-17.2%	-12.8%	-6.2%	-13.6%	-16.4%*
European×MELAA	-17.4%**	-1.4%	2.8%	-7.3%**	-10.2%	-10.9%*
Māori-only	-91.0%**	-78.3%**	-54.4%**	-18.3%**	6.5%*	-13.4%**
Pasifika-only	-89.6%**	-97.7%**	-75.2%**	-35.5%**	-5.1%	-36.3%**
Asian-only	-12.0%**	-16.1%**	-8.9%**	2.6%*	16.6%**	21.4%**
MELAA-only	-36.1%**	-44.0%**	-27.1%**	-5.8%*	8.3%	2.5%
Residual	-56.2%**	-53.8%**	-31.7%**	-18.1%**	-1.2%	-18.4%**
Mean of dependent variable	0.075	0.0065	0.0065	0.464	0.433	0.201
Individual characteristics	Y	Y	Y	Y	Y	Y
Labour market experience	N	N	Y	Y	Y	Y

Estimated entrepreneurship gaps, controlling for covariates, come from: table 6 (column 5) for "being a WP;" table 9 (columns 1 and 4) for "entry into self-employment;" and from the $t+5$ estimates in tables 11-13, respectively, for "five years after entry." Columns (1) and (2) excludes controls for labour market experience. All estimated gaps are expressed as a percentage of the mean of the corresponding dependent variable (reported in the bottom panel of the table), and are directly comparable to the raw gaps presented in table 5 (columns 3 & 6). Ethnicity gaps are relative to European-only. **/* implies underlying coefficient significantly different from zero at the 1/5% level (robust standard errors clustered on individual).

Table 11: Correlates of continued self-employment

	(1) $t+1$	(2) $t+2$	(3) $t+3$	(4) $t+4$	(5) $t+5$
Female	0.006*	0.001	-0.004	-0.006*	-0.008**
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
European×Māori	-0.007	-0.017**	-0.032**	-0.029**	-0.040**
	[0.006]	[0.006]	[0.007]	[0.007]	[0.006]
European×Pasifika	-0.047*	-0.049*	-0.030	-0.005	-0.048*
	[0.018]	[0.019]	[0.019]	[0.019]	[0.019]
European×Asian	-0.011	-0.012	-0.039	-0.042	-0.029
	[0.021]	[0.023]	[0.023]	[0.022]	[0.022]
European×MELAA	-0.014	-0.033*	-0.033*	-0.041**	-0.034**
	[0.013]	[0.014]	[0.014]	[0.013]	[0.013]
Māori-only	-0.023**	-0.060**	-0.069**	-0.080**	-0.085**
	[0.007]	[0.008]	[0.008]	[0.008]	[0.008]
Pasifika-only	-0.070**	-0.100**	-0.137**	-0.160**	-0.165**
	[0.013]	[0.013]	[0.013]	[0.012]	[0.012]
Asian-only	0.019**	0.018**	0.002	0.006	0.012*
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
MELAA-only	-0.012	-0.015	-0.014	-0.004	-0.027*
	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]
Residual	-0.009	-0.057**	-0.078**	-0.073**	-0.084**
	[0.014]	[0.015]	[0.015]	[0.015]	[0.014]
First arrival: 0-5yrs	0.031**	0.038**	0.060**	0.069**	0.074**
	[0.006]	[0.007]	[0.007]	[0.007]	[0.007]
6-10yrs	0.027**	0.028**	0.027**	0.029**	0.027**
	[0.006]	[0.006]	[0.006]	[0.007]	[0.007]
11-20yrs	0.001	0.009	0.018**	0.027**	0.019**
	[0.006]	[0.007]	[0.007]	[0.007]	[0.007]
21-30yrs	-0.001	-0.005	0.002	0.001	-0.005
	[0.008]	[0.009]	[0.009]	[0.009]	[0.009]
31+yrs	-0.011	-0.022**	-0.027**	-0.020**	-0.026**
	[0.006]	[0.007]	[0.007]	[0.007]	[0.007]
unknown	-0.010	-0.001	0.003	0.002	0.021
	[0.016]	[0.017]	[0.017]	[0.017]	[0.017]
Has conversational English	-0.031**	-0.019	-0.008	-0.001	-0.008
	[0.010]	[0.011]	[0.011]	[0.012]	[0.012]
Highest qual: Level 1-3 cert	0.004	0.005	0.015**	0.017**	0.013**
	[0.004]	[0.005]	[0.005]	[0.005]	[0.005]
Level 4 cert	0.018**	0.025**	0.036**	0.034**	0.038**
	[0.005]	[0.005]	[0.006]	[0.006]	[0.006]
Level 5-6 dip	-0.001	-0.009	0.007	0.006	0.004
	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
Bachelor/level 7	-0.002	0.002	0.011*	0.017**	0.019**
	[0.005]	[0.006]	[0.006]	[0.006]	[0.006]
Post-grad/hons	-0.010	-0.010	-0.002	0.001	-0.002
	[0.008]	[0.009]	[0.009]	[0.009]	[0.009]
Masters	-0.006	-0.015	-0.005	-0.006	-0.005
	[0.008]	[0.008]	[0.009]	[0.009]	[0.009]
Doctorate	0.004	0.028*	0.028	0.025	0.047**
	[0.013]	[0.014]	[0.014]	[0.015]	[0.015]
Post-school unknown	0.014	0.025**	0.030**	0.036**	0.033**
	[0.008]	[0.009]	[0.009]	[0.009]	[0.009]
Dependants: One (0-4yr)	0.031**	0.048**	0.053**	0.055**	0.056**
	[0.004]	[0.005]	[0.005]	[0.005]	[0.005]
2+ (0-4yr)	0.033**	0.043**	0.054**	0.063**	0.062**
	[0.006]	[0.006]	[0.007]	[0.007]	[0.007]
One (5-8yr)	0.012*	0.033**	0.043**	0.036**	0.036**
	[0.005]	[0.006]	[0.006]	[0.006]	[0.006]
2+ (5-8yr)	0.027**	0.035**	0.023*	0.026**	0.032**
	[0.009]	[0.010]	[0.010]	[0.010]	[0.010]
2+ (mixed)	0.040**	0.041**	0.052**	0.043**	0.048**
	[0.005]	[0.006]	[0.006]	[0.006]	[0.006]

Table continued on next page

Table continued from previous page

	(1) $t+1$	(2) $t+2$	(3) $t+3$	(4) $t+4$	(5) $t+5$
Absent from NZ: $t-1$	0.019	0.023	0.039*	0.047**	0.032*
	[0.014]	[0.015]	[0.015]	[0.015]	[0.015]
$t-2$	-0.004	-0.017	-0.020*	-0.019*	-0.016
	[0.008]	[0.009]	[0.009]	[0.009]	[0.009]
$t-3$	0.014*	0.019*	0.000	-0.006	-0.006
	[0.007]	[0.008]	[0.008]	[0.008]	[0.008]
$t-4$	0.005	-0.006	-0.007	-0.011	-0.014
	[0.006]	[0.007]	[0.007]	[0.007]	[0.007]
$t-5$	0.013*	0.000	-0.008	-0.015*	-0.020**
	[0.006]	[0.007]	[0.007]	[0.007]	[0.007]
Has benefit: $t-1$	-0.032**	-0.043**	-0.050**	-0.048**	-0.046**
	[0.007]	[0.008]	[0.008]	[0.007]	[0.007]
$t-2$	-0.017*	-0.019*	-0.016	-0.015	-0.013
	[0.007]	[0.008]	[0.008]	[0.008]	[0.008]
$t-3$	-0.006	-0.015*	-0.018*	-0.020**	-0.018*
	[0.007]	[0.007]	[0.008]	[0.007]	[0.007]
$t-4$	0.003	-0.005	-0.007	-0.004	-0.008
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
$t-5$	-0.009	-0.024**	-0.025**	-0.032**	-0.032**
	[0.005]	[0.006]	[0.006]	[0.006]	[0.006]
Has job: $t-1$	0.026**	0.037**	0.043**	0.046**	0.041**
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
$t-2$	0.022**	0.022**	0.022**	0.013**	0.014**
	[0.004]	[0.005]	[0.005]	[0.005]	[0.005]
$t-3$	0.007	0.017**	0.011*	0.020**	0.020**
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
$t-4$	0.013**	0.013*	0.017**	0.015**	0.012*
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
$t-5$	0.020**	0.032**	0.034**	0.036**	0.040**
	[0.004]	[0.004]	[0.005]	[0.005]	[0.004]
Average firm fixed effect: $t-1$	-0.030*	-0.041**	-0.056**	-0.067**	-0.047**
	[0.014]	[0.015]	[0.015]	[0.015]	[0.015]
$t-2$	-0.020	-0.017	-0.017	-0.017	-0.021
	[0.017]	[0.018]	[0.018]	[0.018]	[0.018]
$t-3$	-0.027	-0.036	-0.029	-0.024	-0.020
	[0.017]	[0.019]	[0.019]	[0.019]	[0.019]
$t-4$	-0.005	-0.014	-0.034	-0.053**	-0.058**
	[0.018]	[0.019]	[0.019]	[0.019]	[0.019]
$t-5$	-0.002	-0.002	0.008	0.006	0.000
	[0.015]	[0.016]	[0.016]	[0.016]	[0.016]
Average job churn	-0.008**	-0.001	0.001	0.002	0.003
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Observations	144,618	144,618	144,618	144,618	144,618
R^2	0.029	0.042	0.050	0.058	0.064
Mean of dependent variable	0.718	0.617	0.551	0.502	0.464

Ordinary least squares regression for population of working proprietors who enter in $t \in [2005, 2010]$. Dependent variable is an indicator variable set to one if the individual is a WP (zero otherwise) at time $t+x$. All covariates are as at $t-1$ (ie, the year prior to entry). Robust standard errors in brackets (**/* implies coefficient significantly different from zero at the 1/5% level). Regressions include year dummies, prior job industry dummies (19 ANZSIC'06 divisions), and quartics in age ($a = (\text{age}-41)$) and worker fixed effect. Relationships between continued self-employment at $t+x$ and age/WFE, conditional on other covariates, are reported in figures 9 and 10 respectively. Reference group is male; European-only; NZ-born; no conversational English; no formal qualification; no dependent children. Unreported indicator variables also included for missing: ethnicity; NZ-born status; conversational English; highest qualification; worker fixed effect; number of dependants. First arrival unknown is overseas-born but year of first arrival is unknown. Highest qualification post-school unknown is post-school qualification of unknown level.

