School of Economics
18 September 2023

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

Modelling Public Wages Expenses Across States and Time Using Survey Data

Prepared by
Dr. Christiern Rose, Dr. Lizi Yu, and Professor Alicia Rambaldi
School of Economics, University of Queensland

Contents

Tables

Figures

Background

1. The model used in the Commonwealth Grants Commission's (referred to as "CGC" hereafter) 2025 Methodology Review – Wage Costs Consultation Paper (referred to as "*2025 Review*" hereafter) aims to estimate state-level geographic effects on wages, using an econometric model focusing on private sector workers. Using data from the Characteristics of Employment survey (referred to as "*COE survey*" hereafter), the model estimates the relationship between the logarithm of wage and the state of employment, controlling for non-geographic explanatory variables (e.g., industry, occupation, education).
2. The geographic effects estimated from the above method are then used to assess the non-policy driven, geographic cost pressure differences across states and territories (referred to as "states" hereafter) in public-sector wages. Using effects coding, the model estimates a wage cost factor for each state that measures the percentage difference from the national average wage level.
3. In the *2025 Review*, the CGC has proposed several empirical changes intended to improve the current estimation and has encouraged the states to provide comments and opinions on six questions. The CGC has also engaged Professor Alison Preston as an independent reviewer to provide advice on the proposed methodology in the Wage Costs Consultant Report (referred to as "*Reviewer Report*" hereafter). The remainder of this report addresses responses to the six questions in the *2025 Review* (Section 1) and comments on the eleven recommendations in the *Reviewer Report* (Section 2).

Section 1: Response to Issues Raised in the 2025 Review

Question 1 (Conceptual basis)

Do states agree that the underlying conceptual basis for the wage costs assessment remains sound?

4. Using private-sector wages to estimate geographic differences in public-sector wages is based on two major conceptual foundations:
   a. Labour Market Competition and Mobility of Workers: The skills and qualifications required for many jobs can be similar across the private and public sectors. Workers are generally mobile and can move between sectors based on the best opportunities available, thus the private and public sectors compete for workers in the same labour markets. If private-sector wages are higher in a particular geographic area, the public sector may need to offer competitive wages to attract and retain skilled workers. Conversely, in areas with lower private-sector wages, public sector wages might also be comparatively lower.
   b. Cost of Living: Differences in private-sector wages across states often reflect variations in the cost of living, as higher cost of living might translate to higher wages because employees need more compensation to afford basic necessities. If the private sector adjusts wages based on these factors, the public sector would also need to do so to retain workers.

5. However, using private-sector wages to assess public-sector wages is plausibly hindered by an empirical challenge, which may become more pronounced during the economic downtowns, especially in the aftermath of COVID-19.

   Wage Compression and Wage Premiums in the Public Sector: Studies have documented the existence of wage compression[1] and wage premiums[2] in the public sector. Wage compression refers to the smaller difference between the highest and lowest wages in the public sector, compared to the private sector, making it more difficult to retain high-ability and experienced workers. On the other hand, wage premiums refer to additional compensation factors, such as better benefit packages and better job security, to retain workers in the public sector.

   Consistent with the conceptual foundation illustrated in bullet point 4a: Over the longer term, the wage compression and wage premiums effects can possibly lead to an equilibrium distribution of workers in the public and private sectors, as the benefits of moving to one sector can be offset by the disadvantages. However, cyclical forces can potentially lead to unbalanced states in the short run. Economic booms might bolster private sector wages and opportunities, drawing workers away from the public sector. Conversely, recessions can make public sector jobs with their wage premiums more appealing.

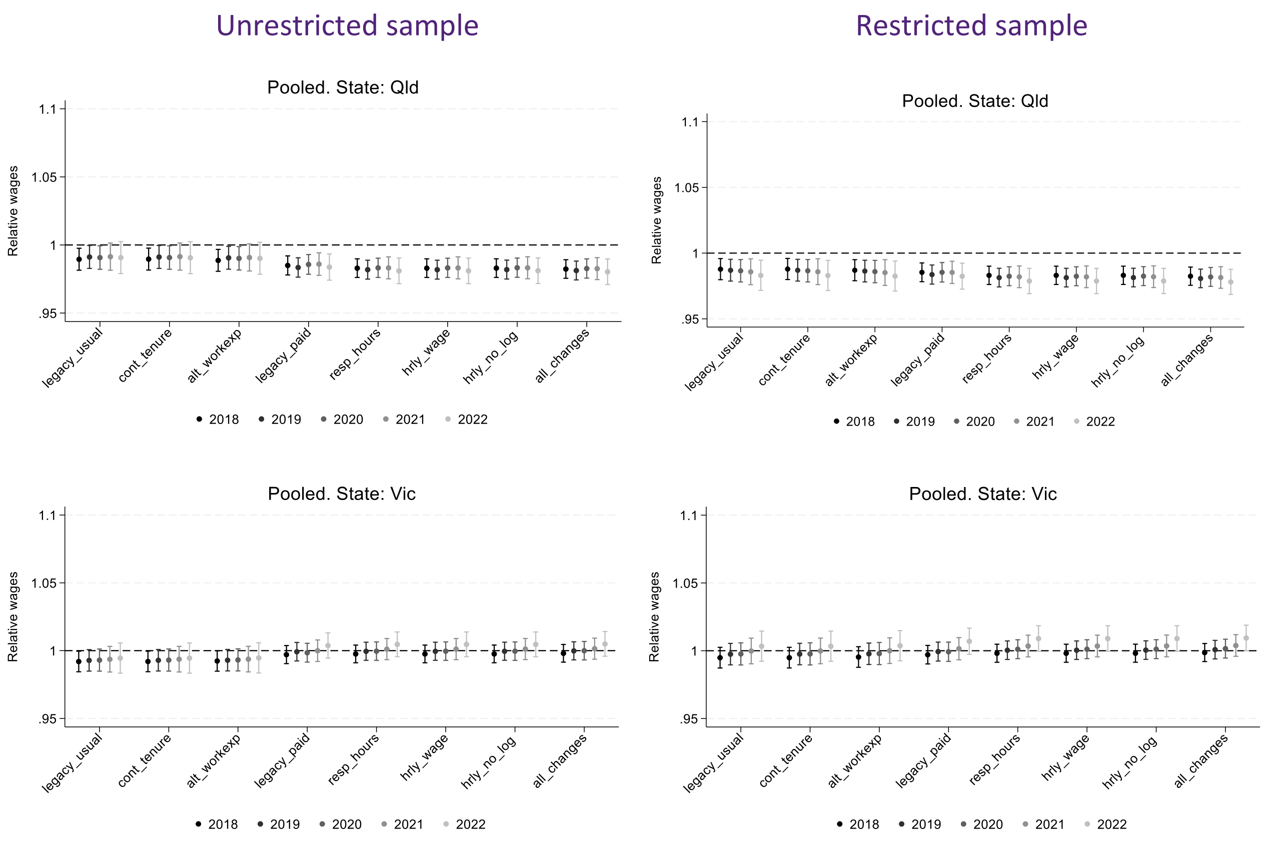
   During the economic downtowns following the COVID-19 pandemic, one may be concerned that the surveyed private-sector workers may not serve as a good counterfactual group for public-sector workers. Over time, however, wage compression and wage premiums effects, in conjunction with cyclical forces, can help maintain equilibrium. Thus, this empirical challenge may be less concerning if we estimate the impact by pooling data over a longer-term period, as we recommend in bullet points 9-12. In addition, as various industries have been affected by the COVID-19 pandemic in different ways, we also recommend excluding some heavily affected industries (such as hospitality and tourism) during pandemic years to address the impact during the COVID-19 pandemic, as illustrated in bullet points 6-7.

6. The CGC have placed restrictions on the sample to attempt to control for lockdown effects. Specifically, they exclude workers who declare themselves to currently be working fewer hours than usual. Since workers most impacted by lockdowns tend to work in industries with low wages (e.g., hospitality) relative to other private sector workers (e.g., those able to work from home), dropping these individuals from the sample increases the average wage of workers included in the sample. Hence, in states that experienced lockdowns, the sample is one of higher wage workers, whereas in states which did not experience lockdowns, the sample is one of all workers. This makes an accurate comparison of wages between states difficult.
7. Figure 1 illustrates this finding for Queensland and Victoria using the pooled sample (a five-year rolling window, discussed in bullet point 9). Queensland experienced relatively few days of lockdown, whereas Victoria experienced a greater number. The left panel shows results using the unrestricted sample. The right panel uses the restricted sample. We observe that the relative wages for Queensland become lower when the sample restriction is imposed, whilst the relative wages in Victoria become higher.

[1] For example, see Borjas (2002), Katz and Krueger (1991).
[2] For example, see Bertola (1990), Freemand and Ichiniowski (2007), and Gomes (2015).

To avoid this issue, we propose to drop workers in all industries strongly affected by lockdowns (e.g., hospitality) in the pandemic years from the sample, irrespective of whether they actually experienced a lockdown. Given that smaller states rely on smaller sample, we consider how excluding the most affected industries (i.e., Accommodation and Food Services, and Arts and Recreation Services) may reduce the sample size. We report the percentage of sample being excluded for each state in the Appendix Table 1, if only the current year of data is used or a pooling sample with 3-year or 5-year rolling window is used (see bullet point 9 below for a discussion on pooling). Overall, the reduction in sample size is relatively modest, especially if we pool the data. We conclude that it is beneficial to exclude industries impacted by lockdowns during the pandemic years, leading to a slightly smaller but unbiased sample for the regression analysis. Alternatively, as the impact of the pandemic recedes, it may be appropriate to return to using the unrestricted sample.

