The Australian Economic Review

The Australian Economic Review, vol. 47, no. 1, pp. 44–62

Occupational Differences and the Australian Gender Wage Gap

Michael B. Coelli*

Abstract

There are conflicting findings in the literature about whether occupational differences contribute to the Australian gender wage gap. Most papers conclude that they do not, while two papers conclude that they can explain a sizeable amount of the gap. These papers use different estimation methods and levels of occupational aggregation. I show that constructing decompositions using much more disaggregated occupational categories than those used in previous Australian research yield large positive contributions of occupational differences to the gap. Differences in estimation method (particularly regarding the inclusion or otherwise of industry indicators) also affect estimates.

* Department of Economics, The University of Melbourne, Victoria 3010 Australia; email <mcoelli@unimelb.edu.au>. Many thanks to David Ong for research assistance and to Jeff Borland, Deborah Cobb-Clark and Paul Miller for useful comments and suggestions. Participants at the 2011 Australasian Meetings of the Econometric Society in Adelaide and the University of New South Wales also provided useful suggestions. This article reports statistical analysis based on unconfidentialised unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Survey project was initiated and is funded by the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this article, however, are those of the author, and should not be attributed to FaHCSIA or the Melbourne Institute. I thank Michelle Summerfield of the Melbourne Institute for her assistance in accessing the In Confidence version of the HILDA Survey. Financial support was provided by the Faculty of Business and Economics, The University of Melbourne.

1. Introduction

Although not large by international standards, there remains a sizeable gap (around 12 per cent) between the average hourly wage rates of men and those of women in Australia. Occupational segregation by gender in Australia, however, remains quite high. Men and women predominantly work in different occupations. A standard Duncan and Duncan (1955) measure of occupational segregation for Australia reveals that over 50 per cent of women would have to change occupations in order to have the same occupational distribution as men.[1]

This article investigates whether these large occupational differences contribute to the Australian gender wage gap. In other words, is it the case that women are overly represented in low-paying occupations in Australia, as the popular press in particular, but also a minority of researchers,[2] might lead us to believe? This is an important question when trying to ascertain whether comparable worth policies (equal pay for work of equal value) have a role to play in the Australian labour market. An important example of the relevance of this question is the 2010 ‘equal remuneration’ case brought before Fair Work Australia by unions representing workers in the Social and Community Services sector. Austen, Jefferson and Preston (2013) provide interesting insights into the competing arguments made in this case and those arguments hinge on whether or not women are overly represented in low-paying occupations.

Prior research in economics has reached distinctly opposite conclusions regarding whether occupational differences contribute to the Australian gender wage gap. The majority of papers concluded that occupational differences do not contribute to the gap (Chapman and Mulvey 1986; Rimmer 1991;

Kidd 1993; Kidd and Meng 1997; Lee and Miller 2004; Kee 2006; Barón and Cobb-Clark 2010; Cobb-Clark and Tan 2011). While women are under-represented in high-paying manager occupations, they are also under-represented in the lowest paying labourer occupations. In fact, some studies concluded that if women had the same occupational distribution as men, the wage gap would actually be larger, not smaller.

Two pieces of research, however, found that the 'femaleness' of an occupation was related to significantly lower wage rates (Miller 1994; Wooden 1999). That is, if the proportion of women in an occupation is high, average wages in the occupation are low. They thus conclude that occupational differences can explain a sizeable proportion of the Australian gender wage gap.[3] This is in accordance with findings for the United States (Sorensen 1990; Groshen 1991; Macpherson and Hirsch 1995) using the same estimation method.

The main objective of this article is to ascertain whether or not occupational differences do indeed contribute to the Australian gender wage gap. Some effort is expended, however, to understand which particular elements of the prior studies may be causing such contrasting findings. The contrasting findings could be due to several elements that vary across studies, particularly the level of occupational aggregation, the estimation technique and the characteristics of individuals and jobs that are controlled for during estimation.

The prior Australian research all grouped occupations quite coarsely, into just seven-to-nine groups (one-digit level) or 40–50 groups (two-digit level). In this investigation, I show that moving from the one-digit to the two-digit level of occupational aggregation increases the proportion of the wage gap attributable to occupational differences, explaining some of the disparity in findings in the literature. In addition, I show that disaggregating occupations even further than the previous literature has—to the three- and four-digit levels—increases the estimated occupational contribution even further. Austen, Jefferson and Preston (2013) argued that women were over-represented in low-paying occupations *within* the broad one- and two-digit occupational groups employed in previous research. The evidence I report here provides support for this argument.

The literature primarily uses the Blinder (1973) and Oaxaca (1973) decomposition methods to isolate the contribution of differences in observable characteristics (including occupational differences) to the mean gender wage gap. The main difference between the estimation techniques employed in the studies mentioned above is how occupation is entered in the wage regressions that underpin the decompositions. The majority of research generally includes individual indicators for each occupational group in the regressions, while both Miller (1994) and Wooden (1999) instead include measures of the proportion female in an occupation. The only difference is thus whether or not a functional form is imposed on the relationship between wages and the proportion female in an occupation.

Imposition of a functional form may seem innocuous, yet it does affect the size of estimates of the occupational contribution. In particular, occupation and industry of employment are closely related, with many occupations defined within industry. Estimates of the occupational contribution change appreciably if industry of employment is also included in the wage regressions. This change is not as large if occupations are entered using a proportion female measure, rather than as individual indicators, due to correlations among covariates being smaller in the former case.

The main finding of this investigation is that occupational differences do contribute to the gender wage gap in Australia when occupation is defined at an appropriately disaggregated level. This finding is reasonably robust to choice of estimation method. Estimates of the occupational contribution do, however, vary depending on whether controls for industry of employment are included.

The outline of the rest of the article is as follows. A description of prior Australian research is provided in Section 2, along with a comparison of the main empirical techniques employed in this research. A simple analysis of the effect of moving towards more disaggregated occupational breakdowns on estimates of the occupational contribution to the gender wage gap is provided in Section 3. Estimates using the main competing decomposition techniques are provided in Section 4. Section 5 concludes.

2. The Australian Literature

The main empirical concept employed to investigate mean gender wage gaps is decomposition analysis (Blinder 1973; Oaxaca 1973). First, log wage regressions are estimated separately for men and women (equations (1) and (2), respectively). The vectors X_i^m and X_i^f almost always include individual productive characteristics, such as education and labour market experience, while marital status and presence of children are also commonly included. Certain job characteristics are also included in some studies: part-time status, union membership, industry, occupation and sector of employment etc.:

$$W_i^m = X_i^m \beta^m + \varepsilon_i^m \quad (1)$$

$$W_i^f = X_i^f \beta^f + \varepsilon_i^f \quad (2)$$

The estimated vectors $\hat{\beta}^m$ and $\hat{\beta}^f$ are then employed to decompose the mean log wage gap into a component attributable to mean differences in individual and job characteristics across genders (first term on right of equation (3), sometimes referred to as 'explained') and a remainder or residual that is often interpreted as 'discrimination'.[4] The residual reflects differences in the estimated coefficients; particularly, the constant term, but also potential differences in the 'labour market returns' to individual and job characteristics across genders. For example, women may be rewarded less for obtaining higher levels of education:

$$\overline{W}^m - \overline{W}^f = \left(\overline{X}^m - \overline{X}^f\right)\hat{\beta}^m + \overline{X}^f\left(\hat{\beta}^m - \hat{\beta}^f\right) \quad (3)$$

In the majority of the relevant Australian literature, occupation is entered in the wage regressions using separate indicators for each group. The resulting decomposition can be written as equation (4), where all characteristics apart from occupation are collected in matrix X_c, while the occupational indicators comprise matrix X_o:

$$\overline{W}^m - \overline{W}^f = \left(\overline{X}_c^m - \overline{X}_c^f\right)\hat{\beta}_c^m + \left(\overline{X}_o^m - \overline{X}_o^f\right)\hat{\beta}_o^m + \overline{X}^f\left(\hat{\beta}^m - \hat{\beta}^f\right) \quad (4)$$

