NERA
ECONOMIC CONSULTING

Benchmarking of NERL Staff Pay

Prepared for National Air Traffic Services (NATS)

21 September 2021

CONFIDENTIALITY

Our clients' industries are extremely competitive, and the maintenance of confidentiality with respect to our clients' plans and data is critical. NERA Economic Consulting rigorously applies internal confidentiality practices to protect the confidentiality of all client information.

Similarly, our industry is very competitive. We view our approaches and insights as proprietary and therefore look to our clients to protect our interests in our proposals, presentations, methodologies and analytical techniques. Under no circumstances should this material be shared with any third party without the prior written consent of NERA Economic Consulting.

Project Team

George Anstey
Zoë Fannon

NERA Economic Consulting
Marble Arch House
66 Seymour Street
London, UK W1H 5BT
+44 207 659 8500
www.nera.com

Contents

List of Tables

List of Figures

Executive Summary

National Air Traffic Services (NATS) commissioned NERA Economic Consulting (NERA) to perform a benchmarking analysis of the wages paid by NATS En Route plc (NERL), in the context of preparing its business plan submission for the next price control period, NR23.

This report contains that analysis. We use wage equations estimated from a large, publicly-available dataset to estimate the market compensation for NERL staff. We compare those estimates to actual current pay for NERL staff groups in negotiated grades subject to collective bargaining. This analysis shows that NERL pay is broadly in line with market compensation.

Our Wage Equation Analysis Shows that NERL Wages are in Line with Benchmark Wages

Wage equations estimate the relationship between hourly pay and factors that drive differences in pay between individuals. These factors, termed "explanatory variables", include qualifications, industry, geographic location, and others. The academic literature contains a long heritage of estimating wage equations to explain the variation in compensation observed within the economy.[1]

At RP3, NATS presented wage equations developed by NERA which benchmarked NERL staff with reference to similarly skilled occupations in the wider economy.[2] In that analysis we estimated wage equations using data from the Labour Force Survey (LFS), a quarterly survey of approximately 40,000 households conducted by the Office for National Statistics (ONS).[3] Our wage equations included indicators for comparator occupations for NERL staff groups. We identified comparator occupations through a job-matching procedure based on Standard Occupational Classification (SOC) codes and descriptions of NERL staff duties. We then combined the estimated wage equations with data provided by NERL on the values of the explanatory variables for NERL staff groups (e.g. average qualification level within a staff group) to estimate benchmark pay for each staff group.

The analysis in this report follows the same approach as that used for RP3, including relying on the same comparator occupations (except in one case where a single comparator code is no longer available). We estimate wage equations using data from the LFS and combine those wage equations with data on the characteristics of NERL staff groups to estimate benchmark pay for each staff group. The staff groups are the negotiated grades subject to collective bargaining, namely ATCO, ATSA, ATCE, MSG, and STAR.

Compared to the previous report, we use updated data from both the LFS and NERL. We use data from sixteen waves of the LFS, from 2017Q2 through 2021Q1. In our previous report we used LFS data from Q4 only, as variables recording union membership and whether pay is

---

1 Notable examples include Mincer (1974), *Schooling, experience, and earnings*, which examined variation in pay driven by education and experience; Oaxaca (1973) *Mage-female wage differentials in urban labour markets*, which examined variation in pay by gender; Blinder (1973) *Wage discrimination: reduced form and structural estimates*, which examined variation in pay by race; and Krueger and Summers (1988), *Efficiency wages and the inter-industry wage structure*, which examined variation in pay by industry.

2 NERA (21 March 2018), Staff Operating Expenditure for Air Traffic Control: Prepared for NERL

3 Office for National Statistics (13 January 2015), Information Paper: Labour Force Survey

influenced by union agreements were only recorded in Q4. In this report, we use imputation to fill in the missing data on union variables in Q1-Q3, which allows us to use data from all quarters. This allows us to include more data from the period affected by COVID-19 to ensure our analysis captures the impact of the pandemic on wages in the economy and comparator occupations.

Our analysis suggests that NERL staff wages are broadly in line with market benchmarks. Figure 1 presents the results of our analysis for each of the five staff groups:

- ATCO pay is **within** the range of benchmark estimates. The benchmark estimates are equal to between 75 and 106 per cent of ATCO actual pay.
- ATSA pay is **above** the range of benchmark estimates. The benchmark estimates are equal to between 69 and 84 per cent of ATSA actual pay.
- ATCE pay is slightly **above** the range of benchmark estimates. The benchmark estimates are equal to between 85 and 94 per cent of ATCE actual pay.
- MSG pay is **within** the range of benchmark estimates. The benchmark estimates are equal to between 88 and 106 per cent of MSG actual pay.
- STAR pay is **within** the range of benchmark estimates. The benchmark estimates are equal to between 100 and 107 per cent of STAR actual pay.

The only two groups for which actual pay is above the range of benchmark estimates are ATSAs and ATCEs.

For ATCEs, this appears to be an artefact of the construction of the LFS variable *hourpay*, used as the outcome variable in our wage equations for comparison with NERL total pay. *Hourpay* is the LFS's measure of total pay per hour. However, *hourpay* may systematically understate the market benchmark for NERL total hourly pay. This is because over 80 per cent of LFS participants report that *hourpay* does not include any additions to basic pay, e.g. annual bonuses and shift premia, which are included in NERL total hourly pay. The dataset does not distinguish between respondents who do not receive additional pay and those who do not report it.[4] However, insofar as more than 20 per cent of LFS participants receive additions to basic pay, *hourpay* may be systematically missing variable components of pay and understate total compensation in the general economy. Moreover, to the extent that NERL staff receive more additional pay for performance and antisocial working hours, *hourpay* may understate the average hourly wages that NERL staff could earn for similar performance and conditions from the market at large. When we repeat the analysis using only those participants who report that *hourpay* does include additions to basic pay, ATCE pay is **within** the range of benchmark estimates.

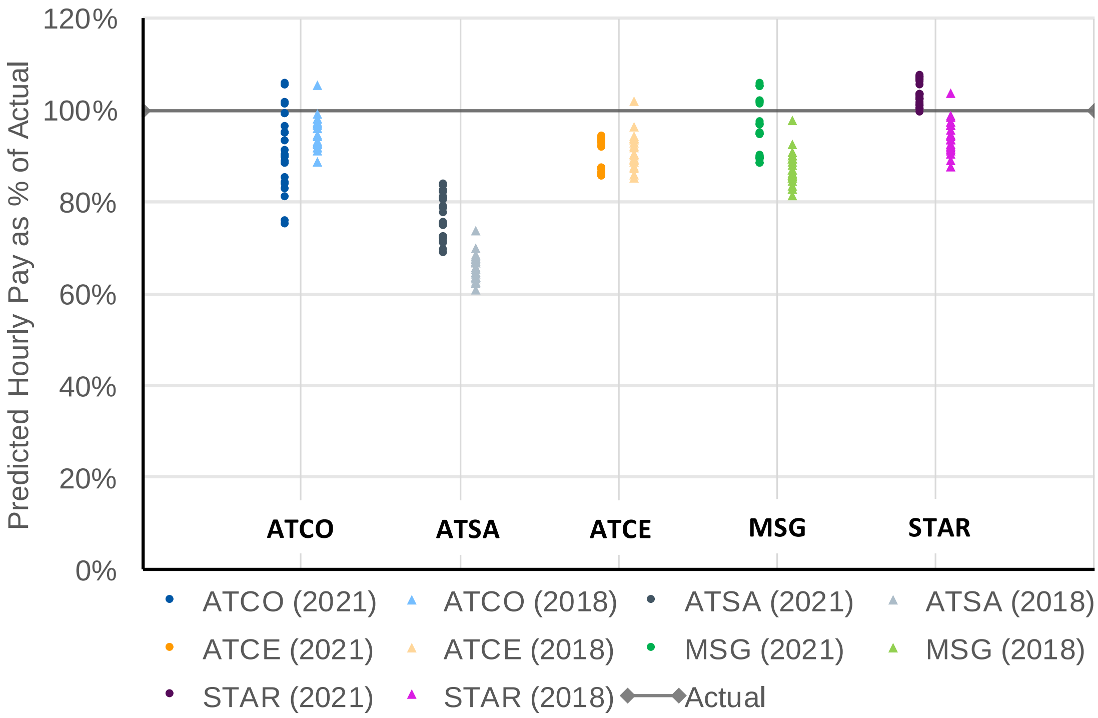
For ATSAs, the discrepancy between actual pay and benchmark pay cannot be fully explained by the construction of *hourpay*. The residual disparity between our estimates of market compensation for ATSAs and ATSA total compensation may reflect the difficulty of finding close comparators for ATSAs. ATSAs perform a wide array of duties and in

[4] There is reason to believe at least some respondents are not *reporting* additions to basic pay rather than not *receiving* additions to basic pay. For approximately 1/3 of the LFS sample, *hourpay* is calculated from gross pay in the preceding month rather than annual gross pay. Gross pay in the preceding month would typically not include annual bonuses.

particular have a responsibility for the safety of others that is not well reflected in our selected comparator occupations.

Overall, the results of this benchmarking exercise are similar to the results of our previous benchmarking exercise for RP3. Both Figure 1 and Table 1 compare the range of benchmark estimates, as a share of NERL actual wages, between the current exercise and the previous exercise. As can be seen from the Figure and Table, the maximum modelled benchmark pay for ATSAs has increased as a share of total pay since our last report in 2018. In other words, the pay of ATSAs and comparator occupations have converged over the period.

Figure 1: Model Predicted Wages as Share of NERL Actual Wages (Circles Show Results of Current Exercise, Triangles Show Results of Previous Exercise)



Source: NERA analysis of LFS and NERL data

Table 1: Model Predicted Wages as Share of NERL Actual Wages - Current and Previous Benchmarking Exercises

	ATCO	ATSA	ATCE	MSG	STAR
Predicted wages as share of NERL actual wages – current project					
Minimum predicted wage	75%	69%	85%	88%	100%
Maximum predicted wage	106%	84%	94%	106%	107%
Predicted wages as share of NERL actual wages – previous project					
Minimum predicted wage	89%	61%	85%	81%	88%
Maximum predicted wage	105%	74%	102%	98%	104%

Source: NERA analysis of LFS and NERL data; NERA 2018 report

1. Introduction

NERA Economic Consulting was commissioned by National Air Traffic Services (NATS) to provide economic advice and analysis to support NATS in preparing its business plan submission for the next price control period, NR23.

The Civil Aviation Authority (CAA) intends that the next price control will cover the period 2023-2027. NATS' business plan submission is due in February 2022.

NATS has asked NERA to prepare a benchmarking analysis of wages paid by NATS En Route plc (NERL) to "the aviation sector, other relevant professional groups and sectors, and the broader economy".[5]

This report is set out as follows:

- Chapter 2 provides an overview of the method we use to perform a benchmarking analysis of NERL staff wages, which is based on wage equations;
- Chapter 3 provides details on our estimation of wage equations. It sets out the dataset and variables we use, our selection of SOC comparators, and the choice of model specifications. Our selection of data, variables, SOC comparators, and model specifications is informed by our previous wage benchmarking study for NERL;[6] and
- Chapter 4 reports the results of our benchmarking exercise.

[5] NATS (13 April 2021), Request for Proposal – Economic advice and financial analysis support for price control reset, p. 4

[6] NERA (21 March 2018), Staff Operating Expenditure for Air Traffic Control: Prepared for NERL

2. We Use Wage Equations to Benchmark Pay

Our benchmarking analysis is based on wage equations. We estimate wage equations using a publicly-available UK dataset, and we use these wage equations to generate predicted market rates of total hourly pay for NERL staff in negotiated grades subject to collective bargaining.

This section provides an overview of our approach:

- Section 2.1 outlines our reasons for using wage equations to benchmark NERL staff wages; and
- Section 2.2 is a high-level overview of the three-step procedure we follow in conducting this benchmarking analysis.

2.1. Introduction to Wage Equations

In addition to any bottom-up scrutiny of expenditure, most cost-assessment in regulated industries relies on comparative benchmarking of cost lines or total costs across similar companies. For instance, Ofgem and Ofwat rely on statistical benchmarking models which aim to draw conclusions on the efficient level of costs, given outputs, for regulated electricity, gas, and water networks. In assessing the costs of local energy networks and water companies, Ofgem and Ofwat benefit from a sample of multiple companies all operating within their jurisdiction (8 Gas Distribution Networks, 14 Distribution Network Operators for electricity, and even more water companies).

NATS is the sole Air Navigation Service Provider (ANSP) operating within UK airspace. Accordingly, CAA would have to rely on international comparators to construct any meaningful statistical benchmarking analysis using similar businesses. However, at Reference Period 3 (RP3) CAA rejected international comparative benchmarking, describing it as "inconclusive at best" due to uncertainty around the selection of comparator ANSPs.[7] The Competition and Markets Authority (CMA) accepted CAA's position, noting that NERLs' planned program of airspace modernisation and technology transformation created additional challenges for this form of international benchmarking.[8]

At RP3, NATS presented wage equations developed by NERA that benchmarked NERL staff with reference to similarly skilled jobs in the wider UK economy. The academic literature contains a long heritage of estimating wage equations to explain the variation in compensation observed within the economy.[9] Wage equations contain factors such as qualifications, experience, industry, and geographic location – all of which contribute to differences in compensation between individuals. As well as a long academic heritage, wage equations have a history of application in regulatory and policy contexts. For example, the Department for Communities and Local Government (DCLG) estimated a "labour cost

7 CAA (2019), Reference to the CMA of NERL RP3 price controls: CAA response to NERL's Statement of Case, para. 5.59

8 CMA (23 July 2020), NATS (En Route) Plc/CAA Regulatory Appeal – Final Report, para. 8.63

9 Notable examples include Mincer (1974), Schooling, experience, and earnings, which examined variation in pay driven by education and experience; Oaxaca (1973) Mage-female wage differentials in urban labour markets, which examined variation in pay by gender; Blinder (1973) Wage discrimination: reduced form and structural estimates, which examined variation in pay by race; and Krueger and Summers (1988), Efficiency wages and the inter-industry wage structure, which examined variation in pay by industry.

adjustment" to take account of differences in wage costs between areas in order to determine local government funding.[10]

Estimating wage equations from an economy-wide dataset provides us with an estimate of how wages vary given the characteristics of employees, employers, and the job in question. Using that model for wages in the economy as a whole, we can then calculate the wages that NERL staff could expect to receive in the broader economy for performing a similar job for a similar employer based on the factors measured in the dataset. These factors include some, but not all, of the characteristics of NERL staff (e.g. education, experience), NERL as an employer (e.g. sector, employer size), and the job performed by each staff member (e.g. hours of work, occupation).

The wages that NERL staff could expect to receive in the broader economy (for a similar job and employer) represent the "outside option" of NERL staff – i.e. what they could reasonably expect to receive elsewhere. This is an appropriate benchmark for NERL wages, as NERL must pay its staff enough that they choose not to exercise their outside option and instead remain (or become, for new hires) employed by NERL. The difference between the benchmark wage and NERL's actual wage would then include any inefficient staff costs as well as the effect of any residual factors that the model does not explain (e.g. characteristics of the individual, employer, or job not measured in the dataset).

This approach based on wage equations establishes an objective benchmark for the wages that NERL should pay by:

- ***Controlling for a wide range of factors that affect compensation.*** Wage equations allow us to control for a large number of factors that may explain the pay of NERL staff. The inclusion of these explanatory variables means that we remove more of the wage variation caused by factors other than inefficiency from the comparison. These additional explanatory variables may include individual human capital measures (such as education, experience) employer characteristics (such as industry, size) and job characteristics (such as location, occupation); and

- ***Relying on the Labour Force Survey (LFS), a large publicly-available dataset.*** We rely on a large publicly-available dataset from the Office for National Statistics (ONS), the LFS, for our wage regressions. As we discuss in detail in Section 3.1, the LFS is a quarterly survey of over 40,000 households, capturing around 800 descriptive characteristics for each respondent, including pay, age, educational qualification and occupation. The LFS is widely used in applied economic analysis.[11] By using the LFS

[10] DCLG, "Methodology Guide for the Area Cost Adjustment 2013/14". http://webarchive.nationalarchives.gov.uk/20140505105916/http://www.local.communities.gov.uk/finance/1314/methacas.pdf In particular, the DCLG ran a regression on hourly earnings excluding overtime payments against a set of variables, including the area where each individual worked and factors it controlled for. The control variables included age, gender, occupation and industry, which were derived from the ASHE dataset. The coefficients on the area variables represent the relative wage in each area, after allowing for differences that are due to the control variables.

[11] See for example Equality and Human Rights Commission (2017), Research report 109: The gender pay gap (link: https://www.equalityhumanrights.com/sites/default/files/research-report-109-the-gender-pay-gap.pdf) and Institute for Fiscal Studies, Working Paper W13/11: What can wages and employment tell us about the UK's productivity puzzle (link: https://ifs.org.uk/wps/wp201311.pdf)

we ensure that our analysis is based on a representative sample of the UK economy, replicable, and easily audited.

2.2. We Follow a Three-step Approach to Benchmark Pay Using Wage Equations

Our approach to benchmarking the wages that NERL pays follows the procedure set out below, which is broadly the same as that used in our 2018 analysis for RP3:

1. First, we develop and estimate a series of models for wages across the economy as a whole. In our previous report on NERL staff pay, we estimated 160 models with different explanatory variables, based on a review of the literature on what variables drive pay and the explanatory variables available in the LFS dataset. We then identified a smaller number of "preferred" specifications which explained more of the variation in wages in the data, i.e. those which included granular measures of occupation and educational qualification as explanatory variables.[12] In this report, we estimate only these "preferred" specifications of wage equations. A simple wage equation could take the following *linear* functional form:[13]

   $$wage = \alpha + (\beta_1 \times age) + (\beta_2 \times qualification) + (\beta_3 \times occupation) + \ldots$$

   Where:

   - $wage$ is a measure of pay (the "dependent variable"); and
   - age, $qualification$, and $occupation$ are measures of the factors that drive pay (the "explanatory variables").

   We then estimate the above equation that best fits the LFS dataset for the economy as a whole. In other words, we estimate coefficients α and β in the above equation, where:

   - α is the amount all workers get paid; and
   - β_1 , β_2 and β_3 are the additional amounts that a particular worker gets paid per additional year of age, if they hold a particular educational qualification, or if they are in a particular occupation, respectively.

2. Having estimated the above wage equations, next we input specific data on the characteristics of NERL staff into the equation (i.e. data on NERL staff's age, $qualification$ and $occupation$) to estimate the predicted market wage.

3. Finally, we compare the market wages of NERL staff – as predicted by the wage equations – with their actual wages. Any difference between our prediction and NERL's wages would reflect (a) any inefficiency as well as (b) factors that influence wages but are not measured in the LFS dataset and therefore do not appear in the model (e.g. specific skills that are required of NERL staff, such as the ability to work under pressure).

[12] NERA (21 March 2018), Staff Operating Expenditure for Air Traffic Control: Prepared for NERL, p. 29

[13] Although, in practice, we estimate wage equations of a *log-linear* functional form (see Section 3.3), which is common to the economic literature.

3. We Estimate Wage Equations Following the Approach Established in Our Previous Pay Benchmarking Exercise

This chapter expands on section 2.2, providing more detail on the approach we follow to benchmark NERL staff hourly pay. We initially developed the approach in our 2018 report on NERL staff pay, and further information on the development of the approach can be found in that report.[14] This chapter proceeds as follows:

- Section 3.1 describes the LFS data we use to estimate our wage equations.
- Section 3.2 explains our selection of the outcome and explanatory variables for the wage equations. In particular, we explain our selection of comparator occupations that we use to benchmark NERL staff groups.
- Section 3.3 explains how we combine the explanatory variables into different model specifications for the wage equations.

3.1. We Use a Well-Known, Publicly-Available Data Set

To estimate the wage equations, we use data from the LFS. The LFS is a publicly-available dataset prepared by the ONS consisting of data collected from a quarterly survey of approximately 40,000 households and 100,000 individuals per quarter.[15] It collects information on over 800 descriptive variables,[16] including key variables of interest, such as hourly pay, age, educational qualifications, and occupation.