Figure 1: Coefficient estimates using unrestricted versus restricted sample (five-year rolling window)



8. Overall, we agree that the fundamental conceptual foundations for wage cost assessments remain robust and valid, although the COVID-19 pandemic and associated lockdowns have introduced volatility and influenced some aspects of the labour market. To address the concerns relating to the COVID-19 and the associated economic downtowns, we recommend:

   a. Dropping some industries to account to address the lockdown impact (as in bullets point 6-7);

   b. Pooling data over a longer-term period (as in bullets points 9-12);

   c. Discount the coefficient estimates to reflect the extent of uncertainty (as in bullet points 19-23).

Question 2 (Combination of estimates across years)

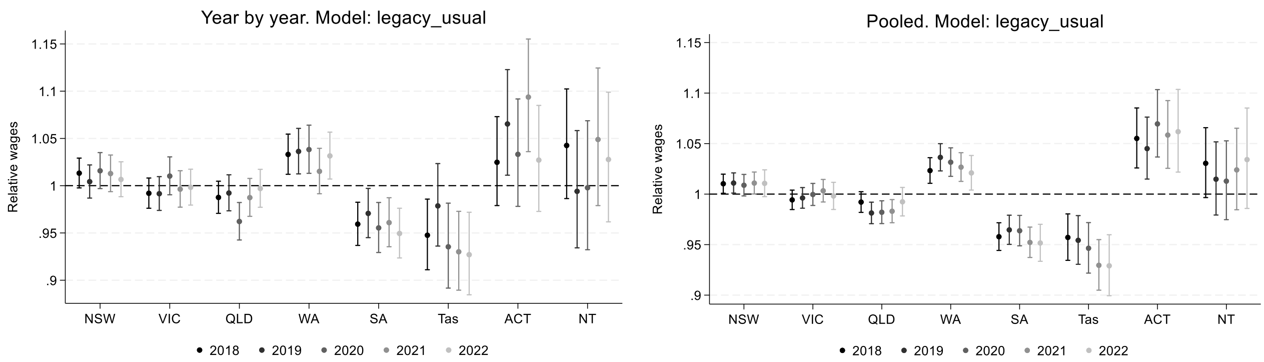
Do states agree with the proposed approach to combine estimates of relative differences in states wages across years?

9. We have compared the results using only one year of data with a pooled sample comprising a rolling window of three or five years centred on the year of interest (e.g., for 2020, use the sample from 2019-2021 inclusive for a three-year window, and 2018-2022 for a five year window). When using pooled data, we include dummies for each year in the regression. These account for wage inflation common to all states. As expected, we find that pooling leads to a substantial decrease in the variability over time.

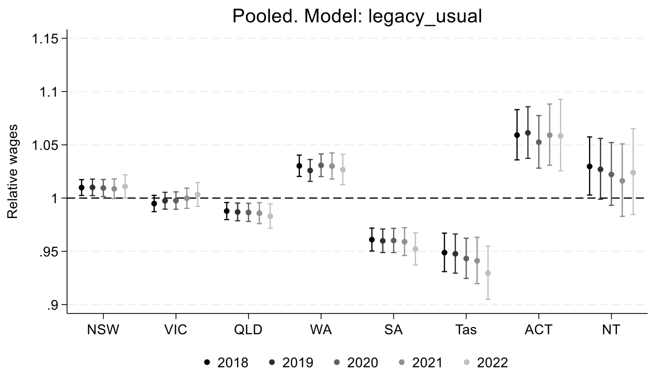
   Figure 2 plots the relative wages (including the 12.5% discount used by the CGC) and 0.95 confidence intervals using the regression model specification and sample definition currently used by the CGC, which is denoted by "*legacy_usual*" (see Table 1 below). The top left panel uses the unpooled sample. The top right panel uses a three-year rolling window. The bottom panel uses a five-year rolling window. As expected, as the width of the window increases, there is considerably less variability between years, which is also reflected by the narrower 0.95 confidence intervals. The appropriate width of the rolling window ultimately depends on the speed at which the labour market evolves. However, given the small sample size involved, it is our view that either a three- or five-year rolling window is more appropriate than a year by year specification.

Figure 2: Coefficient estimates using unpooled versus pooled sample

Year by year | Three year rolling window



Five year rolling window



10. In the *2025 Review*, the CGC is proposing an alternative approach intended to improve accuracy and reduce volatility. The CGC's proposed approach involves performing each regression one year at a time (i.e., on the unpooled samples), and then taking a weighted average of the state effects from each sample. They propose a formula for this which accounts for (a) the distance to the year of interest (e.g., for the 2020 estimates, place most weight on 2020, less on 2019 and 2021, and even less on 2018 and 2022); and (b) the variance of the estimate (higher variance implies lower weight).

    We do not support this approach because the formula could be viewed to be somewhat arbitrary, especially the way in which future and past years are used to construct the weighted average. It is less clear how much weight should be given to years closer to the year of interest, and how much should be given to the years further away. Relatedly, it is also not clear how to reliably calculate standard errors for this proposed approach. Compared to the pooling approach, this proposed approach also has less power because each year's estimate is based on a smaller sample. In contrast, pooling can increase statistical power to provide a more reliable estimate, improving accuracy and reducing volatility as intended.

    Due to the above, we do not recommend the use of the CGC's proposed approach to combine estimates across years; and we view our rolling window sample approach to be somewhat less arbitrary and considerably more transparent. In a recently published paper in the *Journal of Econometrics*, Cai and Juhl (2023) study rolling regressions and tabulate critical values for inference using this method. The motivation for researchers and the journal to publish this work is that rolling regressions are often employed to characterise changing economic relationships over time as they are an intuitive shortcut to more technically demanding time varying parameter models. Their popularity is evidenced by the number of routines written in languages or packages such as in R, STATA, Matlab, and Excel.

11. In the 2025 Review, the CGC proposes several changes to the independent and dependent variables used in the regression model. We replicate the regression models considered in the R files that we received from the CGC, which are described in Table 1. Full regression results for all models are provided in Appendix Table 3.

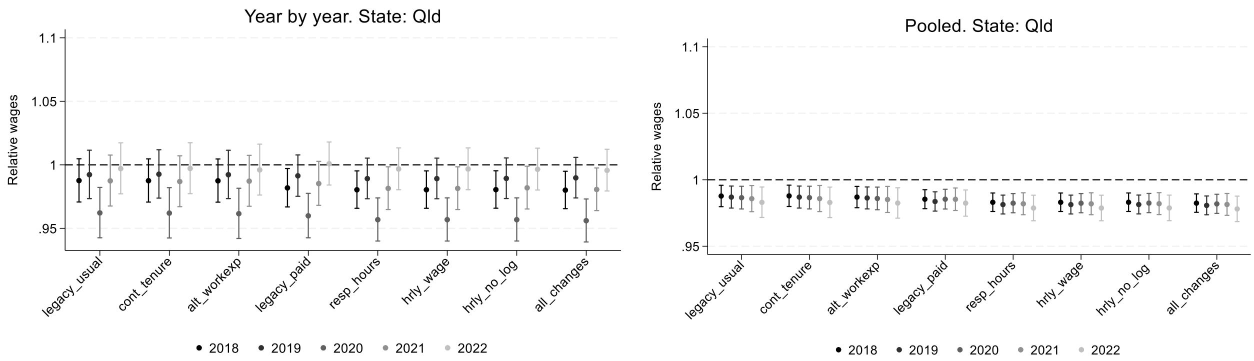
Table 1: Summary of Regression Models

Model name	Description	*2025 Review* question
legacy_usual	The model that has been used to date	-
cont_tenure	Change the measure of tenure to be continuous	5
alt_workexp	Change the measure of work experience to use interactions of age with levels of education	4
legacy_paid	Replace usual hours worked with paid hours worked	3
resp_hours	Use both usual and paid hours, including their interaction and log of paid hours	3
hrly_wage	resp_hours but also changing dependent variable from weekly to hourly wages	3
hrly_no_log	hrly_wage but omitting log of paid hours	3
all_changes	All changes proposed in the *2025 Review* (hrly_no_log but with continuous-tenure and alternative measure of work experience)	

12. Figure 3 displays the results for the Queensland coefficient over different regression models. The left panel uses the unpooled sample. The right panel uses the pooled sample over a five-year rolling window. The proposed changes to the measure of tenure (*cont_tenure*) and work experience (*alt_workexp*) make little difference. The proposed changes to the measures of hours worked (*legacy_paid, resp_hours*) and dependent variable (*hrly_wage, hrly_no_log*) reduce the estimated wage level in Queensland relative to other states.

    Comparing *legacy_usual* with *all_changes*, the overall impact of the changes proposed in the consultation paper is a reduction in the estimated wage level in Queensland relative to other states. This reduction is driven by the modelling decisions concerning usual versus paid hours and the choice of dependent variable (weekly versus hourly wages) addressed in *2025 Review* question 3.

Figure 3: Queensland coefficient estimates from each regression model



Question 3 (Use of hourly wage)

Q3 Do states agree the Commission should:

- **use hourly wages rather than weekly wages as the dependent variable?**
- **include both usual hours worked and paid hours as explanatory variables including as non-linear and interacting terms?**

13. The proposed change to a model predicting log of hourly wages, instead of log of weekly wages, is intended to capture the direct effect of hours worked on earnings (referenced in points 50-51 in the *2025 Review*). In the *COE survey*, hourly wage is calculated by taking the ratio of weekly wage to paid hours. The proposed change in *2025 Review* suggests using the log of hourly wage as the dependent variable, while paid hours, paid hours squared, and an interaction term with usual hours appear as explanatory variables entering the regression. However, this change may introduce significant empirical challenges.