The component $(\overline{X}_o^m - \overline{X}_o^f)\hat{\beta}_o^m$ is the contribution of occupational differences across genders to the gender wage gap.[5] It can be shown (see Appendix 1) that this component can be rewritten, as in equation (5), where P_j^m and P_j^f are the proportions of men and women employed in occupation j respectively. The $\widetilde{W}_j^m$ are measures of mean wages by occupation adjusted for the characteristics included in X_c. Occupational differences across genders contribute to the wage gap if men are more likely to be employed in occupations that pay higher 'productivity-adjusted' wages; that is, $P_j^m > P_j^f$ in occupations where $\widetilde{W}_j^m$ is high:

$$\left(\overline{X}_o^m - \overline{X}_o^f\right)\hat{\beta}_o^m = \sum_{j=1}^{J}\left(P_j^m - P_j^f\right)\widetilde{W}_j^m \quad (5)$$

Prior Australian research using this method includes Chapman and Mulvey (1986) and Cobb-Clark and Tan (2011). Both studies found that occupational differences were a large negative contributor to the wage gap. In other words, women are generally employed in higher rather than lower paying occupational groups. These two studies, however, grouped occupations quite coarsely, using seven and 18 groups, respectively.

A variant of the above decomposition is detailed in equation (6) (see Treiman and Hartman 1981). The terms $\overline{W}_j^m$ and $\overline{W}_j^f$ are unadjusted mean log wages by occupation and gender. This variant is essentially the standard decomposition of equation (3), but only occupational indicators are included in X:

$$\overline{W}^m - \overline{W}^f = \sum_{j=1}^{J}\left(P_j^m - P_j^f\right)\overline{W}_j^m + \sum_{j=1}^{J}P_j^f\left(\overline{W}_j^m - \overline{W}_j^f\right) \quad (6)$$

Rimmer (1991) and Lee and Miller (2004) both employed this variant, using occupations defined at the two-digit level. Rimmer found occupational differences detracted from the wage gap, while Lee and Miller found they explain a small proportion of the gap.[6]

The method employed by Miller (1994) and Wooden (1999) is also based on equation (3). In this case, instead of including individual indicators for each occupational group in the wage regressions, a variable denoting the proportion of employment in the occupation that is female (denoted *PF*) is included. Thus, for an individual employed in occupation *j*, the variable *PF* takes the value:

$$PF_j = \frac{\text{number of women employed in occupation } j}{\text{total number of employees in occupation } j} \tag{7}$$

The decomposition in this case can be written as equation (8). The term $\hat{\beta}^m_{PF}$ is the estimated coefficient on the *PF* variable and $(\overline{PF}^m - \overline{PF}^f)\hat{\beta}^m_{PF}$ is the occupational contribution. This contribution will be positive if $\hat{\beta}^m_{PF}$ is negative (lower wages in occupations dominated by women) as $(\overline{PF}^m - \overline{PF}^f)$ will be negative (women more likely to be employed in high-*PF* occupations):

$$\overline{W}^m - \overline{W}^f = \left(\overline{X}^m_c - \overline{X}^f_c\right)\hat{\beta}^m_c + \left(\overline{PF}^m - \overline{PF}^f\right)\hat{\beta}^m_{PF} + \overline{X}^f\left(\hat{\beta}^m - \hat{\beta}^f\right) \tag{8}$$

It can be shown (again, see Appendix 1) that this component can be written as follows:

$$\left(\overline{PF}^m - \overline{PF}^f\right)\hat{\beta}^m_{PF} = \sum_{j=1}^{n}\left(P^m_j - P^f_j\right)\hat{W}^m_j \tag{9}$$

In this case, the $\hat{W}^m_j = \beta^m_{PF} \times PF_j$ are essentially the predicted productivity-adjusted wages in each occupation from the male log wage regression, constructed using the estimated coefficient on *PF*. The only difference between equations (9) and (5) is whether or not a functional form is imposed on the relationship between wages and *PF*. There is thus no clear indication of whether one method or the other will estimate a larger contribution of occupational differences to the gender wage gap, unless imposing a linear relationship between wages and *PF* is incorrect.

Researchers using *PF* measures, however, often allow for non-linear relationships by employing a small number of indicators that group occupations by *PF* level. Miller (1994) includes two indicators: one for occupations with a *PF* below 0.3 and a second for occupations with a *PF* of between 0.3 and 0.6. Wooden (1999) includes four indicators that split occupations into quintiles based on *PF*. Unlike the majority of Australian research, both Miller (1994) and Wooden (1999) found that occupational differences explain in excess of 40 per cent of the Australian gender wage gap. When constructing their *PF* measures, both authors used occupations defined at the relatively more disaggregated two-digit level.

Finally, Kidd (1993), Kidd and Meng (1997) and Cobb-Clark and Tan (2011) employed a decomposition technique proposed by Brown, Moon and Zoloth (1980). This technique decomposes the wage gap into explained and unexplained components, both across and within separate occupational groups. It essentially breaks down the two components of equation (6) into separate explained and unexplained components within each one. Consistent with the Australian research using the standard Oaxaca–Blinder method and separate occupational indicators, these three studies found that the overall contribution of occupational differences (explained and unexplained) to the wage gap ranged from a large negative to zero. Women were not overly represented in low-paying occupational groups. Note that these studies also defined occupation at an aggregated level (seven or 18 occupational groups).

3. Data and Occupational Disaggregation

Of the Australian studies discussed above, approximately half defined occupations at the very broad one-digit level; that is, just seven-to-nine groups. The other half of the studies employed occupation defined at the more disaggregated two-digit level (40–50 groups). In this section, I use the simple decomposition technique of equation (6) to show that the level of occupational aggregation has a large impact on estimates of the contribution of occupational differences to the gender wage gap.

In the investigation, I employ data from waves 1 to 9 of the Household, Income and Labour Dynamics in Australia (HILDA) Survey (covering the years 2001–09). For a more detailed description of the HILDA Survey, see Watson and Wooden (2010). The HILDA Survey is a nationally representative, longitudinal survey of Australian household members in private dwellings. All members of responding households in wave 1 (2001) form the panel of individuals to be followed over time. All individuals aged 15 years or older in households where original panel members reside are interviewed annually. This includes original panel members plus individuals who join these households, even if temporarily. All children of original panel members are interviewed and followed over time once they turn 15 years old.

The advantage of the HILDA Survey over other potential Australian data sources for this investigation is the ability to construct a continuous usual hourly wage measure and the availability of many productive characteristics of workers. Alternative Australian data sources (Census, Survey of Education and Training) often provide weekly wage and/or weekly hours of work information in ranges only, making it difficult to construct a continuous hourly wage measure. In addition, these alternative sources do not provide a measure of actual work experience, a major determinant of wages.

As with all longitudinal surveys, the HILDA Survey suffers from sample attrition; that is, some panel members cannot be located for interview or do not respond in later years. While re-interview rates in the HILDA Survey are high relative to similar panel surveys overseas (Watson and Wooden 2010), there is still concern that the estimates below may suffer from attrition bias. The dependent variable of interest in this analysis is the hourly wage. If sample attrition is a function of wages earned, then the estimates provided below may suffer from such attrition bias. Goldberger (1981) shows that attrition based on values of the dependent variable of interest lead to attenuation (shrinking towards zero) in regression-based estimates. Of particular interest here are measures of wages by occupation, either in raw form or after controlling for the productive characteristics of workers. Sample attrition may thus result in an underestimate of wage differences by occupation and thus an attenuation of estimates of the contribution of occupational differences across genders to the gender wage gap.

When conducting the analyses below, I focus on individuals aged 22–60 years old in paid employment only (the self-employed and employers are excluded).[7] By focusing on this particular age range (also the focus in Barón and Cobb-Clark 2010), the influence of complexities in the labour market choices and outcomes of young individuals[8] and individuals contemplating early retirement are potentially avoided. Such choices and outcomes may differ considerably across gender and occupation.

