Australian Bureau of Statistics

1351.0.55.055

Research Paper

Use of a Prototype Linked Employer-Employee Database to Describe Characteristics of Productive Firms

New Issue

Research Paper

Use of a Prototype Linked Employer-Employee Database to Describe Characteristics of Productive Firms

Chien-Hung Chien and Andreas Mayer

Data Standards and Methods Branch

AUSTRALIAN BUREAU OF STATISTICS

EMBARGO: 11.30 AM (CANBERRA TIME) THU 14 MAY 2015

ABS Catalogue no. 1351.0.55.055

Produced by the Australian Bureau of Statistics

INQUIRIES

The ABS welcomes comments on the research presented in this paper.
For further information, please contact Ric Clarke, Director, Statistical Informatics Section, on Canberra (02) 6252 6736 or email <analytical.services@abs.gov.au>.

USE OF A PROTOTYPE LINKED EMPLOYER-EMPLOYEE DATABASE TO DESCRIBE CHARACTERISTICS OF PRODUCTIVE FIRMS

Chien-Hung Chien and Andreas Mayer
Data Standards and Methods Branch

EXECUTIVE SUMMARY

This study uses a prototype linked employer-employee database (LEED) to analyse both employee and firm characteristics to identify factors that explain differences in labour productivity across firms and industries. We created the prototype LEED by linking de-identified individual Personal Income Tax and Business Tax data from the Australian Taxation Office with the Australian Bureau of Statistics Business Longitudinal Database (BLD), for the 2010–11 financial year.

We demonstrate the analytical potential of the prototype LEED by constructing multilevel models (two and three-level) to describe employer and employee characteristics of productive firms. We caution readers not to draw any causal conclusions from the analysis because the purpose was descriptive analysis only. This paper has demonstrated the importance of considering both firm and employee dynamics in the analysis of labour productivity.

Our two- and three- level results are broadly consistent. We have found that investment is significantly negative at the industry level but positive at the firm level. Our model results suggest that hours worked may prove a better proxy for labour productivity. We found that age and experience are relevant to explaining firm-level productivity, and our results also indicate that it may be useful to consider job tenure to measure experience. Finally, there are mixed results with the occupation variables – our proxy for skills. Measures of education attainment might provide a better proxy. Therefore, we conclude that it would be useful to supplement the prototype LEED with key variables such as hours worked, firm-level capital stock and education attainment.

We have also extended the study to consider the impact of multiple job holders in the models. The three level model results are similar after we have taken these multiple job holders into account. One of the reasons is that the prevalence of multiple job holders is low (less than 1%) in this prototype LEED. However, this should be considered in the model as it could become an important estimation issue in larger samples.

We conclude that the LEED is a powerful database with many possible analytical uses.

ACKNOWLEDGEMENTS

The results of these studies are based, in part, on tax data supplied by the Australian Taxation Office to the Australian Bureau of Statistics under the *Taxation Administration Act 1953*, which requires that such data is only used for the purpose of administering the *Census and Statistics Act 1905*. Legislative requirements to ensure privacy and secrecy of this data have been adhered to. In accordance with the *Census and Statistics Act 1905*, results have been confidentialised to ensure that they are not likely to enable identification of a particular person or organisation.

Any discussion of data limitations or weaknesses is in the context of using the prototype linked employer-employee database for the productivity analysis project undertaken by the ABS. The discussion is not related to the ability of the data to support the ATO's core operational requirements. Any findings from this paper are not official statistics and the opinions and conclusions expressed in this paper are those of the authors. The ABS takes no responsibility for any omissions or errors in the information contained here.

We gratefully acknowledge several people who provided inputs for this paper: Daniel Elazar, Ric Clarke, Siu-Ming Tam and Peter Rossiter for their comments and guidance at various stage of drafting; Australian National University advisers – Alan Welsh, Robert Breunig, Tue Gorgens and Hoa Nguyen; our ABS colleagues – Derek Burnell, David Taylor and Deepa Wright for their comments; and those who attended our presentation at the Asia Pacific Productivity Conference 2014 and provided comments that helped improve our paper. We remain solely responsible for the views expressed in this article.

CONTENTS

USE OF A PROTOTYPE LINKED EMPLOYER-EMPLOYEE DATABASE TO DESCRIBE CHARACTERISTICS OF PRODUCTIVE FIRMS

Chien-Hung Chien and Andreas Mayer
Data Standards and Methods Branch

ABSTRACT

This study uses a prototype linked employer-employee database (LEED) to analyse both employee and firm characteristics to identify factors that explain differences in labour productivity across firms and industries. We created the prototype LEED by linking de-identified individual Personal Income Tax and Business Tax data from the Australian Taxation Office (ATO) with the Australian Bureau of Statistics (ABS) Business Longitudinal Database (BLD), for the 2010–11 financial year. We demonstrate the analytical potential of the prototype LEED by constructing multilevel models to describe employer and employee characteristics of productive firms. The hierarchical structure of the prototype LEED lends itself to using multilevel models to capture the dynamics between firms and employees. A LEED is a rich database that provides a great opportunity for further labour and productivity research. We have proposed some key areas to further develop this preliminary research.

1. INTRODUCTION

This study uses a prototype linked employer-employee database (LEED) to analyse both employee and firm characteristics to identify factors that describe labour productivity across firms and industries. We created the prototype LEED by linking de-identified Personal Income Tax and Business Tax data from the Australian Taxation Office (ATO) with the Australian Bureau of Statistics (ABS) Business Longitudinal Database, for the 2010–11 financial year. This is the first Australian prototype LEED constructed by linking administrative and ABS survey data sources rather than conducting a survey.[1] This prototype contains employer and employee level information and can therefore be used to study these two interrelated factors in the labour market simultaneously, and their impact on labour productivity.

This paper uses multilevel models that capture the dynamics between firms and employees. We construct both firm-level and person-level multilevel models across industries. The main advantages of using this technique are that it captures the cross-level relationship and it makes better use of the hierarchical structure of the prototype LEED to better understand the statistical relationships. The main drawbacks of using this approach are that it assumes normally distributed error terms and it needs a large sample size at the industry level. We compare results from the models using different estimation methods, including the Bayesian credible and 95% confidence intervals for the estimated coefficients to overcome the violation of the normality assumption. The paper demonstrates the analytical potential of this prototype LEED by using it to describe the firm and employee characteristics of productive firms, rather than providing evidence to explain underlying factors which are associated with labour productivity growth.

1 Chien *et al.* (2012) provided a detailed discussion on the advantages and disadvantages of using different methods to construct a LEED.

2. LITERATURE REVIEW

A linked employer-employee database (LEED), which as its name suggests contains employer and employee level information, can be used to study these interrelated factors in the labour market simultaneously. These interrelated factors can be used to measure (i) supply-side factors including employee outcomes, e.g. wage levels and distributions of workers' characteristics such as age; and (ii) demand-side factors consisting of workplace outcomes, e.g. growth in employment determined by business performance (including profitability and productivity). An extensive review by Abowd and Kramarz (1999) showed that many international studies have used a LEED to gain a better understanding of labour market dynamics. Examples include analysing compensation, mobility, unemployment and productivity (Gray *et al.*, 2005; Leonard *et al.*, 1999).[2]

This prototype Australian LEED contains useful information on both employee characteristics (e.g. Age) and firm-level characteristics (e.g. Turnover) for productivity analysis. Bachmann and David (2009) highlighted the importance of capturing both employee and firm-level heterogeneity in analysing labour market dynamics, which the prototype LEED does. It can be used to better understand how the labour market interacts with the economic environment. The information can be used to derive useful contextual (or explanatory) variables such as per employee profitability and turnover, age/gender profile of the workers, and income profile by firm and industry for statistical modelling (Dixon, 2007).

There are three dimensions distinguishing different types of LEED. First, some are cross sectional databases and others are longitudinal. Second, some data designs emphasise employee samples such as Australia's *Survey of Employment and Unemployment Patterns*[3] (SEUP), while others focus on firms. Lastly, some are constructed by conducting a survey of both employers and employees, e.g. Statistics Canada's *Workplace and Employee Survey*[4] and SEUP, while others use a mixture of surveys and administrative records, e.g. the New Zealand LEED[5] and the prototype LEED used in this paper.

---

2 Gray *et al.* (2005) used longitudinal data from the Australian Survey of Employment and Unemployment Patterns (SEUP) to compare the labour market dynamics of the unemployed, marginally attached and non-attached workers.

3 The survey was a longitudinal survey, which provides information on the dynamics of the Australian labour market, conducted in three waves covering the period September 1994 to September 1997. The target population was people considered to be most likely currently eligible for labour market assistance or to become eligible for assistance in the near future (ABS, 2005).

4 The survey collects statistics on employers and their employees and links data at the micro data level. The employees who respond to the survey are selected from within sampled workplaces. The information from both the supply and demand sides of the labour market is available for analysis (Statistics Canada, 2009).

5 New Zealand integrated the Inland Revenue Department (IRD) Pay as You Earn and income tax data with business data from their Longitudinal Business Frame (Statistics New Zealand, 2003).

A review by Bartelsman and Doms (2000) on the empirical use of longitudinal micro data for productivity analysis divided the type of use into two groups – those describing productivity and those examining the factors behind productivity growth. The first group of papers document the cross sectional distributions of productivity and present the stylised facts on the dispersion of productivity across firms (see Devine *et al.*, 2012).[6] The second group focuses on the fundamental questions in productivity analysis by asking what are the factors underlying productivity growth? Some factors that have been investigated include managerial practice, technology, diversity, quality of inputs and regulation (see Syverson, 2010;[7] Mahlberg *et al.*, 2011[8] and Parrotta *et al.*, 2012[9]).

Hildreth and Pudney (1999) discussed the statistical properties of different methods to create a LEED. They highlighted several major problems in the analysis of these linked databases, including (i) the absence of key variables at the employee level which are important to individual productivity; and (ii) the negative effect of the non-representative samples in the LEED on the estimated model parameters, which resulted from the poor linking process.

This prototype Australian LEED, because of the limited selection of the databases for linking, does not include some key employee variables (e.g. education attainment and employment tenure) that are important for measuring labour quality (Fox and Smeets, 2011). Li (2013) identifies that education attainment and hours worked, which are not yet available in the prototype, are important statistical measures for labour quality. However, the prototype LEED does contain useful information on employee characteristics such as age, sex and occupation etc. for individuals. Moreover, the data quality issues associated with poor linkage do not affect this prototype database because the linking process is deterministic (using unique keys) and hence has excellent matching accuracy.

---

6 Devine *et al.* used a LEED to describe the productivity dispersion within New Zealand industries and found that including labour quality by using wage bills as a proxy can reduce productivity dispersion across different industries.

7 Syverson discussed the importance of capturing the quality of labour capital inputs to explain the underlying productive differences between firms. However, it is an ongoing challenge to provide a finer labour skills measure in a LEED.

8 Mahlberg *et al.* conducted a panel regression on a linked employer-employee database for 2002–2007 to understand the effect of ageing on wages and firm productivity across industry sectors in the Austrian economy. Their results showed that there is a positive correlation between the share of older employees and firm level productivity.

9 Parrotta *et al.* used fixed effect estimation techniques to analyse the Denmark LEED to understand the effects of labour diversity on firm level productivity. They found that labour diversity in education significantly enhances a firm's value added.

This paper adds to the literature by describing firm and person characteristics that explain labour productivity for small and medium size firms in Australia.[10] The aim is to demonstrate the analytical potential of this database rather than providing evidence to explain underlying factors which drive labour productivity growth. It uses a multilevel model framework to describe labour productivity across firms, clustered within industries. There are few examples of productivity studies that use multilevel modelling techniques. Similar studies using this approach look at productivity in the education sector (Hanchane and Mostafa, 2010), in the health sector (Grassetti *et al.*, 2005), or evaluating differentials in individual wage policy setting in firms (Cardoso, 2000). We are not aware of any studies using this approach to describe labour productivity across firms and industries in Australia.

This paper is organised as follows: Section 3 describes the prototype LEED including its sources, creation process and quality issues; Section 4 discusses the model specification; Section 5 shows the model selection process and considers the estimated results; and Section 6 concludes and proposes future directions.

---

10 This study focuses on hiring firms with up to 199 employees. The size of firms is set to be in line with the scope of the Business Longitudinal Database (BLD) which provides the additional firm level characteristics for analysis (ABS, 2013).

3. DATA DESCRIPTION

3.1 Data sources

This prototype LEED is created by linking data from the ABS and the Australian Taxation Office. The ABS' Business Longitudinal Database (BLD) provided the subset of firms that we focused on, as well as a number of detailed firm-level variables. These firms were then linked to business and personal income tax records from the ATO, including the Business Activity Statement (BAS), Business Income Tax data (BIT), Personal Income Tax data (PIT), and Pay As You Go (PAYG) data.[11] Note that the information contained in both the BLD and tax records is not collected for creating a LEED. The discussion on data quality here focuses on the extent to which these datasets are suitable for the production of the prototype LEED – we are not calling into question the suitability of these datasets for the purposes for which they were collected. This section discusses the process of producing the prototype Australian LEED and its quality in the context of this paper, and identifies future statistical opportunities for assisting informed decision making (see Appendix A for detailed descriptions of the data sources).

3.2 Data creation

The LEED is assembled by deterministically linking firm-level records, identified by Australian Business Number (ABN), to person-level variables, identified by Scrambled Tax File Number (STFN), through the PAYG records, which contain ABNs and STFNs.[12] The linking process is of high quality, and issues associated with missing data for particular variables do not affect the linking quality. To create the LEED, we began by linking the BAS and BIT tax data to the BLD data, which formed the subset of Australian firms that we focused on. Next, using the PAYG data, all payment records were selected which had ABNs in the BLD and STFNs in the PIT (excluding non-lodgers). Where multiple records had the same ABN-STFN combination, we combined them into a single record by summing the PAYG wage and salary income for that person from that ABN, resulting in one record per STFN-ABN combination. These were then linked to the firm data using ABN. The true annual employee count for each firm could then be derived by counting the number of STFNs associated with that ABN, and then non-hiring firms and firms with greater than 199 employees were removed. The result is linked employee-employer records for BLD firms employing between 1 and 199 employees (inclusive). We consider a cross section of data, from the 2010–11 financial year.