    Our main concern is the mechanical relationship caused by this structure. Since the dependent variable is directly defined from the ratio of weekly wage to paid hours, any change in paid hours inherently changes the dependent variable. This can lead to endogeneity, where the predictor (paid hours) is highly correlated with the outcome (log of hourly wage) because of the way the outcome is constructed. This change potentially leads to severe bias in the estimation, making it challenging to form an intuitive interpretation of the magnitude and direction of the coefficient estimate. If paid hours increase, then the dependent variable, log of hourly wage, will inherently *decline*. However, empirically, log of hourly wage is expected to *increase* with paid hours because jobs with more paid hours (e.g., more annual leave and less unpaid overtime shifts) are likely better jobs offering higher hourly wage. This issue has also been discussed in the *Reviewer Report* on page 21.
14. Even if paid hours is not added on the right-hand-side as an explanatory variable, we still think using hourly wage as dependent variable may not be an appropriate change for three primary reasons:
    a. Given that the proposed dependent variable, hourly wage, is defined as the ratio of weekly wage to paid hours, it is susceptible to the risk of spurious correlations for the ratio problem. This concern has been extensively discussed in the literature, notably by Kronmal (1993).[3] Specifically, when both the numerator (weekly wage) and the denominator (paid hours) of a ratio share a common correlated variable—such as gender or age—this can create a heavily biased estimate. This situation can be quite common with many variables, which can cause severe bias in coefficient estimation.
    b. The proposed new dependent variable (log hourly wage) largely reduces the R-squared from roughly 0.62 to 0.42, as presented in columns 1 and 6 of Table C1 in the *2025 Review*. A substantial reduction in R-squared indicates that the explanatory variables used in the model explain a much smaller proportion of the variation in the log hourly wage, suggesting that the model fits the data for log hourly wage less well than for log weekly wage. The notable reduction in R-squared is addressed in point 101 of the *2025 Review*. The primary justification for using the log hourly wage is that the main

[3] For recent studies, see Clemens and Hunt (2019) and Bartlett and Partnoy (2020).

coefficients remain unchanged, ignoring that the coefficient estimate of log hours has significantly changed (from 0.99 to -0.01) and the interpretation of this coefficient becomes unintuitive.

c. The hourly wage is generally more relevant for analysing sectors or occupations characterized by significant variations in work hours, which is especially true in contexts where part-time positions and irregular work hours are common. The underlying rationale of this assessment – to use wages of *comparable* private-sector workers as a proxy for estimating public-sector wages – is more aligned with the emphasis of the "usual wage" approach, as in the 2020 Legacy Model.

15. The proposed modification to include both usual hours and paid hours as explanatory variables is aimed to maximize the explanatory power of the wage costs assessment (referenced in point 53 in the *2025 Review*). Yet, given the strong correlation between usual hours and paid hours, this adjustment might introduce a major risk of multicollinearity. Though multicollinearity among such control variables is not likely to bias the coefficients of interest, it implies that there is little to be gained in terms of model fit by including both usual and paid hours.

    a. As presented in columns 5-8 of Table C1 in the *2025 Review*, estimated coefficients of paid hours and usual hours are -0.02 and +0.02, respectively. When combined, their values essentially cancel out, summing to zero. This indicates a high degree of collinearity between these two variables. Including both usual hours and paid hours as well as their non-linear and interacting terms will heavily bias the estimation the coefficients on usual and paid hours. This bias problem has also been discussed in the *Reviewer Report* on page 21.

    b. The coefficient of paid hours is also difficult to interpret. The estimated coefficient (-0.02) implies that one additional unit of paid hours would reduce hourly wage by two percent, holding usual hours fixed. This is quite implausible because higher paid hours are typically indicative of greater benefits such as paid overtime, holidays and sick leave. This suggests that the negative coefficient estimate is likely caused by the mechanical collinearity issue, as mentioned in points 13 and 14a.

16. Another option might be to use paid hours to replace usual hours in the regression, but this may not improve the model as intended. Paid hours may vary significantly within an individual over time. Even for workers with regular schedules, paid hours can fluctuate due to seasonal overtime shifts, sick leave or other factors. As the survey is conducted in the August of each year, this bias can be worse if some industries consistently work more or fewer hours during the survey time.

    In addition, "paid hours" is believed to be a variable with more measurement issues, as survey respondents typically report "paid hours" with less precision than "usual hours". Thus, even only using paid hours and its quadratic term in the regression may introduce measurement error.

17. In light of points 13-16, we consider the 2020 Legacy Model to have an appropriate specification for the dependent variable and treatment of hours worked.

Question 4 (New measure of work experience)

Do states agree the Commission should replace the derived work experience variable with interacting variables of age and level of education?

17. To improve the explanatory power of the model, the CGC proposes to use interacting variables of age and levels of education to replace the derived work experience variable. In particular, the proposed specification is “to include a categorical variable for education level (unchanged), work experience as age minus 15, and all two-way interactions between education level, work experience (linear and quadric), and sex”. This can conceptually address the differences such as the wage effect on a 25-year-old worker with a graduate diploma is different from the effect on a 50-year-old worker with a graduate diploma. We think the estimated coefficients are consistent with the 2020 Legacy Model and this change can potentially improve the model.

Question 5 (Job tenure as a continuous variable)

Do states agree the Commission should treat job tenure as a continuous variable?

18. The CGC proposes to treat job tenure as a continuous variable. The justification is that the empirical model has fewer regressors than one that uses fixed effects and the relationship between wage and tenure is believed to be linear (referenced as points 60-61 in the *2025 Review*). We do not recommend this change and we think the CGC should still treat job tenure as a discrete dummy variable, as in the 2020 Legacy Model.
    a. Consistent with the CGC’s view, we think it is important to include job tenure in the regression, even when experience and its quadratic term have already been included. Experience captures the breadth of knowledge and skills someone have acquired over time that contribute to the human capital effect. Job tenure is the length of time one has been in their current job position or with their current employer; reflecting familiarity with specific job tasks, and especially the understanding of firm-specific knowledge or procedure. Building on the classical works on human capital theory[4], there is a distinction made between *general* human capital and *firm-specific* human capital.
    b. Firm-specific job tenure is seen as a proxy for firm-specific human capital since longer tenure is associated with greater accumulation of knowledge and skills that are uniquely valuable within a given firm. It is observed that wages often rise with job tenure, reflecting the acquisition of firm-specific human capital. Workers with longer tenures might receive higher wages because they have accumulated more firm-specific knowledge, making them more valuable to the company.
    c. The impact of firm-specific job tenure, however, is likely non-linear. Early years in a job might come with rapid wage increases due to promotions or skill acquisition and training. After a certain point, additional years might not lead to as significant wage increases. Many studies[5] in the literature find diminishing returns of job tenure on worker wage. While general skills are transferable to other employers, firm-specific skills are not. Thus, longer job tenure can be associated with the accumulation of more firm-specific human capital but less bargaining power in the job market, and consequently lower wage increases.

[4] Becker (1964); Beaudry and Ninardo (1991), Baker, Gibbs and Holmstrom (1994),
[5] Lazear (2009), Stole and Zwiebel (1996).

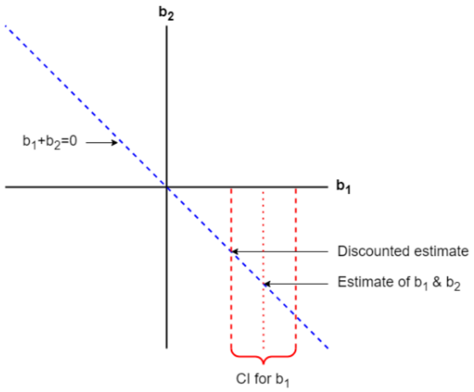
d. Therefore, categorizing tenure can help capture its nonlinear impact on workers' wage, as opposed to treating it as a continuous variable, which may not adequately reflect the nuances of firm-specific job tenure. While including polynomial functions of job tenure could address this issue, this approach may not offer any additional benefit, compared to categorizing tenure, in terms of a simpler model with fewer coefficients to be estimated. Thus, the potential benefits referenced in bullets points 61-62 in the *2025 Review* might not be achieved by treating job tenure as a continuous variable.

Question 6 (Discounting)

Do states agree that a 12.5% discount remains appropriate?

19. When calculating relative wages, the CGC applies a 12.5% discount to reflect the uncertainty in the estimates of the state coefficients. This discounting rule essentially takes the relative wages in each state implied by the regression estimates and shifts them towards parity by 12.5%. States with high relative wages have their relative wages revised downwards. States with low relative wages have their relative wages revised upwards. The principle of discounting to reflect uncertainty is reasonable. However, the approach that is currently used does not reflect the *extent* of the uncertainty. Below we propose an alternative which does this explicitly and compare it to the 12.5% rule.
20. For ease of exposition, suppose that there are just two states. In that case, the regression model would omit one state from the list of regressors (the reference state) and estimate the coefficient for the other state. Suppose that state 2 is the reference state and let the coefficient on state 1 be *b1*. Under the effects coding approach employed by the CGC, the coefficient for state 2 is then *b2=-b1*, because the coefficients of all states must sum to zero (i.e., we know that *b1+b2=0*). This is because the coefficients are interpreted as deviations from the "grand mean", (*b1+b2*)/2, and deviations from the mean must sum to zero.
21. Figure 4 illustrates our proposed approach graphically. From the regression model, we obtain an estimate of *b1* (dashed red line) and obtain an estimate *b2* using the dashed blue line (the feasible values of *b1* and *b2*). The principle behind discounting suggests that to correct for uncertainty, we ought to revise this estimate towards zero. That is, we should find a point on the dashed blue line that is closer to the origin than the estimated point. We would like to select such a point explicitly based on the level of uncertainty. The level of uncertainty is fully described by a confidence interval around the estimate of *b1*, that we can obtain from the regression. We thus propose to take the point where the lower bound of the confidence interval intersects the dashed blue line. Formally, we search for the relative wages that are as close to parity as possible and are not rejected by the data for a given confidence level.

