

EARNINGS HETEROGENEITY AND JOB MATCHING – EVIDENCE FROM LINKED EMPLOYER-EMPLOYEE DATA

David C. Maré

*Motu Economic and Public Policy Research,
Wellington*

Dean R. Hyslop

*Policy Support and Analysis Section
The Treasury, Wellington and
Statistics New Zealand*

Abstract

This paper uses data from Statistics New Zealand's linked employer employee database (LEED) over the six year period April 1999–March 2005 to derive and analyse estimates of two-way worker and firm fixed effects components of job earnings rates. The fixed effects estimates reflect the portable earnings premium that each worker receives in whichever firm they work for, and the time-invariant premium that each firm pays to all the workers it employs. We focus on three issues. First, how much of the variation in job earnings rates is attributable to observable worker demographic factors (age and gender), unobserved worker effects and unobserved firm effects? Second, how much compositional change occurred during this period of substantial employment growth? Third, what is the aggregate pattern of sorting of workers and firms across jobs?

Introduction

Workers' earnings rates may vary because of systematic differences across workers and/or differences across the firms they work for. Disentangling the sources of variation in earnings is important for understanding in several areas, including earnings inequality, productivity differences across firms, the impacts of alternative remuneration policies, etc. In addition, if there are complementarities between workers and firms, then we would expect high-earning workers to be concentrated in high-paying firms. Assessing the relative importance of alternative sources of variation in job earnings, as well as the strength of the matching, requires the simultaneous estimation of unobserved worker and firm effects from longitudinal job-level data that allow workers and firms to be linked (Abowd and Kramarz, 1999).

This paper uses linked employer-employee earnings data to address these two sets of issues for the first time in New Zealand, using Statistics New Zealand's Linked Employer-Employee Data (LEED) over the period 1999–2005.

The primary objective of the paper is to estimate the regression-adjusted joint (two-way) worker and firm fixed effects associated with the full-time equivalent (FTE) annualised job earnings rate, controlling for worker observed (age and sex) demographic differences. We first document the cross-sectional variation in job-

earnings rates over the sample, and then consider various econometric identification and estimation issues associated with estimating the worker and firm effects of interest.

Our second objective is to describe and analyse the patterns of interaction between these estimated components of the job earnings rate. In this analysis we focus on three broad issues: the degree of worker and firm earnings heterogeneity, compositional changes over the period, and the strength of assortative matching of workers and firms across jobs.

First, we document the degree to which earnings variability across jobs reflects differences in workers' observed demographics, versus unobserved systematic worker effects, firm effects and/or idiosyncratic worker-firm job effects? Second, we describe the effects of compositional change over a period of strong employment growth, by focusing on changes in mean firm and person effects. Third, we document the degree of assortative matching between workers and firms, both across the full population and within various subgroups. Our main focus here is on the extent to which high-earning workers work for high-paying firms. A positive correlation between worker and firm effects can arise if there are sufficient complementarities in production.¹

The paper is organised as follows. In the next section we provide a brief overview of the LEED and a discussion of

the derivation of the variables we use in the analysis. In the following section, we outline the econometric framework adopted, and discuss various identification and estimation issues encountered. We then discuss key results on the questions of heterogeneity, matching, and compositional change, before concluding with a summary discussion and suggestions of future research directions.

Data

The analysis in this paper uses Statistics New Zealand's Linked Employer-Employee Data (LEED), which uses information from tax and statistical sources to construct a record of paid jobs. Since April 1999, all employers in New Zealand are required to file a monthly record with Inland Revenue (IRD) called an Employer Monthly Schedule (EMS), which lists all paid employees at that firm during the month, the earnings they received and the amount of tax that was deducted at source. Two types of recipients are covered by EMS: those who have Pay-As-You-Earn (PAYE) tax deducted, who are employees; and those who pay withholding tax, who are a subset of the self-employed. Because the selection and coverage of which self-employed workers have tax withheld is unknown, we use only information on PAYE-deducted (employee) jobs.¹ We use all the available data on PAYE employee jobs in New Zealand during the six March-years from April 1999 to March 2005.

Firms (employers) and workers (employees) are identified by unique confidentialised identifiers based on their respective IRD tax numbers. For workers, this represents a single identifier over time, enabling workers to be tracked longitudinally and across the firms that they work for. Employers are identified as the administrative unit to which the EMS return relates, which may change for legal and/or other administrative reasons without any change in the economic structure of the 'firm'. We use a version of the LEED that has allocated EMS returns to geographic units, identified by a unique identifier (the Permanent Business Number, PBN) in Statistics New Zealand's Longitudinal Business Frame (LBF) (Seyb, 2003), and adopt such geographic units as our concept of firms.

Conceptually, the LEED covers the universe of PAYE employment relationships and earnings in New Zealand over the period. In addition, there is limited information on the characteristics of workers and firms: age, sex, and location of workers; and industry and location of firms. Perhaps the main weakness of the LEED for the current analysis is that the EMS returns report only monthly earnings for each employee and contains no information on hours worked. As a result, we cannot accurately distinguish hourly wages versus hours worked variation on earnings – e.g. low monthly earnings may be due to either a low hourly wage and/or low hours worked.

