Australian Government

Commonwealth Grants Commission

2025 Methodology Review

Wage costs consultation paper

June 2023

| Paper issued	20 June 2023
Commission contact officer	Philip Harber, 02 6218 5726, Philip.Harber@cgc.gov.au
Submissions sought by	13 November 2023
	Submissions should be placed in your private state channel in CGC Engagement Teams, with a post notifying Katrina Baldock.
	Submissions should be in Word format and posted in the relevant state channel of the CGC engagement Team. Submissions more than 10 pages in length should include a summary section.
Confidential material	It is the Commission's normal practice to make state submissions available on its website under the CC BY licence, allowing free use of content by third parties.
	Further information on the CC BY licence can be found on the Creative Commons website (http://creativecommons.org).
	Confidential material contained in submissions must be clearly identified or included in separate attachment/s, and details provided in the covering email. Identified confidential material will not be published on the Commission's website.

CONTENTS

Overview of category

1   State governments employ about 1 in 10 Australian workers. Wages and salaries represent the largest component of recurrent state expenditure and account for a significant share of expenses in every expense category. The wage costs assessment addresses a cross-cutting driver of interstate difference in cost pressures, rather than the expenses associated with an individual category of service delivery (such as schools or health spending).

2   The wage costs assessment recognises that comparable public sector employees in different states are paid different wages, partly due to differences in labour markets beyond the control of state governments.

3   The Commission estimates differences between state wage costs using an econometric model of wages of private sector employees, which uses data from the ABS's Characteristics of Employment survey. Private sector wages are used to estimate a policy neutral measure of public sector wage differences across states. The ABS survey is conducted annually and a new estimate of relative differences in state wages is calculated for each assessment year.

4   The assessment is designed to identify the difference in cost for states to employ similar workers. Non-geographic factors (such as education, skills mix and years of experience) that influence individual wages are controlled for in the model. This leaves only the geographic effects (such as local amenities, climate, attachment to a state, cost of migration).

5   In the 2025 Review, the Commission's review framework focuses on what has changed since the 2020 Review and whether these changes have implications for the assessment method. With regard to the wage costs assessment, the labour market has continued to evolve, including because of the impact of COVID-19, and new data have become available.

6   Noting the importance and complexity of this assessment, the Commission has engaged Professor Alison Preston of the University of Western Australia Business School, a labour market econometrician, to review the wage costs assessment method, including the specification of the econometric model used by the Commission. States will be invited to provide specific questions for the consultant to consider. States will be provided the resulting report in August 2023 so that it can inform their submissions on this consultation paper. Given that the states will receive the external consultant's report on the wage costs assessment in August 2023, the timing for states providing submissions on wage costs has been extended to **13 November 2023**.

Current assessment method – 2020 Review

7   Public sector wage levels vary between states. There are many factors influencing these differences. The Commission's task is to identify differences resulting from factors outside a state's control. It does this by measuring the differences in private sector wages across states and using them as a proxy for the non-policy driven differences across states in public sector wages. The model assumes that geographic effects will have the same impact on public sector wages as on private sector wages.

8   The Commission uses a regression to estimate the differences in wages between individuals attributed to a wide range of characteristics. A state dummy variable is included to estimate the wage difference attributed to state level geographic effects. The model uses extensive controls to account for differences in industry, occupation, education, experience and other, non-geographic factors that influence individual wages. The model excludes all public sector employees to eliminate any direct effects of state government policy on wages, however there is still potential for high public sector wages to drive up private sector wages in a state.

9   The Commission uses a regression to calculate coefficients. It converts these to provide a wage cost factor for each state. A state's wage cost factor reflects the percentage difference from the national average wage level that is driven by geographic cost pressures.

10   In the 2020 Review a 'low' discount of 12.5% was applied to the wage cost factors. This reflected some uncertainty around the reliability of the survey-based coefficient estimates, the precision of the econometric model and the strength of the correlation between private sector and public sector wages. Discounted wage cost factors are shown in Figure 1.

Figure 1   Wage cost factors for states, 2016-17 to 2021-22
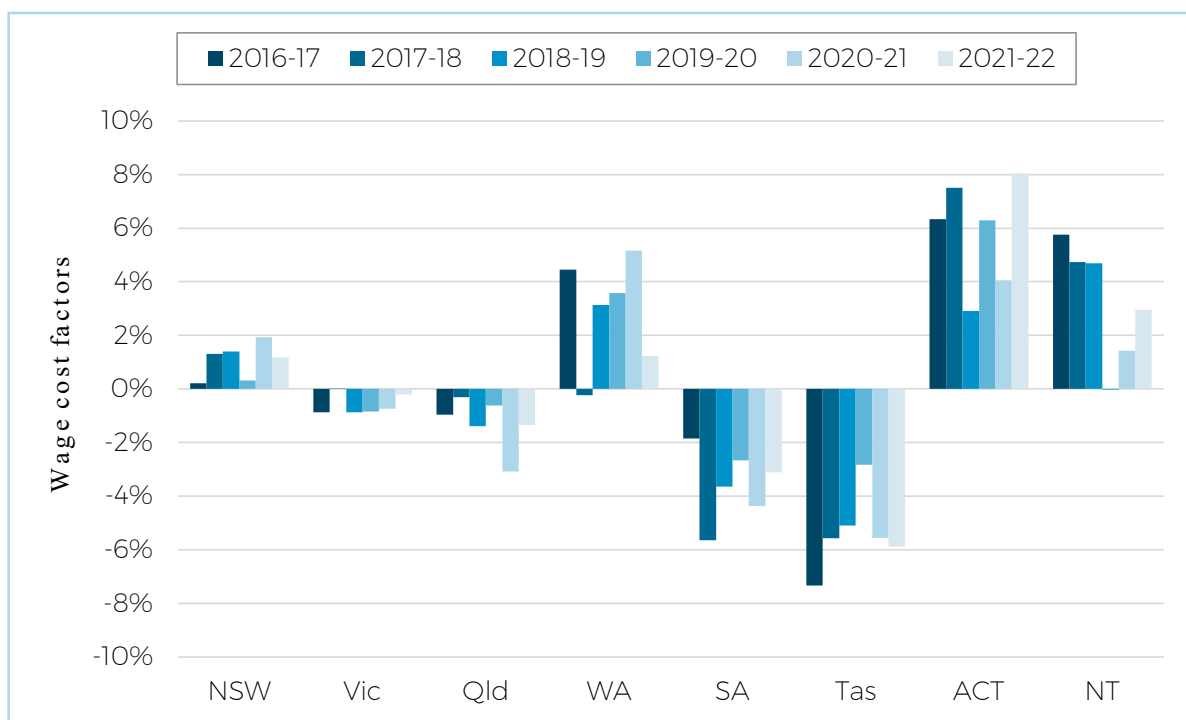
Note: A 12.5% discount has been applied.
Source: Commission modelling based on the Characteristics of Employment survey.

Data used in the assessment

11   The assessment uses data from the ABS Characteristics of Employment survey, which is conducted each year in August as a supplement to the monthly Labour Force Survey.

12   These data provide individual wage income linked to personal characteristics and job characteristics including education level, age, sex, marital status, number of dependent children, migrant status, permanent/casual status, hours worked, tenure in current job, and industry and occupation of employment. This allows for the effects on wages of all those other attributes to be controlled for when measuring the state level geographic effect.

13   Only some of the data are used in the model. The sample is restricted to only include private sector, wage earning employees with valid responses to questions on the key explanatory variables outlined above. The survey includes approximately 50,000 respondents, including 15,359 in scope of the regression model.

Applying wage costs to expenses

14   For the 2020 Review, the proportion of expenses attributed to wage costs for each expense category was estimated as the proportion of direct costs that were wage related. This includes wages, bonuses, superannuation costs, fringe benefits and workers compensation. This proportion was calculated as an average over 2015–16 to

2017–18 and fixed for the entire 2020 Review period. The discounted wage factors are applied to the relevant proportion of expenses in each category.

15   In housing, transport and roads a significant amount of labour costs is classified as other expenses, such as payments to contractors. The proportion of labour costs in these assessments is unclear, so the wages proportion for these categories was estimated as the average proportion for all other categories. Table 1 shows the proportion of direct expenses related to wages in each category, as well as the proportion of total expense in each category that is assessed to be wage related.

Table 1   Wage cost proportions of assessment categories

Category	Proportion of expenses related to wage costs	Assessed proportion
Schools	79.5%	79.5%
Post-secondary education	55.3%	55.3%
Health	65.4%	65.4%
Housing	39.0%	63.3%
Welfare	36.4%	36.4%
Services to communities	44.9%	44.9%
Justice	72.4%	72.4%
Roads	29.6%	63.3%
Transport	12.7%	63.3%
Services to industry	47.0%	47.0%
Other expenses	45.0%	45.0%

Source: Commission calculation.

