Skills, Sorting and the Early-Career Graduate Premium: Evidence from the UK

Max Schroeder*

August 25, 2026

Abstract

Differences in the distribution of graduates and non-graduates across occupations explain a large part of the graduate wage premium. But does graduates' occupational advantage reflect stronger skills or differences in assignment conditional on skill? I estimate a model of latent skill, occupation-specific returns to skill, and occupational choice to separate skill-based sorting from residual assignment conditional on skill. Differences in labour-market skill explain most of graduates' concentration in Professional occupations, although residual assignment also shifts in their favour. Counterfactuals show that occupational sorting is central to the level of the graduate premium but compresses rather than exacerbates it: sorting benefits non-graduates more because their labour-market skill distribution is wider. I then use the model to decompose the decline in the early-career graduate premium. The estimates suggest that this decline is driven primarily by a narrowing skill gap between graduates and non-graduates, while changing occupational wage schedules partly offset this effect.

Keywords: graduate premium, skill, occupational sorting, higher education
JEL Classification: I23, I24, I26, J24

*Durham University Business School, Email: max.j.schroeder@durham.ac.uk

I would like to thank Konstantinos Angelopoulos, Richard Foltyn, Spyridon Lazarakis, Margaret Leighton, and Jim Malley for comments on earlier versions of this paper. I further thank the seminar participants at Durham and Lancaster University, the EEA-ESEM Congress (2024), the European Society of Labour Economists Conference (2024), and the 2nd Toulouse Economics of Education Workshop (2025) for their insights. All remaining errors are my own. This research was enabled by the Hamilton HPC Service of Durham University.

1 Introduction

A substantial part of the graduate wage premium is associated with occupation. Young graduates are more likely than non-graduates to work in Professional occupations, which pay more on average. It is less clear how much of this pattern reflects graduates' privileged access to these jobs and how much reflects differences in skill and the resulting sorting of workers across occupations. This distinction has become particularly relevant as young graduates face a more difficult entry-level labour market and struggle to land *graduate jobs*.[1]

In this paper, I address this question by developing a model of occupational sorting for early-career graduates and non-graduates. Workers draw skill from a distribution that varies by degree status, and choose among occupations that offer differential returns to this skill. The graduate premium therefore reflects differences in the distribution of skill among graduates and non-graduates, as well as differences in the return to that skill, driven by differential sorting of workers across occupations.

I estimate the model using the five-quarter longitudinal Quarterly Labour Force Survey (L-QLFS) for the UK over 2001–18. The estimated model is able to fit the main observed patterns of occupational sorting and wage dynamics over the sample period and generates sensible skill distributions and return schedules. Importantly, the estimates suggest that the early-career graduate premium is primarily a function of differential skill endowments. Graduates are on average higher in the model-implied skill distribution, but their mean advantage falls between 2001–06 and 2013–18 by around 0.17 initial standard deviations.

The estimated returns to skill also display a clear occupational hierarchy: they are highest in Professional occupations and lowest in Service occupations. This ordering is stable, but the return to skill in Routine occupations rises while the Professional return falls, narrowing the gap between the two. The graduate premium is therefore shaped not only by the size of the skill gap, but also by how the market rewards that gap across the occupations into which workers endogenously select. The model also suggests that occupational sorting is strongly stratified by skill within both degree groups. Graduates are exposed to higher skill returns on average, but the correlation between skill and the return in the chosen occupation is consistently stronger among non-graduates. At the same time, skill differences do not account for all sorting patterns. For Professional occupations, the choice-factor component accounts for 21.3% of the graduate–non-graduate employment gap in 2001–06 and 44.0% in 2013–18.

I then use model-based decompositions to relate these objects to the observed decline in the graduate premium. From 2011 onwards, the narrowing skill gap is the main driver of the decline in the graduate premium. Changes in wage schedules partly offset this trend,

[1] Recent examples include discussion of the contraction in graduate opportunities and declining public confidence in the financial value of a degree.

while other factors, such as demographic composition or group specific factors affecting occupational choice play a minor role. This suggests that the decline in the early-career graduate premium is largely a matter of the narrowing skill gap between graduates and non-graduates—a finding that resonates with anecdotal accounts, as well as empirical evidence, that rapid higher education expansion brought students from further down the skill distribution into university and weakened outcomes at the lower end of the graduate distribution (Walker & Zhu, 2008; Lindley & McIntosh, 2015; Carneiro & Lee, 2011).

1.1 Related Literature

This paper adds to three connected literatures. The first studies the UK graduate premium during and after the expansion of higher education. Earlier work found considerable resilience in the average premium, alongside weaker outcomes in parts of the graduate distribution and among newer cohorts (Walker & Zhu, 2008; O'Leary & Sloane, 2011; Blundell et al., 2022).[2] More recent evidence further documents a decline in the graduate premium for early-career workers (Boero et al., 2025). I complement this work by using a model of occupational sorting to decompose both the level and the change in the early-career graduate premium into differences in estimated latent skill, occupation-specific wage schedules and assignment conditional on skill. A model-based decomposition attributes most of the decline to a narrowing graduate–non-graduate skill gap.

The second literature studies heterogeneity in graduate outcomes and the measurement of graduate jobs, overqualification and mismatch. Gottschalk & Hansen (2003) show that, when preferences are heterogeneous, equally productive college graduates may work in both college and non-college jobs. More generally, the graduate share of an occupation is not by itself a clean measure of its skill requirement (Green & Henseke, 2016). Graduate outcomes vary substantially across fields (Altonji et al., 2016), while much of the growth in UK graduate wage inequality has occurred within fields of study (Lindley & McIntosh, 2015). Occupation-based, qualification-based and task-based measures consequently do not deliver the same trend in graduate underemployment or mismatch (Green & Zhu, 2010; Green & Henseke, 2016). O'Leary & Sloane (2016) apply the Gottschalk–Hansen classification approach to young British graduates. I instead model workers' choices among occupations with skill-dependent wage schedules, while allowing assignment conditional on skill to differ by degree status. Recent task-based evidence finds that graduate skill requirements continued to expand through 2017, although their growth slowed after 2006 and estimated assignment and task prices remained broadly stable (Henseke et al., 2025). In this paper I provide suggestive evidence that returns to skill correlate with complex task content, providing complementary evidence to earlier work on the importance of

[2]The corrected descriptive series in Blundell et al. (2024) shows a small decline near the end of the original sample, while leaving the paper's regression results and main conclusions unchanged.

task content for defining graduate jobs. The model further adds a worker-side dimension by showing how occupation-specific wage gradients interact with the graduate and non-graduate distributions of latent skill and with occupational assignment to shape the graduate premium.

The third literature studies how worker skills and occupational task content interact with technology and task prices to shape occupational assignment and wages (Autor, Levy, et al., 2003; Acemoglu & Autor, 2011; Autor & Handel, 2013). The closest methodological analogue is Diegert (2024), who uses short panels of wages and occupational choices from the Survey of Income and Program Participation (SIPP) to identify skill, occupation-specific skill prices and sorting patterns under endogenous occupational selection. I draw on his approach to study the early-career graduate premium in the UK, focussing on the impact of skill heterogeneity across graduate and non-graduate cohorts. Lise & Postel-Vinay (2020) jointly model multidimensional skills, occupational sorting and human-capital accumulation, while Gathmann & Schönberg (2010) show that task-specific human capital transfers more readily between occupations with similar task requirements. I complement this work by quantifying how group–cohort differences in a latent skill distribution interact with occupational wage schedules and assignment conditional on skill to generate the graduate premium.

The rest of the paper is structured as follows. Section 2 presents the data and motivating facts. Section 3 describes the model. Section 4 discusses identification, estimation and model fit. Section 5 presents the estimated skill distributions and occupational wage schedules. Section 6 examines occupational sorting, first through wage incentives and assortative matching and then through the Professional-employment gap. Section 7 studies the decline of the graduate wage premium using counterfactuals and decompositions. Section 8 concludes.

2 Motivating empirics

This section briefly describes the data sample and then documents two connected facts about the early-career graduate premium. First, occupation is a strong descriptive mediator of the level of the graduate premium. Second, the decline in the early-career graduate premium largely survives conditioning on occupation. Occupational allocation is therefore central to the wage gap, but a change in observed occupational shares is not sufficient to account for its decline. This evidence then motivates the economic model described in Section 3.

2.1 Data and sample

My main data source is the five-quarter longitudinal Quarterly Labour Force Survey (L-QLFS). The L-QLFS is a longitudinal survey that follows a rotating panel of households for five quarters, with 20% of the sample being replaced at every wave.[3] Importantly, the L-QLFS combines a rotating-panel structure with information on both wages and occupations, allowing me to study long-run trends using repeated observations on the same workers. I restrict attention to employees aged 25–30 at their first interview whose panel begins between 2001 and 2018.[4] Appendix G.6 reports re-estimates for two alternative wave-1 age windows. I further restrict the sample to individuals who are observed in all five waves, remain in employment, have a valid occupation in every wave and report valid wages at waves 1 and 5.[5] The final sample contains 13,412 people and 67,060 person-wave observations: 5,163 graduates and 8,249 non-graduates.[6]

2.2 Two facts about the early-career graduate premium

I define the reduced-form graduate premium as the coefficient on graduate status in a log-hourly-wage regression. This is an observed conditional wage difference rather than a causal return to university. Figure 1 reports centred three-year moving averages of the annual coefficient under four increasingly saturated control sets. The raw series contains year effects and graduate-by-year interactions only. The common-control series adds age, age squared, region of residence, part-time status and sex. The final two series add either the three broad occupation groups or detailed SOC-2-digit occupation controls.[7] Table 1 reports the corresponding six-year block regressions, while Appendix Table A1 reports the occupation-based decomposition.

[3]This is analogous to the US Current Population Survey (CPS) in that it follows a rotating panel of households, although the L-QLFS interviews households quarterly for five consecutive waves.

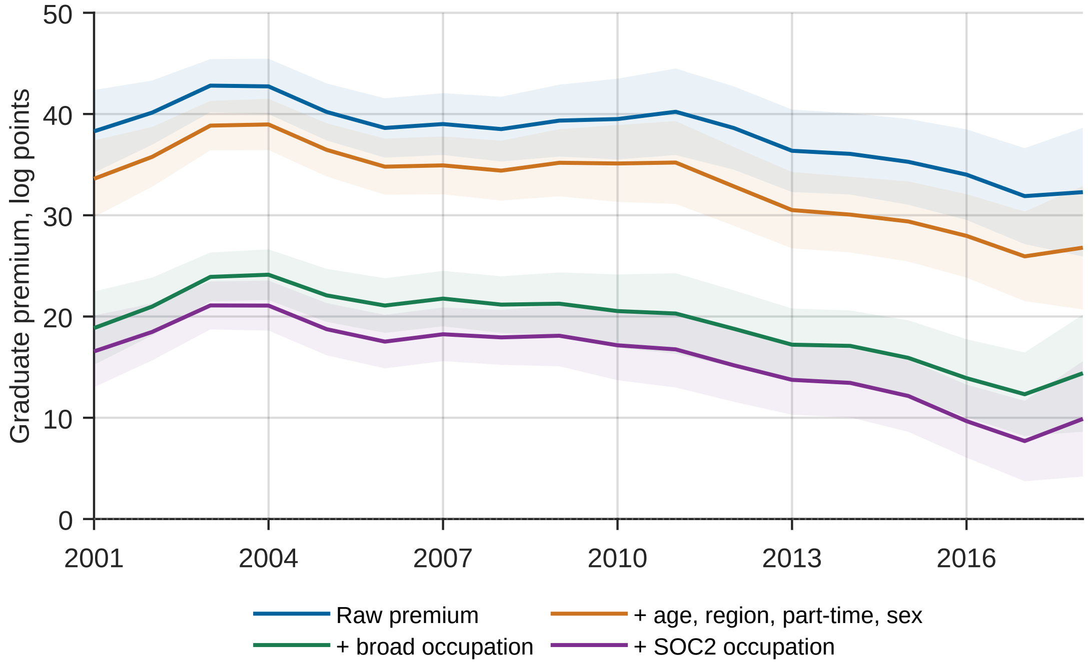
[4]The source sample extends in principle to calendar year 2019. Because the data window ends effectively in 2020, workers observed in the final year contribute only one observed wage measurement within the window rather than the two available elsewhere. The corresponding annual estimates are noticeably noisier, so I exclude the final year from the reported results and report the analysis through 2018.

[5]Wages are observed only in waves 1 and 5. I retain workers with valid wages at both endpoints and generate wage measurements for waves 2–4 using the baseline imputation procedure. Observed wages are deflated by calendar-year CPI (2014=100) and winsorised at the 0.1st and 99.9th percentiles. The narrowing skill gap, the ranking of occupational skill returns and the negative skill-distribution contribution to the premium decline are robust to alternative imputation methods and to using only observed wages; see Appendix D.

[6]For more details on data construction, see Appendix B.

[7]The broad classification aggregates SOC-1-digit major groups into Professional (groups 1–3), Routine (groups 4–5 and 8–9), and Service (groups 6–7) occupations. I crosswalk SOC2000 and SOC2010 occupation codes onto this common grouping; the detailed controls use SOC-2-digit categories under the applicable source classification.

Figure 1: Observed graduate wage premium under alternative control sets



Note: Employed workers aged 25–30 at their first interview. The figure reports centred three-year moving averages of 100 times the annual graduate coefficient from regressions of observed real log hourly wages on year effects, graduate-by-year interactions and the indicated controls. The common controls are age, age squared, residence region, part-time status and sex. Occupation specifications add either the three broad groups or SOC-2-digit controls. Shaded regions are 95% confidence intervals based on person-clustered standard errors.
Source: Quarterly Labour Force Survey, author's own calculations.

Table 1: Graduate wage premia under alternative control sets

	2001–06	2007–12	2013–18
Raw premium	40.6***	39.4***	34.0***
	(1.085)	(1.403)	(1.669)
Common controls	36.5***	34.8***	28.1***
	(1.048)	(1.434)	(1.618)
Common controls + broad occupation	21.7***	20.3***	15.2***
	(1.095)	(1.609)	(1.590)
Common controls + SOC-2-digit occupation	18.2***	16.2***	11.7***
	(1.121)	(1.500)	(1.606)
Observations	13,632	8,058	4,871

Note: Entries are log points; standard errors are in parentheses. The table estimates $\log w_{it} = \alpha_b + \delta_t + \beta_b G_i + X_{it}'\gamma_b + O_{it}'\pi_b + \varepsilon_{it}$ separately within each six-year period b. Here G_i denotes graduate status, δ_t denotes the year fixed effects. The vector X_{it} is omitted from the raw specification and otherwise contains age, age squared, residence region, part-time status and sex. The vector O_{it} is omitted from the first two specifications and contains either broad-occupation or SOC-2-digit indicators in the final two. Regressions use survey weights and standard errors are clustered by individual. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.
Source: L-QLFS and author's calculations.

Fact 1: occupation is a strong mediator of the graduate premium. Figure 1 shows a substantial raw early-career graduate premium throughout the sample. In 2001–06 it averages around 40 log points. Adding the common demographic controls reduces it only modestly, to 36 log points, whereas conditioning on the three broad occupation groups lowers it to around 22 log points. Detailed SOC-2-digit occupation controls reduce it only slightly further, to 19 log points. The main attenuation therefore comes from broad differences in the types of jobs held by graduates and non-graduates, rather than from fine distinctions between narrowly defined occupations. Hence, a substantial part of graduates' wage advantage is associated with the occupations they enter.[8]

Fact 2: the fall in the premium survives occupation controls. Across all four specifications, the early-career graduate premium fluctuates around a broadly stable level during the first part of the sample, before declining after 2010. After controlling for a set of demographic characteristics, the premium falls from an average of around 36 log points in 2001–06 to 28 log points in 2013–18, in line with recent evidence on early-career graduates (Boero et al., 2025). Importantly, the decline remains visible after conditioning on occupation: the broad-occupation premium falls from 22 to 15 log points, while the SOC-2-digit premium falls from 19 to 11 log points. The decline therefore appears to operate primarily through wage changes within occupations rather than solely through changes in allocation.[9]

Together, these facts place occupation at the centre of the early-career graduate premium, but leave open why graduates are more likely to enter higher-paying occupations. One possibility is that their occupational advantage reflects greater labour-market-relevant skill, whether acquired through university or already present among those who enter higher education. Another is an additional degree-based access margin, such as nominal degree requirements, that gives graduates preferential access conditional on the same ability. At the same time, conditioning on observed occupational allocation does not remove the decline in the premium. This suggests that while occupational assignment plays a central role in the level of the graduate premium, it is not sufficient to account for the decline documented for recent years.

In the following section, I develop a model that can rationalise these patterns and help us understand the mechanisms that drive the early-career graduate premium and its decline.

[8] Table 1 reports the corresponding six-year block estimates. In Panel A of Appendix Table A1, the between-occupation component contributes between 13 and 16 log points to the adjusted premium across the three blocks, reinforcing the importance of occupational sorting for its level.

[9] Appendix Table A1 complements these controlled series by decomposing the change between the first and final six-year blocks. Panel B suggests that around two-thirds of the 8.4-log-point decline comes from the within-occupation component, while the remaining third comes from the between-occupation component.

3 Model

The evidence provided in the previous section shows that the early-career graduate premium depends both on occupational allocation and on wage differences that remain within observed occupations. In this section I develop an economic model that can account for both of these patterns. The central modelling object is a measure of labour-market skill, defined as a latent skill index, which is rewarded differently across a number of discrete occupations. Graduates and non-graduates share the same qualitative skill, but draw from different distributions. Workers compare occupation-specific wage offers and non-pecuniary choice factors, and sort across occupations according to their relative attractiveness. The rest of this section presents the model in detail.

3.1 Workers and skills

The model's focus is on early-career workers aged 25–30, who are either university graduates or non-graduates, indicated by $g_i \in \{G, N\}$. Workers are further assigned to a cohort c_i, which depends on the year that they turned 25.[10] Workers are observed for five consecutive quarters, where $t \in \{1, \ldots, 5\}$ denotes the interview wave, and $\tau(t)$ denotes calendar year. Each worker is characterised by a scalar latent skill index s_i that is fixed over the five-quarter panel and further observable characteristics X_{it}.

The model conceives of skill as a latent univariate index that generates persistent wage variation across otherwise observationally equivalent workers.[11] It therefore reflects general labour-market ability, such as the ability to perform challenging and complex tasks.[12] In this sense, s_i is a persistent person effect that is partly transferable across occupations, providing a common measure of skill.

I model the distribution of the skill index flexibly, allowing for variation within cohorts, across graduates and non-graduates, and across cohorts over time. Specifically, workers draw skill from a distribution that varies by graduate status and cohort,

$$s_i \sim F_{g_i c_i}(s), \qquad s_i \in [0, 1]. \tag{1}$$

[10]The four cohorts cover 1996–02, 2003–07, 2008–11 and 2012–18. The common cut-offs are weighted quartiles of the year workers turned 25. Their unequal lengths keep the weighted cohort shares as close to one quarter as possible for both graduates and non-graduates despite the discreteness of calendar years.

[11]A broader empirical literature places persistent worker heterogeneity at the centre of wage dispersion. Using matched worker–firm data, Abowd et al. (1999) find that person effects, especially those unrelated to observed education, are more important for wage variation than firm effects. In a life-cycle model, Huggett et al. (2011) find that differences in initial conditions at age 23 account for more variation in lifetime earnings than shocks received over the working life. Heckman et al. (2006) similarly show that low-dimensional latent cognitive and non-cognitive skills jointly shape wages, employment and occupational choice.

[12]Appendix C provides some suggestive evidence that the return to latent skill is increasing in complex task content.

3.2 Occupations and wages

Let $o \in \mathcal{O} = \{P, R, S\}$ denote Professional, Routine and Service occupations. Occupations are characterised by annual wage schedules that map latent skill and observable characteristics into potential hourly wages.[13] Suppressing the wave index for simplicity, the potential expected log wage of worker i in occupation o and calendar year τ is

$$W_{io\tau} = \eta_{o\tau} + \lambda_{o\tau} s_i + X_{it}^{w\prime} \beta_o^w. \tag{2}$$

The intercept $\eta_{o\tau}$ shifts the occupation-year wage schedule, while $\lambda_{o\tau}$ determines how strongly the occupation rewards the latent index.[14] The vector X_{it}^w contains observed covariates. Potential experience is measured at each interview as age minus 25 and enters non-parametrically. The remaining covariates are sex, part-time status and macro region of residence. All control coefficients are occupation-specific, time invariant and common across graduate-status groups. The return $\lambda_{o\tau}$ is the estimated wage slope over the model-implied latent skill index, and therefore the key parameter governing skill-based sorting in the model.

3.3 Occupational choice

In addition to the pecuniary rewards, workers have systematic preferences over non-pecuniary aspects of occupations. These preferences are captured by

$$R_{io\tau}^g = \omega_{o\tau}^g + X_{it}^{c\prime} \gamma_o^c, \qquad g \in \{G, N\}. \tag{3}$$

Here $\omega_{o\tau}^g$ is the occupation–year–group non-pecuniary choice factor and X_{it}^c contains the same set of observed covariates as in equation (2). Professional occupations are the reference category, so that $\omega_{P\tau}^g = 0$ and $\gamma_P^c = 0$. The non-pecuniary term is deliberately broad. It may include non-wage job characteristics, systematic preferences, search frictions, access constraints, mismatch and systematic offer differences across graduates and non-graduates.[15]

The total value of occupation o for worker i at wave t is then the sum of the expected

[13] Appendix G.5 shows that the central patterns remain when wage and choice parameters are constant within three-year blocks.

[14] The baseline uses a single latent skill with common returns to skill within each occupational group. Appendix C provides some evidence that both groups perform similar tasks within occupations. Appendix F shows that the central results are robust to allowing skill returns to differ by graduate status and to introducing a restricted second skill dimension.

[15] Specifically, $\omega_{o\tau}^g$ will capture any difference in occupational assignment conditional on skill and observable covariates.

log wage, the non-pecuniary reward and a stochastic choice shock ε_{iot}:

$$V^g_{io\tau t} = W_{io\tau} + R^g_{io\tau} + \frac{\varepsilon_{iot}}{\rho_g}, \qquad g \in \{G, N\}. \tag{4}$$

The scale parameter $\rho_g > 0$ governs how sharply group g responds to deterministic value and is allowed to differ by graduate status.[16] I normalise the coefficient on expected log wages to one, so all deterministic choice components are measured in log-wage units.

Following the standard approach in the discrete choice literature (K. E. Train, 2009), I assume that ε_{iot} is independently and identically distributed following a Type I Extreme Value. Hence, the conditional occupation choice probability is

$$\Pr(o_{it} = o \mid s_i, g_i = g, X_{it}) = \frac{e^{\rho_g [W_{io\tau(t)} + R^g_{io\tau(t)}]}}{\sum_{j \in \mathcal{O}} e^{\rho_g [W_{ij\tau(t)} + R^g_{ij\tau(t)}]}}, \qquad g \in \{G, N\}. \tag{5}$$

As in a Roy model, workers sort across occupations according to the interaction between their characteristics and the occupation-specific schedules (Roy, 1951). Conditional on s_i, observed characteristics X_{it} and annual wage schedules, the worker evaluates the three occupations in every quarter and chooses the occupation with the highest value. Occupational choices and wage outcomes are hence indicative of latent skill. Figure 2 below summarises the choice sequence.

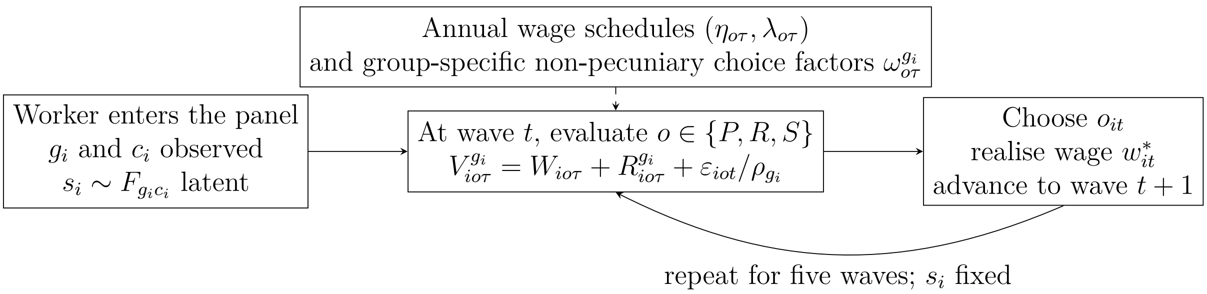


Figure 2: Model timeline

With the economic structure in place, I next turn to how the joint wage and occupational-choice sequences identify the model parameters and how I estimate them.

4 Identification, estimation and model fit

To estimate the model parameters, I exploit the five-wave panel structure of the L-QLFS. Intuitively, the model specified in the last section has a factor structure similar to that in Heckman et al. (2006). Repeated manifestations of a common persistent wage component pin down the latent skill index, while wage variation across individuals within the same

[16] The central results are robust to imposing a common utility scale; see Appendix G.3.

occupation helps identify the occupation-year wage schedules. Conditional on these wage-side objects, the multinomial-logit structure maps remaining systematic differences in occupation probabilities into non-pecuniary choice factors.

4.1 Likelihood and identifying variation

I estimate the completed-data model by maximum likelihood. Each worker panel consists of a five-wave sequence of occupational choices $o_i^* = \{o_{i1}^*, \ldots, o_{i5}^*\}$ and wages $w_i^* = \{w_{i1}^0, \widehat{w}_{i2}, \widehat{w}_{i3}, \widehat{w}_{i4}, w_{i5}^0\}$, where the hat notation indicates imputed values of wages at waves 2–4. Appendix D describes the imputation procedure and the alternative wage-measurement specifications. I discuss the role of the completed wage measurements in the identification argument below.

Skill is fixed over this short panel, while wage schedules and non-pecuniary choice factors vary by calendar year.[17]

I treat each observed or completed wage as a noisy measurement of the worker's potential wage in the selected occupation:

$$w_{it}^* = W_{i,o_{it}^*,\tau(t)} + \nu_{it}. \tag{6}$$

For the observed wages, ν_{it} captures transitory wage variation and ordinary survey measurement error. For the imputed middle-wave wages, it also captures the discrepancy introduced by the imputation procedure.