In order to provide a partial adjustment for the lack of hours information, we develop an algorithm to estimate each worker's relative employment intensity. This algorithm takes into account both the worker's monthly LEED employment earnings and any earnings-tested income they receive from other sources; the algorithm also allocates their total employment across their (multiple) jobs. We first assume that each worker can have up to one unit of employment intensity in any month, and their employment is zero in any month that they have no LEED earnings. We next scale down a worker's monthly employment intensity if either their total monthly earnings are less than full-time minimum wage earnings, and/or they receive any earnings-tested LEED 'non-work payments' income. Specifically, we estimate individual-*i*'s employment intensity in month-*m*, e_{im} , as

$$e_{im} = \min \left\{ 1, \frac{earn_{im}}{(earn_{im} + non_earn_{im})}, \frac{earn_{im}}{FT_mw_earn_{im}} \right\} \quad (1)$$

where $earn_{im}$ is *i*'s total LEED employment earnings in month-*m*, non_earn_{im} is their total (earnings-tested) non-work income in month-*m*, and $FT_mw_earn_{im}$ is the full-time minimum wage earnings level applicable to them in month-*m*. As hourly wages generally exceed both minimum wages and non-work income rates, these adjustments likely overstate the employment intensity of part-time workers and those receiving non-work payments relative to full-time workers.¹

For workers with multiple jobs in a month, their total monthly employment intensity (or "effective employment") is allocated across the jobs they held in that month in proportion to the earnings from each job, to give their effective monthly employment in those jobs. Job earnings and effective employment can be aggregated over a year, and we estimate the full-time-equivalent (FTE) annual earnings rate as the relevant annual earnings divided by estimated annual effective employment.

Table 1 provides a summary of the annual data: for all years pooled, the first year (1999/2000), the last year (2004/05) and the percentage change between the first and last years. Over the six-year period, there are 18,676,300 distinct job-year observations and 8,018,300 FTE job-years associated with 9,729,900 jobs, worked by 2,776,400 workers (employees) in 322,700 firms (PBNs). On average over the six-year period, workers have 3.5 jobs (different firms), firms employ about 30 different workers, and there are 1.9 annual observations per job. In addition, there are 11,716,400 worker-year observations (on average 4.2 per worker) and 1,211,202 firm-year observations (3.8 per firm). The sample period was one of strong employment growth. The number of annual job observations, annual job FTE, number of workers and number of firms increased 14, 17, 16 and 10 percent, respectively, between the first and last years.

Table 1: Sample characteristics.

	Pooled Years	1999/2000	2004/05	Percentage Change
No. Worker-yr Obs	11,716,402	1,825,411	2,110,895	15.6
No. of Firm-yr Obs	1,211,202	195,039	213,760	9.6
No. Job-yr Obs	18,676,324	2,920,760	3,334,000	14.1
No. FTE Job-yr Obs	8,018,349	1,239,560	1,452,851	17.2
FTE Employment	0.782	0.781	0.784	0.5
< FT Employment	0.451	0.459	0.457	-0.4
FTE Earnings	\$44,077	\$43,680	\$45,481	4.1
Age	38.0	37.2	38.6	3.9
Female	0.462	0.462	0.461	-0.1

Note: 2,776,361 workers, 322,713 firms and 9,729,904 jobs observed

Years are April-March – e.g. 2000 refers to April 1999-March 2000. All income values are in December quarter 2005 \$ values, adjusted using the Consumers Price Index (CPI).

The focus of our analysis is on FTE-weighted statistics. The average job-year FTE employment is 0.78, and 45 percent of job-year observations involve less than full-time employment during the year. The employment weighted average age of workers was 38 years (and increased 1.4 years or 4 percent over the period), and 46 percent were female. All earnings and incomes have been adjusted using the Consumers Price Index (CPI) and expressed in constant (December quarter 2005) dollar values. Average annual FTE job earnings are \$44,077.

Statistical Model Specification and Estimation Issues

In this section we outline the statistical framework adopted to analyse the contributions to the job-level earnings rate of a limited set of observable worker characteristics, and time-invariant (unobservable) worker and firm effects. Letting i, j and t index workers, firms, and time (year), and defining job- ij as the employment relationship between worker- i and firm- j , the unit of observation for this analysis is a unique (ijt) job-year combination. We estimate log-linear additive effects models of the following form:

$$y_{ijt} = \theta_i + \psi_j + x_{ijt}\beta + \tau_t + \varepsilon_{ijt} \quad (2)$$

where y_{ijt} is the log(annual FTE earnings rate) of the job- ij in year- t , which is associated with worker- i ($i = 1, \dots, N$) employed in firm- j ($j = 1, \dots, J$); θ_i is the time-invariant effect associated with worker- i , which represents their earnings premium across the firms they work for; similarly, ψ_j is the time-invariant effect associated with firm- j , which represents the earnings premium it pays to all its workers; x_{ijt} is a vector of observable worker and firm-level characteristics that affect earnings, and β is the associated parameter vector; τ_t are time effects; and ε_{ijt} is a residual that captures idiosyncratic job-match effects, measurement errors, etc.