Summary statistics for the estimation sample are provided in Table A1. Brief descriptions of the key variables are provided in Table A2. The dependent variable of interest here is the hourly wage, and in particular, the gender wage gap. This gap is estimated as 0.118 log points (approximately 12 per cent) in these data. Note that the hourly wage is constructed by taking the ratio of usual weekly earnings and usual hours worked in each individual's main job. As in all survey data, both the numerator and denominator in this ratio may suffer from measurement error due to respondents reporting either element incorrectly. If any errors in responses are purely random deviations from the true value (often referred to as 'classical measurement error'), then no systematic biases in the estimates should result.[9] However, if there are systematic biases in responses, the estimates to follow may also suffer from bias. Wooden, Wilkins and McGuinness (2007) show that hourly wages that are constructed using the HILDA Survey data are in alignment with other Australian survey data. This provides some comfort that measurement error is unlikely to be any more of a problem when employing these data than when employing alternative Australian data sources.

Notable differences in characteristics between men and women (see Table A1) include the greater proportion of women with Bachelor degrees, graduate certificates and diplomas and the higher proportion of men with Certificates III/IV. Men have more work experience, occupation tenure and job tenure. Women are more likely to work for government and not-for-profit employers, much more likely to work part-time and more likely to hold casual positions. Women are less likely to work in the industries of agriculture, mining, manufacturing, construction and transport and more likely to work in education and training, health care and social assistance.

At the bottom of Table A1, means are provided for *PF* constructed at each level of occupational aggregation from one digit (eight occupational groups) to four digits (up to 474 occupations).[10] Even at the one-digit level, large differences between men and women in occupation of employment are obvious. The mean *PF* values are 0.57 and 0.40 for women and men, respectively.[11] As one moves to higher levels of occupational disaggregation, the gender difference in average *PF* increases, implying higher levels of occupational segregation.[12]

The results of simple decompositions of the log gender wage gap using equation (6) are presented in Table 1 for all four levels of occupational aggregation. Note that these simple decompositions use mean raw log wages by occupation. Decompositions using log wage regressions that also control for other observable differences across genders (education, experience, industry etc.) will be presented in the next section. Note also that all estimates provided in this article employ the cross-sectional population weights provided with the HILDA Survey data. These weights are provided to ensure that any estimates constructed using the HILDA Survey data reflect the Australian population. Sample attrition is one main reason why respondents to the HILDA Survey may not accurately represent the Australian population. The survey weights help correct for such non-random attrition.[13]

It is clear that the level of occupational aggregation is extremely important when constructing measures of the contribution of occupational differences to the gender wage gap. If occupations are grouped into eight broad one-digit categories, the occupational contribution is negative. This is in line with previous research that also grouped occupations at this broad level. The main reason for this is that women, while under-represented in the high-paying manager group, are also under-represented in the lowest paying labourer group. Women are also over-represented in the reasonably well-paying clerical and administrative group. To see this, Table A3 provides wages by occupational group at the one- and two-digit levels.

If occupations are defined at the more disaggregated two-digit level (44 groups),

Table 1 Decompositions of Hourly Earnings Gap by Level of Occupational Aggregation[a]

Level of aggregation	*Occupations identified*	*Occupational contribution[b]*	*Within-occupation differences*	*Total log wage gap*
One-digit	8	–0.0125	0.1309	0.1184
Two-digit	44	0.0185	0.0999	0.1184
Three-digit	103	0.0356	0.0828	0.1184
Four-digit	326	0.0425	0.0759	0.1184

Notes: (a) Decompositions were based on models estimated using 23,061 and 23,077 observations (on 5,239 and 5,161 individuals) from the Household, Income and Labour Dynamics in Australia Survey's waves 1–9 for female and male employees aged 22–60 years, respectively. Cross-sectional population weights were employed during estimation. The effect of real wage growth over time has been controlled for using unrestricted time indicators.
(b) The occupational contribution was constructed using male average wages by occupation.

occupational differences now contribute positively to the gender wage gap. The finer level of detail allows the estimator to capture important differences within the broad occupational groups. For example, the estimator now employs the detail that women are over-represented in the relatively low-paying carer and aide occupations and under-represented in the higher paying protective services occupations. Both these occupations fall within the middle-paying community and personal service occupational group (Table A3).

If one moves to even more finely disaggregated occupations (three and four digits), the estimated contribution of occupational differences increases even further. At the four-digit level, occupational differences can 'explain' approximately 36 per cent of the gender wage gap. Moving to more disaggregated occupational groupings is vitally important to understanding differences across genders in occupations and wages. For example, it is only by breaking occupations down to the three-digit level that one is able to take account of women being over-represented among less well-paid nursing, while men are over-represented among the very high-paying medical practitioners group. Both occupations are within the 'health professionals' two-digit occupational category.[14]

The occupational categories employed in this investigation may also suffer from measurement error, with two potential sources of error. First, survey respondents may incorrectly report their occupation. Second, there may be errors in how the HILDA Survey team code the verbatim (text) responses provided by survey participants to the standard occupational code-frame used here. Watson and Summerfield (2009) provide evidence that there is considerable error in this coding step for both occupation and industry. Potentially 15 per cent of occupations may be coded incorrectly at the four-digit level, with lower error rates at less disaggregated occupational levels. Such occupational misclassification will most likely result in underestimation of differences in earnings by occupation and thus underestimation of the contribution of occupational differences to the gender wage gap. As such misclassification increases as one moves from the aggregated one-digit level to the disaggregated four-digit level, the extent of underestimation may increase.

It must also be noted that defining occupations at disaggregated levels (three and four digits) using the HILDA Survey data leads to small cell sizes within occupations and gender. For example, at the four-digit level, 54 of the 326 identified occupations[15] had less than 20 individual-wave observations in the sample employed during estimation. To check the robustness of the findings of Table 1, I repeated the exercise using information on average hourly wage rates by occupation provided in the 2008 Employee Earnings and Hours data release of the Australian Bureau of Statistics (2009). This release provides wage information by occupation (as reported by employers), broken down to the three-digit level, for full-time adult non-managerial workers. Using these data, occupational differences were estimated to 'explain' approximately 12 per cent of the gender wage gap using male average wages (21 per cent using female wages).[16] The smaller contribution estimated using these data rather than the HILDA Survey data is primarily due to the exclusion of managerial and part-time employees from the former data source. These particular employees are important for understanding the contribution of occupational differences to the gender wage gap.

4. Estimates by Technique

The discussion of Section 2 noted four main estimation techniques used in the literature:

(i) standard Oaxaca–Blinder with occupation included using separate indicators;

(ii) standard Oaxaca–Blinder with occupation included via the *PF* measure;

(iii) simple decomposition using unadjusted average wages by occupation (equation (6)); and

(iv) Brown–Moon–Zoloth (BMZ) extension of Oaxaca–Blinder.

In terms of estimating the overall contribution of occupational differences to the gender wage gap, the BMZ technique is essentially the same as using equation (6) (see Section 2); that is, both essentially use unadjusted average wages by occupation. Estimates using this technique are presented in Table 1 and were discussed above.

In this section, I provide estimates that employ the first two techniques listed above, both of which employ wages by occupation that have been adjusted for the productive characteristics of workers. As discussed in Section 2, the only real difference between these two techniques is whether or not a functional form is imposed on the relationship between wages by occupation and *PF*. One will see how important this imposition can potentially be. I estimate two versions using *PF* measures: one that includes a linear *PF* term and one that includes separate indicators for occupations that are grouped into three groups based on occupational *PF*, as estimated by Miller (1994). I estimate all models, both including and excluding industry controls. This choice has a significant effect on the estimates.