---

11 The ATO collects this data for compliance purposes, with statistical production not in mind.

12 The ATO does not provide real TFNs to the ABS to protect confidentiality.

3.3 Linked data quality

We consider its coverage and representativeness in terms of statistical analysis. Looking at the coverage of the raw data, all of the variables considered for analysis in our models have a low or negligible degree of missing data. Employee level variables have at most 3% of values missing, and firm-level variables have at most 6% of values missing, with the exception of our derived capital stock measure (14.2% missing).[13] Note that the log transformations were done after adding 1 to the value of each variable, so where the original value was 0 it did not become a missing value.[14] Where different data sources used to construct the LEED were inconsistent (e.g. missing values for the firm's industry division) we used the data source which was most fit for purpose for our analysis and provided the most reliable values. This was a minor issue affecting only a small number of firms.

Tables 3.1 and 3.2 present summary statistics for the variables in both the two and three level models (see Section 4). Where a variable has been divided into ranges/groups, we indicate the summary values for each group. More detailed discussion on the derivation of these variables is in Appendix B. Note that we have more than 5,000 firms and 100,000 employees.[15]

3.1 Summary statistics

	Mean	*St Dev*	*Missing*	*Q1*	*Median*	*Q3*
L.E.Turnover_exGST (Firm)	11.67	1.14	0	11.01	11.67	12.34
LPers_Turnover_exGST (Person base)	11.10	1.77	0	10.08	11.38	12.34
LMM_Turnover_exGST (Person MM)	11.10	1.76	0	10.08	11.38	12.35
L.E.Capital_Stock	9.95	2.41	723	9.04	10.14	11.26
L.E.CapEx	4.67	4.19	0	0.00	5.83	8.34
L.E.OExp_exGST	10.97	1.68	168	10.20	11.05	11.92
PT_pv	0.39	0.37	280	0.00	0.33	0.72
Perm_pv	0.73	0.34	280	0.50	0.88	1.00
Sex_pv_Female	0.39	0.32	0	0.10	0.33	0.60
Age_R_pv_0_to_29	0.32	0.28	0	0.00	0.28	0.50
Age_R_pv_30_to_44	0.31	0.25	0	0.11	0.29	0.43
Income_R_pv_0_to_24999	0.35	0.31	0	0.10	0.29	0.53
Income_R_pv_25000_to_49999	0.35	0.27	0	0.17	0.33	0.50

Source: ABS unpublished prototype LEED

13 See Appendix B for details of derivation.

14 The transformation is necessary in dealing with the large variation in the scale of the variables.

15 We cannot disclose exact total counts due to confidentiality reason.

3.2 Summary statistics – Dummy variables

	Proportion	*Missing*
Dummy variables – Firm		
d_forown	4.9%	245
DExporter	11.5%	0
DProfLoss_R_0_to_9999	32.6%	217
DProfLoss_R_10000plus	37.9%	217
Dummy variables – Person		
DSex_Female	37.7%	63
DAge_R_0_to_29	36.4%	64
DAge_R_30_to_44	32.8%	64
DIncome_R_0_to_24999	31.6%	0
DIncome_R_25000_to_49999	33.9%	0

Source: ABS unpublished prototype LEED

The firms within the data are divided into 14 industry divisions, and considerable differences in the distribution of employees are visible across industries. The mining, construction and transport/storage industries have a low proportion of female employees, whereas most other industries are relatively balanced across the genders. For employee age, the proportion of employees aged 30 to 44 years is relatively similar, but the Accommodation and Food Services (H) and Arts and Recreation Services (R) industries have a much lower proportion of workers aged 45+ years and a higher proportion of workers aged under 30 years. By contrast, the Transport, Postal and Warehousing (I) industry has fewer workers aged under 30 years and a higher proportion of workers aged 45+ years. In terms of employee income, the Agriculture, Forestry and Fishing (A), Accommodation and Food Services (H) and Arts and Recreation Services (R) industries have a high proportion of low income works (earning less than $25,000) and low proportion of high income workers (earning $50,000 plus), whereas for the Mining (B), Construction (E) and Professional, Scientific and Technical Services (M) industries the situation is reversed. This illustrates the need to consider firms as clustered within industries when analysing the data.

In determining the representativeness of the linked data for statistical analysis, the following factors must be considered: first, the BLD contains a subset of firms, drawn from those which have a simple structure, fewer than 200 employees, and excluding certain industries (see Section A.1 in Appendix A); and we have further excluded non-hiring firms (deriving employment count from the PAYG records). This means that the firms present in the prototype LEED are not necessarily representative of the Australian business community as a whole, and hence conclusions drawn from the descriptive analysis we perform may not be applicable to the wider economy. With that in mind, we note that the firms included in the prototype LEED have a broadly equal spread across the 14 industry divisions included. There is a need to compare our results with similar studies with a larger sample size to test if our results are representative of the whole economy.

Second, to conduct firm productivity analysis with causal interpretation, there is a need to expand the prototype database to include other key variables, as discussed previously. In particular, we have no measure of hours worked or educational attainment so we cannot reliably measure labour quality. The intermediate firm inputs available in the prototype data do not distinguish inputs used in the production process or resale stock. The distinction is important to measure productivity using a value added measure. For capital, the data provides a measure of the stock of assets and the flow of capital expenditure, but not directly of the stock of capital. We derive a basic capital stock measure (see Appendix B), though our modelling results suggested that using capital expenditure is more appropriate for our analysis. Finally, in the future this database can be extended to become a longitudinal database which would allow analysts to better model correlation between changes in labour state and firm performance.

The descriptive productivity analysis of the prototype Australian LEED, as a proof of concept, has given the ABS the opportunity to explore many aspects of the linked data to evaluate its potential for statistical production. We have found that the data is very rich and provides an excellent database. The analysis shown in this paper is only one possible way of analysing this database. Chien *et al.* (2012) provides a list of examples including:

- statistics on labour market dynamics for better measuring job creation/ destruction and understanding the relationship between earnings mobility and business competitiveness. Farmakis-Gamboni *et al.* (2012) highlighted some current statistical gaps and proposed the strong need for a LEED in the minimum wages research context.
- information to measure firm-level productivity by capturing supply and demand of human capital and the associated characteristics of employers and employees. Leonard *et al.* (1999) suggested that this information is important to better assess the relationship between pay policies and firm productivity.
- disaggregate data for regional analysis. The data can be used to compare employment and earning trends across geographic areas and detailed industries to assist in developing regional economic or social policies.

4. STATISTICAL MODELS SPECIFICATION

Fox and Smeets (2011) suggested that the differential firm outputs can be decomposed by differences in the inputs such as capital, materials and labour spent, and unexplained residuals in the production process. A simple Cobb-Douglas production function can be expressed as:

$$\log Y_j = \lambda_0 + \lambda_K \log K_j + \lambda_L \log L_j^* + e_j$$

where Y_j is the gross output of firm j, i.e. turnover[16] as the productivity measure, K_j is the physical capital, L_j^* is the labour characteristics and e_j is the residual. λ_K and λ_L are the elasticities of capital and labour. We modified the basic production function by (i) decomposing the labour input into a set of labour characteristics which associate with firm productivity and (ii) introducing a set of firm characteristics which associate with labour productivity as contextual variables.[17] Crépon *et al.* (2003) proposed that labour characteristics, L_j^*, can be decomposed into a weighted sum of different characteristics p of employee j. The weights are represented by an individual productivity factor λ_{pj}.

$$\begin{aligned} L_j^* &= \sum_{p=0}^{P} \lambda_{pj} X_{pj} = \lambda_{0j} X_j \left[1 + \sum_{p=1}^{P}\left(\frac{\lambda_{pj}}{\lambda_{0j}} - 1\right)\frac{X_{pj}}{X_j}\right] \\ \ln\left(L_j^*\right) &= \ln\left(\lambda_{0j}\right) + \ln\left(X_j\right) + \ln\left[1 + \sum_{p=1}^{P}\left(\frac{\lambda_{pj}}{\lambda_{0j}} - 1\right)\frac{X_{pj}}{X_j}\right] \end{aligned} \quad (1)$$

where

- X_{pj} is a set of employee characteristics such as age or sex etc.;
- λ_{0j} is the labour productivity of the reference group of employees;
- $\beta_{pj} = \left(\dfrac{\lambda_{pj}}{\lambda_{0j}} - 1\right)$ implies the relative productivity difference between an employee and the reference group of employees, e.g. the marginal productivity differential of an unskilled worker with a group of skilled workers;
- $\ln\left(1 + \sum_{p=1}^{P} \beta_{pj} \dfrac{X_{pj}}{X_j}\right) \approx \sum_{p=1}^{P} \beta_{pj} \dfrac{X_{pj}}{X_j}$, which indicates output per employee can be estimated by equation (2).

---

16 OECD (2001) discussed gross and value added output measures. We have considered both measures but we have found that the gross measure is easier to explain here because there is no data on the firm level deflators.

17 See Appendix E for an expanded mathematical explanation.

We did not consider economies of scale (firm heterogeneity) and technical efficiency in the paper because the aim of the paper is to describe the characteristics of firms who produce higher output per employee (as a measure for labour productivity). In addition, the prototype LEED misses some key variables (e.g. firm specific intermediate inputs) needed to use the standard approach for analysis. Thus interpretation of the elasticity is not about determining causal effects, rather the coefficients, through their relative signs and magnitudes, indicate the strength of association between these characteristics and firm outputs per employee.[18]

4.1 Two-level model (firm-industry)

We constructed multilevel models that capture the dynamics between firms and employees. The two-level model is specified as:

$$Y_{kj} = \beta_{0k} + \sum_{q=1}^{q_n} \beta_{qk} Z_{qkj} + \sum_{q=q_n+1}^{Q} \beta_{qk} \frac{X_{qkj}}{X_{kj}} + r_{kj} \qquad (2)$$

where

- Y_{kj} is the log of firm-level labour productivity derived by $\ln\left(\frac{\text{Turnover}_{kj}}{\text{Emp}_{kj}}\right)$, where Emp_{jk} is the number of employees for firm j in industry k. Due to a lack of credible firm-level deflators and that we only consider single year, we chose the gross output measure, i.e. Turnover (excluding GST). We also normalise it by dividing the turnover by the number of employees;
- $\left\{Z_{qkj} : q = 1, \ldots, q_n\right\}$ are the q_n firm-level explanatory variables such as investment, operating expenses and profit/loss dummies for firm j in industry k;
- $\left\{\frac{X_{qkj}}{X_{kj}} : q = q_n + 1, \ldots, Q\right\}$ are variables measuring the proportion of employees (for each firm) with a given characteristic, e.g. age, sex, income and occupation at the firm level. These are crude measures of the distribution of employee characteristics across the firm. Other key employee variables such as education and hours worked are not available in this prototype database; though the proportion of part time workers is available at the firm level;
- β_{0k} is the intercept for industry k;

18 Our focus is on the coefficients and thus we do not derive any MFP related measures for the model.

- $\{\beta_{qk} : q = 1, \ldots, Q\}$ are the corresponding firm-level coefficients that indicate the direction and strength of association between each firm characteristic q and the outcome in industry k;
- r_{kj} is the model error term that represents the deviation of firm j's observed outputs in industry k from the predicted outputs based on the firm-level model.[19]

The second-level (at the industry-level) model describes the productivity differences within and across industries. Both the intercept β_{0k} and the slope β_{qk} are industry dependent and can be split into an overall average and an industry specific random effect γ_{q0}, i.e. allowing the slope of the variable to change by industry, which can be expressed as:

$$\beta_{0k} = \gamma_{00} + \sum_{s=1}^{S} \zeta_{s0} V_{sk} + u_{0k} \tag{3}$$

$$\beta_{qk} = \gamma_{q0} + u_{qk} \tag{4}$$

where

- γ_{00} is the average intercept across all industries;
- γ_{q0} is the average regression slope across different industries for firm characteristic q;
- $\{V_{sk} : s = 1, \ldots, S\}$ are the S industry-level explanatory variables, each of which is formed as the industry mean of the firm-level variable for industry k;
- ζ_{s0} are the corresponding industry-level coefficients that indicate the direction and strength of association between each industry characteristic s and the firm output. Note that these coefficients do not vary across industries;
- u_{0k} is the industry dependent deviation for the intercept;[20]
- u_{qk} is the industry random slope effects associated with the firm characteristic q (Bryk and Raudenbush, 1992).[21] We only have one random slope here, firm operating expenses; for the rest of our firm variables u_{qk} is zero.[22]

---

19 r_{kj} is assumed to be normally distributed with $\mu(r_{kj}) = 0$ and $\mathrm{var}(r_{kj}) = \sigma^2$. A discussion on the violation in this assumption can be found in the next section.

20 u_{0k} is assumed to be normally distributed.

21 u_{qk} is assumed to be normally distributed.

22 We tested other variables but the random slopes were not significant for them.

The first level model is nested within the second level model by substituting β_{0k} and β_{qk} then we have:

$$\begin{aligned} Y_{kj} &= \gamma_{00} + \sum_{s=1}^{S} \zeta_{s0} V_{sk} + \sum_{q=1}^{q_n} \gamma_{qk} Z_{qkj} + \sum_{q=q_n+1}^{Q} \gamma_{qk} \frac{X_{qkj}}{X_{kj}} \\ &\quad + u_{0k} + \sum_{q=1}^{q_n} u_{qk} Z_{qkj} + \sum_{q=q_n+1}^{Q} u_{qk} \frac{X_{qkj}}{X_{kj}} + r_{kj} \end{aligned} \tag{5}$$

4.2 Three-level (employee-firm-industry) model

We also constructed a three-level model to see whether this gave a better fit. The person-level model is specified as:

$$Y_{kji} = \alpha_{0kj} + \sum_{p=1}^{P} \alpha_{pkj} X_{pkji} + e_{kji} \tag{6}$$

where

- Y_{kji} is the log of person-level Turnover derived by $\ln\left(\text{Turnover}_{kj} \times \text{WPAYG}_i\right)$ for all employees i who receive a Pay-As-You-Go (PAYG) payment from firm j in industry k.[23] We derive employee-level Turnover by dividing a firm's Turnover between its employees according to their wage share (making the simplifying assumption that a person's contribution to firm production is proportional to their wage received from that firm);
- $\left\{X_{pkji} : p = 1, \dots, P\right\}$ are the person-level characteristics, e.g. age, sex and occupation etc.;[24]
- α_{0kj} is the intercept for firm j in industry k;
- $\left\{\alpha_{pkj} : p = 1, \dots, P\right\}$ are the corresponding employee level coefficients that indicate the direction and strength of association between each employee characteristic and employee-level Turnover;
- e_{kji} is the model error term that represents the deviation of person i's contribution to firm j's output in industry k from the predicted employee-level Turnover.