Table 12: Correlates of employing, conditional on entry at t and WP at $t+x$

	(1) t	(2) $t+1$	(3) $t+2$	(4) $t+3$	(5) $t+4$	(6) $t+5$
Female	0.053**	0.050**	0.050**	0.048**	0.048**	0.046**
	[0.003]	[0.003]	[0.004]	[0.004]	[0.004]	[0.004]
European×Māori	0.029**	0.029**	0.026**	0.025**	0.013	0.016
	[0.006]	[0.007]	[0.008]	[0.009]	[0.009]	[0.010]
European×Pasifika	0.013	0.009	-0.011	-0.028	-0.043	-0.010
	[0.017]	[0.022]	[0.025]	[0.026]	[0.027]	[0.030]
European×Asian	0.001	0.001	0.004	-0.004	-0.029	-0.059
	[0.020]	[0.026]	[0.029]	[0.032]	[0.034]	[0.035]
European×MELAA	0.008	-0.001	-0.012	-0.023	-0.048*	-0.044
	[0.012]	[0.016]	[0.019]	[0.021]	[0.023]	[0.025]
Māori-only	0.063**	0.063**	0.038**	0.024*	0.034**	0.028*
	[0.007]	[0.009]	[0.010]	[0.011]	[0.012]	[0.013]
Pasifika-only	0.017	0.025	0.000	-0.003	-0.003	-0.022
	[0.012]	[0.016]	[0.018]	[0.020]	[0.023]	[0.025]
Asian-only	0.068**	0.061**	0.059**	0.056**	0.058**	0.072**
	[0.005]	[0.006]	[0.006]	[0.007]	[0.007]	[0.008]
MELAA-only	0.026*	0.032*	0.026	0.020	0.023	0.036
	[0.012]	[0.015]	[0.017]	[0.018]	[0.019]	[0.020]
Residual	0.002	-0.010	-0.004	0.011	0.020	-0.005
	[0.013]	[0.018]	[0.021]	[0.023]	[0.025]	[0.028]
First arrival: 0-5yrs	0.009	-0.005	0.004	0.000	-0.001	0.000
	[0.006]	[0.008]	[0.009]	[0.009]	[0.010]	[0.010]
6-10yrs	-0.012*	-0.021**	-0.021**	-0.020*	-0.019*	-0.022*
	[0.006]	[0.007]	[0.008]	[0.008]	[0.009]	[0.009]
11-20yrs	-0.021**	-0.034**	-0.033**	-0.032**	-0.032**	-0.041**
	[0.006]	[0.007]	[0.008]	[0.009]	[0.009]	[0.009]
21-30yrs	-0.037**	-0.049**	-0.045**	-0.033**	-0.039**	-0.039**
	[0.008]	[0.010]	[0.011]	[0.012]	[0.012]	[0.013]
31+yrs	-0.024**	-0.042**	-0.041**	-0.039**	-0.046**	-0.043**
	[0.006]	[0.007]	[0.008]	[0.009]	[0.009]	[0.010]
unknown	0.009	-0.007	0.006	-0.005	0.017	0.006
	[0.016]	[0.021]	[0.022]	[0.024]	[0.025]	[0.025]
Has conversational English	-0.051**	-0.039**	-0.022	-0.025	-0.041**	-0.027
	[0.012]	[0.013]	[0.014]	[0.015]	[0.016]	[0.016]
Highest qual: Level 1-3 cert	0.001	-0.002	0.001	0.002	0.010	0.008
	[0.004]	[0.006]	[0.006]	[0.006]	[0.007]	[0.007]
Level 4 cert	-0.016**	-0.014*	-0.005	0.000	0.012	-0.002
	[0.005]	[0.006]	[0.007]	[0.007]	[0.008]	[0.008]
Level 5-6 dip	-0.022**	-0.030**	-0.023**	-0.010	-0.012	-0.018*
	[0.005]	[0.007]	[0.008]	[0.008]	[0.008]	[0.009]
Bachelor/level 7	-0.028**	-0.035**	-0.031**	-0.030**	-0.017*	-0.024**
	[0.005]	[0.006]	[0.007]	[0.007]	[0.008]	[0.008]
Post-grad/hons	-0.063**	-0.077**	-0.074**	-0.078**	-0.072**	-0.084**
	[0.007]	[0.009]	[0.010]	[0.011]	[0.012]	[0.012]
Masters	-0.077**	-0.091**	-0.080**	-0.084**	-0.073**	-0.084**
	[0.007]	[0.009]	[0.010]	[0.011]	[0.011]	[0.012]
Doctorate	-0.092**	-0.089**	-0.089**	-0.074**	-0.072**	-0.098**
	[0.011]	[0.015]	[0.016]	[0.017]	[0.018]	[0.018]
Post-school unknown	-0.003	-0.014	0.004	0.004	0.015	0.002
	[0.009]	[0.011]	[0.011]	[0.012]	[0.013]	[0.013]
Dependants: One (0-4yr)	0.041**	0.033**	0.036**	0.035**	0.030**	0.029**
	[0.005]	[0.006]	[0.006]	[0.006]	[0.007]	[0.007]
2+ (0-4yr)	0.062**	0.077**	0.072**	0.078**	0.068**	0.066**
	[0.006]	[0.008]	[0.008]	[0.008]	[0.009]	[0.009]
One (5-8yr)	0.023**	0.026**	0.022**	0.023**	0.027**	0.037**
	[0.005]	[0.007]	[0.007]	[0.007]	[0.008]	[0.008]
2+ (5-8yr)	0.041**	0.032**	0.032**	0.043**	0.014	0.023
	[0.009]	[0.011]	[0.012]	[0.013]	[0.013]	[0.014]
2+ (mixed)	0.047**	0.050**	0.054**	0.050**	0.046**	0.051**
	[0.006]	[0.007]	[0.007]	[0.008]	[0.008]	[0.008]

Table continued on next page

Table continued from previous page

	(1) t	(2) $t+1$	(3) $t+2$	(4) $t+3$	(5) $t+4$	(6) $t+5$
Absent from NZ: $t-1$	0.153**	0.175**	0.190**	0.182**	0.172**	0.155**
	[0.016]	[0.019]	[0.020]	[0.022]	[0.022]	[0.024]
$t-2$	-0.001	0.019	0.030**	0.011	0.029*	0.023
	[0.008]	[0.010]	[0.011]	[0.012]	[0.013]	[0.013]
$t-3$	-0.003	-0.013	-0.008	0.001	-0.002	-0.006
	[0.007]	[0.009]	[0.010]	[0.010]	[0.011]	[0.011]
$t-4$	0.005	0.013	0.007	0.012	0.001	0.004
	[0.006]	[0.008]	[0.009]	[0.009]	[0.010]	[0.010]
$t-5$	-0.021**	-0.019*	-0.017*	-0.026**	-0.023*	-0.024*
	[0.006]	[0.008]	[0.008]	[0.009]	[0.010]	[0.010]
Has benefit: $t-1$	0.016*	0.021*	0.013	0.007	-0.002	0.002
	[0.007]	[0.009]	[0.010]	[0.011]	[0.012]	[0.012]
$t-2$	-0.010	-0.009	-0.013	-0.003	-0.011	-0.006
	[0.007]	[0.009]	[0.010]	[0.011]	[0.012]	[0.013]
$t-3$	0.004	0.004	0.015	0.003	0.003	0.012
	[0.007]	[0.009]	[0.010]	[0.010]	[0.011]	[0.012]
$t-4$	0.016*	-0.003	-0.005	-0.003	0.007	-0.015
	[0.007]	[0.008]	[0.009]	[0.010]	[0.010]	[0.011]
$t-5$	-0.012*	-0.005	-0.004	-0.009	-0.012	-0.003
	[0.005]	[0.007]	[0.008]	[0.008]	[0.009]	[0.009]
Has job: $t-1$	0.076**	0.097**	0.102**	0.098**	0.094**	0.104**
	[0.003]	[0.004]	[0.005]	[0.005]	[0.005]	[0.005]
$t-2$	-0.003	-0.012*	-0.009	-0.002	0.010	0.005
	[0.004]	[0.005]	[0.006]	[0.006]	[0.007]	[0.007]
$t-3$	0.002	0.003	0.005	0.008	0.006	0.006
	[0.004]	[0.006]	[0.006]	[0.007]	[0.007]	[0.007]
$t-4$	-0.003	0.003	0.004	-0.002	0.001	0.004
	[0.005]	[0.006]	[0.006]	[0.007]	[0.007]	[0.007]
$t-5$	0.008*	0.008	0.003	0.006	0.006	0.007
	[0.004]	[0.005]	[0.006]	[0.006]	[0.006]	[0.007]
Average firm fixed effect: $t-1$	-0.092**	-0.089**	-0.069**	-0.068**	-0.059**	-0.065**
	[0.013]	[0.016]	[0.018]	[0.020]	[0.020]	[0.021]
$t-2$	-0.022	-0.029	-0.072**	-0.054*	-0.048	-0.032
	[0.015]	[0.019]	[0.022]	[0.024]	[0.024]	[0.026]
$t-3$	-0.053**	-0.080**	-0.022	-0.032	-0.043	-0.071**
	[0.016]	[0.021]	[0.023]	[0.025]	[0.026]	[0.027]
$t-4$	-0.016	0.017	-0.005	-0.003	0.021	0.027
	[0.017]	[0.021]	[0.023]	[0.025]	[0.027]	[0.029]
$t-5$	-0.005	-0.048**	-0.046*	-0.056**	-0.075**	-0.079**
	[0.014]	[0.018]	[0.020]	[0.021]	[0.022]	[0.023]
Average job churn	-0.001	0.000	-0.005	-0.002	0.000	-0.002
	[0.002]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Observations	144,618	103,806	89,250	79,617	72,582	67,152
R^2	0.049	0.058	0.058	0.063	0.065	0.065
Mean of dependent variable	0.273	0.362	0.395	0.413	0.422	0.433

Ordinary least squares regression for population of working proprietors who enter in $t \in [2005, 2010]$ and continue to be a working proprietor at $t+x$. Dependent variable is an indicator variable set to one if the individual is an employer (zero otherwise) at time $t+x$. All covariates are as at $t-1$ (ie, the year prior to entry). Robust standard errors in brackets (**/* implies coefficient significantly different from zero at the 1/5% level). Regressions include year dummies, prior job industry dummies (19 ANZSIC'06 divisions), and quartics in age ($a = (\text{age}-41)$) and worker fixed effect. Relationships between employing at $t+x$ and age/WFE, conditional on other covariates and continuing self-employment, are reported in figures 11 and 12 respectively. Reference group is male; European-only; NZ-born; no conversational English; no formal qualification; no dependent children. Unreported indicator variables also included for missing: ethnicity; NZ-born status; conversational English; highest qualification; worker fixed effect; number of dependants. First arrival unknown is overseas-born but year of first arrival is unknown. Highest qualification post-school unknown is post-school qualification of unknown level.

Table 13: Correlates of employing, conditional on entry at t

	(1) t	(2) $t+1$	(3) $t+2$	(4) $t+3$	(5) $t+4$	(6) $t+5$
Female	0.053**	0.039**	0.031**	0.025**	0.022**	0.018**
	[0.003]	[0.003]	[0.002]	[0.002]	[0.002]	[0.002]
European×Māori	0.029**	0.016**	0.008	-0.002	-0.008	-0.012*
	[0.006]	[0.006]	[0.006]	[0.006]	[0.005]	[0.005]
European×Pasifika	0.013	-0.009	-0.024	-0.025	-0.022	-0.021
	[0.017]	[0.016]	[0.016]	[0.015]	[0.015]	[0.015]
European×Asian	0.001	-0.007	-0.006	-0.018	-0.027	-0.033*
	[0.020]	[0.019]	[0.019]	[0.018]	[0.017]	[0.016]
European×MELAA	0.008	-0.004	-0.016	-0.018	-0.026**	-0.022*
	[0.012]	[0.011]	[0.010]	[0.010]	[0.009]	[0.008]
Māori-only	0.063**	0.035**	-0.003	-0.018**	-0.021**	-0.027**
	[0.007]	[0.007]	[0.007]	[0.006]	[0.006]	[0.006]
Pasifika-only	0.017	-0.008	-0.037**	-0.054**	-0.065**	-0.073**
	[0.012]	[0.011]	[0.010]	[0.010]	[0.009]	[0.008]
Asian-only	0.068**	0.054**	0.047**	0.035**	0.035**	0.043**
	[0.005]	[0.005]	[0.005]	[0.004]	[0.004]	[0.004]
MELAA-only	0.026*	0.018	0.010	0.007	0.010	0.005
	[0.012]	[0.012]	[0.012]	[0.011]	[0.011]	[0.011]
Residual	0.002	-0.008	-0.023	-0.027*	-0.023*	-0.037**
	[0.013]	[0.013]	[0.012]	[0.012]	[0.011]	[0.010]
First arrival: 0-5yrs	0.009	0.008	0.020**	0.026**	0.031**	0.034**
	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
6-10yrs	-0.012*	-0.005	-0.003	-0.001	0.003	0.003
	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
11-20yrs	-0.021**	-0.026**	-0.019**	-0.013*	-0.009	-0.016**
	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
21-30yrs	-0.037**	-0.035**	-0.032**	-0.018*	-0.021**	-0.021**
	[0.008]	[0.008]	[0.007]	[0.007]	[0.007]	[0.007]
31+yrs	-0.024**	-0.032**	-0.031**	-0.030**	-0.030**	-0.030**
	[0.006]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
unknown	0.009	-0.006	0.005	0.001	0.013	0.017
	[0.016]	[0.016]	[0.016]	[0.015]	[0.015]	[0.015]
Has conversational English	-0.051**	-0.048**	-0.029**	-0.028**	-0.032**	-0.030**
	[0.012]	[0.011]	[0.011]	[0.011]	[0.011]	[0.010]
Highest qual: Level 1-3 cert	0.001	0.000	0.004	0.009*	0.014**	0.011**
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Level 4 cert	-0.016**	-0.003	0.007	0.016**	0.022**	0.017**
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Level 5-6 dip	-0.022**	-0.023**	-0.019**	-0.005	-0.005	-0.008
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Bachelor/level 7	-0.028**	-0.028**	-0.022**	-0.016**	-0.005	-0.005
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Post-grad/hons	-0.063**	-0.062**	-0.054**	-0.048**	-0.042**	-0.046**
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
Masters	-0.077**	-0.071**	-0.059**	-0.053**	-0.044**	-0.047**
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.006]
Doctorate	-0.092**	-0.068**	-0.056**	-0.042**	-0.040**	-0.045**
	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.010]
Post-school unknown	-0.003	-0.005	0.011	0.014	0.022**	0.015
	[0.009]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
Dependants: One (0-4yr)	0.041**	0.038**	0.045**	0.046**	0.043**	0.043**
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
2+ (0-4yr)	0.062**	0.071**	0.068**	0.074**	0.071**	0.067**
	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
One (5-8yr)	0.023**	0.024**	0.027**	0.030**	0.027**	0.031**
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
2+ (5-8yr)	0.041**	0.033**	0.036**	0.035**	0.018*	0.024**
	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]
2+ (mixed)	0.047**	0.052**	0.052**	0.052**	0.043**	0.047**
	[0.006]	[0.006]	[0.005]	[0.005]	[0.005]	[0.005]