We use data from sixteen waves of the LFS: 2017Q2 – 2021Q1.[17] We begin the sample from 2017 because in our previous report we considered data up to 2016Q4. By starting in 2017, we ensure that there is no overlap between the dataset used in this report and the dataset used in the previous report. The absence of overlap means that the analysis of this report can be treated as independent of the previous report. This independence is valuable because it means the two reports constitute separate pieces of evidence on whether NERL staff wages are in line with market benchmarks, so that if both reports give the same result the overall conclusion is more robust. We begin from Q2 of 2017, rather than Q1, to ensure that each quarter appears the same number of times in the final dataset. Each quarter appears four times: Q2-Q4 in 2017-2020, and Q1 in 2018-2021. This means that any seasonal effects in our data should be averaged out across the full sample.

In our previous report, we used data from Q4 alone because union-related variables were only collected in Q4. In this analysis, we use *imputation* to fill in the values of union-related variables for Q1-Q3, so we can use data from all quarters.[18] This mitigates the impact of

[14] NERA (21 March 2018), Staff Operating Expenditure for Air Traffic Control: Prepared for NERL, Section 3.

[15] Office for National Statistics (13 January 2015), Information Paper: Labour Force Survey

[16] The raw dataset for 2017Q2 contains 809 variables.

[17] The sixteenth wave includes a very small number of observations that strictly speaking fall within 2021Q2.

[18] Imputation is an econometric technique that can be used to fill in the values of missing variables under certain conditions. It uses the patterns of association between the observed values of the variable of interest (in this case, union-related variables) and the observed values of other variables to generate values for the missing values of the variable of interest based on the other variables. See Appendix B.1 for further details.

seasonal patterns and allows us to include more data reflecting COVID-19 conditions. We also impute data for a small number of missing values for other variables.[19]

Although each LFS collects responses from approximately 100,000 individuals, only around 10,000 participants provide pay data each quarter. We drop all observations for which we are missing pay data.

We could not impute the missing pay data, because the pay variable does not meet one of the conditions required for imputation to be unbiased. For imputation to be unbiased, it must be the case that the probability that the variable is missing for a given individual does not depend on the value of the variable for that individual. This condition is clearly met in the case of variables that are missing because they are not collected for some quarters of the year, such as the unionisation variables considered in our analysis. However, this condition of independence is not met in the case of pay data, because individuals with high or low pay may be more likely to refuse to respond to a survey question on pay (and therefore be missing pay data) than individuals with mid-range pay.[20]

We also drop a small number of observations for which we are missing data on other variables.[21] Imputation of missing data for these variables would have involved complex econometric techniques, because these variables are *categorical* – they record the category to which an individual belongs (e.g. region, industry) and the categories are not ordered.[22] For each variable the number of observations with missing information is small (less than 400 observations) and so dropping the observations is unlikely to cause bias.

This leaves us with a total final sample of 142,803 observations, which is a large dataset by the standards of applied economic analysis.[23]

3.2. We Identify Key Variables for Our Equations from the Economic Literature

In our previous report, we conducted a review of the academic literature on wage equations to select the variables to use in our wage equations. This review informed both our selection of the outcome variable, hourly pay, and the explanatory variables, capturing the factors that drive wages.

---

19 The other variables for which we imputed values were: an indicator for whether the employer was public sector or private sector; an indicator for whether the employer had over 500 employees; the number of years for which an individual had worked for their current employer; and the number of hours worked in excess of usual hours.

20 Lillard, L., Smith, J.P., Welch, F. (1986) *What do we really know about wages? The importance of nonreporting and census imputation, Journal of Political Economy, 94(3).*

21 The variables for which we dropped observations where there was missing data were: highest educational qualification; industry; region; ethnicity; and the indicator for being a full-time employee.

22 Imputation is more straightforward for binary (yes/no) variables, like union membership, and continuous variables, like tenure.

23 Miles, J. and Shevlin, M., *"Applying Regression and Correlation: A Guide for Students and Researchers",* page 119. The authors state that as a rule of thumb, there should be at least 20 observations per independent variable in the sample. Although the authors go on to state that a full power analysis is preferred rather than relying on rules of thumb, with a sample size of 142,803 observations we have over 200 observations per independent variable, even in the largest models we estimate. We are therefore confident in the stability of our model estimates.

The academic literature on wage equations has a long pedigree, and the review in our previous report focused more on the established literature than on recent developments aimed at investigating specific research questions. The established literature has not changed significantly since 2018 and so it was unnecessary to repeat the review exercise for this report. Instead, we rely on the results of our previous review, but make some adjustments to reflect changes to data availability, the COVID-19 pandemic, and specific features of the new LFS dataset.

In our previous report, we selected a longlist of variables that had been used by at least two academic authors in their wage equations. We then filtered that longlist down based on a further three criteria, which we re-apply in the context of the current report:

1. ***Omitting statistically similar variables.*** Many variables included in the academic literature on wage equations have slightly different specifications or descriptions but essentially measured the same underlying economic properties. For instance, "potential experience" appears in several papers – which in turn is constructed from variables "age" minus "age when completed full-time education". Including many variables which are closely related introduces a statistical phenomenon known as multicollinearity, resulting in unreliable inferences from model estimation. As a result, we prune the longlist to include only a set of variables which are potentially jointly meaningful;

2. ***Data availability in the LFS database.*** As we rely on the LFS database to estimate the economy-wide wage equation, we also need to ensure that the variables we include in our equations are also recorded in that database; and

3. ***Data availability from NERL.*** Having estimated the economy-wide wage equation, we then need to input specific data on the characteristics of NERL staff into the wage equation to estimate their predicted market wage. As such, we also need to ensure that any variables we include in our wage equation in the first place are also available for NERL staff.

3.2.1. Our outcome variable is gross total hourly pay

The outcome variable we use is gross total hourly pay. This is the same outcome variable that we used in our previous report.

In the LFS dataset, the variable *hourpay* records gross hourly pay. It is equal to gross pay (i.e. pay before deductions) received in the previous pay period, divided by the sum of basic usual hours worked and paid overtime hours for that pay period.[24]

In the NERL data, we have annual total pay, based on payroll data, and actual hours worked per week, as reported by NERL staff. We divide annual total pay by 52 (the number of weeks in a year) and again by actual hours worked per week to get gross total hourly pay.

[24] The variable *hourpay* does not include unpaid overtime hours, and so may underestimate total actual hours and thus overestimate total hourly pay (compared to NERL total hourly pay). However, we assume that the impact of unpaid overtime is likely to be small, as NERL employees in negotiated grades are paid on a shift basis and are therefore paid for overtime.

3.2.2. We include a range of explanatory variables

We make some adjustments to the set of explanatory variables as compared to the previous report. These changes are a result of the re-application of the filtering process described at the beginning of Section 3.2 to the new LFS and NERL datasets.

In this report we omit variables capturing union membership, recent on-the-job training, and the number of dependent children. We do not have data on these variables for NERL staff. For recent on-the-job training, we determined that this variable would be distorted by COVID-19 and chose not to collect it; for union membership, we instead used a similar variable recording whether wages are subject to union agreement.

We revised the format of variables capturing comparator occupation effects, firm size, hours worked, and region of work. In two cases this was due to constraints of the updated LFS dataset, specifically changes to the Standard Occupational Classification (SOC) and issues of multicollinearity with the variables capturing hours worked. For firm size, the adjustment was to facilitate imputation, and for region of work, the adjustment was to resolve incompatibility between the LFS and NERL datasets.

We explain the adjustments to comparator occupations and hours worked in in Sections 3.2.2.1 and 3.2.2.2. We explain the adjustments to other explanatory variables in Appendix C. The explanatory variables included in our wage equations are listed in Table 3.1.

3.2.2.1. SOC indicators

In the previous report, we included an indicator for each Standard Occupational Classification (SOC) code in the economy. We then used an unweighted average of the estimated coefficients on the SOCs for NERL staff group comparators to get the estimated "occupation effect" for each of the five NERL staff groups (ATCOs, ATSAs, ATCEs, STARs, and MSGs).

In this report, we construct an indicator for each staff group, which captures all SOCs that are comparators for that staff group and estimates a single "occupation effect" for them.

The new approach is in response to data constraints in the LFS. The SOC classification was updated in 2020: new codes were introduced, some codes were retired, and other codes were either split or combined. The 2020 SOC replaces the 2010 SOC in the LFS from 2021Q1 onwards. We could not find a mapping from the 2010 SOC to the 2020 SOC for the whole economy. In any case, relying on a single occupational effect is sufficient for our purposes because we are interested in estimating the combined effect of the benchmark occupations on wages for NERL staff, rather than the wage uplift for each benchmark occupation relative to the economy as a whole.

Table 3.1: Explanatory Variables Included in our Wage Equations

Variable
Indicators for whether the SOC is a comparator occupation for one of NERL's staff groups
Controls relating to employer characteristics
Indicator for large employer (over 500 employees)
Indicator for private sector employer
Indicators for 1-digit Standard Industrial Classification (SIC) code
Controls relating to employee characteristics
Indicators for highest qualification
Age (proxy for work experience)
Tenure at current employer
Indicator for whether wages are influenced by union agreements
Controls relating to job characteristics
Indicator for whether the position is full-time
Basic usual hours worked per week (proxy for contracted hours)
Excess usual hours worked per week (total usual hours less basic usual hours)
Indicator for region of work/residence
Functional form adjustments
Square of tenure
Square of age
Square of basic usual hours
Square of excess usual hours
Time Effects
Time dummies (indicator for quarter of observation)
Time trend
Interaction 1-digit SIC and time dummies
Interaction SOC indicators and time dummies
Interaction 1-digit SIC and time trend
Interaction SOC indicators and time trend
Interactions
Interaction of indicator for full-time position with basic usual hours
Interaction of indicator for full-time position with square of basic usual hours
Demographics
Indicator for sex
Indicators for marital status
Indicators for ethnicity

Source: NERA analysis

3.2.2.2. Hours worked

In the previous report, we included a variable for total usual hours worked (including overtime) as well as a variable for basic usual hours worked (excluding overtime). In this report, we replace the variable for total usual hours worked with a variable capturing excess hours worked (i.e. total hours minus basic hours). We do this to reduce multicollinearity between total usual hours worked and basic usual hours worked.

In the previous report, we did not include squares of hours worked. In this report, we do include squares of hours worked. This is in response to evidence from the academic literature, which suggests a non-linear relationship between hours worked and hourly pay.[25]

In this report, we also include *interactions* between usual hours worked (and its square) and whether the job is full-time. An interaction allows the estimated effect of one variable in the model to depend on the value of a second variable.[26] The interactions between usual hours worked and whether the job is full-time allow the shape of the relationship between hours worked and pay to differ between part-time and full-time employees.

In the absence of this interaction, the estimated shape of the relationship between hours worked and pay could be distorted by the presence of part-time high earners. Part-time high earners could cause wages to appear to "peak" at the level of part-time hours and then fall thereafter. Including these interaction effects eliminates the potential for this distortion.

3.2.3. We Selected Comparator SOCs for NERL Staff Groups Based on Analysis in Our Previous Pay Benchmarking Exercise

The SOC is a system for classifying occupations in the UK (managed by the ONS), which groups jobs in terms of their skill level and skill content with up to 4 levels of granularity. The ONS publishes an SOC description document, which provides job descriptions and summaries of key tasks associated with each SOC (most detailed at the 4-digit level).

In our previous report, we conducted a thorough review of the SOC and NERL staff job descriptions in order to identify candidate comparator SOCs for each NERL staff group. We then applied a set of funnel criteria to restrict the set of candidate comparator SOCs to a final selected set of comparator SOCs. Full details of the procedure can be found in that report.[27]

In this report, we adopt the same set of comparator SOCs for each staff group that were identified in the previous report. These are listed in Table 3.2. There is only one difference relative to the previous report. In the previous report, Design and Development Engineers (2126) were included as a comparator SOC for ATCEs. This code was eliminated in the 2020 update to the SOC and so we exclude it to ensure consistency across the full sample.

---

[25] See for example Goldin, C. (2014), A grand gender convergence: its last chapter.

[26] Consider a wage equation with just hours worked and full-time:

$$wage = \alpha + (\beta_1 \times hours) + (\beta_2 \times fulltime) + \cdots$$

We can add an interaction to make this

$$wage = \alpha + (\beta_1 \times hours) + (\beta_2 \times fulltime) + (\beta_3 \times hours \times fulltime) + \cdots$$

The coefficient β_1 captures the impact of an extra hour worked on wages across the sample; the *interaction* coefficient β_3 captures any change to the impact of an extra hour worked on wages if we consider only those that work full time.

[27] NERA (21 March 2018), Staff Operating Expenditure for Air Traffic Control: Prepared for NERL, Section 3.2

Table 3.2: Comparator SOCs for Each NERL PS Group

NERL Staff Category	Comparator SOCs
ATCO	Aircraft pilots and flight engineers (3512)
ATSA	Administrative occupations: Records (413)
	Administrative occupations: Office managers and supervisors (416)
	Secretarial and related occupations (421)
	Human resources and industrial relations officers (3562)
	Health and safety officers (3567)
ATCE	Mechanical engineers (2122)
	Electrical engineers (2123)
	Electronics engineers (2124)
	IT specialist managers (2133)
	IT project and programme managers (2134)
	IT business analysts, architects and system designers (2135)
	Programmers and software development professionals (2136)
	Quality control and planning engineers (2461)
	Electrical and electronics technicians (3112)
	Engineering technicians (3113)
MSG	Administrative occupations: Finance (412)
	Office managers and supervisors (416)
	Financial and accounting technicians (3537)
	Financial accounts managers (3538)
	Human resources and industrial relations officers (3562)
STAR	Research and development managers (215)
	IT business analysts, architects and system designers (2135)
	Programmers and software development professionals (2136)
	Management consultants and business analysts (2423)
	Business and financial project management professionals (2424)
	Health and safety officers (3567)

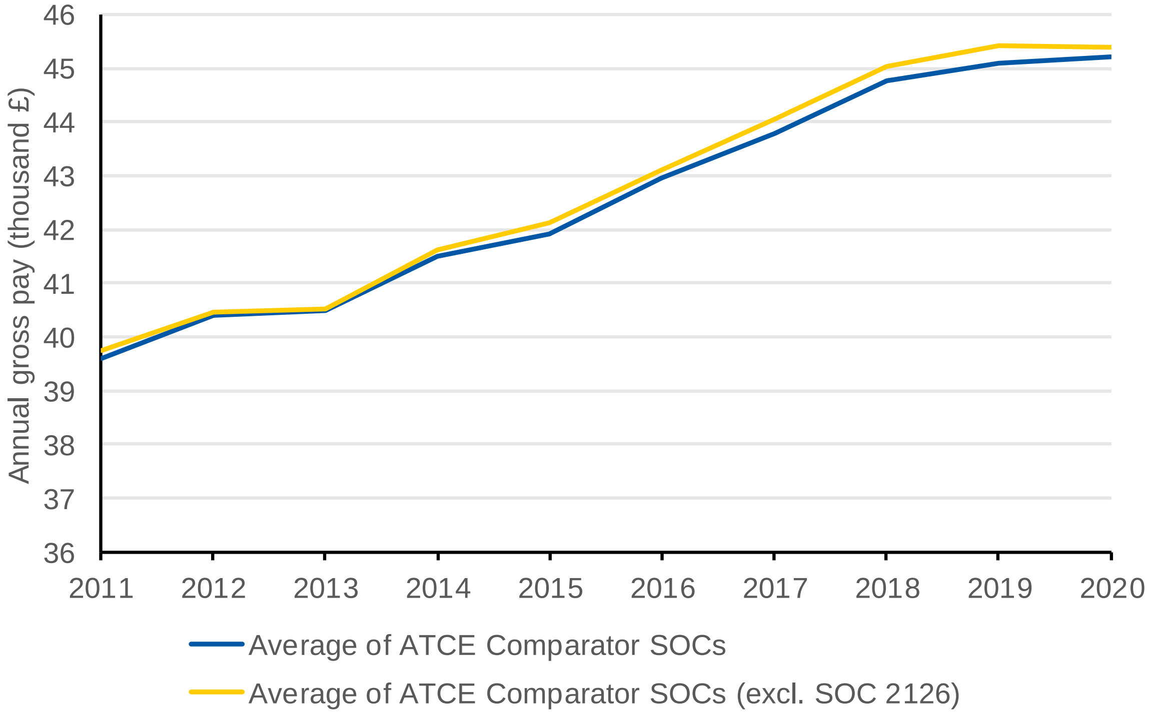
Source: NERA analysis of SOC and NERL job descriptions

The impact of excluding Design and Development Engineers (2126) on estimated market benchmark pay is small, because this SOC code is towards the middle of the ATCE comparator SOCs in terms of pay.[28] Figure 3.1 shows the average of ATCE SOC Comparator annual pay over 2011-2020, both including and excluding Design and Development Engineers (2126).[29] The figure shows that excluding Design and Development Engineers (2126) has a relatively small impact on the average.

[28] Looking at data from the Annual Survey of Hours and Earnings (ASHE) over the past 10 years, we see that Design and Development Engineers (2126) are consistently paid more than 4 of the other 10 ATCE comparator SOCs and less than 6 of the 10 ATCE comparator SOCs.

[29] These are calculated using data on average pay within each SOC from the ASHE. The ASHE is an ONS dataset based on a 1% sample of the HMRC PAYE register. Compared to the LFS, it provides more accurate and reliable data on earnings, as the sample is larger and pay is not self-reported. It is therefore preferable for aggregate analysis such as that in Figure 3.1, but it cannot be used for wage equation analysis in its aggregated, publicly-available format.

Figure 3.1: Removing Design and Development Engineers (2126) from ATCE Comparator SOCs Marginally Increases the Average ATCE Comparator Pay



Source: NERA analysis of ASHE data (in financial years, e.g. 2020 is year ending March 2020)

3.3. We Developed and Estimated a Range of Wage Equations

We estimate wage equations in a *log-linear* functional form, which is common in the economic literature on wage equations.[30] A log-linear wage equation takes on the following general structure:

$$\ln(wage) = \alpha + (\beta_1 \times age) + (\beta_2 \times qualification) + (\beta_3 \times occupation) + \ldots^{31}$$

In this equation, coefficients β_1, β_2, and β_3 have the interpretation of the additional *percentage* that a particular worker gets paid per additional year of age, if they have a particular qualification, or if they are in a particular occupation (respectively).

Specifically, we estimate the wage equations of $\ln(gross\ hourly\ pay)$ on all the explanatory variables we identified in Section 3.2, including our selection of comparator SOCs for each NERL staff category.

In order to show the impact of different approaches on our estimated wages, we run a range of models – 37 in total – with different model specifications. Intuitively, we are not trying to find the artificial "best model" of wages in the general economy. In econometric analysis, relying on just one specification is generally unwise because estimates can be sensitive to

30 See Appendix B.2 for further explanation of the use of the log-linear functional form.

31 ln(wage) is the natural log of the variable wage.

minor changes in model specification. Rather, our approach is to obtain estimates from a range of equally plausible model specifications, which is more reliable.

All specifications include the Employer Controls, Employee Controls and Job Controls listed in Table 3.1, with the exception of the variable reflecting whether wages are set by union agreement, *tucov*, which is sometimes omitted. They also include the 1-digit SIC indicators, SOC comparator indicators, and time (either as a dummy or as a trend). The differences therefore lie in the functional form of time (dummy vs. trend), inclusion/exclusion of functional form adjustments (squared terms), interaction effects, and demographic controls listed in Table 3.2. The specific details of the estimated models are listed in Table F.1.

We estimate substantially fewer models in this report than we did in our previous report. In our previous report, we estimated a total of 160 models. However, in the previous report, we ultimately narrowed this set down to a "preferred" group of 24 models.

The 24 "preferred" models in the previous report included 1-digit SICs, 4-digit SOCs, and a variable recording the highest educational qualifications. The models differed in their treatment of other variables, including interaction terms.

To reduce the complexity of this report, we do not estimate the categories of model that were not selected as "preferred" in our previous report. We do not consider models with levels of SIC and SOC other than the preferred 1-digit and 4-digit, nor do we estimate models that exclude educational qualifications. The narrower range of models presented in this report reflects the relative reliability of the models we had estimated. In our previous report, we determined that the preferred models, which used more granular SOCs and which included educational qualification, were more plausible for at least two reasons:[32]

- ***Education is a key explanatory variable.*** The majority of the papers assessed in the literature review of the previous report included education as an explanatory variable. In the previous report, the education variable was statistically significant across all wage models.
- ***Granular SOCs are more relevant.*** In the previous report, models with 4-digit SOCs explained more of the variation in pay in the economy as a whole than did models with SOCs at lower granularity (1-digit, 2-digit, and 3-digit SOCs). That is, they had a higher adjusted-R2. This indicates that the granularity of the 4-digit SOCs adds value in predicting differences in pay between specific occupations.

