

**2015 REVIEW**

**SIMPLIFYING THE INTERSTATE WAGES REGRESSION MODEL**

**STAFF DISCUSSION PAPER  
CGC 2014-01-S**

**MARCH 2014**

|  |  |
| --- | --- |
| Paper issued | 27 March 2014x |
| Commission contact officer | Tony Nichols, (02) 6229 8833, anthony.nichols@cgc.gov.au |
| Submissions sought by | Due date: 30 April 2014.  Submissions should be emailed in Word format to [secretary@cgc.gov.au](mailto:secretary@cgc.gov.au) .  Submissions of more than 10 pages in length should include a summary section. |
| Confidential material | It is the commission’s normal practice to make State submissions available on its website under the CC BY licence, allowing free use of content by third parties.  Creative Commons - CC BY Attribution  Further information on the CC BY licence can be found at <http://www.creativecommons.org>  Confidential material contained in submissions must be clearly identified or included in separate attachment/s, and details provided in the covering email. Identified confidential material will not be published on the commission’s website. |

### BACKGROUND

* 1. The Interstate wages regression model used in the 2010 Review aims to measure the relative differences in wage levels between the States. The model uses the Australian Bureau of Statistics (ABS) Survey of Education and Training (SET) to regress the log of private sector earnings on State of employment and other measurable labour market influences.
  2. The SET allows for potentially dozens of variables to be included in the regression model. Currently the model includes 219 variables. Because the SET includes a relatively large number of observations, the standard errors for each variable are less sensitive to increases in the number of variables.[[1]](#footnote-1) However, it is important to ensure each variable retains theoretical relevance and assists in explaining the variation in wages.
  3. This paper examines how the regression model used in the 2010 Review can be simplified to include only those variables that have theoretical relevance and provide explanatory power. We also consider how some of the coding can be simplified. Attachment A gives the results of the simplified regression.

### Proposed adjustments

* 1. We have identified a number of areas where the regression model can be simplified and made more transparent. In this section we discuss the impact of reducing the number of variables and how we can adjust how some variables are defined. The GST impacts of the proposed adjustments are shown in the next section (Table 1).

#### Remove effects coding and use simple dummy variables

* 1. In the 2010 Review, the Interstate wages assessment used effects coding, rather than simple dummy variables, as the method for estimating the variation in interstate wages. Both methods effectively produce the same results. However, the interpretation of results is very different. Unlike simple dummy variables, effects coding calculates the difference from the national average of each State within the regression model.[[2]](#footnote-2) The dummy variable method calculates the difference of each State from a reference State (we have used Tasmania as the reference State). A process of standardisation is then used to calculate the deviation from the national average for each State outside the regression model.
  2. The effects coding method was used in the 2010 Review because it was considered simpler to calculate the difference from the national average within the regression model. However, this adds complexity to the regression model and the interpretation of the results is not immediately obvious. For this reason we have decided to use the simple dummy variable approach. This approach produces the same outcome but removes a layer of complexity in the regression coding.[[3]](#footnote-3)

#### Remove female interaction variables

* 1. Interaction variables are commonly used in regression analysis to measure the combined impact of two (or more) variables. For example, the current regression model uses an interaction variable by combining gender and education to test if an undergraduate degree increases wages more for males or females.
  2. In fact, the current model includes a female interaction variable for every variable in the model. This nearly doubles the number of variables in the model from 115 to 219. While theoretically there may be differences in the relationship between productivity influences on wages for males and females we have found that the inclusion of female interaction variables only increases the explanatory power of the model by a small amount. The R squared increases from 0.783 to 0.790 when all 114 female interaction variables are included in the model. This suggests they do not explain much of the change in wages but add complexity and potentially decrease the precision of the model. Therefore, staff propose to remove the female interaction variables from the regression model.

#### Remove the variable of hours worked less than 15 hours and greater than 60 hours

* 1. Currently the model measures the impact of the number hours worked by including a variable that is the log of an employee’s continuous number of hours worked per week. The model includes two additional variables that measure the impact of working less than 15 hours and more than 60 hours. We have found that neither of the additional hours worked variables add any explanatory power beyond the information included in the continuous hours worked variable. Therefore, staff propose to remove the less than 15 and more than 60 hours worked variables.