I assume that, conditional on latent skill, observed characteristics and the selected occupational history, the measurement errors are independent across waves and independent of the shocks governing occupational choice. Their conditional distribution is

$$\nu_{it} \mid (s_i, X_i, o_i^*) \sim \mathcal{N}\left(0, \phi^2_{o_{it}^*,\tau(t)}\right), \qquad \nu_{it} \perp \nu_{ir} \mid (s_i, X_i, o_i^*) \quad \text{for } t \neq r. \tag{7}$$

The baseline therefore uses a common occupation-year measurement-variance process for observed and completed wages.[18]

Under these assumptions, observed and completed wages provide noisy measurements of the same potential-wage object. The conditional wage density is

$$f(w_{i,t}^* \mid s_i, o_{i,t}^* = o, X_{it}) = \frac{e^{-\frac{\left[w_{i,t}^* - \left(\eta_{o,\tau(t)} + \lambda_{o,\tau(t)} s_i + X_{it}^{w\prime} \beta_o^w\right)\right]^2}{2\phi^2_{o,\tau(t)}}}}{\sqrt{2\pi\phi^2_{o,\tau(t)}}}. \tag{8}$$

[17]The assumption of fixed skills seems defensible in the context of a five-quarter panel. The model controls for observable experience to account for common experience or tenure effects on wages.

[18]Appendix G.4 allows the measurement-error variances to differ between observed and completed wages. Appendix D further reports estimates under alternative imputation procedures and using only the directly observed wages at waves 1 and 5.

Combining the choice probability in Equation (5) with the wage density gives the conditional person likelihood contribution

$$\mathcal{L}_i(s_i) = \prod_{t=1}^{5} \Pr(o^*_{i,t} \mid s_i, g_i, X_{it}) f(w^*_{i,t} \mid s_i, o^*_{i,t}, X_{it}). \tag{9}$$

Finally, I integrate over the worker's graduate-status and cohort skill distribution:

$$\mathcal{L}_i = \int \mathcal{L}_i(s_i)\, dF_{g_i c_i}(s_i). \tag{10}$$

Conditional on the maintained measurement and choice assumptions, the same persistent s_i must rationalise the wage measurements and occupational choices, which ties the wage and assignment sides of the model together. Details on the numerical evaluation of the likelihood are provided in Appendix E.

4.2 Identification

Identification rests on two linked structures. The first is the selected wage-measurement system. After removing the observed wage covariates, define

$$\widetilde{w}_{i,t} = w^*_{i,t} - X^{w\prime}_{i,t} \beta^w_{o^*_{it}}. \tag{11}$$

The wage measurement in the selected occupation then satisfies

$$\widetilde{w}_{i,t} = \eta_{o^*_{it},\tau(t)} + \lambda_{o^*_{it},\tau(t)} s_i + \nu_{i,t}, \tag{12}$$

where s_i is the persistent latent skill index, $\eta_{o\tau}$ is the level of the occupation-year wage schedule and $\lambda_{o\tau}$ determines how strongly that schedule rewards latent skill. Repeated wage measurements therefore provide the variation used to distinguish persistent heterogeneity from transitory measurement noise over the five-quarter panel.

Conditional on the maintained measurement assumptions, this model provides a scalar linear counterpart of the selected repeated-measurement system in Diegert (2024). The wage equation separates the time-invariant scalar skill s_i from the time-varying measurement error ν_{it} and is additively separable in logs. With one latent factor and five completed wage measurements, the model satisfies the dimension condition $2K < T$. The restrictions $\lambda_{o\tau} > 0$ make each occupation-year wage schedule injective in skill, while the common support $s_i \in [0, 1]$ and the positive-loading restriction fix the location, scale and orientation of the latent index.

Occupational selection determines which wage schedule is observed at each wave, but does not prevent the selected wage from measuring the same underlying index. The common s_i must therefore account jointly for wage measurements arising under different occupation-year schedules. Variation across these selected measurements disciplines the distribution of s_i together with the intercepts $\eta_{o\tau}$ and loadings $\lambda_{o\tau}$. Under the maintained repeated-measurement restrictions, Diegert (2024) establishes identification of the latent distribution and selected wage schedules, up to the stated normalisations.

The formal result applies to the idealised completed-data model defined above. In the data, wages at waves 2–4 are generated using the observed endpoint wages, current occupations and worker characteristics. The imputation procedure is designed to recover the persistent wage component that is informative about s_i, so the completed measurements provide additional modelled wage-side signal about the same underlying object. Because the three predictions draw partly on common endpoint wage information, however, they may also transmit common residual noise across waves. Completion therefore introduces an additional channel of dependence in the measurement errors, making conditional independence a more demanding assumption than it would be with five directly observed wages. In the baseline, I use the additional signal contained in the completed wages while treating cross-wave independence as a working approximation.[19]

The second identifying structure is the multinomial-logit choice model.[20] Once the wage schedules and group-cohort skill distributions have been recovered, average wage differences across occupations are known functions of s_i and the observed covariates. Relative occupation probabilities then identify the remaining deterministic utility differences under the standard multinomial-logit structure.

4.3 Estimation and model fit

I estimate the model using an EM algorithm, which profiles the group-cohort histogram masses by expectation maximisation while optimising the remaining parameters. For this, I approximate each F_{gc} using a histogram with ten equal-width bins on $[0, 1]$, with a separate probability mass for every graduate-status and cohort cell. This flexible histogram approach follows the treatment of random coefficients in K. Train (2016).[21]

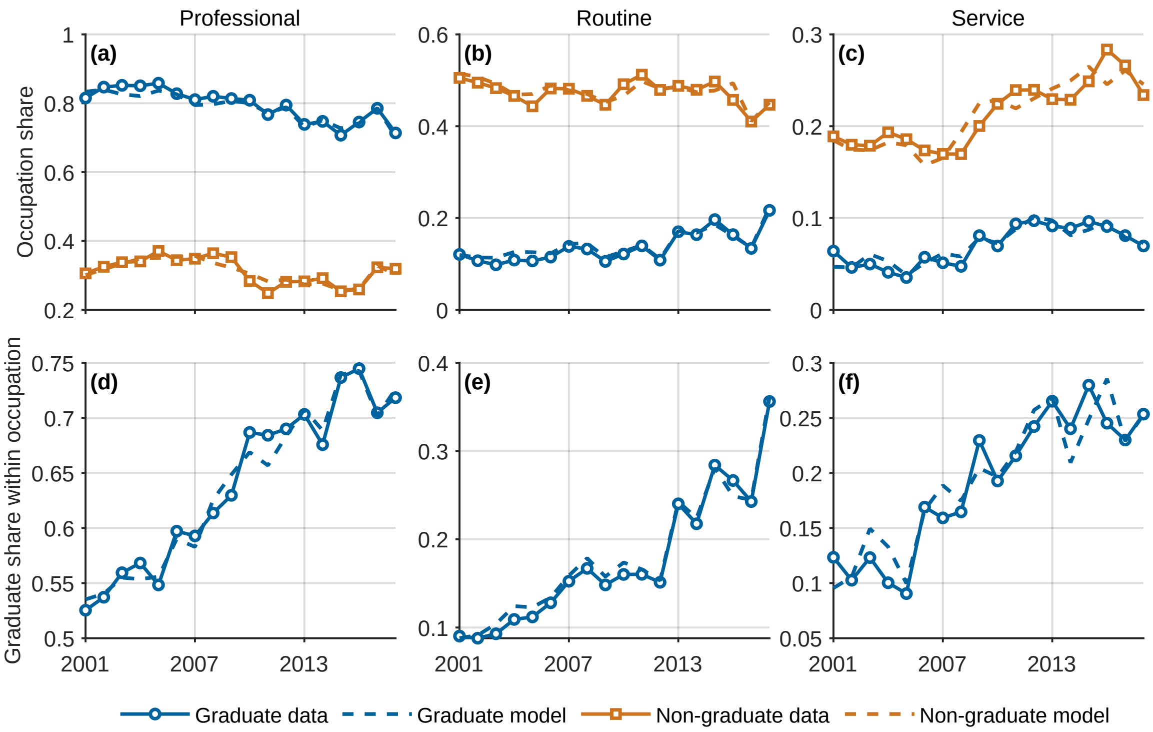
[19] Appendix D assesses the practical importance of this approximation by re-estimating the model under several alternative imputation procedures and using only the two directly observed wage measurements. The narrowing graduate–non-graduate skill gap, the occupational ranking of skill returns and the negative skill-distribution contribution to the premium decline remain robust across these specifications.

[20] Here I depart from Diegert (2024)'s non-parametric treatment of occupational sorting. Diegert estimates the joint distribution of skill and occupation histories, so the conditional probability of an occupation history given skill can be recovered by Bayes' rule without specifying a choice equation. I instead model quarterly occupational choices directly using a multinomial logit. This restriction allows expected wages, observed characteristics and group-specific non-pecuniary choice factors to enter occupational choice separately, which provides the interpretable choice-side objects used in the counterfactual analysis.

[21] The main results are robust to using fifteen equal-width skill bins; see Appendix G.2. I assess practical recovery in a simulation exercise, generating data from known parameters and re-estimating the model.

Figures 3 and 4 compare the estimated model with the corresponding sample moments. The model closely matches the broad occupation shares, mean wages and overall graduate premium. Fit is somewhat less precise for occupation-specific wages, where small over- and under-predictions persist over time. Appendix Table A2 reports the corresponding annual fit diagnostics.

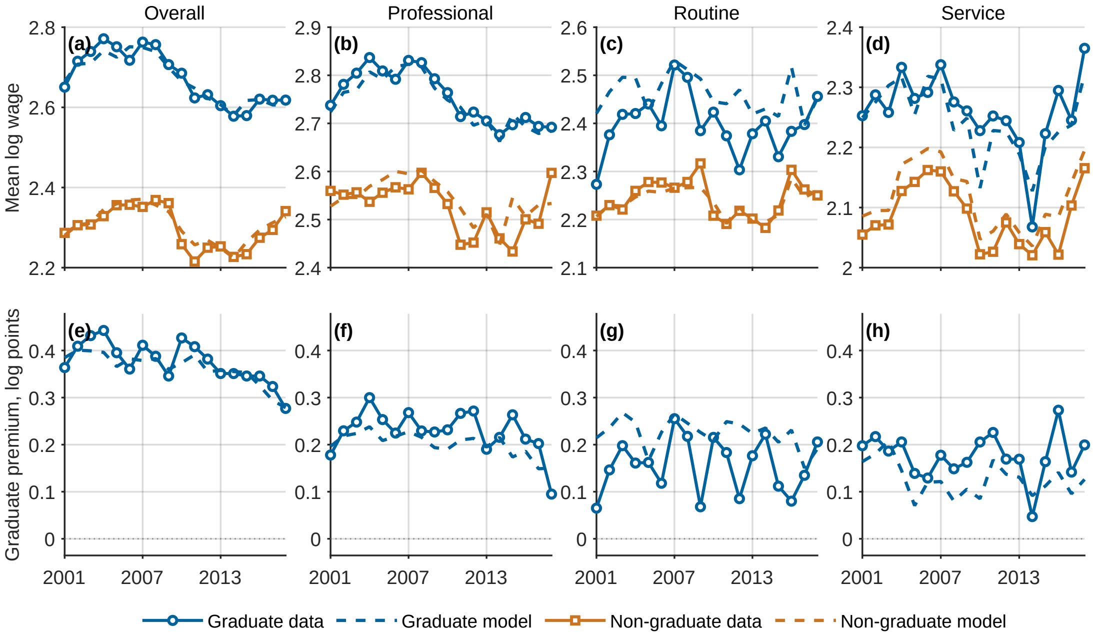
Figure 3: Model fit – occupation shares and graduate representation within occupation



Note: The top row compares annual sample and model-implied occupation shares within graduate-status groups. The bottom row compares the graduate share within each occupation.
Source: L-QLFS and Baseline Model Estimates.

The recovered estimates are centred around their true values. For details, see Appendix E.

Figure 4: Model fit – mean wages and graduate wage premia



Note: The top row compares annual sample and model-implied mean real log hourly wages overall and within occupation. The bottom row reports the corresponding graduate wage premia.
Source: L-QLFS and Baseline Model Estimates.

5 Estimated skill and occupational wage schedules

This section presents the two wage-side objects that underpin the model graduate premium.[22] I first compare the estimated skill distributions of graduates and non-graduates and then examine how occupations reward skill over time. Section 6 turns to the full wage schedules and estimated choice terms that map skill into occupational assignment, while Section 7 quantifies the implications for the level and evolution of the graduate premium.

5.1 Skill distribution

Figure 5 reports annual mean skill and dispersion for graduates and non-graduates. The underlying distributions are estimated by graduate status and cohort and are fixed within each group–cohort cell. The annual series combine these fixed distributions using the group-specific cohort composition observed in each year rather than estimating a separate skill distribution for every calendar year. Each point represents the mean (standard deviation) of the relevant workers observed in that year. For comparability, mean skill is centred on the 2001 mean skill pooled across both groups and scaled by the 2001 pooled standard deviation.

[22]The narrowing graduate–non-graduate skill gap and the occupational ranking of skill returns are robust to alternative age windows, wage-imputation methods and model specifications; see Appendices G.6, D and F.

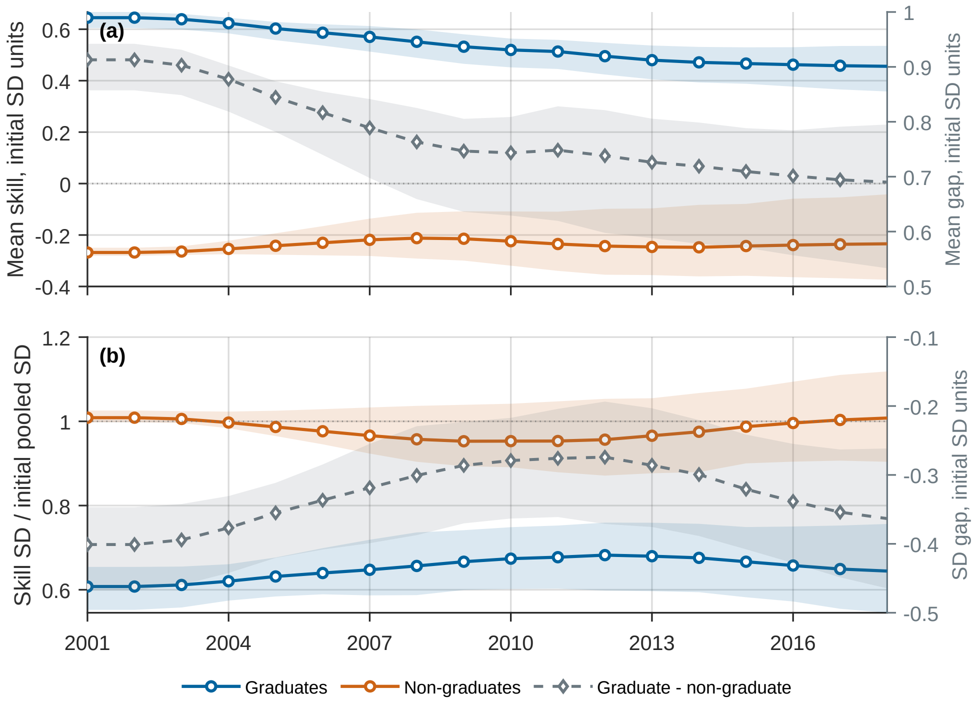
The model estimates suggest that graduates have a higher mean skill throughout the sample, but the gap narrows over time. The graduate mean averages 0.624 initial-standard-deviation units in 2001–06 and 0.466 in 2013–18, a fall of 0.158 initial standard deviations. The non-graduate mean is comparatively stable, increasing marginally from -0.254 to -0.241. As a result, the average gap between the two groups falls from 0.878 to 0.707 initial standard deviations, a narrowing of 0.171 standard deviations. Most of this movement therefore comes from the graduate side of the estimated distributions rather than a corresponding fall among non-graduates.

Within-group skill dispersion is broadly stable over time. The graduate distribution is consistently much less dispersed than the non-graduate distribution, and both standard deviations move only modestly across the sample.[23]

The fall in the estimated graduate mean is consistent with a composition effect from the expansion of higher education, although the model does not identify that historical mechanism directly. UK evidence documents both the scale of the expansion (Walker & Zhu, 2008; Blundell et al., 2022) and greater heterogeneity in prior attainment and occupational destinations among later graduate cohorts (Lindley & McIntosh, 2015). Related US evidence shows that expanded enrolment can lower average measured skill among graduates through changes in selection (Carneiro & Lee, 2011). In the present estimates, however, the main change appears as a shift in the location of the graduate distribution rather than a pronounced increase in its dispersion - although a small increase in graduate skill dispersion is observed.

[23] Figure A1 reports the full estimated distributions for each year-turned-25 cohort group.

Figure 5: Model-implied skill means and dispersion



Note: Annual graduate and non-graduate moments combine cohort distributions using the observed cohort composition. Means are centred on the pooled 2001 mean (0.445), and all moments are scaled by the pooled 2001 standard deviation (0.160). Dashed lines show graduate–non-graduate gaps on the right axes; skill is uniform within histogram bins. Shaded regions are 95% confidence intervals from 500 bootstrap replications.
Source: Baseline Model Estimates and bootstrap replications.

Holding the occupation-specific wage schedules and occupational-choice rules fixed, the narrower mean-skill gap reduces the part of the graduate premium generated directly by differences in skill. Because occupational choice also varies with skill, the same distributional change can affect the premium through assignment across occupations. Subsection 6.1 examines this second channel.

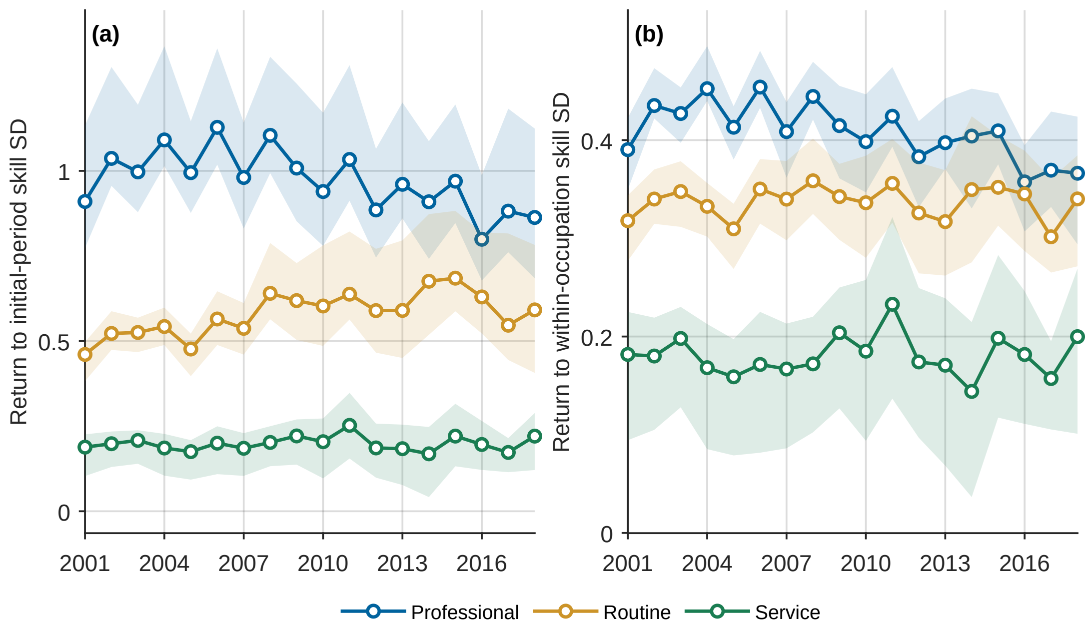
5.2 Returns to skill

Figure 6 reports the estimated within-occupation returns to skill under two normalisations. Panel (a) expresses each slope as the log-wage effect of a one-standard-deviation increase in skill using the initial pooled standard deviation, keeping the skill unit constant across periods and groups. Panel (b) instead scales each slope by the model-implied standard deviation of skill among workers assigned to that occupation in each year.

The figure shows two distinct patterns. First, there is a clear ranking across occupations: Professional work has the largest return to skill, Routine work lies in the middle and Service work has the smallest. Averaged over 2001–06, a one-initial-standard-deviation increase in skill is associated with 1.026 log-wage units in Professional, 0.515 in Routine and 0.193 in Service; by 2013–18, the corresponding values are 0.897, 0.620 and 0.194. This hierarchy is in line with the task literature: non-routine problem-solving and complex communication tasks are characteristic of Professional, managerial and technical work, and measured analytical tasks predict wage differences both across and within occupations (Autor, Levy, et al., 2003; Acemoglu & Autor, 2011; Autor & Handel, 2013). It also aligns with the external task evidence presented in Appendix C, where occupation–year cells with broader task scope, greater complexity and more problem-solving content have larger estimated skill-return gradients. Second, the occupational ranking is stable, but the gaps between occupations change. The Professional return falls by 0.129 log-wage units while the Routine return rises by 0.104 and the Service return remains broadly unchanged, narrowing the gap between Professional and Routine work.

Panel (b) shows that the same Professional–Routine–Service ordering remains when each slope is scaled by the model-implied dispersion of skill within the occupation. The gaps are smaller under this normalisation, but Professional work still offers the largest return to a one-standard-deviation movement in the skill distribution of its assigned workers.

Figure 6: Returns to skill under pooled and within-occupation standardisations



Note: Panel (a) reports $\lambda_{o\tau}$ multiplied by the 2001 pooled skill standard deviation (0.160). Panel (b) reports $\lambda_{o\tau}$ multiplied by the corresponding model-implied standard deviation of skill among workers assigned to occupation o in year τ. Shaded regions are 95% confidence intervals from 500 bootstrap replications.
Source: Baseline Model Estimates and bootstrap replications.

At fixed occupational assignment and within-occupation skill composition, a change in $\lambda_{o\tau}$ changes the graduate–non-graduate wage gap in occupation o in proportion to the difference in mean skill between the two groups in that occupation. The relevant skill compositions are themselves outcomes of occupational sorting and are reported in Appendix Figure A4. Since the Professional return falls slightly while the Routine and

Service returns rise, the aggregate effect on the graduate premium is ambiguous.

Returns can also affect occupational choice because a steeper wage schedule raises an occupation's wage offer more strongly at higher skill values. This is an occupational-sorting channel rather than a direct within-occupation return effect, so I next examine the full wage schedules and the resulting assignment patterns.

6 Skill, returns and occupational sorting

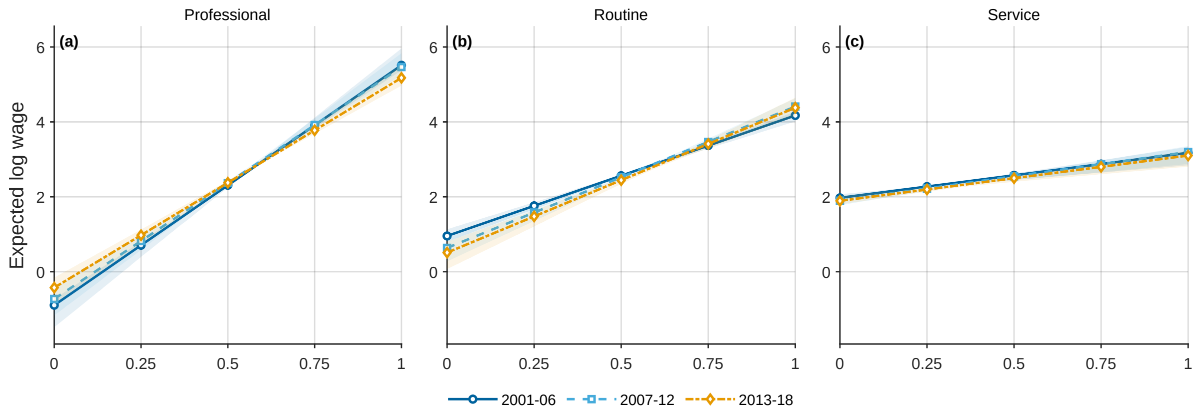
This section turns from the wage-side objects to occupational sorting. I first show how the full wage schedules generate incentives to sort by skill and examine the resulting assortative matching. I then separate the roles of skill composition and conditional assignment in the graduate–non-graduate Professional-employment gap.

6.1 Wage incentives and assortative matching

Occupational sorting is the outcome of the interaction of occupation-specific wage offers with the estimated choice terms. I first show the wage incentives implied by the full occupational schedules, then examine the resulting matching of skill to occupations. The model suggests that occupational sorting is highly stratified by skill, resulting in sharp differences in occupational allocation between graduates and non-graduates. However, skill-based sorting does not explain all occupational choices, and I explore this aspect by examining the remaining graduate–non-graduate difference in Professional assignment at a given skill level.

Figure 7 plots the expected wage offer in each occupation over the common skill support. Unlike Figure 6, which compares the slopes of the schedules, this figure also includes their fixed components and therefore shows how fitted wage offers vary with skill across occupations.

Figure 7: Expected occupational wage schedules over skill



Note: Panels (a)–(c) report the Professional, Routine and Service wage schedules, respectively. Each line gives the average schedule for one of the three broad periods. For each year and skill value, the model wage offer is averaged over the pooled survey-weighted worker composition observed in that year; the annual schedules are then averaged with equal weight within period. Shaded regions are 95% confidence intervals from 500 bootstrap replications.
Source: Baseline Model Estimates and bootstrap replications.

The wage schedules differ primarily in their slopes. The Professional schedule is substantially steeper than the Routine and Service schedules, so the wage incentive to enter Professional work rises strongly with skill. Conversely, the larger fixed components of the flatter schedules make Routine and Service work relatively more attractive at lower skill levels. The wage component of occupational choice therefore generates sorting pressure towards steeper schedules at higher skill levels and flatter schedules at lower skill levels.

These wage incentives are only one part of the assignment mechanism because the worker utilities also include the estimated choice factors. Appendix Figure A5 summarises the resulting assignment patterns.

Panel (a) reports the mean skill gradient attached to workers' assigned occupations. Graduates face a higher return gradient because they are more frequently employed in Professional work. The decline in this mean over the sample combines changes in occupational assignment with changes in the estimated occupation-specific gradients.

Panel (b) provides the more direct summary of assortative matching. The correlation between skill and the skill gradient in the assigned occupation is positive for both groups, indicating that workers with higher estimated skill tend to enter occupations with steeper wage schedules. The correlation is numerically higher for non-graduates. This means that their occupational assignment is more strongly stratified by skill: higher-skill non-graduates are matched more systematically to occupations with steeper schedules, while lower-skill non-graduates are matched to flatter schedules. Occupational allocation therefore transmits more of their within-group skill heterogeneity into the return schedule they face. The rise in the graduate correlation, particularly in the earlier part of the sample, means that graduate assignment also becomes more skill-contingent over time, increasingly separating higher- and lower-skill graduates across occupational return schedules.