Figure 4: Illustration of discounting method accounting for uncertainty

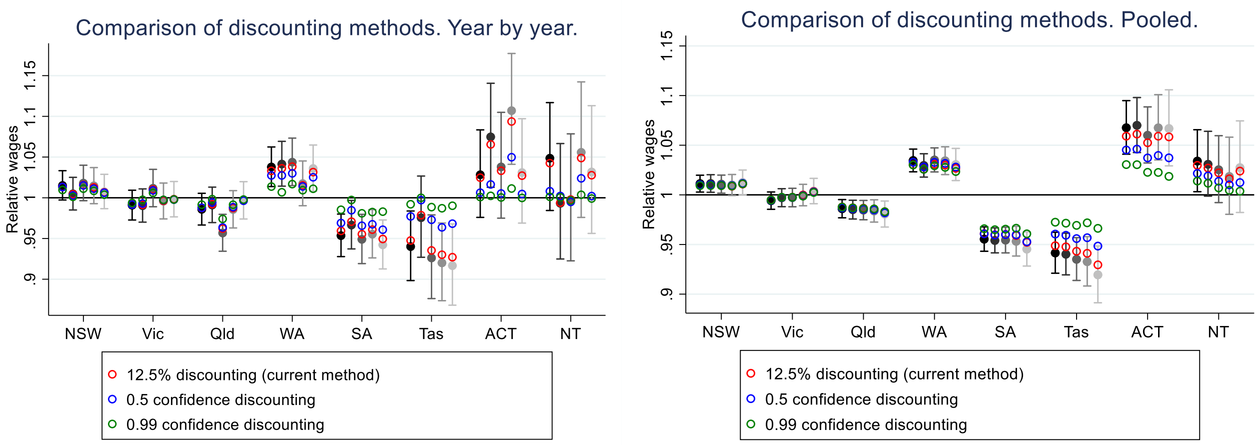


The level of discounting is then varied by changing the confidence level, which determines the width of the confidence interval. The advantage of this approach is that, if we obtained more or better data, there would be less uncertainty (the confidence interval would be narrower), and hence there would be less discounting. In the extreme case of no uncertainty, there would be no discounting.

22. This idea is easily extended to more than two states (see point 23 for technical details). Figure 5 compares our proposed approach to the 12.5% rule for the unpooled (left panel) and pooled five-year rolling window (right panel) samples. The model is the one that is currently used by the CGC (*legacy_usual*). The black/grey dots are the relative wages with no discounting. The red dots apply the 12.5% rule. The blue dots apply our approach using a low confidence level of 0.5 (smaller discounts). The green dots apply our approach using a high confidence level of 0.99 (larger discounts).

    Two features are immediately apparent. First, our approach leads to more discounting for the unpooled sample versus the pooled sample and for the less populous states versus the more populous states. In both cases this is because our discounting rule applies larger discounts when there is more uncertainty. Second, looking at the unpooled sample (left panel), we observe that our discounting rule leads to less volatility between years than the 12.5% rule. This is because the volatility is driven by uncertainty, and the level of discount applied using our rule depends explicitly on the amount of uncertainty.

Figure 5: Comparison of discounting methods



23. Here we provide technical details on the extension to two or more states. Let there be *s* states, and suppose that state *s* (the last state) is the reference state in the regression model. Then, the regression provides us with estimates of the coefficients for states 1 to *s-1*, denoted $\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_{s-1}$ which can be stacked into the *(s-1)*$\times 1$ vector $\hat{\boldsymbol{b}}$. The regression also provides us the $(s-1) \times (s-1)$ variance estimate for $\hat{\boldsymbol{b}}$, denoted by $\hat{\boldsymbol{V}}$**.** Applying the Wald test, the set of values for the state coefficients $\boldsymbol{b}$ which would not be rejected by the data with asymptotic confidence level 1-α is then given by those satisfying the inequality

$$(\boldsymbol{b} - \hat{\boldsymbol{b}})'\hat{\boldsymbol{V}}^{-1}(\boldsymbol{b} - \hat{\boldsymbol{b}}) \leq \boldsymbol{c_{1-\alpha}},$$

where $\boldsymbol{c_{1-\alpha}}$ is the 1-α quantile of the chi-squared distribution with s-1 degrees of freedom. Our proposed discounting method searches for the $\boldsymbol{b}$ which satisfies this inequality and is as close to $\boldsymbol{b} = \boldsymbol{0}$ as possible, whilst ensuring that all states receive a discount. Formally, we solve the quadratic program

$$min_{\boldsymbol{b}} \quad \boldsymbol{b}_1^2 + \boldsymbol{b}_2^2 + \cdots + \boldsymbol{b}_{s-1}^2 + \left(\sum_{k=1}^{s-1} \boldsymbol{b_k}\right)^2$$

subject to

$$(\boldsymbol{b} - \hat{\boldsymbol{b}})'\hat{\boldsymbol{V}}^{-1}(\boldsymbol{b} - \hat{\boldsymbol{b}}) \leq \boldsymbol{c_{1-\alpha}}$$

$$0 \leq b_k \leq \widehat{\mathrm{b_k}} \quad if \quad \widehat{\mathrm{b_k}} \geq 0, \qquad 0 \leq -\textstyle\sum_{k=1}^{s-1} b_k \leq -\sum_{k=1}^{s-1} \widehat{\mathrm{b_k}} \quad if \quad -\sum_{k=1}^{s-1} \widehat{\mathrm{b_k}} \geq 0$$

$$\widehat{\mathrm{b_k}} \leq b_k \leq 0 \quad if \quad \widehat{\mathrm{b_k}} \leq 0, \qquad -\textstyle\sum_{k=1}^{s-1} \widehat{\mathrm{b_k}} \leq -\sum_{k=1}^{s-1} b_k \leq 0 \quad if \quad -\sum_{k=1}^{s-1} \widehat{\mathrm{b_k}} \leq 0$$

The objective function minimizes the Euclidian distance of $\boldsymbol{b}$ from the origin (the point at which all states have the same wages). The final term in the objective function arises because the coefficient for the reference state is $b_s = -\sum_{k=1}^{s-1} b_k$. The first constraint implies that the coefficients are not rejected by the data. The remaining constraints imply that all state coefficients lie between the regression estimate and zero (i.e., they impose that there is discounting), including for the reference state.

Section 2: Response to Recommendations in the Reviewer Report

- ### Recommendation 1

  The Commission continue to use the regional wage structure in the private sector as a proxy for labour market pressures in the state/territory public sector.

24. We agree with the recommendation to continue to use the regional wage structure in the private sector as a proxy for labour market pressures in the public sector. More details regarding our response to this recommendation are elaborated in bullet points 4-8.

- Recommendation 2

  Given the different sex composition of the public and private sectors, the Commission give consideration to using the FEMALE private sector regional wage structure as a proxy for labour market pressures in the state/territory public sector.

25. We do not agree with this recommendation to use the female private sector regional wage structure as a proxy for labour market pressures in the public sector, regardless of the 65% female representation in the public sector. Our disagreement is based on two primary reasons:

    a. Using only female workers' wages introduces bias in the assessment. Studies have found that the gender wage gap is narrower in the public sector than in the private sector, due to stronger regulations, higher unionization, and more wage transparency in the public sector. Thus, using female wage only would be to focus on the relatively lower part of the wage distribution in the private sector, conditional on other variables controlled. Since the gender wage gap correlates with the industry being examined, assessing state labour market pressure using female wage only will cause bias in the assessment, because the estimation would be sensitive to the industry composition in different states.

    b. For example, some states might have more industries or occupations that are female-dominated, so economic fluctuations in these industries or occupations would disproportionately affect female wages. According to the most recent estimate in 2023 May, gender gap notably varies across industries, ranging from 5.2% (Public Administration and Safety) to 22.7% (Professional, Scientific and Technical Services).[6] The percentages of employment in the high-gender-gap industries is 42% in Victoria and only 30% in Australian Capital Territory.[7] Thus, labour market pressure estimated in some states would be higher when estimating using only female wages rather than estimating using all workers' wages. On the contrary, states with more male-dominated industries or occupations (i.e., female wages relatively lower) may be thought as having less labour market pressure, if labour market pressure is only estimated using female worker wages. By excluding male wages, the state effects would be capturing gender-specific shocks that disproportionately affect each state, rather than the overall economic conditions.

    c. Lastly, while the public sector has a notable 65% female representation, state governments are not mandated to hire workers based on criteria like gender, age, or race. Therefore, to assess the broader impact representing labour market conditions across states, there is no justification to only use female private sector wage structure as a proxy.

    d. Use of female workers only will substantially reduce the sample size, leading to increased uncertainty and volatility between years.

- Recommendation 3

  The Commission remain with the COES for estimation purposes.

[6] Australian Government Workplace Gender Equality Agency (2022).

[7] These industries are Agriculture, Forestry and Fishing, Electricity, Gas, Water and Waste Services, Construction, Wholesale Trade, Transport, Postal and Warehousing, Rental, Hiring and Real Estate Services, Financial and Insurance Services, and Professional, Scientific and Technical Services.

26. In the short run, we agree with the recommendation to continue using *COE Survey* for estimation purposes. However, in the longer term, we recommend CGC to consider employing administrative data available from PLIDA (formerly MADIP and recently renamed PLIDA, "person-level integrated data asset"), which provides much more accurate labour market information covering the nationally representative labour force. Administrative data is also less prone to the potential measurement errors commonly associated with self-reporting in survey-based data, and offers many more observations than survey data.