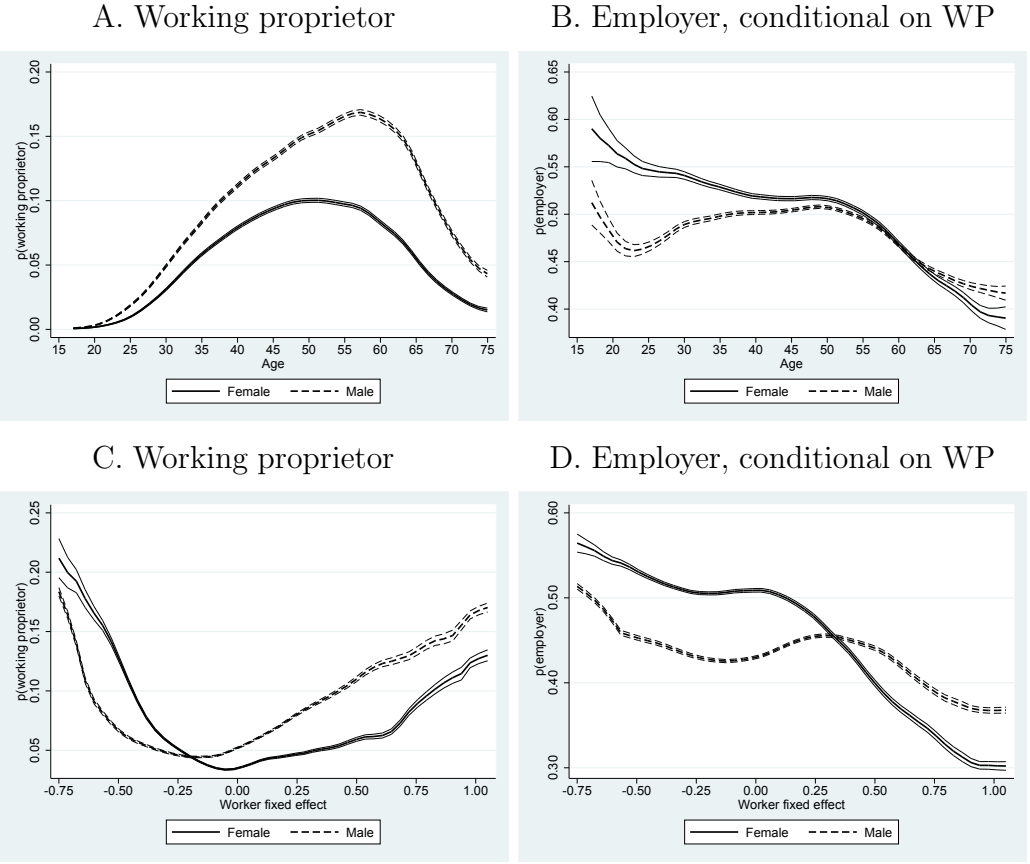
Table continued on next page

Table continued from previous page

	(1) t	(2) $t+1$	(3) $t+2$	(4) $t+3$	(5) $t+4$	(6) $t+5$
Absent from NZ: $t-1$	0.153**	0.138**	0.130**	0.120**	0.112**	0.090**
	[0.016]	[0.016]	[0.015]	[0.015]	[0.014]	[0.014]
$t-2$	-0.001	0.012	0.012	-0.002	0.005	0.002
	[0.008]	[0.008]	[0.008]	[0.008]	[0.007]	[0.007]
$t-3$	-0.003	-0.007	0.000	-0.001	-0.006	-0.007
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.006]
$t-4$	0.005	0.012	0.003	0.003	-0.004	-0.004
	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
$t-5$	-0.021**	-0.010	-0.012*	-0.020**	-0.019**	-0.021**
	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.005]
Has benefit: $t-1$	0.016*	0.001	-0.011	-0.017**	-0.020**	-0.019**
	[0.007]	[0.007]	[0.006]	[0.006]	[0.006]	[0.006]
$t-2$	-0.010	-0.013	-0.015*	-0.009	-0.012	-0.008
	[0.007]	[0.007]	[0.007]	[0.006]	[0.006]	[0.006]
$t-3$	0.004	0.001	0.002	-0.006	-0.007	-0.002
	[0.007]	[0.007]	[0.006]	[0.006]	[0.006]	[0.006]
$t-4$	0.016*	-0.002	-0.006	-0.007	-0.001	-0.011*
	[0.007]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
$t-5$	-0.012*	-0.007	-0.011*	-0.014**	-0.019**	-0.016**
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Has job: $t-1$	0.076**	0.079**	0.077**	0.071**	0.067**	0.066**
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
$t-2$	-0.003	0.000	0.003	0.007	0.009*	0.008*
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
$t-3$	0.002	0.005	0.009*	0.009*	0.011**	0.011**
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
$t-4$	-0.003	0.006	0.008	0.006	0.006	0.007
	[0.005]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
$t-5$	0.008*	0.013**	0.014**	0.017**	0.019**	0.021**
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Average firm fixed effect: $t-1$	-0.092**	-0.082**	-0.069**	-0.072**	-0.068**	-0.059**
	[0.013]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]
$t-2$	-0.022	-0.032*	-0.053**	-0.043**	-0.039**	-0.032*
	[0.015]	[0.015]	[0.015]	[0.014]	[0.014]	[0.014]
$t-3$	-0.053**	-0.066**	-0.027	-0.026	-0.029	-0.040**
	[0.016]	[0.016]	[0.016]	[0.015]	[0.015]	[0.015]
$t-4$	-0.016	0.008	-0.010	-0.018	-0.015	-0.014
	[0.017]	[0.016]	[0.016]	[0.016]	[0.015]	[0.015]
$t-5$	-0.005	-0.033*	-0.031*	-0.030*	-0.036**	-0.041**
	[0.014]	[0.014]	[0.013]	[0.013]	[0.013]	[0.012]
Average job churn	-0.001	-0.004	-0.004	-0.002	0.001	0.000
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Observations	144,618	144,618	144,618	144,618	144,618	144,618
R^2	0.049	0.053	0.052	0.053	0.054	0.055
Mean of dependent variable	0.273	0.260	0.244	0.227	0.212	0.201

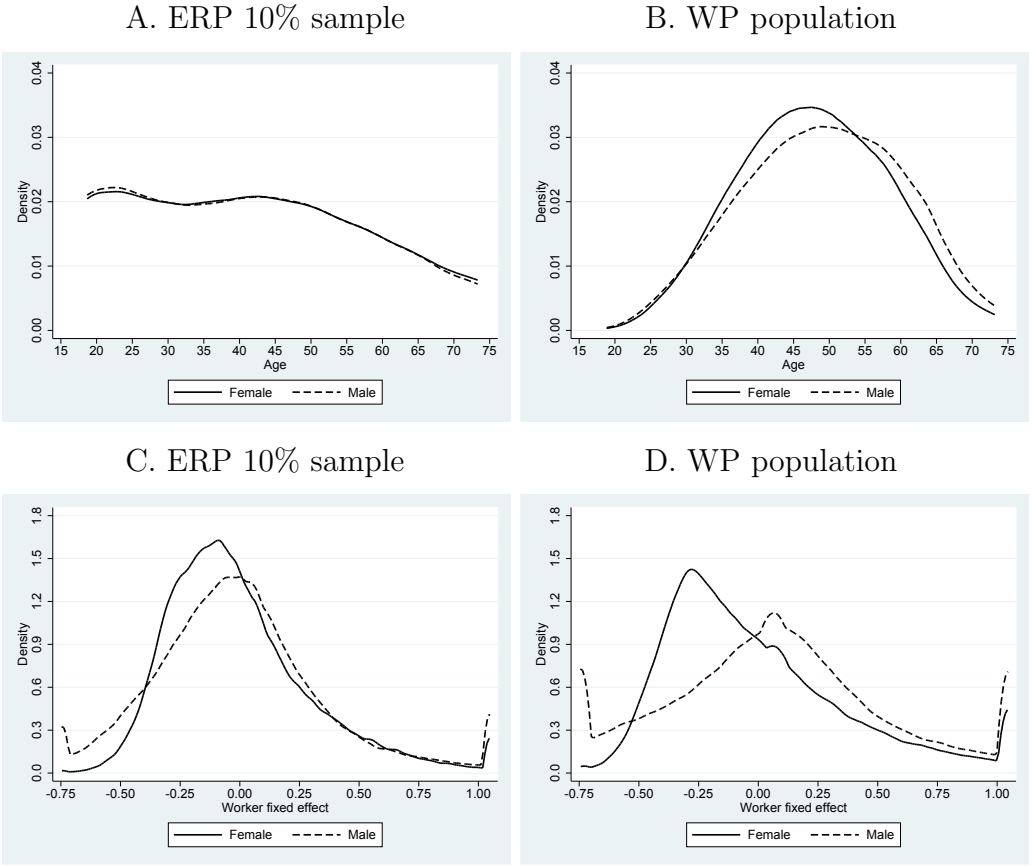
Ordinary least squares regression for population of working proprietors who enter in $t \in [2005, 2010]$. Dependent variable is an indicator variable set to one if the individual is an employer (zero otherwise) at time $t+x$. All covariates are as at $t-1$ (ie, the year prior to entry). Robust standard errors in brackets (**/* implies coefficient significantly different from zero at the 1/5% level). Regressions include year dummies, prior job industry dummies (19 ANZSIC'06 divisions), and quartics in age (a = (age−41)) and worker fixed effect. Relationships between employing at $t+x$ and age/WFE, conditional on other covariates, are reported in figures 13 and 14 respectively. Reference group is male; European-only; NZ-born; no conversational English; no formal qualification; no dependent children. Unreported indicator variables also included for missing: ethnicity; NZ-born status; conversational English; highest qualification; worker fixed effect; number of dependants. First arrival unknown is overseas-born but year of first arrival is unknown. Highest qualification post-school unknown is post-school qualification of unknown level.