3.4. Our Models Allow for the Time-Varying Impact of Explanatory Factors on Wages

We estimate our wage equations using data from 16 quarters, i.e. 2017Q2 through 2021Q1. Wages can reasonably be expected to change over the course of this three-year period, both economy-wide and for specific industries and occupations. To account for this, we include three different types of explanatory variable in our wage equations, as shown below:

$$\ln(wage) = \alpha + (\beta_1 \times factor) + (\beta_2 \times time) + (\beta_3 \times interaction\ between\ time\ and\ factor) + \ldots$$

32 NERA (21 March 2018), Staff Operating Expenditure for Air Traffic Control: Prepared for NERL, Section 4.2

The three types of explanatory variable are: factors of interest (e.g. industry, occupation), time, and interactions between time and the factor of interest. We explain the impact of each of these on estimated wages below.

- ***Factor variable.*** The coefficient β_1 captures the average wage premium associated with a particular value of the factor variable *on average across all 16 quarters* in the dataset. For example, if the factor were industry, then for NATS the coefficient β_1 captures the wage premium associated with the transport industry on average over the 16 quarters.
- ***Time variable.*** The coefficient β_2 captures the average impact on wages associated with a particular time period *for the economy as a whole*. These average impacts would include, for example, the effect of inflation on wages or the economy-wide average impact of COVID-19. When calculating NATS benchmark wages, we set the time variable equal to 2021Q1, so it reflects the average economy-wide impact of being in 2021Q1 rather than in any of the previous quarters (for example, it reflects overall inflation since 2017Q2).
- ***Interaction between time and factor.*** An interaction allows the estimated effect of one variable in the model to depend on the value of a second variable.[33] The coefficient β_3, which is an interaction between time and a factor variable in a wage equation, captures how the relationship between the factor and wage changes over time. For example, an industry-time interaction captures deviation in industry average pay, for a given quarter, from industry-average pay over the full 16 quarters in the sample. This deviation is industry-specific, i.e. not explained by economy-wide factors such as inflation that are already accounted for in β_2. These industry-specific deviations could be driven by industry-specific periods of contraction or growth. In this report, we consider both industry-time (SIC-time) and occupation-time (SOC-time) interactions. These are useful to understand the industry- and occupation-specific impacts of COVID-19.

As a result:

- all of our estimates take account of average wage movements in each quarter. In other words, we benchmark all ATCO wages against average wages prevailing in Q1 2021;
- all of our estimates reflect the average impact of *some* explanatory factors (e.g. education) over the sample period over and above average hourly pay for the economy as a whole. In other words, we benchmark NERL's wages by assuming the wage premium for higher education (and other factors) over average wages is fixed for the entire sample period; and
- some of our estimates allow the impact of *some* explanatory factors to vary over the sample period, depending on the model run. In our modelling, we include time-interactions for SICs and SOCs in only some of the sample of model runs. In other words, we benchmark NERL's wages against:
  - average wages in Q1 2021; plus
  - a higher-education premium (and other premia) that is fixed over the whole period; plus

[33] See the footnote in Section 3.2.2.2 for further details.

- a premium for operating in the air transport sectors or reflecting specific comparator occupations which is estimated based on either (a) Q1 2021 data or (b) the sample period as a whole, depending on the specific model run. The estimates from models that include time-SIC or time-SOC interaction terms will more closely reflect recent fluctuations in the aviation industry and comparator occupations.

4. Our Wage Equation Analysis Shows That NERL Wages are in Line with Benchmark Wages

This chapter reports the results of our benchmarking exercise, which combines the wage equations estimated from the LFS (as described in chapter 3) with data from NERL staff groups on the explanatory variables in those wage equations. The section proceeds as follows:

- Section 4.1 describes the available data on NERL staff groups and explains how we use that data in conjunction with the estimated wage equations to generate benchmark wages;
- Section 4.2 compares the benchmark wages with NERL staff group actual wages;
- Section 4.3 compares the findings from the current benchmarking exercise with the findings of our previous benchmarking exercise, conducted for NERL in 2018;
- Section 4.4 and Section 4.5 outline some limitations of our benchmarking exercise and discuss how interpretation of the results should account for those limitations; and
- Section 4.6 concludes.

Overall, the results of our current benchmarking exercise show that NERL wages are broadly in line with benchmark wages. This is consistent with the results of our previous analysis in 2018.

4.1. We Calculate NERL's Predicted Benchmark Wages by Inputting Data on NERL Staff Groups into Estimated Wage equations

The wage equations described in Section 3.3 allow us to calculate the predicted wage for any given individual with known values of the explanatory variables. We do this by entering the individual values of the explanatory variables into the wage equation. The result is the wage that such an individual could expect to receive in the economy as a whole and therefore a measure of their outside option.

To benchmark wages for NERL staff groups, we calculate the average predicted wage for that staff group. We do this by calculating the average values of each explanatory variable within each NERL staff group and entering these values into the wage equation. The data for each explanatory variable comes from one of three sources: NERL's internal database, a NERL staff survey, and inferred NERL-specific information. The result of the calculation is the average predicted wage for that staff group. This is, on average, the benchmark wage for that staff group, i.e. the wage that NERL needs to pay to recruit and retain employees given labour market conditions.

We would not expect this benchmark wage to hold instantaneously and over all points in time given the rigidities and transactions costs associated with moving jobs in the labour market. However, over the long term, market wages for comparable staff provide the best possible indication of the wages that NERL would need to pay to recruit and retain staff.

Table 4.1 lists the source of NERL-specific data for each explanatory variable included in our wage equations. There are three distinct sources of data, as follows:

- ***Information from NERL's internal databases.*** For the majority of variables, we rely on detailed information obtained from NERL's payroll database. The data collected from payroll includes age, annual total and basic pay, and region of work. We use the average value of each variable within each staff group in our wage equations. The payroll database contains 3,133 observations;
- ***Information from NERL staff survey.*** For variables where we could not obtain relevant information from the NERL database – namely "highest educational qualification" and "actual hours worked" – we rely on a survey of NERL staff. The survey was distributed to all NERL employees and received 812 valid responses.[34] We use the average value of each variable within each staff group in our wage equations;
- ***Inferring NERL-specific information.*** There are also some variables – namely "1-digit SICs", "size of employer workforce" and "public/private sector" – which take the same value for all NERL staff groups, as they derive from employment by NERL. NERL operates in the private sector, employs over 500 people, and, sits within the "Transportation and storage (H)" 1-digit SIC. We also set the value of time variables equal to the most recent quarter in the dataset, i.e. 2021 Q1, since the data provided by NERL reflects conditions in 2021.

We provide descriptive statistics for the data on NERL staff in Appendix A.2.

[34] We excluded 62 responses with missing, incomplete, or clearly inaccurate information. For example , we excluded one observation which stated the number of hours worked in the week as "too many" and another which stated 115 hours.

Table 4.1: Variables from NERL data

Variable	NERL data source
Indicators for whether the SOC is a comparator occupation for one of NERL' PS groups	Inferred NERL-specific value
Controls relating to employer characteristics	
Indicator for large employer (over 500 employees)	Inferred NERL-specific value
Indicator for private sector employer	Inferred NERL-specific value
Indicators for 1-digit SIC	Inferred NERL-specific value
Controls relating to employee characteristics	
Indicators for highest qualification	Staff survey
Age (proxy for work experience)	Internal database
Tenure at current employer	Internal database
Indicator for whether wages are influenced by union agreements	Internal database
Controls relating to job characteristics	
Indicator for whether the position is full-time	Internal database
Basic usual hours worked per week (proxy for contracted hours)	Internal database (contracted hours)
Excess usual hours worked per week (total usual hours less basic usual hours)	Actual hours from staff survey less contracted hours from internal database
Indicator for region of work/residence	Internal database
Functional form adjustments	
Square of tenure	Internal database
Square of age	Internal database
Square of basic usual hours	Internal database (contracted hours)
Square of excess usual hours	Actual hours from staff survey less contracted hours from internal database
Time Effects	
Time dummies (indicator for quarter of observation)	Inferred NERL-specific value
Time trend	
Interaction 1-digit SIC and time dummies	Inferred NERL-specific value
Interaction SOC indicators and time dummies	Inferred NERL-specific value
Interaction 1-digit SIC and time trend	Inferred NERL-specific value
Interaction SOC indicators and time trend	Inferred NERL-specific value
Interactions	
Interaction of indicator for full-time position with basic usual hours	Internal database (contracted hours)
Interaction of indicator for full-time position with square of basic usual hours	Internal database (contracted hours)
Demographics	
Indicator for sex	Internal database
Indicators for marital status	Internal database
Indicators for ethnicity	Internal database (average by PS group)

Source: NERA analysis of NERL data

In the previous report, we used data on actual hours from NERL's internal database. For the current report, it was not possible to retrieve data on actual hours from NERL's internal database and so we relied on data collected from the staff survey.

To check the impact of this change, we compared actual hours per staff group from the 2018 report with actual hours per staff group from the recent staff survey. Table 4.2 shows that the hours per staff group are similar. In particular, the numbers for the recent survey are not systematically higher, indicating that employees did not systematically overestimate their hours when responding to the survey.

Table 4.2: Actual Hours Worked in 2021 are in Line with Actual Hours Worked in 2018

PS Group	2018 Core Hours (Average from Database)	2021 Contract Hours (Average from Database)	2018 Actual Hours (Average from Database)	2021 Actual Hours (Average from Survey)
ATCO	32.86	34.19	35.46	36.61
ATSA	32.86	34.20	36.99	37.55
ATCE	35	34.66	38.98	38.50
MSG	35	33.76	38.16	37.81
STAR	35	33.95	37.36	36.62

Source: NERA analysis of NERL data

Note that we use contracted (core) hours from NERL staff data to replace basic usual hours in the LFS wage equations, while we use the difference between actual and contracted (core) hours to replace excess hours in the LFS wage equations. This is in line with the approach in our previous report.

4.2. We Compare NERL's Predicted Benchmark Wages to NERL's Actual Wages

We calculate NERL's actual hourly pay as described in Section 3.2.1; that is, total annual pay (including supplements such as bonuses), divided by 52 weeks in the year and actual hours worked per week.[35] This is in line with the approach in our previous report.

Table 4.3 reports the minimum and maximum predicted wage for each NERL staff group across these 37 wage models, as well as the actual hourly pay for each staff group. Actual pay is within the predicted range for ATCOs, MSGs, and STARs. For ATSAs actual pay is above the predicted range and for ATCEs it is just above the predicted range.

[35] Note that we use actual hours as the denominator to calculate the NERL equivalent of *hourpay* and in the calculation of the excess hours explanatory variable, while using contracted hours to set reflect the basic hours explanatory variable. This implies an assumption that the difference between actual and contract hours for NERL staff is entirely due to paid, rather than unpaid, overtime. In the LFS dataset, the denominator of *hourpay* is (basic hours + paid overtime hours), while excess hours is (paid overtime hours + unpaid overtime hours). Since the LFS data includes three categories of hour and the NERL data includes only two categories of hour, some assumption is required to align the datasets. We could either assume all overtime hours are paid, or all overtime hours are unpaid. NERL negotiated grades operate shifts and therefore are typically paid for overtime, making the assumption that all overtime hours are paid more reasonable than the assumption that all overtime hours are unpaid.

Table 4.3: Model Predicted and Actual Hourly Pay for NERL Staff Groups

	ATCO	ATSA	ATCE	MSG	STAR
Minimum predicted wage	36.86	20.03	29.03	22.02	27.02
Maximum predicted wage	51.81	24.36	31.99	26.33	29.16
Actual wage	49.04	29.03	33.97	24.91	27.14

Source: NERA analysis of LFS and NERL data

Sections 4.2.1 through 4.2.4 compare the actual wages for each NERL staff category to the predicted benchmark wages from our wage equations. Each section contains a Figure in which actual wages are represented by the orange line. The predicted wages from our wage equations are represented by the blue dots, with one dot for each of our 37 model specifications.

These predicted wages are plotted against the adjusted-R2 ("R-squared") statistic for each estimated wage equation. The adjusted-R2 measures the goodness-of-fit[36] of regression models, corrected for the number of explanatory variables in the model. In general, the higher the adjusted-R2, the more the explanatory variables in the model explain variation in the outcome variable, and so the better the model fits.[37]

Our models typically have an adjusted-R2 around 0.4, meaning that they explain approximately 40 per cent of the variation in wages. This figure compares well with wage equations in the academic literature.[38] It is not surprising that a significant share of variation in wages in the economy as a whole remains unexplained; much wage variation is due to individual-specific characteristics which are either not objectively measurable (e.g. ability to work under pressure) or too specific to be recorded in a large-scale survey like the LFS (e.g. familiarity with domain-specific technology).

Our analysis shows that models with a higher adjusted-R2 (i.e. better fit) predict higher wages, on average, for all NERL staff groups. However, all of the 37 specifications that we consider fit the data similarly well and the differences in adjusted-R2 across models are small compared to the average total adjusted-R2 of the models.

Our analysis also shows that models which include squared terms typically predict higher pay than those which do not. The squares capture non-linear returns to experience (as proxied by age), hours, and tenure. Consider, for example, tenure. Typically pay increases by more, per additional year of tenure, at the beginning of an employee's tenure than at the end of the employee's tenure. This is the dynamic shown by the blue line in Figure 4.1. Including squared terms in the model allows us to capture this dynamic. Without squared terms, the model will assume constant (linear) returns to tenure, as shown by the yellow line. The yellow line (model without squared terms) underestimates the wages of employees with

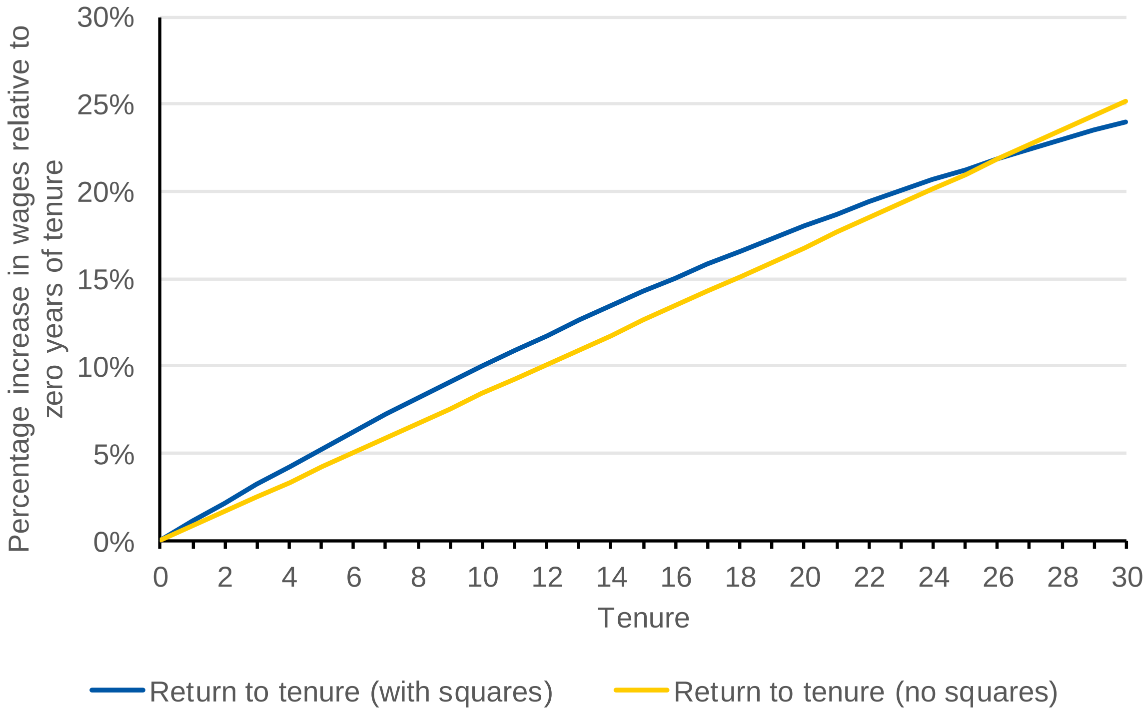
36 The proportion of the variance in the dependent variable that is explained by the explanatory variables, between 0 and 1.

37 Note that the adjusted-R2 should only be compared for models estimated from the same dataset, with the same outcome variable. The adjusted-R2 cannot be used to compare models using different outcome variables or datasets, as the total variation in the denominator will be different.

38 For example, the classic wage decompositions from Oaxaca (1973) report R2 between 22% and 56%, depending on the variables included in the regression.

tenure below 25 years, which is the point at which the blue and yellow lines converge. Similar dynamics exist for total experience (proxied by age) and hours worked.

Figure 4.1: Illustration of Non-Linear Returns to Tenure (Based on Estimated Coefficients in Wage Equations)



Source: NERA analysis of LFS data

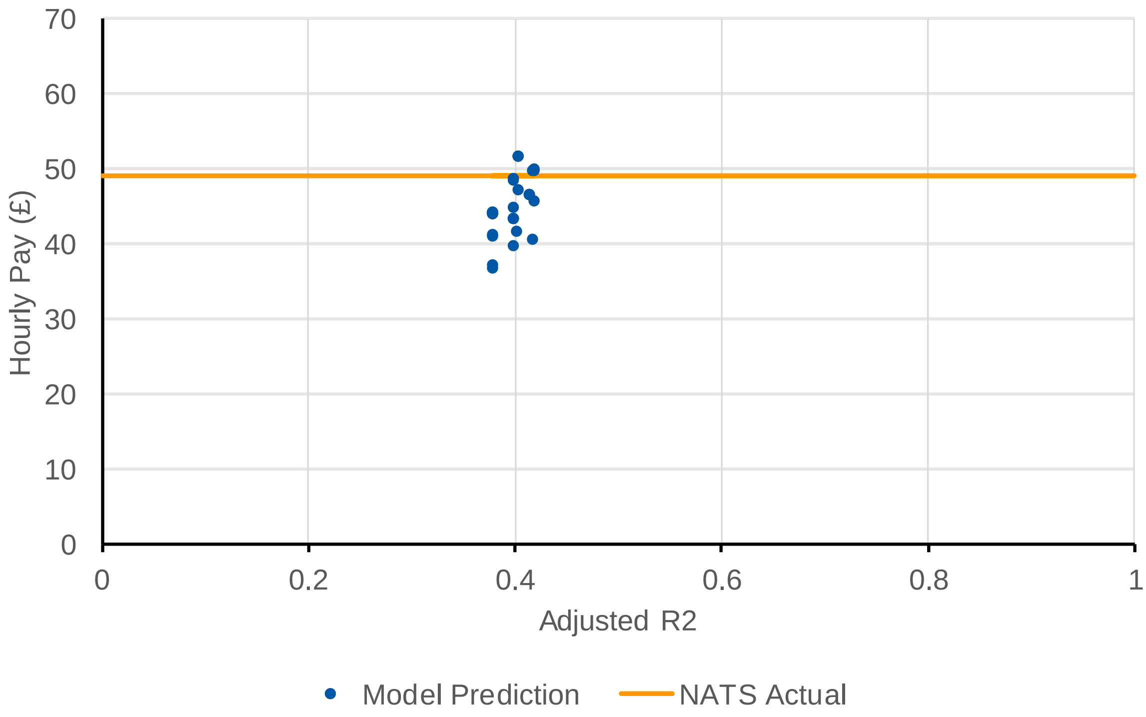
4.2.1. ATCO actual wages are within the range of model predictions

Our models predict benchmark pay equal to between 75-106 per cent of ATCO actual wages. Figure 4.2 shows the range of model predictions.

The models predicting particularly low pay for ATCOs (£40/hour and below) are those with time-SOC interaction terms. As explained in Section 3.4, time-SOC interaction terms capture how the relationship between an SOC and wages change from quarter to quarter.

Models *without* time-SOC interaction terms estimate a single coefficient for the SOC comparator, which captures the average additional impact on wages across the LFS dataset of belonging to a comparator SOC for the NERL staff group in question. In the case of ATCOs, the only SOC comparator is Airline Pilots, and so this single coefficient estimates the average wage premium afforded to pilots (relative to the economy as a whole) across 2017Q2-2021Q1.

Figure 4.2: Predicted ATCO Total Hourly Pay



Source: NERA analysis of LFS and NERL data

Models *with* time-SOC interaction terms estimate a different coefficient for the SOC comparator in each time period (in this case, in each quarter). For ATCOs, these coefficients reflect the wage premium afforded to pilots (relative to the economy a whole) in each quarter from 2017Q2 to 2021Q1. The low predictions of the models with time-SOC interaction terms therefore likely reflect the recent decline in hourly pay of the ATCO comparator SOC, airline pilots, due to COVID-19.