In all estimated models, controls are included for a comprehensive set of characteristics (summarised in Table A1). Instead of reporting coefficient estimates on these additional characteristics for all models, estimates for a base model (with no occupational control) are provided in Table A4. These estimates are in line with estimates in the previous literature. Individuals earn higher wages if they have higher education levels, more work experience and tenure, are married, live in urban areas, are Australian-born or from an English-speaking background, are union members, work in larger firms and do not work for not-for-profit employers.

Decomposition estimates using occupations defined at the two-digit level are provided in Table 2. Estimates with occupations defined at the four-digit level are provided in Table 3 and will be discussed later. The numbers provided are the estimated contributions of each set of covariates to the gender wage gap. The third row in these tables, labelled 'Other variables', provides estimates of the contribution of differences in individual and worker characteristics apart from occupation and industry. Differences between genders in these characteristics are estimated to 'explain' between 19 and 42 per cent of the wage gap, depending on the version of the model being estimated. The particular characteristics that contribute most of this effect are work experience, occupation tenure and employer type (women are more likely to be employed by low-paying, not-for-profit employers).

Decomposition results for the models that include individual occupational indicators are presented in the first two columns of

Table 2 Decompositions by Estimation Method, Using Two-Digit Occupations[a]

	Occupational indicators		*Linear PF[b] term*		*Non-linear PF terms*	
Contribution	*No industry*	*Industry*	*No industry*	*Industry*	*No industry*	*Industry*
Occupation[c]	0.0224	–0.0044	0.0444	0.0141	0.0506	0.0193
Industry	–	0.0550	–	0.0394	–	0.0367
Other variables	0.0352	0.0229	0.0497	0.0347	0.0465	0.0335
Unexplained	0.0608	0.0448	0.0243	0.0302	0.0213	0.0289
Total log gap	0.1184	0.1184	0.1184	0.1184	0.1184	0.1184

Notes: (a) Decompositions were based on models estimated using 23,061 and 23,077 observations (on 5,239 and 5,161 individuals) from the Household, Income and Labour Dynamics in Australia Survey's waves 1–9 for female and male employees aged 22–60 years, respectively. Cross-sectional population weights were employed during estimation. The other covariates that are included in the models are those listed in Table A4.
(b) *PF* denotes proportion of employment in the occupation that is female.
(c) Occupation was defined at the two-digit level and male wage returns were employed.

Table 3 Decompositions by Estimation Method, Using Four-Digit Occupations[a]

	Occupational indicators		*Linear PF[b] term*		*Non-linear PF terms*	
Contribution	*No industry*	*Industry*	*No industry*	*Industry*	*No industry*	*Industry*
Occupation[c]	0.0466	0.0063	0.0597	0.0257	0.0609	0.0352
Industry	–	0.0619	–	0.0375	–	0.0347
Other variables	0.0271	0.0155	0.0487	0.0340	0.0480	0.0330
Unexplained	0.0447	0.0347	0.0100	0.0212	0.0095	0.0155
Total log gap	0.1184	0.1184	0.1184	0.1184	0.1184	0.1184

Notes: (a) Decompositions were based on models estimated using 23,061 and 23,077 observations (on 5,239 and 5,161 individuals) from the Household, Income and Labour Dynamics in Australia Survey's waves 1–9 for female and male employees aged 22–60 years, respectively. Cross-sectional population weights were employed during estimation. The other covariates included in the models are those listed in Table A4.
(b) *PF* denotes proportion of employment in the occupation that is female.
(c) Occupation was defined at the four-digit level and male wage returns were employed.

Table 2. If industry indicators are not included, occupational differences across genders are estimated to contribute 0.0224 log points (19 per cent) to the gender wage gap of 0.1184 log points.[17] If industry indicators are included, however, the occupational contribution is a negative 0.0044 log points. Thus, the inclusion of industry indicators alters the estimates significantly.

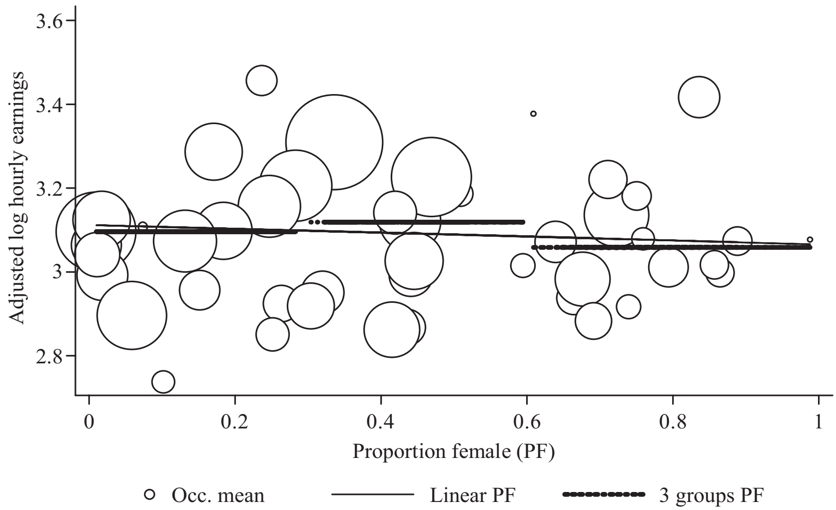
This large change in the estimated occupational contribution is due to the close relationship between industry and occupational indicators in the data. Note the very large contribution of industry in the second column (0.055 log points).[18] When I re-estimated the model after excluding the occupational indicators (results not reported), the industry contribution was a smaller 0.044 log points. Including the occupational indicators has resulted in a boosting of the estimated industry contribution. The industry indicators are soaking up part of the occupational effect due to strong collinearity amongst the two sets of indicators.

Decomposition results for models that include a linear measure of *PF* are presented in the middle two columns of Table 2. If industry controls are excluded, occupational differences are estimated to contribute a large 0.0444 log points (38 per cent) to the gender wage gap.[19] If industry controls are included, a positive occupational contribution is still estimated, albeit the estimate is a smaller 0.0141 log points (12 per cent).[20] In this case, the estimated contribution of industry is still large, at 0.0394 log points, yet smaller than when individual occupational indicators are included. This is due to the less strong relationship between the industry indicators and *PF* than between the same industry indicators and the individual occupational indicators. Note that this industry inclusion effect is more pronounced than that reported by Wooden (1999).[21]

Decomposition results for models that include a non-linear function of *PF* are presented in the final two columns of Table 2. In these two models, indicators for occupations with $PF < 0.3$ and for $0.3 \leq PF < 0.6$ were included instead of a linear *PF* term (the excluded base category was $PF \geq 0.6$). These estimates are similar to those when the linear *PF* measure is used, yet a slightly smaller occupational contribution is estimated. The coefficient estimates on the *PF* indicators revealed a concave relationship between adjusted wages and *PF*. Wages are lower among male-dominated occupations ($PF < 0.3$) than among integrated occupations ($0.3 \leq PF < 0.6$), while wages are lowest among female-dominated occupations ($PF \geq 0.6$).

This concave relationship is illustrated in Figure 1. In this figure, each circle represents a two-digit occupation, with the centre of the circle plotting the mean productivity-adjusted log wage for men in the occupation (vertical axis) against the *PF* value for that occupation (horizontal axis). These productivity-adjusted

Figure 1 Occupation and Adjusted Log Hourly Earnings, Males, Two–Digit Level



Note: *PF* denotes proportion of employment in the occupation that is female.

Source: The Household, Income and Labour Dynamics in Australia Survey, waves 1–9 (2001–09) and author calculations.

log wages by occupation were constructed from the log wage model for men that includes all the characteristics in Table A4, industry and occupational indicators. The coefficients of the occupational indicators were used to construct these productivity-adjusted log wage measures by occupation. The relative size of the circles reflects the number of male employees in that occupation (based on the HILDA Survey weights). One can see relatively larger circles for men in the low-*PF* (male-dominated) technicians and trades and operators and drivers occupations, with smaller circles for the high-*PF* clerical and community and personal services occupations.