23 Note that $\text{WPAYG}_i = \dfrac{\text{PAYG}_i}{\sum_{i=1}^{I} \text{PAYG}_i}$ and $\sum_{i=1}^{I} \text{WPAYG}_i = 1$.

24 The income or earning variables are excluded in the explanatory variables to avoid problem of endogeneity.

The second- or firm-level model describes the productivity differences explained by firm-level variables, while the third- or industry-level model describes the productivity difference across industries and summarises the similarities and differences between firms effectively. The combined firm-level and industry-level models can be expressed as:

$$\alpha_{0kj} = \delta_{000} + \sum_{s=1}^{S} \zeta_{s00} V_{sk} + \sum_{q=1}^{Q} \beta_{qk0} Z_{qkj} + u_{0k0} + v_{0kj} \quad (7)$$

$$\alpha_{pkj} = \delta_{p00} \quad (8)$$

$$\beta_{qk0} = \gamma_{q00} + u_{qk0} \quad (9)$$

where

- δ_{000} is the overall intercept across all industries;
- δ_{p00} is the average regression slope across different industries for person characteristic p;[25]
- γ_{q00} is the average regression slope across different industries for firm characteristic q;
- $\{V_{sk} : s = 1, \dots, S\}$ are the S industry-level explanatory variables, each of which is formed as the industry mean of the firm-level variable for industry k;
- ζ_{s00} are the corresponding industry-level coefficients that indicate the direction and strength of association between each industry characteristic s and the employee-level Turnover;
- $\{\beta_{qk0} : q = 1, \dots, Q\}$ are the corresponding firm-level coefficients that indicate the direction and strength of association between each firm characteristic and employee-level Turnover;[26]
- $\{Z_{qkj} : q = 1, \dots, Q\}$ are the Q firm-level explanatory variables such as investments, operating expenses and foreign ownership dummy for firm j in industry k;
- u_{0k0} is the industry dependent deviation from the total industry intercept;
- v_{0kj} is the firm dependent deviation for firm j from the intercept;

25 For consistency with the two-level model, we do not let any of the person-level variables vary by firm or industry.

26 See Snijders and Bosker (1999) and Bryk and Raudenbush (1992).

- The slope β_{qk0} are industry dependent and can be split into an overall average γ_{q00} and an industry dependent deviation of slope u_{qk0}, i.e. allowing the slope of the firm variable to vary by industry (but not by firm). Note that we allow for each firm within each industry to have a different intercept, but for the random slope we only allow this to differ by industry to ensure the consistency with the two-level model. Again u_{qk0} is zero for all variables except firm operating expenses.

Substituting α_{0kj}, α_{pkj} and β_{qk0} yields:

$$\begin{aligned} Y_{kji} &= \delta_{000} + \sum_{s=1}^{S} \zeta_{s00} V_{sk} + \sum_{q=1}^{Q} \gamma_{q00} Z_{qkj} + \sum_{p=1}^{P} \delta_{p00} X_{pkji} \\ &\quad + u_{0k0} + \sum_{q=1}^{Q} u_{qk0} Z_{qkj} + v_{0kj} + e_{kji} \ . \end{aligned}$$

5. MODEL SELECTION AND ESTIMATION RESULTS

We present a series of models which describe firm and employee characteristics that describe labour productivity, focusing on small and medium size firms. The main goal of this paper is to demonstrate the analytical potential of the prototype LEED and not to draw conclusions on determining causal relationships. The modelling objective is to use the method which can best describe the prototype LEED. It is therefore quite natural to consider using a multilevel modelling framework because of the nested structure of this database, i.e. an employee i works in firm j or firm j operates in industry k, to better capture the within and between group variations (Gelman and Hill, 2006).

The main advantages of using this technique are that it captures the cross-level relationship, and it makes better use of the hierarchical structure of the prototype LEED to better understand the statistical relationships (Snijders and Bosker, 1999, Grassetti *et al.*, 2005). This method does not assume that error terms have equal variance across different industries. When the data has a nested structure, the observations within groups have similar characteristics because of the selection process and it is not appropriate to use OLS regression (Hox, 2010). Lastly, the technique also provides more accurate inferences as it takes into account the homogeneity within a firm or of firms within an industry.

Parameters in the multilevel models, including our models, are often estimated by using the maximum likelihood (ML) estimation method. A key assumption underlying the use of the ML method is that the error terms are distributed normally. If there is a violation of this assumption, the asymptotic errors are incorrect which leads to inaccurate confidence intervals. This is particularly important at the higher level (i.e. industry) for the random coefficient (Maas and Hox, 2004a). In addition, the variance component can be underestimated if the sample size at the higher level is too small.[27] In our case we have 14 industries. We constructed the Bayesian Credible and 95% Confidence Intervals and compared the results between the two and three level models to overcome the violation of the normality assumption. A detailed discussion of the tests and results follow in the next section.

We constructed a base (two-level) model and we also used a variant (three-level) model to verify the base model results. The same model selection process is used for both base and variant models. These models are constructed by regressing dependent variables against a number of variables within the prototype LEED, which fit the Cobb Douglas production function and relate to labour productivity. We then removed

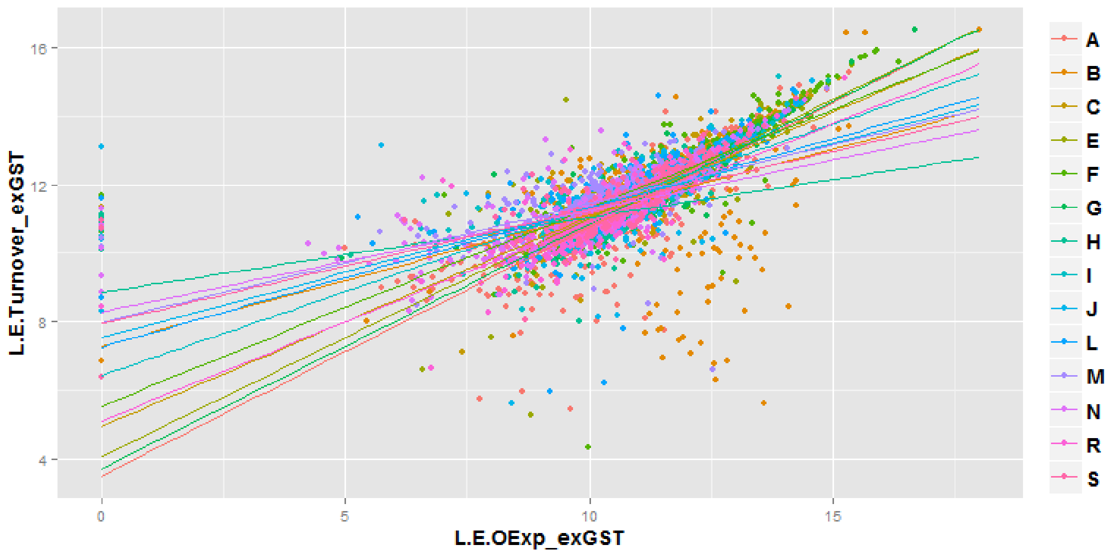
27 The literature on the ideal group size is inconclusive. Browne and Draper (2000), cited in Maas and Hox (2004b), suggested that as few as 6 to 12 is sufficient. In contrast, Van der Leeden *et al.* (1994), also cited in Maas and Hox (2004b), indicated more than 100 groups are needed.

those variables which were statistically insignificant at the 10% level.[28] We also found that doing so lowered the Akaike information criterion (AIC), implying the smaller model better explains the data. A discussion of the variable choice and derivation can be found in Appendix B.

5.1 Why multilevel modelling?

We first demonstrate that there is sufficient variation in each industry to justify the use of multilevel models by considering one of our key firm-level variables, logged per-employee operating expenses.

5.1 Random intercepts and slopes, by industry



See table B.2 in Appendix B for a list of industry codes.
Source: ABS unpublished prototype LEED.

Chart 5.1, logged per-employee Turnover regressed against logged per-employee firm operating expenses, shows that each industry has a different intercept and slope for logged per-employee Turnover. This suggests that we can use multilevel modelling at the industry division level. Empirical estimation is also used to confirm the visual diagnostics by comparing the results with and without random intercepts.[29] The likelihood ratio between these two models showed that they were significantly different at the 1% level. These results suggest that it is appropriate to construct a multilevel model with random intercepts. We likewise compare the multilevel model

---

28 These included the number of firm locations, dummy variables for three firm size categories, firm/employee location variables (at Australian state level), firm internet use, flexible working arrangements offered by the firm, an exporter dummy, employee skill level proportions (derived from employee occupation) and a dummy variable for the firm implementing innovation in the last twelve months. Note that we also compared models with (logged per-employee) Capital Stock and with Capital Expenditure; while Capital Stock was significant in some cases, Capital Expenditure resulted in a much lower AIC value, so we chose to model with Capital Expenditure.

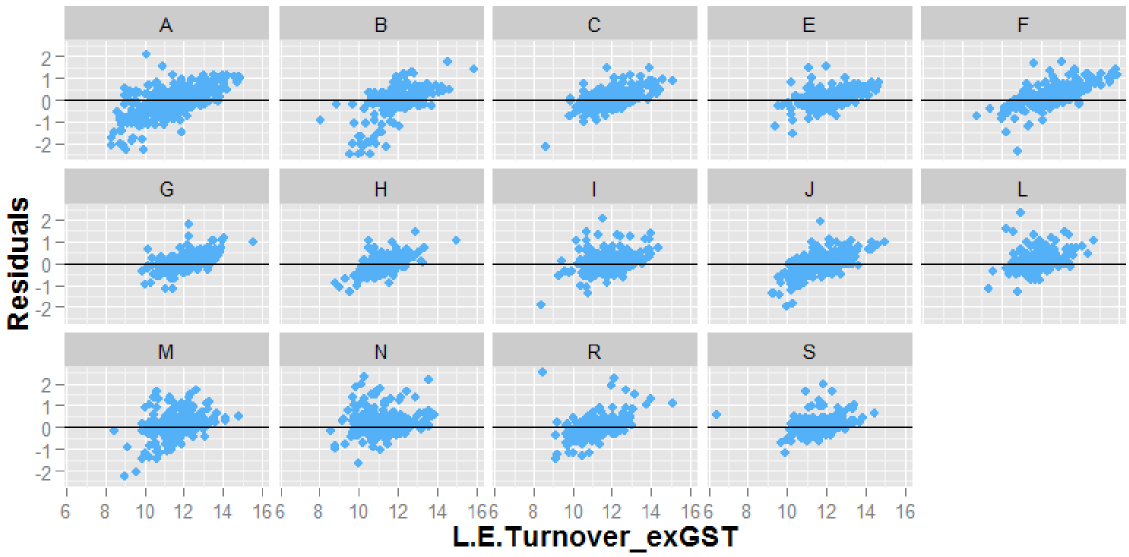
29 We used generalised least squares to estimate a model without random slopes or intercepts.

with just a random intercept with a multilevel model which also includes a random slope for logged per-employee Operating Expenses.[30] These models were also significantly different at the 1% level, showing that there are variations in the slopes for Operating Expenses across different industries.

5.2 Testing and resolving estimation issues for the two-level model

As our next step we consider and remove outliers.[31] The residuals for several firms in the Mining sector in particular have a much wider spread when we plot against logged per-employee Turnover in comparison with other industries. Chart 5.2 shows the results after removing these outliers.

5.2 Residual plots against logged per employee Turnover (excluding outliers), by industry



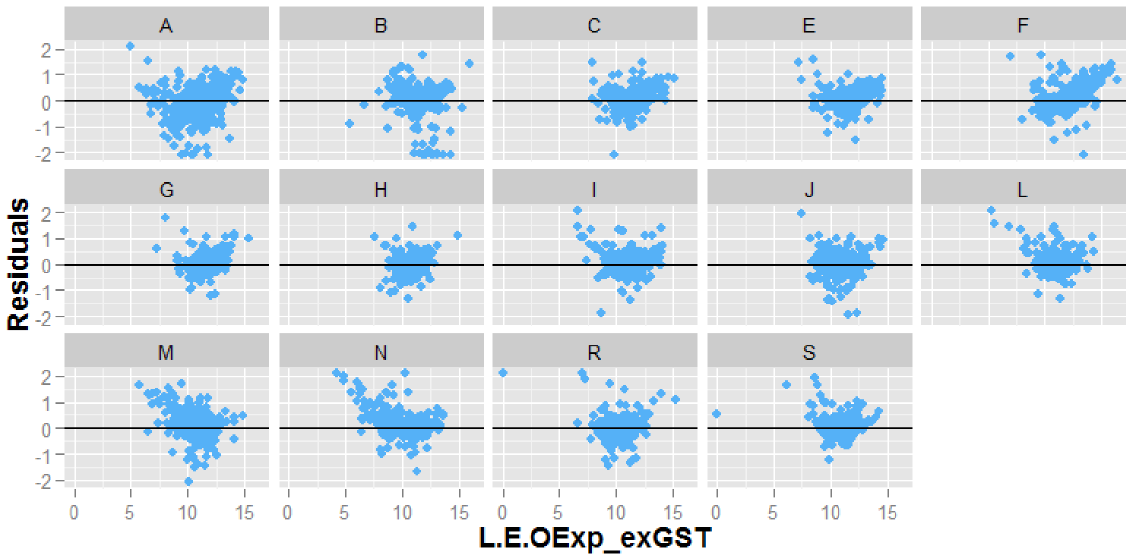
See table B.2 in Appendix B for a list of industry codes.
Source: ABS unpublished prototype LEED.