- ## Recommendation 4

The Commission use hourly wages as the dependent variable.

27. Overall, we do not agree with the recommendation to use hourly wage as the dependent variable. First, we agree with the discussion on Page 21 that adding hour-control variables on the right-hand-side of the regression where hourly wage is the dependent variable can be problematic, and our discussion is provided in bullet point 13. In addition, even if hour-control variables are not added on the right-hand-side, we still think hourly wage is not an appropriate measure in this assessment and the CGC's 2020 Legacy Model using usual wage may better serve the purpose of this assessment. We discuss three primary reasons in bullet point 14.

- ## Recommendation 5

The Commission deals with potential measurement error in hourly wages by excluding sample members who report working less than 5 hours per week in their main job and those working 60 or more hours per week in their main job.

28. We agree with this recommendation to exclude outliers from the estimation. For these outliers, coefficient of each human capital explanatory variable can be quite different from the analysis for the main sample, so excluding them from the analysis can reduce the noise in the estimation. In addition, we think it also appropriate to apply this exclusion on the 2020 Legacy Model whose dependent variable is the usual wage.

- ## Recommendation 6

If the Commission has strong a-priori reason to believe that the hours-wage relationship differs across the distribution the recommendation is to adopt a simpler specification using a dummy variable approach with controls for part-time hours and long-hours.

29. We agree with this. In addition, we think it also appropriate to apply this exclusion on the 2020 Legacy Model whose dependent variable is the usual wage.

- ## Recommendation 7

The Commission should use a series of age dummy variables to capture labour market experience rather than a measure of potential experience.

30. In the empirical literature, it is a more standard practice to include age and its quadratic term in an empirical specification to identify the nonlinear impact of age, assuming there is a "turning point" when one's income maximizes at a golden age. While using age dummy variables could also be a potential solution to address the nonlinear relationship between worker age and worker wage, this specification loses interpretation to

understand how accumulating age (and experience accordingly) can *marginally* affect one's wage. In addition, some effects within an age category may be cancelled out and could not be accurately captured using the age category variable. Thus, we are more inclined to follow CGC' recommendation to examine the impact of potential experience on worker wage.

- ### Recommendation 8

The Commission does not include age-education (interactions) in its model.

31. Related to our response to Recommendation 7, we are more inclined to follow CGC' recommendation to include age-education interaction in the model and we address our rational in bullet point 17.

- ### Recommendation 9

The Commission include tenure as a continuous variable.

32. We do not agree with this recommendation, because the impact of job tenure on wage is not linear. Early years in a job would come with rapid wage increases due to promotion of skill accumulation and training. After a certain point, additional years might not lead to significant wage increases. This is due to diminishing returns of skill accumulation, as well as less bargaining power when an employee has accumulated more of firm-specific human capital. Therefore, we think the original categorizing approach may be more appropriate. More details of this discussion can be found in bullet point 18.

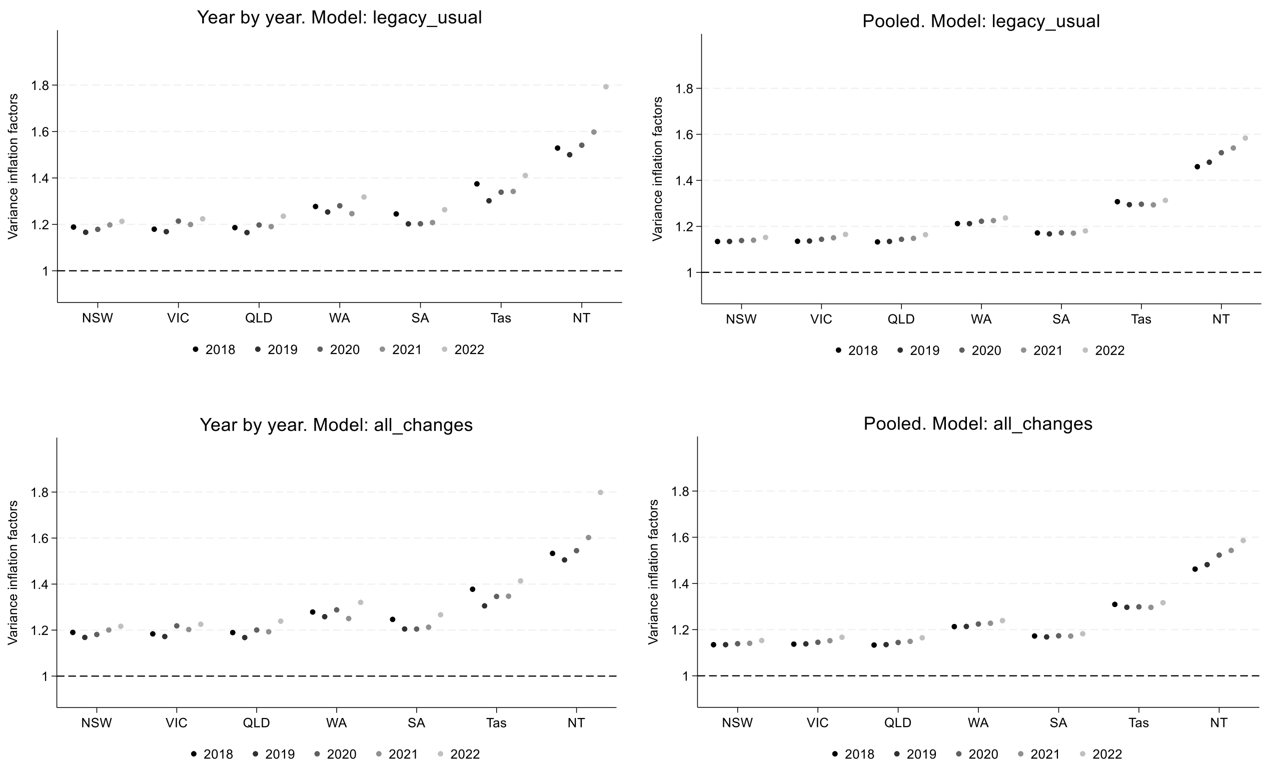
- ### Recommendation 10

The Commission seek to estimate a parsimonious model (fewer predictor variables).

33. The *Reviewer Report* proposes to estimate a model with fewer predictor variables. One of the concerns in the current model is multicollinearity. We test multicollinearity using Variance Inflation Factors. A variance inflation factor for a regressor of interest is 1/(1-Rsq), where Rsq is the R-squared from the regression of the regressor of interest on all other regressors in the model. A variance inflation factor of one means there is no multicollinearity. Researchers might worry about multicollinearity for variance inflation factors which exceed five.

    Figure 5 demonstrates that variance inflation factors for all state coefficients remain below two, regardless of whether the sample is pooled over a five-year rolling window or unpooled, and for all regression models considered. We conclude that there is little evidence that the uncertainty around relative wages is driven by multicollinearity. Note that the reference state in the regression is the ACT, hence we do not report a variance inflation factor for the ACT. Note also that, though we discussed that there is likely to be multicollinearity among usual and paid hours in point 15, since both of those variables are controls, this issue is distinct from multicollinearity concerning the variables of interest.

Figure 6: Variance inflation factors for all regression models considered



- ## Recommendation 11

  To reduce the volatility of the geographic wage relativities the Commission consider alternative approaches such as pooling data over a moving three-year period when estimating the geographic wage structure.

34. We agree with this recommendation to consider pooling data over a rolling window to improve accuracy and reduce volatility of the estimation. Our rationale for endorsing this approach is elaborated in bullet points 9-12.

Appendix

Table Appendix 1: Loss in sample size by excluding most-affected industries during the COVID-19

(Excluding observations in Accommodation and Food Services, and Arts and Recreation Services industries during the years of 2020-2022.)

A. Current year data

	NSW	VIC	QLD	WA	SA	TAS	ACT	NT
2018	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2019	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2020	8.35%	6.26%	8.08%	7.34%	8.27%	9.73%	9.41%	7.92%
2021	6.58%	8.37%	8.66%	7.57%	7.33%	11.17%	7.18%	8.77%
2022	8.55%	7.65%	8.67%	9.00%	6.81%	10.35%	6.12%	10.09%

B. Pooling – 3-year rolling window around the current year

	NSW	VIC	QLD	WA	SA	TAS	ACT	NT
2018	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2019	2.78%	2.09%	2.69%	2.45%	2.76%	3.24%	3.14%	2.64%
2020	4.98%	4.88%	5.58%	4.97%	5.20%	6.97%	5.53%	5.56%
2021	7.83%	7.43%	8.47%	7.97%	7.47%	10.42%	7.57%	8.93%
2022	7.57%	8.01%	8.66%	8.29%	7.07%	10.76%	6.65%	9.43%

C. Pooling – 5-year rolling window around the current year

	NSW	VIC	QLD	WA	SA	TAS	ACT	NT
2018	1.67%	1.25%	1.62%	1.47%	1.65%	1.95%	1.88%	1.58%
2019	2.99%	2.93%	3.35%	2.98%	3.12%	4.18%	3.32%	3.34%
2020	4.70%	4.46%	5.08%	4.78%	4.48%	6.25%	4.54%	5.36%
2021	5.87%	5.57%	6.35%	5.98%	5.60%	7.81%	5.68%	6.69%
2022	7.83%	7.43%	8.47%	7.97%	7.47%	10.42%	7.57%	8.93%

Table Appendix 2a: Adjusted coefficients using different discounting methods (Pooled sample, five year window)