16   The wages factor is also used as an input into the investment assessment. The relative costs of construction in the states are applied to investment. State wage costs are used in combination with the most current Rawlinsons construction cost guide for each assessment year to estimate relative construction costs in the states.[1] Further details on the use of wage costs in the investment assessment will be provided in the consultation paper for the investment assessment.

17   Table 2 shows the total expenses affected by the wage costs assessment.

Table 2   Wage expenses

	2018-19	2019-20	2020-21	2021-22
Wage expenses ($m)	144,210	153,905	165,695	184,650
Proportion of total expenses (%)	64.3	64.0	63.6	63.2

Source: Commission calculation using 2023 Update estimates.

---

[1] In the 2023 Update the Australian Construction Handbook 2020 Edition 38, 2021 Edition 39 and 2022 Edition 40 were used for the 3 assessment years.

COVID-19 adjustments

18   In the 2023 Update there were concerns over the Characteristics of Employment survey data from August 2021. Several major cities were subject to COVID-19 public health orders restricting work during the reference period of the survey. The Commission decided to use a measure of hours worked in the pay period when explaining wages for the 2021–22 year, rather than usual hours of work.

19   Details of how and why the wage costs assessment was adjusted in the 2022 and 2023 Updates due to COVID-19 related issues are explained in New issues in the 2022 Update and New issues in the 2023 Update.

GST Distribution in the 2023 Update

20   Table 3 shows the extent to which the assessment results in a different distribution of GST compared with an equal per capita distribution. In the 2023 Update the assessment distributed $1.6 billion ($61 per capita) away from an equal per capita distribution.

Table 3   GST impact of the wage costs assessment

	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Total Effect
$ million	658	-277	-655	713	-461	-232	190	64	1,625
$ per capita	80	-41	-121	251	-249	-397	404	246	61

Note: Excludes the assessed wage contribution to construction costs
Source: Commission calculation, 2023 Update.

21   Further detail on the underlying conceptual cases for the assessment methods and the scope of the adjusted budget are explained in Volume 2 Chapter 27 – Wage costs of the Report on GST Revenue Sharing Relativities, 2020 Review.

What has changed since the 2020 Review?

The labour market has continued to change

22   Australia's labour market has continued to evolve, including as a result of the impact of COVID-19. Changes include:

- increases in online and remote work[2]

    – In August 2021, 41% of Australians regularly worked from home, up from 32% in 2019.[3] This was partly due to the temporary effects of COVID-19, but also reflects an existing trend towards more remote work.

- low unemployment and labour shortages

---

[2] (National Skills Commission, 2021)

[3] (ABS, 2021)

- The Australian labour market tightened rapidly after the onset of COVID-19. This occurred across most segments of the economy, especially in higher skilled occupations, including those that are well represented in the state public sector. There were 37 occupations in the health profession that were assessed to be in shortage in 2022 that had not been in 2021. This is the largest increase of any professional group. The number of suitable applicants per advertised vacancy for education professionals more than halved between 2021 and 2022.[4]

- an increase in demand for post-secondary qualifications[5]

  - More than 9 out of 10 new jobs to be created in the 5 years to November 2026 will require post-secondary qualifications. The jobs projected to have the largest increases in employment are aged and disabled carers, software and applications programmers, and registered nurses.[6]

- slightly elevated levels of industrial action

  - More working days were lost to industrial action in 2022 than in any year since 2012, potentially reflecting increased bargaining power due to skills shortages. However, the level of industrial action is low in a longer-term context.[7]

- falling real wages.

  - Since 2020, inflation has increased, with CPI rising 3.5% in 2021 and 7.8% in 2022.[8] As nominal wage growth has been substantially below this level, this has led to falling real wages. Nominal wage growth has increased, growing 3.6% in 2022.[9]

Increased volatility in the wage cost assessment

23   Since the ABS produced data enabling annual estimates of relative state wage levels in 2014–15, the assessment has been quite volatile. This volatility does not align with recognised stability of wages. Wages are known to be 'sticky' and slow to respond to shocks.

24   Volatility in the assessment has become increasingly apparent as the time series has extended. Before annual data were available the wage cost factors were calculated every 5 years, not necessarily during reviews, and indexed using the wage price index. This led to a pattern of stability between most updates, with major revisions limited to when new data became available every 5 years.

25   Sampling error can explain most of the assessment's volatility. While updating the estimates annually has the benefit that any unusual estimates are only applied to a

---

[4] (National Skills Commission, 2022)

[5] (National Skills Commission, 2021)

[6] (National Skills Commission, 2022)

[7] (ABS, 2022)

[8] (ABS, 2023)

[9] (ABS, 2023)

single assessment year, this approach has highlighted the level of sampling error associated through the variation in annual estimates.

26 Sampling error can be reduced by expanding the sample size. As wages are relatively stable, estimates from several years can be combined, increasing the effective sample size, reducing sampling error and increasing the accuracy of the estimates. This is discussed further below.

The data environment has improved

27 The ABS has improved access to data since the 2020 Review. In 2022, the ABS released the Characteristics of Employment survey data into its data laboratory. This gave the Commission the ability to interrogate and analyse the data directly.

28 The regression model the Commission uses to estimate differences in state wages relies on 'Hours usually worked' as an important control variable, or predictor of wages. In the 2023 Update, the impact of COVID-19 on the survey data revealed that the dependence on 'Hours usually worked' has the potential to introduce bias into the model. When hours of work were reduced during lockdowns this was captured in the model as reduced wages in lockdown affected states. It is possible that other local economic shocks affecting hours of work, such as from natural disasters, could create a similar bias in Commission estimates of relative wages. Increased access to the unit record data allowed the Commission to identify this issue in the model.

29 Since April 2020, the ABS has produced a data series on weekly payroll jobs and wages, using single touch payroll data. From 2023, the ABS will be using single touch payroll data for their Public Sector Employment and Earnings data series.

30 The Multi-Agency Data Integration Project links administrative data from the Australian Taxation Office and the Department of Social Services to census data. There are also other surveys providing information on wages and individual characteristics associated with earnings potential, such as the Household Income and Labour Dynamics in Australia survey.

31 The increasing availability of data sources, especially the linked administrative data, provides opportunities to investigate alternative data for the wage costs assessment.

Implications for assessment

32 The Commission has identified 4 issues for consideration:
- whether, and to what extent, labour market changes challenge the conceptual basis for the wage costs assessment
- whether the accuracy of the assessment can be improved, and volatility reduced
- whether changes to data used in the assessment would make it more resilient to shocks
- whether to continue to discount the wage costs assessment.

Changes in the labour market and the conceptual basis for the wage costs assessment

33   While there have been considerable changes in Australian labour markets since the 2020 Review, the Commission has not identified any changes that undermine the conceptual basis of the wage costs assessment.

34   As an example, a general tightening of the labour market, including among occupations prevalent in the public sector, is likely to increase pressure on public sector wages. However, the Commission is not aware of any evidence that these effects will be felt more strongly in any state in a way that is not also reflected in that state's private sector wages.

35   The Commission's preliminary view is that, notwithstanding the changes occurring in the labour market, the underlying conceptual basis of the wage costs assessment remains sound.

Consultation question

Q1.   Do states agree that the underlying conceptual basis for the wage costs assessment remains sound?

Improving accuracy and reducing volatility

36   Estimates of relative wage levels of states using the Characteristics of Employment surveys are volatile and imprecise due to the sampling error inherent in surveys.

37   Actual wages are relatively stable over time. The relative wage level of a state in previous years is a good indicator of the likely relative wage level in the current year. It is possible to effectively increase the sample size used in the wage costs assessment by using results from surveys in other years. This would increase the accuracy and reduce the volatility of the assessment.

38   Annual estimates using the Characteristics of Employment surveys can be indexed using the wage price index to account for any differences in wage growth between the states, before being combined to produce an estimate for each assessment year.

39   Annual estimates of relative state wage levels, with 95% confidence intervals, are shown in Figure 2 (survey estimates) and Figure 3 (combined estimates). These represent Commission estimates for relative state wages in the 2023 Update.

40   Each annual estimate in Figure 3 uses the survey estimates from 7 years of data, from 2016–17 to 2022–23. Survey estimates are indexed using the wage price index and are given a lower weight the greater the time elapsed between the survey and the year being estimated. Survey estimates also have higher weights if the standard error for that estimate is lower.