After excluding outliers, we then tested the model for heteroscedasticity, assessed whether the residuals were normally distributed, and tested for endogeneity. To assess heteroscedasticity, we graphed the residuals against each of the explanatory variables, divided into the 14 industry divisions. The only variable with clear patterns in the residuals was logged per-employee Operating Expenses. Chart 5.3 shows that it displayed some curvature, possibly indicating the need for a quadratic term in the model. We resolved this by including a squared term for that variable in our model; this term was significant at the 1% level, and reduced the Akaike Information Criterion, demonstrating a clear increase in explanatory power.

---

30 We also tried a random slope for Capital Expenditure but this was not statistically significant.
31 All observations whose residual was below –2.3 or above +2 were excluded.

5.3 Residual plots against logged per-employee Operating Expenses, by industry



See table B.2 in Appendix B for a list of industry codes.
Source: ABS unpublished prototype LEED.

The continuous firm variables, (logged per-employee) Operating Expenses and Capital Expenditure, could plausibly be endogenously related to our dependent variables. Marschak and Andrew (1944), cited in Fox and Smeets (2011), suggested that more productive firms are likely to use more inputs which can lead to overestimating the input coefficients. We suspected that the investments and operating expenses variables can cause the problem of endogeneity. We followed the steps suggested by Spencer and Fielding (2000) to perform the tests using instrumental variables. We also tested for correlation between the random effects and the fixed predictors (Rice *et al.*, 1997). Our test results showed no presence of endogeneity. Some possible explanations for this include (i) we use proxy variables i.e. Capital Expenditure and Operating Expenses instead of Capital Stock and Intermediate Materials (Levinsohn and Petrin, 2000), (ii) we include the proportions of many employee-level characteristics which can lessen the correlation between model residuals and these firm-level explanatory variables.

Both QQ-plots of the standardised residuals by industry and a Jarque-Bera test suggest that the normality assumption is violated.[32] It is also useful to compare the asymptotic maximum likelihood standard errors with the robust standard errors as a way of appraising the possible effect of model misspecification (Maas and Hox, 2004b). We have seen that these standard errors from the two estimators are similar and hence there appears to be no misspecification problem. We also use Bayesian estimation to determine credible intervals for these parameters, rather than relying on the standard errors from our multilevel modelling. We used flat (non-informative)

32 It is significant at the 1% level.

priors for the fixed effects and random effects (intercept and slope) to determine the robust credible intervals for the parameters. We also checked that the Markov Chain Monte Carlo iterations show no trend and the posterior density estimates of the parameter are normally distributed to ensure convergence (Hadeld, 2014).

5.3 Testing and resolving estimation issues for the three-level model

It is useful to note some key differences between the two- and three- level models here. Firstly, the two-level model uses a firm-level dependent variable, whereas the three-level model uses a person-level dependent variable derived from this. Secondly, the two-level model represents employee characteristics as the proportion of employees in a given firm exhibiting that characteristic, whereas the three-level model directly uses the person-level dummy variables for each characteristic.[33]

We have followed the same model selection process for the three-level model. We began by testing for random intercepts and found that there is evidence to support theuse of random intercepts, as well as a random slope for Operating Expenses at the industry level.[34] Next, to maintain consistency with our firm-level model, we exclude all observations for those firms which were excluded in the two-level model.

As our third step, we assessed heteroscedasticity, normality and endogenity. Our test results show that there is no heteroscedasticity or endogenity.[35] Again, the normality assumption appears to be violated, and a Jarque Bera test confirms this at the 1% level. We also compare the asymptotic maximum likelihood standard errors with the robust standard errors and we did not have evidence of misspecification. Finally, we constructed the Bayesian credible and 95% confidence intervals for the coefficients.

5.4 Discussion of the results

This paper focuses on what we can observe from firm and employee characteristics which may describe differences in labour productivity. As discussed previously, we are running descriptive models and the coefficients should not to be interpreted as robust empirical estimates to make causal conclusions. In this section, we also compare our results with other empirical findings which use different modelling techniques.

---

33 Note that for part time workers, we only have this information at the firm level, and so the part time-female interaction term remains at the firm level in the three-level model.

34 Note that the two-level model has random intercepts at the industry level, whereas the three-level model has random intercepts at both the firm and industry level.

35 After adding a squared term for per-employee Operating Expenses.

Table 5.4 reports the two-level model estimates, including for comparison the same model without allowing the intercept or slope to vary by industry. Note that the person-level variables are aggregated as the proportion of employees with a given characteristic within a firm in the two-level model. This provides a measure of labour characteristics at the firm level. Column (1) shows the result from the fixed intercept and slope model, column (2) shows the results with random intercept and slope included. Some notable observations include:

Industry level results:

We have tested several industry contextual variables and reported variables that have significant results.[36] These contextual variables can be used to account for industry-to-industry variability and highlights the advantage of using multilevel technique for nested data which could show different results at the different levels (Bickel, 2007). We have found that investment (L.E.CapEx_mean) is negatively significant at the industry level, though it is positively significant at the firm level. The interaction term between the proportion of female employees and proportion of part time employees is negative but insignificant at the industry level, though both industry variables were significantly negative when included individually. We show the interaction term for consistency with our firm-level variables (see below).

Firm level results:

Of the age proportion variables, our proxy for employee experience, one is significant and the other insignificant. This result contradicts the finding of Mahlberg *et al.* (2011) where they found a significant positive correlation between labour productivity and age using the Austrian LEED. Note that a direct comparison is not possible because the differences can come from the estimation method and the linked data used. We still consider age and experience relevant to explain firm-level productivity. Our result may be different if we use different criteria to range the age variable or use a better proxy, e.g. job tenure for experience.[37]

Similarly, we found a mixed result with the occupation proportion variables, our proxy to skills. Our results are similar to Turcotte and Rennison (2004)[38] and they also showed mixed results for significance of the occupation proportion variables.[39] The reference group here is mainly the non-skilled employees (apprentices and trainees). There are some occupations associated with lower labour productivity, but none of these are significant in the firm level model. Those occupations with the

36 e.g. industry average capital expenditure (L.E.CapEx_mean) and proportion of part-time employee (PT_pv_mean) etc.

37 The age proportion variable is ranged to ensure each bin contains a similar size for the analysis.

38 Turcotte and Rennison use Canadian Workplace and Employee Survey Microdata.

39 We cannot make direct comparison due to differences in estimation methods and data.

strongest association with high labour productivity are ICT Professionals (26), Consultants (9C), Specialist Managers (13), Machine and Stationary Plant Operators (71), and Personal Assistants and Secretaries (52), all of which are significant at the 1% level. Because we consider labour productivity at the firm level here, this does not necessarily imply that people in these occupations are more productive, but rather that firms which are more productive hire more employees in these occupations.

The analysis of the two-level model shows that, consistent with findings from Lopes and Teixeira (2012) and Earle *et al.* (2012), firms with a higher level of investment are associated with higher productivity.[40] The interpretation of the firm operating expenses is complicated by the quadratic term. We need to derive the overall elasticity using both the firm operating expenses and its square terms.[41] Higher operating expenses are associated with higher productivity and the effects are stronger with larger operating expenses. As expected we found that more profitable firms are more productive than less profitable ones.[42] In addition, firms with a higher proportion of high paid employees are more productive than firms with a higher proportion of low paid employees. There is a strong association with remuneration to productivity.

The estimated results show that the interaction term between the proportion of female employees and proportion of part time employees is negative in this model. This may partly be an artefact of our dependent variable being per-employee, without adjusting for hours worked (not available in this prototype data).[43] In addition, the labour force survey results suggested that, on average, women work fewer hours than men. This implies that it is important to account for individual heterogeneity in the model and we would remind readers not to draw any causal conclusions from the analysis.

Columns (3) and (4) show the results of Bayesian credible intervals and 95% confidence intervals. We observed that the intervals show consistent results and the confidence intervals provide narrower bounds.[44]

---

40 Please note that we can't make direct comparison with these studies because of different data and methods used.

41 The overall elasticity is calculated by $-0.35 + 2 \times 0.04 \times \text{L.E.OExp\_exGST}$ and the first (10.2), median (11.0) and third (11.9) quartiles. The elasticities are 0.466, 0.530 and 0.602 respectively.

42 The reference group is unprofitable firms.

43 This means that if two firms have the same Turnover, but one firm has all full time employees and the other firm has some part time employees and so more employees overall to cover the same number of hours worked, the latter firm will have a lower per-employee Turnover and so be considered less productive.

44 The exception is for the industry-level variables, which have considerably different estimates in the Bayesian credible intervals.

5.4 Two-level model results

	Fixed Intercept and Slope	*Random Intercept and Slope*	*95% Confidence Intervals*	*Bayesian Credible Intervals*
Intercept	10.63 (0.21) ***	10.64 (0.33) ***	[10.00 ; 11.29]	[7.39 ; 16.77]
Industry level:				
L.E.CapEx_mean	–0.09 (0.01) ***	–0.06 (0.02) **	[–0.09 ; –0.03]	[–1.09 ; 0.52]
PT_pv_mean:Sex_pv_Female_mean	–0.36 (0.12) **	–0.18 (0.14)	[–0.46 ; 0.10]	[–11.37 ; 2.16]
Firm level:				
d_forown	–0.17 (0.03) ***	–0.15 (0.03) ***	[–0.22 ; –0.09]	[–0.21 ; –0.08]
Perm_pv	0.13 (0.02) ***	0.11 (0.02) ***	[0.06 ; 0.15]	[0.06 ; 0.15]
L.E.CapEx	0.01 (0.00) ***	0.01 (0.00) ***	[0.00 ; 0.01]	[0.01 ; 0.01]
L.E.OExp_exGST	–0.33 (0.03) ***	–0.35 (0.04) ***	[–0.44 ; –0.27]	[–0.60 ; –0.16]
L.E.OExp_exGST.2	0.04 (0.00) ***	0.04 (0.00) ***	[0.04 ; 0.05]	[0.04 ; 0.05]
DProfLoss_R reference group Loss				
DProfLoss_R_0_to_9999	0.14 (0.02) ***	0.13 (0.02) ***	[0.10 ; 0.17]	[0.10 ; 0.16]
DProfLoss_R_10000plus	0.41 (0.02) ***	0.40 (0.02) ***	[0.37 ; 0.43]	[0.37 ; 0.43]
Age_R_pv reference group 45+				
Age_R_pv_0_to_29	–0.09 (0.03) **	–0.07 (0.03) *	[–0.12 ; –0.01]	[–0.12 ; –0.01]
Age_R_pv_30_to_44	–0.03 (0.03)	–0.03 (0.03)	[–0.08 ; 0.03]	[–0.08 ; 0.03]
Income_R_pv reference group 50000+				
Income_R_pv_0_to_24999	–0.52 (0.03) ***	–0.54 (0.03) ***	[–0.60 ; –0.48]	[–0.6 ; –0.48]
Income_R_pv_25000_to_49999	–0.25 (0.03) ***	–0.27 (0.03) ***	[–0.33 ; –0.21]	[–0.33 ; –0.21]
OCPTN_cd_pv reference group Unskilled				
OCPTN_cd_pv_11	0.18 (0.08) *	0.15 (0.08)	[–0.00 ; 0.31]	[–0.01 ; 0.30]
OCPTN_cd_pv_12	0.35 (0.11) ***	0.23 (0.11) *	[0.03 ; 0.44]	[0.05 ; 0.47]
OCPTN_cd_pv_13	0.35 (0.09) ***	0.33 (0.09) ***	[0.16 ; 0.50]	[0.14 ; 0.49]
. . .	. . .	. . .	. . .	. . .
OCPTN_cd_pv_26	0.47 (0.11) ***	0.35 (0.11) **	[0.13 ; 0.57]	[0.11 ; 0.55]
. . .	. . .	. . .	. . .	. . .
OCPTN_cd_pv_52	0.35 (0.12) **	0.31 (0.12) **	[0.07 ; 0.54]	[0.09 ; 0.58]
. . .	. . .	. . .	. . .	. . .
OCPTN_cd_pv_71	0.30 (0.12) **	0.32 (0.11) **	[0.10 ; 0.55]	[0.12 ; 0.58]
. . .	. . .	. . .	. . .	. . .
OCPTN_cd_pv_9C	0.45 (0.12) ***	0.33 (0.12) **	[0.10 ; 0.57]	[0.09 ; 0.56]
PT_pv:Sex_pv_Female	–0.12 (0.03) ***	–0.11 (0.03) **	[–0.18 ; –0.04]	[–0.18 ; –0.04]
AIC	5,355.55	5,183.91		
BIC	5,738.24	5,585.74		
Log Likelihood	–2,617.78	–2,528.96		

Detailed parameter estimates for the Occupation dummy variables are provided in table D.1 in Appendix D.
Significance Level: ° is 10%, * is 5%, ** is 1%, *** is 0.1%
Source: ABS unpublished prototype LEED.

As discussed, the three-level model is constructed for comparison purposes to check for consistency of results. The variables and estimation methods used in these models are similar; therefore we expect similar results after allowing for differences due to the different derivations for the dependent variable and different representations of the employee characteristics at the different levels.

Table 5.5 shows the results of the three-level model estimates, including for comparison the same model without allowing the intercept or slope to vary by industry. These results are broadly consistent with our two-level model. Some key differences include:

- The interaction term, at the industry level, is insignificant and negatively associated with labour productivity. This result is consistent with the two-level model. However, the interaction term, at the firm level, between the proportion of female employees and proportion of part time employees becomes positive in this model. As discussed previously, we need hours worked to better capture the effect of part time and full time female employees. However, we observe that wage share provides a proxy for hours worked and we observe different results from the two-level model.
- We also found a mixed result with the occupation proportion variables in the three-level model. Here, managers (Farmers and Farm Managers (12), Chief Executives, General Managers and Legislators (11), and Specialist managers (13)) are strongly associated with higher labour productivity, while Education Professionals (24), Protective Service Workers (44), Health and Welfare Support Workers (41) and Health Professionals (25) are strongly associated with lower labour productivities.
- The age coefficients are both significant and negative in the three-level model, though only one was significant in the two-level model. The coefficients for young and middle age workers are negative compared with the older employee reference group. One possible explanation for this is that the three-level model captures more employee level variation, particularly in the dependent variable, i.e. Turnover split by employee PAYG. This suggests that there is a clear interaction between income and age.[45] As we expect, older workers have higher income, likely because they are more experienced and advanced in their careers.

45 We tested the interaction between income and age and found that it was significant, but we have not included it here to avoid overcomplicating the analysis.