Figure 1: Propensity to be entrepreneurial, by sex



Smoothed propensities, including 95 percent confidence intervals, are derived from (Epanechnikov) kernel-weighted local polynomial regressions (using Stata's default rule-of-thumb bandwidth). Panels A and C utilise the 10 percent ERP sample and the dependent variable is an indicator variable for being a working proprietor. Panels B and D utilise the WP population and the dependent variable is an indicator variable for being an employer. The worker fixed effect (WFE, panels C and D) is restricted to the specified range to comply with Stats NZ confidentiality rules. WFEs outside this range are pooled at the relevant endpoints (see figures 2C and 2D). Individuals who never have jobs in the EMS data are excluded from panels C and D.

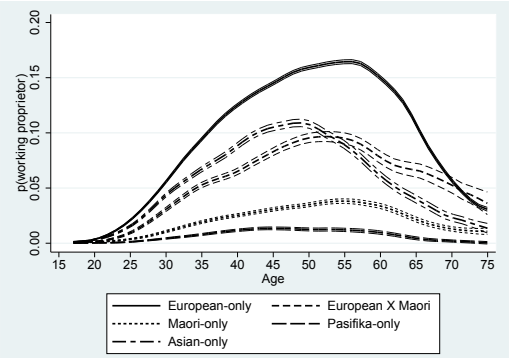
Figure 2: Population density by age, skill (WFE) and sex



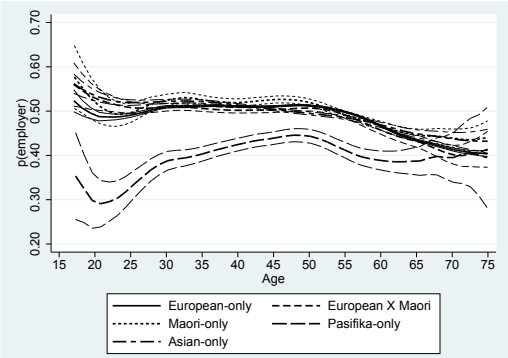
Smoothed density using Epanechnikov kernel with 100 (200) bins for age (WFE). Panels A and C utilise the 10 percent ERP sample, and panels C and D utilise the WP population. Densities for females and males are not rescaled to reflect cross-group differences in average self-employment. The worker fixed effect (WFE, panels C and D) is restricted to the specified range to comply with Stats NZ confidentiality rules. WFEs outside this range are pooled at the relevant endpoints. Individuals who never have jobs in the EMS data are excluded from panels C and D.

Figure 3: Propensity to be entrepreneurial, by ethnicity

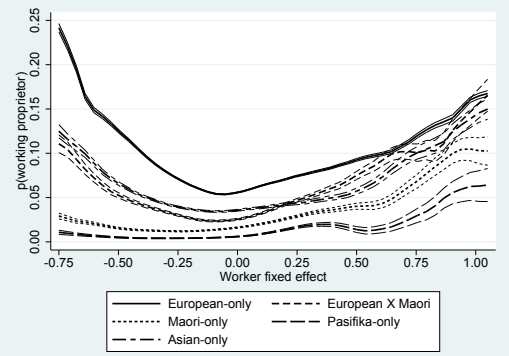
A. Working proprietor



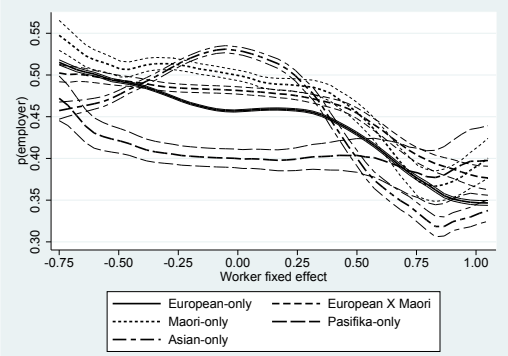
B. Employer, conditional on WP



C. Working proprietor



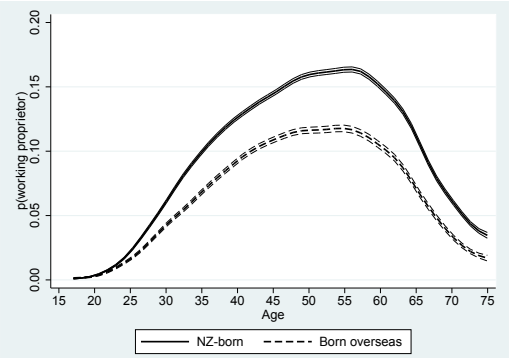
D. Employer, conditional on WP



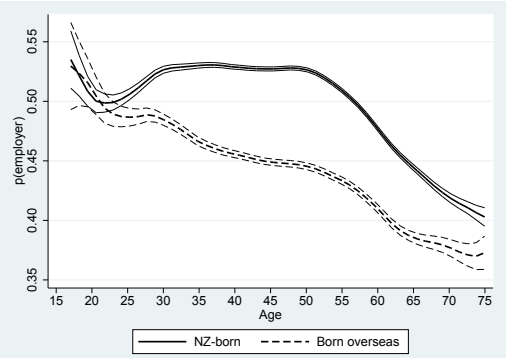
See figure 1 for notes.

Figure 4: Propensity to be entrepreneurial, by New Zealand-born

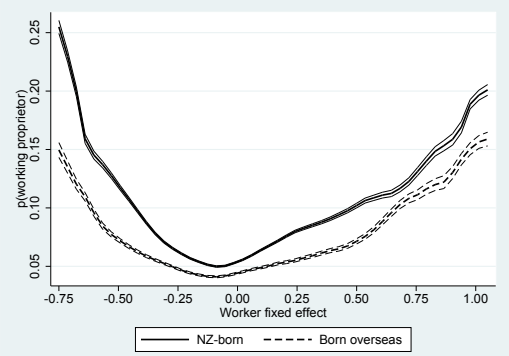
A. Working proprietor



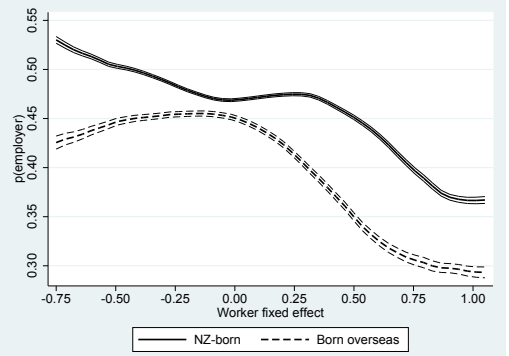
B. Employer, conditional on WP



C. Working proprietor



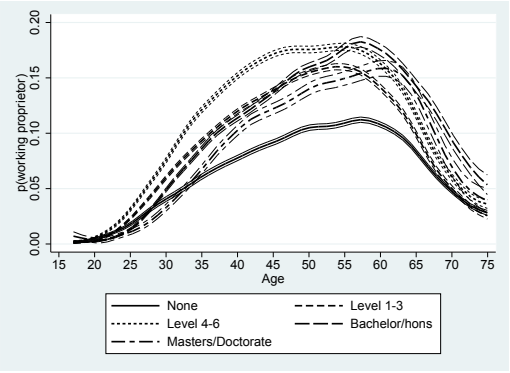
D. Employer, conditional on WP



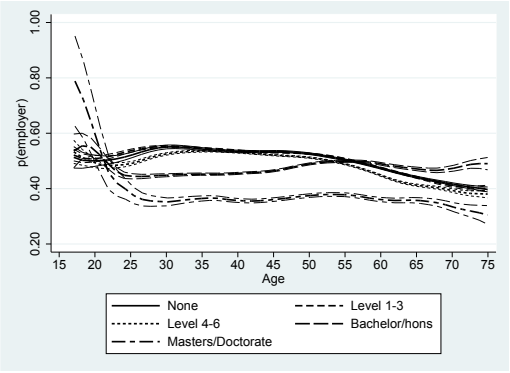
See figure 1 for notes.

Figure 5: Propensity to be entrepreneurial, by highest qualification

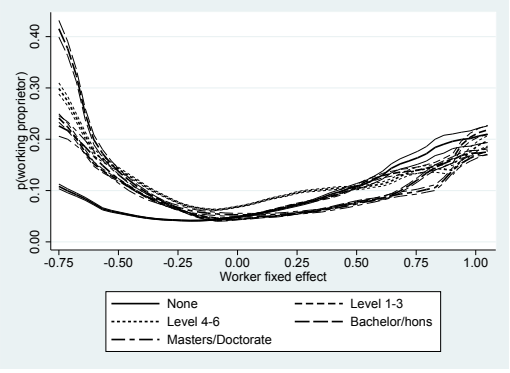
A. Working proprietor



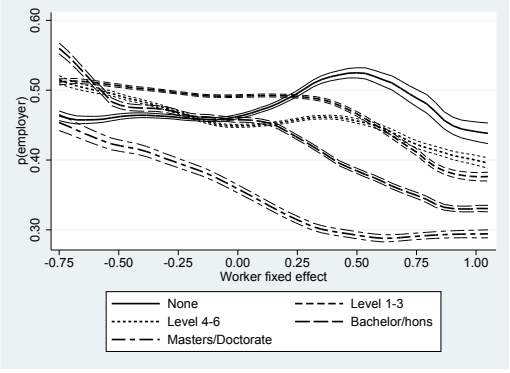
B. Employer, conditional on WP



C. Working proprietor

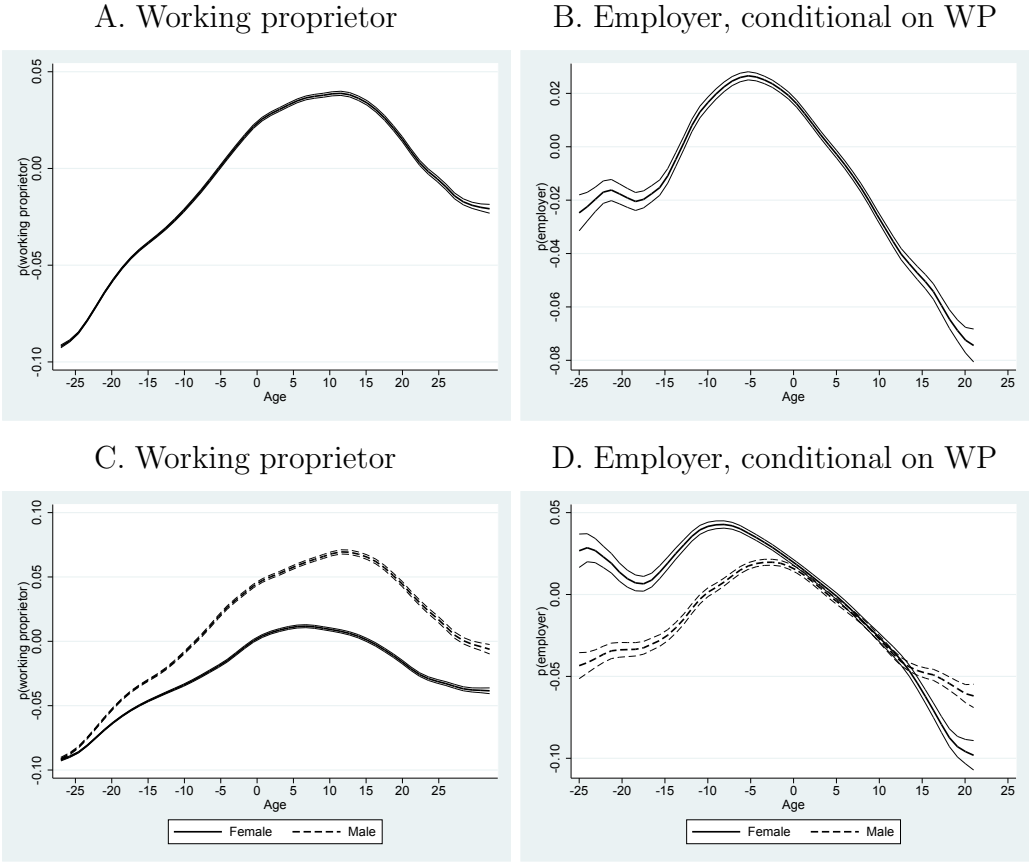


D. Employer, conditional on WP



See figure 1 for notes.

