



AN EXPLORATORY ANALYSIS OF AGGREGATES AND INDIVIDUAL WAGE EARNINGS IN THE LINKED EMPLOYER- EMPLOYEE DATA (LEED)

Eleanor Guzman-Posadas

*Statistical Methodological Services,
Statistics New Zealand, Wellington*

Abstract

Longitudinal analyses with continuous repeated outcomes provide fuller insight into population and individual behaviour over time. Insights into processes of social change can thus be greatly enhanced through a more extensive use of longitudinal data. Using the Linked Employer-Employee Data (LEED), factors associated with wage earnings were explored. Random effects models were investigated to identify fixed population effects as well as to help understand stochastic processes attributed to individual employee variations. It was found that wages vary significantly across region of residence, industry, age groups and gender. Random intercept adjustments provide an effective alternative for exploring wage earning variability over time.

Introduction

Competing in a global economy brings considerable challenge for any government to sustain its edge relative to other economies. To ensure sustainable growth, one of the strategies identified by government was to maintain a highly skilled workforce.

With the growing demand for skilled workers, wages and salaries have been used as an indicator of one's competitive edge over others. To ensure its competitive position in national and international markets, employers must sustain productivity by improving opportunities for workforce training and maintaining a skilled workforce through wages. Over the years, significant disparities in wage earnings have been observed not only in New Zealand but also in other industrialised countries.

A careful analysis of the determinants of wage earnings is necessary, as it could be linked to a series of factors, that, when identified, could lead to policy reforms that would benefit both the New Zealand economy and its residents.

This study, attempts to explore how wages vary across age, gender, region of residence and select industry groups in New Zealand over time. The aim is to draw insight not only on population outcomes but how differences in individual experiences affect our understanding of wage earnings. It takes advantage of the Linked Employer-Employee Data (LEED), an integrated dataset of administrative data from the Inland Revenue PAYE and income tax systems and business data from Statistics New Zealand's Business Frame (Graham 2003), by applying general linear mixed effects models. A particular focus of this research is an understanding of how these models behave by examining population and

individual outcomes using LEED. Consequently, this research adds to existing literature.

Methodology

A number of studies have attempted to examine the determinants of disparities in wage earnings. Bowles, Gintis and Osborne (2001) found that education, heredity, ability to adapt to a new economic condition, attitudes towards risk and other behavioural factors affect earnings. Calistri and Galbraith (2001) investigated the disparities in wage earnings relative to the concept of supply and demand in the labour market. Hyslop and Mare (1999) found that changes in household structure, attributes and employment outcomes each contributed to the observed increase in household income disparities. Papps (2004) evaluated the income gap between men and women, and assessed the factors associated with changes in income inequality between genders. O'Dea (2000) provides a good summary of studies that explain the disparities in income distribution in New Zealand. These studies utilised the results of survey panel data like, the Household Labour Force Survey or the New Zealand Income Survey, or longitudinal datasets.

However, the application of random effects models has not been explored in New Zealand. While cross-sectional analysis would provide an idea of factors associated with wage earnings at a point in time, it is useful to understand how variations in the population and individual experiences affect wages over time. Hyslop's (1999) work on LEED, identified the need to understand income variations and its association with previous or earlier wage earnings. This is quite intuitive as reforms in labour policies, changing immigration flows, varying individual preferences and behavioural factors expected in a

dynamic and evolving labour market ideally should be considered in understanding earning inequality.

While Hyslop (1999) investigated the effect of age and gender on income, this paper explores the significance of place of residence and industry in explaining wage earning disparities between 1999 and 2005. Migration may affect the regional supply and demand leading to wage disparities over time. Labour demand in certain industry sectors likewise may contribute to the increasing wage disparities. This paper tries to evaluate whether deviations in individual experiences play a significant role in understanding wage differences. It should be noted, however, that while the LEED base data does not provide a number of important socio-demographic variables, such as ethnicity, that may be associated with wage earnings. A detailed description of the limitations of LEED is available in the LEED feasibility report (2003).

In this paper, factors associated with wage earnings at baseline and at the end of the study period were investigated. Particularly, the associated effect of age, gender, region of residence, and industry classification of an employer were explored. While this model provides the average wage profile at a point in time, an understanding of the magnitude and/or shift of the associated effect of these factors on wages is the main focus.

The main motivation for considering linear mixed effects models is to understand the variability of wage earnings in terms of fixed population effects and as well as differences in individual experiences over time. In general, a linear mixed effects model is any model which satisfies the following condition (Verbeke and Molenberghs 2000),

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i$$

$$b_i \sim N(0, D)$$

$$\varepsilon_i \sim N(0, \Sigma_i)$$

$$b_1, \dots, b_N, \varepsilon_1, \dots, \varepsilon_N \text{ independent.}$$

where Y_i is the n_i -dimensional response vector for individual i , X_i and Z_i are the $(n_i \times p)$ and $(n_i \times q)$ dimensional matrices of known covariates (variables that explain the response), β is a p -dimensional vector containing the fixed effects (population-average regression coefficients), b_i is the q -dimensional vector containing the random or subject-specific effects (individual employee regression coefficients) and ε_i is an n_i -dimensional vector of residual components. The random effects, b_i , and residual components ε_i are assumed to be independent with distributions $N(0, D)$ and $N(0, \Sigma_i)$ respectively. Estimation was done using maximum and restricted maximum likelihood.