41   This approach means later data can be incorporated into the Commission's estimates. For example, ABS processing means that preliminary weighted estimates of the 2022–23 relative wage levels (based on a survey run in August 2023) could have been used in estimating the factor for 2021–22 and earlier assessment years in the 2023 Update. Under the current approach, these latest data are not used until the following update.

42   For a full description of the proposed method for combining indexed estimates, see Attachment B.
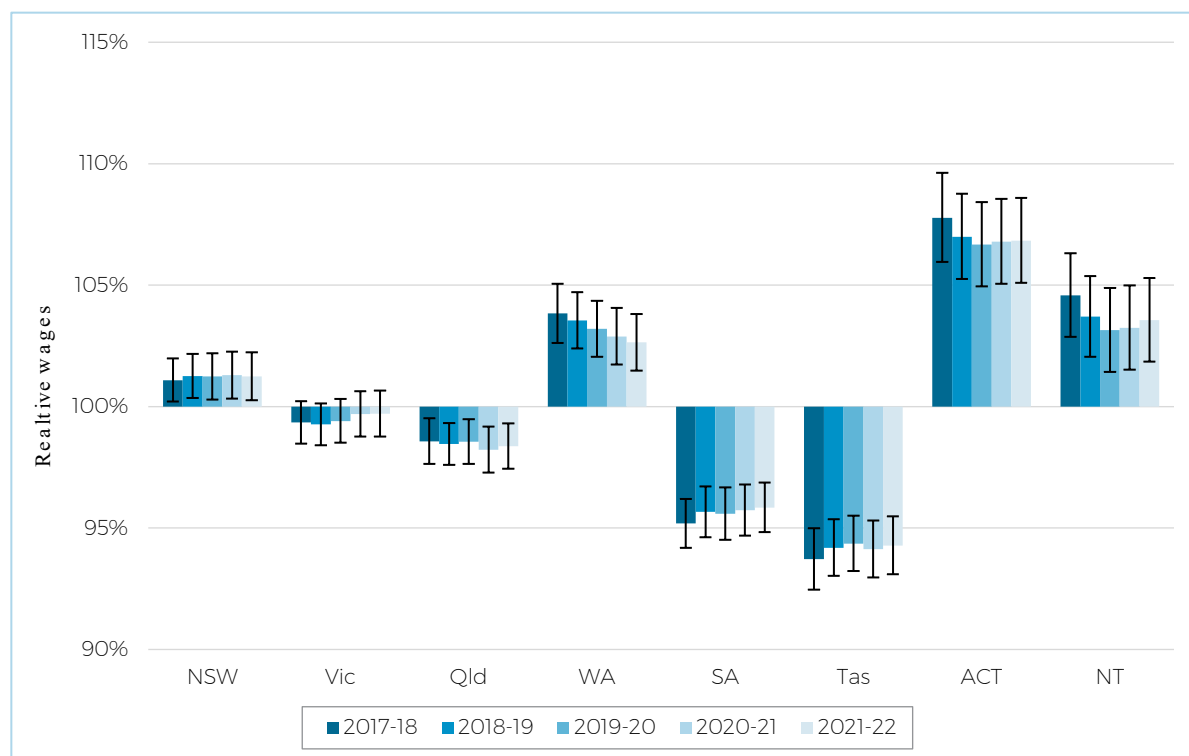
Figure 2   Annual survey estimates of relative state wage levels, 2020 Review methods
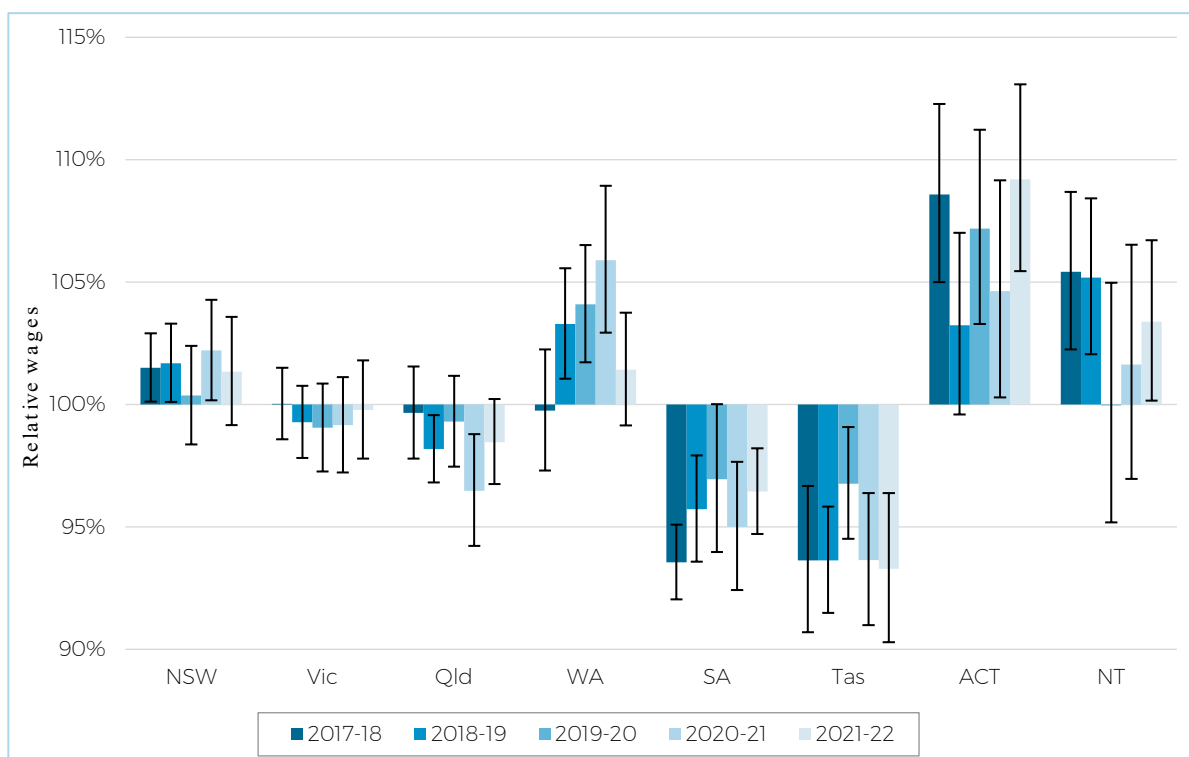
Notes: Annual survey estimates using the methods outlined in the 2020 Review, the 2022 Update (new issues) and 2023 Update (new issues). Error bars show 95% confidence intervals.

Source: Commission calculation.

Figure 3   Annual combined estimates of relative state wage levels

Notes: Annual combined estimates using fixed effects method to combine survey estimates from 2016–17 to 2022–23. Error bars show 95% confidence intervals.

Source: Commission calculation.

43   Each annual estimate in Figure 3 uses all the available survey estimates, including those generated by data from later than the assessment year to which they are applied. In the 2023 Update, the 2020–21 estimate would have been based on survey data from 2015–16 to 2021–22. In the 2024 Update, the 2022–23 survey would also contribute. This means the wage costs assessment would include revisions.

44   Using only data from current and previous years to generate estimates for each assessment year would remove the prospect of revisions. However, it would also reduce the reliability of the assessment by not using all available data.

45   In the 3 updates between the 2020 Review and the 2023 Update, the wage costs assessment led to a change in GST of more than $50 per capita for a state in 8 instances, including 3 changes of more than $100 per capita. Under the proposed approach, there would have been 4 changes of more than $50 per capita and none of more than $100. As the number of surveys contributing to this approach increases, this volatility is expected to reduce further.

46   The Commission's preliminary view is to estimate wage costs using survey results from all years since 2016–17, and to revise assessment year data as more data become available.

Increasing the resilience of the wage costs assessment to shocks and other method improvements

47   The effect of lockdowns on state specific labour markets highlighted a potential bias with the 2020 model and its ability to deal with temporary shocks. Using paid hours rather than usual hours, as per the 2023 Update, does mitigate this.

48   While that adjustment significantly reduced the bias in the model, a model predicting hourly wage rather than weekly wage would eliminate sensitivity of wage estimates to short-term changes in hours worked. Retaining additional variables related to usual and paid hours as well would better reflect the complex relationship between hours worked and pay.

49   Since the 2020 Review, the ABS has provided access to the data underlying the Commission's model. This has enabled the Commission to thoroughly review the detailed specifications in a way that was not possible in previous reviews. In light of this opportunity, and the need to remedy the way hours of work are captured in the model, the Commission has reviewed all the coefficients specified in the model.