In the presence of worker fixed effects, time and age are perfectly collinear in a balanced panel, so these effects are not identified in equation 2. Although we have an

unbalanced panel, and also measure age on an employment-weighted basis within each year, which means age is not perfectly synchronised over time, we believe the resulting identification associated with estimating equation 2 is tenuous at best. Maré and Hyslop (2006) contains a discussion of preliminary results from estimating equation 2 and problems associated with these.

Because of this identification problem, we adopt a two-step estimation procedure. In the first stage, we estimate unrestricted sex-age earnings profiles for each year by regressing job FTE annual earnings rate on a full set of worker sex-age dummies, allowing the coefficients to vary by year. Then, in the second stage, we use the first-stage residuals to estimate the (unobserved) worker and firm effects by projecting the residuals onto full sets of worker, firm and time dummy variables. That is, the estimating equations are:

$$\begin{aligned} y_{ijt} &= \beta_{gAt} + \varepsilon_{ijt} \\ \hat{\varepsilon}_{ijt} &= \theta_i + \psi_j + \tau_t + u_{ijt} \end{aligned} \quad (3)$$

where β_{gAt} is a vector of coefficients on a full set of sex * age * year dummy variables, and $\hat{\varepsilon}_{ijt}$ is the residual from the first-stage regression. This approach identifies the combined second stage dependent variable (i.e. the combined worker, firm, and idiosyncratic job-year effects) as orthogonal to the unrestricted sex-year age profiles estimated in the first-stage. The year dummy variables in the second stage regression (τ_t) are included to control for compositional changes over the period. We use information on all employment jobs observed in the LEED, and weight each job-year observation in the estimation by its estimated FTE employment.

There are three main statistical challenges in estimating the relationships in equation 3. First, not all worker and firm effects parameters can be identified. Abowd, Creedy and Kramarz (2002) (ACK) consider the identification of worker and firm effects that arise with the simultaneous estimation of worker and firm fixed effects models such as that shown in the second line of equation 3. As with

standard fixed effects models, restrictions are required in order to identify the relevant effects of interest. The first step for estimation and identification is to allocate job-year observations into distinct 'connected' groups of firms and workers. A 'connected' group contains "all the workers who ever worked for any of the firms in the group and all the firms at which any of the workers were ever employed" (ACK, p. 3). Suppose there are G distinct non-overlapping groups of connected workers and firms. Within a group g containing N_g persons and J_g firms, it is possible to identify the group mean, $N_g - 1$ worker effects and $J_g - 1$ firm effects, yielding $N_g + J_g - 1$ identified effects. Across all G groups, there are $N + J - G$ estimable effects.

Second, the estimates obtained are not unique, and an explicit identification procedure must be imposed. The non-uniqueness of estimates arises because, within each group, some arbitrary normalisation is required – e.g. omitting the group mean, one of the worker effects or one of the firm effects. Our identification strategy is to restrict the overall mean firm effect to be zero, and the mean worker effect within each group to be zero. Given these restrictions, we can identify the overall mean of the dependent variable, and $N + J - G - 1$ worker and firm fixed effects.

Third, given the large number of person and firm fixed effects parameters in the model (more than 3 million), it is not feasible to use standard direct least squares estimation, which requires inversion of a very large sparse covariate matrix. Instead, we use a weighted variant of the exact solution for estimation of this model, as described in ACK. We adopt their approach of using a preconditioned conjugate-gradient algorithm developed by Dongarra, Duff, Sorensen and Van der Vorst (1991), and implemented in Fortran for this application by ACK. We apply the ACK grouping algorithm to data on all job-years observed during the six years of our data.

Results

Aggregate Contributions to Earnings Variation Across Jobs

Based on the grouping and the two-stage estimation approach described above, Table 2 contains a summary of the main estimation results from FTE-weighted regressions of job-year observations. The first row contains the main FTE-employment weighted estimation results of job-year FTE earnings. The logarithm of the FTE annual job earnings rate has a mean of 10.54 (geometric mean of \$37,800 per year), with a standard deviation of 0.34.

By construction, the first stage residuals have zero mean. The earnings component associated with the "first stage

covariates" has a standard deviation of 0.17, while the worker and firm effects have standard deviations of 0.24 and 0.10, respectively. The variability in each of these components reflects the degree of heterogeneity across workers and firms, as measured across intensity-weighted job-years. These results suggest there is greater systematic variability in job earnings within observable age-sex worker demographic subgroups than there is across the groups – i.e. the standard deviation of worker effects (0.24) exceeds the variability of earnings accounted for by the sex-age profiles (0.17).