### IMPACT OF PROPOSED ADJUSTMENTS

* 1. In the 2013 discussion papers to the States, staff proposed a change from using the whole of State wage to assess interstate wage relativities to using capital city wages. We have used capital city wages as the basis for this analysis.
  2. Figure 1 shows how interstate relative wage levels change with each adjustment to the regression model. The adjustments were made progressively so that removing the under 15 and over 60 hours worked variables includes the impact of the previous adjustments.

Figure 1 Relative wage levels, SET 2009



Source: SET 2009.

* 1. Table 1 shows the impact from adjusting the model on the redistribution from EPC if the changes were applied in the 2014 Update. Changing to the dummy variable approach does not have an impact on the GST distribution. The proposed adjustments increase the impact of the Wages assessment by $7.6 per capita but do not have a material impact for any State.

Table 1 Impact from adjusting the model, redistribution from EPC, U2014 (a)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | NSW | Vic | Qld | WA | SA | Tas | ACT | NT | Redist. |
|  | $pc | $pc | $pc | $pc | $pc | $pc | $pc | $pc | $pc |
| 2014 Update assessment using capital city | 101.9 | -45.4 | -143.7 | 160.0 | -163.1 | -173.7 | 166.2 | 287.1 | 55.8 |
| Female interaction variables removed | 126.2 | -52.9 | -172.6 | 160.1 | -157.7 | -182.7 | 159.8 | 272.7 | 63.3 |
| Remove under 15 and over 60 hours worked logged variables | 126.3 | -52.8 | -172.8 | 160.2 | -158.2 | -182.0 | 160.0 | 272.8 | 63.4 |
| Difference from U2014 (b) | 24.4 | -7.4 | -29.2 | 0.2 | 4.9 | -8.2 | -6.2 | -14.3 | 7.6 |

(a) Adjustments were made progressively so that removing the under 15 and over 60 hours worked variables includes the impact of the previous adjustments.

(b) This is the difference between the 2014 Update Wages assessment using capital city and the combined impact from including all adjustments to the regression model.

Source: SET 2009.

### CONCLUSIONS

* 1. Staff have found that the female interaction variables and the under 15 and over 60 hours worked variables do not add explanatory power to the model. By removing them, the regression model is significantly simplified. We have also found that using the dummy variable approach to measure interstate wage differences simplifies the coding but produces the same outcome.
  2. Staff consider this simpler and more transparent model specification continues to produce a reliable estimation of interstate wage levels.

|  |
| --- |
| Staff propose to recommend the Commission:   * remove effects coding and use simple dummy variables * remove the female interaction variables * remove the variable hours worked less than 15 and greater than 60. |