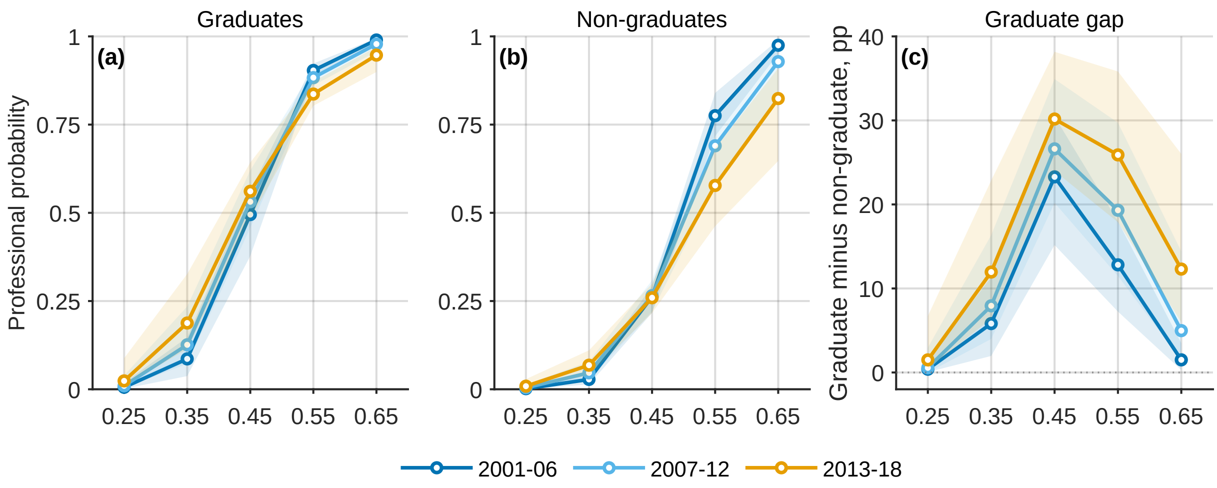
Thus, graduates face higher assigned returns on average, but non-graduates display tighter within-group matching between skill and returns. Appendix Figure A4 reports the corresponding skill composition within occupations. Mean skill is highest in Professional work, followed by Routine and then Service work, for both graduates and non-graduates, while graduates have higher mean skill within each occupation. Mean skill also falls over time in Professional work and rises in Routine work. These patterns are consistent with positive skill–return matching.

Taken together, the estimates show positive matching of skill to high-return occupations. I next ask how much of the graduate advantage in Professional employment reflects this skill-based sorting and how much remains in assignment conditional on skill.

6.2 Professional assignment over skill

The observed graduate–non-graduate gap in Professional employment is large throughout the sample: it is 49.6 percentage points in 2001–06 and remains 45.7 percentage points in 2013–18. A central question is whether this gap arises because graduates and non-graduates have different skill distributions, or because they continue to be assigned differently even if skill is held fixed. If the skill distributions account for the gap, higher graduate Professional employment primarily reflects the greater concentration of graduates in the upper part of the skill distribution, where the steep Professional wage schedule is relatively attractive. Convergence in the skill distributions should then narrow the employment gap even without a change in assignment conditional on skill. If graduates remain more likely to enter Professional work at the same skill level, however, group-specific choice factors sustain occupational separation beyond skill. This second case is consistent with degree-based access, different preferences, or search and recruitment frictions, and allows the employment gap to persist even as the skill distributions converge. Figure 8 first shows how Professional assignment varies over the common skill support.

Figure 8: Professional assignment over skill by graduate status



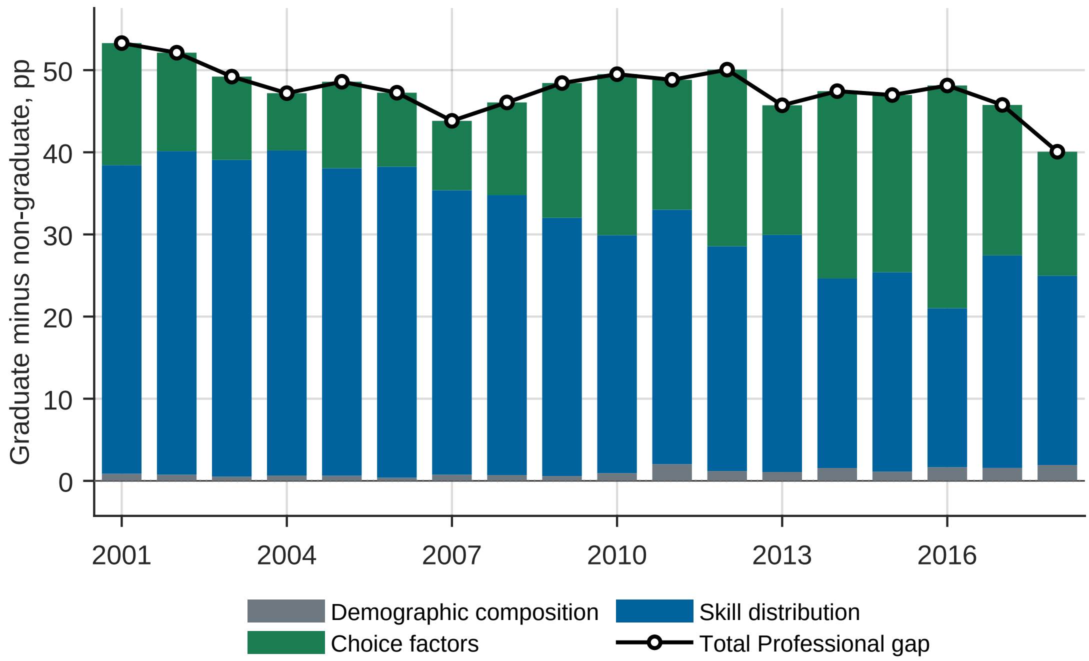
Note: Panels (a) and (b) report year averages for 2001–06, 2007–12 and 2013–18 of exact multinomial-logit Professional employment probabilities for graduates and non-graduates, respectively. Panel (c) reports the graduate-minus-non-graduate difference in percentage points. Shaded regions are 95% confidence intervals from 500 bootstrap replications.
Source: Baseline Model Estimates and bootstrap replications.

Panels (a) and (b) show that the predicted probability of Professional employment rises sharply with skill for both groups. Since the graduate skill distribution lies further to the right, this common gradient gives differences in skill an important role in the aggregate employment gap. Panel (c), however, shows that skill does not account for the entire difference: throughout the common skill support, graduates remain more likely to enter Professional work. At $s = 0.45$, the fitted graduate–non-graduate probability gap ranges from 23.3 to 30.2 percentage points across the three periods.

6.3 Decomposing the Professional-employment gap

Figure 9 converts these patterns into an annual aggregate decomposition, while Table 2 reports the corresponding three-year averages and 95% bootstrap confidence intervals. The skill-distribution component measures the contribution associated with the groups' different skill distributions. The choice-factor component measures the contribution associated with the group-specific non-wage occupational-choice terms and utility scale. These terms may reflect preferences, credentials, frictions or access constraints, although the model does not separate these interpretations. The remaining component captures differences in demographic composition.

Figure 9: Shapley decomposition of the graduate–non-graduate Professional-employment gap



Note: The figure reports annual percentage-point contributions to the fitted graduate-minus-non-graduate difference in Professional employment. Contributions are Shapley values over the demographic-composition, skill-distribution and choice-factor blocks and sum to the fitted gap in each year. Choice factors comprise the group-specific non-wage occupational-choice terms and utility scale.
Source: Baseline Model Estimates.

Table 2: Decomposition of the Professional-employment gap with 95% confidence intervals

Period	Total gap	Demographic	Skill distribution	Choice factors
2001–03	51.5	0.7	38.5	12.3
	[48.7, 54.4]	[0.2, 1.4]	[34.9, 42.3]	[7.4, 16.9]
2004–06	47.7	0.6	38.3	8.8
	[44.8, 49.7]	[0.1, 1.1]	[33.9, 41.8]	[4.2, 13.7]
2007–09	46.1	0.7	33.4	12.0
	[42.6, 49.0]	[0.1, 1.7]	[26.1, 37.2]	[6.5, 20.3]
2010–12	49.5	1.4	29.1	19.0
	[45.2, 54.1]	[0.6, 2.6]	[21.0, 34.5]	[11.9, 29.1]
2013–15	46.7	1.3	25.4	20.0
	[42.6, 50.6]	[0.3, 2.5]	[16.9, 32.7]	[11.9, 30.1]
2016–18	44.7	1.7	22.8	20.2
	[39.3, 48.7]	[0.7, 3.2]	[15.3, 29.1]	[12.6, 29.1]

Note: Point estimates with 95% bootstrap confidence intervals in brackets. Entries are three-year averages in percentage points. The total is the fitted graduate-minus-non-graduate Professional-employment gap. The three Shapley components change demographic composition, the latent skill distributions and residual occupation-choice factors. Point components sum to the total. Intervals are the 2.5th and 97.5th percentiles across 500 bootstrap replications.
Source: Baseline Model Estimates and bootstrap replications.

Differences in the skill distributions remain the largest source of the Professional-employment gap, but their importance declines substantially across the sample. The skill-distribution contribution falls from 38.5 percentage points, or 75% of the gap, in 2001–03 to 22.8 points, or 51%, in 2016–18. The choice-factor contribution rises from 12.3 to 20.2 points over the same periods. Demographic composition remains small, contributing between 0.6 and 1.7 percentage points across the six three-year blocks.

The confidence intervals leave the skill-distribution and choice-factor contributions positive in every three-year period. Their intervals overlap in the final two periods, however, so the point-estimate ranking between these components is less sharp at the end of the sample than it is initially.

The modest narrowing of the total employment gap therefore masks a larger decline in the skill-distribution component, partly offset by a stronger contribution from the remaining choice-factor margin. The model attributes most of the level difference in Professional employment to skill, but a growing share to assignment factors not summarised by skill or demographic composition.

External evidence provides a plausible credential-based interpretation of part of this increase. Using the British Skills and Employment Survey, Henseke et al. (2025) find that worker-reported degree requirements rose from 15% of jobs in 1997 to 29% in 2017. After 2006, degree requirements rose much faster than task-based graduate-skill requirements: changes in task profiles accounted for only about 2.2 percentage points of a 9-point increase in degree requirements, whereas they had accounted for the full increase before 2006. This divergence is consistent with degree credentials becoming more important in occupational access beyond the jobs' changing task content.

Taken together, the Professional-employment gap remains primarily a skill-composition gap, but conditional assignment becomes increasingly important as the skill distributions converge.

Appendix Figure A2 reports the corresponding decompositions for Routine and Service employment. Routine provides the closest complement to Professional employment: the skill-distribution contribution to the graduate–non-graduate Routine-employment gap shrinks from −20.0 percentage points in 2001–06 to −4.7 points in 2013–18, while the choice-factor contribution becomes more negative, from −16.4 to −23.2 points. Thus, as the skill distributions converge, skill accounts for much less of graduates' lower Routine employment, while conditional assignment increasingly sustains the gap. Taken across Professional and Routine employment, occupational separation is becoming less a consequence of the groups' different skill distributions and more a consequence of residual choice factors operating conditional on skill.

The results presented in this section answer one of the paper's central questions: whether graduates' greater concentration in Professional occupations reflects differences in skill or other assignment factors. Skill composition remains the main source of graduates' higher Professional employment, but residual assignment conditional on skill becomes increasingly important as the skill distributions converge.

7 Occupations and the graduate wage premium

Sections 5 and 6 described the skill distributions, occupational wage schedules and assignment patterns that shape the graduate premium. This section turns to the premium itself. I first compare graduate and non-graduate wages under different occupational choice regimes, in order to assess the importance of sorting for the graduate wage premium. I then use block counterfactuals and a Shapley decomposition to explain how the premium changes over time.[24]

7.1 Occupational assignment and the graduate premium

Positive assortative matching does not by itself determine whether occupational sorting raises or lowers the graduate wage premium. Sorting lowers the premium if occupation-specific assignment raises non-graduate wages by more than graduate wages, and raises it in the opposite case. Figure 10 evaluates this comparison using three counterfactual assignment rules. Panels (a) and (b) report how each rule changes graduate and non-graduate mean log wages relative to the fitted assignment schedule, while panel (c) reports the resulting change in the graduate wage premium. Each counterfactual holds the group-specific skill distributions, observed composition and occupational wage schedules fixed, but changes how workers are assigned across occupations.

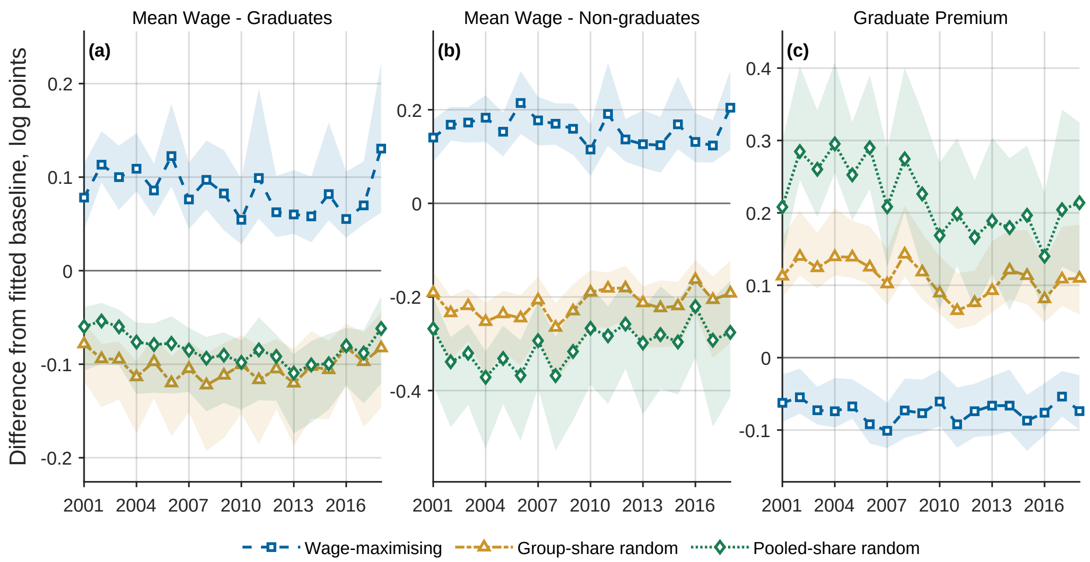
I consider three alternative assignment regimes: wage-maximising assignment, group-share random assignment and pooled-share random assignment. Wage-maximising assignment places each worker in the occupation offering the highest predicted wage given their characteristics and latent skill. This removes any residual choice factors or stochastic factors from occupational assignment, so that the only remaining source of occupational sorting is the interaction of skill with the occupation-specific wage schedules.

The other two alternative assignment rules allocate workers randomly across occupations, but differ in how they treat the two groups. Group-share random assignment randomly allocates workers according to the survey-weighted occupation shares observed for their own group in each year. It therefore removes within-group sorting by skill while preserving between-group differences in occupational allocation. Pooled-share random

[24]The central decomposition result—that the narrowing skill distribution provides the largest negative contribution to the premium decline—is robust to alternative age windows, wage-imputation methods and model specifications; see Appendices G.6, D and F.

assignment instead gives both groups the same overall annual occupation choice probabilities, removing both within-group sorting by skill and between-group differences in occupational allocation.

Figure 10: Occupational assignment, mean wages and the graduate wage premium



Note: Panels (a) and (b) report differences from fitted assignment in graduate and non-graduate mean log wages, respectively. Panel (c) reports the corresponding difference in the graduate wage premium, equal to the graduate wage effect in panel (a) minus the non-graduate wage effect in panel (b). Shaded regions are 95% confidence intervals from 500 bootstrap replications. *Source:* Baseline Model Counterfactuals and bootstrap replications.

The first two panels make the wage mechanics behind the premium effects explicit. Pooled-share and group-share random assignment lower the mean wage of both groups relative to fitted assignment, but lower non-graduate wages by more and therefore raise the premium. Wage-maximising assignment raises both groups' wages, but raises non-graduate wages by more and therefore lowers the premium. The premium effects consequently reflect asymmetric group wage gains from occupational assignment rather than a simple change in graduates' occupation shares.

The pooled-share random counterfactual raises the premium by 0.208 log points relative to fitted assignment in 2001 and by 0.214 log points in 2018. Occupational sorting in the fitted model therefore raises non-graduate wages by more than graduate wages and lowers the premium. The stability of this wedge also shows that sorting affects the level of the premium but accounts for little of its decline.

The group-share random counterfactual raises the premium by 0.113 log points in 2001 and by 0.110 log points in 2018, suggesting that within-group sorting also benefits non-graduates more. Their lower-centred and wider skill distribution gives lower-skill non-graduates scope to enter the flatter, higher-intercept Routine and Service occupations, while higher-skill non-graduates can enter the steeper Professional occupation. The same mechanism operates for graduates, but their more concentrated skill distribution leaves less scope to gain across the two tails.

Finally, the wage-maximising counterfactual lowers the premium by 0.062 log points relative to fitted assignment in 2001 and by 0.074 log points in 2018. Residual choice factors therefore raise the graduate premium on balance, implying that they constrain non-graduates' wage-maximising assignment more strongly. These factors can reflect preferences, access or other occupational frictions. One interpretation is that high-skilled non-graduates have weaker access to high-return Professional work than similarly skilled graduates.[25]

The analysis suggests that sorting is an important part of the graduate premium. Generally, sorting lowers the graduate premium, since it allows non-graduates to make occupational matches that are more advantageous than random assignment. However, given sorting the analysis also shows that residual choice factors raise the premium, suggesting that non-graduates face stronger constraints in their occupational assignment than graduates. In any case, the effect that sorting has on the level of the premium is stable over time, so it does not explain the observed decline in the graduate premium.

7.2 Simple counterfactual paths

The assignment counterfactuals establish how occupational sorting shapes the level of the graduate premium while holding the skill distributions, observed composition and wage schedules at their fitted values. In this subsection I relax the latter restriction and examine how the model components affect the level and evolution of the premium.

For this purpose, it is convenient to collect the model inputs in year τ into four blocks. Demographic composition $\mathcal{D}_\tau$ contains the group-specific distributions of sex, experience, part-time status and region. The skill block $\mathcal{S}_\tau$ contains the annual graduate and non-graduate skill distributions. The wage-schedule block $\mathcal{W}_\tau$ contains the occupation-specific wage intercepts and returns to skill. Finally, the choice-factor block $\mathcal{R}_\tau$ contains the non-pecuniary choice factors and group utility scales.

For any combination of these blocks, the model-implied mean log wage of group g can be written as

$$\overline{W}_g(\mathcal{D},\mathcal{S},\mathcal{W},\mathcal{R}) = \int_{\mathcal{X}}\int_0^1 \sum_{o\in\mathcal{O}} P_{og}(s,x;\mathcal{W},\mathcal{R})W_o(s,x;\mathcal{W})\,dF_g^{\mathcal{S}}(s)\,dH_g^{\mathcal{D}}(x). \tag{13}$$

Here, $H_g^{\mathcal{D}}$ describes the group's observed composition, while $F_g^{\mathcal{S}}$ is its skill distribution. $W_o(s,x;\mathcal{W})$ is the occupation-specific wage offer in Equation (2), and $P_{og}(s,x;\mathcal{W},\mathcal{R})$ is the occupational-choice probability implied by Equation (5). The model graduate premium

[25] Section 6 examined this possibility directly by comparing Professional employment across degree groups at a given skill level.

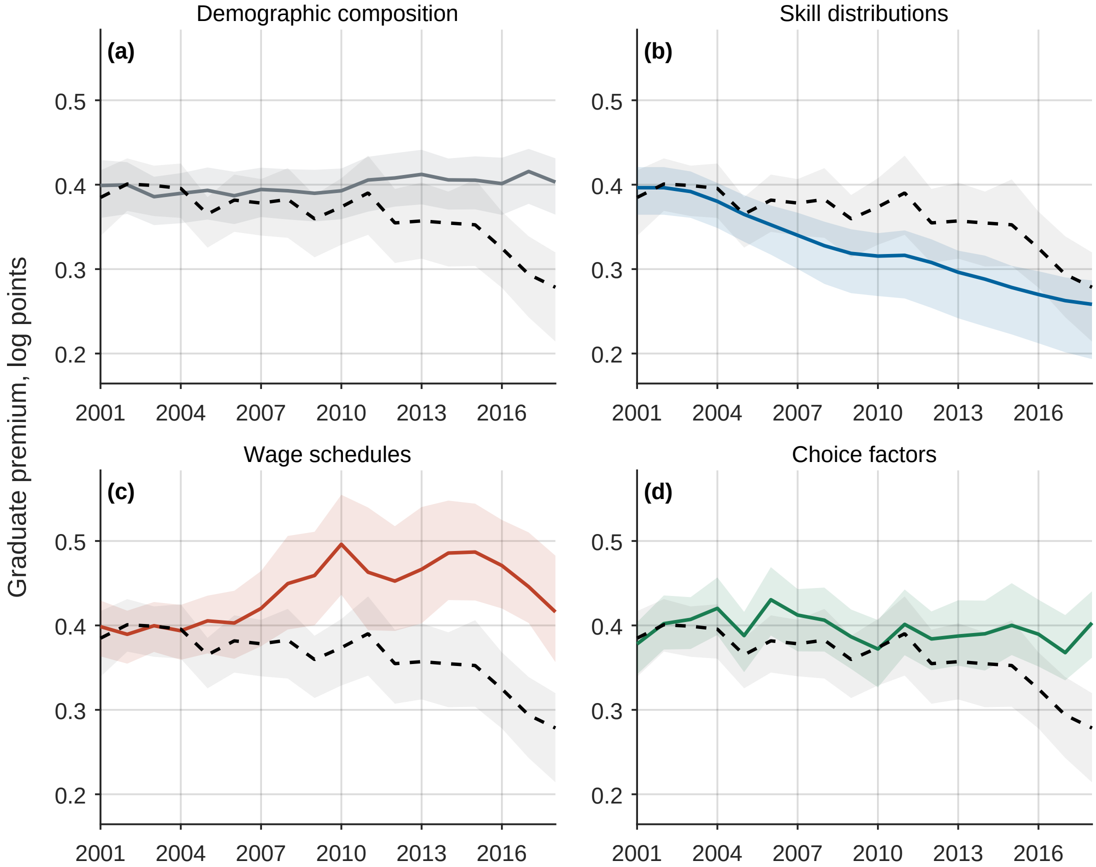
generated by any combination of blocks is therefore

$$GP(\mathcal{D}, \mathcal{S}, \mathcal{W}, \mathcal{R}) = \overline{W}_G(\mathcal{D}, \mathcal{S}, \mathcal{W}, \mathcal{R}) - \overline{W}_N(\mathcal{D}, \mathcal{S}, \mathcal{W}, \mathcal{R}). \tag{14}$$

The fitted premium in year τ uses $\mathcal{D}_\tau$, $\mathcal{S}_\tau$, $\mathcal{W}_\tau$ and $\mathcal{R}_\tau$. I use 2001–03 as the common reference period. Specifically, I construct each counterfactual three times, using the complete 2001, 2002 and 2003 model configurations in turn as the reference environment, and average the resulting model outcomes. I therefore average predicted wages rather than parameter vectors. I use two complementary exercises to examine how the model components affect the graduate premium. First, the one-block counterfactuals allow one component to follow its estimated annual path while holding all other components at each reference-year configuration. Second, the Shapley decomposition assigns the total change in the model premium across the four blocks by averaging each block's marginal contribution over all possible orderings and the three reference environments.

The one-block paths establish how each model block moves the premium in a common reference environment (Figure 11). For each of the three reference years, every model block is fixed at that year's value except for the named block, which follows its estimated annual path. The plotted path averages the three resulting outcomes. This isolates whether each component pushes the premium up or down, and when, without conflating it with simultaneous changes in the other blocks.

Figure 11: Simple counterfactuals



Note: Panels (a)–(d) allow only the named block to follow its estimated annual path, while holding the remaining blocks at the reference environment. Each path averages the counterfactual outcomes obtained using the complete 2001, 2002 and 2003 configurations in turn as that environment. Shaded regions are 95% confidence intervals from 500 bootstrap replications. The dashed line is the model-fitted graduate premium.
Source: Baseline Model Counterfactuals and bootstrap replications.

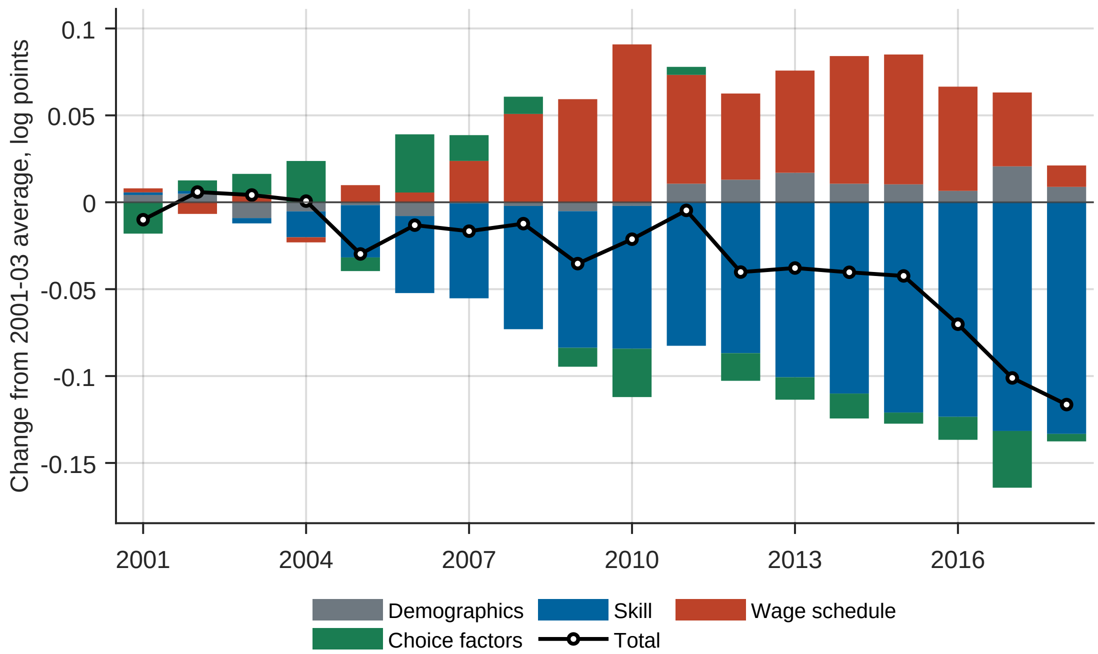
The mechanism-specific paths reinforce the preceding results. Observed composition alone leaves the premium close to its reference-period level, making composition an unlikely source of the sustained decline. Changing only the skill distributions instead produces a smooth fall, in line with the narrowing graduate–non-graduate skill gap documented above. The wage-schedule path moves in the opposite direction after 2007 before falling back towards the end of the sample. Its mid-sample rise coincides with the increasing Routine return, although the path changes all occupation-specific intercepts and skill returns and includes the occupational reallocation induced by those changes. Finally, updating only the choice factors moves the premium non-monotonically and remains close to the reference-period level on average. The estimated evolution of conditional assignment therefore does not by itself reproduce the decline.