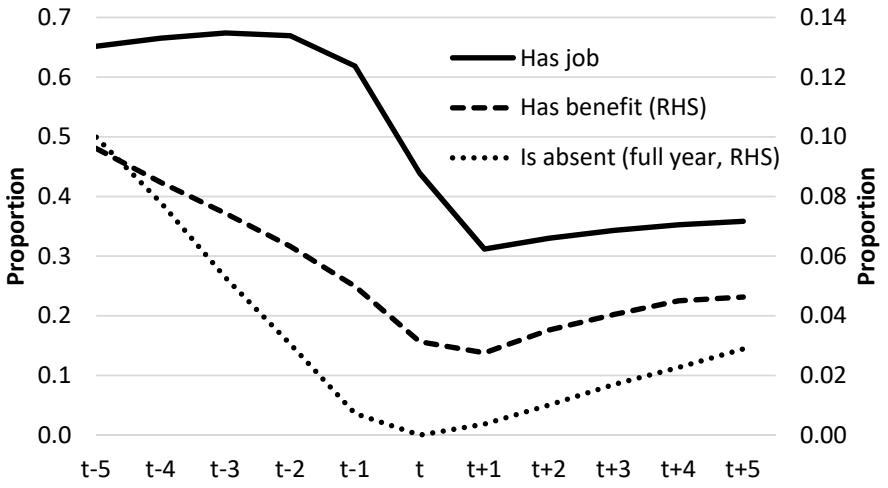
Figure 6: Estimated relationship between self-employment and age



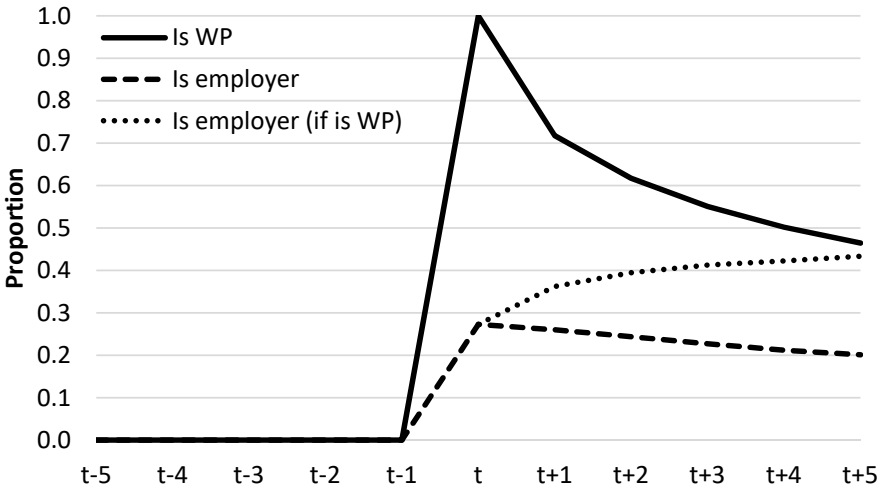
Smoothed propensities, including 95 percent confidence intervals, are derived from (Epanechnikov) kernel-weighted local polynomial regressions (using Stata's default rule-of-thumb bandwidth). Dependent and independent variables are estimated residuals from regressions on all other covariates in column (5) of tables 6 (panel A) and 7 (panel B) to approximate the estimated quartic age controls without directly imposing this functional form. Panels C and D re-estimate the smoothed propensities by sex – mimicking the inclusion of sex-specific age controls – to enable comparison with figure 1, panels A and B respectively. In panels C and D, the female indicator is excluded from the covariates so that mean difference between sexes are included in the age profiles, consistent with figure 1. The top and bottom 1% of observations of (residual) age are trimmed for confidentiality and presentation purposes.

Figure 7: Pre- and post-entry dynamics with self-employment at t

A. Job, benefit and absence from New Zealand

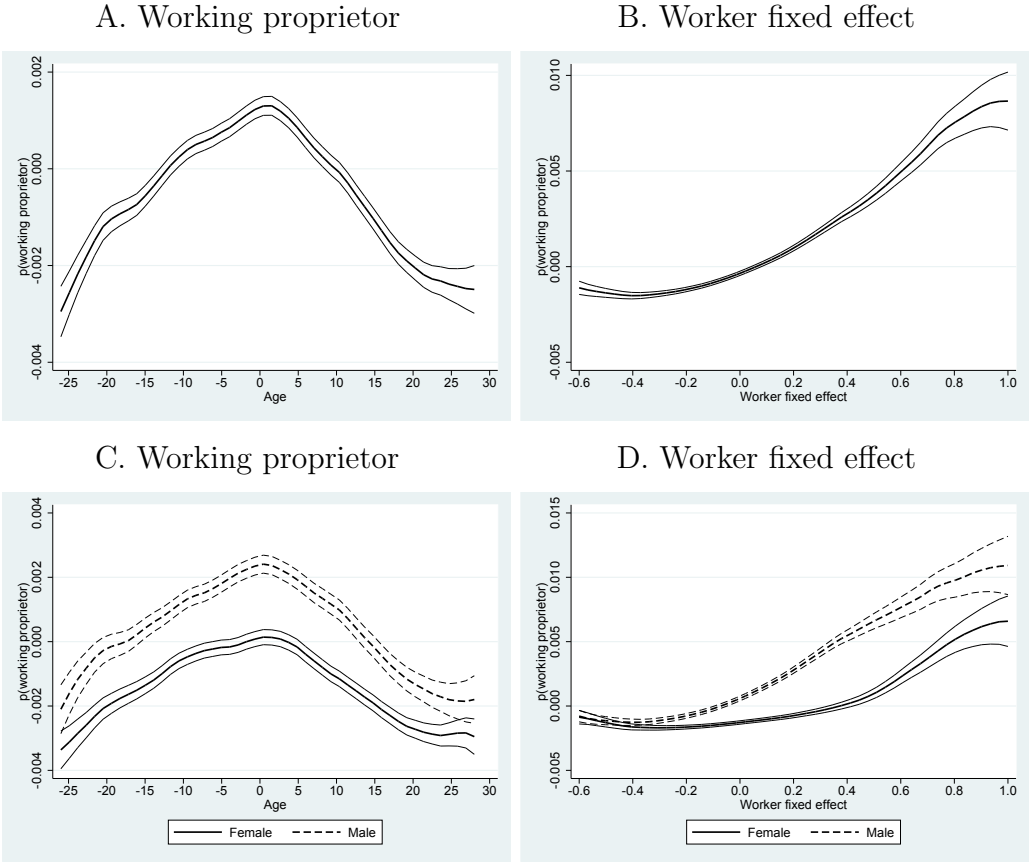


B. Self-employment and employer



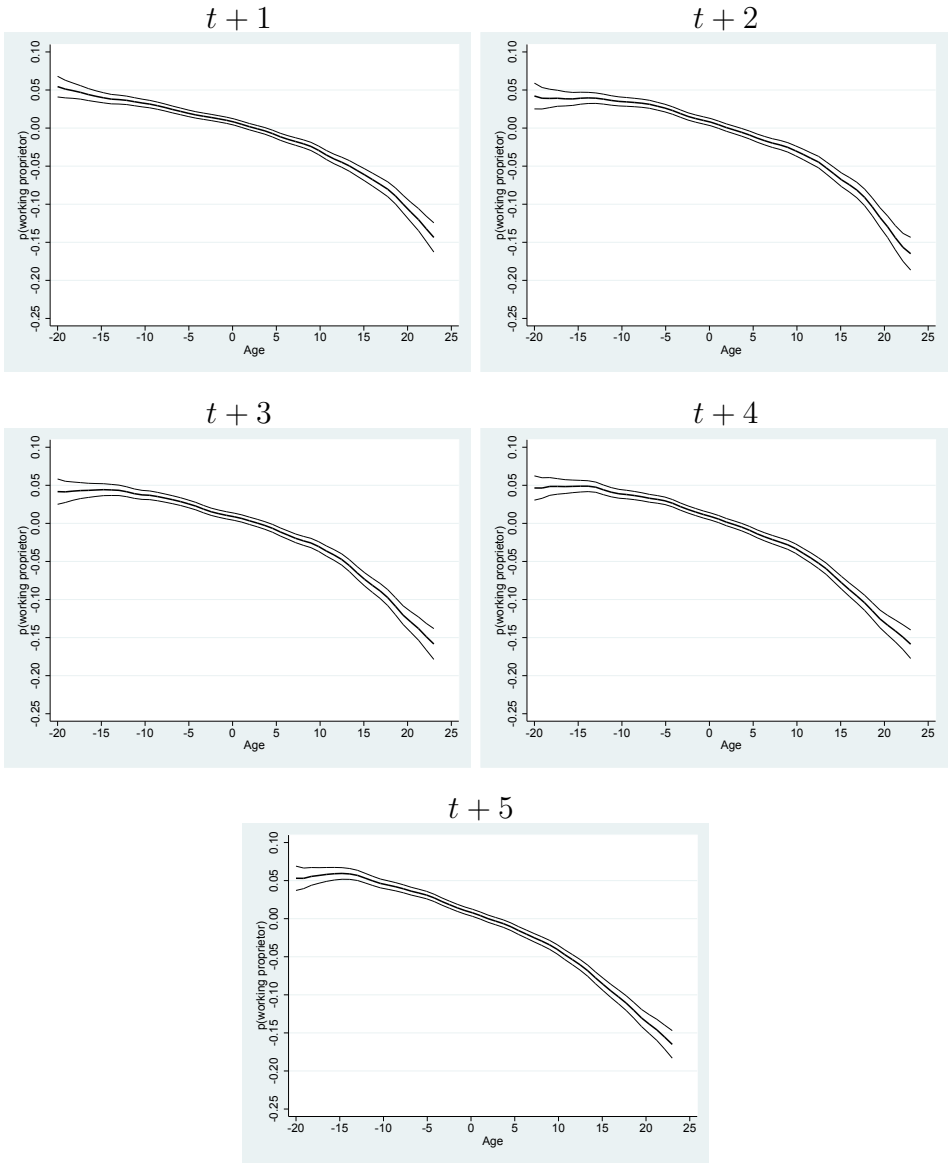
Proportions derived from population of working proprietors who enter at $t \in [2005, 2010]$. By definition, these individuals are not WPs for the five years prior to t, are a WP at time t, and cannot be absent from New Zealand for the entirety of t (which would place them outside the Estimated Resident Population). "Has job" refers to working as an employee in a firm *not* owned by the individual.