Data

The Linked Employer-Employee Data (LEED) is an integrated longitudinal dataset linking employer and employee series from Statistics New Zealand's Business Frame and the Employer Monthly Schedule payroll data from Inland Revenue. It consists of unit record data at the employer and employee level. A unique identifier between the two datasets makes the integration feasible. In this paper, individual level data was used covering every taxpayer that pays tax at source. Being effectively a census, the data can be used to produce information on special groups of interest (Graham, 2003).

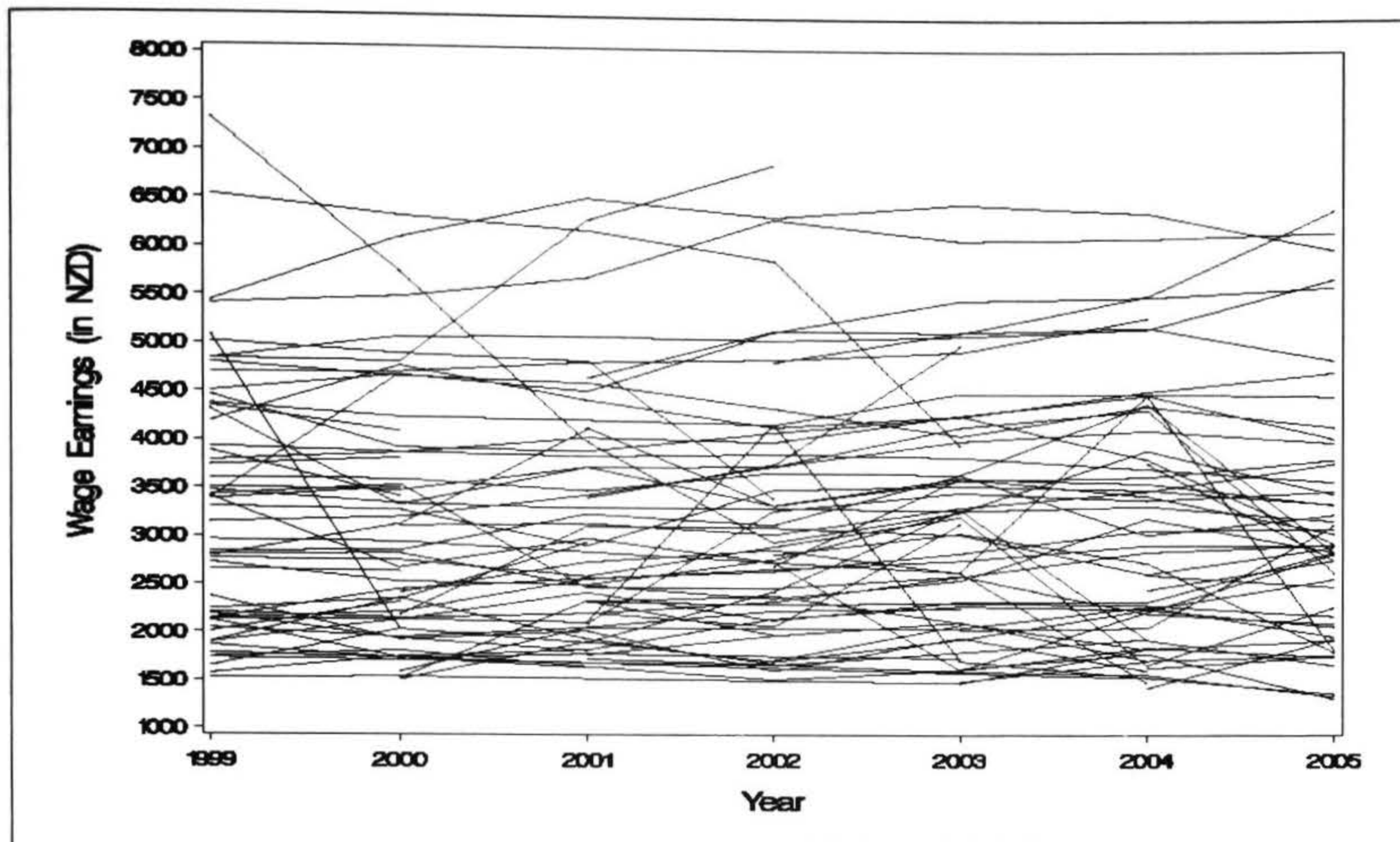
While the use of complete, rather than sample, data is feasible in LEED, the dataset used here is a random sample of 4035 people who received wages from April 1999 to July 2005. The study population was limited to those people 15 years and over receiving wages and salaries. In LEED, gross earnings were based on income generated from accident compensation, benefit, pension, wages and salary, student allowance, paid parental leave and withholding payment.

In the absence of an indicator of temporary or part time employment, wage earnings were limited to those between 1500 and 10000 NZD per month. Wage earnings below 1500 NZD per month were considered earnings from temporary or part time employment and those greater than 10,000 NZD per month were considered extreme cases. Earnings from self-employment, i.e., income from withholding payment (WHP), were likewise excluded due to the limited number of observations currently available in LEED.