7.3 Shapley decomposition of the changing graduate premium

The Shapley decomposition measures how much each model block contributes to the overall change in the graduate premium when the blocks interact. Intuitively, for each reference year it considers every possible sequence in which the model moves from that configuration to the configuration in year τ. I average the resulting state values across the three reference years before assigning each block its average marginal contribution across sequences (Shorrocks, 2013).[26]

The decomposition uses the 2001–03 average model premium as the common anchor. Figure 12 reports the annual decomposition, while Table 3 reports unweighted three-year averages of the observed and model premium changes and the four Shapley contributions, together with 95% bootstrap confidence intervals. By construction, each component averages to zero over 2001–03.

Figure 12: Annual Shapley decomposition of the change in the model graduate premium



Note: Bars report annual Shapley contributions to the change in the model graduate premium relative to the 2001–03 average. The black line reports the total change. Demographic composition, the skill distributions, wage schedules and choice factors sum to the total in each year.
Source: Baseline Model Counterfactuals.

[26]For each target year and each reference year, I evaluate all combinations of reference and current blocks. Averaging state values across the three reference years before applying the Shapley formula is equivalent to averaging the three reference-year Shapley decompositions. This allocates interactions symmetrically and makes the four contributions add exactly to the change from the mean 2001–03 model premium. Changes in the skill and wage blocks can induce occupational reallocation, so this re-sorting is assigned to those blocks rather than exclusively to the choice-factor block.

Table 3: Graduate-premium changes and Shapley contributions with 95% confidence intervals

Period	Observed change	Model change	Demographic	Skill distributions	Wage schedules	Choice factors
2004–06	-0.002 [-0.046, 0.045]	-0.014 [-0.041, 0.012]	-0.005 [-0.018, 0.009]	-0.030 [-0.047, -0.014]	0.004 [-0.020, 0.027]	0.016 [-0.001, 0.035]
2007–09	-0.020 [-0.066, 0.027]	-0.021 [-0.065, 0.016]	-0.003 [-0.016, 0.012]	-0.068 [-0.107, -0.036]	0.045 [0.004, 0.090]	0.005 [-0.016, 0.025]
2010–12	0.004 [-0.052, 0.059]	-0.022 [-0.067, 0.025]	0.007 [-0.008, 0.022]	-0.084 [-0.134, -0.046]	0.068 [0.025, 0.124]	-0.013 [-0.032, 0.011]
2013–15	-0.052 [-0.103, 0.005]	-0.040 [-0.089, 0.007]	0.013 [-0.002, 0.029]	-0.111 [-0.165, -0.074]	0.069 [0.031, 0.115]	-0.011 [-0.027, 0.017]
2016–18	-0.086 [-0.148, -0.030]	-0.096 [-0.150, -0.049]	0.012 [-0.004, 0.030]	-0.129 [-0.188, -0.091]	0.038 [0.009, 0.085]	-0.017 [-0.033, 0.008]

Note: Point estimates with 95% confidence intervals from bootstrap replications in brackets. Entries are unweighted three-year averages. All columns report changes relative to the 2001–03 average.
Source: Baseline Model Estimates, Baseline Model Counterfactuals and bootstrap replications.

The decomposition attributes the decline in the model graduate premium primarily to a competition between the changing skill distributions and wage schedules. Relative to the 2001–03 average, the narrowing graduate–non-graduate skill gap lowers the premium by 0.076 log-wage units in 2007–12 and by 0.120 in 2013–18, driven mainly by the downward shift in the graduate distribution. The wage schedules work in the opposite direction, offsetting 0.056 log-wage units in the middle period and 0.054 in the final period. They therefore almost neutralise the skill-driven compression initially, but fall behind as the negative skill contribution grows. Choice factors contribute -0.004 and -0.014 log-wage units, while demographic composition contributes 0.002 and 0.012. The result is a 0.022-log-point decline in 2007–12 and a 0.068-log-point decline in 2013–18.

The bootstrap intervals keep the skill-distribution contribution below zero and the wage-schedule contribution above zero in every three-year period from 2007 onwards. The interval for the total model change excludes zero only in 2016–18, when the skill-driven compression is no longer offset by the other blocks.

Table 4 reports unweighted three-year averages of the nested annual contributions plotted in Appendix Figure A6. It further decomposes the two main drivers of the premium: the skill-distribution component by degree group in Panel A and the wage-schedule component by occupation in Panel B. Every underlying annual contribution uses the 2001–03 average as its common base. Panel A shows that the graduate distribution accounts for almost all of the skill component's growing negative effect: it contributes -0.026 log-wage units in 2004–06, -0.060 in 2007–09 and -0.117 in 2016–18. The non-graduate contribution remains much smaller, reaching -0.012 in the final period. The compression generated by the changing skill distributions therefore comes primarily from the graduate distribution shifting down. This corresponds directly to the estimates in Section 5.1, where the closing of the mean graduate–non-graduate skill gap is mainly driven by reductions in the graduate mean.

Panel B splits the offsetting wage-schedule contribution by occupation. The Professional schedule drives the initial offset: it contributes 0.063 log-wage units in 2007–09, while the Routine and Service schedules together subtract 0.019. The contribution broadens thereafter. By 2013–15, Professional contributes 0.020 log-wage units and Routine and Service together contribute a further 0.049. In 2016–18, Professional and Routine contribute 0.026 and 0.014, while Service contributes −0.002. The wage schedules therefore initially offset the skill-driven compression through Professional work, before the contribution becomes more broadly distributed across occupations. The early offset is concentrated in Professional work, whose return to skill remains above its 2001–03 average through 2007–09 and therefore favours the higher-skilled graduate group disproportionately employed there; the lower Professional intercept offsets part of this slope channel. After the GFC, the Professional slope falls and that channel weakens, although the rise in the Professional intercept helps keep its total contribution positive. The post-GFC Routine and Service contributions reflect a Routine schedule that rotates towards skill—a higher slope paired with a markedly lower intercept—and a Service schedule that shifts down at an almost unchanged slope through 2013–15. These changes raise the premium given non-graduates' greater exposure to these occupations and lower-centred skills.

Table 4: Nested Shapley contributions with 95% confidence intervals

Component	2004–06	2007–09	2010–12	2013–15	2016–18
Panel A: Skill-distribution component by degree group					
Graduates	-0.026	-0.060	-0.084	-0.110	-0.117
	[-0.043, -0.009]	[-0.098, -0.024]	[-0.126, -0.044]	[-0.158, -0.065]	[-0.172, -0.067]
Non-graduates	-0.004	-0.008	0.000	-0.001	-0.012
	[-0.015, 0.005]	[-0.036, 0.013]	[-0.039, 0.030]	[-0.047, 0.032]	[-0.066, 0.023]
Total	-0.030	-0.068	-0.084	-0.111	-0.129
	[-0.047, -0.014]	[-0.107, -0.036]	[-0.134, -0.046]	[-0.165, -0.074]	[-0.188, -0.091]
Panel B: Wage-schedule component by occupation					
Professional	0.045	0.063	0.032	0.020	0.026
	[0.027, 0.064]	[0.036, 0.096]	[-0.002, 0.070]	[-0.009, 0.054]	[-0.005, 0.062]
Routine	-0.015	-0.007	0.017	0.025	0.014
	[-0.027, -0.006]	[-0.023, 0.014]	[0.001, 0.040]	[0.010, 0.046]	[0.001, 0.031]
Service	-0.026	-0.012	0.019	0.024	-0.002
	[-0.039, -0.013]	[-0.030, 0.006]	[-0.004, 0.048]	[-0.001, 0.048]	[-0.021, 0.021]
Total	0.004	0.045	0.068	0.069	0.038
	[-0.020, 0.027]	[0.004, 0.090]	[0.025, 0.124]	[0.031, 0.115]	[0.009, 0.085]

Note: Point estimates with 95% confidence confidence intervals from bootstrap replications in brackets. Entries are unweighted three-year averages of the nested annual Shapley contributions plotted in Figure A6. Every annual contribution is measured relative to the 2001–03 average, in log-wage units.
Source: Baseline Model Counterfactuals and bootstrap replications.

Overall, the two exercises point to the same account of the declining graduate premium. The narrowing skill gap is the dominant force, driven mainly by the downward shift in the graduate skill distribution. Wage schedules offset much of this pressure, first through Professional occupations and later through Routine and Service occupations, but the offset is not large enough by the end of the sample. Choice factors make a small additional negative contribution, while observed composition provides a small offset.

8 Conclusion

This paper has investigated how occupational sorting shapes the early-career graduate wage premium in the UK, and how that relationship changed in the first two decades of the 21st century. The model features differences in skill, differences in how occupations reward that skill, and differences in occupational assignment conditional on skill. I use this framework to analyse the role of skill differentials in shaping occupational sorting and the graduate wage premium.

The results point to four broad findings. First, graduates remain higher in the estimated distribution of skill, but their advantage narrows by around 0.17 initial standard deviations over the sample period. This convergence comes almost entirely from a downward shift in the graduate distribution rather than a substantial change among non-graduates. Second, occupations continue to reward skill very differently. Professional occupations offer the steepest wage schedules, Service occupations the flattest, and Routine occupations lie between them. This hierarchy remains stable, although the gap between Professional and Routine work narrows as the Professional return falls and the Routine return rises. Third, workers sort strongly across these schedules: those with higher estimated skill tend to enter occupations that reward skill more strongly. Fourth, the interaction between skill differences, occupational wage schedules, and sorting patterns is central to understanding changes in the graduate wage premium, with the declining skill gap emerging as the dominant force behind its reduction over time.

Skill and occupational rewards therefore explain much of the sorting between graduates and non-graduates. Graduates enter Professional work more often partly because they occupy a higher position in the skill distribution and gain more from its steep wage schedule. But this mechanism becomes less dominant over time. Skill accounts for 75% of the graduate–non-graduate Professional-employment gap in 2001–03 and 51% in 2016–18, while the contribution of other choice factors rises from 24% to 45%. Occupational sorting remains strongly structured by skill, but differences in assignment conditional on skill increasingly sustain the graduate advantage in Professional employment.

Sorting has a less straightforward effect on the graduate wage premium itself. Relative to the pooled-share random occupational assignment, the sorting generated by the model lowers the premium by around 0.21 log points in both 2001 and 2018. Both groups benefit from matching their skills to suitable occupations, but non-graduates gain more. Sorting therefore compresses the level of the graduate premium, while its stability means that it contributes little to the premium's decline.

A decomposition instead assigns the decline primarily to the narrowing skill gap. Relative to the 2001–03 average, changes in the skill distributions lower the model premium by 0.120 log points in 2013–18, driven overwhelmingly by the graduate distribution. Changes in occupational wage schedules offset part of this decline, pushing in the opposite direction, while other factors play a relatively minor role. The decline therefore originates mainly in the changing graduate skill distribution; occupational wage schedules and sorting determine how that change reaches wages.

Taken together, the findings point to two broader implications. First, changes in the distribution of labour-market-relevant skill among graduates are central to the decline in the premium, although the model does not distinguish changes in selection into higher education from changes in skills acquired during higher education. Second, differences in Professional employment are increasingly sustained by assignment conditional on skill. This directs attention towards degree requirements, recruitment practices and search frictions as possible mechanisms, that might be worthy of the attention of policy makers going forward.

References

Abowd, John M., Francis Kramarz & David N. Margolis (1999). "High Wage Workers and High Wage Firms". In: *Econometrica* 67.2, pp. 251–333. DOI: 10.1111/1468-0262.00020.

Acemoglu, Daron & David H. Autor (2011). "Skills, Tasks and Technologies: Implications for Employment and Earnings". In: *Handbook of Labor Economics*. Ed. by Orley Ashenfelter & David Card. Vol. 4. Elsevier, pp. 1043–1171.

Altonji, Joseph G., Peter Arcidiacono & Arnaud Maurel (2016). "The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects". In: *Handbook of the Economics of Education*. Ed. by Eric A. Hanushek, Stephen J. Machin & Ludger Woessmann. Vol. 5. Elsevier, pp. 305–396.

Autor, David H. & Michael J. Handel (2013). "Putting Tasks to the Test: Human Capital, Job Tasks, and Wages". In: *Journal of Labor Economics* 31.S1, S59–S96.

Autor, David H., Frank Levy & Richard J. Murnane (2003). "The Skill Content of Recent Technological Change: An Empirical Exploration". In: *The Quarterly Journal of Economics* 118.4, pp. 1279–1333.

Blundell, Richard, David A. Green & Wenchao Jin (2022). "The U.K. as a Technological Follower: Higher Education Expansion and the College Wage Premium". In: *The Review of Economic Studies* 89.1, pp. 142–180. DOI: 10.1093/restud/rdab034.

— (2024). "Correction to: The U.K. as a Technological Follower: Higher Education Expansion and the College Wage Premium". In: *The Review of Economic Studies* 91.4, pp. 2538–2541. DOI: 10.1093/restud/rdae029.

Boero, Gianna et al. (2025). "The College Wage Premium in the UK: Decline and Fall?" In: *Oxford Economic Papers* 77.1, pp. 1–18. DOI: 10.1093/oep/gpae014.

Carneiro, Pedro & Sokbae Lee (2011). "Trends in Quality-Adjusted Skill Premia in the United States, 1960–2000". In: *American Economic Review* 101.6, pp. 2309–2349.

Diegert, Paul (2024). "The Role of Skills and Sorting in Explaining Wage Inequality". Unpublished manuscript.

Gathmann, Christina & Uta Schönberg (2010). "How General Is Human Capital? A Task-Based Approach". In: *Journal of Labor Economics* 28.1, pp. 1–49. DOI: 10.1086/649786.

Gottschalk, Peter & Michael Hansen (2003). "Is the Proportion of College Workers in Noncollege Jobs Increasing?" In: *Journal of Labor Economics* 21.2, pp. 449–471. DOI: 10.1086/345564.

Green, Francis & Golo Henseke (2016). "The Changing Graduate Labour Market: Analysis Using a New Indicator of Graduate Jobs". In: *IZA Journal of Labor Policy* 5.14, pp. 1–25. DOI: 10.1186/s40173-016-0070-0.

Green, Francis & Yu Zhu (2010). “Overqualification, Job Dissatisfaction, and Increasing Dispersion in the Returns to Graduate Education”. In: *Oxford Economic Papers* 62.4, pp. 740–763. DOI: 10.1093/oep/gpq002.

Heckman, James J., Jora Stixrud & Sergio Urzua (2006). “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior”. In: *Journal of Labor Economics* 24.3, pp. 411–482.

Henseke, Golo et al. (2025). “Degrees of demand: a task-based analysis of the British graduate labour market”. In: *Oxford Economic Papers* 77.1, pp. 144–165. DOI: 10.1093/oep/gpae020.

Huggett, Mark, Gustavo Ventura & Amir Yaron (2011). “Sources of Lifetime Inequality”. In: *American Economic Review* 101.7, pp. 2923–2954. DOI: 10.1257/aer.101.7.2923.

Lindley, Joanne & Steven McIntosh (2015). “Growth in Within Graduate Wage Inequality: The Role of Subjects, Cognitive Skill Dispersion and Occupational Concentration”. In: *Labour Economics* 37, pp. 101–111. DOI: 10.1016/j.labeco.2015.03.015.

Lise, Jeremy & Fabien Postel-Vinay (2020). “Multidimensional Skills, Sorting, and Human Capital Accumulation”. In: *American Economic Review* 110.8, pp. 2328–2376. DOI: 10.1257/aer.20162002.

O’Leary, Nigel & Peter Sloane (2011). “The Wage Premium for University Education in Great Britain During a Decade of Change”. In: *The Manchester School* 79.4, pp. 740–764. DOI: 10.1111/j.1467-9957.2010.02189.x.

— (2016). “Too Many Graduates? An Application of the Gottschalk–Hansen Model to Young British Graduates between 2001–2010”. In: *Oxford Economic Papers* 68.4, pp. 945–967. DOI: 10.1093/oep/gpw027.

Roy, Andrew Donald (1951). “Some Thoughts on the Distribution of Earnings”. In: *Oxford Economic Papers* 3.2, pp. 135–146.

Shorrocks, Anthony F. (2013). “Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value”. In: *The Journal of Economic Inequality* 11.1, pp. 99–126. DOI: 10.1007/s10888-011-9214-z.

Train, Kenneth (2016). “Mixed Logit with a Flexible Mixing Distribution”. In: *Journal of Choice Modelling* 19, pp. 40–53.

Train, Kenneth E. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Walker, Ian & Yu Zhu (2008). “The College Wage Premium and the Expansion of Higher Education in the UK”. In: *The Scandinavian Journal of Economics* 110.4, pp. 695–709. DOI: 10.1111/j.1467-9442.2008.00557.x.

A Additional Figures and Tables

Table A1: Occupational decomposition of the graduate wage premium

	2001–06	2007–12	2013–18
Panel A: Symmetric Oaxaca–Blinder decomposition			
Adjusted graduate premium	36.5***	34.8***	28.1***
	(1.079)	(1.370)	(1.611)
Between-occupation component	16.4***	15.1***	13.4***
	(0.655)	(0.847)	(0.885)
Within-occupation component	20.1***	19.7***	14.7***
	(1.124)	(1.531)	(1.646)
Panel B: Change decomposition, 2013–18 minus 2001–06			
	Change	**Share (%)**	
Change in adjusted graduate premium	−8.4***	100.0	
	(2.038)		
Change in between-occupation component	−3.0***	35.2	
	(1.104)		
Change in within-occupation component	−5.5***	64.8	
	(2.050)		

Note: Entries are log points; standard errors are in parentheses. Panel A applies a symmetric twofold Oaxaca–Blinder decomposition to wages adjusted for age, age squared, residence region, part-time status, sex and year effects. The between component uses occupation-share differences evaluated at average adjusted occupation means; the within component uses adjusted wage gaps evaluated at average occupation shares. Panel B reports the corresponding change from 2001–06 to 2013–18. Decomposition standard errors use 399 individual-cluster bootstrap repetitions; the Panel B bootstrap jointly spans both periods. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.
Source: L-QLFS and author's calculations.

Table A2: Model-fit diagnostics

Graduate occupation share (%)	**Data**	**Model**	**Diff.**	**Diff. (%)**	**MAE**
Professional	79.464	78.781	-0.683	-0.859	1.196
Routine	13.587	14.169	0.581	4.279	0.744
Service	6.949	7.050	0.101	1.460	0.662
Non-graduate occupation share (%)	**Data**	**Model**	**Diff.**	**Diff. (%)**	**MAE**
Professional	31.311	31.089	-0.222	-0.708	1.293
Routine	47.390	47.679	0.289	0.610	1.072
Service	21.299	21.232	-0.068	-0.317	1.326
Graduate share within occupation (%)	**Data**	**Model**	**Diff.**	**Diff. (%)**	**MAE**
Professional	63.975	63.901	-0.074	-0.116	0.955
Routine	17.593	18.078	0.485	2.758	0.745
Service	19.031	19.350	0.319	1.674	1.642
Graduate mean log wage	**Data**	**Model**	**Diff.**	**Diff. (%)**	**MAE**
Mean overall	2.674	2.670	-0.004	-0.140	0.017
Mean Professional	2.755	2.745	-0.010	-0.361	0.018
Mean Routine	2.399	2.461	0.063	2.611	0.065
Mean Service	2.262	2.244	-0.017	-0.761	0.032
Non-graduate mean log wage	**Data**	**Model**	**Diff.**	**Diff. (%)**	**MAE**
Mean overall	2.298	2.307	0.009	0.377	0.015
Mean Professional	2.527	2.544	0.017	0.665	0.031
Mean Routine	2.243	2.237	-0.006	-0.250	0.013
Mean Service	2.086	2.118	0.032	1.510	0.032
Graduate wage premium	**Data**	**Model**	**Diff.**	**Diff. (%)**	**MAE**
Overall	0.375	0.363	-0.012	-3.306	0.021
Within Professional	0.228	0.201	-0.027	-11.735	0.035
Within Routine	0.156	0.224	0.068	43.777	0.071
Within Service	0.175	0.127	-0.049	-27.754	0.056

Note: Data and Model report unweighted averages of the annual moments over 2001–2018. Diff. is Model minus Data, while Diff. (%) expresses that difference relative to the average data moment. MAE is the mean absolute annual model–data difference. Occupation-share levels are percentages and their differences and MAEs are percentage points; wage and premium entries are log-wage units. The statistics describe point-estimate fit and are not uncertainty measures.
Source: L-QLFS and Baseline Model Estimates.

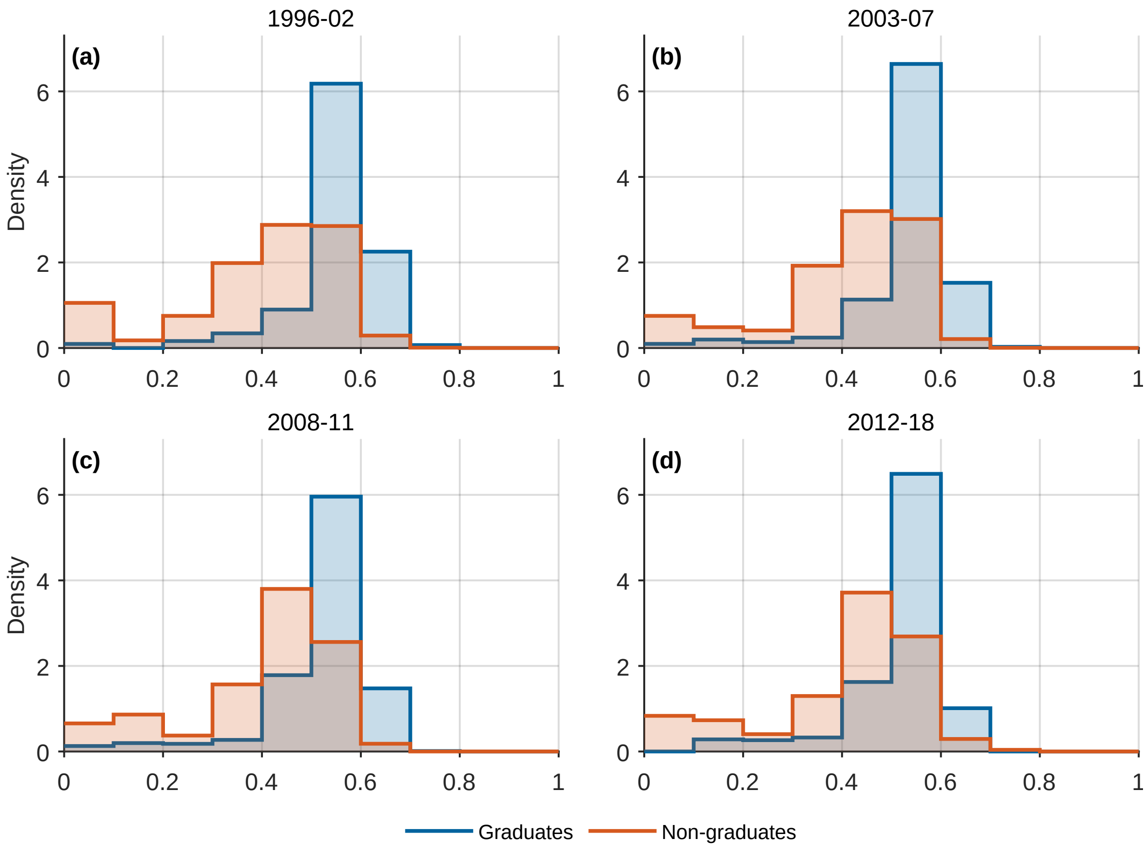


Figure A1: Cohort skill distributions by graduate status

Note: The figure reports the estimated ten-bin distributions of skill for graduates and non-graduates in the 1996–2002, 2003–07, 2008–11 and 2012–18 year-turned-25 cohorts. Bars show probability mass. Skill is uniformly distributed within each bin on the maintained common support $[0, 1]$.
Source: Baseline Model Estimates.

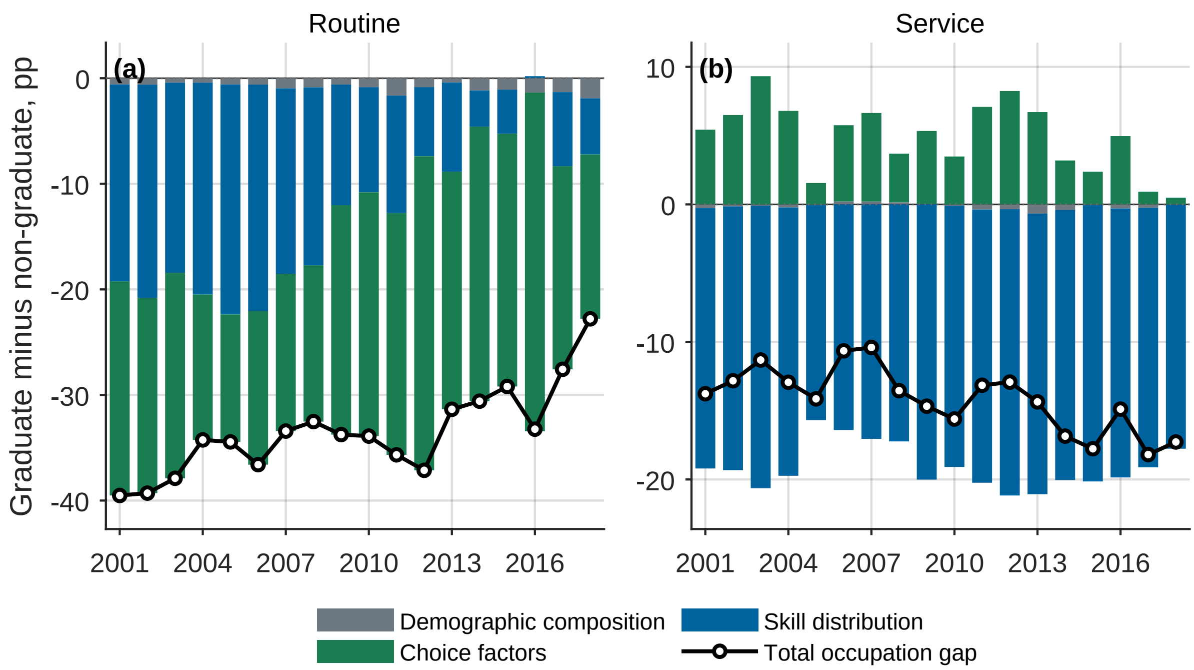


Figure A2: Shapley decompositions of the graduate–non-graduate Routine- and Service-employment gaps

Note: The panels extend Figure 9 to Routine and Service employment. Bars report signed Shapley contributions, in percentage points, from demographic composition, the skill distributions and choice factors; black lines report the corresponding total graduate-minus-non-graduate occupation-share gaps. Positive values indicate a higher occupation share among graduates and negative values a lower share.
Source: Baseline Model Estimates.