Figure 8: Estimated relationship between entry into self-employment, and age and worker fixed effect



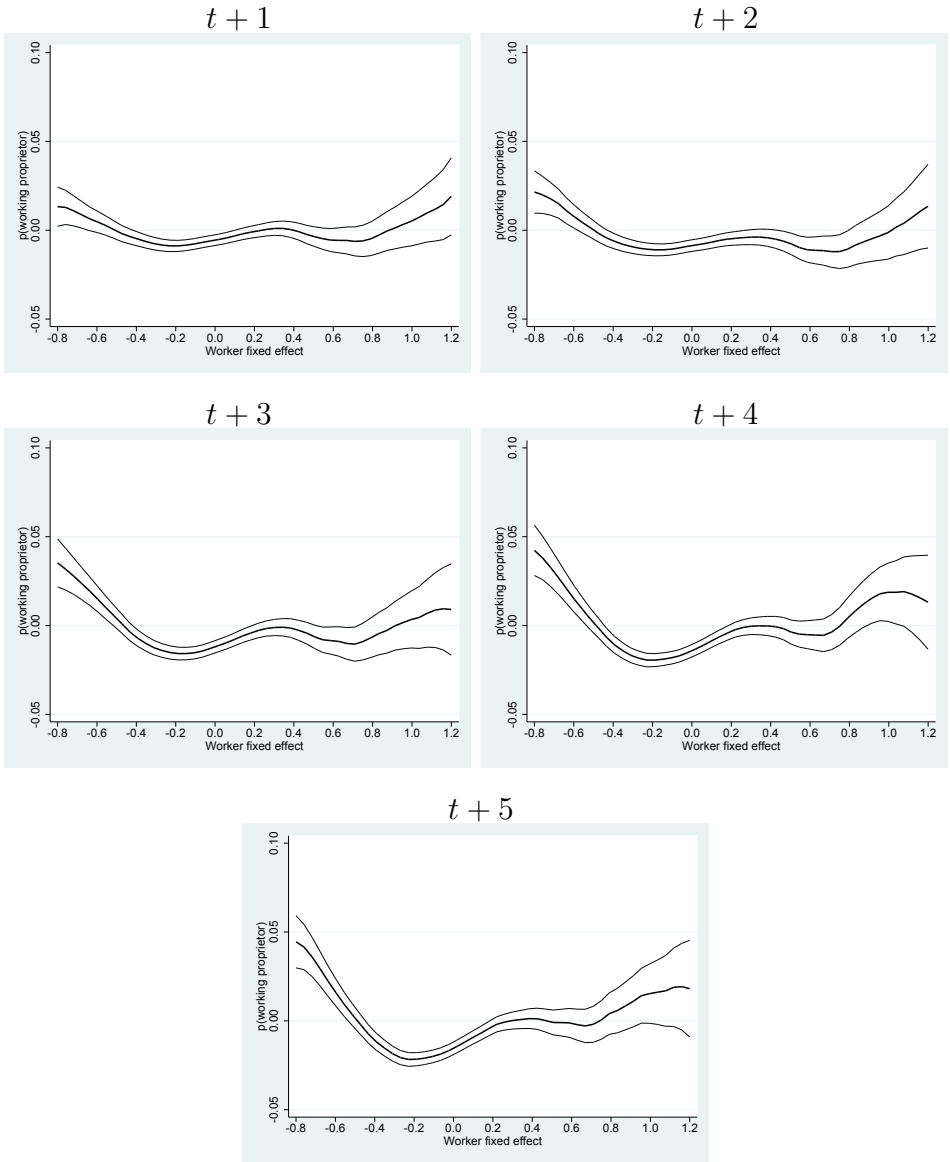
Smoothed propensities, including 95 percent confidence intervals, are derived from (Epanechnikov) kernel-weighted local polynomial regressions (using Stata's default rule-of-thumb bandwidth). Dependent and independent variables are estimated residuals from regressions on all other covariates in column (4) of table 9 to approximate the estimated quartic controls without directly imposing this functional form (panels A and B). Panels C and D re-estimate the smoothed propensities by sex – mimicking the inclusion of sex-specific age/WFE controls. In panels C and D, the female indicator is excluded from the covariates so that mean difference between sexes are included in the profiles. The top and bottom 1% of observations of (residual) age/WFE are trimmed for confidentiality and presentation purposes. Individuals who never have jobs in the EMS data are excluded from panels B and D.

Figure 9: Estimated relationship between continued self-employment and age



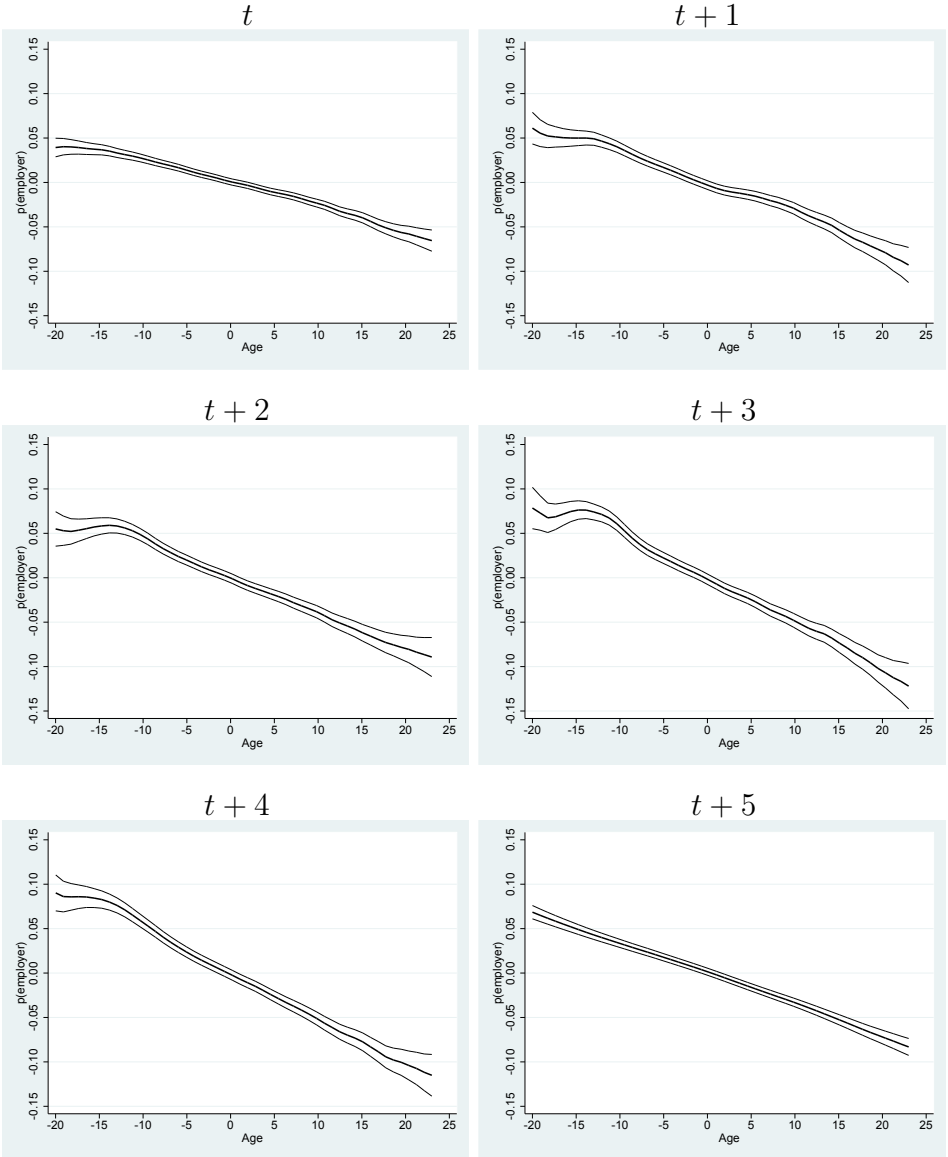
Smoothed propensities, including 95 percent confidence intervals, are derived from (Epanechnikov) kernel-weighted local polynomial regressions (using Stata's default rule-of-thumb bandwidth). Dependent and independent variables are estimated residuals from regressions on all other covariates in each column of table 11 to approximate the estimated quartic age controls at $t+x$ without directly imposing this functional form. The top and bottom 1% of observations of the independent variables are trimmed for confidentiality and presentation purposes.

Figure 10: Estimated relationship between continued self-employment and worker fixed effect



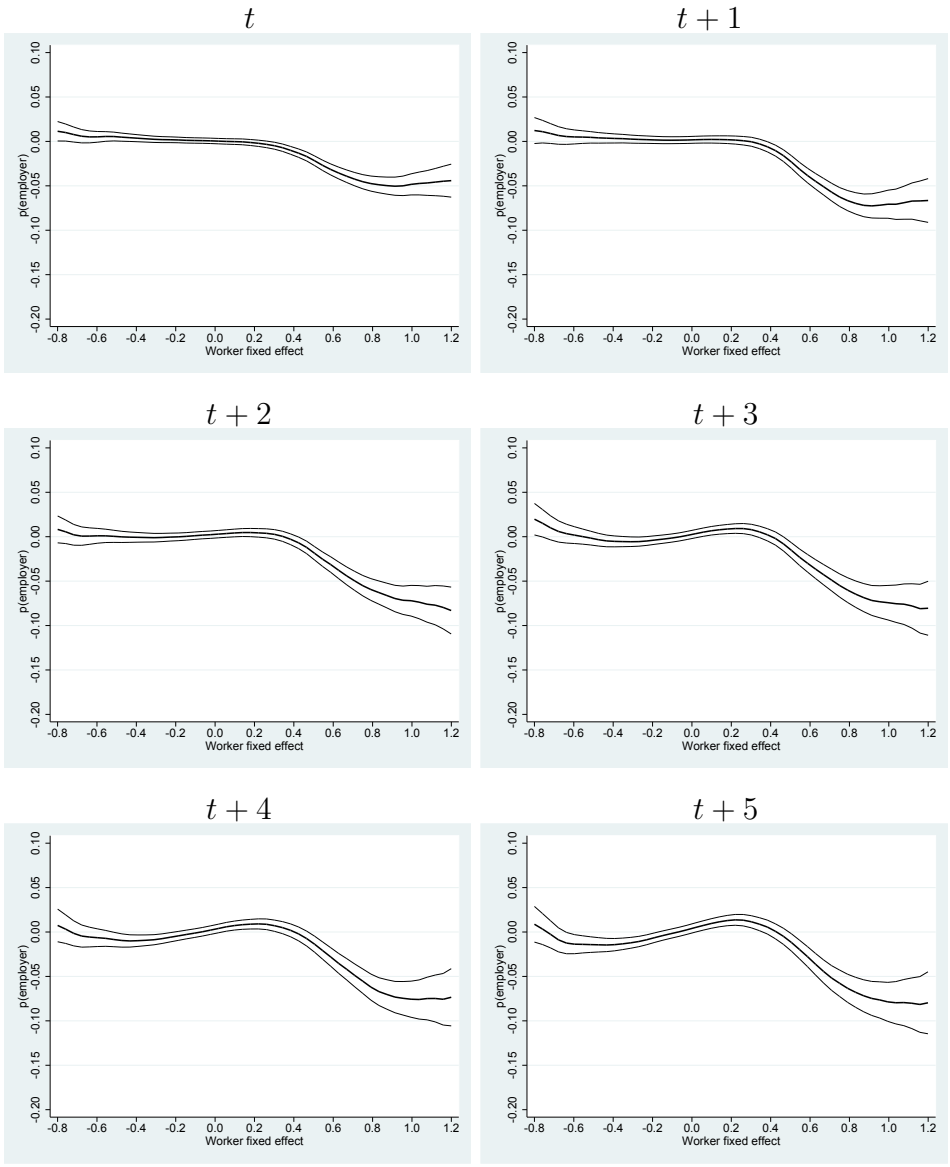
Smoothed propensities, including 95 percent confidence intervals, are derived from (Epanechnikov) kernel-weighted local polynomial regressions (using Stata's default rule-of-thumb bandwidth). Dependent and independent variables are estimated residuals from regressions on all other covariates in each column of table 11 to approximate the estimated quartic WFE controls at $t+x$ without directly imposing this functional form. The top and bottom 1% of observations of the independent variables are trimmed for confidentiality and presentation purposes. Individuals who never have jobs in the EMS data are excluded from all panels.