The downward sloping thin line in Figure 1 represents the slope of the coefficient on the *PF* measure in the log wage model for men, again including the industry indicators. This coefficient estimate was –0.0462 and can be interpreted as follows. Wages in an occupation that is completely male-dominated (*PF* of zero) are approximately 4.6 per cent higher than wages in an occupation that is completely female-dominated (*PF* of one). The circles for each occupation depict a relationship with *PF* that is both noisy and concave. The thick horizontal lines represent average productivity-adjusted wages for the three groups of occupations grouped by *PF* and are based on the estimates of the model in the last column of Table 2. As noted above, wages in male-dominated occupations ($PF < 0.3$) are slightly lower than wages in 'integrated' occupations ($0.3 \leq PF < 0.6$). Wages in female-dominated occupations ($PF \geq 0.6$) are the lowest on average.

Given that the inclusion of industry indicators has such a marked effect on these estimates of the contribution of occupational differences to the gender wage gap, is it important to include them? Some researchers would argue that wage differences across industries reflect compensating wage differentials; that is, work in some industries is more distasteful (dirty, dangerous) than work in others. Mining and construction may be important examples. Thus, these differences should be accounted for if one is to obtain clean estimates of the role of occupational differences alone on the gender wage gap. Other researchers may argue that women are essentially kept out of male-dominated, high-paying occupations in certain industries. By removing wage differences across industries, an important source of wage differences across occupations is also being removed. These differences are important in understanding the gender wage gap and its sources. There thus appears to be no consensus on whether industry should be controlled for in such studies.

As the simple decompositions of Table 1 indicated, larger occupational contributions were evident when occupations were defined at more disaggregated levels. This finding carries over to the estimates of this section, where decompositions adjust for productivity differences across workers. Decomposition results with occupation defined at the disaggregated four-digit level are provided in Table 3. In all six estimated versions, the occupational contribution is estimated to be larger than when occupation is defined at the two-digit level. Even with individual occupational indicators and including industry indicators, a positive occupational contribution is estimated (second column). The occupational contribution is again higher when using *PF* measures (last four columns).

Including industry indicators has large effects on estimates of the occupational contribution provided in Table 3, particularly in the first two columns. The overlap between individual occupational indicators at the four-digit level and industry indicators is very large. As a consequence, the estimates of column 2 may not be particularly reliable. The estimates that employ the *PF* measures break this overlap between individual occupational indicators and the industry indicators to a large degree and thus may provide more reliable estimates of the occupational contribution if controlling for industry differences is deemed important.

Irrespective of whether or not industry controls are included, the decompositions that employ *PF* measures rather than individual occupational indicators consistently yielded a larger estimate of the occupational contribution to the gender wage gap. The discussion of Section 2 did not imply that such a finding was necessarily to be expected. Determining whether this finding is specific to this particular dataset or is more general (using other datasets at different time periods or for different countries) would require considerable further theoretical and empirical research.[22]

5. Conclusions

The main objective of this article was to ascertain whether occupational differences across genders contribute to the gender wage gap in Australia. The majority of prior Australian research on this issue concluded that occupational differences do not contribute to the gap, while two prior studies concluded quite the opposite. The overarching conclusion to be drawn from the results of the very detailed analysis conducted in this article is that occupational differences do indeed contribute to the gender wage gap. This finding, however, required defining occupations at an appropriately disaggregated level, rather than grouping occupations into a small number of categories as the prior literature had generally done.

My estimates of the proportion of the gender wage gap attributable to occupational differences, as defined at the disaggregated four-digit level, range from 0.05 to 0.51. The inclusion of controls for industry of employment had a large effect on the estimated occupational contribution, reducing it considerably. This effect is driven by the close overlap of occupational and industry categories. If industry is not controlled for, estimates of the occupational contribution range from 0.39 to 0.51, depending on estimation method employed. These are large estimates, irrespective of method of estimation (individual occupational indicators or occupational *PF* measures).

Although having more robust knowledge that occupational differences across genders do contribute to the gender wage gap in Australia is important, we are still a long way from an answer of what the correct policy response should be. Before policy-makers attempt to raise the wages of workers in low-paid, female-dominated occupations, we need to understand why wages are lower in those occupations. Is it due to occupational segregation and discrimination, to compensating wage differentials or to unobserved productivity differences?

Cobb-Clark and Tan (2011) investigate the role of personality and locus of control in determining occupational choice across genders in Australia. International research has investigated differences by gender in a number of job attributes, including the probability of death (DeLeire and Levy 2004; Grazier and Sloane 2008), earnings risk (Bonin et al. 2007) and job displacement risk (Dan 2010).[23] Macpherson and Hirsch (1995) investigate a number of occupational characteristics, including the prevalence of part-time work, hazards, noise, fumes, strength, whether the job requires outside work, plus job turnover rates and the time required to gain access to particular occupations. Coelli (2012) investigates the role of commuting time by occupation. Rosenbloom et al. (2008) investigate the role of preferences (or personality) in occupational choice differences across genders (particularly for information technology jobs). This research highlights many potential factors determining occupational differences and wages across men and women. More work is thus required to fully understand the role of occupation in the gender wage gap in Australia.

First version received March 2013; final version accepted July 2013 (Eds).

Appendix 1: Proofs of Equations (5) and (9)

In equation (5), the equality of $(\overline{X}^m_o - \overline{X}^f_o)\hat{\beta}^m_o$ and $\sum_{j=1}^{n}\left(P^m_j - P^f_j\right)\widetilde{W}^m_j$ was posited. The vectors $\overline{X}^m_o$ and $\overline{X}^f_o$ are means of occupational indicators and thus equal the proportion of a particular gender employed in each occupation; for example, $\overline{X}^m_j = P^m_j$. The elements of $\hat{\beta}^m_o$ are the coefficients on the occupational indicators for men. One occupational indicator must be dropped to avoid collinearity with the regression equation's constant term, so the elements of $\hat{\beta}^m_o$ are the differences in (productivity-adjusted) wages between each occupation and the dropped occupation, which will be denoted occupation 1, without loss of generality. Thus, one can write:

$$\left(\overline{X}^m_o - \overline{X}^f_o\right)\hat{\beta}^m_o = \sum_{j=2}^{J}\left(\overline{X}^m_j - \overline{X}^f_j\right)\hat{\beta}^m_j = \sum_{j=2}^{J}\left(P^m_j - P^f_j\right)\left(\widetilde{W}^m_j - \widetilde{W}^m_1\right)$$

It follows that:

$$\begin{aligned}\textstyle\sum_{j=2}^{J}&\left(P^m_j - P^f_j\right)\left(\widetilde{W}^m_j - \widetilde{W}^m_1\right)\\ &= \textstyle\sum_{j=2}^{J}\left(P^m_j - P^f_j\right)\widetilde{W}^m_j - \left(\sum_{j=2}^{J}P^m_j - \sum_{j=2}^{J}P^f_j\right)\widetilde{W}^m_1\\ &= \textstyle\sum_{j=2}^{J}\left(P^m_j - P^f_j\right)\widetilde{W}^m_j - \left(\left(1-P^m_1\right) - \left(1-P^f_1\right)\right)\widetilde{W}^m_1\\ &= \textstyle\sum_{j=1}^{J}\left(P^m_j - P^f_j\right)\widetilde{W}^m_j\end{aligned}$$

In equation (9), the equality of $(\overline{PF}^m - \overline{PF}^f)\,\hat{\beta}^m_{PF}$ and $\sum_{j=1}^{J}(P^m_j - P^f_j)\,\hat{W}^m_j$ was posited. The terms $\overline{PF}^m$ and $\overline{PF}^f$ are means of PF for men and women, respectively. Recall that PF takes the same value for each individual working in a specific occupation. Thus, the mean $\overline{PF}^m$ ($\overline{PF}^f$ similarly) can be constructed as follows:

$$\overline{PF}^m = \sum_{j=1}^{J}P^m_j \times PF_j$$

Thus, the following can be written, noting that $PF_j \times \hat{\beta}^m_{PF}$ are the predicted values for average (productivity-adjusted) wages by occupation ($\hat{W}^m_j$) from the regression:

$$\left(\overline{PF}^m - \overline{PF}^f\right)\hat{\beta}^m_{PF} = \sum_{j=1}^{J}\left(P^m_j - P^f_j\right)PF_j \times \hat{\beta}^m_{PF} = \sum_{j=1}^{j}\left(P^m_j - P^f_j\right)\hat{W}^m_j$$

Note that a constant can be added to these predicted values $\hat{W}^m_j$ to rescale them to the correct level without affecting the value of the term $\sum_{j=1}^{J}(P^m_j - P^f_j)\hat{W}^m_j$.