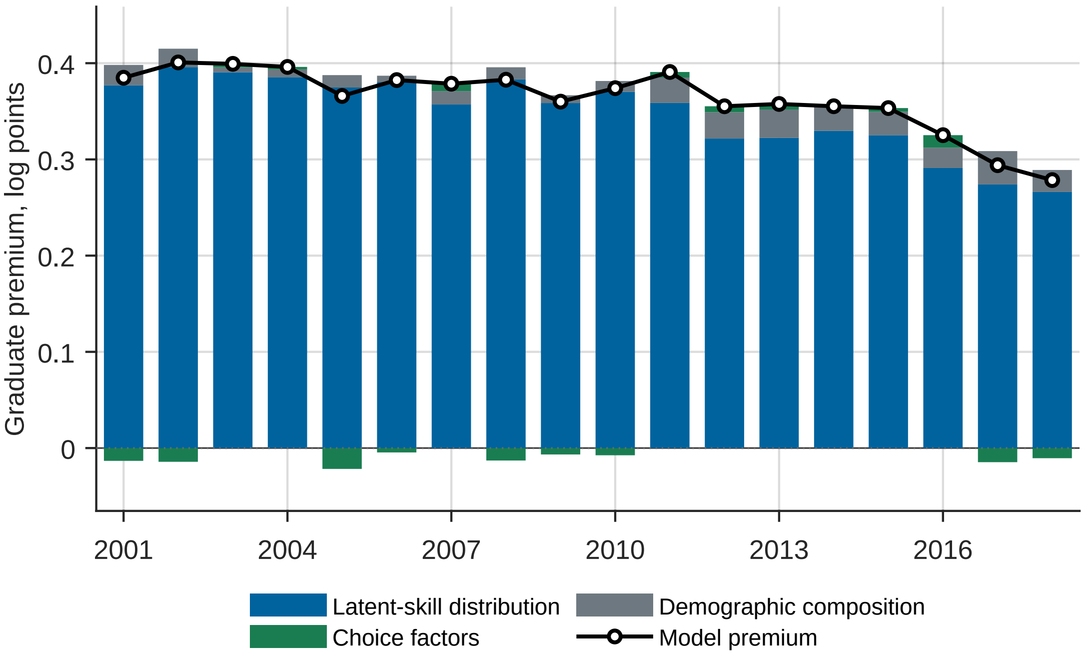


Figure A3: Level decomposition of the graduate wage premium

Note: The figure reports an exact aggregation of the annual primitive Shapley decomposition of the model-implied graduate wage premium.
Source: Baseline Model Estimates and Counterfactual Models.

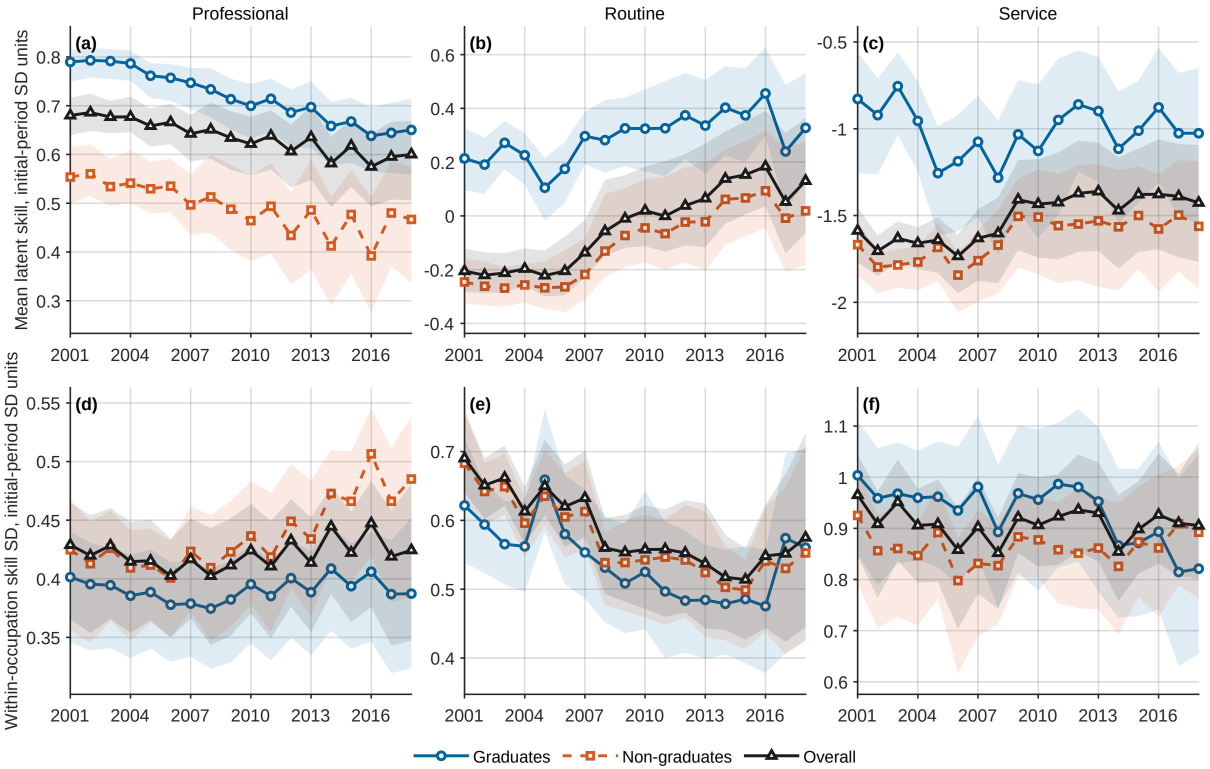


Figure A4: Mean skill within broad occupations

Note: The figure reports mean skill within each broad occupation, split by graduate status and for the pooled sample. Skill is centred on the pooled 2001 mean and divided by the corresponding pooled standard deviation. The paths report model-implied selected-cell compositions. Shaded regions are 95% confidence intervals from 500 bootstrap replications.

Source: Baseline Model Estimates and bootstrap replications.

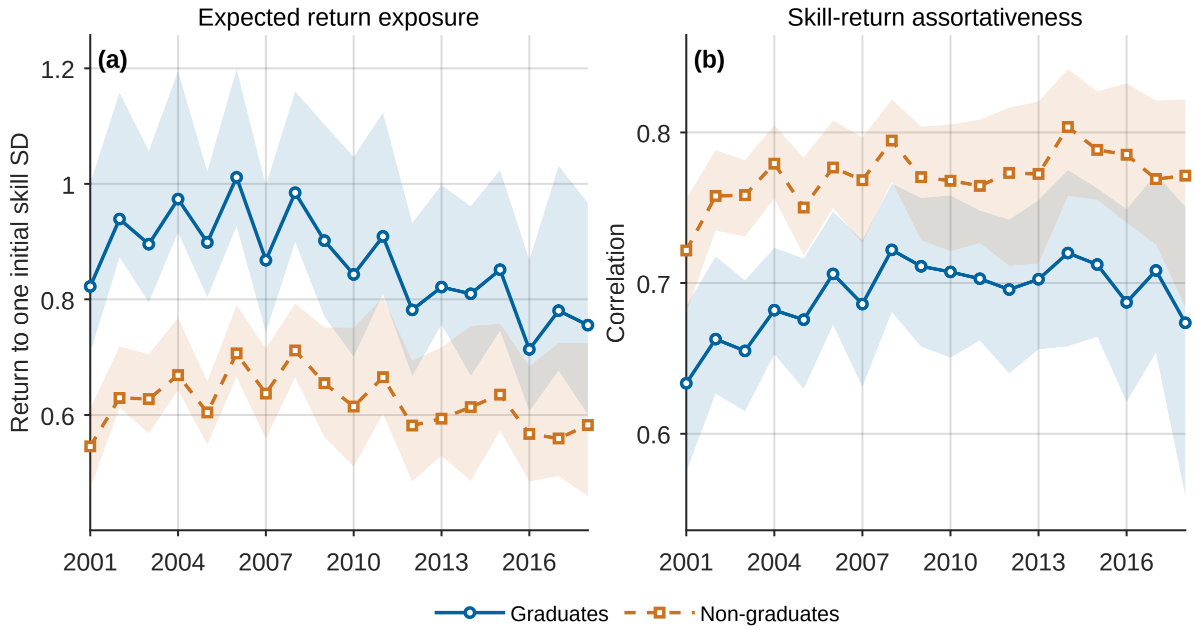


Figure A5: Expected return exposure and skill-return assortativeness

Note: Panel (a) reports each group's average return in its assigned occupation, multiplied by the pooled 2001 skill standard deviation. Panel (b) reports the model-implied correlation between skill and the return in the chosen occupation. Shaded regions are 95% confidence intervals from 500 bootstrap replications.
Source: Baseline Model Estimates and bootstrap replications.

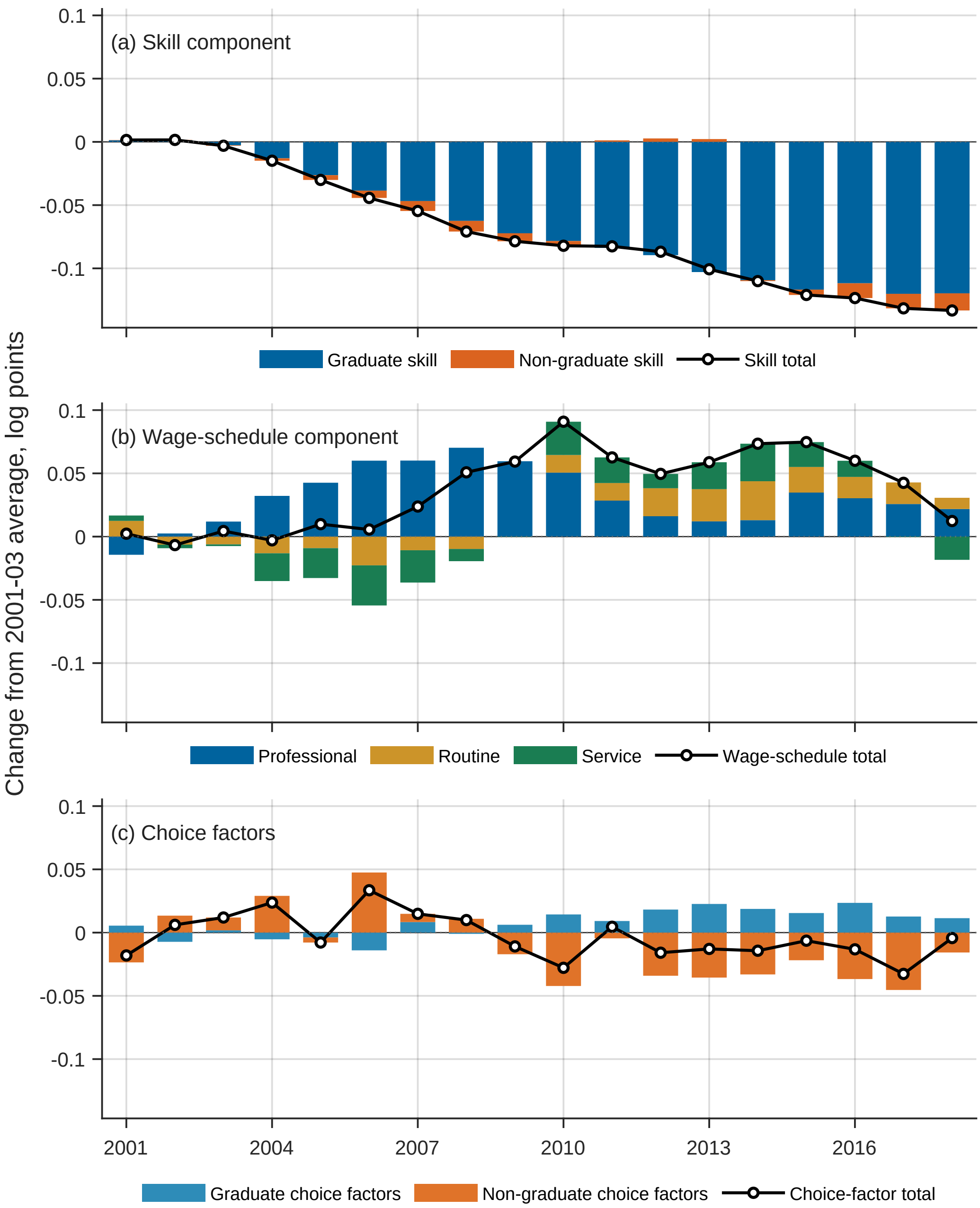


Figure A6: Detailed components of the annual graduate-premium change

Note: All panels report nested annual Shapley contributions to the change in the model premium relative to the 2001–03 average. Panel (a) splits the skill component by graduate status, panel (b) splits the wage-schedule component by occupation, and panel (c) splits the choice factors by graduate status.
Source: Counterfactual Models.

B L-QLFS Data and Sample Construction

This appendix describes the source data, the harmonisation and cleaning of the variables used in the analysis, and the sequential sample restrictions.

B.1 Panel structure and variable harmonisation

I use the quarterly five-wave L-QLFS extracts whose first interviews run from 2001Q1 to 2018Q4. Each source record contains five consecutive quarterly interviews. The extracts are available to registered users under the relevant access conditions through the UK Data Service Labour Force Survey series (SN 2000026), 10th Release (Office for National Statistics, 2024).

Education is measured at wave 1 and held fixed over the panel. I harmonise the changing HIQUAL and HIQUAP classifications into a binary graduate indicator: graduates have a degree-level qualification, while all other valid qualifications are classified as non-graduate. I retain wave-5 education only as a consistency check and exclude people who are full-time students at wave 1.

Employment status, age, full-time or part-time status, residence region and occupation are observed in each wave. I retain people who remain employees throughout all five waves and have valid full-time or part-time status and residence region in every wave. I harmonise residence into seven macro regions: North England, the Midlands, London, South and East England, Wales, Scotland and Northern Ireland. The empirical specifications use the wave-specific region and part-time indicators as observed controls.

I harmonise occupation codes across the SOC2000 and SOC2010 source variables. I retain a consistent SOC-1-digit occupation code and record whether any available SOC-2-digit code follows SOC2000 or SOC2010, thereby keeping the 2011 classification break explicit. The change from SOC90 to SOC2000 at 2001Q1 leaves some SOC-1-digit occupation cells missing around the 2000Q1–2001Q1 transition. For these cells, and only these cells, I use the nearest valid occupation observed for the same person in another wave. I then aggregate the SOC-1-digit codes into Professional (SOC-1-digit major groups 1–3), Routine (groups 4–5 and 8–9), and Service (groups 6–7) occupations.

The wage measure is reported gross hourly pay. I preserve the nominal value and deflate it using calendar-year CPI, with the CPI index normalised to 2014=100, so that real wages are expressed in 2014-price units. I set non-positive or otherwise invalid wages to missing and define the log wage only for valid observed values. In the relevant L-QLFS extracts, wages are reported at the first and fifth interviews; the intermediate measurements are missing by design and are imputed only after the final sample is selected, as described in Appendix D.

B.2 Baseline sample selection

I define the baseline sample using the first interview. I retain people aged 25–30 at wave 1 whose first interview occurs between 2001Q1 and 2018Q4. I then require people to remain employees throughout all five waves and to have five consecutive quarterly observations, valid wave-1 education, no full-time student status at wave 1, and valid full-time or part-time status, macro residence region and harmonised SOC-1-digit occupation in every wave. I also require a plausible quarterly age sequence. Table B1 reports the sequential selection counts.

Table B1: Construction of the baseline estimation sample

Sample definition	Persons	Person-wave observations
All target panels with valid wave-1 education	29,677	148,385
Continuously employed	19,815	99,075
Two valid wage observations (waves 1 and 5)	13,740	68,700
Valid other data (final estimation sample)	13,412	67,060

Note: Counts are unweighted. All observations are five-wave panels for people aged 25–30 at wave 1 whose first interview occurs between 2001Q1 and 2018Q4. The continuous-employment restriction requires workers to remain employees throughout all five waves. The two-wage restriction requires valid observed wages at waves 1 and 5. Valid other data requires valid wave-1 education, non-student status at wave 1, stable sex, a plausible age sequence, and valid part-time status, macro residence region and harmonised SOC-1-digit occupation in all five waves.
Source: L-QLFS and author's calculations.

Within this baseline panel, I winsorise observed real log hourly wages at the unweighted 0.1st and 99.9th percentiles. I retain people with exactly two valid observed wages and verify that these occur at waves 1 and 5. The final sample contains 13,412 people, 67,060 person-wave observations and 26,824 observed wage rows.

B.3 Baseline sample composition

The L-QLFS longitudinal weights account for attrition over the five-wave panel. I use them for the weighted descriptive statistics and in the survey-weighted structural likelihood.

Table B2 summarises the baseline sample by graduate status and panel-entry period. Women make up 46.9% of the weighted sample, and the weighted mean age at wave 1 is 27.57. The two education groups differ more markedly in where and how they work: 77.6% of graduates work in Professional occupations at wave 1, compared with 30.4% of non-graduates. Graduates are also more likely to live in London and less likely to work part-time. The graduate share rises from 34.5% among panels whose first interview occurs in 2001–06 to 50.0% among those beginning in 2013–18, while the female share and mean age at wave 1 remain comparatively stable.

A notable feature of the L-QLFS data is the steady decline of sampled participants, falling from over 7,000 in 2001–06 to around 2,300 in 2013–18. This trend is reflective of broader falling response rates across the QLFS and the UK household survey system more generally, and is not specific to the five-wave panel. The L-QLFS longitudinal weights account for this attrition, and the following subsection shows that my sample restrictions are not substantially affecting the composition of the estimation sample over time.

Table B2: Baseline sample characteristics by graduate status and panel-entry period

	Graduate status			Panel-entry period		
	All	Graduates	Non-graduates	2001–06	2007–12	2013–18
People	13,412	5,163	8,249	7,304	3,781	2,327
Person-wave observations	67,060	25,815	41,245	36,520	18,905	11,635
Female (%)	46.9	49.7	44.8	47.7	47.2	45.9
Mean age at wave 1	27.57	27.54	27.59	27.57	27.57	27.55
London residence (%)	13.3	20.3	8.0	12.1	12.9	14.7
Part-time at wave 1 (%)	14.5	6.5	20.5	13.6	14.5	15.2
Professional at wave 1 (%)	50.5	77.6	30.4	50.0	51.1	50.3
Routine at wave 1 (%)	33.7	14.8	47.8	36.1	33.2	32.2
Service at wave 1 (%)	15.8	7.7	21.8	13.9	15.7	17.5
Mean observed real log hourly wage	2.454	2.664	2.297	2.470	2.458	2.435

Note: Entry period is defined by the calendar year of wave 1. Counts are unweighted. Shares and mean age are wave-1 statistics weighted by the L-QLFS longitudinal weight. Mean log wages use survey-weighted observed wages at waves 1 and 5 after winsorisation. Professional comprises SOC-1-digit major groups 1–3, Routine groups 4–5 and 8–9, and Service groups 6–7.
Source: L-QLFS and author's calculations.

B.4 Impact of sample restrictions

Table B3 compares the corresponding graduate and non-graduate characteristics across the wider target panels and the intermediate restrictions.

The most impactful sample restriction is the requirement of continuous employment, which reduces the sample by around 30% and disproportionately removes women and part time workers - particularly among the non-graduate group. The remaining restrictions have smaller effects, and the final sample is broadly representative of the wider target panels.

Figure B1 shows that the sample restrictions have little effect on either the level or the decline of the observed graduate wage premium. The share of workers continuously employed is broadly stable before the Great Financial Crisis, falls during the recession, and then recovers more quickly for graduates than for non-graduates. The share with two valid wage observations is also stable over time, apart from 2001, when the missing wage observation in 2001Q1 lowers this share. Finally, the share of the wider target panel retained in the final estimation sample changes little over time. Taken together, these

Table B3: Sample characteristics by graduate status across sample restrictions

	All	Graduates	Non-graduates
All target panels with valid wave-1 education			
People	29,677	8,861	20,816
Person-wave observations	148,385	44,305	104,080
Female (%)	50.2	51.8	49.4
Mean age at wave 1	27.49	27.48	27.50
London residence (%)	15.4	22.1	12.0
Part-time at wave 1 (%)	19.4	11.6	23.6
Professional at wave 1 (%)	44.8	74.7	26.3
Routine at wave 1 (%)	38.5	16.4	52.2
Service at wave 1 (%)	16.7	8.9	21.5
Mean observed real log hourly wage	2.416	2.633	2.265
Continuously employed			
People	19,815	7,078	12,737
Person-wave observations	99,075	35,390	63,685
Female (%)	45.7	50.3	42.7
Mean age at wave 1	27.51	27.49	27.52
London residence (%)	13.9	20.9	9.3
Part-time at wave 1 (%)	15.0	7.9	19.6
Professional at wave 1 (%)	46.8	75.8	27.6
Routine at wave 1 (%)	36.6	15.7	50.2
Service at wave 1 (%)	16.7	8.4	22.1
Mean observed real log hourly wage	2.436	2.650	2.284
Two valid wage observations (waves 1 and 5)			
People	13,740	5,316	8,424
Person-wave observations	68,700	26,580	42,120
Female (%)	47.0	49.7	45.0
Mean age at wave 1	27.56	27.52	27.58
London residence (%)	13.3	20.3	8.1
Part-time at wave 1 (%)	15.1	7.2	21.1
Professional at wave 1 (%)	50.5	77.4	30.3
Routine at wave 1 (%)	33.7	14.8	47.9
Service at wave 1 (%)	15.8	7.8	21.9
Mean observed real log hourly wage	2.451	2.658	2.296
Valid other data (final estimation sample)			
People	13,412	5,163	8,249
Person-wave observations	67,060	25,815	41,245
Female (%)	46.9	49.7	44.8
Mean age at wave 1	27.57	27.54	27.59
London residence (%)	13.3	20.3	8.0
Part-time at wave 1 (%)	14.5	6.5	20.5
Professional at wave 1 (%)	50.5	77.6	30.4
Routine at wave 1 (%)	33.7	14.8	47.8
Service at wave 1 (%)	15.8	7.7	21.8
Mean observed real log hourly wage	2.454	2.664	2.297

Note: Counts are unweighted. The All column equals the sum of the graduate and non-graduate columns and corresponds directly to the restriction-specific counts in Table B1. Shares and mean age are measured at wave 1 and weighted by the L-QLFS longitudinal weight. The mean log wage uses survey-weighted observed CPI-deflated wages at waves 1 and 5 before winsorisation. Professional comprises SOC-1-digit major groups 1–3, Routine groups 4–5 and 8–9, and Service groups 6–7.
Source: L-QLFS and author's calculations.

patterns indicate that the restrictions do not generate the observed dynamics through changing sample selection.

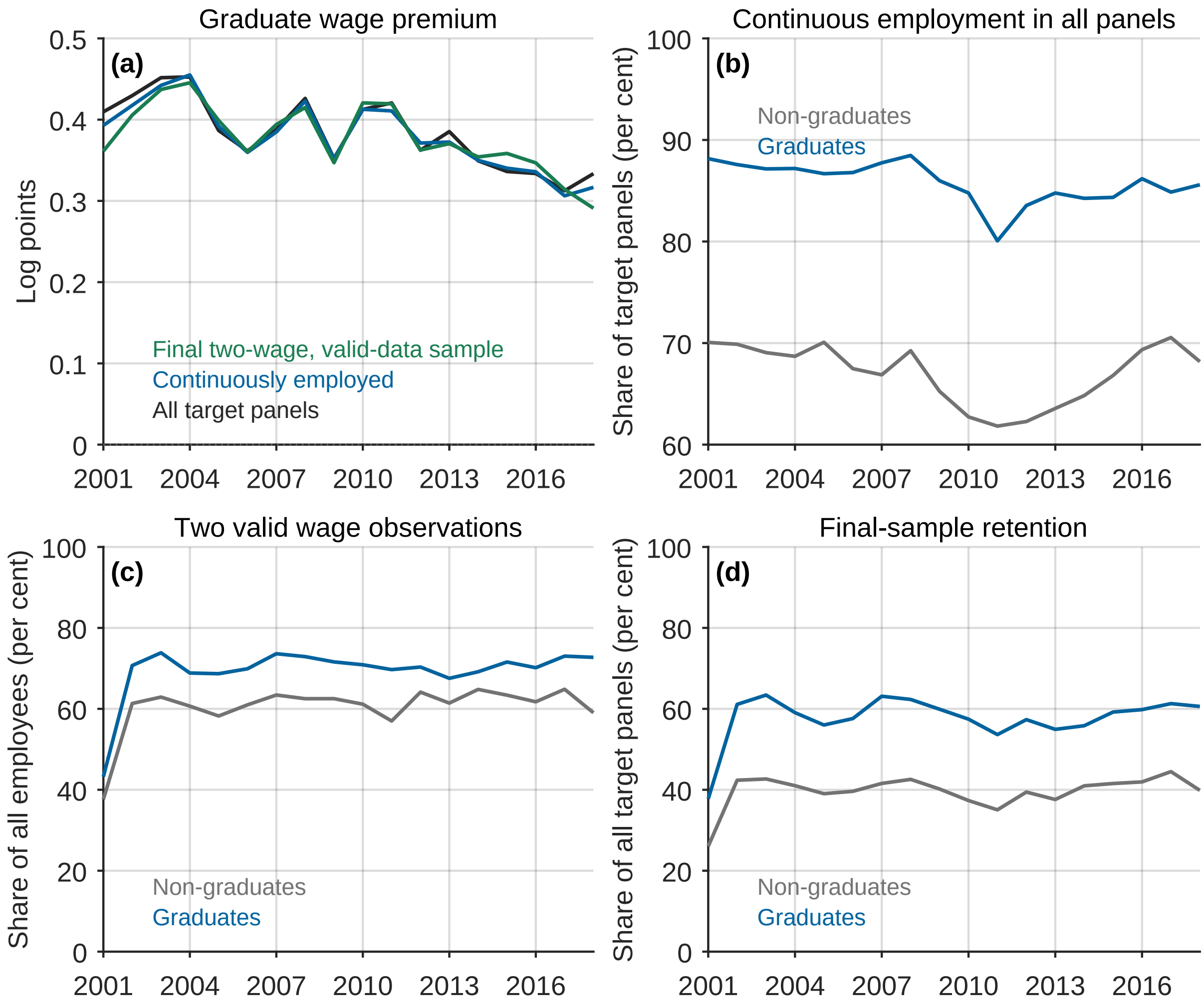


Figure B1: Selection diagnostics for the baseline estimation sample

Note: Panel (a) reports survey-weighted differences in mean observed log hourly wages between graduates and non-graduates. The three lines compare the broad target panels, the continuously employed sample and the final two-wage sample. The continuously employed series includes workers with one or two valid endpoint wages; the final series additionally imposes the two-wage and valid-other-data restrictions. Panel (b) reports the continuous-employment share in the broad target panels. Panel (c) reports the share of employed person-wave observations in the broad target panel that meet the two-valid-wage restriction. Panel (d) reports the share of all person-wave observations in the broad target panel that are retained in the final estimation sample.
Source: L-QLFS and author's calculations.