Initially, monthly employee profiles were followed from April 1999 to July 2005. In the analysis, as wage earnings were observed to be relatively stable between months, the average monthly earnings per quarter was used in the analysis instead. Quarters were defined within the calendar year to account for seasonal variations. It should be noted that while unit record data was used in the analysis, strict adherence to privacy, security and confidentiality protocols, defined in the LEED feasibility report (2003), were followed.

As an indication of varying dispersion of wage earnings between and within employees over time, the employee-specific profiles of 99 employees, randomly selected from the 4035 sample, are plotted in Figure 1. Wage earnings were observed to be quite varied between individuals for the baseline year (1999). Generally, wages were relatively stable with some individuals observed to have significant peaks and troughs in their earnings over time. Data gaps in the profile or dropouts were likewise observed which may be attributed to people shifting from work to other income sources and back to wage earnings within the study period.

Figure 1: Employee specific profile of wage earnings.



The average and median monthly wage earnings per year are shown in Table 1. We expected wage earnings to increase over time but wage earnings in the sample indicate a decreasing trend except for the increase in 2003 and 2004. The decreasing trend is not consistent with economy-wide data and may be due to a number of factors. One possible reason is the varied sample composition per period due to dropouts. Workers were observed to dropout, on average, around 4% per quarter. Maximum loss of information was observed at the end of the study period, with about 15% of the sample dropping out leading to an increase of about 17% in the variance. Assessing the average evolution of wage earnings by region and industry, may provide additional input unto this trend. The average evolution of wage earnings by subgroups, describe how the profile for a number of relevant subgroups evolves over time.

Table 1: Average monthly wage earnings from 1999 to 2005* (in NZD).

Year	Mean (Std Error)	Median
1999	3,319.67 (19.36)	2,947.88
2000	3,205.36 (16.43)	2,840.08
2001	3,167.36 (16.08)	2,809.71
2002	3,149.24 (15.87)	2,803.69
2003	3,191.03 (15.75)	2,858.16
2004	3,219.19 (15.69)	2,905.40
2005	3,174.93 (18.04)	2,864.38

* Wages are in constant (June 1999) dollar values, adjusted using the Consumer Price Index

In LEED the geo-coded address of the employee was used to determine the Territorial Local Authority and Regional Council in which the employee resides. For purposes of analysis three region breakdown was used; Auckland, Wellington and all other regions. A person's region of residence is time-specific, it varies for individuals over time.

The average profile of wage earnings by region is shown in Figure 2. Note the varying regional gap over time. Generally, the average wage trend of residents from the Auckland region has a similar time profile with those from other regions. Average wage earnings decreased until 2002 and increased through 2004. On the other hand, the average wage earnings in Wellington were observed to dip slightly in 2004. Over time, wage earnings of residents in Auckland (p-value = 0.023) were significantly higher than those in the Wellington region. On the other hand, wage earnings of those in other regions (p-value = 0.0015) were significantly lower than those in Wellington. These differences are uncontrolled for workers and employers attributes and may therefore be the result of changing sample composition over time.

The wage earning profile by industry group were also explored. In LEED, the industry from which an employer was classified was based on the 2006 Australian and New Zealand Standard Industrial Classification (ANZSIC). For purposes of the analyses, focus was on the following industry groups: Manufacturing, Retail Trade, Professional, Scientific and Technical Services, Education, Health Care and Arts and Recreation. For reference, the following codes were used in some plots to denote the industry classification.

Figure 2: Average evolution of wage earnings by region.