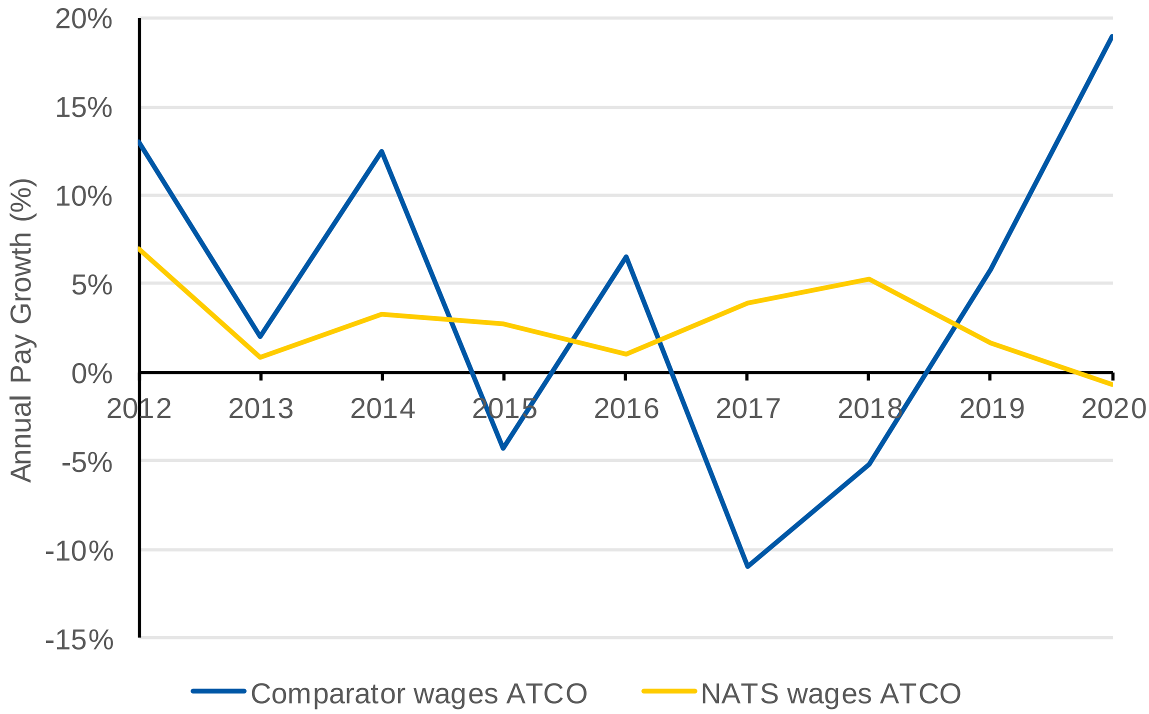
The models *without* time-SOC interaction terms are preferable to benchmark ATCO pay, particularly in the context of an assessment for the NR23 price control that covers a five-year period from 2023-2027.

- The models *without* time-SOC interaction terms allow us to assess whether current ATCO pay is in line with comparator pay on average over a longer time horizon, rather than being driven by short-term events. In recent quarters, the pay of the ATCO comparator group, airline pilots, has been reduced due to the short-term impact of COVID-19 on aviation. These short-term reductions should not be reflected in benchmark ATCO pay in the context of a price control that will run until the end of 2027, i.e. six years from now and three years after Eurocontrol predicts a recovery of aviation traffic to 2019 levels.[39]
- It is particularly important, when comparing ATCOs to airline pilots, to use a longer time horizon because average pay for airline pilots is more volatile than average pay for ATCOs. Figure 4.3 compares the year-on-year growth over the past ten years of average airline pilot pay, based on ASHE data, and average ATCO pay, based on NATS data.

[39] Eurocontrol (21 May 2021) Forecast Update 2021-2024. The central scenario predicts traffic levels at 95 per cent of 2019 levels in 2024. Link: https://www.eurocontrol.int/publication/eurocontrol-forecast-update-2021-2024

Airline pilot pay exhibits large swings, with growth exceeding 10 percentage points of magnitude in four of the ten years. ATCO pay growth never exceeds 10 percentage points of magnitude. The volatility of pilot pay may be because airlines operate in fiercely competitive labour markets with multiple competitors and volatile end-user demand that affects the demand for their services, often over short time horizons. The provision of Air Traffic Control services is restricted to NERL for the UK's complex airspace and is an essential service whatever the level of traffic, at least in the short term. Competition between NATS and rival employers for its staff is therefore longer-term in nature. Whatever the reason for the volatility, it makes the quarter-to-quarter benchmarking implied by models with time-SOC interaction terms undesirable.

Figure 4.3: Airline Pilot Pay is More Volatile, Year-on-Year, than NERL ATCO Pay



Source: NERA analysis of ASHE and NERL data (in financial years, e.g. 2020 is year ending March 2020)

4.2.2. ATSA actual wages are above the range of model predictions

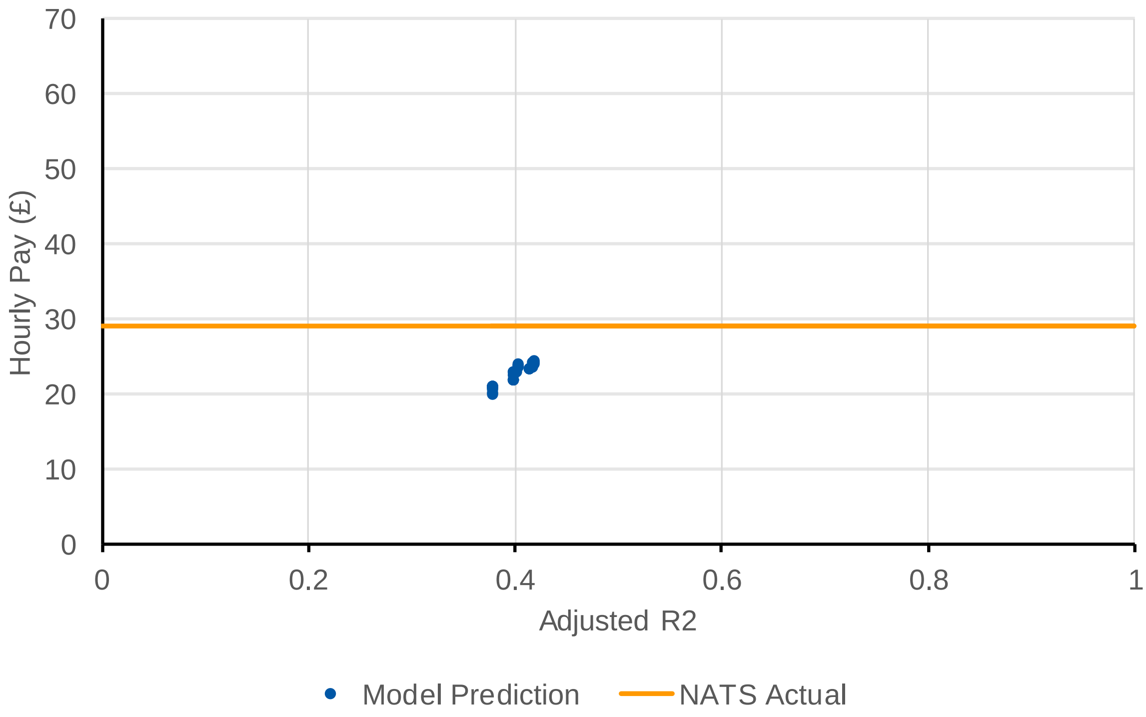
Our models predict pay equal to between 69-84 per cent of ATSA actual wages.

The fact that ATSA pay is not in line with the benchmark estimated based on selected comparator SOCs likely reflects the difficulty, identified in our previous report, of selecting appropriate comparator SOCs for the ATSA group.

The ATSA group covers a wide range of functions, depending in part on the grade within the group. The key skills for ATSAs in grade 3 and below are secretarial/clerical skills, whereas for ATSAs in grade 4 and above the key skills are strategy/leadership/advice skills.[40]

40 NERA (21 March 2018), Staff Operating Expenditure for Air Traffic Control: Prepared for NERL, p. 16

Figure 4.4: Predicted ATSA Total Hourly Pay



Source: NERA analysis of LFS and NERL data

ATSAs in grade 3 and below are therefore most comparable to SOCs within the 3-digit SOC group "Clerks and Assistants" (413), who had an average gross annual salary of £22,293 in 2020.[41] ATSAs in grade 4 and above, meanwhile, are more comparable to the 3-digit SOC group "Office managers and supervisors" (416), with an average gross annual salary of £30,862 in 2020. Since both 413 and 416 groups are included in the ATSA comparator SOC, the relative impact of each on benchmark pay will depend on their relative prevalence in the economy as a whole (specifically in respondents to the LFS), which may not reflect the relative prevalence of grade 3 and below vs. grade 4 and above at NERL.

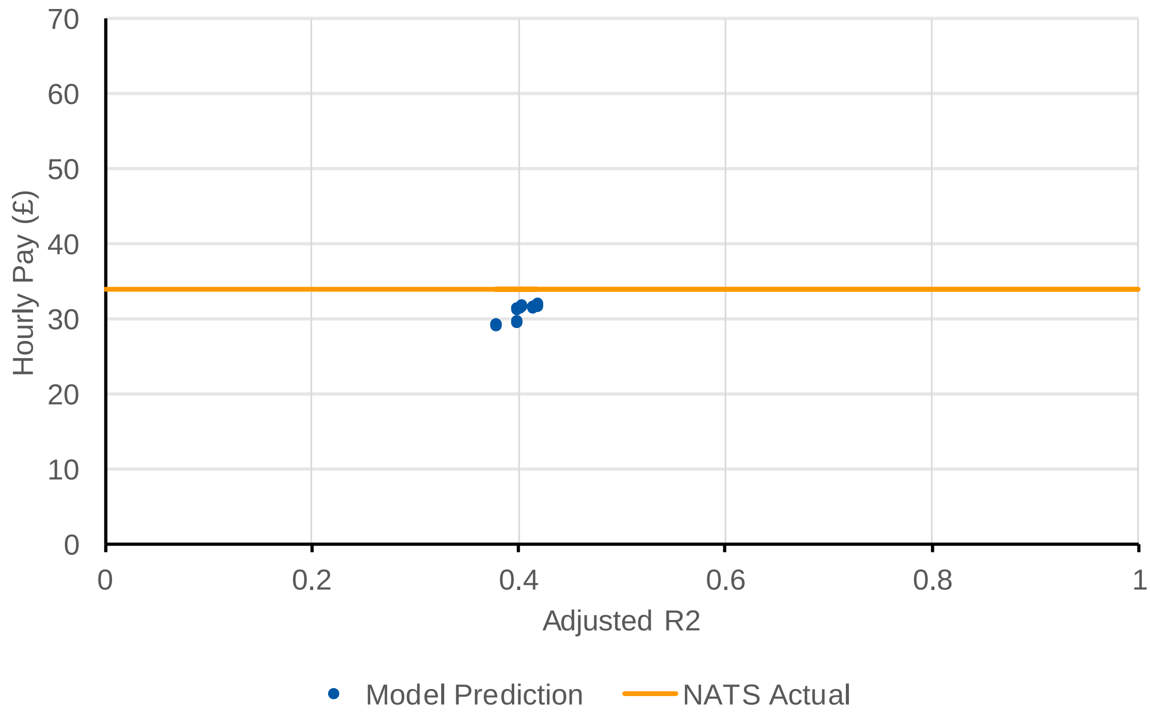
In addition to the difficulty of selecting an appropriate balance of comparators to reflect the wide range of functions performed by ATSAs, it was particularly difficult to find a comparator with a similar degree of responsibility for safety to ATSAs. Responsibility for the safety of others requires a particular set of skills, such as the ability to work calmly under pressure, that are also valued in other contexts and therefore command a wage premium but which are not recorded in publicly-available datasets such as the LFS.

4.2.3. ATCE actual wages are above the range of model predictions

Our models predict pay equal to between 85-94 per cent of ATCE actual wages. All models predict pay within a relatively limited range.

[41] ONS (3 November 2020), Earnings and hours worked, occupation by four-digit SOC: ASHE Table 14.7a. Link: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/occupation4digitsoc2010ashetable14

Figure 4.5: Predicted ATCE Total Hourly Pay



Source: NERA analysis of LFS and NERL data

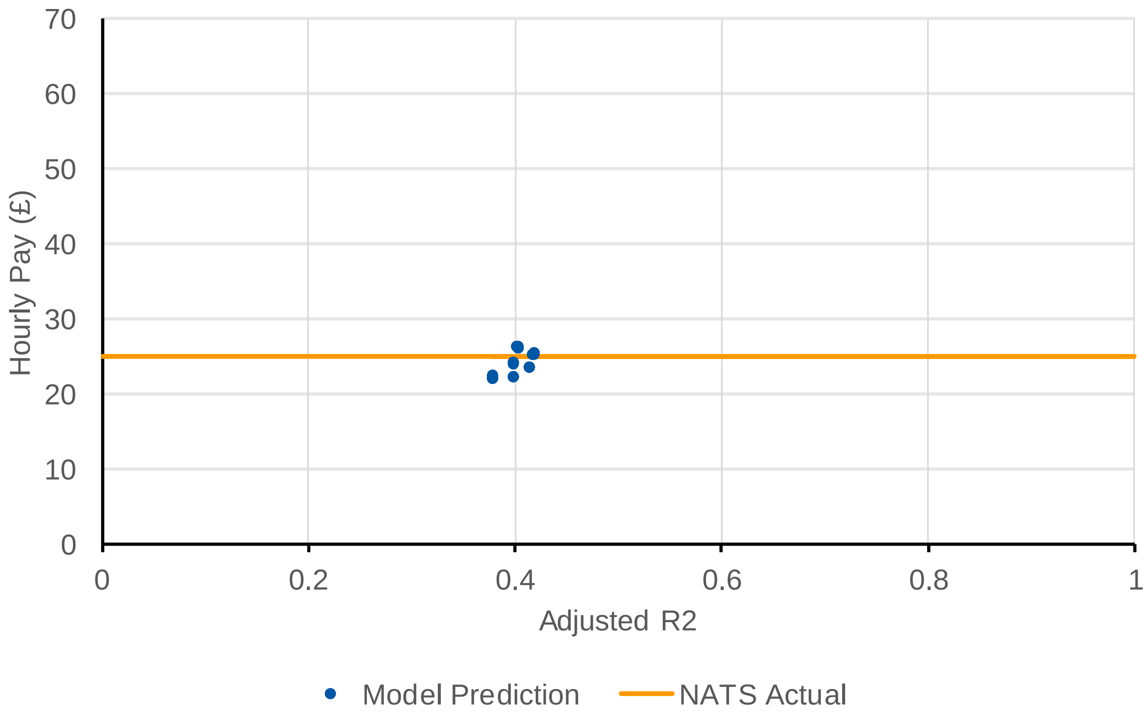
The fact that ATCE pay is not in line with comparator pay may reflect the changing nature of the ATCE role. We understand from NERL that historically, ATCEs performed work that required a mechanical engineering skillset ("nuts and bolts" engineers). NERL informs us that in recent years, the skills required of ATCEs have shifted more towards software and IT. The ASHE data shows that SOCs like "IT specialist managers (2132)" and "IT business analysts, architects and system designers (2133)" command higher average pay than SOCs like "Mechanical engineers (2122)". We include all of these SOCs as ATCE comparators, but the weight on each reflects their relative prevalence in the LFS rather than their relative prevalence among NERL staff. If mechanical engineers are more prevalent in the LFS than software and IT engineers, but the latter are more prevalent among NERL staff, the modelled benchmark would underestimate ATCE pay.

The discrepancy between model predicted and ATCE actual hourly pay may also be partly driven by the construction of the LFS variable *hourpay*, used to estimate the wage equations. The construction of the LFS variable *hourpay* means that the wage equation model predictions are likely to understate benchmark total hourly pay. This is discussed further in Section 4.4. Our analysis shows that ATCEs are particularly affected by the construction of *hourpay*, because the conclusion as to whether their actual pay is within the range of model predicted pay changes when we account for the construction of *hourpay*.

4.2.4. MSG and STAR actual wages are within the range of model predictions

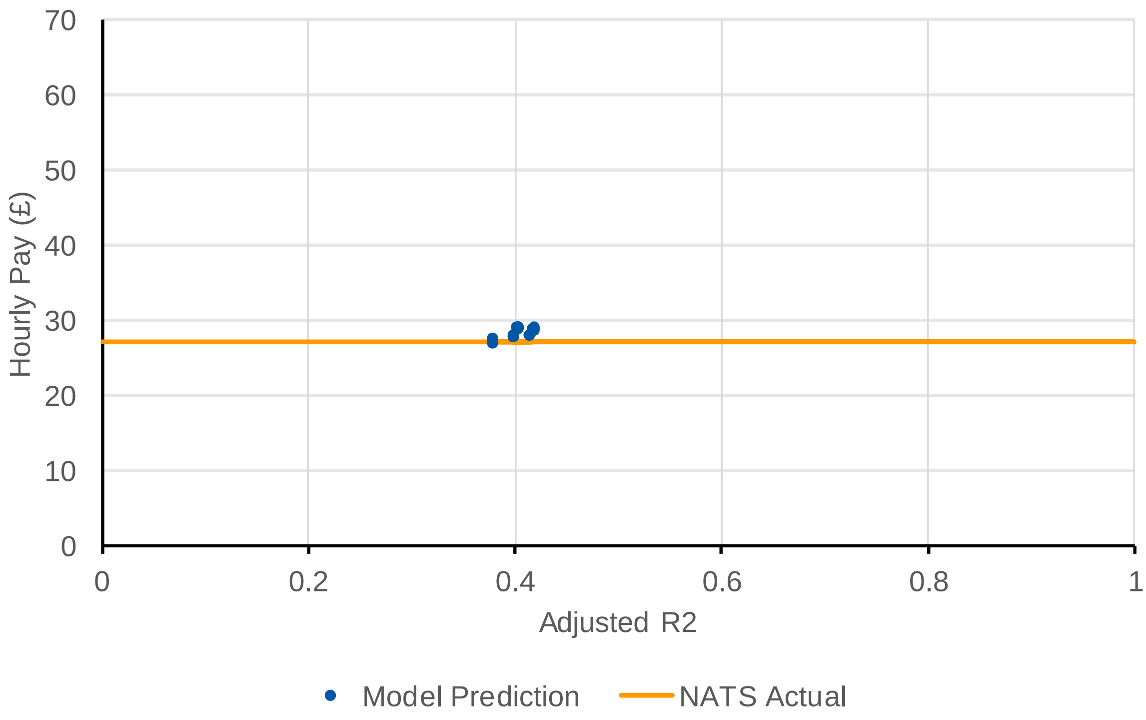
Our models predict pay equal to between 88-106 per cent of MSG actual wages, and between 100-107 per cent of STAR actual wages.

Figure 4.6: Predicted MSG Total Hourly Pay



Source: NERA analysis of LFS and NERL data

Figure 4.7: Predicted STAR Total Hourly Pay



Source: NERA analysis of LFS and NERL data

4.3. Comparison to Results of Previous Report

The results from this analysis are broadly in line with the results of our 2018 report. In that report, we found that actual NERL staff pay was within the range of model predictions for ATCOs, ATCEs, MSGs, and STARs, but above the range of model predictions for ATSAs.

Across all NERL staff groups, the adjusted-R2 of the models in the current report is lower than the adjusted-R2 of the models in the previous report. However, this does not imply that the quality of the models in this report, for the purposes of estimating NERL benchmark wages, is worse than the quality of models in the previous report.

The adjusted-R2 is a measure of the share of the variation in pay in the LFS dataset that can be explained by the wage equation model. There are two reasons that the lower adjusted-R2 of the wage equations in this report compared to the previous report do not indicate that the quality of the models is worse:

- ***The total variation in the new dataset may be higher.*** The dataset we use in this report has over twice as many observations as that used in the previous report.[42] We therefore expect that the total variation in the new dataset is higher. Further, there may be additional variation in the new dataset due to pay volatility as a result of COVID-19.