C Skills and Employment Survey evidence

C.1 Data source, waves and occupation groups

The UK Skills and Employment Survey (SES) is a repeated cross-sectional survey of workers in Britain. It records the tasks performed on the job, the skills those tasks require, and wider features of work organisation, autonomy, training and job quality. The worker-reported measures are comparable across waves and therefore provide a useful description of the task environments associated with the broad occupations in the model. The series dataset used here is available from the UK Data Service SES repository (SN 8589).

I use the 2001, 2006, 2012 and 2017 waves throughout the two exercises below. I restrict the SES sample to workers aged 25–30 in line with the baseline sample. I use the SES occupation classification to construct the same three groups as in the L-QLFS: Professional occupations are SOC-1-digit major groups 1–3, Routine occupations are groups 4–5 and 8–9, and Service occupations are groups 6–7. The regression exercise also reports specifications with the three broad groups, SOC-1-digit major groups and SOC-2-digit groups as occupation fixed effects. All SES means and estimations use the appropriate SES survey weights.

C.2 Task measures

The task measures are respondent-level summaries of ordered SES questions. I recode five-category importance questions so that the highest category is one and the lowest is zero, use the analogous zero-to-one scaling for the 0–4 skill indices, and orient every measure so that larger values denote greater use or importance. The regressions use these measures in their zero-to-one units, without further standardisation. Survey weights enter both the respondent-level regressions and the occupation–wave means used in Figure C1.

I use four measures in both exercises. *Problem-solving skills* is the SES skill-importance index for problem solving, rescaled to the unit interval. *Analytical/abstract skills* is the respondent-level mean of the problem-solving skill item, analysing complex problems in depth, arithmetic, percentage and statistics tasks, and the importance and complexity of computer use. *Quantitative and statistical task use* is the respondent-level mean of the arithmetic, percentage/fraction and advanced mathematics/statistics items. *Complex and self-directed task use* gives equal weight to four domains: problem-solving skills, the analytical/quantitative task-use domain, computer complexity and self-planning skills.

C.3 Differential task use

I first ask whether graduates and non-graduates report different tasks within similar occupations. Table C1 reports weighted respondent-level regressions for the four measures defined above. Specifically, I estimate

$$T_{iw} = \alpha + \beta G_i + X_i'\gamma + \delta_w + \Gamma_{o_i}^{(c)} + \varepsilon_{iw}, \tag{C1}$$

where T_{iw} is one of the zero-to-one task measures for respondent i in SES wave w, G_i is a graduate indicator, and X_i contains sex, age and age squared. The term δ_w captures SES-wave fixed effects, while $\Gamma_{o_i}^{(c)}$ denotes the occupation fixed effects in column c. These are omitted in the first column and then added for the three broad groups used in the model, SOC-1-digit major groups and SOC-2-digit groups. I estimate the equation using SES survey weights. The coefficient β is therefore the graduate–non-graduate difference in the task measure after the controls in the relevant column.

Without occupation controls, graduates report more problem-solving skills (0.066), analytical/abstract skills (0.136), quantitative and statistical task use (0.086) and complex and self-directed task use (0.127). The three broad occupation groups eliminate the problem-solving and quantitative/statistical gaps and reduce the analytical/abstract and complex/self-directed gaps by around two thirds, to 0.045 and 0.042, respectively; the latter two differences remain statistically significant. More detailed occupation controls remove the remaining differences. With SOC-2-digit controls, the corresponding coefficients are -0.009, 0.012, -0.027 and 0.016, and none is statistically distinguishable from zero. Occupation is therefore the primary driver of differential task use between graduates and non-graduates, while the remaining differences within broad occupations are small.

Table C1: Graduate differences in SES task use, ages 25–30

	Graduate coefficient			
Task measure	Wave, sex, age	+ Broad occ.	+ SOC-1-digit	+ SOC-2-digit
Problem-solving skills	0.066***	0.001	−0.008	−0.009
	(0.013)	(0.013)	(0.013)	(0.013)
Analytical/abstract skills	0.136***	0.045***	0.017	0.012
	(0.013)	(0.013)	(0.013)	(0.013)
Quantitative and statistical task use	0.086***	0.006	−0.026	−0.027
	(0.018)	(0.019)	(0.020)	(0.019)
Complex and self-directed task use	0.127***	0.042***	0.021**	0.016
	(0.011)	(0.011)	(0.010)	(0.010)

Note: The table reports the coefficient on a graduate indicator from weighted respondent-level regressions in the Skills and Employment Survey. The sample contains workers aged 25–30 in the 2001, 2006, 2012 and 2017 waves. All specifications include wave fixed effects, sex, age and age squared. The occupation columns add fixed effects for the three broad occupations used in the model, SOC-1-digit major groups, or SOC-2-digit groups. The four task measures are scaled from zero to one, with larger values denoting greater task use or skill importance. Quantitative and statistical task use is the respondent-level mean of the arithmetic, percentage/fraction and advanced mathematics/statistics items. Complex and self-directed task use gives equal weight to problem-solving skills, the analytical/quantitative task-use domain, computer complexity and self-planning skills; no source item is repeated within this index. The number of observations varies slightly with item non-response. Standard errors are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.
Source: Skills and Employment Survey.

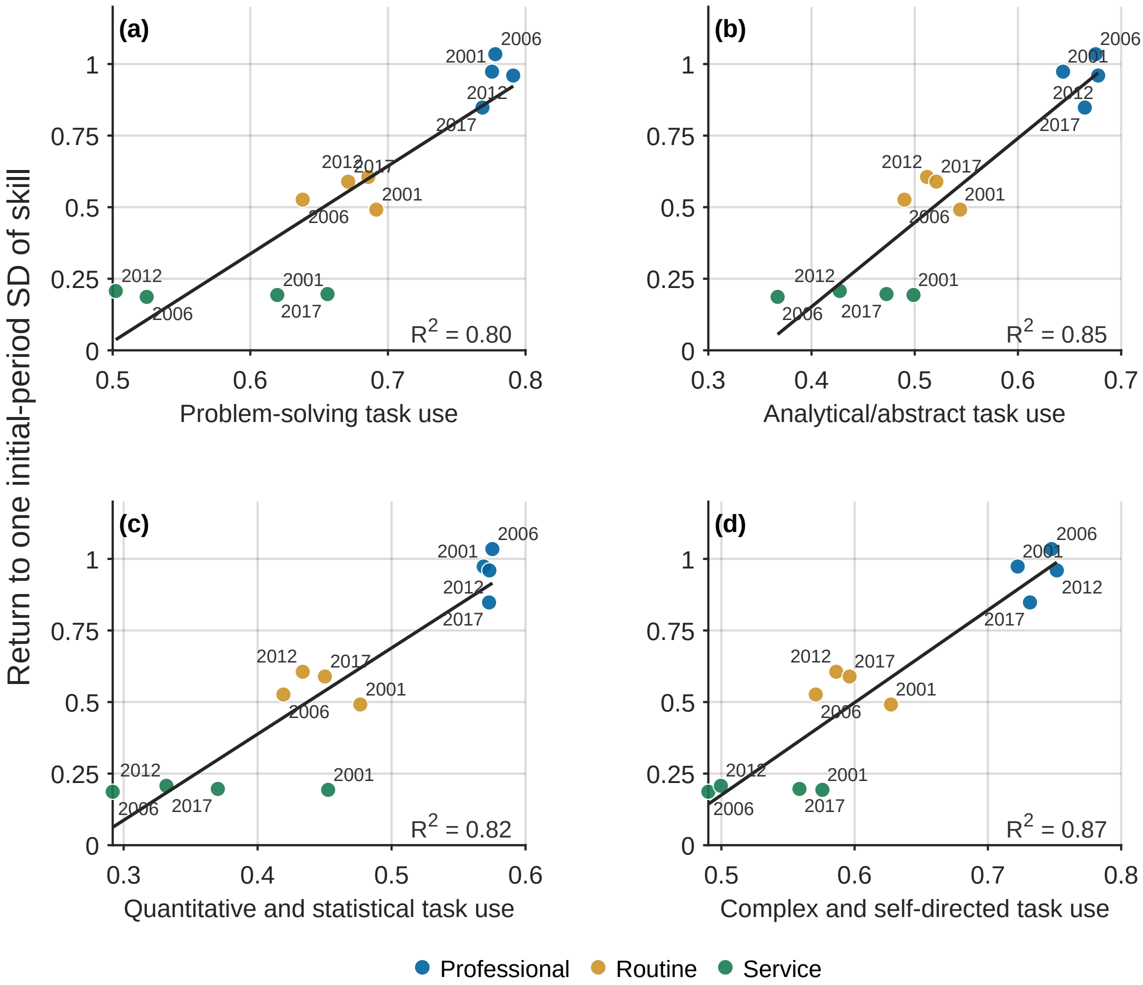
C.4 Relationship between task use and modelled returns to skill

Figure C1 provides a simple level comparison between reported task content and the model's return to skill. I first calculate a weighted task mean for each task, SES wave, broad occupation and degree-status cell. I then average the graduate and non-graduate cell means, leaving one observation for each of the three occupations in each of the four waves.

Additionally, I use the model return to a one-standard-deviation movement in skill, with the standard deviation fixed at its initial-period value. I align the annual model output with the less frequent SES waves by averaging the return uniformly over the survey year and its adjacent years: 2001–02 for the 2001 survey, 2005–07 for 2006, 2011–13 for 2012 and 2016–18 for 2017. I overlay a linear fit and report the squared correlation, or R^2, across the 12 occupation–wave cells.

The level comparison shows a clear positive association between each task measure and the estimated skill-return gradient. The R^2 values are 0.80 for problem-solving skills, 0.85 for analytical/abstract skills, 0.82 for quantitative and statistical task use, and 0.87 for complex and self-directed task use. Professional occupation–wave cells lie towards the high-task, high-return end of each panel, Service cells lie towards the low-return end, and Routine cells are generally between them.

Figure C1: SES task content and model skill-return gradients



Note: The figure plots problem-solving skills, analytical/abstract skills, quantitative and statistical task use, and complex and self-directed task use against the baseline model's occupation-year return to one initial-period standard deviation of skill. The SES sample contains workers aged 25–30 in the 2001, 2006, 2012 and 2017 waves. Each point is a broad occupation–wave mean, averaging the weighted graduate and non-graduate cell values; point labels identify the SES survey year. Model returns are averaged uniformly over the survey year and adjacent years. The black line in each panel is a linear fit across the 12 occupation–wave cells.
Source: Baseline Model Estimates and Skills and Employment Survey.

Overall, the SES evidence supports the role assigned to occupation in the main analysis. Broad occupations capture large differences in task environments, and these differences align with the model's estimated skill-return gradients. At the same time, detailed occupation removes most, but not all, graduate–non-graduate differences in reported task use.

D Wage Measurement and Imputation

The L-QLFS records wages only in waves 1 and 5. The structural model uses the five-wave sequence of occupational choices and wage measurements, so the wages in waves 2–4 must be imputed before estimating the model. I use a Mundlak random-effects–Random-Forest (RE–RF) procedure as the baseline imputation. I then compare the results with an RE–RF procedure without Mundlak means, a best linear unbiased prediction (BLUP), a Mean/within prediction, and an observed-wages-only specification.

The sample construction, wage definition, winsorisation and endpoint-wage restriction are described in Appendix B. All imputation methods below are fitted after those restrictions, leave the observed wages at waves 1 and 5 unchanged, and generate values only for waves 2–4.

D.1 Baseline Mundlak RE–RF imputation

I first estimate a linear mixed wage equation on the observed wage rows. Let y_{it} denote real log hourly wages, let X_{it} contain graduate status, experience, experience squared, year, sex, current broad occupation, part-time status, higher-degree status and macro-region indicators, and let $\overline{Z}_i$ collect person-level means of experience, the Professional and Routine occupation indicators, and part-time status. The Service occupation mean is omitted because the three occupation shares sum to one. The first-stage equation is

$$y_{it} = X_{it}'\beta + \overline{Z}_i'\pi + b_i + u_{it}, \qquad b_i \sim \mathcal{N}(0, \sigma_b^2). \tag{D1}$$

The person means allow the persistent component b_i to be correlated with the individual's average experience, occupational exposure and part-time status. I use the empirical-Bayes estimate $\widehat{b}_i$ from this equation as a person-level wage signal.

In the second stage, I estimate a regression forest on the same observed wage rows. Its predictors are $\widehat{b}_i$, the current values in X_{it} and the person means $\overline{Z}_i$. For a missing wage in waves 2–4, the generated value is

$$\widehat{y}_{it}^{\text{Mundlak RE-RF}} = \widehat{m}_{\text{RF}}\left(\widehat{b}_i, X_{it}, \overline{Z}_i\right). \tag{D2}$$

The forest uses 200 regression trees and a minimum leaf size of one. Its point prediction allows flexible interactions between the person-level wage component, current occupation, time and demographics, while the Mundlak means relax the random-effects orthogonality restriction. I refer to this procedure as the Mundlak RE–RF imputation, and use it for the baseline estimates.

D.2 Alternative imputations and the observed-wage check

The alternative imputation methods keep the sample, observed endpoint wages and current controls fixed, but change how persistent wage heterogeneity and time variation are carried into waves 2–4.

RE–RF without Mundlak means. The first alternative removes $\overline{Z}_i$ from both stages. I estimate a mixed model with a person random intercept and then use a regression forest with $\widehat{b}_i$ and the current controls as predictors:

$$\widehat{y}_{it}^{\text{RE-RF}} = \widehat{m}_{\text{RF}}\left(\widehat{b}_i, X_{it}\right). \tag{D3}$$

This comparison isolates the role of the Mundlak correction.

Best linear unbiased prediction. The BLUP alternative uses the linear mixed model without the forest or the Mundlak means. Its generated value is the conditional mean

$$\widehat{y}_{it}^{\text{BLUP}} = X_{it}'\widehat{\beta} + \widehat{b}_i. \tag{D4}$$

Relative to the forest methods, BLUP imposes an additive linear wage equation. It therefore tests whether the results depend on the nonlinearities and interactions captured by the random forest second stage.

Mean/within prediction. The Mean/within alternative anchors each person's wage profile to the mean of the two observed endpoint wages. I first estimate the pooled linear wage equation $y_{it} = X_{it}'\beta + \nu_{it}$ and obtain the fixed prediction $\widehat{m}_{it} = X_{it}'\widehat{\beta}$ for all five waves. The generated value is

$$\widehat{y}_{it}^{\text{MW}} = \overline{y}_i^{\text{obs}} + \left(\widehat{m}_{it} - \overline{\widehat{m}}_i\right), \qquad \overline{y}_i^{\text{obs}} = \frac{1}{2}\sum_{t\in\{1,5\}} y_{it}, \qquad \overline{\widehat{m}}_i = \frac{1}{5}\sum_{t=1}^{5}\widehat{m}_{it}. \tag{D5}$$

This method preserves the observed person-level wage location and uses the pooled equation only to allocate predicted deviations across the five waves.

Observed wages only. Finally, I re-estimate the model without using any imputation. This check uses only the observed wages at waves 1 and 5, while retaining all five occupational-choice observations. It changes the amount of wage information entering the likelihood without changing the people in the sample, the observed wage endpoints or the occupational histories.

D.3 Fit of imputed wages

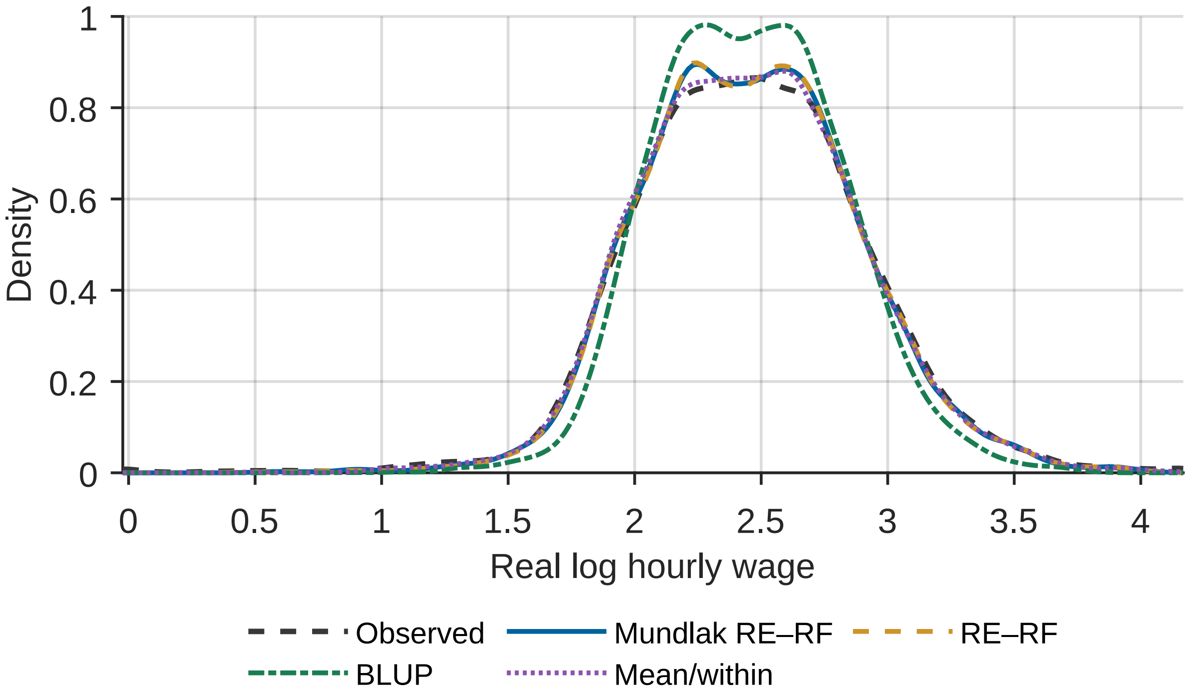
Figure D1 compares the survey-weighted density of the observed endpoint wages with the densities of the generated wages at waves 2–4. The generated means are close, ranging from 2.454 to 2.459 log-wage units. The main difference is dispersion: the standard deviation is 0.4219 under Mundlak RE–RF, 0.4215 under RE–RF and 0.4260 under Mean/within, compared with 0.3611 under BLUP. This pattern is expected because BLUP is a conditional mean, whereas the forest and Mean/within procedures retain more of the observed person-level dispersion.

D.4 Structural robustness to wage measurement

I next re-estimate the structural model under the four alternative wage-measurement specifications. Figures D2–D5 report the same four objects for each specification: mean skill, occupation-specific returns to skill, assignment-rule effects and the Shapley decomposition of the graduate-premium change. The sample, observed endpoint wages, occupational histories and controls are otherwise held fixed. The baseline Mundlak results are reported in the main text and are not reproduced here.

Across the four alternatives, the graduate mean skill declines while the non-graduate mean moves less, so the graduate–non-graduate skill gap narrows. Professional occupations retain the steepest wage schedule, Routine occupations lie in the middle and Service occupations have the lowest return to skill. Each specification also implies a lower model premium in 2018 than the 2001–03 average, and each decomposition assigns a negative contribution to the changing skill distribution. The wage-measurement choice changes the estimated levels and the division of the decline across wage-schedule and choice-factor channels more than the direction of the main mechanisms.

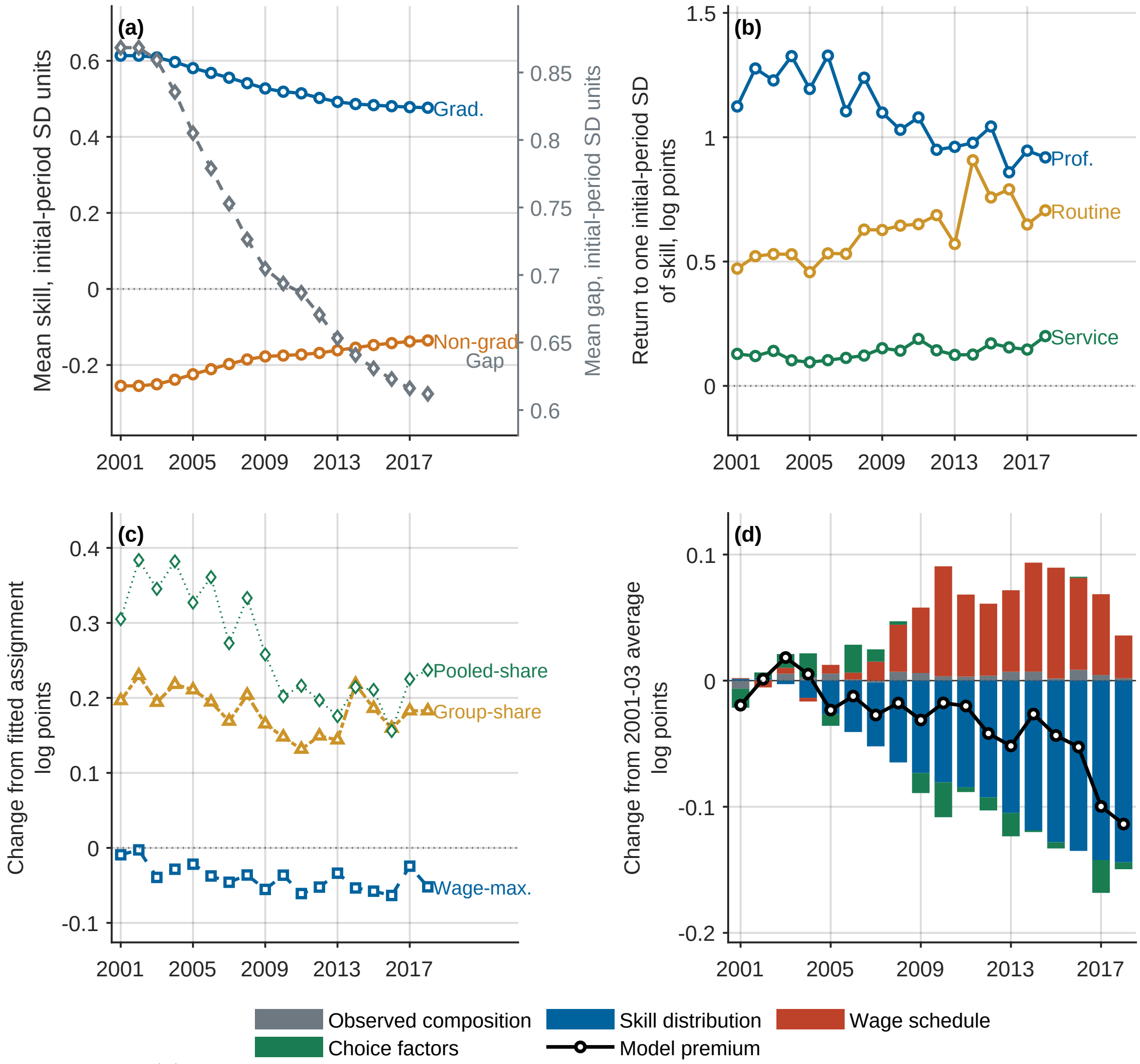
Figure D1: Wage measurements under alternative imputation methods



Note: The figure compares the survey-weighted kernel density of the observed wages at waves 1 and 5 with the densities of generated wages at waves 2–4. All densities use the bandwidth selected for the observed-wage distribution. Observed wages are winsorised at the sample-specific 0.1st and 99.9th percentiles.

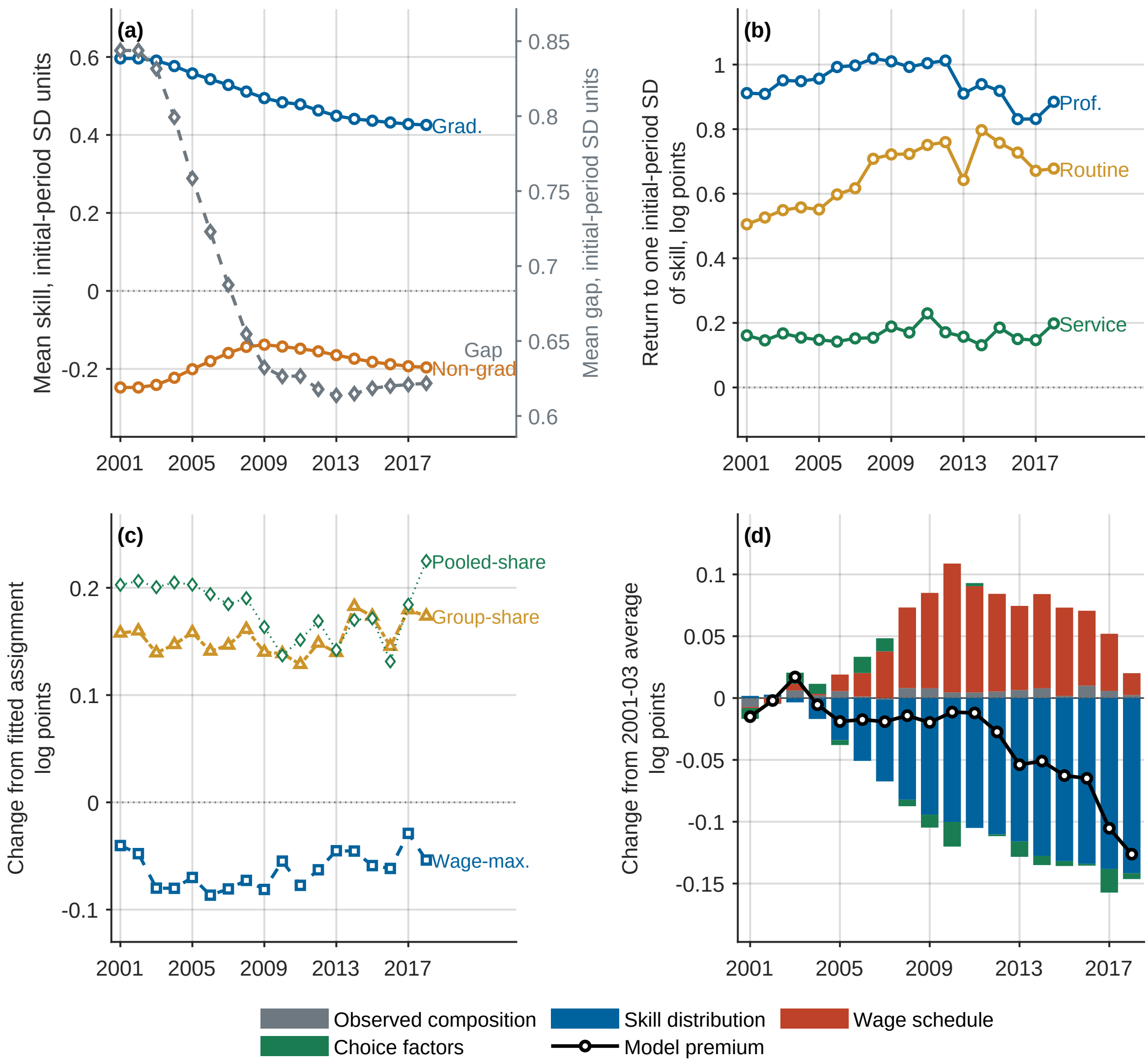
Source: L-QLFS, 2001–18, and author's calculations.