Overall, the model accounts for 90.3 percent of the variation in FTE annual job earnings rates. To gauge the importance of the various components to job earnings we measure how much each contributes to the total variation in job earnings. For each component, we have calculated both the "simple R^2 " from the regression of $\log(\text{job earnings rate})$ on that component and also the "marginal R^2 ", which is the increase in R-squared associated with adding that component to the model that already includes all other components. The simple R^2 s associated with the observable variation across sex, age and/or years, worker effects and firm effects, respectively, are 0.26, 0.49 and 0.25, while the marginal R^2 s of each of these is 0.24, 0.49 and 0.09. Based on these estimates, the worker effects make the largest contribution, followed by the worker demographics. Given that we do not observe education levels (and other common worker characteristics) in our data, the worker effect component absorbs the impact of human capital variation that is included as part of observable interpersonal variation in some other studies. Also, the lower contribution of the firm effects reflects, in part, the greater clustering of jobs across firms, with an average of 58 job-year observations per firm, compared with only 7 job-year observations per person. As we will see subsequently, the similarity of the simple and marginal R^2 s for the worker effects and observable effects, and the substantial drop in marginal versus simple R^2 for the firm effects, is explained by the finding that these two effects are negatively correlated, while each is positively correlated with the firm effects.

Aggregate Correlations Across Earnings Components

In Table 3, we present the (FTE-weighted) estimated correlations between the various components described in Table 2 across job-year observations. Perhaps the main result of interest in this panel is the correlation between the estimated worker and firm effects. The correlation between the weighted estimates is 0.12. This correlation is greater than the 0.08 estimated by Abowd, Kramarz and Pérez-Duarte (2003) for seven US states, and -0.03 and -0.28 for Washington state and France, respectively, estimated by ACK, suggesting possibly greater positive assortative matching of workers and firms in New Zealand.

Table 2: Summary of earnings components.

	Log(Earning Rate) (y_{ijt})	First-stage Covariates (β_{gAt})	Time Effects (τ_t)	Worker Effects (θ_i)	Firm Effects (ψ_j)	Residual (u_{ijt})
1. FTE Wgtd ($R^2=0.903$)	10.54 (0.34)	10.54 (0.17)	0.00 (0.01)	0.00 (0.24)	0.00 (0.10)	0.00 (0.10)
Simple R^2	...	0.26	...	0.49	0.25	...
Marginal R^2	...	0.24	...	0.49	0.09	...

Note: Numbers in parentheses are standard deviations

Table 3: Correlation Between Earnings Components.

FTE Weighted	y_{ijt}	β_{gAt}	θ_i	ψ_j
Log(Earns) (y_{ijt})	1	0.507	0.697	0.495
1st stage Obs (β_{gAt})		1	-0.088	0.202
Worker effect (θ_i)			1	0.119
Firm effect (ψ_j)				1

In addition, the table shows that the worker effects are negatively correlated with the observed sex-age profiles, suggesting that (lower earning) females and non-prime-aged workers, on average, have positive effects, and males and prime-aged workers have negative effects. Counterbalancing this effect, there is a positive correlation between firm effects and worker demographics, which suggests that higher-paying firms predominantly employ males and prime-aged workers.

Employment Composition Effects over the Period

Between 1999/2000 and 2004/05 there was strong growth in the number of workers and effective annual employment over the sample period (16–17 percent), and in the number of active firms (10 percent). Against this backdrop, we describe the compositional changes associated with the workforce and firms over the sample period, summarised in Table 4. The first row of Table 3 reports the means and variability of the components of earnings variation for the full sample of job-years. The following rows describe the annual composition effects for each year over the period. Average FTE annual earnings rate increased about 6 percent (0.06 log-points) over the period, and this increase is attributed to the sex-age profiles and aggregate time dummies in the first-stage regression (first-stage covariates).¹

The estimates of the second-stage components in the subsequent columns characterise the composition changes over the period, conditional on the observed year-specific sex-age profiles. Although the first-stage regression ensures that the mean of the dependent variable in the second stage (first-stage residual, $\hat{\varepsilon}_{ijt}$) is zero for each year, the composition of workers and firms is changing,

as reflected in their (average) estimated effects. Annual mean worker effects decline from 2 percent above the 6-year mean in the first year (1999/2000) to 3 percent below in the last year (2004/05). There is a smaller decline in the average of firm effects by year from about 1 percent above average in 1999/2000 to marginally less than average in 2004/05. The second-stage estimated time effects act to balance out the impact of these declining average worker and firm effects over the sample, and show an increase from -3 percent in 1999/2000 to 3 percent in 2004/05.¹ In terms of possible compositional changes over the sample period, the correlation between worker and firm effects across jobs is roughly constant over the period, although falls somewhat over the final two years.

These results are consistent with a simple hypothesis that, during a business cycle upswing, the composition of the workforce changes as it expands, and less productive workers and firms are drawn into employment.⁶ However, it should be noted that our analysis is based on earnings and potentially confounds labour productivity with labour supply effects.