## ATTACHMENT A – 2009 SET regression results

Table A - 2009 SET regression results including proposed adjustments

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Description of variable | Parameter Estimate | Standard Error | t value | P > |t| |
|  |  |  |  |  |
| Intercept | 2.44166 | 0.09632 | 25.35 | <.0001 |
| **State of residence** |  |  |  |  |
| New South Wales | 0.07441 | 0.04174 | 1.78 | 0.0747 |
| Victoria | 0.03414 | 0.04178 | 0.82 | 0.4139 |
| Queensland | 0.00313 | 0.0427 | 0.07 | 0.9416 |
| Western Australia | 0.06961 | 0.04288 | 1.62 | 0.1046 |
| South Australia | 0.01317 | 0.04373 | 0.3 | 0.7633 |
| **\*\***Tasmania |  |  |  |  |
| ACT | 0.08169 | 0.05344 | 1.53 | 0.1264 |
| Northern Territory | 0.09278 | 0.06018 | 1.54 | 0.1232 |
| **Sex** |  |  |  |  |
| Female | -0.12196 | 0.01188 | -10.27 | <.0001 |
| \*\*Male |  |  |  |  |
| **Marital status** |  |  |  |  |
| Married | 0.06129 | 0.01157 | 5.3 | <.0001 |
| **\*\***Not married |  |  |  |  |
| **Whether had any young children** |  |  |  |  |
| With children under 15 years old | -0.00346 | 0.01142 | -0.3 | 0.7618 |
| **\*\***Without children under 15 year old |  |  |  |  |
| **Whether permanent or casual** |  |  |  |  |
| Permanent with main period employer | 0.05839 | 0.01474 | 3.96 | <.0001 |
| **\*\***Casual with main period employer |  |  |  |  |
| **Hours usually worked per week** |  |  |  |  |
| Log of Number of hours | 0.96298 | 0.01294 | 74.41 | <.0001 |
| **Migrant status** |  |  |  |  |
| Born in ESC lived in Australia >20 years | 0.06057 | 0.02005 | 3.02 | 0.0025 |
| Born in ESC, lived in Australia 10-20 years | 0.12242 | 0.0343 | 3.57 | 0.0004 |
| Born in ESC, lived in Australia < 10 years | 0.0928 | 0.02623 | 3.54 | 0.0004 |
| Born in NESC, lived in Australia more than 20 years | -0.06534 | 0.01739 | -3.76 | 0.0002 |
| Born in NESC lived in Australia between 10-20 years | -0.12963 | 0.02139 | -6.06 | <.0001 |
| Born in NESC, lived in Australia less than 10 years | -0.13022 | 0.02057 | -6.33 | <.0001 |
| \*\*Born in Australia |  |  |  |  |
| **Size of firm (number of employees)** |  |  |  |  |
| Less than 20 | -0.00856 | 0.03116 | -0.27 | 0.7836 |
| 20-99 | 0.06395 | 0.03157 | 2.03 | 0.0428 |
| 100 and over | 0.13485 | 0.03165 | 4.26 | <.0001 |

Table A - 1 2009 SET regression results including proposed adjustments (continued)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Description of variable | Parameter Estimate | Standard Error | t value | P > |t| |
| **\*\***Number unknown |  |  |  |  |
| **Trade union membership** |  |  |  |  |
| Had trade union membership | 0.01032 | 0.01431 | 0.72 | 0.4708 |
| **\*\***Did not have trade union membership |  |  |  |  |
| **Detailed Occupation** |  |  |  |  |
| Managers nfd | 0.54598 | 0.08032 | 6.8 | <.0001 |
| Chief Executives, General Managers and Legislators | 0.87512 | 0.07785 | 11.24 | <.0001 |
| Farmers and Farm Managers | -0.07818 | 0.14308 | -0.55 | 0.5848 |
| Specialist Managers | 0.45574 | 0.04342 | 10.5 | <.0001 |
| Hospitality, Retail and Service Managers | 0.23431 | 0.04523 | 5.18 | <.0001 |
| Professionals nfd | 0.31208 | 0.14613 | 2.14 | 0.0327 |
| Arts and Media Professionals | 0.27977 | 0.08364 | 3.35 | 0.0008 |
| Business, Human Resource and Marketing Professionals | 0.36927 | 0.04424 | 8.35 | <.0001 |
| Design, Engineering, Science and Transport Professionals | 0.30609 | 0.04762 | 6.43 | <.0001 |
| Education Professionals | 0.22145 | 0.06356 | 3.48 | 0.0005 |
| Health Professionals | 0.37253 | 0.05796 | 6.43 | <.0001 |
| ICT Professionals | 0.39078 | 0.0505 | 7.74 | <.0001 |
| Legal, Social and Welfare Professionals | 0.29574 | 0.06739 | 4.39 | <.0001 |
| Engineering, ICT and Science Technicians | 0.22406 | 0.04912 | 4.56 | <.0001 |
| Automotive and Engineering Trades Workers | 0.06358 | 0.04704 | 1.35 | 0.1766 |
| Construction Trades Workers | 0.02258 | 0.05486 | 0.41 | 0.6806 |
| Electrotech and Telecommunications Trades Workers | 0.09176 | 0.05124 | 1.79 | 0.0734 |
| Food Trades Workers | -0.02728 | 0.05675 | -0.48 | 0.6307 |
| Skilled Animal and Horticultural Workers | -0.04719 | 0.07081 | -0.67 | 0.5051 |
| Other Technicians and Trades Workers | 0.08243 | 0.05126 | 1.61 | 0.1079 |
| Health and Welfare Support Workers | 0.20196 | 0.0674 | 3 | 0.0027 |
| Carers and Aides | 0.0958 | 0.04978 | 1.92 | 0.0543 |
| Hospitality Workers | 0.1095 | 0.05096 | 2.15 | 0.0317 |
| Protective Service Workers | 0.17796 | 0.07329 | 2.43 | 0.0152 |
| Sports and Personal Service Workers | 0.22276 | 0.05488 | 4.06 | <.0001 |
| Office Managers and Program Administrators | 0.31843 | 0.04802 | 6.63 | <.0001 |
| Personal Assistants and Secretaries | 0.22579 | 0.0541 | 4.17 | <.0001 |
| General Clerical Workers | 0.10692 | 0.05249 | 2.04 | 0.0417 |
| Inquiry Clerks and Receptionists | 0.10406 | 0.04632 | 2.25 | 0.0247 |
| Numerical Clerks | 0.0782 | 0.04563 | 1.71 | 0.0866 |
| Clerical and Office Support Workers | -0.07475 | 0.07036 | -1.06 | 0.288 |
| Other Clerical and Administrative Workers | 0.0702 | 0.04759 | 1.48 | 0.1402 |
| Sales Representatives and Agents | 0.19331 | 0.05131 | 3.77 | 0.0002 |
| Sales Assistants and Salespersons | 0.05321 | 0.0411 | 1.29 | 0.1955 |
| Sales Support Workers | 0.02986 | 0.04902 | 0.61 | 0.5424 |
| Machinery Operators and Drivers nfd | -0.22509 | 0.16255 | -1.38 | 0.1662 |