		Legacy Model estimates	12.5% discounting	0.5 confidence discounting	0.99 confidence discounting
2018	NSW	1.0113	1.0098	1.0101	1.0093
	VIC	0.9941	0.9949	0.9945	0.9948
	QLD	0.9860	0.9878	0.9869	0.9877
	WA	1.0346	1.0303	1.0328	1.0301
	SA	0.9554	0.9610	0.9608	0.9659
	TAS	0.9415	0.9488	0.9605	0.9725
	ACT	1.0676	1.0591	1.0452	1.0305
	NT	1.0340	1.0297	1.0216	1.0139
2019	NSW	1.0114	1.0100	1.0102	1.0094
	VIC	0.9971	0.9975	0.9974	0.9976
	QLD	0.9851	0.9870	0.9860	0.9869
	WA	1.0296	1.0259	1.0281	1.0257
	SA	0.9541	0.9598	0.9596	0.9649
	TAS	0.9402	0.9477	0.9593	0.9715
	ACT	1.0700	1.0612	1.0461	1.0304
	NT	1.0308	1.0270	1.0191	1.0117
2020	NSW	1.0108	1.0094	1.0098	1.0093
	VIC	0.9973	0.9976	0.9975	0.9975
	QLD	0.9846	0.9865	0.9855	0.9863
	WA	1.0351	1.0308	1.0327	1.0293
	SA	0.9545	0.9602	0.9600	0.9655
	TAS	0.9351	0.9432	0.9560	0.9693
	ACT	1.0600	1.0525	1.0372	1.0227
	NT	1.0253	1.0222	1.0137	1.0070
2021	NSW	1.0099	1.0087	1.0093	1.0089
	VIC	0.9998	0.9998	0.9991	0.9987
	QLD	0.9837	0.9858	0.9848	0.9859
	WA	1.0343	1.0300	1.0317	1.0277
	SA	0.9532	0.9591	0.9597	0.9664
	TAS	0.9327	0.9411	0.9569	0.9719
	ACT	1.0675	1.0591	1.0397	1.0226
	NT	1.0185	1.0162	1.0101	1.0046
2022	NSW	1.0125	1.0109	1.0116	1.0111
	VIC	1.0038	1.0033	1.0029	1.0024
	QLD	0.9806	0.9830	0.9818	0.9830
	WA	1.0305	1.0267	1.0280	1.0235
	SA	0.9453	0.9522	0.9528	0.9608
	TAS	0.9195	0.9295	0.9484	0.9663
	ACT	1.0668	1.0584	1.0373	1.0186
	NT	1.0274	1.0239	1.0124	1.0037

Table Appendix 2b: Adjusted coefficients using different discounting methods (Unpooled sample)

		Legacy Model estimates	12.5% discounting	0.5 confidence discounting	0.99 confidence discounting
2018	NSW	1.0151	1.0132	1.0132	1.0096
	VIC	0.9908	0.9917	0.9917	0.9937
	QLD	0.9858	0.9878	0.9878	0.9916
	WA	1.0378	1.0274	1.0274	1.0138
	SA	0.9536	0.9691	0.9691	0.9853
	TAS	0.9401	0.9772	0.9772	0.9921
	ACT	1.0283	1.0063	1.0063	1.0009
	NT	1.0486	1.0081	1.0081	1.0012
2019	NSW	1.0049	1.0029	1.0029	1.0012
	VIC	0.9903	0.9930	0.9930	0.9976
	QLD	0.9911	0.9943	0.9943	0.9983
	WA	1.0414	1.0274	1.0274	1.0068
	SA	0.9665	0.9847	0.9847	0.9973
	TAS	0.9756	0.9972	0.9972	1.0000
	ACT	1.0748	1.0164	1.0164	1.0024
	NT	0.9932	1.0022	1.0022	1.0005
2020	NSW	1.0180	1.0152	1.0152	1.0112
	VIC	1.0117	1.0091	1.0091	1.0060
	QLD	0.9568	0.9636	0.9636	0.9740
	WA	1.0437	1.0300	1.0300	1.0168
	SA	0.9490	0.9657	0.9657	0.9809
	TAS	0.9262	0.9730	0.9730	0.9886
	ACT	1.0379	1.0053	1.0053	1.0004
	NT	0.9975	0.9950	0.9950	0.9968
2021	NSW	1.0147	1.0112	1.0112	1.0085
	VIC	0.9958	0.9973	0.9973	0.9975
	QLD	0.9856	0.9885	0.9885	0.9915
	WA	1.0174	1.0139	1.0139	1.0093
	SA	0.9554	0.9675	0.9675	0.9825
	TAS	0.9201	0.9637	0.9637	0.9872
	ACT	1.1070	1.0498	1.0498	1.0114
	NT	1.0558	1.0239	1.0239	1.0037
2022	NSW	1.0075	1.0062	1.0062	1.0038
	VIC	0.9980	0.9977	0.9977	0.9979
	QLD	0.9966	0.9964	0.9964	0.9978
	WA	1.0360	1.0252	1.0252	1.0111
	SA	0.9422	0.9608	0.9608	0.9830
	TAS	0.9167	0.9682	0.9682	0.9903
	ACT	1.0310	1.0047	1.0047	0.9998
	NT	1.0317	1.0021	1.0021	0.9991

Table Appendix 3: State coefficients over different regression models and samples

Model: legacy_usual, Sample: year by year	2018	2019	2020	2021	2022
NSW	0.0157*	0.000705	0.0261**	0.00955	0.0134
	(0.00905)	(0.0102)	(0.0109)	(0.0111)	(0.0107)
Vic	-0.00855	-0.0139	0.0199*	-0.00922	0.00393
	(0.00940)	(0.0105)	(0.0116)	(0.0113)	(0.0111)
Qld	-0.0136	-0.0131	-0.0359***	-0.0195*	0.00251
	(0.0101)	(0.0112)	(0.0121)	(0.0119)	(0.0117)
WA	0.0378***	0.0364***	0.0511***	0.0122	0.0413***
	(0.0119)	(0.0135)	(0.0142)	(0.0138)	(0.0140)
SA	-0.0469***	-0.0382**	-0.0441***	-0.0507***	-0.0536***
	(0.0140)	(0.0157)	(0.0163)	(0.0158)	(0.0163)
Tas	-0.0611***	-0.0289	-0.0684**	-0.0883***	-0.0811***
	(0.0232)	(0.0261)	(0.0283)	(0.0264)	(0.0278)
NT	0.0481	-0.011	0.00581	0.0493	0.0371
	(0.0322)	(0.0364)	(0.0398)	(0.0402)	(0.0387)
R squared	0.677	0.651	0.633	0.618	0.6
Observations	16846	14890	12754	13738	15272

Standard errors in parentheses. *: p<0.1, **:p<0.05, ***:p<0.01

Model: legacy_usual, Sample: Pooled five years	2018	2019	2020	2021	2022
NSW	0.00893**	0.00989**	0.0111**	0.0107**	0.0151**
	(0.00430)	(0.00446)	(0.00460)	(0.00533)	(0.00625)
Vic	-0.00816*	-0.0044	-0.00229	0.000626	0.00649
	(0.00447)	(0.00462)	(0.00477)	(0.00552)	(0.00648)
Qld	-0.0163***	-0.0165***	-0.0151***	-0.0155***	-0.0168**
	(0.00474)	(0.00490)	(0.00505)	(0.00582)	(0.00682)
WA	0.0317***	0.0276***	0.0350***	0.0346***	0.0328***
	(0.00564)	(0.00581)	(0.00598)	(0.00690)	(0.00803)
SA	-0.0479***	-0.0485***	-0.0462***	-0.0470***	-0.0535***
	(0.00656)	(0.00673)	(0.00695)	(0.00799)	(0.00928)
Tas	-0.0625***	-0.0632***	-0.0667***	-0.0688***	-0.0812***
	(0.0112)	(0.0114)	(0.0118)	(0.0136)	(0.0159)
NT	0.0311**	0.0289*	0.0255	0.0192	0.0297
	(0.0154)	(0.0161)	(0.0168)	(0.0194)	(0.0229)
R squared	0.654	0.636	0.616	0.605	0.597
Observations	75726	74011	73500	56654	41764

Standard errors in parentheses. *: p<0.1, **:p<0.05, ***:p<0.01

Model: cont_tenure, Sample: year by year	2018	2019	2020	2021	2022
NSW	0.0158*	0.000245	0.0259**	0.00951	0.0135
	(0.00905)	(0.0102)	(0.0109)	(0.0111)	(0.0107)
Vic	-0.00855	-0.0137	0.0205*	-0.00959	0.00373
	(0.00940)	(0.0105)	(0.0115)	(0.0113)	(0.0111)
Qld	-0.0137	-0.0126	-0.0361***	-0.0205*	0.00258
	(0.0101)	(0.0112)	(0.0121)	(0.0119)	(0.0117)
WA	0.0380***	0.0369***	0.0509***	0.0116	0.0411***
	(0.0119)	(0.0135)	(0.0142)	(0.0138)	(0.0140)
SA	-0.0466***	-0.0386**	-0.0445***	-0.0504***	-0.0538***
	(0.0140)	(0.0157)	(0.0163)	(0.0158)	(0.0163)
Tas	-0.0614***	-0.0287	-0.0684**	-0.0882***	-0.0815***
	(0.0232)	(0.0261)	(0.0283)	(0.0264)	(0.0277)
NT	0.0478	-0.0107	0.00606	0.0493	0.0376
	(0.0322)	(0.0364)	(0.0398)	(0.0402)	(0.0387)
R squared	0.677	0.651	0.633	0.618	0.6
Observations	16846	14890	12754	13738	15272