5.5 Three-level model results

	Fixed Intercept and Slope	*Random Intercept and Slope*	*95% Confidence Intervals*	*Bayesian Credible Intervals*
Intercept	10.77 (0.12) ***	11.26 (0.49) ***	[10.31 ; 12.22]	[8.56 ; 18.97]
Industry level:				
L.E.CapEx_mean	–0.13 (0.01) ***	–0.10 (0.05)	[–0.18 ; –0.01]	[–1.35 ; 0.49]
PT_pv_mean: Sex_pv_Female_mean	–0.22 (0.07) **	–0.19 (0.37)	[–0.91 ; 0.53]	[–13.38 ; 0.88]
Firm level:				
d_forown	–0.09 (0.02) ***	–0.23 (0.04) ***	[–0.30 ; –0.15]	[–0.31 ; –0.16]
Perm_pv	0.15 (0.01) ***	0.12 (0.03) ***	[0.06 ; 0.19]	[0.07 ; 0.19]
L.E.CapEx	0.00 (0.00) *	0.00 (0.00)	[–0.00 ; 0.01]	[0.00 ; 0.01]
L.E.OExp_exGST	–0.31 (0.02) ***	–0.38 (0.06) ***	[–0.51 ; –0.26]	[–0.49 ; –0.24]
L.E.OExp_exGST ^ 2	0.04 (0.00) ***	0.04 (0.00) ***	[0.04 ; 0.05]	[0.04 ; 0.05]
DProfLoss_R reference group Loss				
DProfLoss_R_0_to_9999	0.10 (0.01) ***	0.11 (0.02) ***	[0.06 ; 0.15]	[0.06 ; 0.15]
DProfLoss_R_10000plus	0.28 (0.01) ***	0.33 (0.02) ***	[0.29 ; 0.38]	[0.29 ; 0.38]
PT_pv:Sex_pv_Female	0.07 (0.02) **	0.17 (0.04) ***	[0.08 ; 0.26]	[0.10 ; 0.27]
Person level:				
DAge_R reference group 45+				
DAge_R_0_to_29	–0.34 (0.01) ***	–0.36 (0.01) ***	[–0.38 ; –0.34]	[–0.38 ; –0.34]
DAge_R_30_to_44	–0.15 (0.01) ***	–0.15 (0.01) ***	[–0.17 ; –0.13]	[–0.17 ; –0.14]
DIncome_R reference group 50000+				
DIncome_R_0_to_24999	–1.55 (0.01) ***	–1.61 (0.01) ***	[–1.64 ; –1.59]	[–1.63 ; –1.59]
DIncome_R_25000_to_49999	–0.53 (0.01) ***	–0.60 (0.01) ***	[–0.62 ; –0.58]	[–0.63 ; –0.58]
DOCPTN_cd reference group Unskilled				
DOCPTN_cd_11	0.48 (0.04) ***	0.49 (0.04) ***	[0.41 ; 0.57]	[0.41 ; 0.57]
DOCPTN_cd_12	0.60 (0.06) ***	0.55 (0.06) ***	[0.43 ; 0.66]	[0.44 ; 0.67]
DOCPTN_cd_13	0.32 (0.04) ***	0.30 (0.04) ***	[0.22 ; 0.38]	[0.22 ; 0.38]
. . .	. . .	. . .	. . .	. . .
DOCPTN_cd_24	–1.20 (0.07) ***	–1.08 (0.07) ***	[–1.22 ; –0.95]	[–1.21 ; –0.95]
DOCPTN_cd_25	–0.56 (0.06) ***	–0.57 (0.06) ***	[–0.68 ; –0.45]	[–0.69 ; –0.45]
. . .	. . .	. . .	. . .	. . .
DOCPTN_cd_41	–0.49 (0.10) ***	–0.57 (0.09) ***	[–0.75 ; –0.39]	[–0.73 ; –0.37]
. . .	. . .	. . .	. . .	. . .
DOCPTN_cd_44	–0.49 (0.07) ***	–0.67 (0.07) ***	[–0.81 ; –0.54]	[–0.80 ; –0.54]
. . .	. . .	. . .	. . .	. . .
DOCPTN_cd_9C	0.09 (0.06)	0.17 (0.06) **	[0.04 ; 0.29]	[0.05 ; 0.29]
AIC	294,050	285,089		
BIC	294,615	285,720		
Log Likelihood	–146,965	–142,477		

Detailed parameter estimates for the Occupation dummy variables are provided in table D.2 in Appendix D.
Significance Level: ° is 10%, * is 5%, ** is 1%, *** is 0.1%
Source: ABS unpublished prototype LEED.

6. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, we have used a prototype LEED to describe the characteristics of more productive firms. The hierarchical structure of the prototype LEED lends itself to using multilevel models to capture the dynamics between firms and employees. A LEED is a rich database that provides a great opportunity for further labour and productivity research. Even though we cannot make direct comparisons with other studies due to differences in techniques and data used, our results are broadly consistent with other findings such as Lopes and Teixeira (2012) and Earle *et al.* (2012). We would remind readers not to draw any causal conclusions from the analysis because the purpose was descriptive analysis only. This paper has demonstrated the importance of considering both firm and employee dynamics in the analysis of labour productivity. This preliminary research has many areas for potential future research:

First, we have extended the study to consider the impact of multiple job holders in the models. The three level model results are similar after we have taken these multiple job holders into account (see Appendix C). One of the reasons is that the prevalence of multiple job holders is low (less than 1%) in this prototype LEED. However, this should be considered in the model as it could become an important estimation issue in larger samples.

Second, it would also be worth expanding the prototype LEED to include key variables, such as hours worked (labour inputs), firm-level capital stock (firm investment) using perpetual inventory method, and education attainment (labour skill) rather than using proxies. However, there would be significant methodological challenges to link and expand the prototype LEED to other ABS surveys or administrative data sources.

Finally, a longitudinal LEED could be used to compare employment and earning trends across finer geographic areas, detailed industries and age groups to assist in developing regional economic or social policies. However, producing finer level statistics with existing statistical infrastructure would be challenging.

We conclude that the LEED is a powerful database with many possible analytical uses.

REFERENCES

Abowd, J.M. and Kramarz, F. (1999) "The Analysis of Labor Markets Using Matched Employer-Employee Data", Chapter 40 in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics, Volume 3*, pp. 2629–2710, Elsevier.

Australian Bureau of Statistics (2005) *Technical Manual: Australians' Employment and Unemployment Patterns, Expanded Confidentialised Unit Record File*, cat. no. 6286.0.55.002, ABS, Canberra. <http://www.abs.gov.au/ausstats/abs@.nsf/mf/6286.0.55.002>

—— (2013) *Microdata: Business Longitudinal Database, Australia, 2006–07 to 2010–11*, cat. no. 8168.0.55.001, ABS, Canberra. <http://www.abs.gov.au/ausstats/abs@.nsf/mf/8168.0.55.001>

Bachmann, R. and David, P. (2009) "The Importance of Two-Sided Heterogeneity for the Cyclicality of Labour Market Dynamics", *Ruhr Economic Papers*, No. 124, Ruhr Graduate School in Economics.

Bartelsman, E.J. and Doms, M. (2000) "Understanding Productivity: Lessons from Longitudinal Microdata", *Journal of Economic Literature*, 38, pp. 569–594.

Bickel, R. (2007) *Multilevel Analysis for Applied Research: It's Just Regression!* Guilford Press, New York.

Browne, W.J. and Draper, D. (2000) "Implementation and Performance Issues in the Bayesian and Likelihood Fitting of Multilevel Models", *Computational Statistics*, 15(3), pp. 391–420

Bryk, A.S. and Raudenbush, S.W. (1992) *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage Publications, Newbury Park, CA.

Cardoso, A.R. (2000) "Wage Differentials Across Firms: An Application of Multilevel Modelling", *Journal of Applied Econometrics*, 15(4), pp. 343–354.

Chien, C.-H.; Clarke, C. and Amarasinghe, A. (2012) "Enhancing the ABS's Use of Personal Income Tax Data", *Methodology Advisory Committee Papers*, No. 126, Unpublished, Australian Bureau of Statistics, Canberra.

Crépon, B.; Deniau, N. and Pérez-Duarte, S. (2003) *Wages, Productivity and Worker Characteristics : A French Perspective*, Centre de Recherche en Economie et Statistique, Paris.

Devine, H.; Doan, T. and Stevens, P. (2012) *Explaining Productivity Distribution in New Zealand Industries: The Effects of Input Quality on Firm Productivity Differences*, Paper presented to the New Zealand Association of Economists Annual Conference, Palmerston North, 27–29 June 2012.

Dixon, S. (2007) *The Measurement of Labour Productivity Using the Annual Enterprise Survey and LEED*, Statistics New Zealand. <http://www.stats.govt.nz/browse_for_stats/income-and-work/employment_and_unemployment/the-measure-of-labour-productivity.aspx>

Earle, J.S.; Telegdy, Á. and Antal, G. (2012) "FDI and Wages: Evidence from Firm-Level and Linked Employer-Employee Data in Hungary, 1986–2008", *Budapest Working Papers on the Labour Market*, BWP 2012/9, Hungarian Academy of Sciences.

Farmakis-Gamboni, S.; Rozenbes, D. and Yuen, K. (2012) *Award-Reliant Small Businesses*, Research Report 1/2012, Fair Work Australia. <http://www.fwc.gov.au/documents/sites/wagereview2012/research/1_2012.pdf>

Fox, J.T. And Smeets, V. (2011) "Does Input Quality Drive Measured Differences in Firm Productivity?", *NBER Working Paper Series*, No. 16853, National Bureau of Economic Research, Inc.

Gelman, A. and Hill, J. (2006) *Data Analysis Using Regression and Multilevel/ Hierarchical Models*, Cambridge University Press.

Grassetti, L.; Gori, E. and Minotti, S.C. (2005) "Multilevel Flexible Specification of the Production Function in Health Economics", *IMA Journal of Management Mathematics*, 16(4), pp. 383–398.

Gray, M.; Heath, A. and Hunter, B. (2005) "The Labour Force Dynamics of the Marginally Attached", *Australian Economic Papers*, 44(1), pp. 1–14.

Hadeld, J. (2014) *MCMCglmm Course Notes.* <http://cran.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf>

Hanchane, S. and Mostafa, T. (2010) "Endogeneity Problems in Multilevel Estimation of Education Production Functions: An Analysis Using PISA Data", *LLAKES Research Papers*, 14, Centre for Learning and Life Chances in Knowledge Economies and Societies, London.

Hildreth, A. and Pudney, S. (1999) "Econometric Issues in the Analysis of Linked Worker-Employer Surveys", in Biffignandi, S. (ed.), *Micro- and Macrodata of Firms: Statistical Analysis and International Comparison*, Physica-Verlag HD.

Hox, J.J. (2010) "The Basic Two-Level Regression Model", Chapter 2 in *Multilevel Analysis Techniques and Applications*.

Leonard, J.S.; Mulkay, B. And Audenrode, M.V. (1999) "Compensation Policies and Firm Productivity", in J.C. Haltiwanger, J.I. Lane, J.R. Spletzer, J.J.M. Theeuwes and K.R. Troske (eds.), *The Creation and Analysis of Employer-Employee Matched Data*, pp. 79–114, Emerald Group Publishing Limited.

Levinsohn, J. and Petrin, A. (2000) "Estimating Production Functions Using Inputs to Control for Unobservables", *NBER Working Paper Series*, No. 7819, National Bureau of Economic Research, Inc.

Li, J. (2013) *Introduction to Productivity Measurement Framework*, Working Paper 02–13, Department of Industry, Innovation, Science, Research and Tertiary Education Canberra.

Lopes, A.S. and Teixeira, P. (2012) "Worker Productivity and Wages: Evidence from Linked Employer-Employee Data", *IZA Discussion Papers*, No. 7036, Institute for the Study of Labor (IZA), Bonn.

Maas, C.J.M. and Hox, J.J. (2004a) "The Influence of Violations of Assumptions on Multilevel Parameter Estimates and Their Standard Errors", *Computational Statistics and Data Analysis*, 46(3), pp. 427–440.

Maas, C.J.M. and Hox, J.J. (2004b) "Robustness Issues in Multilevel Regression Analysis", *Statistica Neerlandica*, 58(2), pp. 127–137.

Mahlberg, B.; Freund, I. and Prskawetz, A. (2011) "Ageing, Productivity and Wages in Austria: Evidence From a Matched Employer-Employee Data Set at the Sector Level", *ECON Working Paper Series*, 02/2011, Institute of Mathematical Methods in Economics, Vienna University of Technology.

Marschak, J. and Andrews, W.H. (1944) "Random Simultaneous Equations and the Theory of Production", *Econometrica*, 12(3/4), pp. 143–205.

OECD (2001) *Measuring Productivity – OECD Manual*, OECD Publishing.

Olley, G.S. and Pakes, A. (1996) "The Dynamics of Productivity in the Telecomm-unications Equipment Industry", *Econometrica*, 64(6), pp. 1263–1297.

Parrotta, P.; Pozzoli, D. and Pytliková.M. (2012) "Does Labor Diversity Affect Firm Productivity?", *IZA Discussion Papers*, No. 6973, Institute for the Study of Labor (IZA), Bonn.

Rice, N.; Jones, A. and Goldstein, H. (1997) "Multilevel Models Where the Random Effects are Correlated With the Fixed Predictors: A Conditioned Iterative Generalised Least Squares Estimator (CIGLS)", *CHE Technical Paper Series 10*, Centre for Health Economics, University of York <http://www.york.ac.uk/media/che/documents/papers/technicalpapers/CHE%20Technical%20Paper%2010.pdf>

Snijders, T.A.B. and Bosker, R.J. (1999) *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, SAGE Publications Ltd, London.

Spencer, N.H. and Fielding, A. (2000) "An Instrumental Variable Consistent Estimation Procedure to Overcome the Problem of Endogenous Variables in Multilevel Models", *Multilevel Modelling Newsletter*, 12(1), pp. 4–7.

Statistics Canada (2009) *Workplace and Employee Survey (WES)*, Statistics Canada, Vancouver.

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

Syverson, C. (2010) "What Determines Productivity?", *NBER Working Paper Series*, No. 15712, National Bureau of Economic Research, Inc.