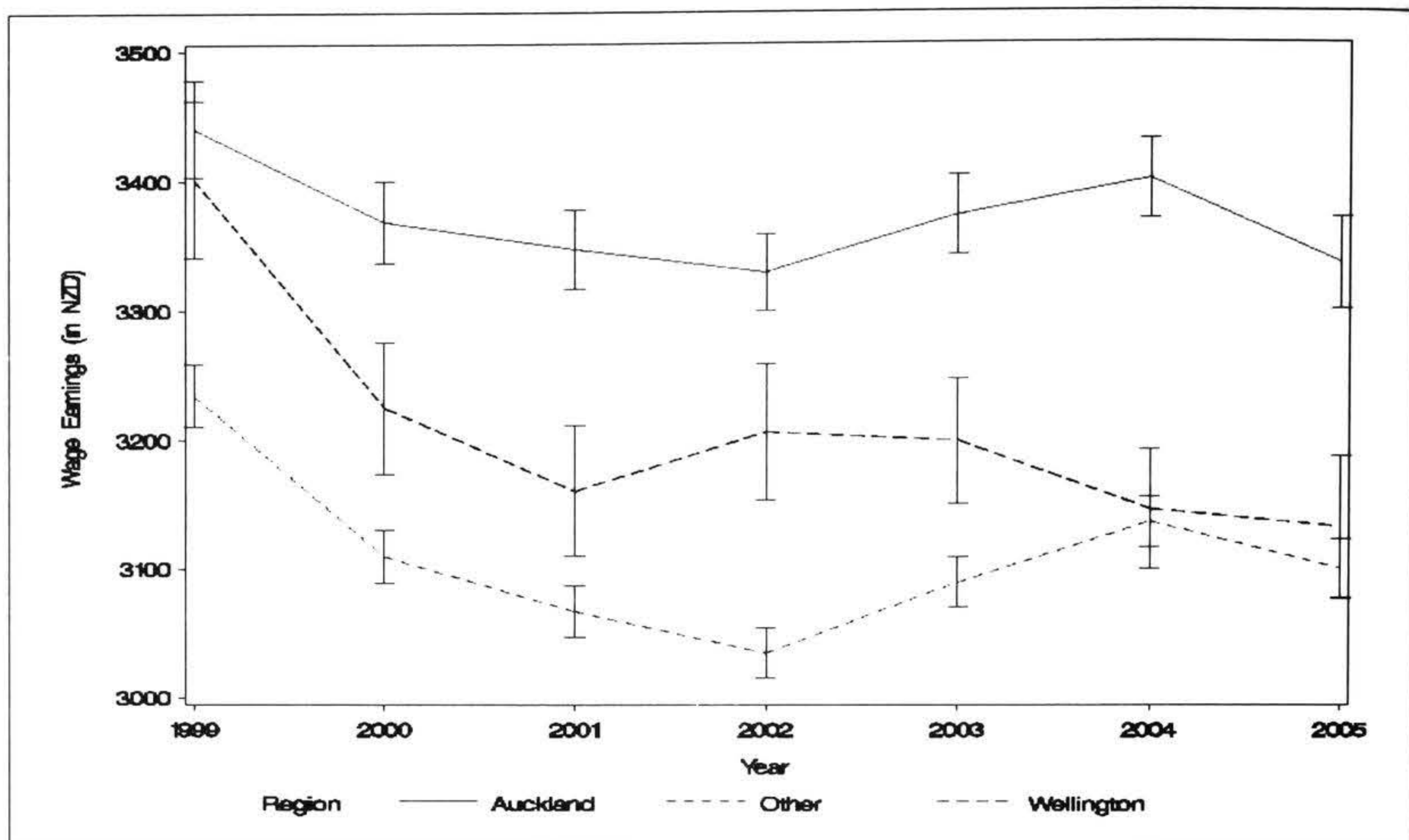


Table 2: Classification Code of Selected Industries.

Code	Industry Name
C	Manufacturing
G	Retail Trade
M	Professional, Scientific and Technical Services
P	Education and Training
Q	Health Care and Social Assistance
R	Arts and Recreation Services

The trend in average wage earnings by industry are shown in Figure 3 below. On average, wage earnings of workers in the Professional, Scientific and Technical Services (M) industry remain the highest across each year and industry groups followed by those in the Education and Training (P) and Manufacturing (C). Average wage earnings of workers from the Health Care and Social Assistance (Q) and Arts and Recreation Services (R) industry were shown to be relatively the same.

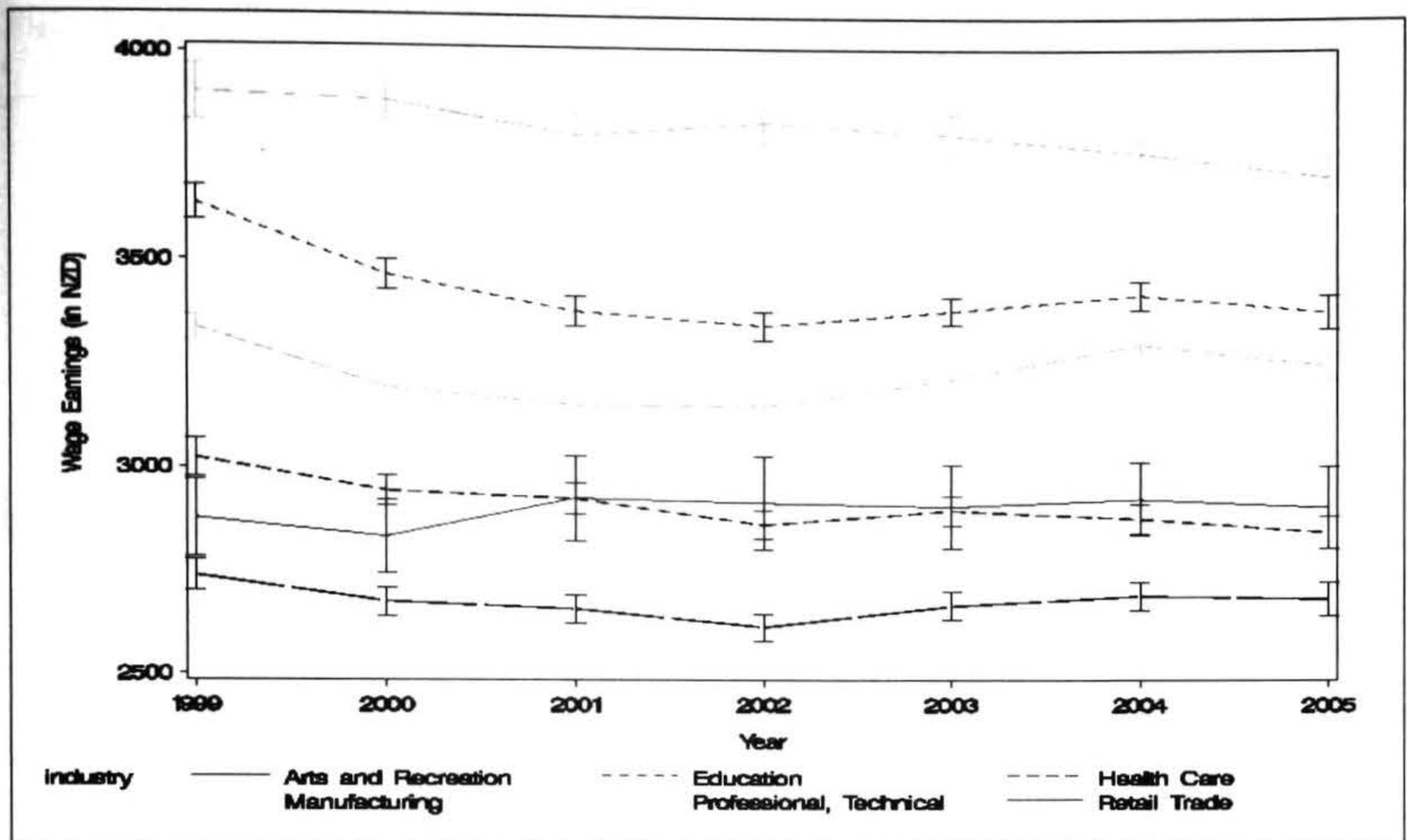
Over time, wages from Manufacturing (p-value=0.0001), and Professional, Scientific and Technical Services (p-value=0.0006) were significantly higher than Arts and recreation while wage earnings of workers from the retail trade industry were significantly lower (p-value = 0.03).