Standard errors in parentheses. *: p<0.1, **:p<0.05, ***:p<0.01

Model: cont_tenure, Sample: Pooled five years	2018	2019	2020	2021	2022
NSW	0.00874**	0.00976**	0.0110**	0.0105**	0.0151**
	(0.00430)	(0.00446)	(0.00460)	(0.00533)	(0.00625)
Vic	-0.00817*	-0.00438	-0.00231	0.000605	0.00641
	(0.00447)	(0.00462)	(0.00477)	(0.00552)	(0.00648)
Qld	-0.0162***	-0.0166***	-0.0151***	-0.0155***	-0.0170**
	(0.00474)	(0.00490)	(0.00505)	(0.00582)	(0.00682)
WA	0.0320***	0.0278***	0.0351***	0.0347***	0.0326***
	(0.00564)	(0.00581)	(0.00598)	(0.00690)	(0.00803)
SA	-0.0480***	-0.0486***	-0.0463***	-0.0473***	-0.0536***
	(0.00656)	(0.00673)	(0.00695)	(0.00799)	(0.00928)
Tas	-0.0628***	-0.0634***	-0.0668***	-0.0689***	-0.0814***
	(0.0112)	(0.0114)	(0.0118)	(0.0136)	(0.0159)
NT	0.0316**	0.0292*	0.0257	0.0195	0.03
	(0.0154)	(0.0161)	(0.0168)	(0.0194)	(0.0229)
R squared	0.654	0.636	0.616	0.605	0.597
Observations	75726	74011	73500	56654	41764

Standard errors in parentheses. *: p<0.1, **:p<0.05, ***:p<0.01

Model: alt_workexp, Sample: year by year	2018	2019	2020	2021	2022
NSW	0.0171*	0.000533	0.0264**	0.0102	0.0128
	(0.00905)	(0.0102)	(0.0109)	(0.0111)	(0.0107)
Vic	-0.00759	-0.0145	0.0211*	-0.0101	0.00497
	(0.00940)	(0.0105)	(0.0115)	(0.0113)	(0.0111)
Qld	-0.0131	-0.0134	-0.0363***	-0.0197*	0.00121
	(0.0101)	(0.0112)	(0.0120)	(0.0119)	(0.0117)
WA	0.0369***	0.0367***	0.0511***	0.013	0.0418***
	(0.0119)	(0.0135)	(0.0142)	(0.0138)	(0.0139)
SA	-0.0469***	-0.0390**	-0.0447***	-0.0490***	-0.0536***
	(0.0140)	(0.0157)	(0.0162)	(0.0158)	(0.0163)
Tas	-0.0591**	-0.0278	-0.0639**	-0.0869***	-0.0788***
	(0.0231)	(0.0260)	(0.0282)	(0.0264)	(0.0277)
NT	0.0471	-0.00922	0.00217	0.0467	0.0353
	(0.0322)	(0.0363)	(0.0397)	(0.0401)	(0.0387)
R squared	0.679	0.654	0.637	0.62	0.603
Observations	16848	14890	12754	13738	15272

Standard errors in parentheses. *: p<0.1, **:p<0.05, ***:p<0.01

Model: alt_workexp, Sample: Pooled five years	2018	2019	2020	2021	2022
NSW	0.00927**	0.0104**	0.0114**	0.0109**	0.0155**
	(0.00429)	(0.00446)	(0.00460)	(0.00532)	(0.00624)
Vic	-0.00764*	-0.00403	-0.00197	0.000825	0.00712
	(0.00446)	(0.00462)	(0.00476)	(0.00551)	(0.00647)
Qld	-0.0173***	-0.0172***	-0.0158***	-0.0163***	-0.0173**
	(0.00473)	(0.00489)	(0.00504)	(0.00581)	(0.00680)
WA	0.0308***	0.0271***	0.0347***	0.0344***	0.0330***
	(0.00563)	(0.00580)	(0.00597)	(0.00689)	(0.00802)
SA	-0.0479***	-0.0484***	-0.0463***	-0.0470***	-0.0533***
	(0.00654)	(0.00672)	(0.00694)	(0.00798)	(0.00926)
Tas	-0.0608***	-0.0618***	-0.0647***	-0.0669***	-0.0793***
	(0.0112)	(0.0114)	(0.0117)	(0.0136)	(0.0158)
NT	0.0307**	0.0282*	0.0245	0.0181	0.0276
	(0.0154)	(0.0161)	(0.0167)	(0.0194)	(0.0229)
R squared	0.656	0.638	0.618	0.607	0.599
Observations	75728	74013	73502	56654	41764

Standard errors in parentheses. *: p<0.1, **:p<0.05, ***:p<0.01

Model: legacy_paid, Sample: year by year	2018	2019	2020	2021	2022
NSW	0.0178**	0.000105	0.0135	0.0174*	0.0171*
	(0.00810)	(0.00872)	(0.00965)	(0.00963)	(0.00898)
Vic	-0.0082	-0.0190**	0.0254**	0.000213	0.0043
	(0.00842)	(0.00901)	(0.0102)	(0.00976)	(0.00933)
Qld	-0.0214**	-0.0153	-0.0443***	-0.0174*	0.00933
	(0.00900)	(0.00960)	(0.0107)	(0.0103)	(0.00987)
WA	0.0295***	0.0467***	0.0367***	0.00326	0.0356***
	(0.0107)	(0.0115)	(0.0126)	(0.0119)	(0.0117)
SA	-0.0414***	-0.0438***	-0.0514***	-0.0466***	-0.0467***
	(0.0125)	(0.0135)	(0.0144)	(0.0137)	(0.0137)
Tas	-0.0558***	-0.0286	-0.0493**	-0.0735***	-0.0827***
	(0.0207)	(0.0223)	(0.0250)	(0.0228)	(0.0233)
NT	0.0461	-0.00502	0.0162	0.0239	0.0141
	(0.0288)	(0.0312)	(0.0353)	(0.0347)	(0.0327)
R squared	0.743	0.745	0.714	0.716	0.72
Observations	16767	14831	12708	13667	15193

Standard errors in parentheses. *: p<0.1, **:p<0.05, ***:p<0.01

Model: legacy_paid, Sample: Pooled five years	2018	2019	2020	2021	2022
NSW	0.00554	0.00875**	0.0111***	0.0104**	0.0157***
	(0.00384)	(0.00394)	(0.00401)	(0.00460)	(0.00539)
Vic	-0.00728*	-0.00299	-0.000519	0.00257	0.0114**
	(0.00399)	(0.00408)	(0.00416)	(0.00476)	(0.00559)
Qld	-0.0208***	-0.0211***	-0.0165***	-0.0160***	-0.0168***
	(0.00423)	(0.00432)	(0.00439)	(0.00502)	(0.00588)
WA	0.0312***	0.0279***	0.0306***	0.0304***	0.0234***
	(0.00503)	(0.00512)	(0.00521)	(0.00595)	(0.00692)
SA	-0.0482***	-0.0487***	-0.0442***	-0.0462***	-0.0506***
	(0.00585)	(0.00594)	(0.00605)	(0.00690)	(0.00800)
Tas	-0.0583***	-0.0550***	-0.0594***	-0.0608***	-0.0714***
	(0.00997)	(0.0100)	(0.0102)	(0.0117)	(0.0137)
NT	0.0299**	0.0234*	0.0179	0.0101	0.0151
	(0.0137)	(0.0142)	(0.0146)	(0.0168)	(0.0198)
R squared	0.726	0.718	0.711	0.708	0.702
Observations	75450	73700	73166	56399	41568

Standard errors in parentheses. *: p<0.1, **:p<0.05, ***:p<0.01

Model: resp_hours, Sample: year by year	2018	2019	2020	2021	2022
NSW	0.0173**	0.00198	0.0169*	0.0179*	0.0151*
	(0.00791)	(0.00857)	(0.00942)	(0.00940)	(0.00877)
Vic	-0.00572	-0.0172*	0.0294***	0.0000853	0.00816
	(0.00822)	(0.00886)	(0.00998)	(0.00952)	(0.00910)
Qld	-0.0223**	-0.0176*	-0.0462***	-0.0223**	0.00458
	(0.00879)	(0.00944)	(0.0104)	(0.0100)	(0.00964)
WA	0.0353***	0.0420***	0.0364***	0.0056	0.0413***
	(0.0104)	(0.0113)	(0.0123)	(0.0116)	(0.0115)
SA	-0.0415***	-0.0419***	-0.0506***	-0.0458***	-0.0479***
	(0.0122)	(0.0132)	(0.0140)	(0.0134)	(0.0134)
Tas	-0.0532***	-0.0276	-0.0420*	-0.0683***	-0.0761***
	(0.0202)	(0.0219)	(0.0244)	(0.0223)	(0.0227)
NT	0.0363	-0.00703	0.00493	0.0182	0.0064
	(0.0281)	(0.0307)	(0.0344)	(0.0338)	(0.0319)
R squared	0.755	0.754	0.728	0.73	0.733
Observations	16767	14831	12708	13667	15193