Turcotte, J. and Rennison, L.W. (2004) *Productivity and Wages: Measuring the Effect of Human Capital and Technology Use from Linked Employer-Employee Data*, Working Paper 2004-01, Department of Finance, Ottawa. <http://www.fin.gc.ca/pub/pdfs/wp2004-01e.pdf>

Van der Leeden, R.; Busing, F. and Meijer, E. (1997) *Applications of Bootstrap Methods for Two-level Models*, Paper presented at the Multilevel Conference, Amsterdam, 1–2 April 1997.

All URLs last viewed on 1 May 2015.

APPENDIXES

A. DESCRIPTIONS OF DATA SOURCES

A.1 Business Longitudinal Database

The Business Longitudinal Database (BLD) is a rolling panel, containing actively trading businesses in the Australian economy that have fewer than 200 employees, have a simple structure, and are not in certain ANZISC 06 divisions, specifically: Electricity, Gas and Water Supply (D), Finance and Insurance (K), Public Administration and Safety (O), Education and Training (P), Health Care and Social Assistance (Q) and some of Other Services (S).[46] Each year a new wave is initiated that is representative of the Australian business population at that point in time, and each wave remains in the BLD for five years. For the purposes of the LEED, all observations for a given financial year (i.e. across the four waves covering that financial year) are combined into a single cross-sectional database. Although the BLD data was not designed to be used in this manner, combining in this way allows for the largest possible sample size for a given year. However, it should be noted that this dataset is not designed to produce annual population estimates.

The size of employment of BLD firms is the main issue in creating the LEED. Not all BLD firms have employees. Linking with PAYG records indicates that a number of BLD firms are non-employing, and so are outside the scope of the LEED. Some firms also grow beyond 200 employees, which is not recorded in the BLD data but can be determined from linked PAYG records. In this case, these firms are also outside the scope of the analysis. For our analysis, we focused on the 2010–11 BLD cross-section data.

A.2 Business Activity Statement

The Business Activity Statement (BAS) is a single form used by businesses to report their taxation obligations and remit their entitlements and obligations for Goods and Services Tax (GST), Pay As You Go (PAYG), Fringe Benefits Tax (FBT), Wine Equalisation Tax (WET), and Luxury Car Tax (LCT). Depending on the business and their reporting requirements, it may be reported in monthly, quarterly or annual statements.

46 For full details on the restrictions see ABS (2013).

Within each of these reporting periods, a given ABN might appear multiple times (beyond what would be expected from that frequency – e.g. 5+ times for quarterly data etc.), which reflects multiple Client Activity Centres (CACs) for one business reporting under one ABN. As they are all the same overall business, multiple statements within each period were summed for each ABN. Some ABNs also reported over multiple periods (e.g. reporting GST monthly and wages quarterly). This means that we can potentially over-estimate the BAS reported items if we sum the values for an ABN over all periods. We resolved this by taking the overall (summed) values for each ABN at each periodicity, and taking the maximum value across the three periodicities for each variable for each ABN. Please note that these are not standard methods used by the ABS for publications.

A.3 Business Income Tax

The Business Income Tax (BIT) files contain unit record data for all businesses that have lodged income tax returns to the ATO by the date on which the files are produced. Each set of files consists of records relating to the questions of the different business tax form types: Companies, Partnerships, Trusts and Individuals. It is provided to the ABS at 12 and 18 month extracts, and these are merged together to form the ABS database. We assemble the most recent observations by taking the observations of all ABNs in the 18 month extract, and those ABNs which only appear in the 12 month extract.

A.4 Pay As You Go

The Pay As You Go (PAYG) data contains information on the wages and salaries paid by companies and businesses to their employees. This includes a scrambled Tax File Number (STFN) and ABN through which the Personal Income Tax (PIT) data can be linked to the business data. For this project, we also derive wage and salary information from the PAYG records.

Several issues needed to be addressed with the PAYG data. First is the existence of records where a person is paid zero salary from a given firm – analysis suggests these records are due to other, non-salary payments, and so we exclude them from our analysis (and in particular, do not use them in deriving employee counts for each firm). Second is when multiple PAYG payments are received by the same employee from the same firm, e.g. when a contract is renewed, a new PAYG record may be created. For the purposes of the LEED, we combine all records for the same STFN-ABN combination into a single record, and sum the salary received from each, to provide a single annual record that is then linked to the business and PIT data. Please note that these are not standard methods used by the ABS for publications.

A.5 Personal Income Tax

The Personal Income Tax (PIT) data comprises all personal income tax records from Australia for that financial year which have been submitted within sixteen months of the end of the given financial year. The file does not contain name and address information but postcodes provide an indication of the individual's address location. All useful employee-level characteristics for the LEED, except wages and salaries, are sourced from the PIT data.

The PIT data presents several conceptual issues for the LEED. Firstly, individuals with income below the tax-free threshold are not required to lodge tax returns, and so may appear in the PAYG data but not in the PIT data. This analysis only included individuals for which the ABS had both PIT and PAYG data. A second conceptual issue is that the ATO may edit an individual's tax return for taxation compliance purposes, but may not edit related fields (e.g. not editing subtotals to match an edited grand total).

B. VARIABLE DESCRIPTIONS AND DERIVATIONS

B.1 Variable descriptions and derivations

Variable name	*Derivation formulas*	*Descriptions*
d_forown		Dummy for if a firm is foreign owned.
L.E.Turnover_exGST	$= \ln\left(\frac{\text{Turnover}_{kj}}{\text{Emp}_{kj}}\right)$	Turnover_{kj} is the firm Turnover excluding goods and services tax (GST) (please note that not every firm attracts the same amount of GST in the Turnover across different industries, so we choose this measure, which ensures consistency for across industries comparison). Emp_{kj} is the number of employees, for firm j in industry k. That is, our labour productivity measure at the firm level is the natural log of per-employee turnover.
L.E.CapEx[47]	$= \ln\left(\frac{\text{CapEx}_{kj}}{\text{Emp}_{kj}}\right)$	CapEx_{kj} is the firm Capital Expenditure, and Emp_{kj} is the number of employees, for firm j in industry k.
L.E.OExp_exGST and L.E.OExp_exGST^2	$= \ln\left(\frac{\text{OExp}_{kj}}{\text{Emp}_{kj}}\right)$	OExp_{kj} is the firm Operating Expenses (excluding GST), and Emp_{kj} is the number of employees, for firm j in industry k. L.E.OExp_exGST^2 is the square of the logged per-employee Operating Expenses.
DProfLoss_R_X (X=0_to_9999, 10000plus)	$= \begin{cases} 1 & \text{if } \frac{\text{ProfLoss}_{kj}}{\text{Emp}_{kj}} \text{ is in range X} \\ 0 & \text{otherwise} \end{cases}$	ProfLoss_{kj} is the firm Profit/Loss, and Emp_{kj} is the number of employees, for firm j in industry k. We then divide the per-employee Profit/Loss into three groups of roughly equal size: negative, \$0–\$9999, \$10000+. The reference is group is firms with negative per-employee Profit/Loss.
DSex_Female	$= \begin{cases} 1 & \text{if female} \\ 0 & \text{otherwise} \end{cases}$	Dummy variable indicates sex. The reference group is males.
DAge_R_X (X=0_to_29, 30_to_44)	$= \begin{cases} 1 & \text{if Age is in range X} \\ 0 & \text{otherwise} \end{cases}$	We divide employee age into three groups of roughly equal size: 0–29, 30–44, 45+ (note that there are no observations with age below 7). The reference group is people aged 45+
DIncome_R_X (X=0_to_24999, 25000_to_49999)	$= \begin{cases} 1 & \text{if PAYG income is in range X} \\ 0 & \text{otherwise} \end{cases}$	We divide total employee PAYG income (across all firms) into three groups of roughly equal size: \$0–\$24999, \$25000–\$49999, \$50000+. The reference group is people with PAYG income of \$50000+
DOCPTN_cd_X (X=11,...,9C)	$= \begin{cases} 1 & \text{if Occupation Code is X} \\ 0 & \text{otherwise} \end{cases}$	For occupation code, we divide employees into 44 groups based on the first two digits of their occupation code, The reference group is group 9A – Apprentices and Trainees.
Perm_pv		Proportion of permanent employees.
PT_pv		Proportion of part time employees.

47 We also derived Capital Stock for a given firm using this formula: $K_t = A_{t-1} - D_t + C_t$, where K_t is the derived capital stock measure at time t, A_{t-1} are non-current assets at time $t-1$, D_t is depreciation at time t, and C_t is capital expenditure at time t (Olley and Pakes, 1996). However, this variable was not significant in some models.

B.1 Variable descriptions and derivations – continued

Variable name	*Derivation formulas*	*Descriptions*
Sex_pv_Female, Age_pv_X (X=0_to_29, 30_to_44), Income_R_pv_X (X=0_to_24999, 25000_to_49999), OCPTN_cd_pv_X (X=11,...,9C)	$= \frac{\sum_i \mathrm{DSex\_Female}_{kji}}{\mathrm{Emp}_{kj}}$, etc.	For our four employee variables (Sex, Age, Income and Occupation Code) we use the groupings defined above and calculate the proportion of employees in each group for each firm. For example, a firm may have 40% of its workforce female and 60% male, so would have the value of 0.4 for the proportion of females and 0.6 for the proportion of males. We then use these proportions in our firm-level models, in each case excluding one group for each variable so as to avoid the proportional equivalent of the dummy variable trap.
LPers_Turnover_exGST	$= \ln\left(\mathrm{Turnover}_{kj} \times \frac{\mathrm{PAYG}_i}{\sum_{i=1}^{I} \mathrm{PAYG}_i}\right)$	$\mathrm{Turnover}_{kj}$ is the firm Turnover (excluding GST) and PAYG_i is the employee PAYG wage, for employee i in firm j in industry k. That is, our labour productivity measure at the employee level is the natural log of the employee's share of firm turnover (shared out by wage share). We thus assume that an employee's contribution to firm turnover is proportional to their PAYG wage from that firm.

B.2 Industry codes

Code	*Industry*
A	Agriculture, Forestry and Fishing
B	Mining
C	Manufacturing
D	Electricity, Gas, Water and Waste Services*
E	Construction
F	Wholesale Trade
G	Retail Trade
H	Accommodation and Food Services
I	Transport, Postal and Warehousing
J	Information Media and Telecommunications
K	Financial and Insurance Services*
L	Rental, Hiring and Real Estate Services
M	Professional, Scientific and Technical Services
N	Administrative and Support Services
O	Public Administration and Safety*
P	Education and Training*
Q	Health Care and Social Assistance*
R	Arts and Recreation Services
S	Other Services

*Industries excluded from analysis

B.3 Occupation codes

Descriptions
For occupation code, we divide employees into 44 groups based on the first two digits of their occupation code, which corresponds to: 9A – Apprentices and Trainees 11 – Chief Executives, General Managers and Legislators 12 – Farmers and Farm Managers 13 – Specialist Managers 14 – Hospitality, Retail and Service Managers 21 – Arts and Media Professionals 22 – Business, Human Resource & Marketing Professionals 23 – Design, Engineering, Science & Transport Professionals 24 – Education Professionals 25 – Health Professionals 26 – ICT Professionals 27 – Legal, Social and Welfare Professionals 31 – Engineering, ICT and Science Technicians 32 – Automotive and Engineering Trades Workers 33 – Construction Trades Workers 34 – Electrotechnology & Telecommunications Trades Workers 35 – Food Trades Workers 36 – Skilled Animal and Horticultural Workers 39 – Other Technicians and Trades Workers 41 – Health and Welfare Support Workers 42 – Carers and Aides 43 – Hospitality Workers 44 – Protective Service Workers 45 – Sports and Personal Service Workers 51 – Office Managers and Program Administrators 52 – Personal Assistants and Secretaries 53 – General Clerical Workers 54 – Inquiry Clerks and Receptionists 55 – Numerical Clerks 56 – Clerical and Office Support Workers 59 – Other Clerical and Administrative Workers 61 – Sales Representatives and Agents 62 – Sales Assistants and Salespersons 63 – Sales Support Workers 71 – Machine and Stationary Plant Operators 72 – Mobile Plant Operators 73 – Road and Rail Drivers 74 – Store persons 81 – Cleaners and Laundry Workers 82 – Construction and Mining Labourers 83 – Factory Process Workers 84 – Farm, Forestry and Garden Workers 85 – Food Preparation Assistants 89 – Other Labourers 9C – Consultants

C. THREE-LEVEL (EMPLOYEE-FIRM-INDUSTRY) MODEL WITH MULTIPLE JOB HOLDERS

The person-level model (for $i =$ employees, $j =$ firms and $k =$ industries) is specified as:

$$Y_{k\{j\}i} = \alpha_{0k\{j\}} + \sum_{p=1}^{P} \alpha_{pk\{j\}} X_{pk\{j\}i} + e_{k\{j\}i} \tag{1}$$

- $\{j\}$ means the set of firms hiring person i (which may be a single firm, or multiple firms in the case of multiple job holders).
- $Y_{k\{j\}i}$ represents the contribution to firm j's output of employee i, who may work for a set of firms $\{j\}$. This is the log of total person-level Turnover (across all their jobs) derived by

  $$\ln\left(\sum\nolimits_{h\in\{j\}} \text{Turnover}_h \times \text{WPAYG}_{hi}\right)^{48}$$

  for all employees who receive Pay-As-You-Go (PAYG) payment from a set of firms $\{j\}$ in industry k.[49] We derive employee-level Turnover by dividing a firm's Turnover between its employees according to their wage share within that firm (making the simplifying assumption that a person's contribution to firm production is proportional to their wage received from that firm). The multiple membership weights are also derived using the employee-level Turnover, as the share of employee-level Turnover contributed by each firm[50];
- $\left\{X_{pj\{j\}i} : p = 1, \ldots, P\right\}$ are the person-level characteristics, e.g. age, sex and occupation etc.;[51]
- $\alpha_{0k\{j\}}$ is the intercept for a set of firms $\{j\}$ in industry k;
- $\left\{\alpha_{pk\{j\}} : p = 1, \ldots, P\right\}$ are the corresponding employee level coefficients that indicate the direction and strength of association between each employee characteristic and employee-level Turnover;

---

48 Note that $\text{WPAYG}_i = \dfrac{\text{PAYG}_{hi}}{\sum_{i=1}^{I} \text{PAYG}_{hi}}$ and $\sum_{i=1}^{I} \text{WPAYG}_{hi} = 1$ for a given firm h.