- ***The LFS dataset includes variation in pay that has no bearing on NERL staff benchmark pay, and the adjusted-R2 reflects all the variation in the data (not just data relevant to NERL).*** For example, there is variation in pay between SOCs that are not comparators for NERL staff groups (e.g. between journalists and restaurant staff), or variation in pay between employers with workforce of size 50 and workforce of size 100. In our previous report, we included more additional variables in our benchmarking analysis that had no bearing on NERL staff benchmark pay (e.g. all 4-digit SOCs, and measures of workforce size other than an indicator for having over 500 employees). Therefore, it is unsurprising that the models in the current report explain less of the variation in the LFS dataset as a whole than the models in the previous report, but are similarly successful in explaining NERL actual wages.

Table 4.4 reports the minimum and maximum predicted wages for each NERL staff group, as a share of actual pay, from both the current analysis and the 2018 wage benchmarking analysis. The shares reported from the 2018 analysis come directly from Table 4.2 of that report.[43]

[42] The dataset used in the previous report had approximately 60,000 observations. See NERA (21 March 2018), Staff Operating Expenditure for Air Traffic Control: Prepared for NERL, p. 22

[43] NERA (21 March 2018), Staff Operating Expenditure for Air Traffic Control: Prepared for NERL, p. 35

Table 4.4: Model Predicted Wages as Share of NERL Actual Wages - Current and Previous Benchmarking Exercises

	ATCO	ATSA	ATCE	MSG	STAR
Predicted wages as share of NERL actual wages – current project					
Minimum predicted wage	75%	69%	85%	88%	100%
Maximum predicted wage	106%	84%	94%	106%	107%
Predicted wages as share of NERL actual wages – previous project					
Minimum predicted wage	89%	61%	85%	81%	88%
Maximum predicted wage	105%	74%	102%	98%	104%

Source: NERA analysis of LFS and NERL data; NERA 2018 report

4.3.1. ATCO pay is in line with benchmark pay in both reports

For ATCOs, the range of predicted wages is wider than it was in the previous report. This is partially driven by the inclusion, in the current analysis, of models with interactions between time dummies and SOC indicators. Such models were not estimated in the previous report. If we exclude them here, the minimum predicted ATCO wage increases from £36.86 to £41.11, which is 84 per cent of the current NERL wage.

4.3.2. ATSA pay has converged towards benchmark pay since the previous report

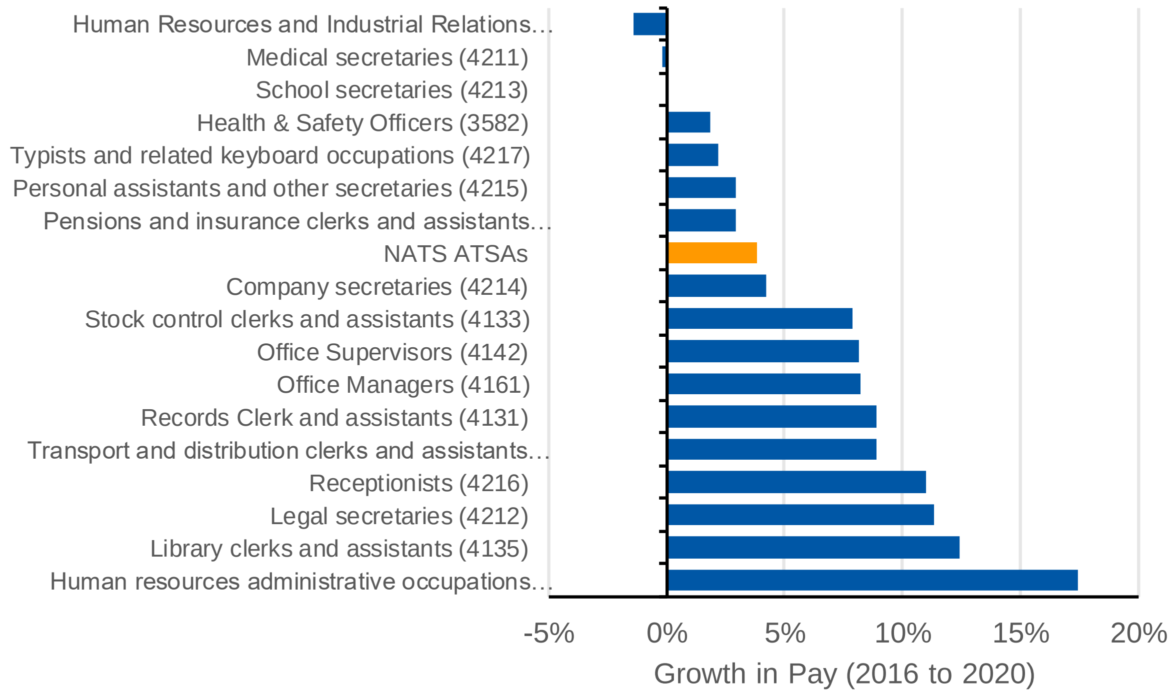
For ATSAs, the range of predictions is slightly wider than it was in the previous report and shifted upwards. In the current analysis, the difference between the maximum predicted wage for ATSAs and the actual ATSA wage is £4.67. In the previous analysis, the maximum predicted wage for ATSAs was £21.46 and the actual wage was £29.03, so the difference was £7.57.[44] This indicates that the gap between NERL actual pay and benchmark pay is decreasing.

Looking at data over the past ten years, we see that the growth rate of pay for most ATSA comparators has outstripped the growth rate of ATSA pay. Figure 4.8 shows the growth rate in average pay for comparator SOCs and NATS ATSAs between 2016 (one year prior to the year used for benchmarking in the previous project) and 2020 (one year prior to the year used for benchmarking in the current project, and the most recently available year in the ASHE dataset). The growth rate for NATS ATSA pay, at 3.83 per cent, is below the median pay growth for ATSA comparator SOCs.

This difference in historical growth rates can explain the convergence between NERL actual pay and benchmark pay from the previous report to the current report. This may be driven by other employers recognising that they need to increase their rates of compensation to recruit good employees with the required skillset. This sort of adjustment typically happens over a relatively long period of time, as recruitment occurs only when an employee needs to be replaced or a company is expanding.

[44] The maximum predicted wage for ATSAs (of those models selected as “preferred”) in the 2018 report was from the model “SIC1, SOC4, Education”. This was model 51. It included controls, education, a time trend, and squares of age and tenure; but excluded interactions and variables recording union membership or whether pay was set by union agreements.

Figure 4.8: ATSA Pay Growth from 2016 to 2020 is Below Median Pay Growth Among Comparator SOCs



Source: NERA analysis of ASHE and NERL data (in financial years, e.g. 2020 is year ending March 2020)

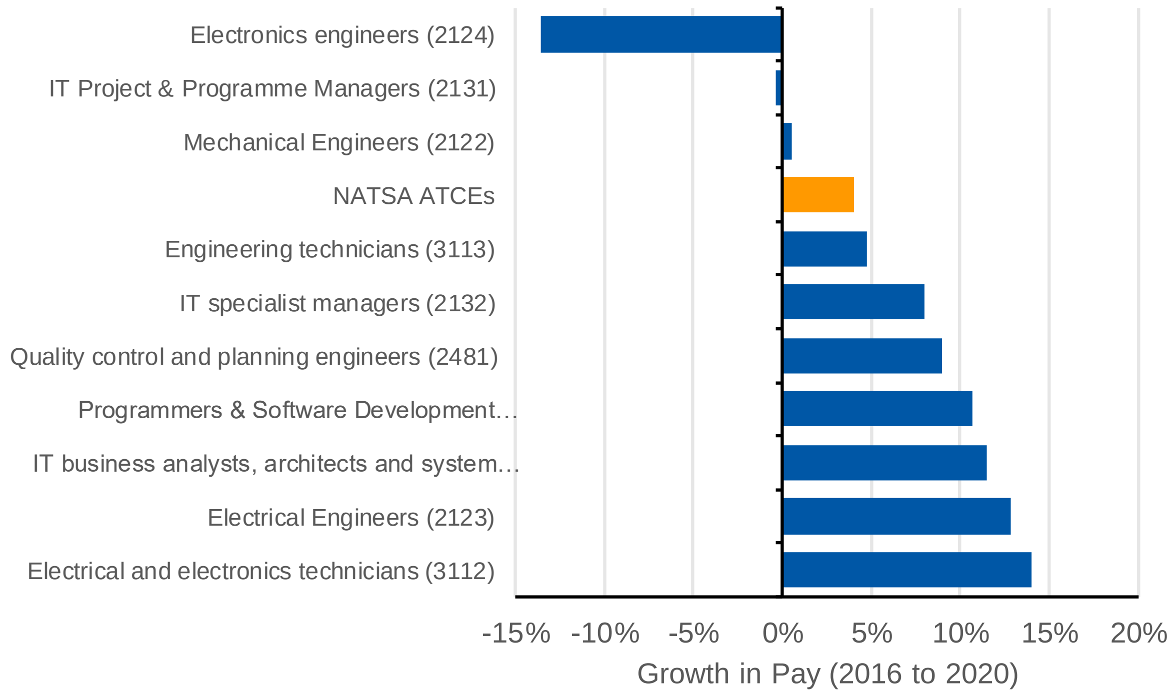
4.3.3. ATCE pay as a share of benchmark pay is above the level in the previous report

For ATCEs, the range of predictions is narrower than it was in the previous report, with the maximum shifted downwards.

From Figure 4.9, we see that ATCE pay growth from 2016 to 2020 is below median pay growth for comparator SOCs. This might be considered surprising, as in the previous report we found that ATCE pay was within the range of model predictions whereas in our current report ATCE pay is slightly above the range of model predictions. This change in result from the previous to the current report would suggest that ATCE pay has grown faster than comparator pay, on average.

One way to reconcile relatively low ATCE pay growth among comparators with the finding that ATCE pay is slightly above model predictions is to consider the relative prevalence of SOC comparators within the LFS. If most ATCE comparators observed in the LFS come from one of the SOC codes that saw slower pay growth than ATCEs (e.g. Mechanical Engineers (2122)), then the average pay growth of ATCE comparators within the LFS may be below the pay growth of ATCEs.

Figure 4.9: ATCE Pay Growth From 2016 to 2020 is Below Median Pay Growth Among Comparator SOCs



Source: NERA analysis of ASHE and NERL data (in financial years, e.g. 2020 is year ending March 2020)

4.3.4. MSG pay as a share of benchmark pay is below the level in our previous report

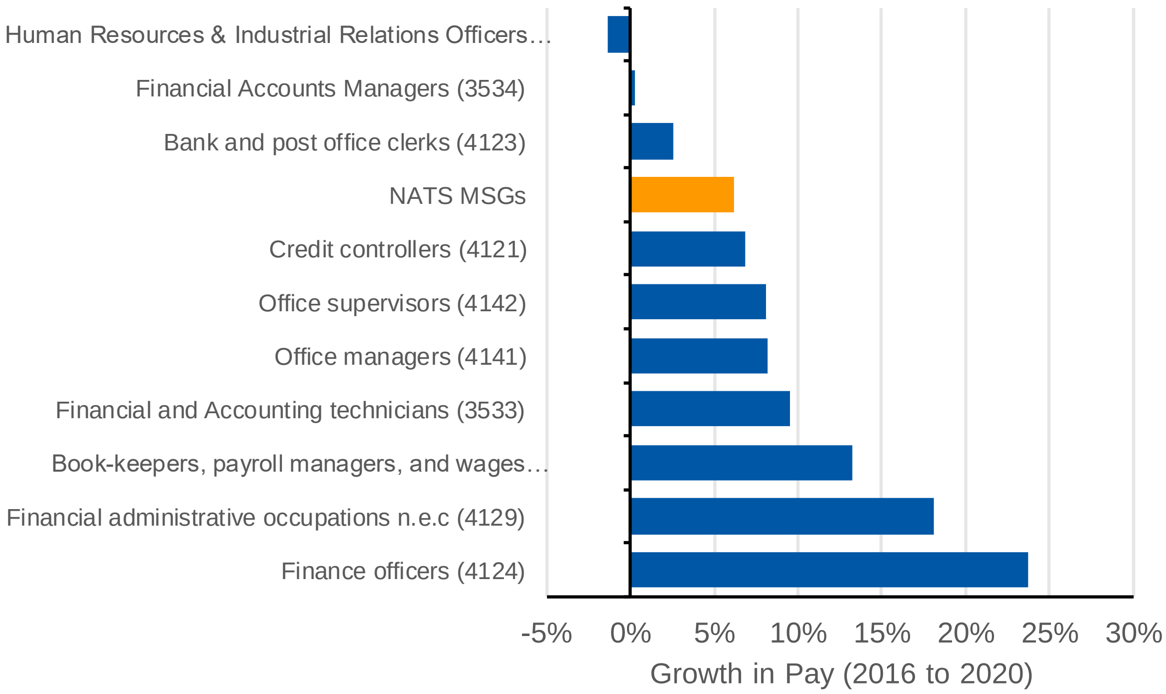
For MSGs, the range of predictions is similar to what it was in the previous report and shifted slightly upwards.

One factor driving the increase in predicted pay for MSGs may be that in this report we exclude demographics (gender, ethnicity, and marital status) from some of the models, whereas in the previous report we included demographics in all models. Models with demographics typically predict lower pay for MSGs than models without demographics because 74 per cent of MSGs are women, and our analysis shows that in the economy as a whole, women receive lower pay than men.[45]

Figure 4.10 shows that MSG pay growth has been below median pay growth for comparator SOCs in the recent past, which also contributes to explaining the convergence between NERL actual pay and benchmark pay.

45 In models that include demographics, the coefficient on the indicator for being male is between 0.10 and 0.14, indicating that on average in the economy as a whole men receive 10-14 per cent higher pay than women.

Figure 4.10: MSG Pay Growth from 2016 to 2020 is Below Median Pay Growth Among Comparator SOCs



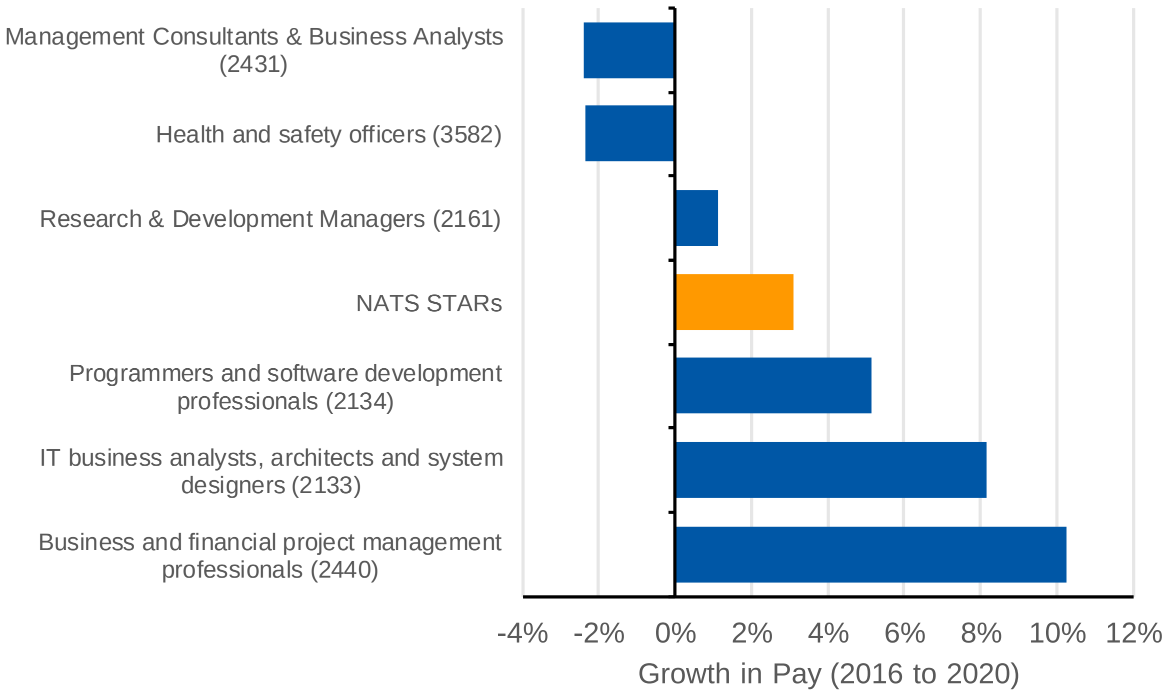
Source: NERA analysis of ASHE and NERL data (in financial years, e.g. 2020 is year ending March 2020)

4.3.5. STAR pay is in line with benchmark pay in both reports

For STARs, the range of predictions is narrower than it was in the previous report and shifted slightly upwards. This indicates that NERL actual pay is closer to benchmark pay than it was previously for STARs.

Figure 4.11 shows that STAR pay growth has been below pay growth for a number of comparator SOCs in the recent past, which may explain the convergence between NERL actual pay and benchmark pay if most STAR comparators observed in the LFS data come from one of the SOC codes that saw faster pay growth than STARs (e.g. Programmers and Software development professionals (2134).

Figure 4.11: STAR Pay Growth From 2016 to 2020 is in Line with Median Pay Growth Among Comparator SOCs



Source: NERA analysis of ASHE and NERL data (in financial years, e.g. 2020 is year ending March 2020)

4.4. Our Wage Equations May Understate NERL Benchmark Pay Due to the Construction of the LFS Outcome Variable

Interpreting our results requires careful consideration of the necessary limitations of the analysis. Our approach has been to use publicly available datasets to estimate economy-wide wages. There will be many factors for which we do not have data for the labour force as a whole and for which we cannot control. Accordingly, it would not be correct to suggest that our results need to exactly coincide with staff pay at NERL to provide reassurance to the CAA that overall pay is proportionate.

Underreporting of hourly pay in the publicly available dataset is one example of known bias which will reduce the level of total pay that we estimate for the general economy and tend to make NERL staff look relatively well paid.

The LFS variable that measures total pay per hour is *hourpay*. However, *hourpay* may systematically understate the market benchmark for NERL total hourly pay. This is because over 80 per cent of LFS participants report that *hourpay* does not include any additions to basic pay, e.g. annual bonuses and shift premia, which are included in NERL total hourly pay.

The dataset does not distinguish between respondents who do not receive additional pay and those who do not report it. There is reason to believe at least some respondents are not *reporting* additions to basic pay rather than not *receiving* additions to basic pay. For approximately 1/3 of the LFS sample, *hourpay* is calculated from total pay in the preceding month rather than annual total pay. Total pay in the preceding month would typically not include annual bonuses, and therefore *hourpay* calculated from pay in the preceding month would understate total hourly pay in the general economy.

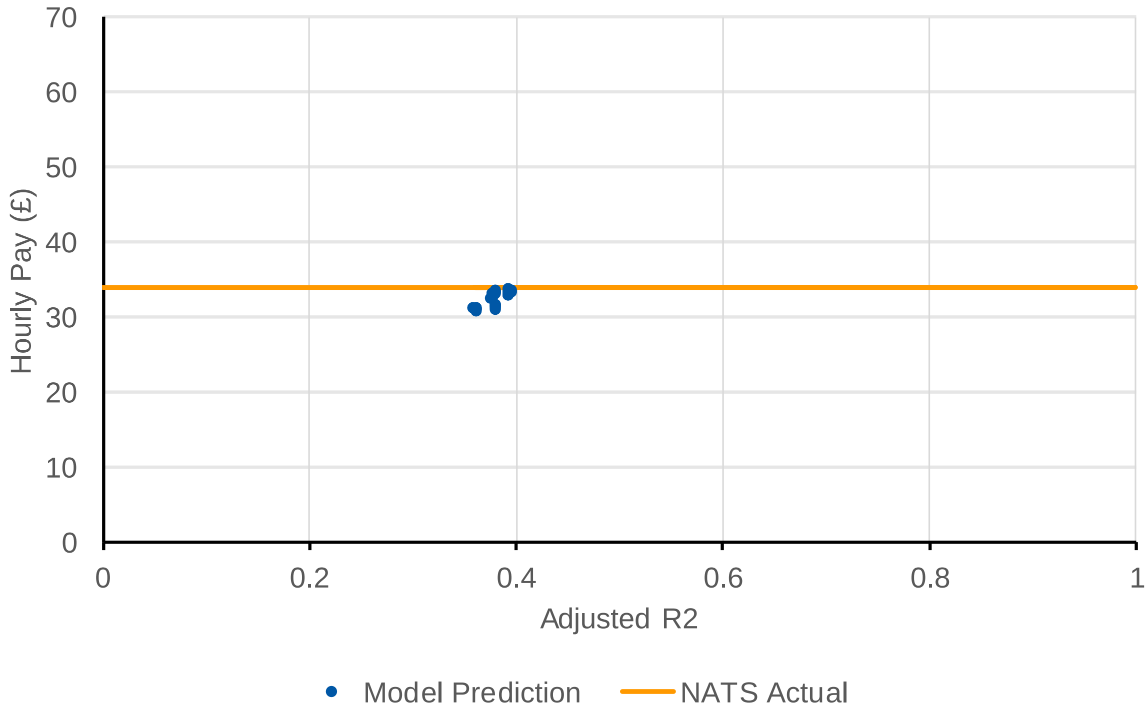
To test the impact of underreporting of pay on our results, we restricted the dataset to the 14.30 per cent of respondents who report that *hourpay* does include additions to basic pay and repeated our wage equation analysis on this smaller sample.