Table A1 Summary Statistics of Employees, the Household, Income and Labour Dynamics in Australia Survey, 2001–09

	Women		*Men*	
Variable	*Mean*	*Standard deviation*	*Mean*	*Standard deviation*
Real usual hourly wage	24.64	19.05	28.15	17.14
Log real usual hourly wage	3.11	0.42	3.23	0.47
Age	39.86	10.54	39.18	10.49
Married	0.55		0.55	
De facto	0.14		0.13	
Separated	0.04		0.02	
Divorced	0.08		0.04	
Widowed	0.01		0.00	
Never married/not de facto	0.19		0.26	
Major city	0.71		0.71	
Inner regional	0.19		0.19	
Outer regional	0.08		0.09	
Remote	0.01		0.01	
Immigrant: English-speaking	0.10		0.10	
Immigrant: Non-English-speaking	0.15		0.15	
Aboriginal	0.02		0.01	
Master's or PhD degree	0.04		0.05	
Graduate certificate or diploma	0.08		0.06	
Bachelor degree	0.21		0.16	
Diploma	0.11		0.09	
Certificate III/IV	0.14		0.28	
Certificate I/II	0.02		0.01	
Certificate not further defined	0.01		0.00	
Year 12	0.16		0.15	
Year 11 or below	0.24		0.20	
Work experience	17.81	10.07	20.18	11.16
Job tenure	6.32	6.96	7.22	8.11
Occupation tenure	8.56	8.85	9.60	9.38
Resident children: 0–4 years old	0.11		0.16	
Resident children: 5–14 years old	0.27		0.23	
Resident children: 15–24 years old	0.26		0.17	
Government employee	0.32		0.23	
Private for-profit employee	0.57		0.72	
Not-for-profit employee	0.09		0.04	
Other commercial employee	0.00		0.00	
Other non-commercial employee	0.01		0.01	
Part-time employee	0.43		0.10	
Union member	0.29		0.32	
Casual employee	0.23		0.14	
Workplace: less than 5 employed	0.10		0.09	
Workplace: 5–9 employed	0.20		0.11	
Workplace: 10–19 employed	0.09		0.09	
Workplace: 20–49 employed	0.11		0.11	
Workplace: 50–99 employed	0.13		0.14	
Workplace: 100–199 employed	0.19		0.18	
Workplace: 200–499 employed	0.14		0.12	
Workplace: 500 or more employed	0.10		0.12	
Agriculture, forestry and fishing	0.01		0.02	
Mining	0.00		0.03	
Manufacturing	0.06		0.17	
Electricity, gas, water and waste services	0.00		0.02	
Construction	0.01		0.08	

Continued

Table A1 Continued

	Women		Men	
Variable	*Mean*	*Standard deviation*	*Mean*	*Standard deviation*
Wholesale trade	0.03		0.05	
Retail trade	0.10		0.07	
Accommodation and food services	0.05		0.04	
Transport, postal and warehousing	0.02		0.07	
Information media and telecommunications	0.03		0.03	
Financial and insurance services	0.05		0.04	
Rental, hiring and real estate services	0.02		0.01	
Professional, scientific and technical services	0.07		0.07	
Administrative and support services	0.03		0.02	
Public administration and safety	0.06		0.10	
Education and training	0.18		0.06	
Health care and social assistance	0.23		0.05	
Arts and recreation services	0.01		0.02	
Other services	0.02		0.04	
Proportion female in occupation: one digit	0.57	0.15	0.40	0.21
Proportion female in occupation: two digits	0.64	0.21	0.33	0.24
Proportion female in occupation: three digits	0.66	0.22	0.31	0.25
Proportion female in occupation: four digits	0.70	0.23	0.28	0.26
Observations	23,061		23,077	

Notes: Sample includes employees aged 22–60 years old with valid reports of all characteristics. Statistics were constructed using cross-sectional population weights.

Table A2 Definition of Key Variables Employed in the Household, Income and Labour Dynamics in Australia Survey

Variable	*Definition*
Real usual hourly wage	Usual weekly gross wage or salary in main job divided by hours per week usually worked in main job, adjusted using September Consumer Price Index to reflect real wages in September 2009 levels
Age	Age last birthday at date of interview
Marital status	Self-reported registered marital status, de facto status etc. at date of interview
Immigrant	Country of birth other than Australia
	English-speaking—a main English-speaking country of birth: United Kingdom, New Zealand, Canada, United States, Ireland or South Africa
Education level	Highest education level achieved
Work experience	Years in paid work since leaving full-time education
Job tenure	Years with current employer
Occupation tenure	Years in current occupation
Resident children	If own children of each age group reside with individual
Sector of employment	Self-reported description of main employer from list provided
Part-time employee	Less than 35 hours per week usually worked in main job
Union member	If individual positively identifies as belonging to a trade union or employee association
Casual employee	Australian Bureau of Statistics' definition of no paid holiday leave or paid sick leave
Workplace size	Number of people employed at place of work (that is, employed people at specific work location, not all people employed by employer)
Industry of employment (one digit)	Current industry of main job, coded from text responses to the Australian and New Zealand Standard Industrial Classification (Australian Bureau of Statistics 2008)
Occupation of employment	Current occupation of main job, coded from text responses to the Australian and New Zealand Standard Classification of Occupations (Australian Bureau of Statistics 2006)

Table A3 Usual Hourly Earnings by Occupation, the Household, Income and Labour Dynamics in Australia (HILDA) Survey, 2001–09

	Occupation	*Female wage*[a]	*Male wage*[a]	*Proportion female*[b]
1	Managers	30.18	36.08	0.34
11	Chief executives, general managers and legislators	39.09	46.16	0.20
12	Farmers and farm managers	28.36	19.53	0.30
13	Specialist managers	33.94	39.31	0.31
14	Hospitality, retail and service managers	23.15	27.85	0.45
2	Professionals	31.18	35.94	0.53
21	Arts and media professionals	27.80	32.81	0.45
22	Business, human resource and marketing professionals	30.86	37.13	0.47
23	Design, engineering, science and transport professionals	29.55	35.97	0.27
24	Education professionals	30.63	32.28	0.69
25	Health professionals	32.94	41.25	0.75
26	Information and communications technology professionals	34.04	38.26	0.19
27	Legal, social and welfare professionals	29.50	31.06	0.56
3	Technicians and trades	22.36	28.00	0.15
31	Engineering, information and communications technology and science technicians	25.33	32.17	0.26
32	Automotive and engineering trades	22.62	27.95	0.01
33	Construction trades	15.03	27.51	0.02
34	Electrotechnology and telecommunications trades	25.93	29.94	0.02
35	Food trades	19.94	23.05	0.32
36	Skilled animal and horticultural	19.50	21.94	0.28
39	Other technicians and trades	22.00	25.42	0.41
4	Community and personal service	22.78	27.51	0.69
41	Health and welfare support	25.29	27.68	0.71
42	Carers and aides	20.95	23.42	0.89
43	Hospitality	23.49	23.25	0.70
44	Protective services	27.55	29.69	0.18
45	Sports and personal services	25.16	29.19	0.64
5	Clerical and administrative	24.87	27.85	0.77
51	Office managers and program administrators	26.74	34.31	0.73
52	Personal assistants and secretaries	25.88	33.79	0.98
53	General clerical	24.36	26.38	0.85
54	Enquiry clerks and receptionists	22.71	25.07	0.86
55	Numerical clerks	25.58	28.46	0.80
56	Clerical and office support	22.61	24.84	0.48
59	Other clerical and administrative	24.72	26.84	0.52
6	Sales	21.41	25.12	0.62
61	Sales representatives and agents	26.50	29.69	0.37
62	Sales assistants and salespersons	20.27	22.32	0.66
63	Sales support workers	20.96	24.30	0.75
7	Machinery operators and drivers	20.11	26.30	0.10
71	Machine and stationary plant operators	19.90	30.35	0.18
72	Mobile plant operators	26.51	25.59	0.03
73	Road and rail drivers	20.14	24.14	0.06
74	Storepersons	19.72	24.01	0.16
8	Labourers	20.09	22.87	0.36
81	Cleaners and laundry workers	19.44	20.49	0.62
82	Construction and mining labourers	20.61	26.96	0.02
83	Factory process workers	20.57	23.39	0.36
84	Farm, forestry and garden workers	18.97	19.70	0.25
85	Food preparation assistants	20.16	20.27	0.55
89	Other labourers	20.41	22.14	0.26
	Total	26.15	30.10	0.46