49 Note that different firms within $\{j\}$ might be in different industries; for simplicity of notation we leave k outside the {}– its affect is accounted for in the combined second and third level models below.

50 That is, $w_{hi} = \dfrac{\text{Turnover}_h \times \text{WPAYG}_{hi}}{\sum_{h\in\{j\}} \text{Turnover}_h \times \text{WPAYG}_{hi}}$ for a given person i and firm h.

51 The income or earning variables are excluded in the explanatory variables to avoid the problem of endogeneity.

- $e_{k\{j\}i}$ is the model error term that represents the deviation of employee i's contribution from the mean output $\{j\}$ in industry k from the predicted outputs after adjusting for the employee predictor, $\left\{X_{pk\{j\}} : p = 1, \ldots, P\right\}$.

The second- or firm level model describes the productivity differences explained by firm-level variables, while the third- or industry-level model describes the productivity difference across industries and summarises the similarities and differences between firms effectively. Both account for multiple job holders working in multiple firms. The combined firm-level and industry-level models can be expressed as:[52]

$$\alpha_{0k\{j\}} = \delta_{000} + \sum_{h \in \{j\}} w_{hi} \left(\sum_{s=1}^{S} \zeta_{s00} V_{sk} + \sum_{q=1}^{Q} \beta_{qk0} Z_{qkh} + u_{0k0} + v_{0kh} \right) \quad (2)$$

$$\alpha_{pk\{j\}} = \delta_{p00} \quad (3)$$

$$\beta_{qk0} = \gamma_{q00} + u_{qk0} \quad (4)$$

- δ_{000} is the overall intercept across all industries.
- w_{hi} represents employee i's weight associated with a particular firm $h \in \{j\}$. This means that for multiple job holders, the contributions to their labour productivity from firm-level explanatory characteristics are weighted for each firm they worked for. The weights associated with the set of firm level units $\{j\}$ in equation (2) add to one. Note that this weight is different from WPAYG_{hi}, which is a within-firm weight. As different firms within $\{j\}$ can be in different industries k, the weights are applied to both the firm and industry variables.
- δ_{p00} is the average regression slope across different industries for person characteristic p.
- γ_{q00} is the average regression slope across different industries for firm characteristic q.
- $\left\{V_{sk} : s = 1, \ldots, S\right\}$ are the S industry-level explanatory variables, each of which is formed as the industry mean of the firm-level variable for industry k.
- ζ_{s00} are the corresponding industry-level coefficients that indicate the direction and strength of association between each industry characteristic s and the employee-level Turnover.

52 See Snijders and Bosker (1999) and Bryk and Raudenbush (1992).

- $\left\{\beta_{qk0} : q = 1, \dots, Q\right\}$ are the corresponding firm level coefficients that indicate the direction and strength of association between each firm characteristic and employee-level Turnover. The relationship between the firm characteristics and outputs does not vary at the firm level because the multiple membership structure only exists at the person level.
- $\left\{Z_{qkh} : q = 1, \dots, Q\right\}$ are the Q firm-level explanatory variables such as investments, operating expenses and foreign ownership dummy for firm $h \in \{j\}$ in industry k.
- u_{0k0} is the industry dependent deviation from the total industry intercept.
- v_{0kh} is the firm dependent deviation, for a particular firm $h \in \{j\}$, from the total firm intercept.
- The slopes β_{qk0} are industry dependent and can be split into an overall average β_{q00} and an industry dependent deviation of slope u_{qk0}, i.e. allowing the slope of the firm operating expenses to vary by industry (but not by firm).[53] Note that we allow for each firm within each industry to have a different intercept, but for the random slope we only allow this to differ by industry to ensure the consistency with the two-level model. Again u_{qk0} is zero for all variables except firm operating expenses.

Substituting $\alpha_{0k\{j\}}$, $\alpha_{pk\{j\}}$ and β_{qk0} yields

$$Y_{k\{j\}i} = \delta_{000} + \sum_{h \in \{j\}} w_{hi} \left(\sum_{s=1}^{S} \zeta_{s00} V_{sk} + \sum_{q=1}^{Q} \gamma_{q00} Z_{qkh} + u_{0k0} + \sum_{q=1}^{Q} u_{qk0} Z_{qkh} + v_{0kh} \right) + \sum_{p=1}^{P} \delta_{p00} X_{pk\{j\}i} + e_{k\{j\}i}$$

The prototype LEED data contains multiple job holders. These are employees who work for multiple firms at the same time. We used a mixed membership model to distinguish their contributions to these firms' productivity. Table C.1 shows the estimation results and they are consistent with the three level model. This is because there is a low prevalence of multiple job holders in the prototype data. The multiple job holders account for less than 1% in the sample. This will become a significant estimation issue for a larger sample.

53 This is to ensure consistency of the model specification between two- and three-level models.

C.1 Three-level model with multiple job holders

	Unweighted for MJH	*Weighted for MJH*	*Bayesian Credible Intervals*
Intercept	12.82 ***	11.94 ***	[7.00 ; 16.35]
Industry level:			
L.E.CapEx_mean	–0.27	–0.03	[–0.74 ; 0.88]
PT_pv_mean: Sex_pv_Female_mean	–4.81	–3.02	[–9.52 ; 3.59]
Firm level:			
d_forown	–0.25 ***	–0.25 ***	[–0.33 ; –0.19]
Perm_pv	0.11 **	0.12 **	[0.06 ; 0.18]
L.E.CapEx	0.00	0.00	[0.00 ; 0.01]
L.E.OExp_exGST	–0.38 ***	–0.41 ***	[–0.53 ; –0.30]
L.E.OExp_exGST ^ 2	0.04 ***	0.04 ***	[0.04 ; 0.05]
DProfLoss_R reference group Loss			
DProfLoss_R_0_to_9999	0.11 ***	0.10 ***	[0.06 ; 0.14]
DProfLoss_R_10000plus	0.34 ***	0.33 ***	[0.28 ; 0.37]
PT_pv:Sex_pv_Female	0.20 ***	0.20 ***	[0.12 ; 0.29]
Person level:			
DAge_R reference group 45+			
DAge_R_0_to_29	–0.35 ***	–0.35 ***	[–0.37 ; –0.33]
DAge_R_30_to_44	–0.15 ***	–0.15 ***	[–0.17 ; –0.13]
DIncome_R reference group 50000+			
DIncome_R_0_to_24999	–1.62 ***	–1.62 ***	[–1.64 ; –1.60]
DIncome_R_25000_to_49999	–0.61 ***	–0.61 ***	[–0.63 ; –0.59]
DOCPTN_cd reference group Unskilled			
DOCPTN_cd_11	0.49 ***	0.49 ***	[0.41 ; 0.57]
DOCPTN_cd_12	0.54 ***	0.55 ***	[0.44 ; 0.66]
DOCPTN_cd_13	0.30 ***	0.30 ***	[0.22 ; 0.37]
. . .	. . .	. . .	. . .
DOCPTN_cd_89	–0.03	–0.03	[–0.37 ; –0.33]
DOCPTN_cd_9C	0.16 **	0.16 **	[–0.17 ; –0.13]
DIC	281,426.1	281,024.2	

Detailed parameter estimates for the Occupation dummy variables are provided in table D.3 in Appendix D.
Significance Level: ° is 10%, * is 5%, ** is 1%, *** is 0.1%
Source: ABS unpublished prototype LEED.

D. PARAMETER ESTIMATES – OCCUPATION DETAILS

D.1 Two-level model results – complete occupation parameter estimates (see table 5.4)

	Fixed Intercept and Slope	*Random Intercept and Slope*	*95% Confidence Intervals*	*Bayesian Credible Intervals*
OCPTN_cd_pv_11	0.18 (0.08) *	0.15 (0.08) °	[–0.00 ; 0.31]	[–0.01 ; 0.30]
OCPTN_cd_pv_12	0.35 (0.11) ***	0.23 (0.11) *	[0.03 ; 0.44]	[0.05 ; 0.47]
OCPTN_cd_pv_13	0.35 (0.09) ***	0.33 (0.09) ***	[0.16 ; 0.50]	[0.14 ; 0.49]
OCPTN_cd_pv_14	0.18 (0.11) °	0.14 (0.10)	[–0.06 ; 0.35]	[–0.09 ; 0.32]
OCPTN_cd_pv_21	0.13 (0.09)	0.10 (0.09)	[–0.07 ; 0.28]	[–0.09 ; 0.27]
OCPTN_cd_pv_22	0.04 (0.09)	–0.03 (0.09)	[–0.21 ; 0.15]	[–0.22 ; 0.14]
OCPTN_cd_pv_23	0.30 (0.09) ***	0.22 (0.09) **	[0.05 ; 0.39]	[0.05 ; 0.41]
OCPTN_cd_pv_24	–0.33 (0.21)	–0.32 (0.21)	[–0.73 ; 0.09]	[–0.76 ; 0.06]
OCPTN_cd_pv_25	0.24 (0.13)	0.16 (0.12)	[–0.09 ; 0.40]	[–0.08 ; 0.43]
OCPTN_cd_pv_26	0.47 (0.11) ***	0.35 (0.11) **	[0.13 ; 0.57]	[0.11 ; 0.55]
OCPTN_cd_pv_27	0.13 (0.17)	0.10 (0.17)	[–0.23 ; 0.43]	[–0.21 ; 0.47]
OCPTN_cd_pv_31	0.32 (0.14) *	0.23 (0.13) °	[–0.03 ; 0.50]	[–0.03 ; 0.49]
OCPTN_cd_pv_32	0.06 (0.08)	0.05 (0.08)	[–0.11 ; 0.21]	[–0.13 ; 0.20]
OCPTN_cd_pv_33	0.10 (0.09)	0.09 (0.09)	[–0.08 ; 0.27]	[–0.07 ; 0.29]
OCPTN_cd_pv_34	0.11 (0.11)	0.11 (0.10)	[–0.10 ; 0.31]	[–0.08 ; 0.33]
OCPTN_cd_pv_35	0.16 (0.11)	0.16 (0.11)	[–0.05 ; 0.37]	[–0.03 ; 0.38]
OCPTN_cd_pv_36	–0.11 (0.09)	–0.07 (0.08)	[–0.23 ; 0.10]	[–0.23 ; 0.12]
OCPTN_cd_pv_39	0.10 (0.09)	0.13 (0.09)	[–0.04 ; 0.30]	[–0.04 ; 0.31]
OCPTN_cd_pv_41	0.14 (0.31)	0.17 (0.30)	[–0.42 ; 0.76]	[–0.47 ; 0.71]
OCPTN_cd_pv_42	0.05 (0.13)	–0.03 (0.13)	[–0.28 ; 0.23]	[–0.28 ; 0.22]
OCPTN_cd_pv_43	0.18 (0.10)	0.17 (0.10) °	[–0.02 ; 0.37]	[–0.03 ; 0.36]
OCPTN_cd_pv_44	0.07 (0.20)	–0.02 (0.20)	[–0.42 ; 0.37]	[–0.39 ; 0.39]
OCPTN_cd_pv_45	0.06 (0.09)	0.09 (0.09)	[–0.09 ; 0.27]	[–0.10 ; 0.25]
OCPTN_cd_pv_51	0.21 (0.09) *	0.19 (0.09) *	[0.01 ; 0.36]	[–0.01 ; 0.35]
OCPTN_cd_pv_52	0.35 (0.12) **	0.31 (0.12) **	[0.07 ; 0.54]	[0.09 ; 0.58]
OCPTN_cd_pv_53	0.26 (0.09) **	0.23 (0.09) *	[0.05 ; 0.42]	[0.04 ; 0.40]
OCPTN_cd_pv_54	0.04 (0.15)	0.00 (0.15)	[–0.29 ; 0.29]	[–0.34 ; 0.25]
OCPTN_cd_pv_55	–0.07 (0.11)	–0.11 (0.11)	[–0.32 ; 0.11]	[–0.32 ; 0.10]
OCPTN_cd_pv_56	0.16 (0.14)	0.14 (0.13)	[–0.12 ; 0.40]	[–0.15 ; 0.37]
OCPTN_cd_pv_59	0.20 (0.11)	0.13 (0.11)	[–0.08 ; 0.34]	[–0.12 ; 0.31]
OCPTN_cd_pv_61	0.31 (0.08) ***	0.22 (0.09) *	[0.05 ; 0.38]	[–0.03 ; 0.32]
OCPTN_cd_pv_62	0.30 (0.08) ***	0.27 (0.08) **	[0.10 ; 0.43]	[0.07 ; 0.43]
OCPTN_cd_pv_63	0.07 (0.17)	–0.06 (0.17)	[–0.39 ; 0.27]	[–0.42 ; 0.21]
OCPTN_cd_pv_71	0.30 (0.12) **	0.32 (0.11) **	[0.10 ; 0.55]	[0.12 ; 0.58]
OCPTN_cd_pv_72	0.29 (0.11) **	0.30 (0.10) **	[0.10 ; 0.50]	[0.13 ; 0.53]
OCPTN_cd_pv_73	0.23 (0.08) **	0.18 (0.08) *	[0.03 ; 0.34]	[0.01 ; 0.34]
OCPTN_cd_pv_74	0.23 (0.16)	0.13 (0.16)	[–0.18 ; 0.44]	[–0.23 ; 0.40]
OCPTN_cd_pv_81	0.01 (0.09)	–0.03 (0.09)	[–0.22 ; 0.15]	[–0.20 ; 0.17]
OCPTN_cd_pv_82	0.16 (0.10) °	0.14 (0.10)	[–0.05 ; 0.33]	[–0.01 ; 0.37]
OCPTN_cd_pv_83	0.20 (0.10) *	0.17 (0.10) °	[–0.02 ; 0.35]	[–0.05 ; 0.33]
OCPTN_cd_pv_84	0.21 (0.08) **	0.17 (0.08) *	[0.01 ; 0.32]	[0.03 ; 0.34]
OCPTN_cd_pv_85	0.20 (0.13)	0.22 (0.12) °	[–0.02 ; 0.46]	[–0.03 ; 0.48]
OCPTN_cd_pv_89	0.16 (0.10)	0.11 (0.10)	[–0.08 ; 0.30]	[–0.07 ; 0.31]
OCPTN_cd_pv_9C	0.45 (0.12) ***	0.33 (0.12) **	[0.10 ; 0.57]	[0.09 ; 0.56]

Significance Level: ° is 10%, * is 5%, ** is 1%, *** is 0.1%

Source: ABS unpublished prototype LEED.