Estimating hourly rather than weekly wages

50   The correlation between hours worked and weekly earnings can be separated into 2 effects:

- Working for an additional hour earns an additional hour's pay. In this way, hours worked directly affects weekly pay at a constant hourly wage rate.

- Working more hours reflects individual or job-related characteristics likely to affect pay, such as the number of hours of experience in a position, or an implicit expectation to regularly work unpaid overtime. These indirect effects are above and beyond the direct effect and will affect hourly pay.

51   The change to a model predicting log of hourly wages, instead of log of weekly wages, will capture the direct effect. The indirect effect is complex, with both usual hours and paid hours having strong, separate, and non-linear relations. The proposed model specified in Table 4 better reflects this relationship (the R squared value increases considerably) and does not have arbitrary breaks between those working 15 and 16 hours or between 59 and 60 hours.

52   There is a strong conceptual case that both hours variables have an independent relationship to wages. Working longer usual hours increases hourly pay, with diminishing marginal returns. This is consistent with both some acknowledgement of unpaid overtime in people's pay as well as more shifts offered to more valuable

workers. Increased paid hours decreases hourly pay at a decreasing rate. This could reflect that people who work a penalty rate shift tend to work fewer hours than those working standard-work weeks.

53   In periods of economic shock, including lockdowns, paid hours can differ significantly from usual hours. Allowing for the complex relationship between these related concepts should maximise the capacity of the wage costs assessment to reliably estimate unbiased state wage differentials.

54   The Commission's preliminary view is to use hourly rather than weekly wages as the dependent variable, and to use both hours paid and usual hours worked as explanatory variables, as shown in Table 4. Attachment C describes the model in more detail.

Table 4   Proposed changes to specification of hours in the model

	R2020	Proposed
Dependent variable	Log of weekly wage	Log of hourly wages [a]
Independent variables	Log of usual hours	Usual hours
	Log of usual hours if usual hours <16	Paid hours
	Log of usual hours if usual hours >59	Usual hours squared
		Paid hours squared
		Usual hours * Paid hours

(a)   based on paid hours.
Source: Commission decision.

Consultation question

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?

Education and age interactions

55   The decline of the traditional organisational career has led to increased organisational, occupational, and industrial mobility. This has implications for the causal effect of work experience on earnings, as individuals do not move vertically through the promotion path determined by their employer over their lifetime.

56   To facilitate a career transition, individuals may undergo further education. If so, education may have a more complex relationship to earnings, as different individuals undergo education at different points in their career progression.

57   Historically, the Commission has modelled work experience as current age less an estimate of the worker's age when they completed their education. This approach

involved strong assumptions regarding the time spent on educational attainment and time spent in full-time employment.

58　This approach is also dependent on the assumption that work experience-earnings profiles are parallel across different education levels, which is not supported in the empirical literature.[10]

59　There is international[11] and Australian[12] evidence to suggest that individuals' returns to schooling vary with their ages. This can be addressed through the interaction of education with work experience variables, which estimates a unique age-earnings profile for each education level, shown in Figure 4.[13] These variables also have a gender interaction term. The relationship for males is similar to the female relationship shown Figure 4.

Figure 4　　Estimated education-dependent age-earnings profiles, females
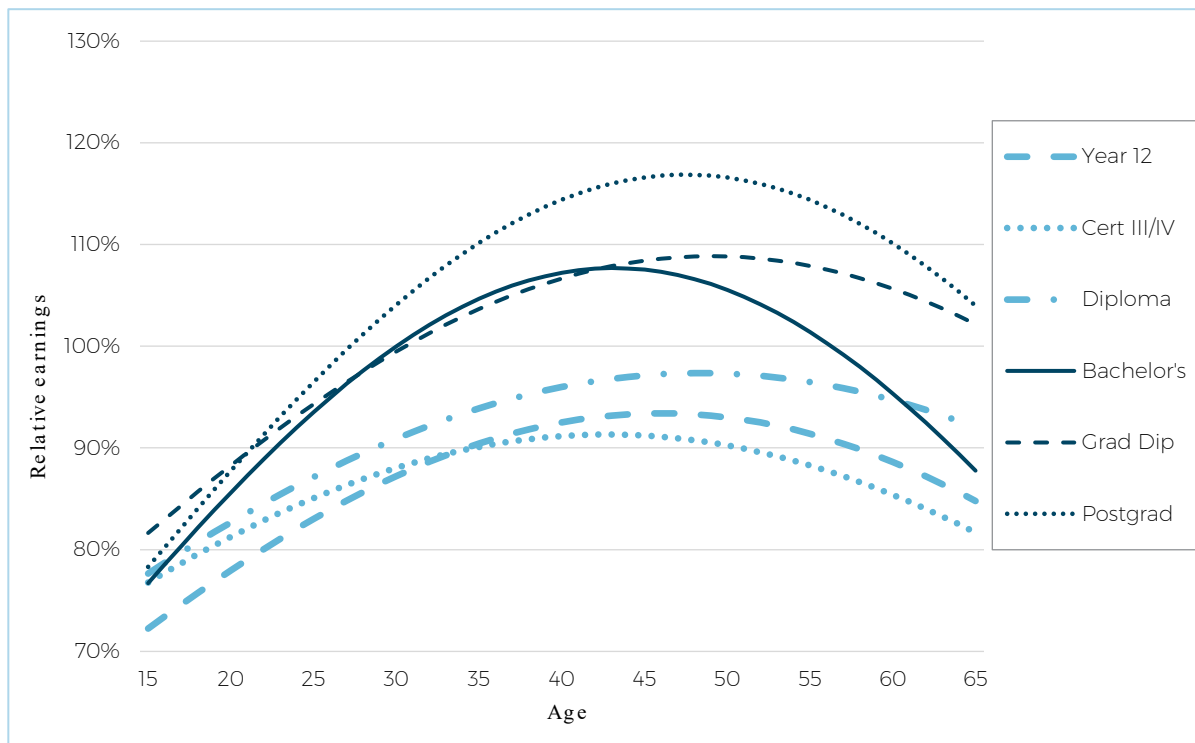
Note:　　The graph above shows the effects of age and education on the wages of women who are otherwise similar.
Source: Commission calculation.

60　The Commission's preliminary view is to replace the imputed experience variable with an education dependent age-earning profile. This specification explains more of the individual level differences in wages. Attachment C describes the model in more detail.

---

[10] (Heckman, Lochner, & Todd, 2003)

[11] (Bhuller, Mogstad, & Salvanes, 2017)

[12] (Perales & Chesters, 2017)

[13] (Card, 1999)

Q4. Do states agree the Commission should replace the derived work experience variable with interacting variables of age and level of education?

Including job tenure as a continuous variable

61 The Commission currently models job tenure as a categorical variable with 4 ranges. Modelling tenure as a continuous variable is standard practice unless the relationship between wage and tenure is found not to be linear. It minimises the arbitrary discontinuity of categorical variables and produces a simpler model with increased explanatory power.

62 The Commission's preliminary view is to replace the 4 categorical variables on tenure with a single variable on years of tenure. This specification explains more of the individual level differences in wages. Attachment C describes the model in more detail.

Q5. Do states agree the Commission should treat job tenure as a continuous variable?

Discounting of the wage costs assessment

63 The current wage costs assessment includes a 12.5% discount, reflecting some uncertainty around the reliability of the survey-based coefficient estimates, the precision of the econometric model and the strength of the correlation between private sector and public sector wages.

64 The Commission's preliminary views, outlined in this paper, include changes designed to reduce volatility and improve the reliability of the assessment. While the wage costs assessment is likely to continue to involve some level of uncertainty, the proposed improvements raise the question of whether a continuation of the 12.5% discount is necessary.

65 The Commission's preliminary view is that, while proposed changes improve reliability and reduce volatility, a low level of uncertainty attributable to other aspects of the assessment remains, particularly the use of private sector wages as a proxy for public sector wage costs. As such, the low-level discount of 12.5% remains appropriate.

Q6. Do states agree that a 12.5% discount remains appropriate?

Proposed assessment

Differences from the 2020 Review approach

66   Subject to state views the Commission proposes 2 changes:

- combining annual survey estimates to increase the reliability and reduce the volatility of estimated relative state wage levels

- changing the specification of the regression model to increase the accuracy and robustness of survey estimates.

67   The Commission proposes to use the full time series of available survey estimates of relative state wage costs, beginning from 2016–17, to estimate relative wage costs in each assessment year. These estimates would be generated by indexing and weighting the estimates from each contributing year as outlined in Attachment B.