Subgroup Analyses

Table 5 documents the variation in earnings rates, worker effects, firm effects, and the correlation between worker and firm effects across various worker and firm-level dimensions.¹

We first describe the patterns across worker sex and age subgroups. The average job earnings rate of males is about 30 percent (27 log-points) higher than for females. Conditional on the respective estimated sex-age profiles, on average, males work in jobs with 2 percent higher firm effects than average. As a consequence of this and of the identification restrictions,¹ females work in jobs with 2 percent lower firm effects, and the average worker effects of males and females are -2 and +2 percent, respectively. The average job earnings rate across age groups describes a concave age earnings profile, with younger workers earning substantially less than, and older workers also earning less than, prime-aged workers. In addition, both young and old workers, on average, have jobs with low firm effects (7 percent lower than average for those aged under 20 years, and 2 percent lower for those aged 60–69 years).

Table 4: Time Variation in Earning Components and Matching.

	Log(Earning Rate) (y_{ijt})	First-stage Covariates ($\beta_{g,lt}$)	Time Effects (τ_t)	Worker Effects (θ_i)	Firm Effects (ψ_j)	Corr(θ_i, ψ_j)
Full sample	10.54	10.54	0.00	0.00	0.00	0.119
1999/2000	10.52	10.52	-0.03	0.02	0.01	0.123
2000/01	10.52	10.52	-0.02	0.02	0.00	0.125
2001/02	10.53	10.53	-0.01	0.01	0.00	0.127
2002/03	10.54	10.54	0.00	-0.00	-0.00	0.125
2003/04	10.57	10.57	0.02	-0.01	-0.00	0.113
2004/05	10.58	10.58	0.03	-0.03	-0.00	0.099

Table 5: Subgroup Variation in Earnings Components and Matching.

	Log(Earning Rate) (y_{ijt})	First-stage Covariates ($\beta_{g,lt}$)	Worker Effects (θ_i)	Firm Effects (ψ_j)	Corr(θ_i, ψ_j)
Full sample	10.54	10.54	0.00	0.00	0.119
Male	10.67	10.67	-0.02	0.02	0.126
Female	10.40	10.40	0.02	-0.02	0.131
Aged (in years):					
Under 20	9.94	9.94	0.06	-0.07	0.036
20–29	10.42	10.42	0.00	-0.00	0.121
30–39	10.65	10.65	-0.02	0.02	0.186
40–49	10.66	10.66	-0.01	0.01	0.136
50–59	10.64	10.64	0.00	-0.00	0.097
60–69	10.54	10.54	0.04	-0.02	0.050
Location					
Auckland	10.62	10.55	0.04	0.04	0.129
Wellington	10.64	10.55	0.06	0.03	0.195
Christchurch	10.49	10.55	-0.03	-0.03	0.045
Other	10.46	10.54	-0.04	-0.03	0.044
Employment Intensity and Stability					
Firms:					
Part year, Parttime	10.44	10.51	-0.03	-0.04	0.110
Full year, Parttime	10.50	10.54	-0.00	-0.04	0.046
Part year, Fulltime	10.82	10.61	0.08	0.12	0.119
Full year, Fulltime	10.81	10.63	0.06	0.11	0.001
Workers:					
Part year, Parttime	10.14	10.42	-0.19	-0.05	-0.012
Full year, Parttime	10.18	10.45	-0.18	-0.06	-0.040
Part year, Fulltime	10.77	10.57	0.11	0.04	0.037
Full year, Fulltime	10.77	10.61	0.11	0.04	0.018

The estimated correlations for male and female workers are similar to the overall correlation. Over the age profile, the correlation is strong for prime-age workers (e.g. for 30–39 year olds, the correlation is 0.19), and weaker for young and old workers (correlations of 0.04 and 0.05 for workers aged under 20 years and 60–69 years, respectively). For young workers, the low correlation between worker and firm effects may be due to early labour market “job shopping”, which results in more “random” matching between workers and firms, and also associated with relatively more higher-ability

workers spending time in non-career jobs (e.g. students) than later in life. For older workers, the lower correlation may be due to greater non-wage compensation associated with older cohorts and/or reflect that the earnings rate is a less important measure of the attractiveness of a job towards the later in the working life. Matching may thus contribute to the shape of the age-earnings profile, with peak earnings levels partly reflecting better job matches.

The principal firm characteristic that we observe in the LEED is the industry the firm operates in, and we

describe the pattern of results across 1-digit industries. Figure 1 summarises the between-industry variation. In this figure, industries are ordered from left to right by increasing average log(job earnings rates). Unsurprisingly, there is substantial cross-industry variation in job earnings rates. For example, the average industry-level log(FTE annual job-earnings rate) varies 0.78 across industries from a low of 10.16 (\$25,800) in the Accommodation, Cafes and Restaurants industry to a high of 10.94 (\$56,400) in the Electricity, Gas and Water Supply industry.