Notes: (a) Mean wages were constructed using employees aged 22–60 years old responding in waves 1–9 of the HILDA Survey, using cross-sectional population weights and with the effect of nominal wage growth removed (wages in 2009 levels). (b) Proportion female numbers were constructed using Census 2006 data on employment by gender and occupation.

Table A4 Log Real Hourly Wage Regressions, Excluding Occupational Indicators[a]

Variable	*Women* Coefficient	*Women* Standard error[b]	*Men* Coefficient	*Men* Standard error
Age	–0.0040	0.0065	–0.0116	0.0087
Age squared/100	0.0007	0.0076	0.0043	0.0104
Base—Single, never married				
Married	0.0808***[c]	0.0178	0.1208***	0.0200
De facto	0.0712***	0.0159	0.0976***	0.0184
Separated	0.0327	0.0222	0.0986***	0.0285
Divorced	0.0607***	0.0229	0.0451	0.0434
Widowed	0.1013***	0.0353	0.0505	0.0437
Base—Metropolitan				
Inner regional	0.0104	0.0463	–0.0326	0.0424
Outer regional	–0.0303	0.0473	–0.0991**	0.0423
Remote	–0.0524	0.0490	–0.1013**	0.0432
Base—Non-immigrant				
Immigrant: English-speaking	0.0343**	0.0150	0.0261	0.0167
Immigrant: Non-English-speaking	–0.0541***	0.0160	–0.0608***	0.0203
Aboriginal	0.0707***	0.0263	0.0528	0.0429
Base—Less than year 12				
Master's or PhD degree	0.3859***	0.0323	0.5008***	0.0355
Graduate certificate or diploma	0.3380***	0.0199	0.3986***	0.0317
Bachelor degree	0.2909***	0.0172	0.3705***	0.0275
Diploma	0.1795***	0.0182	0.2210***	0.0259
Certificate III/IV	0.0517***	0.0149	0.1230***	0.0177
Certificate I/II	–0.0220	0.0320	–0.1187	0.0786
Certificate not further defined	–0.0278	0.0553	–0.0826	0.0842
Year 12	0.0851***	0.0163	0.1254***	0.0251
Work experience	0.0196***	0.0036	0.0216***	0.0045
Work experience squared/100	–0.0371***	0.0079	–0.0233***	0.0087
Job tenure	0.0041**	0.0019	0.0013	0.0021
Job tenure squared/100	–0.0074	0.0070	–0.0003	0.0066
Occupation tenure	0.0063***	0.0014	0.0107***	0.0017
Occupation tenure squared/100	–0.0060	0.0043	–0.0214***	0.0051
Base—No resident children				
Resident children: 0–4 years	0.0316**	0.0127	0.0090	0.0122
Resident children: 5–14 years	–0.0207*	0.0104	0.0204	0.0126
Resident children: 15–24 years	0.0037	0.0109	0.0059	0.0151
Base—Private for-profit employee				
Government employee	0.0240	0.0150	–0.0194	0.0189
Not-for-profit employee	–0.0679***	0.0187	–0.1463***	0.0345
Other commercial employee	–0.0442	0.0385	0.0573	0.1339
Part-time employee	0.0304***	0.0096	–0.0205	0.0282
Union member	0.0259**	0.0108	0.0438***	0.0138
Casual employee	0.0012	0.0129	0.0512***	0.0171
Base—Workplace: less than 5 employed				
Workplace: 5–9 employed	0.0067	0.0191	0.0497***	0.0187
Workplace: 10–19 employed	0.0082	0.0184	0.0999***	0.0195
Workplace: 20–49 employed	0.0236	0.0178	0.1265***	0.0203
Workplace: 50–99 employed	0.0399**	0.0199	0.1542***	0.0217
Workplace: 100–199 employed	0.0332*	0.0201	0.1732***	0.0207
Workplace: 200–499 employed	0.0731***	0.0209	0.2095***	0.0210
Workplace: 500 or more employed	0.1347***	0.0201	0.2504***	0.0226
State indicators: 7	Included		Included	
Industry indicators: 18	Included		Included	
Time indicators: 8	Included		Included	
Observations	23,061		23,077	

Notes: (a) Estimates were based on employees aged 22–60 years old responding in waves 1–9 of the Household, Income and Labour Dynamics in Australia Survey. Population weights were employed during estimation.
(b) Standard errors were constructed to be robust to heteroscedasticity and were clustered at the individual level.
(c) *, ** and *** denote significance at the 0.10 per cent, 0.05 per cent and 0.01 per cent level, respectively.

Endnotes

1. This measure of 50 per cent was constructed using 2006 Australian Census data on all employed individuals with occupations defined at the disaggregated four-digit level (474 individual occupational groups) according to the Australian and New Zealand Standard Classification of Occupations (ANZSCO). See Australian Bureau of Statistics (2006) for details.

2. Austen, Jefferson and Preston (2013) make this argument and the Australian Bureau of Statistics (2005, pp. 153–4) also makes this point using a small amount of supporting evidence.

3. Jones (1983) included elements of both types of analyses: separate occupational indicators for six aggregated occupational groups, plus indicators representing 'proportion female' in disaggregated occupations. The results were consistent with the findings of both strands of the literature. Differences in aggregated occupational groups across genders were estimated to detract from the gender income gap, while differences in the proportion female in disaggregated occupations were estimated to contribute to the gap.

4. Using male or female coefficients to construct the 'explained' contribution, or some linear combination of the two, is valid. Several researchers have stated a preference for using male returns, as it is an intuitive way of thinking about potential discrimination.

5. Kee (2006) and Barón and Cobb-Clark (2010) decompose differences in the entire wage distributions of men and women, rather than just the mean wage gap, and find that adding occupational indicators to the models does not increase the explained proportion of differences in the wage distributions. They thus conclude that occupational differences do not contribute to the gender wage gap.

6. Managerial employees were excluded from Rimmer's (1991) analysis.

7. Owner-managers of incorporated businesses are also excluded from the estimation sample, as their wages may not purely reflect payment for labour services.

8. The employment and wage outcomes of young people are in many cases affected by decisions regarding study and training, plus the payment of youth wages (those aged under 21 years are often paid a percentage of the standard adult wage).

9. An exception is estimates of the relationship between part-time employment status (usual weekly work hours are less than 35) and hourly earnings. Errors in reporting usual hours of work will bias up estimates of the coefficient on part-time status in hourly earnings regressions.

10. Occupational groups were defined according to ANZSCO (see Australian Bureau of Statistics 2006).

11. If women and men were distributed across occupations in the same manner, mean *PF* values for men and women would equal the overall proportion of women in the workforce.

12. Access to the In Confidence version of the HILDA Survey data was required to obtain information on occupations disaggregated at the three- and four-digit levels.