Figure D2: Robustness to RE–RF wage imputation without Mundlak means



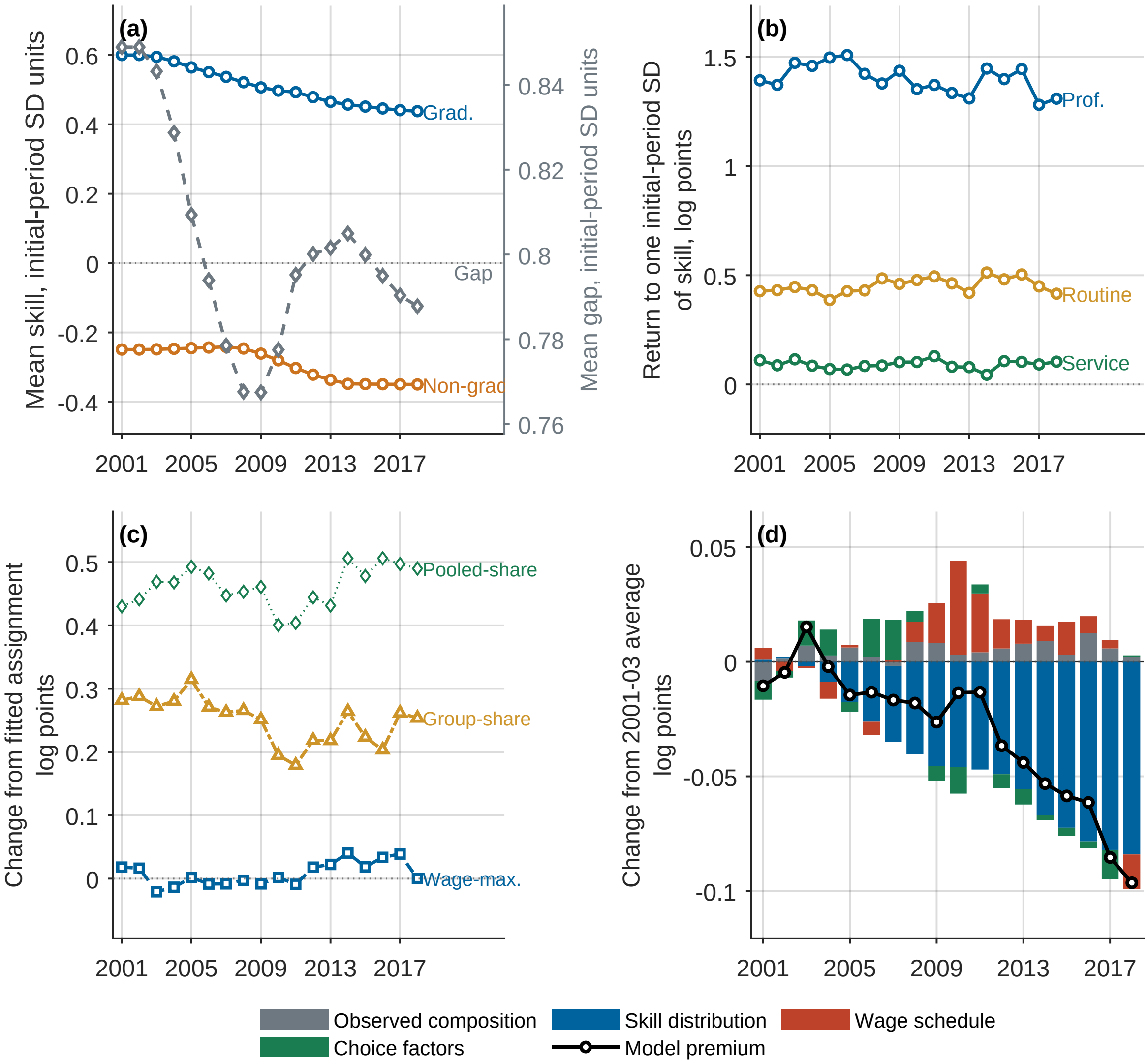
Note: Panel (a) reports graduate and non-graduate mean skill in units of this specification's pooled 2001 skill standard deviation; the dashed series on the right axis is the graduate-minus-non-graduate gap. Panel (b) reports the occupation-specific return to one pooled 2001 skill standard deviation. Panel (c) reports changes in the model graduate premium under wage-maximising, group-share-random and pooled-share-random assignment, relative to fitted assignment. Panel (d) reports the annual Shapley decomposition of the model graduate-premium change, relative to the 2001–03 average, into observed composition, the skill distribution, the wage schedule and choice factors. Wages in waves 2–4 are generated by RE–RF without Mundlak means.
Source: RE–RF Model Estimates and Counterfactual Models.

Figure D3: Robustness to Mean/within wage imputation



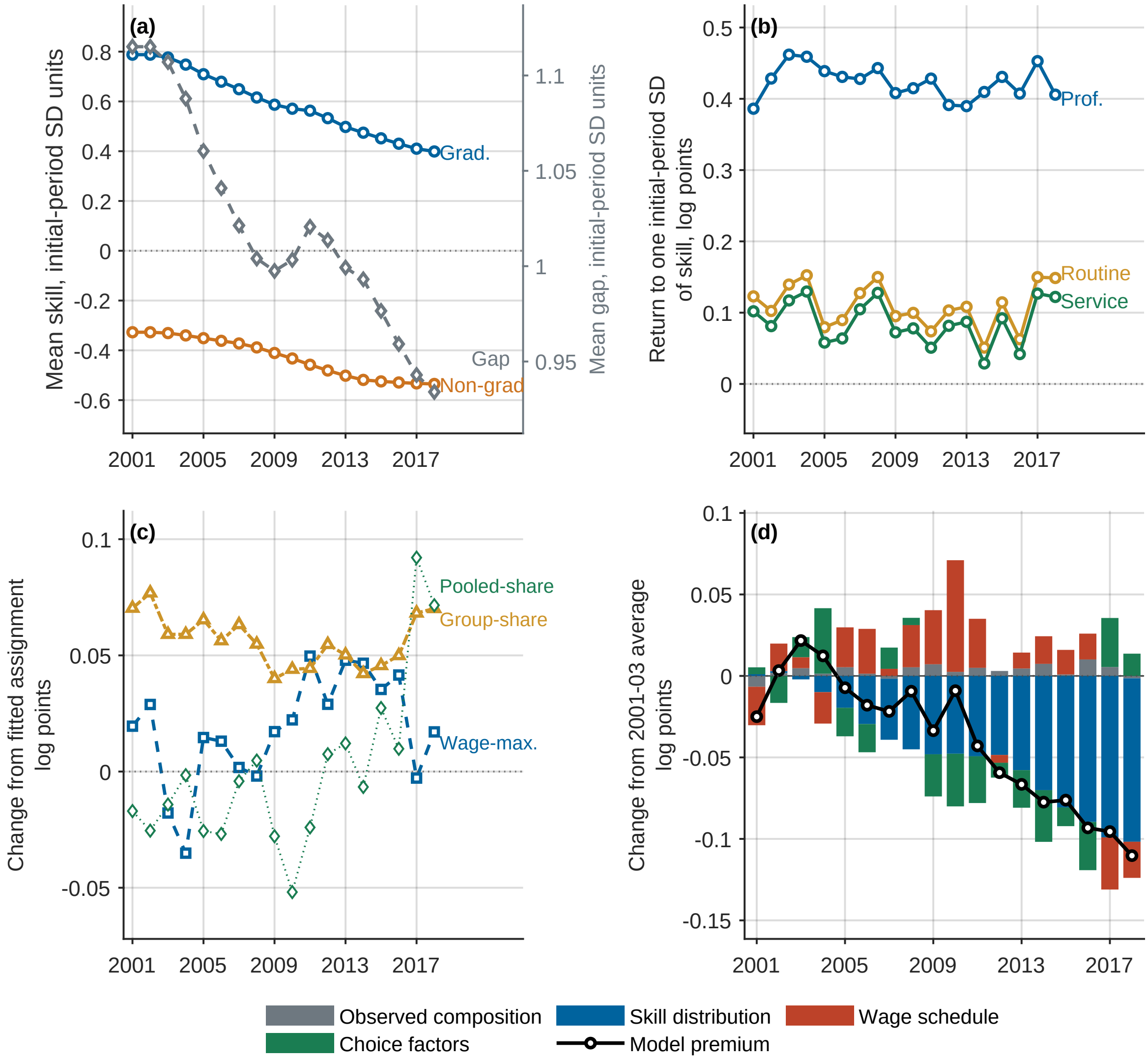
Note: Panel (a) reports graduate and non-graduate mean skill in units of this specification's pooled 2001 skill standard deviation; the dashed series on the right axis is the graduate-minus-non-graduate gap. Panel (b) reports the occupation-specific return to one pooled 2001 skill standard deviation. Panel (c) reports changes in the model graduate premium under wage-maximising, group-share-random and pooled-share-random assignment, relative to fitted assignment. Panel (d) reports the annual Shapley decomposition of the model graduate-premium change, relative to the 2001–03 average, into observed composition, the skill distribution, the wage schedule and choice factors. Wages in waves 2–4 are generated by the Mean/within procedure.
Source: Mean/within Model Estimates and Counterfactual Models.

Figure D4: Robustness to BLUP wage imputation



Note: Panel (a) reports graduate and non-graduate mean skill in units of this specification's pooled 2001 skill standard deviation; the dashed series on the right axis is the graduate-minus-non-graduate gap. Panel (b) reports the occupation-specific return to one pooled 2001 skill standard deviation. Panel (c) reports changes in the model graduate premium under wage-maximising, group-share-random and pooled-share-random assignment, relative to fitted assignment. Panel (d) reports the annual Shapley decomposition of the model graduate-premium change, relative to the 2001–03 average, into observed composition, the skill distribution, the wage schedule and choice factors. Wages in waves 2–4 are generated by the BLUP procedure.
Source: BLUP Model Estimates and Counterfactual Models.

Figure D5: Robustness to using only observed wages



Note: Panel (a) reports graduate and non-graduate mean skill in units of this specification's pooled 2001 skill standard deviation; the dashed series on the right axis is the graduate-minus-non-graduate gap. Panel (b) reports the occupation-specific return to one pooled 2001 skill standard deviation. Panel (c) reports changes in the model graduate premium under wage-maximising, group-share-random and pooled-share-random assignment, relative to fitted assignment. Panel (d) reports the annual Shapley decomposition of the model graduate-premium change, relative to the 2001–03 average, into observed composition, the skill distribution, the wage schedule and choice factors. The likelihood uses observed wages at waves 1 and 5 and occupation choices at all five interviews.
Source: Observed-Wages-Only Model Estimates and Counterfactual Models.

E Numerical Estimation

This section describes the numerical implementation of the model estimation in Section 4.

E.1 Quadrature, EM, and profile likelihood

Let ψ collect the wage, choice and variance parameters, and let π_{gcl} denote the probability mass assigned to skill bin l for degree group g and cohort c. I divide the common support $[0,1]$ into $L = 10$ equal-width bins and use $Q = 10$ Gauss–Legendre nodes, denoted s_{lq}, with within-bin weights a_q. The resulting quadrature rule has $L * Q = 100$ fixed nodes.

For worker i, define the likelihood conditional on being in bin l as

$$A_{il}(\psi) = \sum_{q=1}^{Q} a_q \prod_{t=1}^{5} \Pr(o^*_{it} \mid s_{lq}, g_i, X_{it}; \psi)\, f(w^*_{it} \mid s_{lq}, o^*_{it}, X_{it}; \psi)\,. \tag{E1}$$

I evaluate the likelihood on the log scale. Let d_i be the person-level survey weight and $D = \sum_i d_i$. The quadrature-approximated log likelihood is

$$\ell^Q(\psi, \pi) = \frac{1}{D} \sum_i d_i \log \left[\sum_{l=1}^{L} \pi_{g_i c_i l} A_{il}(\psi) \right]. \tag{E2}$$

For a fixed ψ, EM maximises this criterion over the mass vector in each group–cohort cell. Given a current mass vector π, the E-step posterior probability that worker i belongs to bin l is

$$r_{il} = \frac{\pi_{g_i c_i l} A_{il}(\psi)}{\sum_{k=1}^{L} \pi_{g_i c_i k} A_{ik}(\psi)}. \tag{E3}$$

The survey-weighted M-step is

$$\pi^{\text{new}}_{gcl} = \frac{\sum_{i: g_i = g,\, c_i = c} d_i r_{il}}{\sum_{i: g_i = g,\, c_i = c} d_i}. \tag{E4}$$

Equations (E3) and (E4) define the inner profile step. For each trial outer vector $\psi^{(m)}$, I alternate the E- and M-steps until the EM objective and the maximum change in the bin masses satisfy their convergence tolerances. The converged masses define $\widehat{\pi}(\psi^{(m)})$. The nonlinear routine then updates ψ using the profiled criterion $\ell^Q(\psi, \widehat{\pi}(\psi))$. I repeat the inner profiling and outer update until the outer optimisation criteria are satisfied, so both $\widehat{\psi}$ and its associated mass vector $\widehat{\pi}(\widehat{\psi})$ meet their respective convergence tolerances.

The inner EM problem and the outer profile-likelihood problem are

$$\widehat{\pi}(\psi) = \arg\max_{\pi} \ell^Q(\psi, \pi), \qquad \widehat{\psi} = \arg\max_{\psi} \ell^Q(\psi, \widehat{\pi}(\psi))\,. \tag{E5}$$

The baseline estimate attains $\ell^Q(\widehat{\psi}, \widehat{\pi}(\widehat{\psi})) = -1.82$, or equivalently a survey-weighted mean negative profile log likelihood of 1.82.

E.2 Details of numerical implementation

I begin the optimisation by evaluating the likelihood at 10,000 deterministic scrambled-Halton draws over the starting ranges. I remove points with invalid likelihood evaluations and use the remaining points to construct a likelihood-weighted average starting point, which is passed to local refinement. The refinement uses MATLAB's quasi-Newton `fminunc` routine with forward numerical derivatives. The search is checkpointed in chunks of at most 100 iterations. The local refinement uses optimality, function and step tolerances of 10^{-6}.

For the profiled skill distribution, I parameterise each group–cohort mass vector using bin log ratios and normalise the first log ratio to zero. The EM is initialised from a uniform mass vector; after each update I impose a mass floor of 10^{-12} and renormalise within each group–cohort cell. The numerical tolerances are 10^{-8} for the EM objective-change criterion, 10^{-6} for the maximum EM bin-mass change, and 500 for the maximum number of EM iterations.

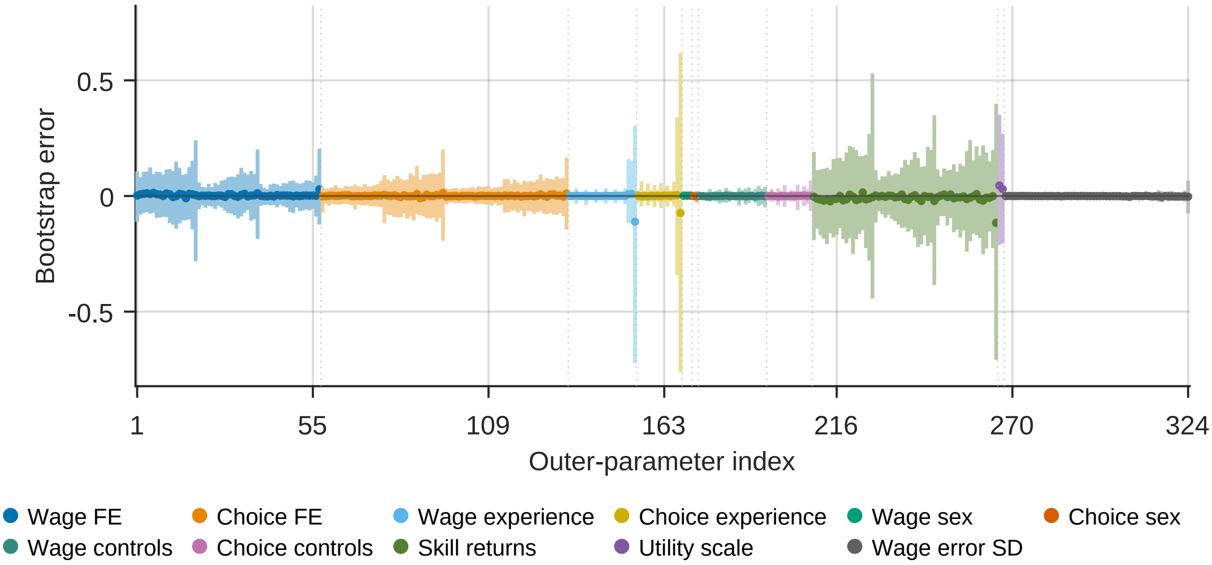
To obtain numerical standard errors for the estimated parameters, I hold the estimated outer vector fixed, reprofile the skill distribution, and evaluate forward-difference derivatives of the profile criterion. I use the inverse of the scaled outer Hessian to obtain model-based numerical standard errors and a numerical delta method for nonlinear transformations and period averages.

E.3 Parametric-bootstrap recovery diagnostic

As a feasibility exercise, I use a parametric bootstrap to check whether the numerical procedure recovers the baseline parameter vector when the model is the true data-generating process. I retain the observed five-wave panel structure, survey weights, covariates and cohort assignments for the 13,412 workers. Across 100 repetitions, I draw one skill value for each worker from the relevant graduate-status and cohort-specific baseline distribution, then simulate occupation choices and wages from the model.

I re-estimate each synthetic panel using the same profile criterion, quadrature rule and EM tolerances. The root mean squared error of the mean bootstrap vector is 0.012, compared with a mean within-replication root mean squared deviation of 0.061; the median absolute mean error is 0.07 bootstrap standard deviations, and zero lies inside the empirical 2.5–97.5 percentile interval for every parameter shown. The bootstrap exercise shows that the numerical procedure recovers the baseline parameter vector with good accuracy and precision.

Figure E1: Parametric-bootstrap recovery of the baseline parameter vector



Note: The figure reports 100 parametric-bootstrap estimates relative to the exact baseline data-generating vector stored in each replication. The horizontal axis follows the parameter ordering of the 324 outer parameters.
Source: Baseline Model Estimates and parametric-bootstrap replications.

E.4 Bootstrap procedure

The parametric exercise above tests whether the numerical procedure recovers the baseline parameters when the model is the data-generating process. I separately use a bootstrap to assess uncertainty in the substantive model outputs while allowing both the worker sample and the generated wage measurements to vary. I resample complete five-wave worker panels with replacement within graduate-status by wave-1-calendar-year strata, renormalise survey weights within each stratum, re-estimate the baseline Mundlak RE–RF wage imputation on each resampled panel, and then re-estimate the structural model from the baseline estimate.

E.5 Numerical estimates

Table E1 gives the full parameter accounting. Tables E2–E4 report the 324 outer parameters, with point estimates and model-based numerical standard-error diagnostics. Table E5 reports the 80 profiled skill-bin masses, which are the economically interpretable representation of the 72 free mass log ratios.

Table E1: Numerical parameter accounting

Object	Count	Maintained structure
Skill-distribution log ratios	72	Nine free bin log ratios for each of two degree groups and four cohorts; profiled by EM.
Occupation-year wage fixed effects, $\eta_{o\tau}$	57	Nineteen years by three occupations; common across degree groups.
Choice-side residuals, $\omega_{o\tau g}$	76	Nineteen years by two non-reference occupations and two degree groups.
Experience effects	35	Seven non-reference experience levels in three wage and two choice equations; common across groups.
Sex effects	5	Three wage and two choice-equation coefficients; common across groups.
Part-time and region effects	35	One part-time indicator and six region indicators in three wage and two choice equations; common across groups.
Occupation-year skill returns, $\lambda_{o\tau}$	57	Nineteen years by three occupations; common across degree groups.
Utility scales, ρ_g	2	One parameter for each degree group.
Occupation-year wage standard deviations, $\phi_{o\tau}$	57	Nineteen years by three occupations; common across degree groups.
Complete parameter vector	396	Includes the 72 profiled skill-distribution log ratios.
Outer profile-likelihood vector	324	All wage, choice and variance parameters; excludes the profiled log ratios.

Note: Professional employment is the reference occupation in the choice equation. The baseline specification uses annual periods over the full estimation window, three broad occupations, two degree groups and four year-turned-25 cohorts.
Source: Baseline Model Estimates.

Table E2: Reported annual wage-equation and wage-error parameters

	Wage fixed effect, η			Skill gradient, λ			Wage-error SD, ϕ		
Year	Professional	Routine	Service	Professional	Routine	Service	Professional	Routine	Service
2001	-0.695	0.945	1.874	5.681	2.875	1.176	0.223	0.177	0.285
	(0.096)	(0.048)	(0.037)	(0.171)	(0.104)	(0.116)	(0.007)	(0.008)	(0.017)
2002	-1.119	0.815	1.889	6.472	3.261	1.238	0.066	0.151	0.257
	(0.057)	(0.037)	(0.026)	(0.101)	(0.072)	(0.076)	(0.002)	(0.005)	(0.010)
2003	-0.962	0.805	1.876	6.223	3.279	1.300	0.152	0.180	0.281
	(0.060)	(0.036)	(0.026)	(0.107)	(0.074)	(0.075)	(0.004)	(0.006)	(0.010)
2004	-1.261	0.777	1.969	6.810	3.388	1.160	0.081	0.135	0.294
	(0.068)	(0.040)	(0.027)	(0.119)	(0.080)	(0.085)	(0.003)	(0.005)	(0.011)
2005	-0.902	0.966	1.984	6.209	2.975	1.093	0.164	0.179	0.280
	(0.056)	(0.036)	(0.026)	(0.100)	(0.073)	(0.080)	(0.004)	(0.005)	(0.011)
2006	-1.343	0.738	1.993	7.041	3.526	1.248	0.064	0.147	0.305
	(0.041)	(0.037)	(0.028)	(0.075)	(0.075)	(0.094)	(0.002)	(0.005)	(0.012)
2007	-0.818	0.787	1.986	6.121	3.355	1.155	0.196	0.164	0.245
	(0.056)	(0.035)	(0.025)	(0.100)	(0.071)	(0.075)	(0.004)	(0.005)	(0.009)
2008	-1.238	0.477	1.909	6.894	3.996	1.262	0.071	0.138	0.263
	(0.054)	(0.040)	(0.028)	(0.092)	(0.077)	(0.087)	(0.002)	(0.005)	(0.010)
2009	-0.918	0.499	1.853	6.294	3.863	1.381	0.129	0.136	0.263
	(0.066)	(0.042)	(0.028)	(0.118)	(0.082)	(0.080)	(0.003)	(0.005)	(0.010)
2010	-0.688	0.498	1.782	5.864	3.764	1.275	0.163	0.157	0.302
	(0.060)	(0.041)	(0.029)	(0.107)	(0.080)	(0.082)	(0.004)	(0.005)	(0.010)
2011	-1.061	0.377	1.751	6.452	3.981	1.575	0.076	0.113	0.348
	(0.061)	(0.042)	(0.031)	(0.112)	(0.079)	(0.088)	(0.002)	(0.004)	(0.011)
2012	-0.557	0.511	1.852	5.526	3.679	1.161	0.210	0.161	0.239
	(0.056)	(0.045)	(0.026)	(0.102)	(0.088)	(0.068)	(0.004)	(0.006)	(0.008)
2013	-0.829	0.480	1.823	5.996	3.683	1.147	0.071	0.129	0.259
	(0.055)	(0.042)	(0.028)	(0.099)	(0.079)	(0.075)	(0.002)	(0.006)	(0.009)
2014	-0.645	0.186	1.820	5.675	4.218	1.053	0.182	0.184	0.263
	(0.062)	(0.051)	(0.030)	(0.112)	(0.099)	(0.093)	(0.004)	(0.007)	(0.009)
2015	-0.816	0.160	1.791	6.053	4.275	1.378	0.068	0.128	0.281
	(0.055)	(0.043)	(0.027)	(0.099)	(0.081)	(0.076)	(0.002)	(0.008)	(0.008)
2016	-0.221	0.380	1.837	4.988	3.930	1.224	0.140	0.081	0.230
	(0.047)	(0.044)	(0.024)	(0.084)	(0.080)	(0.062)	(0.003)	(0.006)	(0.007)
2017	-0.528	0.644	1.911	5.504	3.412	1.078	0.077	0.193	0.166
	(0.054)	(0.041)	(0.020)	(0.095)	(0.076)	(0.046)	(0.002)	(0.005)	(0.005)
2018	-0.465	0.507	1.909	5.388	3.693	1.378	0.235	0.138	0.244
	(0.061)	(0.054)	(0.028)	(0.106)	(0.095)	(0.073)	(0.005)	(0.004)	(0.009)
2019	-2.307	-0.260	1.835	8.777	5.402	1.912	0.060	0.120	0.397
	(0.073)	(0.077)	(0.056)	(0.123)	(0.136)	(0.188)	(0.002)	(0.008)	(0.020)

Note: Each cell reports the point estimate above its model-based numerical standard error in parentheses. The table reports the annual wage fixed effects, skill gradients and wage-error standard deviations estimated in the baseline model. These parameters are common across graduate status. Standard errors use the rank-324 outer information matrix and the numerical delta method described in the text.
Source: Baseline Model Estimates.