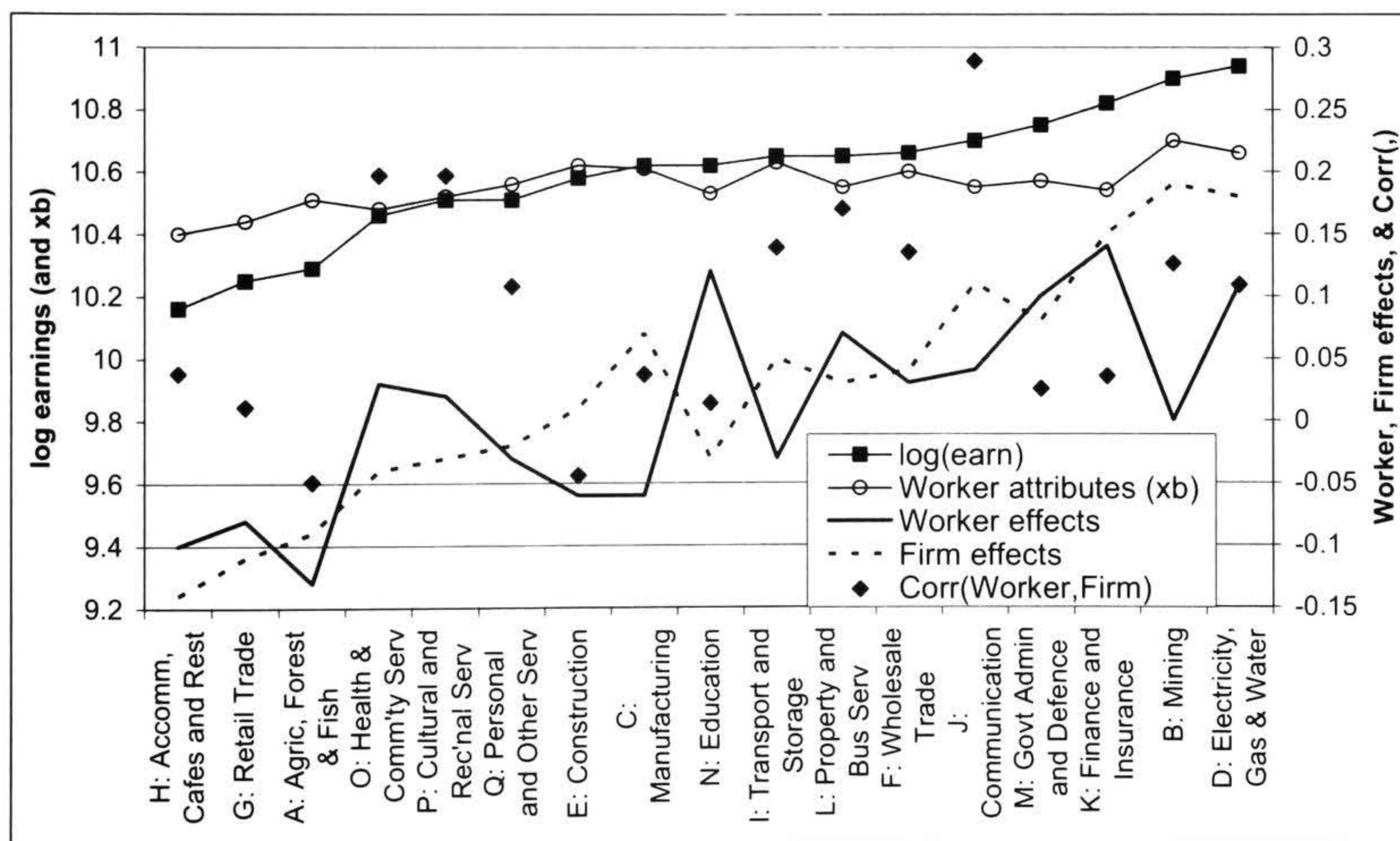
Although there is quite a lot of variation, the raw industry earnings differentials generally reflect all three of the demographic, worker and firm effects. That is, workers in industries with high raw earnings also tend to have demographics associated with higher than average earnings, although the cross-industry demographic relationship is weaker than the raw earnings pattern. Furthermore, the average worker effects and average firm effects tend to be higher in higher earning industries. For example, conditional on observable worker demographics, the average worker effect varies from a low of about 13 percent below the overall average in Agriculture, Forestry and Fishing to a high of 14 percent above the average in Finance and Insurance, while the average firm effect ranges from a low of 14 percent below the overall average in Accommodation, Cafes and Restaurants to a high of 19 percent above the average in Mining. There is also substantial variation in the dispersion of firm effects *within* industries. The standard

deviation of firm effects (not shown) ranges from a low of 0.05 in Accommodation, Cafes and Restaurants, in Government Administration and Defence, and in Education to a high of 0.15 in Mining.

There is substantial variation in the worker and firm effect correlations across industries. The correlation is negative (about -0.05) in Agriculture, Forestry and Fishing, and in Construction, and quite low in several other industries, while there is a strong correlation (0.29) in Communication Services, and also quite high correlation in some other industries.