For all staff groups except ATCOs, this exercise caused an upward shift in the range of model predicted pay (for ATCOs, the exercise caused the range of model predicted pay to narrow but actual pay remains within the range; see Appendix E for details). In particular, for ATCEs the analysis changes our conclusion as to whether NERL pay aligns with benchmark pay. When we restrict our analysis to respondents who report that *hourpay* does include additions to basic pay, we find that ATCE pay is **within** the range of benchmark estimates, as seen in Figure 4.12.

One limitation of this analysis is the relatively small size of the restricted sample, at only 14.30 per cent of our total LFS sample. This sample may create problems in estimating some coefficients, in particular the time-SOC interaction coefficients which are estimated from the observations on SOC comparators in each quarter.

A second limitation of this analysis is that it may overstate benchmark pay, insofar as it excludes any comparators who do not *receive* additions to basic pay as well as those comparators who do not *report* additions to basic pay.

Figure 4.12: Predicted ATCE Total Hourly Pay From Wage Equations, with Sample Restricted to LFS Participants Who Report that Pay Includes Additions to Basic Pay



Source: NERA analysis of LFS and NERL data

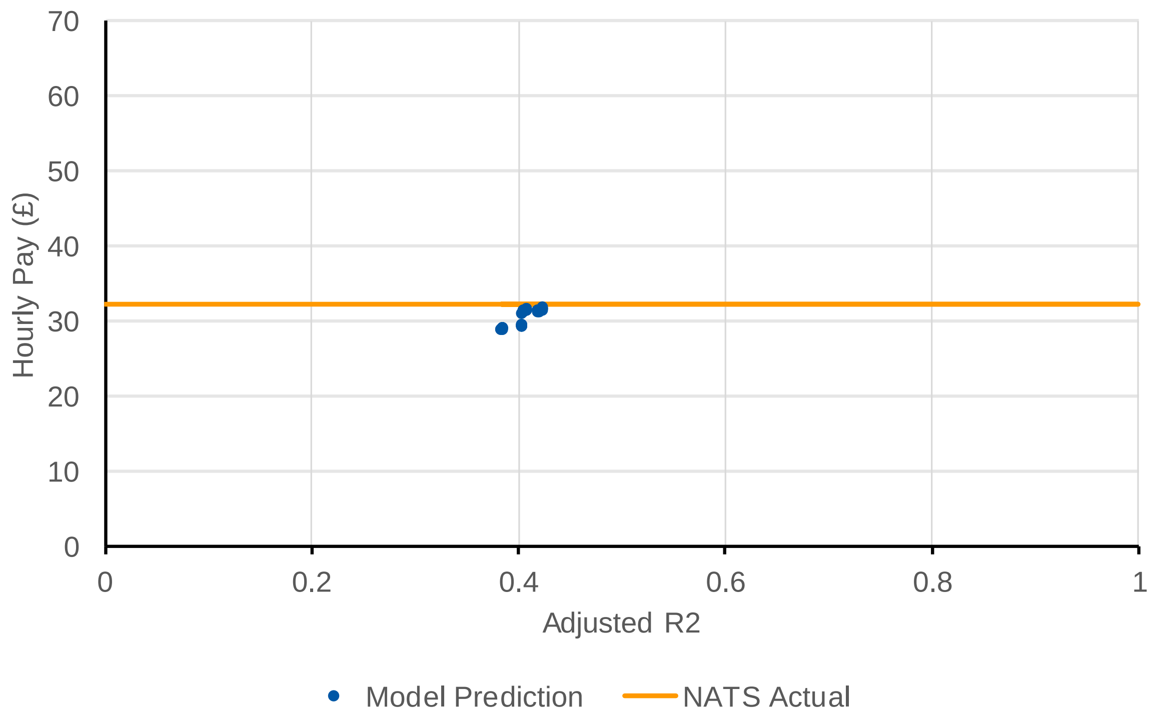
As a second test of the impact of the construction of *hourpay* on our results, and to address our concerns about the first test, we conduct a further benchmarking exercise of NERL actual *basic* pay. Basic pay excludes bonuses and other variable pay and so is not affected by possible underreporting of variable pay in the LFS.

We restrict the LFS sample to only the 84.92 per cent of respondents who report that *hourpay* does not include additions to basic pay, and repeat our wage equation analysis. We then compare the results to NERL basic hourly pay.

Again, for ATCEs this analysis changes our conclusion as to whether NERL pay is in line with benchmark pay. When we benchmark basic hourly pay, we find that ATCE pay is **within** the range of benchmark estimates, as seen in Figure 4.13.

The combined implication of these two tests is that the initial finding reported in Section 4.2, that ATCE total hourly pay exceeds the range of model predictions, may be driven by underreporting of variable pay by ATCE comparators in the LFS. If that is the case, then ATCE pay is in line with benchmark pay in the wider economy.

Figure 4.13: Predicted ATCE *Basic* Hourly Pay From Wage Equations Including *Only* LFS Respondents for Whom Gross Pay Did not Include Additions to Basic Pay



Source: NERA analysis of LFS and NERL data

4.5. Our Analysis Omits "Special Factors" Which Also Drive Wages

Our wage equation models do not account for all the factors that drive NERL staff wages. There may be "special" or "omitted" factors, which influence wages but are not easily measured or are not reflected by the set of variables available in the LFS. Such factors may be related to characteristics of the employer, the employee, or the job performed.

In our previous report, we identified a range of special factors that are likely to influence NERL wages but are not reflected in our wage equation models. These special factors remain relevant to NERL wages today.

We provide a brief review of these special factors here. Further detail on these special factors can be found in our previous report.

4.5.1. Non-liquid market for specialist staff

Our wage models did not consider the liquidity of the market for NERL staff on pay. In practice, the market for NERL staff, particularly ATCOs and ATSAs, is not very liquid.

Outside of NERL, the pool of individuals that have the necessary skills and training to perform NERL specialist roles and are both able and willing to work in the UK is extremely limited. The UK's exit from the European Union in January 2021 has further reduced the size of this pool by making it more difficult for qualified ATCOs and ATSAs from continental Europe to relocate to the UK.

Even if NERL can find suitably qualified external candidates, additional training is required before those external hires can be certified to operate NERL systems.

NERL therefore typically recruits most new staff at entry level and provides all necessary training internally. This approach involves long lead times between recruitment and operation: it takes years of full-time training to become a fully qualified ATCO.

The illiquid structure of the market for NERL specialist staff makes staff retention crucial and requires NERL to pay sufficiently high wages to support retention.

4.5.2. Highly unionised industry

Our wage equations control for the union influence through the variable *tucov*, which records whether wages are affected by union agreements. However, union influence is affected by more than simple union participation and penetration rates.

Union influence also depends on the potential for disruption due to industrial action, including strike action. Air traffic control strikes are particularly disruptive, due to knock-on effects on scheduling both before and after the strike, and the impact not only on air traffic into the UK but on traffic passing through the UK which must be re-routed to avoid UK airspace during any strike action.

The disruptive potential of air traffic control strikes means that there is a high degree of union influence at NERL, such that unions have the ability to negotiate higher wages.

4.5.3. Premia for education and training

Our wage equations include the highest educational qualification obtained as a key explanatory variable. Educational qualifications often do not have an intrinsic value to employers, but rather act as a signal of a worker's general capability. The premium paid for a university degree is typically a premium for evidence of diligence, general intelligence, communication skills, the capacity to understand and interpret guidance, and the capacity to work independently rather than a premium for subject matter knowledge.

NERL ATCOs typically do not require a university degree, but the rigorous selection and training process requires successful ATCOs to have all the characteristics listed above (diligence, general intelligence, etc.). Successful ATCOs are also required to have additional characteristics that would be valuable to other employers and are essential in the safety-critical roles that ATCOs perform, such as the ability to work calmly under pressure and strong attention to detail.

Successful ATCOs can therefore command a wage premium in the general economy at or above the premium afforded to individuals with university degrees, as their successful completion of the ATCO selection and training process serves as a signal of ability to other employers. In order to retain these high-value individuals, NERL must pay an appropriate wage premium.

4.5.4. Shift premium

Our wage equations did not take into account the impact of shift work on pay. In practice, compared to most professional jobs which offer regular working hours, NERL staff (particularly ATCOs and ATSAs) work shifts, including night shifts, and may justifiably demand a premium for their unusual working hours.

For example, the shift premium for junior doctors working night shifts is 37 per cent, while their shift premium for weekend work is between 10 and 15 per cent.[46] Junior doctors are comparable to air traffic control staff in determining an appropriate shift premium, due to the specialized knowledge required of the role and its safety-criticality.

4.6. Conclusion

NERL's staff undertake highly specific, frequently highly-trained jobs requiring skills which pay a premium over wages in the general economy. Our preferred models attempt to control for the activities that NERL's staff undertake by benchmarking wages of NERL staff against comparator SOCs within the broader economy.

Our models typically explain around 40 per cent of the variation in wages. All of the 37 specifications that we consider explain a similar share of variation in wages; the differences in the share of variation explained between specifications are small compared to the average total variation explained.

Whilst the 40 per cent figure compares well with many wage equations in the literature,[47] it is clear that a large share of variation in wages in the economy as a whole remains unexplained by our wage equation models. This is unsurprising. First, we do not include SOC codes to explain variation in wages between individuals in the economy that are not NERL staff comparators. Second, wages are also influenced by a range of factors that are not measured in the LFS data, e.g. familiarity with domain-specific technology. This second fact means that it is not realistic to anticipate that any econometric analysis would exactly and uniquely explain NERL's total wage levels.

---

[46] British Medical Association (January 2020), Junior doctor contract comparison. Link: https://www.bma.org.uk/media/2000/bma-junior-doctor-contract-comparison-jan2020.pdf

[47] For example, the classic wage decompositions from Oaxaca (1973) report R2 between 22% and 56%, depending on the variables included in the regression.

Overall, our modelling suggests that the CAA would not have a basis for concluding that NERL's wages are significantly above market levels. Our models include results which exceed NERL's wage levels for most categories of staff. Our analysis of the construction of the LFS *hourpay* variable also shows that the *hourpay* variable is not directly comparable to NERL total hourly pay and may lead the wage equations to understate the true market benchmark pay.

Further, our analysis of special factors that influence NERL wages demonstrates that it may be necessary for NERL to pay wages slightly above market benchmarks, to ensure that it can meet its service obligations. NERL must pay sufficient wages to retain existing staff, as there are costs and delays associated with training new staff; and must avoid the risk of strikes, which would cause severe disruptions to the aviation sector.

Appendix A. Data Used for Wage Benchmarking

This appendix includes descriptive statistics for both the LFS and NERL datasets.

Overall, NERL staff characteristics are broadly in line with their comparator groups, with a few exceptions:

- NERL staff are, on average, more highly educated than their comparators. As higher education yields a wage premium, this would lead NERL staff to be better paid on average than comparators.
- STARs are typically younger than their comparators. Age (proxying for experience) has a positive effect on wages, so this would lead STARs to be less well paid on average than comparators.
- ATCOs, ATCEs, and ATSAs typically have longer tenure at their current employer (NERL) than their comparators. Tenure has a positive effect on wages, so this would lead these staff groups to be better paid than comparators.
- NERL staff typically work slightly longer hours than their comparators.
- NERL staff are more likely to be full-time than their comparators.
- NERL staff are concentrated in the South of England and in Scotland.
- Demographically, NERL has a higher proportion of women than comparators in ATCO, MSG, and STAR grades, versus a higher proportion of men than comparators in ATSA and ATCE grades. NERL staff are also slightly more likely to be ethnically white than comparator occupations.

A.1. LFS Data

Table A.1: Summary Statistics from LFS Data

	ATCO	ATSA	ATCE	MSG	STAR
Number of observations	105	8,028	5,551	6,924	5,191
Average hourly pay	38.68	13.30	23.78	15.91	25.39
Controls relating to employer characteristics					
Over 500 employees	59.0%	20.2%	34.2%	18.2%	37.1%
Private sector employer	85.6%	67.8%	88.6%	85.3%	82.6%
SIC A: Agriculture, forestry, and fishing	0.0%	0.4%	0.1%	0.6%	0.1%
SIC B, D, E: Energy and water	0.0%	1.6%	4.2%	1.7%	3.3%
SIC C: Manufacturing	1.9%	7.9%	18.0%	9.0%	9.5%
SIC F: Construction	0.0%	4.2%	4.3%	5.1%	3.9%
SIC G, I: Distribution, hotels and restaurants	1.0%	10.6%	4.9%	11.5%	4.2%
SIC H, J: Transport and communication	67.6%	6.4%	33.1%	7.1%	26.4%
SIC K, L, M, N: Banking and finance	3.8%	23.6%	21.1%	40.2%	30.9%
SIC O, P, Q: Public admin, education and health	25.7%	40.2%	12.0%	19.7%	19.0%
SIC R, S, T, U: Other services	0.0%	5.2%	2.4%	5.0%	2.8%

Controls relating to employee characteristics					
Qualification 1: Degree or equivalent	57.1%	25.5%	56.7%	34.3%	66.4%
Qualification 2: Higher education	7.6%	10.9%	12.4%	9.8%	9.0%
Qualification 3: GCE A level or equivalent	27.6%	26.3%	18.7%	23.6%	15.0%
Qualification 4: GCSE grades A*-C or equivalent	4.8%	28.4%	8.8%	24.8%	7.5%
Qualification 5: Other	2.9%	5.2%	2.4%	4.3%	1.5%
Qualification 6: None	0.0%	2.8%	0.6%	2.4%	0.4%
Qualification 7: Don't know	0.0%	0.9%	0.4%	0.7%	0.2%
Age	43.8	45.1	42.1	43.3	41.9
Tenure	10.4	9.6	9.2	9.5	8.7
Pay/conditions affected by union agreements	67.9%	26.0%	24.1%	19.6%	22.9%
Controls relating to job characteristics					
Full-time	89.5%	66.2%	95.9%	73.3%	92.1%
Usual hours worked	39.6	31.8	38.1	33.2	37.3
Excess hours worked	0.9	1.4	2.5	1.7	2.5
Region 1: North	14.3%	22.8%	20.5%	23.5%	20.4%
Region 2: Midlands and East	20.0%	19.5%	18.8%	18.9%	16.8%
Region 3: London	28.6%	11.3%	15.9%	13.7%	19.6%
Region 4: South	21.0%	29.4%	29.7%	27.5%	28.8%
Region 5: Wales	2.9%	4.2%	3.2%	3.6%	3.2%
Region 6: Scotland	7.6%	6.9%	7.2%	6.9%	7.1%
Region 7: Northern Ireland	1.9%	5.8%	4.3%	5.8%	3.9%
Region 8: Outside UK	3.8%	0.0%	0.4%	0.1%	0.2%
Time Effects					
2017Q2	3.8%	6.9%	5.1%	5.8%	5.0%
2017Q3	5.7%	6.4%	5.6%	6.8%	5.9%
2017Q4	8.6%	7.3%	6.8%	7.3%	6.2%
2018Q1	10.5%	6.0%	5.6%	5.8%	5.6%
2018Q2	4.8%	7.2%	6.4%	6.7%	6.1%
2018Q3	7.6%	6.7%	6.3%	6.6%	5.9%
2018Q4	6.7%	6.5%	6.2%	6.7%	6.6%
2019Q1	3.8%	6.0%	6.2%	6.0%	6.2%
2019Q2	3.8%	6.1%	6.8%	6.6%	6.6%
2019Q3	8.6%	5.9%	7.4%	6.2%	6.8%
2019Q4	8.6%	6.1%	6.6%	6.0%	6.9%
2020Q1	3.8%	5.5%	5.7%	5.0%	5.7%
2020Q2	1.9%	4.8%	4.5%	5.0%	4.4%
2020Q3	7.6%	5.9%	7.1%	6.3%	7.2%
2020Q4	3.8%	5.8%	6.6%	6.2%	6.8%
2021Q1	10.5%	6.6%	6.7%	6.7%	7.6%
2021Q2	0.0%	0.5%	0.5%	0.4%	0.4%
Demographics					
Sex = male	94.3%	21.3%	84.2%	29.0%	68.5%

Marital status 1: Single, never married	28.6%	30.8%	32.6%	32.7%	31.4%
Marital status 2: Married, living with spouse	67.6%	53.6%	58.8%	53.6%	59.7%
Marital status 3: Married, separated from spouse	2.9%	2.8%	1.9%	2.9%	2.1%
Marital status 4: Divorced	1.0%	10.6%	5.7%	8.8%	5.7%
Marital status 5: Widowed	0.0%	1.9%	0.6%	1.8%	0.6%
Marital status 6: Currently or previously in civil partnership	0.0%	0.2%	0.4%	0.3%	0.5%
Ethnicity 1: White	95.2%	94.2%	87.4%	91.2%	86.6%
Ethnicity 2: Mixed/multiple ethnic groups	1.9%	0.8%	1.4%	0.7%	1.3%
Ethnicity 3: Indian	0.0%	1.3%	5.7%	2.6%	6.0%
Ethnicity 4: Pakistani	0.0%	0.5%	0.9%	0.8%	0.9%
Ethnicity 5: Bangladeshi	0.0%	0.2%	0.2%	0.4%	0.3%
Ethnicity 6: Chinese	1.0%	0.3%	1.0%	0.6%	1.1%
Ethnicity 7: Any other Asian background	0.0%	0.4%	1.0%	0.7%	1.1%
Ethnicity 8: Black/African/Caribbean/Black British	1.0%	1.8%	1.4%	1.9%	1.5%
Ethnicity 9: Other ethnic group	1.0%	0.4%	1.0%	1.0%	1.2%

Source: NERA analysis of LFS data

A.2. Data Provided by NERL

Table A.2: Summary Statistics from NERL Data

	Source	ATCO	ATSA	ATCE	MSG	STAR
Number of observations	Database	1,266	462	621	396	123
Number of observations	Survey	103	76	248	204	52
Average hourly pay	Both*	49.04	29.03	33.97	24.91	27.14
Controls relating to employer characteristics						
Over 500 employees	Inferred	100%	100%	100%	100%	100%
Private sector employer	Inferred	100%	100%	100%	100%	100%
SIC A: Agriculture, forestry, and fishing	Inferred	0%	0%	0%	0%	0%
SIC B, D, E: Energy and water	Inferred	0%	0%	0%	0%	0%
SIC C: Manufacturing	Inferred	0%	0%	0%	0%	0%
SIC F: Construction	Inferred	0%	0%	0%	0%	0%
SIC G, I: Distribution, hotels and restaurants	Inferred	0%	0%	0%	0%	0%
SIC H, J: Transport and communication	Inferred	100%	100%	100%	100%	100%
SIC K, L, M, N: Banking and finance	Inferred	0%	0%	0%	0%	0%
SIC O, P, Q: Public admin, education and health	Inferred	0%	0%	0%	0%	0%
SIC R, S, T, U: Other services	Inferred	0%	0%	0%	0%	0%
Controls relating to employee characteristics						
Qualification 1: Degree or equivalent	Survey	63.1%	52.6%	71.4%	55.4%	100%
Qualification 2: Higher education	Survey	10.7%	15.8%	24.6%	12.7%	0.0%
Qualification 3: GCE A level or equivalent	Survey	17.5%	17.1%	2.0%	16.2%	0.0%