On the other hand, there was no evidence to suggest a significant difference in earnings between those in the Education (p-value = 0.54) and Health Industry (p-value = 0.69) and those from the Arts and Recreation Services.

In LEED, age and gender are derived variables. Age is based on the date of birth in the IRD records while gender was either based on the sex-specific titles of each individual or on the individuals' first name. Less than 4% of employees in the IRD records have missing date of birth and were imputed in LEED. For gender, 98.5% of the records were derived from the title while 1.5% was derived from first name (Graham, 2003). In this paper, the analysis was limited to wage earnings of people 15 years and over. As an explanatory variable, a ten-year age grouping was used.

It is feasible to use other variables in LEED and derive variables to improve our understanding on employer-employee dynamics. However, certain socio-economic covariates, such as ethnicity, education, and hours of work, which were shown to be associated with wage earnings in similar studies, are still unavailable. The feasibility of integrating this information on LEED is currently being reviewed.

Figure 3: Trends in average wage earnings by industry.



Dropout

Within the study period, it is expected that some employees would dropout from the work pool or may shift towards other income sources, e.g. from wages and salaries to benefit, resulting in data gaps in the series. As previously pointed out, sampled workers were dropping, on average, around 4% per quarter. Evidently, the depletion of the study subjects increases the variability and consequently reduces the precision of our estimates. Note that while the general linear mixed framework takes this into account, i.e., it is not necessary to have a balanced structure in linear mixed framework, sensitivity analysis and an assessment of the impact of missing values in the analysis, may be explored in future research work. Complete case analysis was employed in the analysis of endpoints.

Results

The aim of the study is to explore factors associated with wage earnings of working adults in New Zealand between 1999 and 2005. As wage earnings depend on the changing supply and demand for certain skills and individual attributes and employment outcomes, interest is on both fixed population effects and individual experiences over time.

The associated effects of the selected socio-economic covariates on wage earnings were initially investigated at baseline and at the end of the study period. A comparative summary of the parameter estimates for both study periods is shown in Table 3.

Table 3: Cross-sectional association of covariates with log wage earnings at baseline and end of study period by linear regression analysis.

Parameter	Baseline	Endpoint
	Estimate (Std. Error)	Estimate (Std. Error)
Gender		
Male	7.99 (0.082)**	7.69 (0.15)**
Female	7.74 (0.083)**	7.46 (0.15)**
Age Group		
15-24	- 0.21 (0.062)*	0.029 (0.14)
25-34	0.03 (0.059)	0.28 (0.13)*
35-44	0.10 (0.059)	0.34 (0.13)*
45-64	0.079 (0.059)	0.31 (0.13)*
65+	.	.
Region		
Auckland	0.004 (0.028)	0.072 (0.032)*
Other	- 0.048 (0.026)	0.007 (0.030)
Wellington	.	.
Industry		
Manufacturing	0.066 (0.059)	0.060 (0.061)
Retail	- 0.049 (0.061)	- 0.058 (0.063)
Professional	0.26 (0.061)**	0.27(0.063)**
Education	0.28 (0.060)**	0.18 (0.062)*
Health	0.093 (0.061)	0.036 (0.063)
Arts and Recreation	.	.
R-square (in %)	21	17

** significant ($p < 0.05$); * highly significant ($p < 0.0001$)

At baseline, associated effect of gender (p-value = 0.0001), age (p-value = 0.0001), region (p-value = 0.0104) and industry (p-value = 0.0001) on wage earnings were found to be significant. Notably, the average wage earnings of male workers were significantly higher than female workers, while controlling for age, region and industry. While wage earnings of workers in the Auckland (p-value = 0.90) and other regions (p-value = 0.072) were not significantly different from those in Wellington, wages of workers from other regions were different from those in the Auckland region. Moreover, there was evidence to suggest that wages of workers in the Professional, Scientific and Technical Services (p-value = 0.0001) and Education (p-value = 0.0001) industry were significantly different from those working in the Arts and Recreation Services while wage earnings of those in the Manufacturing (p-value = 0.26), Retail Trade (p-value = 0.42) and Health Care and Social Services Industry (p-value = 0.13) were not statistically different from this industry. Jointly, these covariates explain about 21% of wage earnings variability at baseline.