The next subgroup dimension we consider in Table 5 is based on the geographic location of the firms. Consistent with other research (e.g. Lewis and Stillman, 2005) this shows that jobs located in Auckland and Wellington have 8–10 percent higher earnings rates than the overall average. Our estimates attribute the differences almost entirely to (unobserved) worker and firm effects. In Auckland, both the average worker and the average firm effects are about 4 percent higher than their respective overall averages. In Wellington, the average worker and firm effect differences are 6 percent and 3 percent, respectively. Across regions, the worker-firm effect correlation is higher in Auckland (0.13) and Wellington (0.20) and relatively low in Christchurch (0.05) and "Other" (0.04) areas, providing some support for the hypothesis that labour market matching is more efficient in dense urban areas.

Figure 1: Between-Industry Variation in Earnings Components and Matching.



The final panel of Table 5 shows variation between groups defined by the stability of employment of workers and firms. For this purpose, in each year, we stratify workers by whether they worked full year and/or full-time in every month they worked; and somewhat analogously, we stratify firms by whether the annual employment in a firm consisted predominantly of full-time and/or full-year workers.¹ For workers, the part-time characterisation appears to be the primary dimension along which job earnings rates differ. Part-time workers earn on the order of 40 percent (strictly, 40 log-points) lower job earnings than the overall worker average. The lower earnings are due to 9–12 percent lower earnings associated with observable demographics, 18–19 percent lower worker effects, and 5–6 percent lower firm effects. However, given that our measurement of effective employment is biased upwards (and FTE job earnings biased downwards) for part-time workers, these results should be interpreted with some caution.

A similar, though more muted, pattern applies to firms' employment stability. For example, the job earnings in firms with predominantly "part-time" employment are 4–10 percent lower than the overall average, which is due to 0–3 percent lower earnings associated with worker demographics, 0–3 percent lower worker effects, and 4 percent lower firm effects. The latter suggests that firms that use a relatively large fraction of part-time employment pay lower earnings rates than other firms (who employ the same workers).

The correlation between worker and firm effects is negative for the subgroups of part-time workers (-0.01 and -0.04 for part-year and full-year subgroups, respectively). It is also relatively low for the subgroups of full-time workers (0.04 and 0.02 for the part-year and full-year subgroups). Thus, it appears, that much of the overall correlation between worker and firm effects is associated with the full-time / part-time dimension of employment stability, and that there's little evidence of matching within these groups. The correlation for the subgroups of "full-year" firms are also relatively low (0.05 for the "part-time" and 0.00 for the "full-time" firms).

Concluding Discussion

In this paper we have documented the joint estimation of worker and firm effects, together with observable worker demographic effects, associated with the FTE annual earnings rate of jobs in Statistics New Zealand's LEED. The analysis of these factors has focused on three broad themes. First, we examine how much of the variation in job earnings rates is attributable to observable worker demographic factors (age and sex), unobserved worker effects and unobserved firm effects, and document the degree of heterogeneity in each factor across various dimensions. We find that the unobserved worker fixed effects account for about one-half of the variance in job earnings, while worker demographics account for one-quarter, and the firm fixed effects account for 10–25 percent of the variance.

Second, we explore the compositional changes in workers and firms over the period. Based on changes in the annual average of worker and firm effects over the period associated with such compositional changes, we estimate that worker and firm effects declined about 5 and 1 percent, respectively, over the period. The declining average worker effect is consistent with the hypothesis that there are compositional changes in employment and the labour force over the business cycle that lowers average worker productivity during booms compared to recessions, and suggests there has been about a 1 percent annual decline over the sample period.

Third, we document patterns of sorting of workers and firms across jobs over the sample period. The correlation between worker and firm effects is 0.12, which implies there is a tendency for high-earning workers to work for high-paying firms, suggesting positive complementarities between workers and firms. The 0.12 correlation between worker and firm effects is relatively higher than international estimates. We found quite strong between-group sorting along the full-time / part-time dimension of employment. Sorting appears to be stronger in Auckland and Wellington than other regions. Perhaps unsurprisingly, there is also substantial variation across industries. We find generally positive associations between industries' job earnings rates and each of the average earnings based on worker demographics, average worker effects and average firm effects, although there is also heterogeneity between different industries.

Future Research

This paper is the first to estimate two-way worker and firm fixed effects models for New Zealand. The scope for future research is broad. A discussion of the research possibilities opened up by the availability of linked employer employee data can be found in Abowd and Kramarz (1999).

There are two streams of potential research emerging from the current study. The first is methodological, and includes further analysis of the identification issues that led us to undertake two-stage estimation, the impact of the strength of connectedness between firms and workers, and the stability of fixed effects estimates over time.

The second stream entails further investigation of some of our key findings: analysis of the contribution of firm and worker entry and exit to the reported composition changes over time, and analysis of the patterns of matching, which appear to be largely within rather than between full-time and part-time jobs.

The value of the LEED for further research would be enhanced by inclusion of additional worker and firm covariates, and by some measure of hours. In the absence of such information, the estimated fixed effects produced as part of the current project can serve as proxies for unobserved productive worker and firm attributes in other research projects examining outcomes for workers and firms.

It is worth noting that access to the LEED is restricted. Only researchers employed by Statistics New Zealand are able to access the data, and then only on Statistics New Zealand premises for approved projects. These restrictions reflect the sensitive nature of the data, and give effect to the protections in the Statistics Act and the Income Tax Administration Act.