68   The Commission proposes to make the following changes to the model specification:

- use hourly rather than weekly wages as the dependent variable

- include as independent variables both 'Hours usually worked' and 'Hours from last payslip', with quadratic and interaction terms

- replace the constructed work experience variable with an interaction between highest education level and age

- include job tenure as a continuous variable.

New data requirements

69   No new data will be required from states.

70   The wage cost proportions for each expense category will be recalculated from the ABS' Government Finance Statistics data, following the same method as the 2020 Review.

Consultation

71   The Commission welcomes state views on the consultation questions identified in this paper (outlined below) and the proposed assessment. State submissions should accord with the 2025 Review framework. States are welcome to raise other relevant issues with the Commission.

72   States will be given the opportunity to separately provide input on Professor Preston's consultancy. As outlined in paragraph 6, states will receive the consultant's report in August 2023 and state submissions on the wages costs assessment paper should be with the Commission by 13 November 2023.

Q1. Do states agree that the underlying conceptual basis for the wage costs assessment remains sound?

Q2. Do states agree with the proposed approach to combine estimates of relative differences in states wages across years?

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?

Q4. Do states agree the Commission should replace the derived work experience variable with interacting variables of age and level of education?

Q5. Do states agree the Commission should treat job tenure as a continuous variable?

Q6. Do states agree that a 12.5% discount remains appropriate?

References

ABS. (2021, December). Working arrangements. Canberra. Retrieved from abs.gov.au.

ABS. (2022, December). Industrial Disputes, Australia. Canberra.

ABS. (2023). *Consumer Price Index, Austalia.* Canberra: Australian Bureau of Statistics.

ABS. (2023). *Wage price index, Australia.* Canberra: Australian Bureau of Statistics.

Bhuller, M., Mogstad, M., & Salvanes, K. G. (2017). Life-Cycle Earnings, Education Premiums, and Internal Rates of Return. *Journal of Labor Economics*, 993–1030.

Card, D. (1999). The Causal Effect of Education on Earnings. In O. C. Ashenfelter, & D. Card, *Handbook of Labor Economics* (Vol. 3A, pp 1801–1863). Elsevier Science B.V.

Combes, P.-P., & Gobillon, L. (2015). The Empirics of Agglomeration Economies. In *Handbook of Regional and Urban Economics, Volume 5* (pp 247–348). Elsevier.

Combes, P.-P., Duranton, G., & Gobillon, L. (2008). Spatial Wage Disparities: Sorting Matters. *Journal of Urban Economics*, 723–742.

Fox, J., & Monette, G. (1992, March). Generalized Collinearity Diagnostics. *Journal of the American Statistical Association, 87* (417), 178–183. doi:10.2307/2290467

Groot, S., de Groot, H., & Smit, M. (2014). Regional Wage Differences in the Netherlands: Micro Evidence on Agglomeration Externalities. *Journal of Regional Science*, 503–523.

Heckman, J. J., Lochner, L. J., & Todd, P. E. (2003). Fifty Years of Mincer Earnings Regressions. *NBER Working Paper Series*, 9732-.

Kirk, I. (2022, February 24). *YouGov*. Retrieved from YouGov: https://yougov.co.uk/topics/international/articles-reports/2022/02/24/how-has-covid-19-pandemic-changed-attitudes-workin

Murillo Huertas, I. P., Ramos, R., & Simòn, H. (2020). Revisiting interregional wage differentials: New evidence from Spain with matched employer-employee data. *Journal of Regional Science*, 296–347.

National Skills Commission. (2021). *The Transformation of Australia's Labour Market.* Canberra: National Skills Commission.

National Skills Commission. (2022). *2022 Skills Priority List.* Canberra: National Skills Commission.

National Skills Commission. (2022, March 29). *Projecting employment to 2026.* Retrieved from National Skills Commission: https://www.nationalskillscommission.gov.au/insights/projecting-employment-2026

Perales, F., & Chesters, J. (2017). The returns to mature-age education in Australia. *International Journal of Educational Research*, 87–98.

Vyas, L. (2022). "New normal" at work in a post-COVID world: work-life balance and labor markets. *Politics and Society,* Volume 41, Issue 1, 155–167.

Webster, E., & Tseng, Y.-P. (2000). The Determinants of Relative Wage Change in Australia. *Melboune Instiutute Working Paper Series*.

Attachment A: Model validation

73   There are several key questions regarding the regression model in this attachment:

- Does the model detect and measure a significant difference in wage levels between states?
- Is the choice of control variables appropriate. In particular, is there an issue with multicollinearity or overfitting affecting the reliability of estimated state wage levels?
  - If multicollinearity is a problem, then the variances of estimates will be inflated and estimates will be unreliable.
  - If the model has been overfitted then the coefficients represent the individuals in the sample only, rather than the population as a whole.

Significance

74   State coefficients represent the effect on wages attributed to state of usual residence. They are constructed using effects coding, which gives the intercept the interpretation as the grand mean, or the average of all states. State coefficients are then each state's deviation in wage levels from the national average. If a state coefficient is significantly different from zero, this implies this state has significantly different wage level effects than an average state.

75   Figure A1 shows that estimated wage levels for South Australia are significantly below the grand mean in every survey year, and estimated wage levels for Western Australia and the ACT are significantly above the grand mean in every survey year when using the proposed model. Some individual states in some years are not significantly different to the national average, reflecting that those states have wage levels moderately close to average.

Figure A1   Annual survey estimates of relative state wage levels
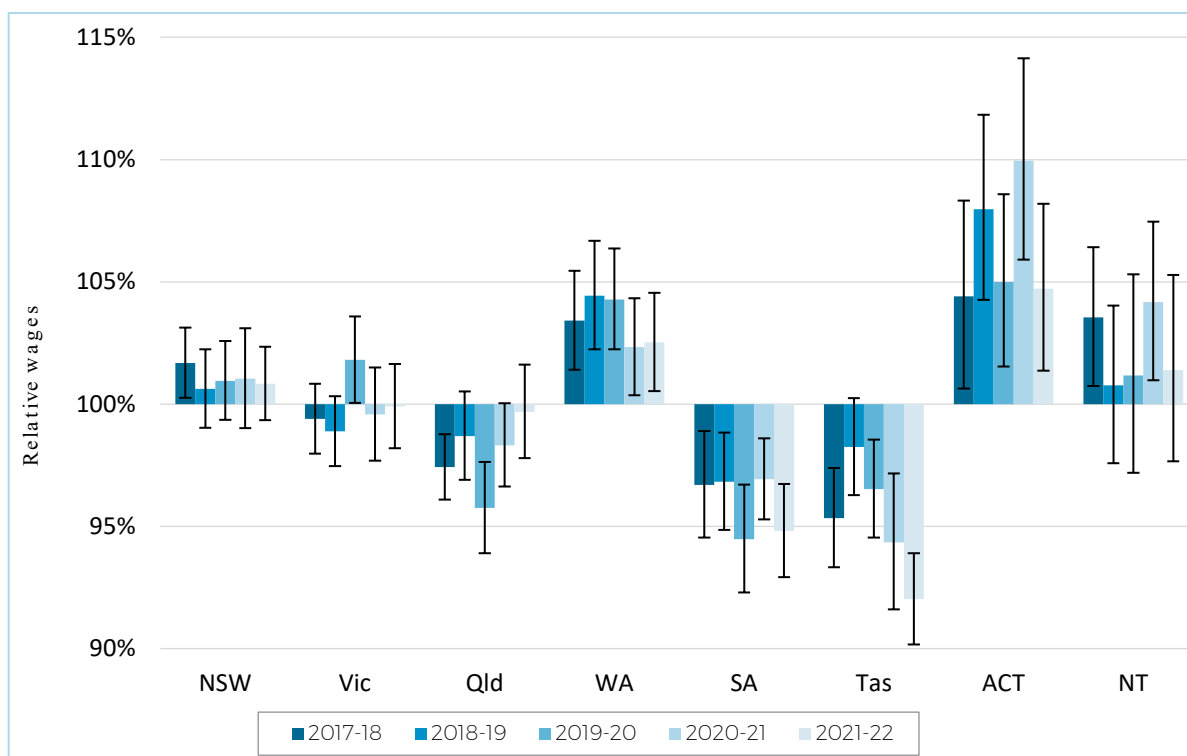
Notes: Annual survey estimates using the methods proposed in this paper. Error bars show 95% confidence intervals.
Source: Commission calculation.

Multicollinearity

76   Multicollinearity may affect the precision, and therefore reliability, of coefficient estimates in a model. When 2 or more variables are highly correlated, it is difficult to separate the effect of one variable from the other. While the joint effect between these variables is robust, small changes in the sample data can vastly change the estimates for each coefficient. This inflates the variance of affected coefficients.