Table A - 1 2009 SET regression results including proposed adjustments (continued)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Description of variable | Parameter Estimate | Standard Error | t value | P > |t| |
| Machine and Stationary Plant Operators | 0.06045 | 0.05331 | 1.13 | 0.2568 |
| Mobile Plant Operators | 0.05833 | 0.05332 | 1.09 | 0.274 |
| Road and Rail Drivers | -0.05452 | 0.04943 | -1.1 | 0.2701 |
| Storepersons | -0.01096 | 0.05494 | -0.2 | 0.8419 |
| Cleaners and Laundry Workers | -0.06699 | 0.04936 | -1.36 | 0.1748 |
| Construction and Mining Labourers | 0.16235 | 0.05754 | 2.82 | 0.0048 |
| Factory Process Workers | -0.20878 | 0.04781 | -4.37 | <.0001 |
| Farm, Forestry and Garden Workers | -0.05238 | 0.07832 | -0.67 | 0.5036 |
| Food Preparation Assistants | 0.02245 | 0.05554 | 0.4 | 0.6861 |
| Inadequately described | 0.09871 | 0.09769 | 1.01 | 0.3123 |
| **Industry** |  |  |  |  |
| Agriculture, Forestry and Fishing | 0.03081 | 0.06828 | 0.45 | 0.6518 |
| Mining | 0.46166 | 0.04755 | 9.71 | <.0001 |
| Manufacturing | 0.16133 | 0.02865 | 5.63 | <.0001 |
| Electricity, Gas, Water and Waste Services | 0.34303 | 0.04875 | 7.04 | <.0001 |
| Construction | 0.18985 | 0.03369 | 5.64 | <.0001 |
| Wholesale trade | 0.15398 | 0.03277 | 4.7 | <.0001 |
| Retail trade | 0.08902 | 0.03047 | 2.92 | 0.0035 |
| Accommodation and Food Services | 0.04924 | 0.03611 | 1.36 | 0.1728 |
| Transport, Postal and Warehousing | 0.19385 | 0.03556 | 5.45 | <.0001 |
| Information Media and Telecommunications | 0.18157 | 0.03933 | 4.62 | <.0001 |
| Financial and Insurance Services | 0.24782 | 0.03295 | 7.52 | <.0001 |
| Rental, Hiring and Real Estate Services | 0.14361 | 0.04451 | 3.23 | 0.0013 |
| Professional, Scientific and Technical Services | 0.17821 | 0.03074 | 5.8 | <.0001 |
| Administrative and Support Services | 0.14317 | 0.03522 | 4.06 | <.0001 |
| Public Administration and Safety | 0.04752 | 0.05556 | 0.86 | 0.3924 |
| Education and Training | 0.02758 | 0.04584 | 0.6 | 0.5475 |
| Health Care and Social Assistance | 0.04479 | 0.03217 | 1.39 | 0.1639 |
| Arts and Recreation Services | 0.06776 | 0.04959 | 1.37 | 0.1719 |
| Inadequately described | -0.01378 | 0.11641 | -0.12 | 0.9058 |
| \*\*Other Services |  |  |  |  |
| **Level of highest education attainment** |  |  |  |  |
| Higher degree | 0.47845 | 0.05475 | 8.74 | <.0001 |
| Postgraduate diploma | 0.41063 | 0.05761 | 7.13 | <.0001 |
| Bachelor degree | 0.37277 | 0.05002 | 7.45 | <.0001 |
| Advanced diploma/diploma | 0.18219 | 0.05011 | 3.64 | 0.0003 |
| Certificate III or IV | 0.16794 | 0.05042 | 3.33 | 0.0009 |
| Certificate I or II | 0.0001423 | 0.06756 | 0 | 0.9983 |
| Certificate not defined | -0.02297 | 0.16561 | -0.14 | 0.8897 |
| Year 12 | 0.05655 | 0.01581 | 3.58 | 0.0004 |
| \*\*Did not complete year 12/unknown |  |  |  |  |