Notes

- 1 The basic ideas are captured by Becker (1973), who highlights the central role of complementarity in household production as a basis for positive assortative matching of marriage partners.
- 2 In addition to regular firm-worker employment jobs being identified in the LEED, several other relationships involving PAYE tax deductions can also be identified by particular "employer" identifiers. These are working-age social welfare taxable benefits; earnings-related accident compensation payments from the Accident Compensation Corporation (ACC); Student Allowance payments (SA); Paid Parental Leave (PPL) payments; and New Zealand Superannuation (NZS) retirement pensions. In what follows, we make a distinction between LEED earnings from employment-jobs and other LEED income from these other (non-employment) sources.
- 3 We have compared the estimated average employment intensity and the fraction estimated to be full-time with analogous estimates using Household Labour Force Survey (HLFS) data for workers over the sample period (See Maré and Hyslop, 2006). The results confirm that, first, the LEED employment intensity construct has similar properties to analogous survey estimates and, second, in the absence of any direct hours measure, it provides a useful first-order adjustment for estimating differing levels of employment intensity across workers.
- 4 Most of this change appears to be due to aggregate time effects. For example, allowing unrestricted age earnings profiles by age and across males and females, but restricting these to be parallel in different years, we estimate the aggregate time effect to be 6 percent (0.06 log-points) between the first and last years. Relaxing these specification to allow separate sex-year dummy variables, we estimate male aggregate (time) earnings growth of 4.4 percent, and female earnings growth of 7.9 percent.
- 5 The pattern of changing time effects reflects the dynamics of worker and firm inflows and outflows over the period. Maré and Hyslop (2006) contains a more detailed discussion, analysing patterns separately for entering, exiting, and continuing firms and workers.

- 6 Solon, Barsky and Parker (1994) show that the usual macroeconomic finding of weakly cyclical wages over the business cycle is substantially affected by composition bias: controlling for compositional changes in employment over the business cycle, real wages are strongly procyclical.
- 7 In considering the subgroup correlations it is important to realise that these are "within-group" correlations (i.e. the correlations are calculated relative to the subgroup worker and firm effect means), and exclude any between-group correlation effect that is included in the overall correlation estimate.
- 8 Literally, the symmetry is exact if we have equally balanced male and female subsamples.
- 9 More specifically, we classify a firm's employment in a year as "full-time" if the number of observed worker-months in LEED is at least 75 percent of the potential number of months given the number of workers employed by the firm during the year, and we classify the firm as "full-time" if the level of FTE employment in the firm is at least 95 percent of the number of worker-months. See Hyslop and Maré (2006) for a more detailed description and discussion of these employment stability measures.

References

- Abowd, J.M., Creecy, R.H. and Kramarz, F. (2002). *Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data*. LEHD Technical Paper 2002-06, available from www.lehd.census.gov
- Abowd, J.M. and Kramarz, F. (1999). The analysis of labor markets using matched employer-employee data. In Ashenfelter, O. and Card, D. (eds) *Handbook of Labor Economics*, Vol. 3B, Chapter 40.
- Abowd, J.M., Kramarz, F. and Pérez-Duarte, S. (2003). *Sorting Workers Between and Within Industries*, February 2003, available at <http://cep.lse.ac.uk/seminarpapers/28-02-03-KRA1.pdf>
- Becker, G.S. (1973). A theory of marriage: Part I. *Journal of Political Economy*, **81**, 813-46.
- Dongarra, J.J., Duff, I., Sorensen, D. and Van der Vorst, H. (1991). *Solving Linear Systems on Vector and Shared Memory Computers*. SIAM, Philadelphia.
- Hyslop, D. and Maré, D. (2006). *Patterns of Labour Market Fragmentation, and Worker and Firm Employment and Earnings Change, 1999-2005*, Wellington: Statistics New Zealand.

Lewis, G. and Stillman, S. (2005). *Regional Economic Performance in New Zealand: How Does Auckland Compare?*, New Zealand Treasury Working Paper, WP 05/08. Wellington: The Treasury.

Maré, D. and Hyslop, D. (2006) *Worker-Firm Heterogeneity and Matching: An Analysis Using Worker and Firm Fixed Effects Estimated from LEED*, Leed Research Report. Statistics New Zealand; available at www.stats.govt.nz.

Seyb, A. (2003). *The Longitudinal Business Frame*, Statistics New Zealand, available from www.stats.govt.nz

Solon, G., Barsky, R. and Parker, J. (1994). Measuring the cyclicalty of real wages: How important is composition bias?, *Quarterly Journal of Economics*, **109**(1), 1-25.

Authors

David C. Maré
Senior Fellow
Motu Economic and Public Policy Research
P.O. Box 24390
Wellington
Dave.Mare@motu.org.nz

Dean R. Hyslop
Principle Advisor
Policy Support and Analysis Section
The Treasury
P.O. Box 3724
Wellington
Dean.Hyslop@treasury.govt.nz