Similarly, the associated effect of gender (p-value = 0.0001), age (p-value = 0.0001), region (p-value = 0.00216) and industry (p-value = 0.0001) on wage earnings were found to be significant at the end of the study period. However, parameter estimates of these fixed population effects were different compared with those at baseline. One notable difference is the significant gap that opened up between wage earnings of residents in the Auckland region from those in Wellington region, i.e., Auckland residents earning more, at the end of the study period. Additionally, except for workers between 15 and 24 years, wage earnings of workers between 25 to 34, 35 to 44 and 45 to 64 were found to be significantly higher than those 65 years and over.

The observed differences in wage estimates at baseline and end of the study period may be expected. Disparities in labour demand and supply may have affected wage earnings between the two time periods. Conversely, up-skilling through training or further studies, changes in personal circumstances, i.e., having children, may have contributed to higher or lower earnings over time. An investigation of the associated effect of previous work earnings on current earnings would be useful in understanding wage disparities over another.

In this respect, we consider a random intercepts model. A random intercepts model assumes that employee-specific wage profiles are linear and fixed with respect to the selected covariates and only random intercept or varying employee-specific baseline earnings explains the stochastic process. Furthermore, this model assumes constant variance over time as well as equal positive correlation between two wage measurements for the same worker. This covariance structure is referred in literature as compound symmetry. While we initially expected wage earnings to vary over time, warranting a random slopes model, this was not generally observed in the randomly selected wage profiles in our sample. Using the same mean structure as the cross-sectional regression models, the resulting random intercepts model is shown

in Table 4. Noticeably, the high intraclass correlation implies that wage disparities between workers are large relative to disparities within individuals.

Table 4: Result of general linear mixed analysis for fixed and random effects of log wage earnings.

	Estimate (Std. Error)	Pr > t
Gender Effect		
Male	7.91 (0.022)	<.0001
Female	7.70 (0.022)	<.0001
Age Group Effect		
15-24	-0.13 (0.0095)	<.0001
	-0.0076	
25-34	(0.0080)	0.34
35-44	0.032 (0.0076)	<.0001
45-64	0.051 (0.0079)	<.0001
65+	.	.
Region Effect		
Auckland	0.044 (0.014)	0.0019
Other	-0.038 (0.013)	0.0031
Wellington	.	.
Industry Effect		
Manufacturing	0.069 (0.017)	<.0001
Retail	-0.043 (0.017)	0.012
Professional	0.089 (0.017)	<.0001
Education	0.042 (0.018)	0.0213
Health	0.038 (0.018)	0.0372
Arts and Recreation	.	.
Time Effect	0.0032	
(no. of quarters)	(0.00014)	<.0001
Random Effects		
Intercept variance	0.092	
Residual variance	0.037	
Intra-class correlation	0.72	

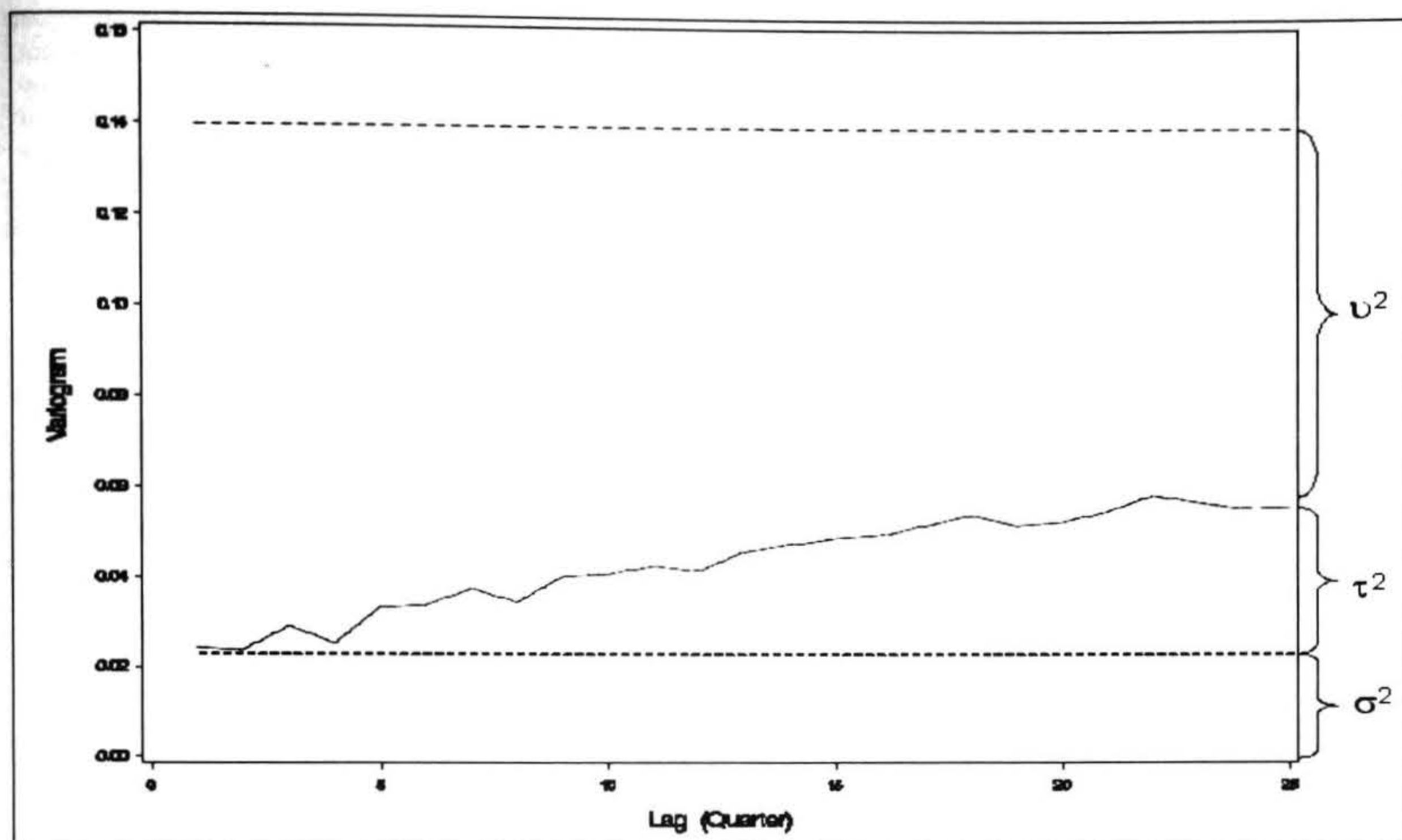
Furthermore, the variogram of a model accounting for age group, region, gender and industry is shown in Figure 4. A variogram plot provides evidence of measurement error, serial correlation and individual effects. Evidence of serial correlation suggest that part of a worker's wage profile is a response to time-varying stochastic processes operating within that worker which supports further, the use of a random effects model.