Figure 11: Estimated relationship between employing and age, conditional on entry at t and WP at $t + x$



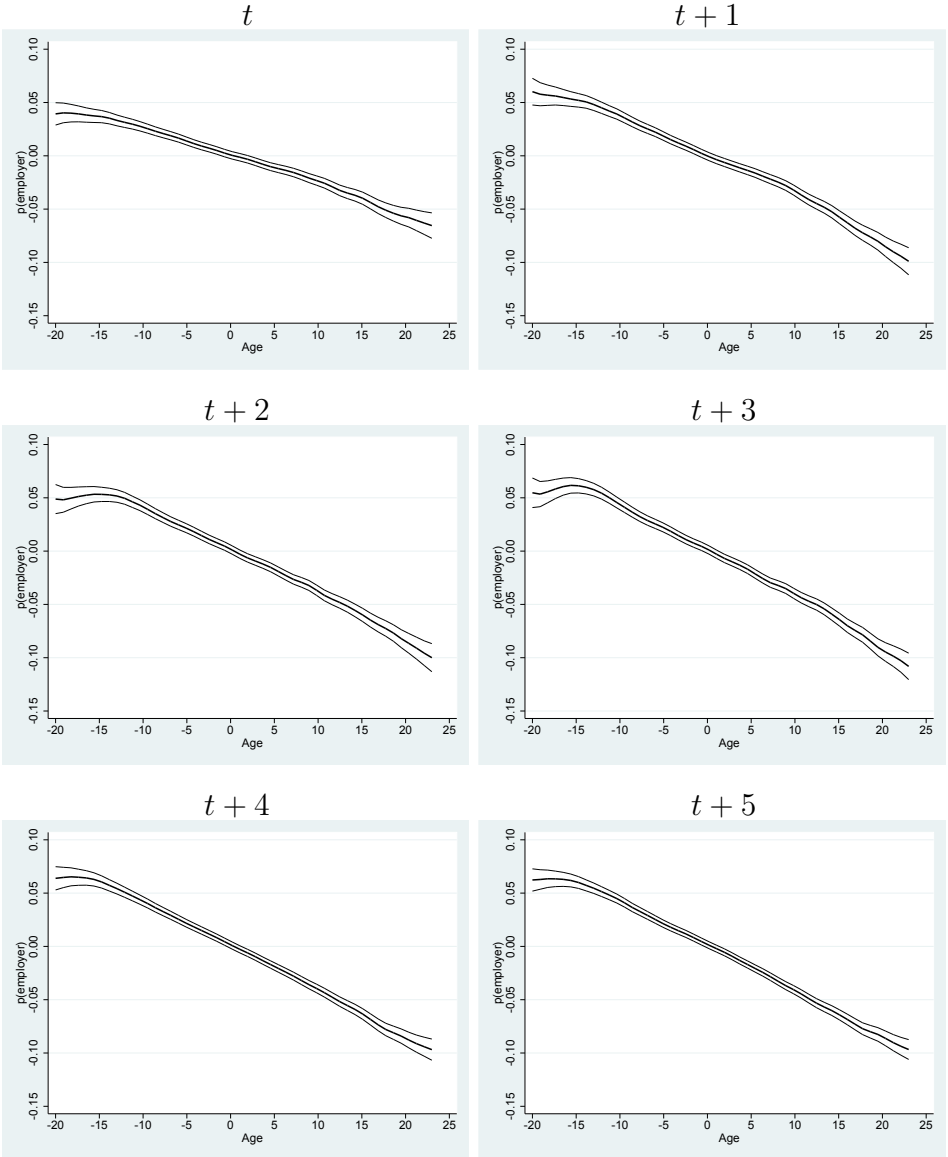
Smoothed propensities, including 95 percent confidence intervals, are derived from (Epanechnikov) kernel-weighted local polynomial regressions (using Stata's default rule-of-thumb bandwidth). Dependent and independent variables are estimated residuals from regressions on all other covariates in each column of table 12 to approximate the estimated quartic age controls at $t + x$ without directly imposing this functional form. The top and bottom 1% of observations of the independent variables are trimmed for confidentiality and presentation purposes.

Figure 12: Estimated relationship between employing and worker fixed effect, conditional on entry at t and WP at $t + x$



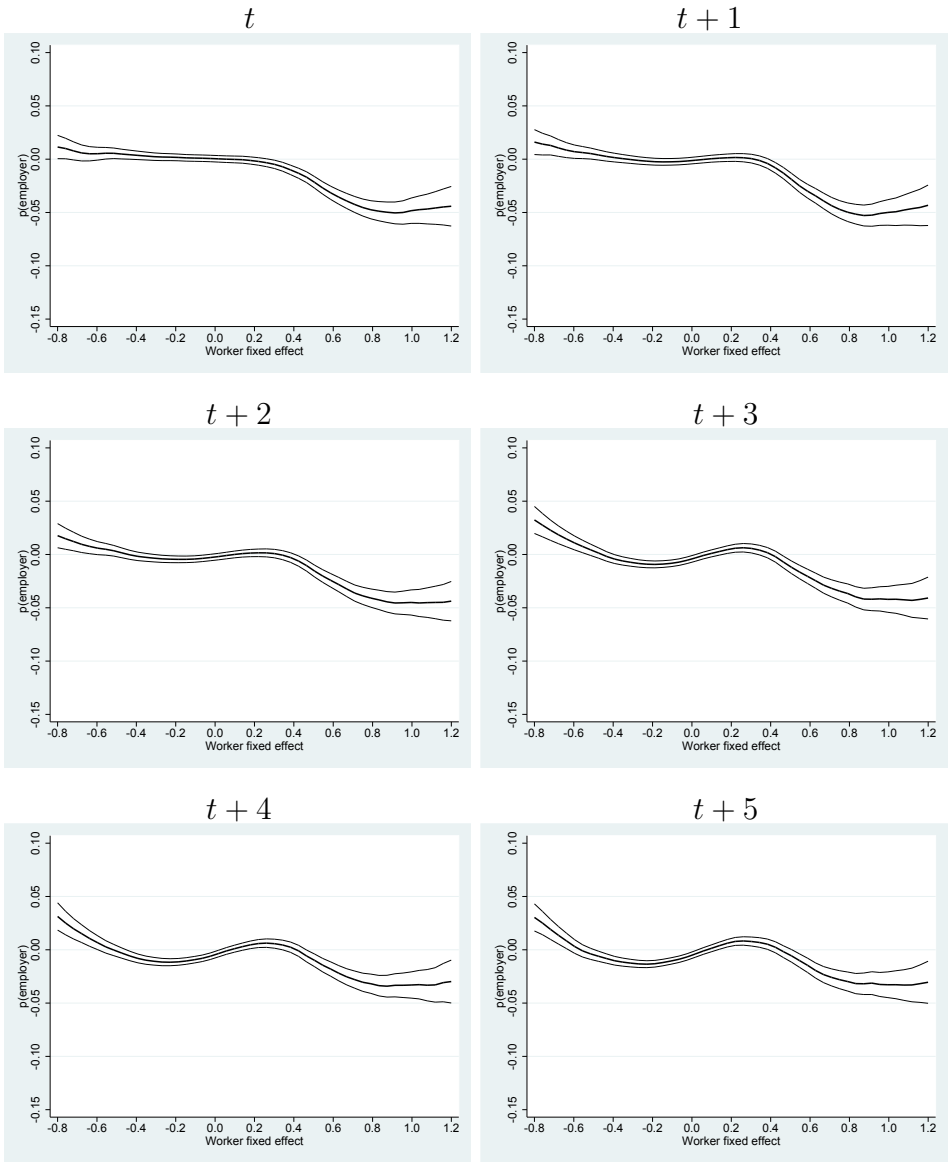
Smoothed propensities, including 95 percent confidence intervals, are derived from (Epanechnikov) kernel-weighted local polynomial regressions (using Stata's default rule-of-thumb bandwidth). Dependent and independent variables are estimated residuals from regressions on all other covariates in each column of table 12 to approximate the estimated quartic WFE controls at $t+x$ without directly imposing this functional form. The top and bottom 1% of observations of the independent variables are trimmed for confidentiality and presentation purposes. Individuals who never have jobs in the EMS data are excluded from all panels.

Figure 13: Estimated relationship between employing and age, conditional on entry at t



Smoothed propensities, including 95 percent confidence intervals, are derived from (Epanechnikov) kernel-weighted local polynomial regressions (using Stata's default rule-of-thumb bandwidth). Dependent and independent variables are estimated residuals from regressions on all other covariates in each column of table 13 to approximate the estimated quartic age controls at $t+x$ without directly imposing this functional form. The top and bottom 1% of observations of the independent variables are trimmed for confidentiality and presentation purposes.

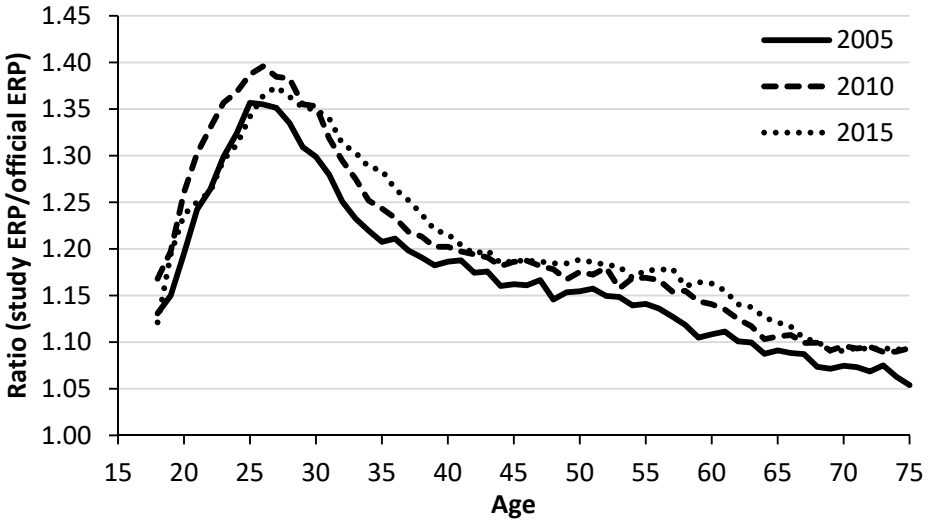
Figure 14: Estimated relationship between employing and worker fixed effect, conditional on entry at t



Smoothed propensities, including 95 percent confidence intervals, are derived from (Epanechnikov) kernel-weighted local polynomial regressions (using Stata's default rule-of-thumb bandwidth). Dependent and independent variables are estimated residuals from regressions on all other covariates in each column of table 13 to approximate the estimated quartic WFE controls at $t + x$ without directly imposing this functional form. The top and bottom 1% of observations of the independent variables are trimmed for confidentiality and presentation purposes. Individuals who never have jobs in the EMS data are excluded from all panels.

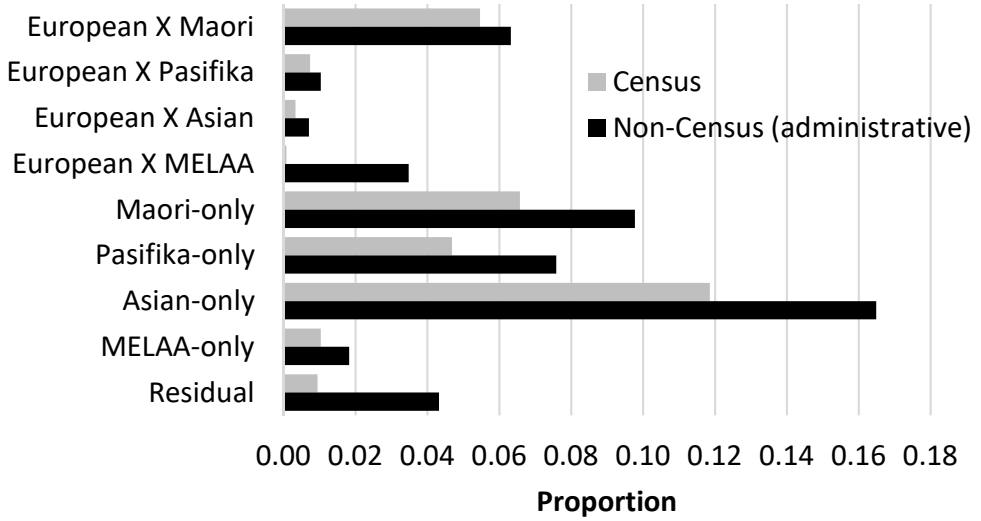
Appendix A – Extra figures

Figure A.1: Comparison of study and official ERP



Study Estimated Resident Population (ERP) includes all individuals present in New Zealand for any day during the relevant March year, whereas official statistic are for those present during the (March) reference month.