77   Some variables in the model, such as related occupations and industries, are multicollinear with each other. This does not affect the validity of state coefficient estimates and can be ignored, so long as they are not multicollinear with state.[14] For example, the model has difficulty determining whether truck drivers in the road freight industry have high wages (given their education and other attributes) because of their occupation or because of their industry.

78   In the Commission's model, multicollinearity is only a concern if state coefficients are multicollinear with control variables. This is because the state coefficients represent the effects the Commission is attempting to measure accurately. For instance, if one state were largely comprised of a particular industry, and that

---

[14] Wooldridge, J. M. (2009). *Introductory Econometrics: A Modern Approach* (4th ed.). Mason: South-Western Cengage Learning, pp 95–99.

industry primarily existed within that state, it would be difficult to separate the industry effect from the state effect. This is not the case.

79   Intuitively, the state coefficients must predict the variance in wages that are not explained by control variables. Multicollinearity only harms the decomposition of causal effect and does not affect prediction. Therefore, so long as the effect of control variables can be separated from state effects, multicollinearity is not a concern.

80   Generalised variance inflation factor (GVIF) tests for multicollinearity between a categorical variable of interest (such as state) and all control variables in a model.[15] GVIF is typically standardised to account for the number of degrees of freedom taken up by the variable, in this case the number of state coefficients. The lower bound is 1, which indicates no multicollinearity, while values around 5 or higher indicate some multicollinearity, and values above 10 indicate severe multicollinearity.

81   These standardised GVIF values are shown in Table A1, for the specifications used in updates and the proposed specification. All values in each model are close to one, which indicates the precision of state coefficients in both models are unaffected by multicollinearity.

Table A1   Standardised GVIF values of state variables

Model	Sector	2018	2019	2020	2021	2022
2020 Review methods	Private	1.246	1.254	1.262	1.267	1.257
	Public	1.358	1.315	1.354	1.335	1.348
Proposed model	Private	1.247	1.254	1.263	1.268	1.258
	Public	1.363	1.322	1.361	1.343	1.355

Source: Commission calculation.

Overfitting

82   A regression with as many observations as independent variables will perfectly predict the dependent variable. This will not necessarily reflect the underlying relationship in the population, but rather the specific pattern of the sample. While the specific value for when overfitting becomes problematic, it is generally regarded as being avoided if a model has more than 10 times as many observations as variables.[16] The wage costs regression has more than 20 times as many observations as variables in every survey year.

83   Stability of the estimates across alternative samples can also provide some confidence that there is no problem of overfitting. When applying the proposed

---

[15] Fox, J., & Monette, G. (1992). 'Generalised Collinearity Diagnostics', *Journal of the American Statistical Association*, 87 (417), 178–183.

[16] Norman, G. R., & Streiner, D. L. (2014). *Biostatistics: The bare essentials* (4th ed.) Shelton, Connecticut: People's Medical Publishing House.

model independently to 5 annual surveys, the estimated coefficients are generally relatively stable and consistent. For example, Figure A1 shows that the estimates for states are generally not significantly different to estimates from previous years. Coefficients for education, hours, migrant status and most occupations and industries are generally comparatively stable. Coefficients for some industries and occupations with very few observations do vary widely, but that is to be expected given the small numbers.

Attachment B: Combining estimates

84   The proposed method for combining survey estimates from multiple years is a fixed effects method. This method is appropriate for combining estimates of the same effect using the same methods but different samples.

85   Combined estimates are generated for each assessment year, using data from every survey year. To ensure that estimates from different survey years are comparable, the survey estimates are indexed to the assessment year using the wage price index.

86   The combined estimate for an assessment year is then a weighted average of these indexed estimates. The weight for each survey estimate is the inverse of the variance of the indexed estimate, adjusted to provide a 'penalty' for the amount of time between the survey year and the assessment year.

87   Figure B1 shows the average annual weights used to create the combined estimates in Figure B3. Survey estimates are weighted according to their variance. Estimates with lower variations are given relatively higher weights, leading to the irregular patterns. Figure B2 shows the weights for state estimates used to create the averages for 2020–21 in Figure B1. The variance for smaller states associated with survey estimates are much more inconsistent between years than for larger states.

Figure B1   Average weights for survey estimates in annual combined estimates
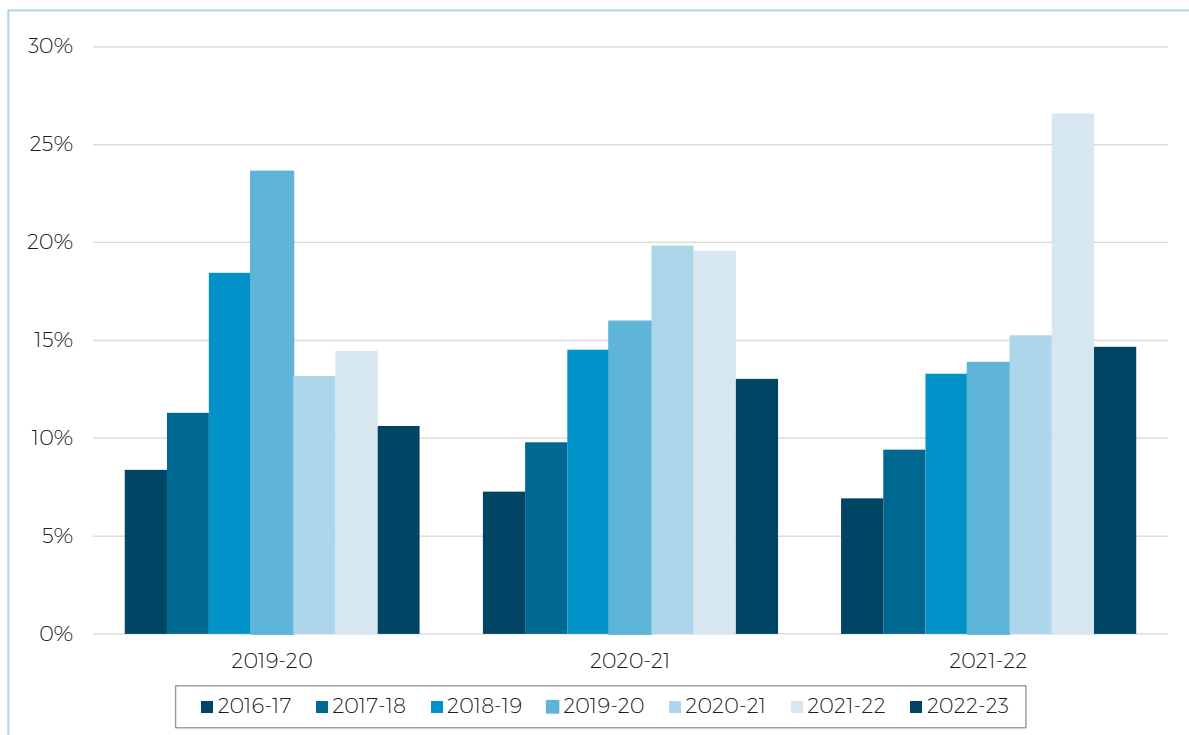
Source: Commission calculation.

Figure B2   Weights for survey estimates of 2020–21 state wage levels
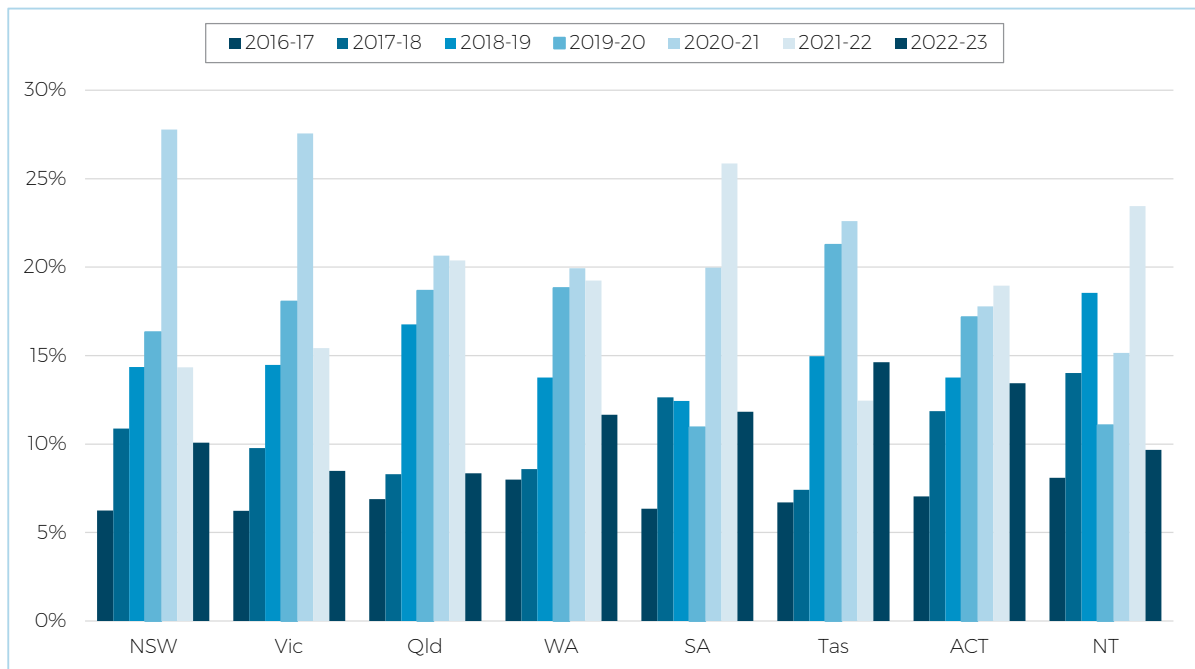
Source: Commission calculation.