In notation, this model can be expressed as,

$$Y_{ij} = (\beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5) Age_i + (\beta_6 + \beta_7) Gender_i + (\beta_8 + \beta_9 + \beta_{10}) Region_i + (\beta_{11} + \beta_{12} + \beta_{13} + \beta_{14} + \beta_{15} + \beta_{16}) Industry_i + t_{ij} + b_{1i} + \varepsilon_{ij}$$

for worker $i = 1 \dots 4035$ and time $j = 1, \dots, 27$. This model suggests a strong link between wage earnings, gender, industry, age and region. While controlling for age, region and industry, wages of male workers were shown to be significantly different from female workers over time.

Figure 4: Variogram.



Notes: The variogram for the linear random-intercepts model. σ^2 , τ^2 , u^2 represent the variability of the measurement error, the serial correlation component and the random intercepts respectively.

As a proxy for experience, as one grows older, wage progressively increases. Between regions, wages of residents in the Auckland region earn more than those in the Wellington region. Conversely, wages of residents from other regions were significantly lower than those in Wellington. Between industries, except for wages of workers in the Retail Services industry, wages of those in the Manufacturing, Professional, Education and Health were significantly higher than those in the Arts and Recreation industry. The resulting predicted and observed profiles for this model have shown that the empirical Bayes estimates for the random intercepts clearly correct the population average profile toward the observed.

While this model seems to suggest gender differences in wage earnings while controlling for age, industry and region, other factors may be contributing to these differences. Previous work has suggested that educational level, ethnicity and other socio-economic covariates may help explain these differences. However, these socio-economic variables are still not available in LEED. Therefore these results cannot address the question of gender wage disparities. The distribution of the empirical Bayes estimates, further suggest the need to evaluate other variables that may be associated with wage earnings. The distribution suggests a bimodal distribution whilst controlling for the effect of gender. While procedures in analysing mixtures (i.e., heterogeneous populations) are well documented (Bohning 1999), this will not be included in this report.

Strengths and Limitations

The longitudinal analysis of the Linked Employer-Employee Data has a number of advantages and disadvantages. Its main strength is in maximising the information that can be used in LEED. This may not previously have been feasible, for instance, using the general linear framework, as data gaps usually result in loss of information and therefore loss of power for inferential work. Through longitudinal analysis, information for each individual in the population is maximised.

The application of linear mixed effects models however was not straightforward. For instance, it may be feasible to run simultaneous regression models on each employee in our sample and use the estimated regression coefficients to assess how each employee experiences varies within the population. In literature, this is sometimes referred to as the two-stage approach. In linear mixed, we needed to understand the average profile of the population (i.e., fixed population effects or how the covariates relate with wage earnings on average) and whether individual stochastic processes (i.e., random effects or how individual records vary over time) may further explain changing wage earnings over time. The latter is rather more intuitive and more informative as we understand how individual experiences influences wage earnings. However, as was previously pointed out, the underlying trend in the empirical Bayes estimates for

random intercepts, suggest the need to identify other potential covariates that would explain the disparities on wage earnings over time.