Figure A.2: Comparison of Level 1 ethnicity by data source (2015 year only)



Analysis is restricted to the Estimated Resident Population (ERP) with non-missing ethnicity. Non-Census (administrative) data is for individuals without a Census response to the ethnicity question and is taken from the IDI source-ranked ethnicity table. European-only group excluded for scale reasons – the Census (non-Census) European-only ethnicity share is 68.3% (48.5%).

Appendix B – Extra tables

Table B.1: Sex-specific coefficients for number of dependants

	(1)	(2)	(3)
Specification [table.column]	[T6.5]	[T7.5]	[T9.1]
Female	-0.049**	0.013**	-0.00349**
	[0.001]	[0.002]	[0.00013]
Male: One dependant (0-4yr)	0.027**	0.019**	0.00760**
	[0.002]	[0.003]	[0.00062]
2+ (0-4yr)	0.055**	0.060**	0.01061**
	[0.003]	[0.003]	[0.00095]
One (5-8yr)	0.040**	0.030**	0.00356**
	[0.003]	[0.003]	[0.00063]
2+ (5-8yr)	0.053**	0.059**	0.00420**
	[0.004]	[0.004]	[0.00115]
2+ (mixed)	0.052**	0.070**	0.00503**
	[0.003]	[0.003]	[0.00067]
Female: One dependant (0-4yr)	0.022**	0.042**	0.00328**
	[0.001]	[0.004]	[0.00040]
2+ (0-4yr)	0.044**	0.061**	0.00556**
	[0.002]	[0.004]	[0.00064]
One (5-8yr)	0.031**	0.022**	0.00136**
	[0.002]	[0.003]	[0.00042]
2+ (5-8yr)	0.049**	0.041**	0.00386**
	[0.003]	[0.005]	[0.00087]
2+ (mixed)	0.045**	0.062**	0.00283**
	[0.002]	[0.004]	[0.00047]
Observations	4,010,454	3,005,475	3,567,669
R^2	0.060	0.012	0.005
Mean of dependent variable	0.075	0.494	0.007

Supplemental OLS regression coefficients for individual specifications in tables 6, 7 and 9 as indicated in top panel of table. In each case, the main table specification is changed to allow the coefficients on dependants to differ by sex (all other independent variables remain as in the original tables). Coefficients on dependants, together with the coefficient on the female indicator variable, are reported though the latter is not directly comparable to the main specification because of the inclusion of sex-specific coefficients on dependants. See footnotes to main tables for further information.

Table B.2: Distribution of WPs by industry

	Proportion of WPs with firm in industry												
	A	C	E	F	G	H	I	L	M	N	Q	S	Other
2005	0.249	0.071	0.153	0.043	0.084	0.051	0.041	0.061	0.119	0.031	0.036	0.050	0.043
2006	0.237	0.070	0.159	0.043	0.083	0.051	0.041	0.059	0.126	0.032	0.037	0.050	0.044
2007	0.229	0.069	0.162	0.042	0.081	0.050	0.041	0.057	0.132	0.034	0.039	0.049	0.045
2008	0.224	0.068	0.165	0.040	0.078	0.049	0.041	0.055	0.137	0.034	0.041	0.049	0.046
2009	0.218	0.067	0.165	0.040	0.076	0.048	0.041	0.050	0.144	0.035	0.044	0.049	0.047
2010	0.215	0.066	0.161	0.040	0.076	0.049	0.040	0.051	0.145	0.035	0.046	0.050	0.048
2011	0.213	0.064	0.161	0.039	0.076	0.050	0.040	0.051	0.149	0.036	0.046	0.050	0.048
2012	0.212	0.064	0.161	0.039	0.074	0.049	0.040	0.051	0.151	0.036	0.046	0.050	0.049
2013	0.210	0.063	0.163	0.038	0.073	0.049	0.039	0.050	0.153	0.037	0.047	0.050	0.049
2014	0.213	0.062	0.167	0.037	0.071	0.048	0.039	0.050	0.153	0.037	0.047	0.050	0.048
2015	0.211	0.062	0.168	0.037	0.070	0.048	0.038	0.049	0.154	0.036	0.048	0.050	0.048

Industry is ANZSIC'06 division: A – Agriculture, forestry & fishing; B – Mining; C – Manufacturing; D – Electricity, gas, water & waste services; E – Construction; F – Wholesale trade; G – Retail trade; H – Accommodation; I – Transport, postal & warehousing; J – Information media & telecommunications; K – Financial & insurance services; L – Rental, hiring & real estate services; M – Professional, scientific & technical services; N – Administrative & support services; O – Public administration & safety; P – Education & training; Q – Health care & social assistance; R – Arts & recreation services; S – Other services. Column labelled "Other" aggregates the seven ANZSIC divisions not separately itemised (B, D, J, K, O, P, R), each of which individually contribute less than 2% of the total. Denominator is total number of private-for-profit WPs in ERP/age-restricted population (column 1, table 4). Row sums add to more than one because WPs can own multiple firms during a year, resulting in double-counting of WPs when owned firms are in different industries.

Table B.3: Time-varying coefficients for number of dependants

	(1) $t+1$	(2) $t+2$	(3) $t+3$	(4) $t+4$	(5) $t+5$
			Table 11		
Female	0.006*	0.001	-0.005	-0.007*	-0.008**
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Dependants: One (0-4yr)	0.029**	0.046**	0.053**	0.052**	0.044**
(at $t+x-1$)	[0.004]	[0.005]	[0.005]	[0.006]	[0.006]
2+ (0-4yr)	0.038**	0.077**	0.106**	0.116**	0.109**
	[0.006]	[0.006]	[0.006]	[0.006]	[0.007]
One (5-8yr)	0.018**	0.029**	0.036**	0.038**	0.041**
	[0.005]	[0.006]	[0.006]	[0.006]	[0.006]
2+ (5-8yr)	0.042**	0.045**	0.061**	0.064**	0.042**
	[0.008]	[0.009]	[0.009]	[0.009]	[0.010]
2+ (mixed)	0.042**	0.050**	0.065**	0.081**	0.101**
	[0.005]	[0.005]	[0.006]	[0.006]	[0.006]
			Table 12		
Female	0.050**	0.049**	0.048**	0.048**	0.047**
	[0.003]	[0.004]	[0.004]	[0.004]	[0.004]
Dependants: One (0-4yr)	0.036**	0.052**	0.044**	0.044**	0.044**
(at $t+x-1$)	[0.006]	[0.006]	[0.007]	[0.007]	[0.009]
2+ (0-4yr)	0.078**	0.073**	0.073**	0.077**	0.063**
	[0.007]	[0.008]	[0.008]	[0.008]	[0.009]
One (5-8yr)	0.026**	0.024**	0.026**	0.027**	0.043**
	[0.007]	[0.007]	[0.007]	[0.008]	[0.008]
2+ (5-8yr)	0.039**	0.052**	0.085**	0.073**	0.068**
	[0.011]	[0.011]	[0.012]	[0.011]	[0.013]
2+ (mixed)	0.061**	0.068**	0.073**	0.068**	0.065**
	[0.007]	[0.007]	[0.007]	[0.007]	[0.008]
			Table 13		
Female	0.039**	0.031**	0.025**	0.021**	0.018**
	[0.003]	[0.002]	[0.002]	[0.002]	[0.002]
Dependants: One (0-4yr)	0.040**	0.058**	0.054**	0.053**	0.048**
(at $t+x-1$)	[0.005]	[0.005]	[0.005]	[0.005]	[0.006]
2+ (0-4yr)	0.074**	0.084**	0.099**	0.107**	0.095**
	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
One (5-8yr)	0.027**	0.027**	0.030**	0.030**	0.042**
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
2+ (5-8yr)	0.044**	0.051**	0.079**	0.071**	0.056**
	[0.009]	[0.009]	[0.009]	[0.008]	[0.008]
2+ (mixed)	0.062**	0.068**	0.076**	0.080**	0.088**
	[0.006]	[0.005]	[0.005]	[0.005]	[0.005]

Supplemental OLS regression coefficients for each $t+x$ specification in tables 11-13 as indicated in each panel of the table. In each case, the main table specification is changed to allow the coefficients on dependants to be time-varying (ie, at $t+x-1$ values) rather than fixed at $t-1$ values as in the main table (all other independent variables remain and are held at pre-entry values as in the original tables). Coefficients on dependants are reported, together with the coefficients on the female indicator variable (which are directly comparable to main specification estimates). See footnotes to main tables for further information.

Recent Motu Working Papers

All papers in the Motu Working Paper Series are available on our website www.motu.nz, or by contacting us on info@motu.org.nz or +64 4 939 4250.

18-12 Fabling, Richard. 2018. "Entrepreneurial beginnings: Transitions to self-employment and the creation of jobs."

18-11 Fleming, David A and Kate Preston. 2018. "International agricultural mitigation research and the impacts and value of two SLMACC research projects." (also a Ministry for Primary Industries publication)

18-10 Hyslop, Dean and David Rea. 2018. "Do housing allowances increase rents? Evidence from a discrete policy change."

18-09 Fleming, David A., Ilan Noy, Jacob Pástor-Paz and Sally Owen. 2018. "Public insurance and climate change (part one): Past trends in weather-related insurance in New Zealand."

18-08 Sin, Isabelle, Kabir Dasgupta and Gail Pacheco. 2018. "Parenthood and labour market outcomes." (also a Ministry for Women Report)

18-07 Grimes, Arthur and Dennis Wesselbaum. 2018. "Moving towards happiness."

18-06 Qasim, Mubashir and Arthur Grimes. 2018. "Sustainable economic policy and well-being: The relationship between adjusted net savings and subjective well-being."

18-05 Clay, K Chad, Ryan Bakker, Anne-Marie Brook, Daniel W Hill Jr and Amanda Murdie. 2018. "HRMI Civil and Political Rights Metrics: 2018 Technical Note."

18-04 Apatov, Eyal, Nathan Chappell and Arthur Grimes. 2018. "Is internet on the right track? The digital divide, path dependence, and the rollout of New Zealand's ultra-fast broadband." (forthcoming)

18-03 Sin, Isabelle, Eyal Apatov and David C Maré. 2018. "How did removing student allowances for postgraduate study affect students' choices?"

18-02 Jaffe, Adam B and Nathan Chappell. 2018. "Worker flows, entry, and productivity in New Zealand's construction industry."

18-01 Harris, Richard and Trinh Le. 2018. "Absorptive capacity in New Zealand firms: Measurement and importance."

17-15 Sin, Isabelle, Steven Stillman and Richard Fabling. 2017. "What drives the gender wage gap? Examining the roles of sorting, productivity differences, and discrimination."

17-14 MacCulloch, Robert. 2017. "Political systems, social welfare policies, income security and unemployment."

17-13 Fleming, David A., Arthur Grimes, Laurent Lebreton, David C Maré and Peter Nunns. 2017. "Valuing sunshine."

17-12 Hyslop, Dean and Wilbur Townsend. 2017. "The longer term impacts of job displacement on labour market outcomes."

17-11 Carver, Thomas, Patrick Dawson and Suzi Kerr. 2017. "Including forestry in an Emissions Trading Scheme: Lessons from New Zealand."

17-10 Daigneault, Adam, Sandy Elliott, Suzie Greenhalgh, Suzi Kerr, Edmund Lou, Leah Murphy, Levente Timar and Sanjay Wadhwa. 2017. "Modelling the potential impact of New Zealand's freshwater reforms on land-based Greenhouse Gas emissions"

17-09 Coleman, Andrew. 2017. "Housing, the 'Great Income Tax Experiment', and the intergenerational consequences of the lease"

17-08 Preston, Kate and Arthur Grimes. 2017. "Migration and Gender: Who Gains and in Which Ways?"

17-07 Grimes, Arthur, Judd Ormsby and Kate Preston. 2017. "Wages, Wellbeing and Location: Slaving Away in Sydney or Cruising on the Gold Coast."

17-06 Leining, Catherine, Judd Ormsby and Suzi Kerr. 2017. "Evolution of the New Zealand Emissions Trading Scheme: Linking."


Motu