Table A - 1 2009 SET regression results including proposed adjustments (continued)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Description of variable | Parameter Estimate | Standard Error | t value | P > |t| |
| **Main field of highest educational attainment** |  |  |  |  |
| Natural and physical sciences | -0.00353 | 0.03699 | -0.1 | 0.9239 |
| Information technology | -0.00262 | 0.03779 | -0.07 | 0.9447 |
| Engineering and related technologies | 0.10686 | 0.02686 | 3.98 | <.0001 |
| Architecture and building | 0.13148 | 0.04052 | 3.24 | 0.0012 |
| Agriculture, environmental and related studies | -0.00304 | 0.0524 | -0.06 | 0.9537 |
| Health | 0.07203 | 0.03614 | 1.99 | 0.0463 |
| Education | 0.03262 | 0.03979 | 0.82 | 0.4124 |
| Management and commerce | 0.0567 | 0.02282 | 2.48 | 0.013 |
| Creative arts | -0.02978 | 0.03347 | -0.89 | 0.3736 |
| Food, hospitality and personal services | 0.07194 | 0.03685 | 1.95 | 0.051 |
| Mixed program or unknown | 0.13552 | 0.05125 | 2.64 | 0.0082 |
| \*\*Society and culture |  |  |  |  |
| **Cumulative duration of employment** |  |  |  |  |
| Under 1 year | -0.06498 | 0.02534 | -2.56 | 0.0104 |
| 1–4 years | -0.05741 | 0.02391 | -2.4 | 0.0164 |
| 5–9 years | -0.01147 | 0.02441 | -0.47 | 0.6384 |
| 10–19 years | 0.01652 | 0.02536 | 0.65 | 0.5147 |
| **\*\***20 years and over |  |  |  |  |
| **Estimated work experience (years)** |  |  |  |  |
| Experience | 0.02535 | 0.00149 | 17 | <.0001 |
| Experience square | -0.0004589 | 3.12E-05 | -14.71 | <.0001 |
| Note: Experience is calculated by first finding 'age-age left school'. Then we deduct the time spent in training beyond 12 yrs at school. For example, 5 years for post-graduate, 4 years for graduate diploma and so on. |  |  |  |  |

Note: \*\* is the reference variable.

Source: SET 2009.

1. This number includes all dummy variables within each category and all female interaction variables. [↑](#footnote-ref-1)
2. Effects coding still requires a reference State to be excluded in the calculation. [↑](#footnote-ref-2)
3. The dummy variable approach produces different State coefficients but the relative differences are the same. [↑](#footnote-ref-3)