88   No measure of variance or standard error is published for the wage price index. To ensure that estimates from different years are given appropriate weight, some measure of variance for the indexation must be assumed.

89   To penalise estimates where greater time has elapsed between the survey and the assessment year, the variance attributed to that wage price index is deliberately overestimated.

90   The annual variance between the wage price indexations for every state in every year is imputed as the variance of all the annual indexation factors. The imputed standard error for a state's change relative to the national average is 0.3%. This provides an upper limit of the plausible variance. This variance is combined with the variance of the survey estimates as though they were perfectly correlated, to overestimate any variance associated with indexations and penalise less contemporaneous data.

91   This total variance is used to generate the weights for each indexed survey estimate to calculate combined estimates for each assessment year.

92   Each combined annual estimate also now has an associated (maximum) variance – the inverse of the sum of weights of all the estimates used to calculate it. This allows the provision of a measure of confidence for the combined estimate, shown by the 95% confidence intervals in Figure B3.

Attachment C: Comparing model specifications

93   Models were validated by testing on a 5-year pooled sample, containing observations from 2018 to 2022. Year dummy variables account for inflation. Pooling the data may be inappropriate for accurate estimation of parameters, due to possible structural breaks. Here it was pooled simply to create a representative set of data for testing model specifications.

94   Changes are recommended on the basis of explanatory power when tested on the 5-year sample as well as simplicity and conceptual validity.

Measuring goodness of fit

95   Five statistics have been used to compare the performance of candidate models: R-squared, adjusted R-squared, log likelihood, AIC and BIC.

96   Adjusted R-squared measures the proportion of variation in the dependent variable which can be explained by variation in the independent variables, controlling for the number of variables in the model.

97   Log likelihood measures the degree to which the observed data reflects the function predicted by regression, where a higher value indicates better fit. This statistic has no interpretation in isolation; it is only useful when comparing between different models predicting the same variable from the same sample.

98   AIC and BIC combine likelihood values with a penalty for overfitting. They have the opposite interpretation, where a lower value indicates better fit. Like log likelihood, they only have an interpretation when comparing models.

Comparing the different specifications

99   The regression results for each proposed change in specifications are shown in Table C1. These all use pooled data and the usual hours variable for hours worked, unless otherwise indicated.

Table C1   Regression results for alternative specifications using pooled data

		2020 Model	Tenure	Experience	Paid Hours	Respecify Hours	Hourly wages	No log(hours)	All changes
(Intercept)		2.4718	2.4568	2.4642	2.6458	2.2585	2.2583	2.2462	2.2287
	s.e.	-0.9552	-0.9551	-0.9540	-0.8012	-0.7850	-0.7849	-0.7849	-0.7834
NSW		0.0027	0.0025	0.0027	0.0092	0.0079	0.0079	0.0078	0.0077
	s.e.	-0.0045	-0.0045	-0.0045	-0.0038	-0.0037	-0.0037	-0.0037	-0.0037
VIC		-0.0135	-0.0135	-0.0132	-0.0032	-0.0026	-0.0026	-0.0026	-0.0023
	s.e.	-0.0046	-0.0046	-0.0046	-0.0039	-0.0038	-0.0038	-0.0038	-0.0038
QLD		-0.0159	-0.0160	-0.0165	-0.0196	-0.0226	-0.0226	-0.0225	-0.0232
	s.e.	-0.0049	-0.0049	-0.0049	-0.0041	-0.0041	-0.0041	-0.0041	-0.0040
WA		0.0377	0.0378	0.0373	0.0284	0.0304	0.0304	0.0304	0.0301
	s.e.	-0.0059	-0.0059	-0.0059	-0.0049	-0.0048	-0.0048	-0.0048	-0.0048
SA		-0.0471	-0.0472	-0.0469	-0.0446	-0.0441	-0.0441	-0.0440	-0.0438
	s.e.	-0.0068	-0.0068	-0.0068	-0.0057	-0.0056	-0.0056	-0.0056	-0.0056
TAS		-0.0660	-0.0661	-0.0646	-0.0597	-0.0547	-0.0547	-0.0546	-0.0533
	s.e.	-0.0115	-0.0115	-0.0115	-0.0096	-0.0094	-0.0094	-0.0094	-0.0094
ACT		0.0619	0.0621	0.0617	0.0632	0.0640	0.0640	0.0640	0.0635
	s.e.	-0.0132	-0.0132	-0.0132	-0.0111	-0.0108	-0.0108	-0.0108	-0.0108
male		0.4394	0.4512	0.4748	0.0315	-0.0274	-0.0271	0.0883	0.0917
	s.e.	-0.6273	-0.6273	-0.6271	-0.5257	-0.5153	-0.5153	-0.5148	-0.5142
log(hours)		0.8648	0.8650	0.8650	0.8606	0.9900	-0.0100	-	-
	s.e.	-0.0077	-0.0077	-0.0077	-0.0050	-0.0127	-0.0127	-	-
log(hours < 16)		-0.0111	-0.0109	-0.0105	0.0132	-	-	-	-
	s.e.	-0.0046	-0.0046	-0.0046	-0.0034	-	-	-	-
log(hours > 59)		-0.0443	-0.0445	-0.0452	-0.0349	-	-	-	-
	s.e.	-0.0056	-0.0056	-0.0056	-0.0056	-	-	-	-
Usual hours		-	-	-	-	0.0203	0.0203	0.0202	0.0203
	s.e.	-	-	-	-	-0.0007	-0.0007	-0.0007	-0.0007
Paid hours		-	-	-	-	-0.0214	-0.0213	-0.0222	-0.0221
	s.e.	-	-	-	-	-0.0013	-0.0013	-0.0007	-0.0007
Usual hours^2		-	-	-	-	-0.0001	-0.0001	-0.0001	-0.0001
	s.e.	-	-	-	-	0.0000	0.0000	0.0000	0.0000
Paid hours^2		-	-	-	-	0.0001	0.0001	0.0001	0.0001
	s.e.	-	-	-	-	0.0000	0.0000	0.0000	0.0000
Male*usual hours		-	-	-	-	-0.0041	-0.0041	-0.0041	-0.0041
	s.e.	-	-	-	-	-0.0009	-0.0009	-0.0009	-0.0009
Male*paid hours		-	-	-	-	-0.0022	-0.0022	0.0035	0.0033
	s.e.	-	-	-	-	-0.0017	-0.0017	-0.0009	-0.0009
Num.Obs.		82,214	82,214	82,214	82,214	82,214	82,214	82,214	82,214
R2		0.619	0.619	0.620	0.732	0.743	0.419	0.419	0.422
R2 Adj.		0.615	0.615	0.616	0.729	0.740	0.413	0.413	0.416
AIC		127,655	127,643	127,405	98,793	95,402	95,400	95,426	95,068
BIC		135,332	135,265	135,343	106,470	103,135	103,133	103,141	102,988
Log.Lik.		-63,003	-63,004	-62,850	-48,572	-46,871	-46,870	-46,885	-46,684
RMSE		0.48	0.48	0.48	0.40	0.40	0.40	0.40	0.39

Note: All 8 models contain around 300 additional variables not shown here but described in Table 2.

Source: Commission calculation.