Qualification 4: GCSE grades A*-C or equivalent	Survey	3.9%	10.5%	0.8%	9.3%	0.0%
Qualification 5: Other	Survey	4.9%	3.9%	1.2%	6.4%	0.0%
Qualification 6: None	Survey	0.0%	0.0%	0.0%	0.0%	0.0%
Qualification 7: Don't know	Survey	0.0%	0.0%	0.0%	0.0%	0.0%
Age	Database	41.2	44.7	44.2	42.1	33.8
Tenure	Database	16.8	17.5	13.5	9.1	6.4
Pay/conditions affected by union agreements	Database	100%	100%	100%	100%	100%
Controls relating to job characteristics						
Full-time	Database	91.2%	90.9%	96.5%	84.8%	90.2%
Usual hours worked	Survey	36.6	37.6	38.5	37.4	36.6
Excess hours worked	Both*	2.4	3.4	3.8	3.6	2.7
Region 1: North	Database	3.7%	0.6%	1.0%	1.3%	1.6%
Region 2: Midlands and East	Database	3.3%	1.5%	1.3%	2.0%	3.3%
Region 3: London	Database	0.8%	0.6%	1.9%	1.3%	0.8%
Region 4: South	Database	67.1%	75.5%	82.6%	82.8%	88.6%
Region 5: Wales	Database	0.9%	0.4%	0.3%	0.5%	0.8%
Region 6: Scotland	Database	23.2%	21.0%	12.7%	11.4%	4.9%
Region 7: Northern Ireland	Database	0.6%	0.0%	0.2%	0.5%	0.0%
Region 8: Outside UK	Database	0.2%	0.0%	0.0%	0.3%	0.0%
Time Effects						
2017Q2	Inferred	0%	0%	0%	0%	0%
2017Q3	Inferred	0%	0%	0%	0%	0%
2017Q4	Inferred	0%	0%	0%	0%	0%
2018Q1	Inferred	0%	0%	0%	0%	0%
2018Q2	Inferred	0%	0%	0%	0%	0%
2018Q3	Inferred	0%	0%	0%	0%	0%
2018Q4	Inferred	0%	0%	0%	0%	0%
2019Q1	Inferred	0%	0%	0%	0%	0%
2019Q2	Inferred	0%	0%	0%	0%	0%
2019Q3	Inferred	0%	0%	0%	0%	0%
2019Q4	Inferred	0%	0%	0%	0%	0%
2020Q1	Inferred	0%	0%	0%	0%	0%
2020Q2	Inferred	0%	0%	0%	0%	0%
2020Q3	Inferred	0%	0%	0%	0%	0%
2020Q4	Inferred	0%	0%	0%	0%	0%
2021Q1	Inferred	100%	100%	100%	100%	100%
2021Q2	Inferred	0%	0%	0%	0%	0%
Demographics						
Sex = male	Database	79.1%	67.1%	89.7%	26.0%	53.7%
Marital status 1: Single, never married	Database	15.0%	15.4%	16.6%	11.4%	13.8%

Marital status 2: Married, living with spouse	Database	45.3%	53.9%	55.1%	48.7%	38.2%
Marital status 3: Married, separated from spouse	Database	1.2%	2.8%	1.0%	1.5%	0.0%
Marital status 4: Divorced	Database	2.7%	3.7%	3.1%	4.5%	0.0%
Marital status 5: Widowed	Database	0.1%	0.0%	0.2%	0.5%	0.0%
Marital status 6: Currently or previously in civil partnership	Database	0.3%	0.2%	0.0%	0.5%	0.0%
Marital status 7: Unknown*	Database	35.4%	24.0%	24.2%	32.8%	48.0%
Ethnicity 1: White	Database	98.4%	98.0%	91.8%	95.4%	89.9%
Ethnicity 2: Mixed/multiple ethnic groups	Database	0.2%	0.2%	1.7%	0.5%	1.1%
Ethnicity 3: Indian	Database	0.2%	1.0%	1.9%	1.1%	1.1%
Ethnicity 4: Pakistani	Database	0.2%	0.2%	0.3%	0.5%	0.0%
Ethnicity 5: Bangladeshi	Database	0.0%	0.0%	0.3%	0.3%	0.0%
Ethnicity 6: Chinese	Database	0.2%	0.0%	0.2%	0.5%	2.2%
Ethnicity 7: Any other Asian background	Database	0.2%	0.0%	0.3%	0.3%	0.0%
Ethnicity 8: Black/African/Caribbean/Black British	Database	0.1%	0.2%	1.7%	0.8%	2.2%
Ethnicity 9: Other ethnic group	Database	0.5%	0.2%	1.7%	0.5%	3.4%

Source: NERA analysis of NERL data
Notes:

(1) *For both average hourly pay and excess hours worked, we combine data from the NERL internal database and the NERL staff survey.*

(2) *NERL's internal database reports marital status as "unknown" for some workers. To align our analysis with the LFS data, we recalculated the marital status shares of NERL staff over the population for whom marital status was known.*

Table A.3: Number of Respondents to the NERL Staff Survey by Grade

Grade	ATCO	ATSA	ATCE	MSG	STAR
1	38	<10	19	<10	<10
2	44	<10	42	33	<10
3	<10	10	90	38	16
4	<10	22	69	56	15
5	<10	<10	23	25	13
6	<10	<10	<10	26	<10
7	<10	<10	<10	13	<10
unknown	17	37	<10	<10	<10

Source: NERA analysis of NERL data
Note: Where there are fewer than 10 respondents, the exact number is not reported to preserve anonymity

Appendix B. Methodological Appendix

B.1. Imputation

In our wage equations, an important explanatory variable is the indicator for whether wages are subject to union agreements. In the LFS dataset, this is *tucov*.

Unfortunately, the LFS only asks about union membership and whether wages are subject to union agreements in the fourth quarter of each year. This means that information on these variables is missing in all other quarters.

In our previous wage benchmarking exercise for NERL, we dealt with this limitation of the LFS by using only data from the fourth quarter in each year.[48] However, this approach is undesirable in this exercise, where we want to use as much data from more recent periods as possible to reflect the impact of the COVID-19 pandemic.

In this study, we deal with this limitation of the LFS by *imputing* the missing values for *tucov* in the first, second, and third quarters. Imputation is an econometric technique that generates replacements for the missing values of tucov based on the available data for that observation on other variables. We give a very simple example in Box A below.

A variety of imputation techniques are available. The technique we adopt is *multiple imputation*. This is the preferred approach in the econometric literature.[49] Rather than generating a single replacement for the missing values of *tucov*, we generate many possible replacements for the missing values of *tucov*. We estimate a wage equation with each set of replacements. The final wage equation is an average of these wage equations. This approach reduces the risk that findings could be driven by the particular choice of imputed value.

The key assumption required for multiple imputation is that the missing values must be *missing at random (MAR)*. MAR means that the fact that a particular observation has a missing value for *tucov* must not be related to the true (missing) value of *tucov*.[50]

This assumption is credible for the *tucov* variable. The missing observations for *tucov* are in quarters 1-3, when the LFS did not ask about union membership and the impact of union agreements on wages. There is no reason to think that the propensity for wages to depend on union agreements varies systematically depending on the quarter of the year.

This assumption would not be credible for log hourly wage. The LFS asks about wages in every quarter, so missing observations for wages typically arise because the respondent chose not to give information on their wages. Evidence suggests that respondents may be less willing to answer questions about pay if they are either high earners or low earners.[51] This means that missing observations for wages are likely to be relatively high or low wages, i.e. there is a relationship between the value and the fact that the value is missing, which violates the MAR assumption. We therefore do not impute missing data on log hourly wages.

---

48 NERA (21 March 2018), Staff Operating Expenditure for Air Traffic Control: Prepared for NERL, p. 22

49 Cameron, A.C. and Trivedi, P.K. (2005), *Microeconometrics: Methods and Applications*, p939 (Ch. 27.9).

50 Formally, MAR requires that tucov is independent of the probability that tucov is missing.

51 See for example Lillard, L., Smith, J.P., Welch, F. (1986) *What do we really know about wages? The importance of nonreporting and census imputation, Journal of Political Economy, 94(3).*

BOX A

Consider a dataset (below) with two variables: the indicator for union influence on wages *tucov* and an indicator for whether the employer is a public sector employer *publicr*. The dataset contains twenty observations on *publicr* but is missing data on *tucov* for ten of those observations (observations with ID 11-20).

We can **impute** the missing values of *tucov*, based on the relationship we observe between *tucov* and *publicr* in the portion of the dataset where neither variable is missing (observations with ID 1-10). Looking at observations 1-10, we see that 80% of respondents in the public sector (*publicr* = Yes) have wages influenced by unions (*tucov* = Yes), but only 20% of respondents in the private sector (*publicr* = No) do. These probabilities constitute our **imputation model**, which describes the relationship between the missing variable *tucov* and the observed variable *publicr*. For a more complex dataset the imputation model would typically be a regression model.

We then apply these probabilities to generate imputed values for *tucov* for observations 11-20. Since the imputation is probabilistic, different imputations are possible (*tucov_imput1* and *tucov_imput2* are two examples). Both imputations show that 80 per cent of the public sector workers who do not report whether they are covered by union wage agreements (selected at random) are in fact covered by union wage agreements. And both show that 20 per cent of private sector workers who do not report whether they are covered by union wage agreements (selected at random) are in fact covered by union wage agreements.

Multiple imputation accounts for the existence of many different possible imputations by constructing several imputations, performing any secondary analysis (e.g. a regression of a third variable, *hourpay*, on *publicr* and *tucov*) with each imputation, and then taking the average result of the secondary analysis.

ID	publicr	tucov	tucov_imput1	tucov_imput2
1	Yes	Yes	Yes	Yes
2	Yes	Yes	Yes	Yes
3	Yes	Yes	Yes	Yes
4	Yes	Yes	Yes	Yes
5	Yes	No	No	No
6	No	Yes	Yes	Yes
7	No	No	No	No
8	No	No	No	No
9	No	No	No	No
10	No	No	No	No
11	Yes	-	*Yes*	*Yes*
12	Yes	-	*No*	*Yes*
13	Yes	-	*Yes*	*Yes*
14	Yes	-	*Yes*	*No*
15	Yes	-	*Yes*	*Yes*
16	No	-	*No*	*No*
17	No	-	*Yes*	*No*
18	No	-	*No*	*Yes*
19	No	-	*No*	*No*
20	No	-	*No*	*No*

We performed imputation using the "mi" suite of commands in the statistical software Stata (Version 16). The variables for which some missing observations were imputed were as follows: indicators for whether pay is set by union agreements, private sector employer, and an employer with over 500 employees; and continuous variables tenure, excess hours, and their squares. Our imputation model also included the outcome variable (log hourpay), as well as all explanatory variables listed in Table 3.1, with the exception of time-SIC and time-SOC interactions. We excluded time-SIC and time-SOC interactions due to multicollinearity problems arising in the imputation model. We performed 30 imputations.[52]

The "mi" suite of commands relies on an iterative algorithm to estimate the imputation model and thus to produce imputed values for missing data. One can evaluate the performance of the imputation procedure by checking for convergence of the iterative algorithm. We check convergence by examining trace plots of the mean and standard deviation of the imputed values.[53]

Figure B.14 shows the trace plots for the mean and standard deviation of the imputed values of *tucov*. There are 30 lines on each plot, one line for each imputation. The X-axis reports the iteration number and the Y-axes report the values of the mean and standard deviation at each iteration.

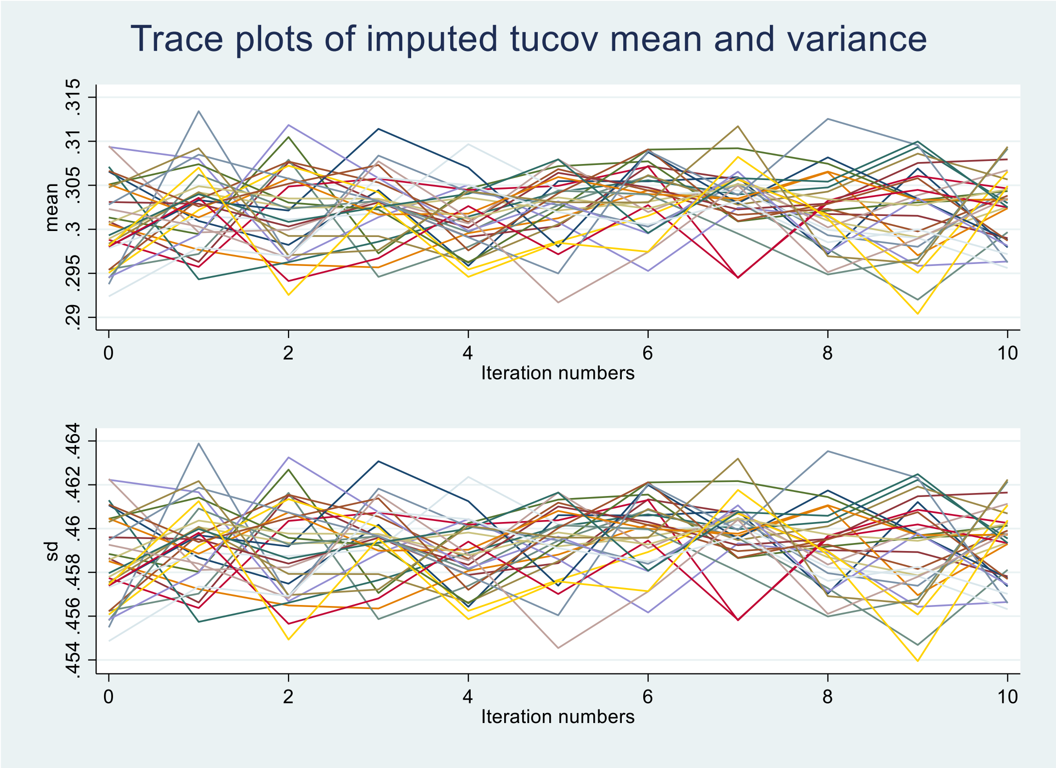
A systematic trend in the mean and standard deviation of the imputed variables over the course of multiple iterations of estimation would suggest that the imputation procedure had not converged on a reasonable set of imputed data. For instance, if the lines on the plot showed a systematically increasing mean, further iterations might increase the mean further. As a result, the imputed dataset from the 10th iteration would not necessarily be a reasonable imputed dataset for the purpose of our analysis.

In practice, none of the lines shows a clear trend. The lack of a systematic trend in any of the lines indicates successful convergence of the algorithm after ten iterations for each of the 30 imputations. The imputation procedure therefore performs well for the variable *tucov*. The trace plots of the other imputed variables also show convergence.

52 The standard rule of thumb is that the number of iterations should be at least equal to the percentage of observations with missing data – see Nguyen, C.D., Carlin, J.B, and Lee, K.J (2017) *Model checking in multiple imputation: an overview and case study, Emerging Themes in Epidemiology 14(8)*. In our sample, we are missing data for *tucov* from 76 per cent of observations; therefore, we originally performed 80 imputations. However, we achieved very similar results with both 30 and 80 imputations and so for ease of estimation we reduced the total number of imputations to 30.

53 StataCorp LLC (2021), Stata Multiple-Imputation Reference Manual Release 17, p. 146. Link: https://www.stata.com/manuals/mi.pdf

Figure B.14: The Trace Plot for tucov Shows that the Imputation Algorithm has Converged Successfully



Source: NERA analysis of LFS data

B.2. Conversion from Logs to Levels

Our wage equations use the log of hourly pay as the outcome variable to be explained, rather than the level of hourly pay. It is standard practice in wage equation analysis to use the log of the pay variable as the outcome.[54] This is for two reasons. First, using the log of pay as the outcome means that the model allows changes to the explanatory variable to impact wages in percentage terms, rather than level terms. This is more consistent with actual practice (for example, wage increases with tenure are typically defined as a percentage increase relative to the previous year). Second, the distribution of wages typically has a long right tail (i.e. a small number of people have wages much higher than the population average). A regression using the level of hourly pay would place very high weight on these observations. Taking the log of pay corrects for this long right tail and reduces the excess influence of these observations on the regression.

The initial predictions of NERL wages that we derive from our wage equations are therefore predictions of the log of hourly pay, rather than the level of hourly pay. In order to get predictions of the level of hourly pay, we must apply a reverse transformation to the predicted wage in logs.

There are a number of reverse transformation options available.

[54] See for example Mincer, J. (1974). *Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2.*

1. ***Naïve transformation.*** A naïve approach would be to simply take the exponent of the predicted wage in logs. However, due to Jensen's inequality this approach systematically underestimates the true predicted level implied by the regression in logs. Despite its limitations, the naïve estimator can give reasonable results in some situations.[55]

$$\hat{y}_i = \exp\left(\widehat{log\, y_i}\right) = \exp\left(\hat{\alpha} + \hat{\beta} X_i\right)$$

2. ***Retransformation under assumption of normality.*** To address the bias in the naïve transformation, if the outcome variable $\log y$ is close to normally distributed then the prediction can be adjusted by the estimated variance of $\log y$.[56]

$$\hat{y}_i = \exp\left(\widehat{log\, y_i} + \frac{s^2}{2}\right); s^2 = \frac{1}{n-k}\sum_i \left(\log y_i - \widehat{\log y_i}\right)^2$$

   This is the approach we used in the previous project, with the equation expressed as:

$$\hat{y}_i = \exp\left(\widehat{log\, y_i}\right) \times \exp\left(\frac{RMSE^2}{2}\right); RMSE = \sqrt{\frac{1}{n-k}\sum_i \left(\log y_i - \widehat{\log y_i}\right)^2}$$

3. ***Retransformation without assumption of normality.*** If an assumption of normality is not credible, Duan's smearing estimator can be used.[57] This is

$$\hat{y}_i = \exp\left(\widehat{log\, y_i}\right) \times \frac{1}{n}\sum_i \exp\left(\log y_i - \widehat{\log y_i}\right)$$

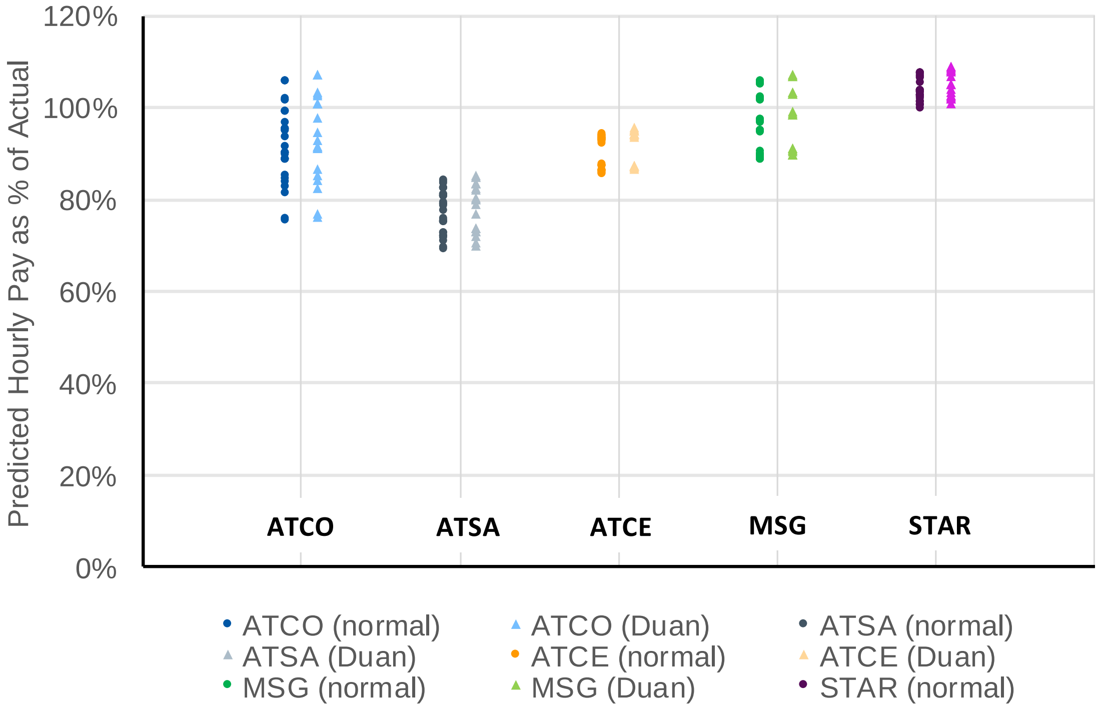
The results reported in Section 4.2 are based on Option 2, retransformation under the assumption of normality. However, as a robustness check we also examined Options 1 and 3. Option 1 predicted wages below the levels predicted by Option 2, as expected. Option 3 (Duan's smearing estimator) produced very similar results to Option 2, with the predicted wages marginally higher as seen in Figure B.15. This suggests that our results are not sensitive assuming that the residuals (or unexplained "error" terms) follow a normal distribution. That lack of sensitivity in turn suggests that the residuals of our wage equations are indeed approximately normal.

55 Wooldridge, J.M. (2009), *Introductory Econometrics: A Modern Approach (4th international student edition)*, p. 232

56 Cameron, A.C. and Trivedi, P.K. (2005), *Microeconometrics: Methods and Applications*, p546

57 Duan, N. (1983). Smearing estimate: a nonparametric retransformation method. *Journal of the American Statistical Association, 78(383)*, pp. 605-610.