Standard errors in parentheses. *: p<0.1, **:p<0.05, ***:p<0.01

Model: resp_hours, Sample: Pooled five years	2018	2019	2020	2021	2022
NSW	0.00679*	0.00963**	0.0111***	0.0107**	0.0157***
	(0.00375)	(0.00385)	(0.00392)	(0.00449)	(0.00526)
Vic	-0.00548	-0.00143	0.00187	0.00498	0.0139**
	(0.00390)	(0.00398)	(0.00406)	(0.00465)	(0.00545)
Qld	-0.0229***	-0.0234***	-0.0197***	-0.0198***	-0.0209***
	(0.00414)	(0.00422)	(0.00429)	(0.00491)	(0.00574)
WA	0.0320***	0.0285***	0.0324***	0.0316***	0.0265***
	(0.00492)	(0.00500)	(0.00509)	(0.00581)	(0.00676)
SA	-0.0476***	-0.0477***	-0.0440***	-0.0459***	-0.0506***
	(0.00572)	(0.00580)	(0.00591)	(0.00674)	(0.00781)
Tas	-0.0537***	-0.0500***	-0.0548***	-0.0560***	-0.0651***
	(0.00974)	(0.00981)	(0.0100)	(0.0114)	(0.0133)
NT	0.0239*	0.0174	0.0114	0.00459	0.00816
	(0.0134)	(0.0139)	(0.0143)	(0.0164)	(0.0193)
R squared	0.738	0.731	0.724	0.721	0.716
Observations	75450	73700	73166	56399	41568

Standard errors in parentheses. *: p<0.1, **:p<0.05, ***:p<0.01

Model: hourly_wage, Sample: year by year	2018	2019	2020	2021	2022
NSW	0.0173**	0.00198	0.0169*	0.0179*	0.0151*
	(0.00791)	(0.00857)	(0.00942)	(0.00940)	(0.00877)
Vic	-0.00572	-0.0172*	0.0294***	0.0000843	0.00816
	(0.00822)	(0.00886)	(0.00998)	(0.00952)	(0.00910)
Qld	-0.0223**	-0.0176*	-0.0462***	-0.0223**	0.00458
	(0.00879)	(0.00944)	(0.0104)	(0.0100)	(0.00964)
WA	0.0353***	0.0420***	0.0364***	0.00561	0.0413***
	(0.0104)	(0.0113)	(0.0123)	(0.0116)	(0.0115)
SA	-0.0415***	-0.0419***	-0.0506***	-0.0458***	-0.0479***
	(0.0122)	(0.0132)	(0.0140)	(0.0134)	(0.0134)
Tas	-0.0532***	-0.0276	-0.0420*	-0.0683***	-0.0761***
	(0.0202)	(0.0219)	(0.0244)	(0.0223)	(0.0227)
NT	0.0363	-0.00703	0.00493	0.0182	0.00639
	(0.0281)	(0.0307)	(0.0344)	(0.0338)	(0.0319)
R squared	0.438	0.455	0.455	0.451	0.447
Observations	16767	14831	12708	13667	15193

Standard errors in parentheses. *: p<0.1, **:p<0.05, ***:p<0.01

Model: hourly_wage, Sample: Pooled five years	2018	2019	2020	2021	2022
NSW	0.00679*	0.00963**	0.0111***	0.0107**	0.0157***
	(0.00375)	(0.00385)	(0.00392)	(0.00449)	(0.00526)
Vic	-0.00548	-0.00143	0.00188	0.00498	0.0139**
	(0.00390)	(0.00398)	(0.00406)	(0.00465)	(0.00545)
Qld	-0.0229***	-0.0234***	-0.0197***	-0.0198***	-0.0209***
	(0.00414)	(0.00422)	(0.00429)	(0.00491)	(0.00574)
WA	0.0320***	0.0285***	0.0324***	0.0316***	0.0265***
	(0.00492)	(0.00500)	(0.00509)	(0.00581)	(0.00676)
SA	-0.0476***	-0.0477***	-0.0440***	-0.0459***	-0.0506***
	(0.00572)	(0.00580)	(0.00591)	(0.00674)	(0.00781)
Tas	-0.0537***	-0.0500***	-0.0548***	-0.0560***	-0.0651***
	(0.00974)	(0.00981)	(0.0100)	(0.0114)	(0.0133)
NT	0.0239*	0.0174	0.0114	0.00459	0.00816
	(0.0134)	(0.0139)	(0.0143)	(0.0164)	(0.0193)
R squared	0.427	0.426	0.42	0.422	0.422
Observations	75450	73700	73166	56399	41568

Standard errors in parentheses. *: p<0.1, **:p<0.05, ***:p<0.01

Model: all_changes, Sample: year by year	2018	2019	2020	2021	2022
NSW	0.0186**	0.000777	0.0167*	0.0181*	0.0144*
	(0.00789)	(0.00855)	(0.00939)	(0.00939)	(0.00875)
Vic	-0.00476	-0.0170*	0.0312***	-0.00117	0.00861
	(0.00821)	(0.00884)	(0.00995)	(0.00952)	(0.00908)
Qld	-0.0221**	-0.0172*	-0.0469***	-0.0236**	0.00299
	(0.00878)	(0.00942)	(0.0104)	(0.0100)	(0.00962)
WA	0.0340***	0.0419***	0.0362***	0.00616	0.0412***
	(0.0104)	(0.0113)	(0.0122)	(0.0116)	(0.0114)
SA	-0.0417***	-0.0434***	-0.0516***	-0.0443***	-0.0473***
	(0.0122)	(0.0132)	(0.0140)	(0.0134)	(0.0133)
Tas	-0.0512**	-0.0257	-0.0385	-0.0680***	-0.0740***
	(0.0202)	(0.0219)	(0.0243)	(0.0223)	(0.0227)
NT	0.0357	-0.00604	0.00271	0.017	0.00627
	(0.0281)	(0.0306)	(0.0343)	(0.0338)	(0.0318)
R squared	0.441	0.459	0.46	0.454	0.451
Observations	16769	14831	12708	13667	15193

Standard errors in parentheses. *: p<0.1, **:p<0.05, ***:p<0.01

Model: all_changes, Sample: Pooled five years	2018	2019	2020	2021	2022
NSW	0.00692*	0.00989***	0.0111***	0.0105**	0.0158***
	(0.00374)	(0.00384)	(0.00391)	(0.00448)	(0.00525)
Vic	-0.00491	-0.000932	0.00226	0.00528	0.0145***
	(0.00389)	(0.00398)	(0.00405)	(0.00464)	(0.00544)
Qld	-0.0236***	-0.0241***	-0.0204***	-0.0206***	-0.0216***
	(0.00412)	(0.00421)	(0.00428)	(0.00490)	(0.00572)
WA	0.0312***	0.0280***	0.0320***	0.0313***	0.0266***
	(0.00490)	(0.00499)	(0.00507)	(0.00580)	(0.00674)
SA	-0.0478***	-0.0477***	-0.0441***	-0.0459***	-0.0504***
	(0.00570)	(0.00579)	(0.00590)	(0.00673)	(0.00779)
Tas	-0.0521***	-0.0489***	-0.0529***	-0.0542***	-0.0636***
	(0.00971)	(0.00979)	(0.00998)	(0.0114)	(0.0133)
NT	0.0238*	0.017	0.011	0.00419	0.00698
	(0.0134)	(0.0139)	(0.0142)	(0.0164)	(0.0192)
R squared	0.43	0.43	0.423	0.425	0.426
Observations	75452	73702	73168	56399	41568

Standard errors in parentheses. *: p<0.1, **:p<0.05, ***:p<0.01

References

Australian Government Workplace Gender Equality Agency (2022). “Australia’s Gender Equality Scorecaard”. https://www.wgea.gov.au/sites/default/files/documents/WGEA-Gender-Equality-Scorecard-2022.pdf

Baker, George, Michael Gibbs, and Bengt Holmstrom. "The internal economics of the firm: Evidence from personnel data." *The Quarterly Journal of Economics* 109, no. 4 (1994): 881-919.

Bartlett, Robert P., and Frank Partnoy. "The ratio problem." *Available at SSRN 3605606* (2020).

Beaudry, Paul, and John DiNardo. "The effect of implicit contracts on the movement of wages over the business cycle: Evidence from micro data." *Journal of political Economy* 99, no. 4 (1991): 665-688.

Becker, Gary S. *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago press, 1964.

Bertola, Giuseppe. "Job security, employment and wages." *European economic review* 34, no. 4 (1990): 851-879.

Borjas, George J. "The wage structure and the sorting of workers into the public sector." (2002).

Cai, Zongwu and Ted Juhl. “The distribution of rolling regression estimators.” *Journal of Econometrics* 235, (2023): 1447-1463.

Clemens, Michael A., and Jennifer Hunt. "The labor market effects of refugee waves: reconciling conflicting results." *ILR Review* 72, no. 4 (2019): 818-857.

Freeman, Richard B., and Casey Ichniowski, eds. *When public sector workers unionize*. University of Chicago Press, 2007.

Gomes, Pedro. "Optimal public sector wages." *The Economic Journal* 125, no. 587 (2015): 1425-1451.

Katz, Lawrence F., and Alan B. Krueger. "Changes in the Structure of Wages in the Public and Private Sectors." (1991).

Lazear, Edward P. "Firm-specific human capital: A skill-weights approach." *Journal of political economy* 117, no. 5 (2009): 914-940.

Leonardi, Marco, and Giovanni Pica. "Who pays for it? The heterogeneous wage effects of employment protection legislation." *The Economic Journal* 123, no. 573 (2013): 1236-1278.

Kronmal, Richard A. "Spurious correlation and the fallacy of the ratio standard revisited." *Journal of the Royal Statistical Society Series A: Statistics in Society* 156, no. 3 (1993): 379-392.

Stole, Lars A., and Jeffrey Zwiebel. "Intra-firm bargaining under non-binding contracts." *The Review of Economic Studies* 63, no. 3 (1996): 375-410.

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

Contact details

Dr. Christiern Rose

T +61 7 336 56770
E christiern.rose@uq.edu.au

Dr. Lizi Yu

T +61 7 334 67054
E lizi.yu@uq.edu.au

Prof. Alicia Rambaldi

T +61 7 336 56576
E a.rambaldi@uq.edu.au

School of Economics | The University of Queensland

CRICOS Provider 00025B • TEQSA PRV12080