100 'Tenure' refers to the 2020 model but with the categorical tenure variables replaced with the single continuous tenure measure. 'Experience' refers to the 2020 model but with the imputed work experience replaced with interactions between age and education. 'Paid hours' replaces usual hours with paid hours in the original model. 'Respecify hours' replaces the discontinuity in the hours specification with linear, quadratic and interaction terms for both paid and usual hours. 'Hourly wages' is the same specification as 'Respecify hours' except with hourly wage rather than weekly wage as the dependent variable for the model. 'No log(hours)' is the same model as 'Hourly wages', with the log(hours) term dropped. 'All changes' is the final proposed model.

## Hourly wages specification and R-squared

101 Replacing the weekly wages with hourly wages as the dependent variable decreases R-squared, since a large amount of variation in weekly wage is explained by hours paid for. In fact, all variation in logarithmic weekly wage over and above logarithmic hourly wage is perfectly explained by the inclusion of a logarithmic hours paid for term. This does not change other coefficient estimates or predictive performance. This can be clearly seen comparing the 'Respecify hours' and 'Hourly wages' columns of Table C1. All the coefficients and standard errors are identical between the two models; only the R-squared has changed.

102 It is incorrect to assume that hourly wage models perform worse than weekly wage models. The decrease in R-squared can be interpreted as the variation in weekly wage over and above hourly wage being accounted for in the data construction, instead of by the model.

Improved hours specifications

103 Using hourly wage or weekly wage with log(hours) as a regressor can equivalently explain the direct link between hours worked and wages (by which hours worked only affects weekly wage). By instead predicting hourly wage, we are free to replace log(hours) with other functional forms which better explain secondary effects by which hours worked affects hourly wage.

104 The current specification of hours uses discontinuous terms typically used in sharp regression discontinuity and is equivalent to only including the interaction terms between log(hours) and high hours/low hours dummies, without including the main effects of high/low hours. There is no reason to believe the specification should use regression discontinuity rather than simple non-linearity, or that the dummies' main effects, the differences in intercept, should be excluded.

105 There is, however, a conceptual reason to assume the response to an additional hour of work differs for those working very few or many hours. For this reason, it is appropriate to include non-linear terms.

106   A simpler approach which improves explanatory power is to include linear, quadratic and interaction terms. As outlined, there are conceptual reasons to believe both 'Hours usually worked' and 'Hours covered by payslip' affect hourly pay at a decreasing marginal rate, and all linear and quadratic terms are significant. This has the benefit that only log-linear or log-quadratic response to covariates is seen in the model (as in the Mincer earnings function), and no log-log response.

107   The improvement of model fit between the 'Paid hours' and the 'Respecify hours' model in Table C1 demonstrates that this approach is a much better fit of the data. Once the model is switched to hourly wage as the dependent variable in 'Hourly wages' the log(hours) term can be dropped. Removing the log(hours) term has almost no detriment to the fit of the model, illustrated in the difference between the fit statistics for 'Hourly wages' and 'No log(hours)'.

108   The two measures of hours worked are both conceptually linked to wage levels, and measure distinct concepts. The following examples illustrate some of the potential mechanisms by which the two different aspects of hours worked can affect hourly pay, and hence why the Commission has used both aspects:

- Part-time workers accrue experience at a slower rate than full-time workers, leading to lower return to part-time work.

- People who work unpaid overtime may be more likely to be promoted or negotiate higher hourly wages.

- Casual employees who work shifts attracting penalty rates may tend to work fewer weekly hours than those who work standard shifts.

109   The net overall effect of these different mechanisms is difficult to predict.

110   Figure C1 shows that main effects for both hours variables are significant and stable between years. Being paid for more hours in a period is associated with lower hourly earnings, while usually working more hours is associated with higher hourly earnings. These effects are both more pronounced for females than for males.

Figure C1   Annual survey estimates of hours main effects
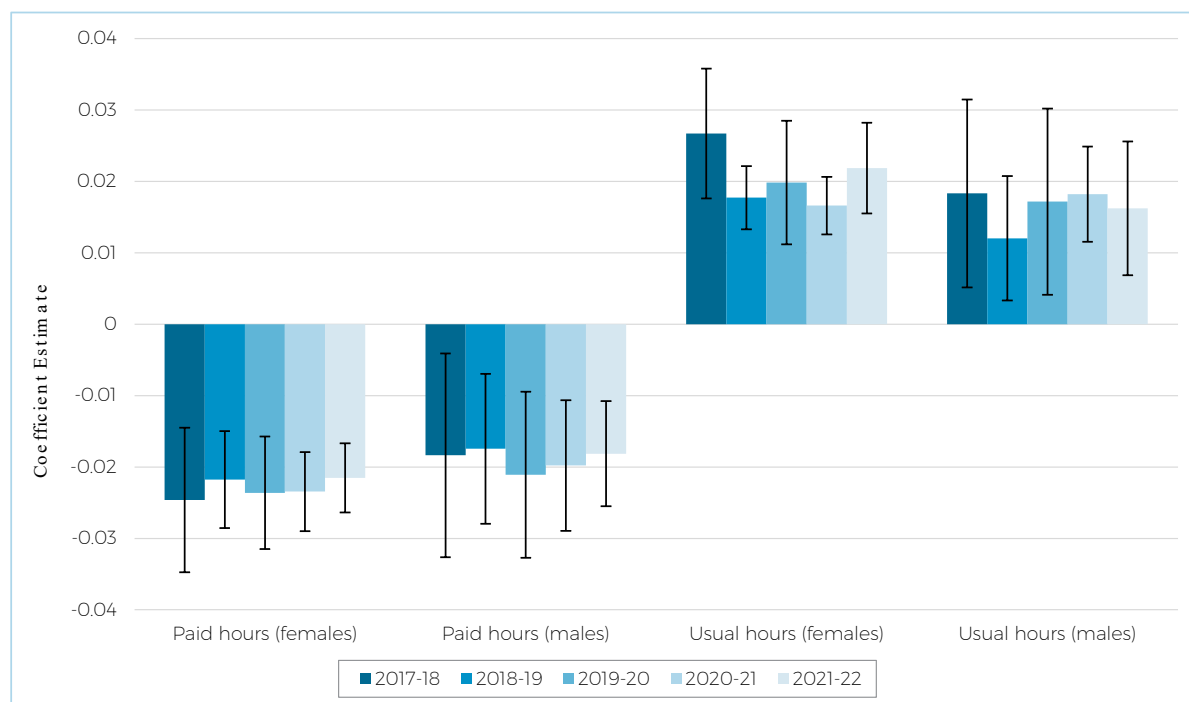
Note: Error bars show 95% confidence intervals.
Source: Commission calculation.

111   Coefficients for paid hours and for usual hours worked are consistent across each annual sample and the coefficients for the quadratic terms of both hours variables consistently reflect diminishing marginal effects. This indicates that these coefficients are measuring consistent effects and the specification chosen for hours worked is appropriate to predict the effect of hours worked on wage levels.

Other regressors

112   The proposed specification of education and work experience 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 quadratic), and sex. This predicts a quadratic age-earnings profile based on education level and sex, with returns beginning at the age of 15. This has 2 advantages over the 2020 specification. First, we no longer need to impute a separate variable for work experience based on assumptions about the individuals in the sample. Second, the fit of the model is improved based on all the diagnostic statistics.

113   Job tenure is included in the COES as a continuous variable. The current method constructs a categorical variable based on ranges which is parameterised with a set of dummy variables. This effectively removes information and increases the number of parameters. It is only of potential value if the wage response to tenure cannot be appropriately modelled with a continuous variable, that is if the error term is correlated with tenure, for linear and/or non-linear specifications.

114　The coefficient estimates for categorical tenure appear roughly linear, and the use of a linear term marginally increases model performance by all metrics, justifying the use of continuous tenure as given.

Table C2　Full model specification

	R2020	Proposed
Dependent variable	Log of weekly wage	Log of hourly wages
Variable of interest	State of usual residence	State of usual residence
Control variables	Log of usual hours	usual hours
	Log of usual hours if usual hours <16	paid hours
	Log of usual hours if usual hours >59	usual hours squared
		Paid hours squared
		Usual hours * Paid hours
	Education (7 categories)	Education (7 categories)
	Imputed work experience	Age minus 15
	Imputed work experience squared	(Age minus 15) squared
		Education*(age minus 15)
		Education*(age minus 15) squared
	Tenure (5 categories)	Tenure (continuous)
	Migrant status (7 categories)	Migrant status (7 categories)
	Marital status	Marital status
	Dependent child (dummy)	Dependent child (dummy)
	Occupation (~120 categories)	Occupation (~120 categories)
	Industry (~260 categories)	Industry (~260 categories)
	Male	Male
	Male*(every other control)	Male*(every other control)

Source: Commission decision.