13. As a robustness check, all estimates were reconstructed without using population weights. While some estimates changed slightly, the conclusions of the article are very much robust to the decision to weight. These unweighted estimates are available upon request.

14. The decompositions in Table 1 use male average wages by occupation. If female wages are used, the estimated occupational contributions to the wage gap are smaller, but still have the pattern of increasing and positive contributions as occupations are defined more finely.

15. An identified occupation is one in which both men and women are observed in the estimation sample. Both genders must be observed for an occupation to be employed in the decomposition, as a measure of average wages in the occupation for both genders is required. Note that at the four-digit level, 326 of potentially 474 occupations were identified here.

16. Employment weights and probabilities by occupation were calculated using data from the full 2006 Census to construct these decompositions.

17. This estimate is similar in magnitude to the estimate of Table 1 at the two-digit level.

18. This is due to women being under-represented in well-paying industries, such as mining and construction, and over-represented in lower paying industries (relative to the education levels of workers), such as education, health care and social assistance.

19. The estimated coefficients on the *PF* term in the log wage regressions were –0.0633 (standard error, SE of 0.0242) and –0.1457 (SE of 0.0225) for women and men, respectively, in this case.

20. The estimated coefficients of the *PF* term in the log wage regressions were –0.0704 (SE of 0.0259) and –0.0462 (SE of 0.0261) for women and men, respectively, in this case.

21. Wooden (1999) found a marked negative effect on the occupational contribution from including industry indicators when using male returns in the decomposition, but also a marked contrary positive effect when using female returns. He thus reports only a small effect when averaged over both genders. I report results using male returns only in Table 2, but note that a contrary positive effect was not found when using female returns.

22. A brief analysis of the dataset used by Miller (1994), however, did show a similar result; that is, if individual occupational indicators were used, the estimated occupational contribution was much smaller than when occupational measures based on *PF* were used.

23. Wilkins and Wooden (2013) also show that job displacement probabilities are higher for men than for women in Australia and that these differences are related to differences in occupation and industry of employment.

References

Austen, S., Jefferson, T. and Preston, A. 2013, 'Contrasting economic analyses of equal remuneration: The Social and Community Services (SACS) case', *Journal of Industrial Relations*, vol. 55, pp. 60–79.

Australian Bureau of Statistics 2005, *Australian Social Trends 2005*, Cat. no. 4102.0, ABS, Canberra.

Australian Bureau of Statistics 2006, *ANZSCO – Australian and New Zealand Standard Classification of Occupations, First Edition, 2006*, Cat. no. 1220.0, ABS, Canberra.

Australian Bureau of Statistics 2008, *Australian and New Zealand Standard Industrial Classification (ANZSIC) 2006 (Revision 1.0)*, Cat. no. 1292.0, ABS, Canberra.

Australian Bureau of Statistics 2009, *Employee Earnings and Hours, Australia, August 2008*, Cat. no. 6306.0, ABS, Canberra.

Barón, J. and Cobb-Clark, D. 2010, 'Occupational segregation and the gender wage gap in private- and public-sector employment: A distributional analysis', *Economic Record*, vol. 86, pp. 227–46.

Blinder, A. 1973, 'Wage discrimination: Reduced form and structural estimates', *Journal of Human Resources*, vol. 8, pp. 436–55.

Bonin, H., Dohmen, T., Falk, A., Huffman, D. and Sunde, U. 2007, 'Cross-sectional earnings risk and occupational sorting: The role of risk attitudes', *Labour Economics*, vol. 14, pp. 926–37.

Brown, R., Moon, M. and Zoloth, B. 1980, 'Incorporating occupational attainment in studies of male–female earnings differentials', *Journal of Human Resources*, vol. 15, pp. 3–28.

Chapman, B. and Mulvey, C. 1986, 'An analysis of the origins of sex differences in Australian wages', *Journal of Industrial Relations*, vol. 28, pp. 504–20.

Cobb-Clark, D. and Tan, M. 2011, 'Non-cognitive skills, occupational attainment and relative wages', *Labour Economics*, vol. 18, pp. 1–14.

Coelli, M. 2012, 'Family friendly occupations and the US gender wage gap', unpublished paper, Department of Economics, University of Melbourne.

Dan, I. 2010, 'Gender based self-selection into industries and occupations: Evidence from displaced workers', unpublished paper, Department of Economics, University of Toronto.

DeLeire, T. and Levy, H. 2004, 'Worker sorting and the risk of death on the job', *Journal of Labor Economics*, vol. 22, pp. 925–53.

Duncan, O. D. and Duncan, B. 1955, 'A methodological analysis of segregation indexes', *American Sociological Review*, vol. 20, pp. 210–17.

Goldberger, A. 1981, 'Linear regression after selection', *Journal of Econometrics*, vol. 15, pp. 357–66.

Grazier, S. and Sloane, P. 2008, 'Accident risk, gender, family status and occupational choice in the UK', *Labour Economics*, vol. 15, pp. 938–57.

Groshen, E. 1991, 'The structure of the female/male wage differential: Is it who you are, what you do, or where you work?', *Journal of Human Resources*, vol. 26, pp. 457–72.

Jones, F. L. 1983, 'Sources of gender inequality in income: What the Australian Census says', *Social Forces*, vol. 62, pp. 134–52.

Kee, H. J. 2006, 'Glass ceiling or sticky floor? Exploring the Australian gender pay gap', *Economic Record*, vol. 82, pp. 408–27.

Kidd, M. 1993, 'Sex discrimination and occupational segregation in the Australian labour market', *Economic Record*, vol. 69, pp. 44–55.

Kidd, M. and Meng, X. 1997, 'Trends in the Australian gender wage differential over the 1980s: Some evidence on the effectiveness of legislative reform', *Australian Economic Review*, vol. 30, pp. 31–44.

Lee, Y. L. and Miller, P. W. 2004, 'Occupational segregation on the basis of gender: The role of entry level jobs', *Australian Journal of Labour Economics*, vol. 7, pp. 355–74.

Macpherson, D. A. and Hirsch, B. T. 1995, 'Wages and gender composition: Why do women's jobs pay less?', *Journal of Labour Economics*, vol. 13, pp. 426–71.

Miller, P. 1994, 'Occupational segregation and wages in Australia', *Economics Letters*, vol. 45, pp. 367–71.

Oaxaca, R. 1973, 'Male–female wage differentials in urban labor markets', *International Economic Review*, vol. 14, pp. 693–709.

Rimmer, S. M. 1991, 'Occupational segregation, earnings differentials and status among Australian workers', *Economic Record*, vol. 67, pp. 205–16.

Rosenbloom, J., Ash, R., Dupont, B. and Coder, L. 2008, 'Why are there so few women in information technology? Assessing the role of personality in career choices', *Journal of Economic Psychology*, vol. 29, pp. 543–54.

Sorensen, E. 1990, 'The crowding hypothesis and comparable worth', *Journal of Human Resources*, vol. 25, pp. 55–89.

Treiman, D. J. and Hartman, H. I. (eds) 1981, *Women, Work and Wages: Equal Pay for Jobs of Equal Value*, National Academy Press, Washington, DC.

Watson, N. and Summerfield, M. 2009, 'Quality of the occupation and industry coding in the HILDA Survey', HILDA Project Discussion Paper no. 3/09, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Watson, N. and Wooden, M. 2010, 'The HILDA Survey: Progress and future developments', *Australian Economic Review*, vol. 43, pp. 326–36.

Wilkins, R. and Wooden, M. 2013, 'Gender differences in involuntary job loss: Why are men more likely to lose their jobs?', *Industrial Relations*, vol. 52, pp. 582–608.

Wooden, M. 1999, 'Gender pay equity and comparable worth in Australia: A reassessment', *Australian Economic Review*, vol. 32, pp. 157–71.

Wooden, M., Wilkins, R. and McGuinness, S. 2007, 'Minimum wages and the "working poor"', *Economic Papers*, vol. 26, pp. 295–307.