D.2 Three-level model results – complete occupation parameter estimates (see table 5.5)

	Fixed Intercept and Slope		*Random Intercept and Slope*		*95% Confidence Intervals*	*Bayesian Credible Intervals*
DOCPTN_cd_11	0.48	(0.04) ***	0.49	(0.04) ***	[0.41 ; 0.57]	[0.41 ; 0.57]
DOCPTN_cd_12	0.60	(0.06) ***	0.55	(0.06) ***	[0.43 ; 0.66]	[0.44 ; 0.67]
DOCPTN_cd_13	0.32	(0.04) ***	0.30	(0.04) ***	[0.22 ; 0.38]	[0.22 ; 0.38]
DOCPTN_cd_14	0.27	(0.04) ***	0.21	(0.04) ***	[0.12 ; 0.29]	[0.13 ; 0.28]
DOCPTN_cd_21	–0.55	(0.05) ***	–0.02	(0.05)	[–0.12 ; 0.09]	[–0.11 ; 0.09]
DOCPTN_cd_22	0.02	(0.04)	0.09	(0.04) *	[0.01 ; 0.18]	[0.00 ; 0.17]
DOCPTN_cd_23	0.14	(0.04) **	0.16	(0.04) ***	[0.08 ; 0.24]	[0.07 ; 0.24]
DOCPTN_cd_24	–1.20	(0.07) ***	–1.08	(0.07) ***	[–1.22 ; –0.95]	[–1.21 ; –0.95]
DOCPTN_cd_25	–0.56	(0.06) ***	–0.57	(0.06) ***	[–0.68 ; –0.45]	[–0.69 ; –0.45]
DOCPTN_cd_26	0.14	(0.05) *	0.15	(0.05) **	[0.05 ; 0.26]	[0.07 ; 0.27]
DOCPTN_cd_27	0.00	(0.07)	–0.06	(0.07)	[–0.20 ; 0.09]	[–0.20 ; 0.08]
DOCPTN_cd_31	0.07	(0.05)	0.08	(0.05) °	[–0.01 ; 0.17]	[–0.01 ; 0.16]
DOCPTN_cd_32	–0.03	(0.04)	0.01	(0.04)	[–0.07 ; 0.09]	[–0.06 ; 0.09]
DOCPTN_cd_33	0.02	(0.05)	0.02	(0.05)	[–0.07 ; 0.12]	[–0.07 ; 0.12]
DOCPTN_cd_34	0.03	(0.05)	0.05	(0.05)	[–0.04 ; 0.15]	[–0.05 ; 0.14]
DOCPTN_cd_35	0.31	(0.05) ***	0.22	(0.05) ***	[0.13 ; 0.31]	[0.13 ; 0.31]
DOCPTN_cd_36	–0.10	(0.05) *	0.01	(0.05)	[–0.08 ; 0.11]	[–0.07 ; 0.11]
DOCPTN_cd_39	–0.04	(0.05)	0.04	(0.05)	[–0.05 ; 0.13]	[–0.04 ; 0.13]
DOCPTN_cd_41	–0.49	(0.10) ***	–0.57	(0.09) ***	[–0.75 ; –0.39]	[–0.73 ; –0.37]
DOCPTN_cd_42	–0.40	(0.05) ***	–0.51	(0.05) ***	[–0.62 ; –0.40]	[–0.61 ; –0.40]
DOCPTN_cd_43	0.16	(0.04) ***	0.08	(0.04) °	[–0.00 ; 0.16]	[–0.01 ; 0.15]
DOCPTN_cd_44	–0.49	(0.07) ***	–0.67	(0.07) ***	[–0.81 ; –0.54]	[–0.80 ; –0.54]
DOCPTN_cd_45	0.00	(0.05)	0.10	(0.05) *	[0.01 ; 0.20]	[0.00 ; 0.20]
DOCPTN_cd_51	0.18	(0.04) ***	0.16	(0.04) ***	[0.08 ; 0.24]	[0.08 ; 0.24]
DOCPTN_cd_52	0.01	(0.05)	0.00	(0.05)	[–0.11 ; 0.10]	[–0.10 ; 0.10]
DOCPTN_cd_53	–0.06	(0.04)	–0.04	(0.04)	[–0.12 ; 0.04]	[–0.12 ; 0.04]
DOCPTN_cd_54	–0.02	(0.05)	–0.01	(0.05)	[–0.10 ; 0.09]	[–0.10 ; 0.09]
DOCPTN_cd_55	–0.02	(0.05)	–0.03	(0.05)	[–0.12 ; 0.06]	[–0.12 ; 0.05]
DOCPTN_cd_56	–0.13	(0.06) *	–0.13	(0.06) *	[–0.25 ; –0.00]	[–0.26 ; –0.02]
DOCPTN_cd_59	0.08	(0.05) °	0.05	(0.04)	[–0.04 ; 0.14]	[–0.03 ; 0.14]
DOCPTN_cd_61	0.20	(0.04) ***	0.15	(0.04) ***	[0.07 ; 0.24]	[0.07 ; 0.24]
DOCPTN_cd_62	0.07	(0.04) °	–0.04	(0.04)	[–0.12 ; 0.04]	[–0.11 ; 0.04]
DOCPTN_cd_63	0.06	(0.05)	–0.04	(0.05)	[–0.13 ; 0.05]	[–0.13 ; 0.05]
DOCPTN_cd_71	0.12	(0.04) **	0.06	(0.04)	[–0.03 ; 0.14]	[–0.04 ; 0.13]
DOCPTN_cd_72	–0.01	(0.05)	0.01	(0.05)	[–0.08 ; 0.10]	[–0.07 ; 0.09]
DOCPTN_cd_73	–0.02	(0.04)	0.02	(0.04)	[–0.06 ; 0.10]	[–0.06 ; 0.09]
DOCPTN_cd_74	0.02	(0.05)	–0.02	(0.05)	[–0.11 ; 0.07]	[–0.11 ; 0.07]
DOCPTN_cd_81	0.12	(0.05) **	0.08	(0.05) °	[–0.01 ; 0.17]	[–0.01 ; 0.17]
DOCPTN_cd_82	–0.03	(0.04)	–0.07	(0.04)	[–0.15 ; 0.02]	[–0.15 ; 0.01]
DOCPTN_cd_83	0.15	(0.04) ***	0.05	(0.04)	[–0.03 ; 0.13]	[–0.03 ; 0.13]
DOCPTN_cd_84	0.19	(0.04) ***	0.23	(0.04) ***	[0.14 ; 0.31]	[0.15 ; 0.30]
DOCPTN_cd_85	0.29	(0.05) ***	0.12	(0.05) *	[0.02 ; 0.22]	[0.01 ; 0.20]
DOCPTN_cd_89	–0.07	(0.04)	–0.04	(0.04)	[–0.13 ; 0.04]	[–0.12 ; 0.04]
DOCPTN_cd_9C	0.09	(0.06)	0.17	(0.06) **	[0.04 ; 0.29]	[0.05 ; 0.29]

Source: ABS unpublished prototype LEED.

Significance Level: ° is 10%, * is 5%, ** is 1%, *** is 0.1%

D.3 Three-level model with multiple job holders – complete occupation parameter estimates (see table C.1)

	Unweighted for MJH	*Weighted for MJH*	*Bayesian Credible Intervals*
DOCPTN_cd_11	0.49 ***	0.49 ***	[0.41 ; 0.57]
DOCPTN_cd_12	0.54 ***	0.55 ***	[0.44 ; 0.66]
DOCPTN_cd_13	0.30 ***	0.30 ***	[0.22 ; 0.37]
DOCPTN_cd_14	0.21 ***	0.21 ***	[0.13 ; 0.29]
DOCPTN_cd_21	–0.01	–0.01	[–0.12 ; 0.08]
DOCPTN_cd_22	0.09 *	0.09 *	[0.01 ; 0.17]
DOCPTN_cd_23	0.16 ***	0.16 ***	[0.08 ; 0.24]
DOCPTN_cd_24	–1.10 ***	–1.09 ***	[–1.22 ; –0.95]
DOCPTN_cd_25	–0.57 ***	–0.57 ***	[–0.69 ; –0.47]
DOCPTN_cd_26	0.17 **	0.17 **	[0.06 ; 0.27]
DOCPTN_cd_27	–0.07	–0.06	[–0.20 ; 0.08]
DOCPTN_cd_31	0.08 °	0.08 °	[–0.01 ; 0.16]
DOCPTN_cd_32	0.01	0.01	[–0.07 ; 0.09]
DOCPTN_cd_33	0.02	0.02	[–0.08 ; 0.11]
DOCPTN_cd_34	0.05	0.06	[–0.03 ; 0.15]
DOCPTN_cd_35	0.22 ***	0.23 ***	[0.14 ; 0.30]
DOCPTN_cd_36	0.06	0.04	[–0.05 ; 0.13]
DOCPTN_cd_39	0.08	0.06	[–0.02 ; 0.16]
DOCPTN_cd_41	–0.59 ***	–0.58 ***	[–0.75 ; –0.39]
DOCPTN_cd_42	–0.54 ***	–0.53 ***	[–0.63 ; –0.42]
DOCPTN_cd_43	0.08 *	0.08 *	[0.00 ; 0.16]
DOCPTN_cd_44	–0.67 ***	–0.67 ***	[–0.79 ; –0.54]
DOCPTN_cd_45	0.11 *	0.11 *	[0.02 ; 0.21]
DOCPTN_cd_51	0.16 ***	0.16 ***	[0.08 ; 0.24]
DOCPTN_cd_52	0.00	0.00	[–0.09 ; 0.11]
DOCPTN_cd_53	–0.03	–0.03	[–0.10 ; 0.05]
DOCPTN_cd_54	–0.01	–0.01	[–0.10 ; 0.08]
DOCPTN_cd_55	–0.03	–0.03	[–0.10 ; 0.07]
DOCPTN_cd_56	–0.11 °	–0.11 °	[–0.22 ; 0.01]
DOCPTN_cd_59	0.06	0.06	[–0.03 ; 0.13]
DOCPTN_cd_61	0.15 **	0.15 **	[0.08 ; 0.24]
DOCPTN_cd_62	–0.04	–0.04	[–0.11 ; 0.04]
DOCPTN_cd_63	–0.04	–0.04	[–0.14 ; 0.05]
DOCPTN_cd_71	0.06	0.06	[–0.03 ; 0.14]
DOCPTN_cd_72	0.02	0.03	[–0.07 ; 0.11]
DOCPTN_cd_73	0.03	0.03	[–0.05 ; 0.10]
DOCPTN_cd_74	–0.01	–0.01	[–0.10 ; 0.07]
DOCPTN_cd_81	0.09 °	0.09 °	[0.010 ; 0.18]
DOCPTN_cd_82	–0.05	–0.05	[–0.14 ; 0.02]
DOCPTN_cd_83	0.05	0.05	[–0.03 ; 0.13]
DOCPTN_cd_84	0.22 ***	0.23 ***	[0.15 ; 0.31]
DOCPTN_cd_85	0.12 *	0.12 *	[0.03 ; 0.21]
DOCPTN_cd_89	–0.03	–0.03	[–0.11 ; 0.06]
DOCPTN_cd_9C	0.16 **	0.16 **	[0.05 ; 0.29]

Source: ABS unpublished prototype LEED.

Significance Level: ° is 10%, * is 5%, ** is 1%, *** is 0.1%

E. DECOMPOSITION OF LABOUR CHARACTERISTICS

Further mathematical detail for the derivation of equation (1) in Section 4.

We use a single labour characteristic e.g. age for exposition:

$$
\begin{aligned}
L_j^* &= \sum_{p=0}^{P} \lambda_{pj} X_{pj} \\
&= \lambda_{0j} X_{0j} + \sum_{p=1}^{P} \lambda_{pj} X_{pj} \\
&= \lambda_{0j} X_j \left[\frac{\lambda_{0j} X_{0j}}{\lambda_{0j} X_j} + \frac{\sum_{p=1}^{P} \lambda_{pj} X_{pj}}{\lambda_{0j} X_j} \right] \\
&= \lambda_{0j} X_j \left[\frac{\lambda_{0j} X_{0j} + \sum_{p=1}^{P} \lambda_{0j} X_{pj}}{\lambda_{0j} X_j} + \frac{\sum_{p=1}^{P} \lambda_{pj} X_{pj} - \sum_{p=1}^{P} \lambda_{0j} X_{pj}}{\lambda_{0j} X_j} \right] \\
&= \lambda_{0j} X_j \left[\frac{\lambda_{0j} X_{0j} + \sum_{p=1}^{P} \lambda_{0j} X_{pj}}{\lambda_{0j} X_j} + \frac{\sum_{p=1}^{P} \left(\lambda_{pj} - \lambda_{0j} \right) X_{pj}}{\lambda_{0j} X_j} \right] \\
&= \lambda_{0j} X_j \left[1 + \sum_{p=1}^{P} \left(\frac{\lambda_{pj}}{\lambda_{0j}} - 1 \right) \frac{X_{pj}}{X_j} \right]
\end{aligned}
$$

FOR MORE INFORMATION . . .

INTERNET **www.abs.gov.au** The ABS website is the best place for data from our publications and information about the ABS.

LIBRARY A range of ABS publications are available from public and tertiary libraries Australia wide. Contact your nearest library to determine whether it has the ABS statistics you require, or visit our website for a list of libraries.

INFORMATION AND REFERRAL SERVICE

Our consultants can help you access the full range of information published by the ABS that is available free of charge from our website, or purchase a hard copy publication. Information tailored to your needs can also be requested as a 'user pays' service. Specialists are on hand to help you with analytical or methodological advice.

PHONE 1300 135 070

EMAIL client.services@abs.gov.au

FAX 1300 135 211

POST Client Services, ABS, GPO Box 796, Sydney NSW 2001

FREE ACCESS TO STATISTICS

All statistics on the ABS website can be downloaded free of charge.

WEB ADDRESS www.abs.gov.au