The Linked Employer-Employee Data offers a wealth of data. Both cross-sectional and longitudinal analysis can be undertaken. Analysis of endpoints was undertaken initially to assess whether longitudinal analysis is warranted. This approach may be useful particularly in intervention analysis where you compare baseline results with 'after-intervention' outcomes to assess the effectiveness of such interventions. Note however, that this result in loss of information by disregarding records between these measures as previously pointed out.

As it stands, the selected model describes how age, region of residence, gender and industry relate with wage earnings over time. The results were quite interesting. However, this cannot be used as basis to support gender differences in wage earnings due to several key variables not been available in LEED. Intuitively, wages would be affected by the number of hours spent, particularly for temporary or part time work, for a particular job. However, an indicator of part time or temporary employment or the number of hours spent for the wages received is not available in LEED.

Adopting Calistri and Galbraith's (2001) perspective of wage disparities resulting from changing supply and demand, it would be useful to determine the effect of up-skilling through further education and type of occupation in the selected industries. This would also help explain the observed wage gap between regions and industry.

Future Research

While the results provide evidence of the suitability of undertaking longitudinal analysis in LEED, particularly of random effects models, there is a need to understand further the effect of missing values or dropouts in analysing longitudinal data. Shrinkage effect, i.e., observed information shrunken towards the average profile, which results from an unbalanced data, should be evaluated.

In understanding wage disparities over time, there is a need to further exploit the feasibility of integrating other information sources that would provide information on educational attainment, occupation, civil status and other socio-economic covariates, number of hours worked, indicators of permanent or part time employment among other variables, that would further enhance models describing wage earning disparities over time. These variables would also be useful in understanding other areas such as income transitions, employment and labour dynamics, and other outcomes.

Note

1 The results presented in this paper are the work of the author, and does not necessarily reflect the views of Statistics New Zealand. Strict adherence

to privacy, security and confidentiality protocols were followed in undertaking this research. Only people authorised by the Statistics Act 1975 were allowed to see data about a particular person or firm. The data was used solely for statistical purposes and no individual information is provided back to the Inland Revenue Department for administrative or regulatory purposes. Any discussion of data limitation is in the context of using the Linked Employer-Employee Data for statistical purposes and is not related to the ability of the data to support IRD's core operational requirements.

The author acknowledges the assistance of David Mare, Dean Hyslop, Tas Papadopoulos, Michael Slyuzberg, participants of Statistics New Zealand LEED Research Forum and Longitudinal Analysis and Survey Design Network and most specially to Walter Davis for their valuable comments on this paper, and to Nairn MacGibbon, and Steve Wong for assisting us in understanding the complex LEED data structure and threshing the main objectives of this paper.

References

- Arnold, R and Liu, I.** (2006). *Longitudinal Data Analysis. Training notes on longitudinal data analysis*, 10-18 May 2006.
- Bertola, G.** (2003). Earnings disparities in OECD countries: Structural trends and institutional influences. Paper presented to the joint *OECD/NDRC Understanding China's Income Disparities Seminar*. Paris, 20-21 October 2003.
- Bowles, S., Gintis, H. and Osborne, M.** (2001). The determinants of earnings: A behavioural approach. *Journal of Economic Literature*, **39**(4), 1137-1176.
- Calistri, A. and Galbraith, J.** (2001). *Wage Flexibility and Unemployment: A Panel Data Analysis of OECD Countries*. University of Texas Inequality Project Work Paper No. 18. University of Texas.
- D'Souza, K.** (2004). *A Profile of Persons Not in the Labour Force*. Key Statistics: May 2004. Statistics New Zealand.
- Daly A., Kawaguchi A., Mumford K., and Meng X.** (2006). *The Gender Wage Gap in Four Countries: A Rehearsal of the Gregory Contribution*. Institute for the Study of Labor (IZA) Discussion Paper No. 1921, January 2006.
- Gannon B., Plasman R., Rycx F., Tojerow I.** (2005). *Inter-Industry Wage Differentials and the Gender Wage Gap: Evidence from European Countries*. Institute for the Study of Labor (IZA) Discussion Paper 1563, April 2005.

Graham, P. (2003). *The IRD Base Data - Quality and Characteristics and LEED Derivation, Edit and Imputation Methods*. Wellington: Statistics New Zealand

Hyslop D., and Maré D. (1999). *Understanding Changes in the Distribution of Household Incomes Between 1983-86 and 1995-98*. Department of Labour and NZ Treasury, Unpublished preliminary results as at September 1999.

Katz, L. and Autor, D. (1998). Changes in the wage structure and earnings inequality. Chapter prepared for Orley Ashelfelter and David Crd (eds.), *Handbook of Labor Economics*.

O'Dea, D (2000). *The Changes in New Zealand's Income Distribution*. Treasury Working Paper 00/13. Wellington: The Treasury.

Papps, K.L. (2004). *Income Inequality and Gender in New Zealand, 1998-2003*. Institute for the Study of Labor (IZA) Discussion Paper No. 1365, October 2004.

Statistics New Zealand (2003). *Feasibility Report: Linked Employer-Employee Data (LEED)*. Wellington: Statistics New Zealand.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data.*, New York: Springer-Verlag.

Author

Eleanor Guzman-Posadas
Senior Methodologist
Statistical Methodological Services
Statistics New Zealand
P.O. Box 2922
Wellington
Eleanor.Posadas@stats.govt.nz