Figure B.15: Options 2 and 3 for Transforming Predicted Wages in Logs to Predicted Wages in Levels Produced Similar Results



Source: NERA analysis of LFS and NERL data

Appendix C. Changes to Explanatory Variables Relative to Previous Report

C.1. Firm Size

In the previous report, this was a categorical variable. In this report, we change it to be a binary variable. Rather than estimating coefficients for the various categories of firm size, we focus on a single coefficient for whether the firm size exceeds 500 employees. Since NERL has over 500 employees, this is the main effect of interest in estimating market benchmark wages for NERL staff – there is no need, for benchmarking purposes, to estimate coefficients for other employer sizes. The benefit of having a binary firm size variable rather than a categorical firm size variable is that it allows us to impute missing data, rather than dropping observations (see the discussion of imputation in Section 3.1).

C.2. Union Membership

This variable was included in our previous report alongside the variable *tucov*, reflecting whether the pay received by an individual is subject to union negotiation. We omit it here for several reasons. First, it is statistically similar to *tucov*, and therefore the need for both variables is unclear. We chose to include *tucov* rather than union membership because there were fewer missing observations on *tucov* than on union membership, and because *tucov* is more directly linked to pay than union membership. Second, we did not have data from NERL on union membership for this iteration of the report. We do not anticipate that the revision to the treatment of unionisation in the wage equations will have had a material impact on the results.

C.3. Training in the Last Three Months

This variable was included in our previous report. For this report, we exclude the training variable as it is likely to be distorted by the impact of COVID-19. Employees on furlough are unlikely to have received training, and other employers may have cut back on training to reduce costs. Therefore, in our current dataset the connection between training, required skill level of a job, and market pay rate is likely to be weaker than it would be in the absence of COVID-19.

C.4. Region

In the previous benchmarking exercise, we estimated coefficients for 22 different regions, based on the values of the LFS variable *regwkr*. For this updated exercise, the region data from NERL did not correspond to the values of the LFS variable. We therefore used a less granular classification of regions in this report, to which we could map both the LFS data and the NERL data.

C.5. Dependent children

This variable was included in our previous benchmarking exercise, but we omit it here because we do not have data on dependent children for NERL staff.

Appendix D. Evaluation of Model Performance

We evaluate model performance through a number of simple checks. Our simple checks indicate that our models perform well, in the sense that they are statistically and economically sound and therefore present a reliable basis for calculating predicted benchmark wages of NERL staff. The performance of these models is similar to that of the models estimated in our 2018 benchmarking exercise.

D.1. Assessment of Estimated Coefficients

We conduct a "sense-check" of estimated coefficients, to ensure that they are in line with expected results based on the existing literature. Coefficients broadly in line with the expectations from the literature indicate a well-performing model, whereas coefficients out of line with the literature suggest some error in model specification or problem with the dataset.

Table D.1 reports the results of the sense-checking exercise. We consider only variables for which we have a clear expectation, guided by either consensus in the literature or more general awareness of economic conditions (in the case of the time variables).

Table D.1: Estimated Coefficients from our Wage Equation Models are in Line with Expectations

Variable	Expectation	Results Consistent with Expectation?
Controls relating to employee characteristics		
Indicators for highest qualification	Higher qualifications result in higher wages.	**Yes** – relative to baseline (degree or equivalent) all other qualifications had negative coefficients, indicating lower wages.
Age (proxy for work experience)	Positive coefficient, wages increase with age.	**Yes** – coefficient positive in all models.
Tenure at current employer	Positive coefficient, wages increase with tenure.	**Yes** – coefficient positive in all models.
Controls relating to job characteristics		
Indicator for region of work/residence	"London premium" – wages are higher in London.	**Yes** – relative to baseline (North) London workers have wages about 25 per cent higher on average. The coefficients for being located in Wales or Northern Ireland are negative, indicating lower wages, also in line with expectations.
Functional form adjustments		
Square of tenure	Negative coefficient, wage growth with age begins to slow with tenure (diminishing marginal returns)	**Yes** – coefficient negative in all models.
Square of age	Negative coefficient, wage growth with age begins to slow at higher ages (diminishing marginal returns)	**Yes** – coefficient negative in all models.
Time Effects		
Time dummies (indicator for quarter of observation)	Coefficients show negative impact of COVID on wages.	**Partial** – coefficients for 2020Q1 always lower than coefficients for 2019Q4 (and in some models negative), but pattern of coefficients for subsequent quarters less clear.
Time trend	Positive coefficient, wages grow over time (e.g. due to inflation)	**Yes** – coefficient positive in all models with a time trend.
Demographics		
Indicator for sex	Women receive lower wages.	**Yes** – coefficient on indicator for being male positive in all models, being male increases wages by around 10 per cent on average.
Indicators for ethnicity	Non-white people receive lower wages than white people.	**Yes** – relative to baseline (white) all other ethnicities had negative coefficients, indicating lower wages.

Source: NERA analysis of wage equation coefficients

The estimated coefficients from our wage equation models are consistent with expectations in all but one case. The coefficients that are not quite consistent with expectations are those for the coefficients on the time dummies. We had expected that the time dummies would show a strong negative impact from COVID-19; this is true in 2020Q1 but not in subsequent quarters. It is possible that the impact of COVID-19 on wages is masked by furlough payments. While recent waves of the LFS have collected data on furlough payments, these variables are not included in the publicly-available version of the LFS[58] and so we cannot test this explanation.

Our previous report does not include an explicit assessment of estimated coefficients, but it does report that some coefficients are in line with expectations from the literature (e.g. the coefficients on highest educational qualification is significant at the 1% level).[59]

D.2. Consistency of Results Across Reports

The fact that the results of the benchmarking exercise are consistent across the two reports provides supporting evidence that the models in both reports are well-performing models.

Overall, both modelling exercises find that NERL wages are broadly in line with market benchmarks. However, they are based on completely independent datasets, and there are some non-trivial differences in the selection of explanatory variables as described in Section 3.2.2 and Appendix C.

The fact that two sets of models, using different explanatory variables and different datasets, produce similar results indicates that the results and general modelling approach are robust to changes in model specification or datasets. Such robustness is an indicator of a good quality model.

D.3. Statistical Evaluation of Model Fit

As reported in Section 4.2, most models that we estimate have an adjusted-R2 in the region of 0.4. This means that approximately 40 per cent of the total variation in hourly pay in the LFS dataset is explained by the models.

We also note that the overall F-test for all models was rejected at the 1% significance level. This indicates that the variables in the model are jointly significant as determinants of the outcome variable, hourly pay.

These two findings indicate that our models have non-trivial explanatory power for wages in the economy as a whole, which is a positive indicator of model performance.

The adjusted-R2 for the models estimated in this report is typically lower than the adjusted-R2 of the models estimated in the previous report. However, as explained in Section 4.3, this does not mean that the models estimated in the current report are of poorer quality than the models estimated in the previous report.

D.4. Statistical Assessment of Residuals

It is often considered best practice in econometric analysis to report the results of a barrage of "standard" statistical tests on residuals. These tests include the Jarque-Bera test for residual

[58] Office for National Statistics (2021), Labour Force Survey User Guide Volume 3 – Details of LFS Variables – Version 1 – January to March 2021, p. 3

[59] NERA (21 March 2018), Staff Operating Expenditure for Air Traffic Control: Prepared for NERL, p. 30

normality; the White test for homoskedasticity; and the RESET test for functional form specification.

However, when working with large datasets such as this one, these standard tests are in fact not particularly useful indicators of model performance. As a general rule, there will always be some degree of imperfection in any model, and so the true value of the test statistics for these standard statistical tests will not be zero. This creates a problem when working with large sample sizes. With large sample sizes, the confidence interval around any test statistic becomes very narrow, such that a test statistic may be statistically significant but meaningfully negligible.

The statistical literature reports that statistical tests are likely to report significance for even negligible effects with sample sizes on the order of tens of thousands.[60] Since our current dataset includes over 100,000 observations, it is almost inevitable that the standard statistical tests on residuals would report the presence of imperfections in our models, even if those imperfections were negligible. As the standard tests provide no guidance on the magnitude of the imperfections, they are not particularly useful indicators of model performance when working with large datasets.

We therefore do not conduct a formal statistical assessment of the residuals as part of our evaluation of model fit, instead relying on the more heuristic sense-checks already discussed above.

We can use other evidence to address some of the concerns that the standard statistical tests are intended to evaluate. In particular, we can address the concern of residual non-normality from the results of the application of Duan's smearing estimator to transform predicted wages in logs to predicted wages in levels, as reported in Appendix B.2. The Duan's smearing estimator is explicitly designed to correct for residual non-normality. In the presence of residual non-normality, it would be expected to yield quite different results than the standard transformation (option 2 in Appendix B.2). Since both transformations yield very similar results from our wage equations, this suggests that the residuals of our wage equations are indeed approximately normal.

[60] Sullivan, G. and Feinn, R. (2012) Using Effect Size – or Why the P Value is Not Enough. *Journal of Graduate Medical Education 4(3)*, pp. 279-282.

Appendix E. Impact of the Construction of the LFS Variable *hourpay* on Benchmark Pay

We estimate our wage equations using the LFS variable *hourpay* as the outcome. This means that predicted wages from these equations are really predicted values of the variable *hourpay*, for a given set of values for the explanatory variables.

By comparing the predicted wages from our wage equations with NERL total hourly pay, we implicitly assume that the LFS variable *hourpay* is equivalent in construction to NERL total hourly pay. In fact, this assumption is not correct. There are differences between the construction of the LFS variable *hourpay* and NERL total hourly pay, such that *hourpay* is likely to be systematically lower than NERL total hourly pay.

To calculate NERL total hourly pay for a staff group, we use the average total *annual* pay of NERL staff in that staff group. Total annual pay includes annual bonuses and other additions to basic pay. We convert annual pay to hourly pay by dividing by 52 (weeks in a year), and then dividing by the average weekly hours worked by NERL staff within the staff group.

To calculate the LFS variable *hourpay*, the ONS uses total pay *in the last pay period*. The last pay period varies by survey respondent. Some LFS participants report their total pay in the last year, in which case total pay includes all annual bonuses and other additions to basic pay. For those survey participants, the LFS variable *hourpay* is comparable to our calculation of NERL actual hourly pay.

Other LFS participants report their total pay over another period. The most common alternative period used is the previous month. In such cases, the total pay reported will include some additions to basic pay but is unlikely to include annual bonuses (as annual bonuses are typically paid in a single annual payment, rather than in additions to the monthly paycheck). This means that *hourpay* for these LFS participants is not comparable to our calculation of NERL total hourly pay and in fact is systematically *lower* than would be implied by the calculation of NERL total hourly pay, because of annual bonuses.

Our analysis of the LFS data indicates that this problem is pervasive. Table E.1 shows that approximately 1/3 of LFS survey respondents in NERL staff group comparator SOCs provide gross pay for the calendar month rather than the year. This means that *hourpay* is likely to be systematically below the equivalent of NERL total hourly pay for about 1/3 of the sample used to benchmark NERL total pay.

Table E.1: Pay Period Reported by LFS Respondents

Gross Pay Period	Non-PS Group Comparator SOCs	PS Group Comparators						Full LFS Sample
		ATCO	ATSA	ATCE	MSG	STAR	PCG	
Year	49%	50%	51%	69%	60%	70%	64%	51%
Month	38%	48%	42%	29%	35%	28%	34%	38%
Other	13%	2%	7%	3%	5%	1%	2%	11%

Source: NERA analysis of LFS data

To understand the impact of the possible underreporting of variable pay on our analysis, we rely on another variable in the LFS dataset. The variable *ernfilt* records the LFS participant's response to the question "did your last (gross) pay contain any additions to basic pay?".[61] 14.30 per cent of our total LFS sample of 142,801 observations answered "yes" to this question, 84.92 per cent answered "no", and 0.77 per cent responded "don't know".

To test the impact of underreporting of pay on our results, we restricted the dataset to the 14.30 per cent of respondents who answered "yes" to this question. The value of *hourpay* for these respondents includes additions to basic pay, so is more likely to include annual bonuses and therefore be comparable to NERL total hourly pay.

We repeat our entire analysis using just this dataset. That is, we re-estimate the wage equations using this dataset alone, and re-calculate the NERL predicted wages using the resulting wage equations and the data from NERL on explanatory variables (e.g. tenure).

The results of this exercise are summarised in Table E.2.

- For ATCOs, the exercise slightly narrows the range of model predicted pay.
- For ATSAs, the exercise shifts the range of model predicted pay slightly upwards.
- For ATCEs, the exercise shifts the range of model predicted pay upwards. As explained in Section 4.4, NERL actual pay for ATCEs now lies **within** the range of model predictions.
- For MSGs, the exercise shifts the range of model predicted pay slightly upwards.
- For STARs, the exercise shifts the range of model predicted pay upwards. NERL actual pay for STARs now lies **above** the range of model predictions.

Table E.2: Model Predicted and Actual Hourly Pay for NERL Staff Groups

	ATCO	ATSA	ATCE	MSG	STAR
NERL total hourly pay	49.04	29.03	33.97	24.91	27.14
Model predicted hourly pay (full sample)					
Minimum	36.86	20.03	29.03	22.02	27.02
Maximum	51.81	24.36	31.99	26.33	29.16
Model predicted hourly pay (restricted sample, reporting that pay includes additions to basic pay)					
Minimum	37.47	20.72	30.89	22.84	28.40
Maximum	50.52	24.47	33.70	27.98	31.25

Source: NERA analysis of LFS and NERL data

One limitation of this analysis is the relatively small size of the restricted sample, at only 14.30 per cent of our total LFS sample. This sample may create problems in estimating some coefficients, in particular the time-SOC interaction coefficients which are estimated from the observations on SOC comparators in each quarter.

61 It is clear from the phrasing and order of the LFS questionnaire that the "last (gross) pay" referred to here is the same payment used to calculate *hourpay*.

A second limitation of this analysis is that it may overstate benchmark pay, insofar as it excludes any comparators who do not *receive* additions to basic pay as well as those comparators who do not *report* additions to basic pay.

As a second test of the impact of the construction of *hourpay* on our results, and to address our concern about the small sample size in our first test, we conduct a further benchmarking exercise of NERL actual *basic* pay. Basic pay is less than total pay, because it excludes bonuses and other additional payments. We define NERL actual basic pay using annual basic pay, which we divide by 52 (the number of weeks in a year) and then again by the actual hours worked.

We restrict the LFS sample to only the 84.92 per cent of LFs participants who respond “no” to the *ernfilt* variable, i.e. only those who are certain that the pay they report does not include additions to basic pay. The *hourpay* variable in this restricted sample is comparable to NERL actual basic pay.

We repeat our entire analysis using just this dataset. That is, we re-estimate the wage equations using this dataset alone, and re-calculate the NERL predicted wages using the resulting wage equations and the data from NERL on explanatory variables (e.g. tenure). We then compare the results to NERL *basic* hourly pay.

Table E.3 summarises the results of this exercise. The range of model predicted pay is similar to, but slightly below, the range reported for the analysis based on the full sample. This is as expected, given that Table E.3 is based on 84.92 per cent of the full sample. NERL basic hourly pay is below NERL total hourly pay.

We conclude that NERL basic hourly pay is **within** the range of model predictions for ATCOs, ATCEs, and MSGs; slightly **below** the range of model predictions for STARs; and **above** the range of model predictions for ATSAs.

Table E.3: Model Predicted and Actual Basic Hourly Pay for NERL Staff Groups

	ATCO	ATSA	ATCE	MSG	STAR
NERL basic hourly pay	44.42	26.11	32.13	24.79	25.80
Minimum predicted hourly pay	35.80	19.93	28.82	21.87	26.81
Maximum predicted hourly pay	52.62	24.44	31.87	26.28	28.95

Source: NERA analysis of LFS and NERL data

The combined implication of these two exercises is that the initial finding reported in Section 4.2, that ATCE total hourly pay exceeds the range of model predictions, is likely incorrect. That finding was based on an LFS dataset in which a large proportion (84.92 per cent) of the sample had pay data that was not comparable to, and systematically lower than, NERL total pay because it did not include additions to basic pay. When we restrict the LFS dataset to consider only those who report basic pay, or only those who report total pay, it is clear that the relevant NERL actual pay for ATCEs (basic and hourly, respectively) is within the range of model predictions.

Appendix F. Model Specifications

Table F.1: We Estimated 37 Model Specifications

Model	Demographics	Union agreements	Squares*	Full time - Basic Usual Hours	Full time - Basic Usual Hours Squared	Time Dummies	Time Trend	SIC - Time Dummies	SIC - Time Trend	SOC - Time Dummies	SOC - Time Trend
1	Y	Y	Y	Y	Y	Y	N	N	N	N	N
2	N	Y	Y	Y	Y	Y	N	N	N	N	N
3	N	Y	Y	N	N	Y	N	N	N	N	N
4	N	Y	N	Y	N	Y	N	N	N	N	N
5	N	Y	N	Y	N	Y	N	N	N	N	N
6	Y	Y	Y	Y	Y	Y	N	Y	N	N	N
7	N	Y	Y	Y	Y	Y	N	Y	N	N	N
8	N	Y	Y	N	N	Y	N	Y	N	N	N
9	N	Y	N	Y	N	Y	N	Y	N	N	N
10	N	Y	N	Y	N	Y	N	Y	N	N	N
11	Y	Y	Y	Y	Y	Y	N	N	Y	N	N
12	N	Y	Y	Y	Y	Y	N	N	Y	N	N
13	N	Y	Y	N	N	Y	N	N	Y	N	N
14	N	Y	N	Y	N	Y	N	N	Y	N	N
15	N	Y	N	Y	N	Y	N	N	Y	N	N
16	Y	Y	Y	Y	Y	Y	N	N	N	Y	N
17	N	Y	Y	Y	Y	Y	N	N	N	Y	N
18	N	Y	Y	N	N	Y	N	N	N	Y	N
19	N	Y	N	Y	N	Y	N	N	N	Y	N
20	N	Y	N	Y	N	Y	N	N	N	Y	N
21	Y	Y	Y	Y	Y	Y	N	N	N	N	Y
22	N	Y	Y	Y	Y	Y	N	N	N	N	Y
23	N	Y	Y	N	N	Y	N	N	N	N	Y
24	N	Y	N	Y	N	Y	N	N	N	N	Y
25	N	Y	N	Y	N	Y	N	N	N	N	Y
26	Y	N	N	N	N	N	Y	N	N	N	N
27	Y	N	Y	N	N	N	Y	N	N	N	N
28	Y	N	N	N	N	Y	N	N	N	N	N
29	Y	N	Y	N	N	Y	N	N	N	N	N
30	Y	Y	N	N	N	N	Y	N	N	N	N
31	Y	Y	Y	N	N	N	Y	N	N	N	N
32	Y	Y	N	N	N	Y	N	N	N	N	N
33	Y	Y	Y	N	N	Y	N	N	N	N	N
34	Y	N	N	N	N	N	Y	N	N	Y	N
35	Y	N	Y	N	N	N	Y	N	N	Y	N
36	Y	Y	N	N	N	N	Y	N	N	Y	N
37	Y	Y	Y	N	N	N	Y	N	N	Y	N

Qualifications, assumptions and limiting conditions

This report is for the exclusive use of the NERA Economic Consulting client named herein. This report is not intended for general circulation or publication, nor is it to be reproduced, quoted or distributed for any purpose without the prior written permission of NERA Economic Consulting. There are no third party beneficiaries with respect to this report, and NERA Economic Consulting does not accept any liability to any third party.

Information furnished by others, upon which all or portions of this report are based, is believed to be reliable but has not been independently verified, unless otherwise expressly indicated. Public information and industry and statistical data are from sources we deem to be reliable; however, we make no representation as to the accuracy or completeness of such information. The findings contained in this report may contain predictions based on current data and historical trends. Any such predictions are subject to inherent risks and uncertainties. NERA Economic Consulting accepts no responsibility for actual results or future events.

The opinions expressed in this report are valid only for the purpose stated herein and as of the date of this report. No obligation is assumed to revise this report to reflect changes, events or conditions, which occur subsequent to the date hereof.

All decisions in connection with the implementation or use of advice or recommendations contained in this report are the sole responsibility of the client. This report does not represent investment advice nor does it provide an opinion regarding the fairness of any transaction to any and all parties.

NERA Economic Consulting
Marble Arch House
66 Seymour Street
London, UK W1H 5BT
+44 207 659 8500
www.nera.com