Table E3: Reported annual non-pecuniary choice-factor parameters

	Graduates		Non-graduates	
Year	Routine	Service	Routine	Service
2001	-0.356 (0.036)	-0.768 (0.067)	-0.160 (0.027)	-0.739 (0.049)
2002	-0.444 (0.022)	-0.878 (0.044)	-0.266 (0.019)	-0.913 (0.032)
2003	-0.404 (0.023)	-0.776 (0.044)	-0.247 (0.020)	-0.899 (0.034)
2004	-0.421 (0.022)	-0.879 (0.048)	-0.304 (0.020)	-0.981 (0.035)
2005	-0.379 (0.023)	-0.927 (0.051)	-0.228 (0.019)	-0.834 (0.035)
2006	-0.441 (0.021)	-1.001 (0.046)	-0.300 (0.018)	-1.066 (0.035)
2007	-0.304 (0.020)	-0.784 (0.041)	-0.189 (0.018)	-0.887 (0.034)
2008	-0.355 (0.019)	-0.908 (0.046)	-0.215 (0.018)	-0.908 (0.034)
2009	-0.338 (0.020)	-0.776 (0.040)	-0.140 (0.018)	-0.770 (0.033)
2010	-0.270 (0.019)	-0.678 (0.040)	-0.062 (0.018)	-0.620 (0.033)
2011	-0.296 (0.018)	-0.776 (0.039)	-0.113 (0.018)	-0.816 (0.033)
2012	-0.293 (0.019)	-0.615 (0.034)	-0.054 (0.018)	-0.668 (0.031)
2013	-0.221 (0.017)	-0.613 (0.034)	-0.051 (0.017)	-0.651 (0.031)
2014	-0.196 (0.019)	-0.670 (0.042)	0.003 (0.018)	-0.636 (0.035)
2015	-0.158 (0.016)	-0.738 (0.034)	0.034 (0.016)	-0.659 (0.031)
2016	-0.181 (0.016)	-0.584 (0.031)	0.049 (0.016)	-0.577 (0.030)
2017	-0.262 (0.018)	-0.728 (0.032)	-0.072 (0.018)	-0.612 (0.029)
2018	-0.168 (0.018)	-0.828 (0.036)	-0.040 (0.019)	-0.737 (0.033)
2019	-0.395 (0.027)	-1.240 (0.062)	-0.285 (0.029)	-1.276 (0.061)

Note: Each cell reports ω above its model-based numerical standard error in parentheses. The table reports the free annual non-pecuniary choice-factor parameters estimated in the baseline model. Professional employment is the reference occupation, so its non-pecuniary choice factor is normalised to zero and is not an estimated parameter. Standard errors use the rank-324 outer information matrix.
Source: Baseline Model Estimates.

Table E4: Complete control and utility-scale parameters

Control	Wage equation: Professional	Wage equation: Routine	Wage equation: Service	Choice equation: Routine	Choice equation: Service
Experience 1	0.044 (0.003)	0.058 (0.005)	0.017 (0.012)	-0.041 (0.009)	-0.008 (0.019)
Experience 2	0.086 (0.004)	0.075 (0.006)	0.037 (0.013)	-0.030 (0.010)	-0.033 (0.020)
Experience 3	0.125 (0.005)	0.123 (0.007)	0.052 (0.013)	-0.051 (0.010)	-0.017 (0.020)
Experience 4	0.173 (0.005)	0.152 (0.007)	0.070 (0.013)	-0.039 (0.011)	-0.033 (0.020)
Experience 5	0.192 (0.006)	0.162 (0.007)	0.119 (0.013)	-0.046 (0.011)	-0.107 (0.021)
Experience 6	0.214 (0.006)	0.201 (0.008)	0.136 (0.015)	-0.085 (0.013)	-0.109 (0.023)
Experience 7	0.495 (0.098)	0.354 (0.045)	0.166 (0.357)	0.191 (0.180)	0.023 (0.469)
Female	-0.115 (0.005)	-0.122 (0.006)	-0.112 (0.007)	0.010 (0.007)	0.183 (0.012)
Part-time	0.035 (0.006)	-0.074 (0.006)	-0.046 (0.007)	0.175 (0.010)	0.171 (0.013)
Midlands	0.035 (0.009)	0.024 (0.009)	0.006 (0.010)	-0.000 (0.011)	-0.007 (0.017)
London	0.279 (0.010)	0.236 (0.009)	0.147 (0.011)	0.010 (0.011)	0.027 (0.019)
South/East England	0.112 (0.009)	0.107 (0.008)	0.047 (0.009)	-0.048 (0.009)	-0.018 (0.015)
Wales	-0.022 (0.014)	-0.074 (0.013)	-0.078 (0.015)	0.084 (0.016)	0.140 (0.025)
Scotland	0.085 (0.010)	0.049 (0.011)	0.104 (0.012)	0.027 (0.014)	-0.028 (0.022)
Northern Ireland	-0.144 (0.016)	0.027 (0.023)	0.051 (0.025)	-0.132 (0.032)	-0.300 (0.047)

Utility scale, ρ	Graduates	Non-graduates
Estimate	7.366 (0.152)	8.179 (0.143)

Note: Each cell reports the point estimate above its model-based numerical standard error in parentheses. The table contains all 75 common control parameters and both group-specific utility scales. Experience zero, male, full-time work, North England and Professional employment in the choice equation are the relevant reference categories. Standard errors use the rank-324 outer information matrix.
Source: Baseline Model Estimates.

Table E5: Complete profiled skill-bin distributions

	Graduates				Non-graduates			
Skill bin	1996–02	2003–07	2008–11	2012–18	1996–02	2003–07	2008–11	2012–18
1: [0.0,0.1)	0.0095	0.0096	0.0127	0.0000	0.1055	0.0751	0.0656	0.0833
2: [0.1,0.2)	0.0000	0.0198	0.0196	0.0283	0.0178	0.0486	0.0863	0.0730
3: [0.2,0.3)	0.0163	0.0139	0.0179	0.0264	0.0751	0.0410	0.0372	0.0405
4: [0.3,0.4)	0.0342	0.0243	0.0271	0.0326	0.1986	0.1923	0.1568	0.1297
5: [0.4,0.5)	0.0898	0.1131	0.1784	0.1625	0.2879	0.3200	0.3801	0.3715
6: [0.5,0.6)	0.6180	0.6641	0.5958	0.6491	0.2852	0.3016	0.2560	0.2689
7: [0.6,0.7)	0.2253	0.1524	0.1475	0.1012	0.0291	0.0209	0.0181	0.0292
8: [0.7,0.8)	0.0068	0.0027	0.0010	0.0000	0.0008	0.0005	0.0000	0.0038
9: [0.8,0.9)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
10: [0.9,1.0]	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: The table reports the complete profiled distribution of latent skill on the common support $[0,1]$. Within each of the eight graduate-status–cohort cells, the ten masses sum to one. These 80 constrained masses are the economically interpretable representation of the 72 free log ratios in the parameter vector. The first bin is the log-ratio reference. Numerical standard errors are not reported because the 324-by-324 outer Hessian does not contain the profiled mass-information rows.
Source: Baseline Model Estimates.

F Alternative Model Specifications

This appendix examines whether the main results depend on two restrictions in the baseline potential-wage schedule. First, I allow graduates and non-graduates to earn different returns to the same skill within an occupation. Second, I introduce a restricted second skill dimension. Both exercises preserve the central result: the graduate–non-graduate skill gap narrows, and the changing skill distribution remains the largest negative contribution to the decline in the modelled graduate premium.

F.1 Graduate-status-specific returns to skill

The baseline model imposed a common skill return within occupations, but conceivably graduates and non-graduates may receive different tasks or responsibilities within the same broad occupation and therefore earn different returns to the same skill. The SES evidence provided in Appendix C suggests that this margin is likely to be small. I nevertheless relax the common-return restriction directly to investigate the effect of this assumption.

Specifically, I replace the baseline potential-wage schedule with

$$W_{io\tau} = \eta_{o\tau} + \lambda^{g_i}_{o\tau} s_i + X^{w\prime}_{it} \beta^w_o, \tag{F1}$$

where $g_i \in \{G, N\}$ denotes graduate status. The specification therefore allows the slope of the within-occupation wage schedules to differ by graduate status.

Figure F1 reports the results. The graduate–non-graduate mean skill gap falls from 0.692 initial-period standard deviations in 2001–06 to 0.481 in 2013–18, a drop of 0.21 initial standard deviations. This result mirrors the baseline finding of a narrowing skill gap. The group-specific skill returns also mirror the baseline results, with Professional returns falling and Routine returns rising for both groups. Throughout the sample, differences between the returns earned by graduates and non-graduates are small, with non-graduates earning marginally smaller returns across all occupations.

Finally, the premium decomposition is close to the baseline specification. Relative to the 2001–03 average, the model premium is 0.043 log points lower in 2013–18. The changing skill distribution contributes -0.106 log points, while group-specific returns offset 0.048 and observed composition offsets 0.014. The choice-factor contribution is 0.001 log points. Allowing graduates and non-graduates to face different skill prices therefore changes the wage-schedule contribution but leaves the narrowing skill distribution as the dominant negative force.

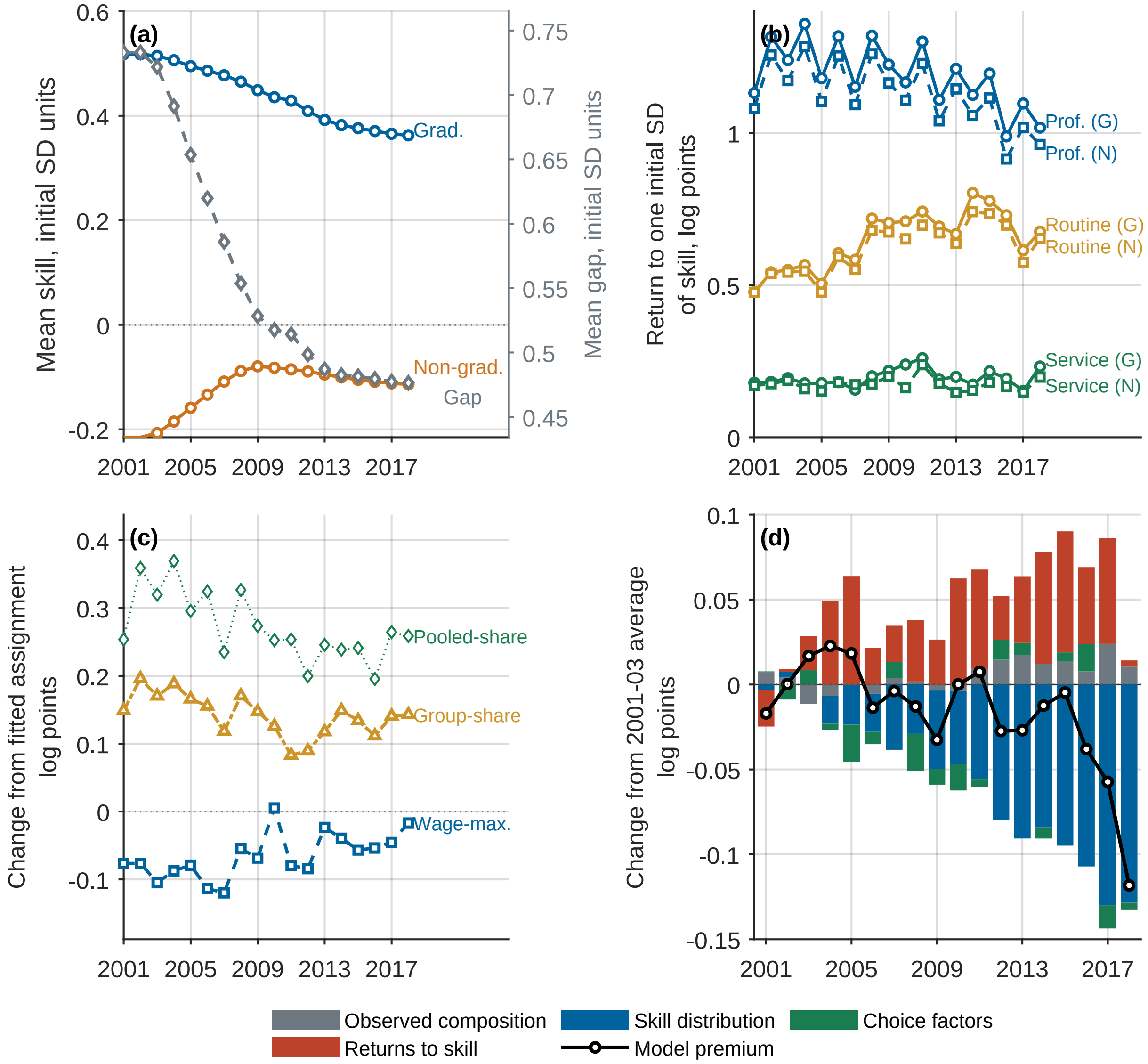


Figure F1: Graduate-status-specific returns to skill

Note: Panel (a) reports graduate and non-graduate mean skill in units of the pooled 2001 skill standard deviation; the dashed series on the right axis is the graduate-minus-non-graduate gap. Panel (b) reports the log-wage return to a one-2001-standard-deviation increase in skill by occupation and graduate status. Solid lines and circular markers denote graduates; dashed lines and square markers denote non-graduates. Panel (c) reports the change in the model graduate premium under three alternative assignment regimes, relative to fitted assignment: wage-maximising assignment, random assignment preserving each education group's occupation shares, and random assignment imposing the pooled occupation shares on both groups. Panel (d) reports the annual Shapley decomposition of the model graduate-premium change, relative to the 2001–03 average, into observed composition, the group–cohort skill distributions, choice factors and group-specific returns to skill.
Source: Group-Specific-Return Model Estimates and Counterfactual Models.

F.2 Restricted two-skill model

The baseline model assumes a singular index of labour market skill. More generally, workers may possess distinct skills, groups may differ in their joint endowments, and occupations may reward those skills differently. Technically, five waves meet the requirement used by Diegert (2024) to identify two skills if all five wage measurements satisfy the repeated-measurement conditions. In the L-QLFS, however, only the first and fifth wages are observed directly; the three intermediate wages are generated by the imputation procedure. I therefore do not use this dimension count to claim identification of a full two-skill model.[27] Instead, I consider a restricted specification that introduces a second skill dimension without estimating its within-group distribution.

The unrestricted two-skill potential-wage schedule is

$$W_{io\tau} = \eta_{o\tau} + \lambda^1_{o\tau} s^1_i + \lambda^2_{o\tau} s^2_i + X^{w\prime}_{it}\beta^w_o. \tag{F2}$$

The two-dimensional factor is invariant to a common nonsingular transformation, so its location, scale and orientation must be normalised. To restrict the model, I first impose that the second skill has no within-group heterogeneity and is instead represented by a group-specific mean. Let $s^2_i = \mu^2_{g_i}$ denote the second skill of worker i in group g_i. Its common wage contribution is absorbed into $\eta_{o\tau}$, so the separately identified second-skill object is the graduate–non-graduate intercept gap

$$\delta_{o\tau} = \lambda^2_{o\tau}(\mu^2_G - \mu^2_N), \tag{F3}$$

rather than the loading and group-mean difference separately. I fix the remaining factor orientation with the single normalisation $\delta_{P1} = 0$, where $\tau = 1$ denotes 2001. The Professional wage intercept is therefore common across graduates and non-graduates in 2001 only. Its graduate–non-graduate gap is unrestricted from 2002 onward, while the Routine and Service gaps are unrestricted in every year. This anchors the second-skill channel in the first Professional wage schedule without setting that channel to zero for Professional work throughout the sample. The resulting restricted two-skill potential-wage schedule is

$$W_{io\tau} = \eta_{o\tau} + \lambda^1_{o\tau} s^1_i + \delta_{o\tau}\mathbf{1}\{g_i = G\} + X^{w\prime}_{it}\beta^w_o. \tag{F4}$$

Figure F2 shows that the primary-skill gap narrows from 0.723 initial-period standard deviations in 2001–06 to 0.470 in 2013–18, a fall of 0.253. The primary-skill returns retain the same ordering as in the baseline. The Professional return falls by 0.105 log points per

[27] Estimating a group–cohort joint distribution of two skills would also add a substantial computational burden.

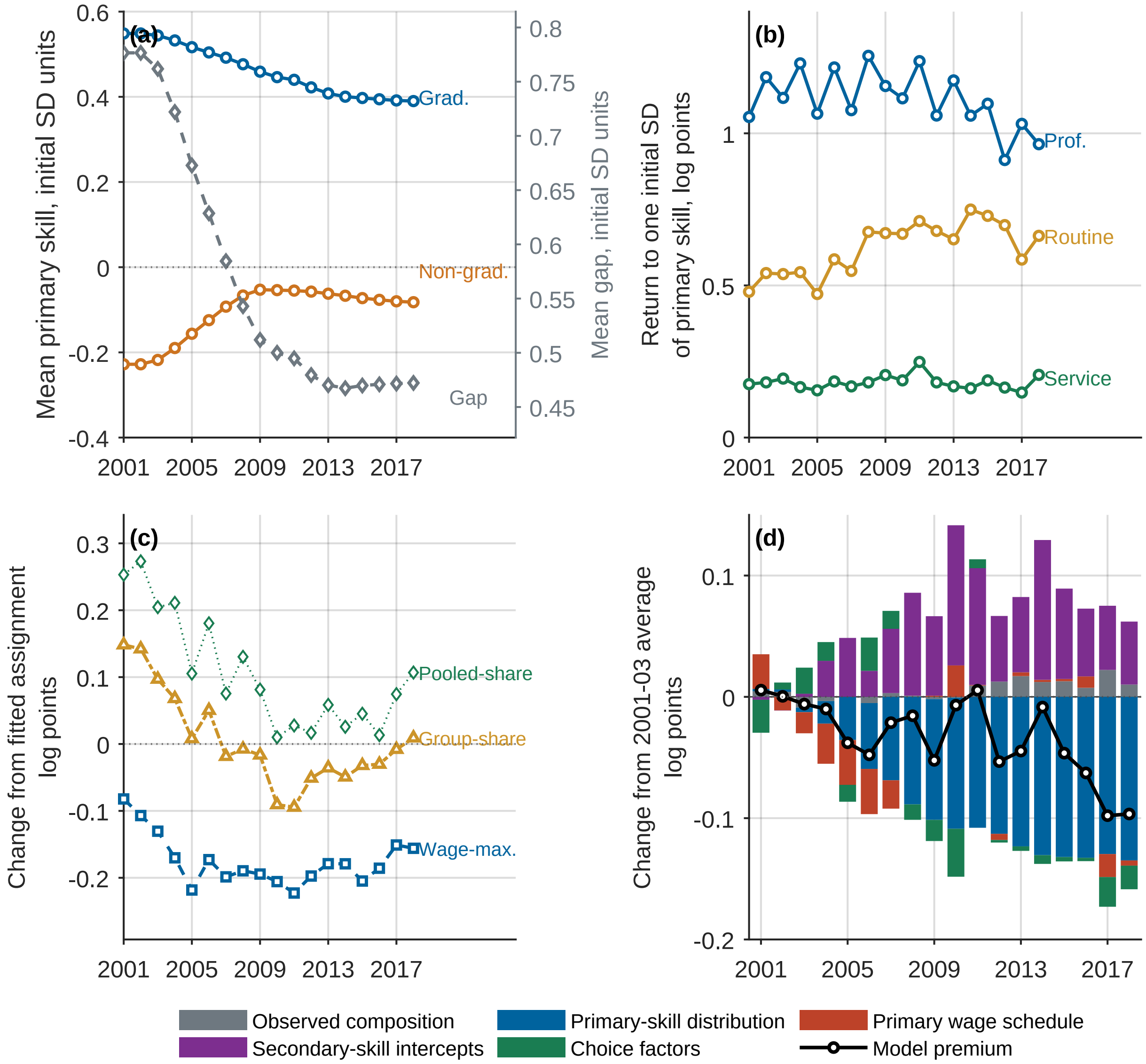


Figure F2: Restricted two-skill model with a 2001 Professional intercept-gap normalisation

Note: Panel (a) reports graduate and non-graduate mean primary skill in units of the pooled 2001 primary-skill standard deviation; the dashed series on the right axis is the graduate-minus-non-graduate gap. Panel (b) reports the common return to a one-2001-standard-deviation increase in primary skill by occupation. Panel (c) reports the change in the model graduate premium under wage-maximising, group-share-random and pooled-share-random assignment, relative to fitted assignment. Panel (d) reports the annual Shapley decomposition of the model graduate-premium change, relative to the 2001–03 average, into observed composition, the primary-skill distribution, the primary wage schedule, the group-specific occupational intercepts representing the second-skill channel, and choice factors. Only the Professional graduate–non-graduate intercept gap in 2001 is normalised to zero; the Professional gap is freely estimated from 2002 onward, and the Routine and Service gaps are freely estimated in every year.
Source: Restricted Two-Skill Model Estimates and Counterfactual Models.

initial primary-skill standard deviation, while the Routine return rises by 0.153 and the Service return is approximately unchanged.

The restricted two-skill model also leads to the same broad conclusion with respect to the graduate premium decomposition. Relative to the 2001–03 average, its model premium is 0.059 log-wage units lower in 2013–18. The primary-skill distribution contributes −0.130 log-wage units, while the primary wage schedule contributes −0.001. The group-specific occupational intercepts representing the second-skill channel offset 0.069 log-wage units and observed composition offsets 0.014, while choice factors contribute −0.010. This second-skill contribution includes the freely estimated Professional intercept gap from 2002 onward, together with the Routine and Service gaps in every year. The narrowing primary-skill distribution nevertheless remains the dominant negative force.

Taken together, the two exercises show that the narrowing skill gap and its central role in the premium decomposition do not depend on either a common graduate-status return to skill or the absence of a restricted second skill dimension.

G Robustness

This appendix reports robustness checks on the baseline model's cohort construction, discretisation of latent skill, utility scale, wage-error process and time aggregation. Each check changes one restriction while retaining the baseline sample, wage imputation, occupational classification and one-skill structure. I then examine alternative age windows.

G.1 Equally spaced cohorts

The baseline model divides workers into four year-25 cohorts with approximately equal weighted mass. This allows the calendar width of a cohort to vary with the distribution of workers across birth years. As a first check, I instead use four equally spaced cohorts, with edges at 1996, 2002, 2008, 2014 and 2020. The remaining model objects continue to vary annually.

Figure G1 shows that this change leaves the main patterns intact. Between 2001–06 and 2013–18, the graduate–non-graduate mean skill gap falls from 0.879 to 0.740 initial-period standard deviations, a decline of 0.140. Over the same periods, the Professional return falls by 0.153 log points per initial skill standard deviation, while the Routine return rises by 0.085.

Relative to the 2001–03 average, the model premium is 0.073 log points lower in 2013–18. The changing skill distribution contributes −0.113 log points, while the wage schedule offsets 0.050 and observed composition offsets 0.005. Choice factors contribute −0.016 log points. The skill-distribution channel therefore remains the largest negative contribution when cohort widths are fixed in calendar time.

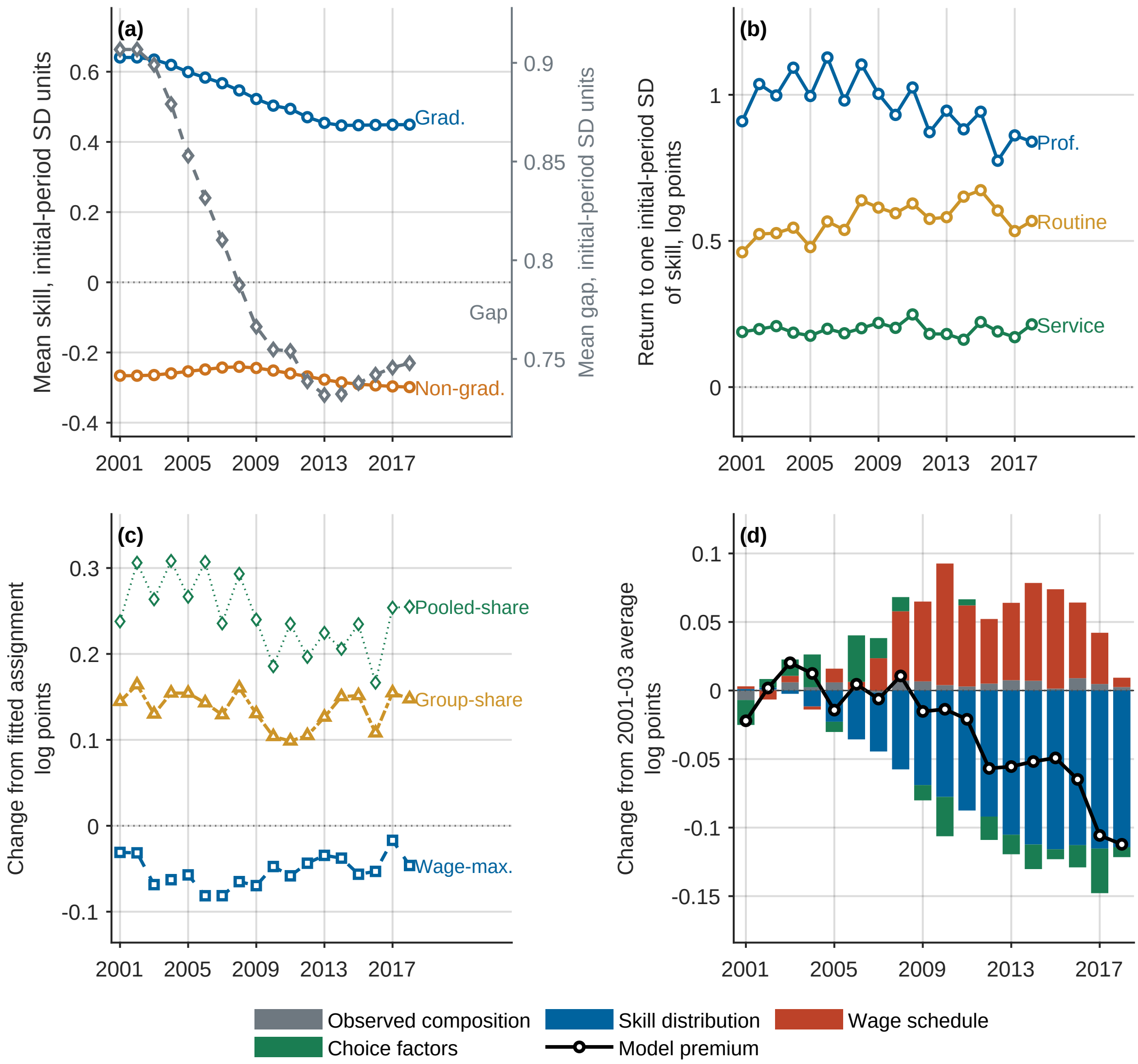


Figure G1: Robustness to equally spaced cohorts

Note: Panel (a) reports graduate and non-graduate mean skill in units of the pooled 2001 skill standard deviation; the dashed series on the right axis is the graduate-minus-non-graduate gap. Panel (b) reports the log-wage return to a one-2001-standard-deviation increase in skill by occupation. Panel (c) reports the change in the model graduate premium under wage-maximising, group-share-random and pooled-share-random assignment, relative to fitted assignment. Panel (d) reports the annual Shapley decomposition of the model graduate-premium change, relative to the 2001–03 average, into observed composition, the skill distribution, the wage schedule and choice factors. Cohorts are defined by equally spaced year-25 intervals with edges at 1996, 2002, 2008, 2014 and 2020.
Source: Equally-Spaced-Cohort Model Estimates and Counterfactual Models.

G.2 Alternative skill-bin count

The baseline model approximates each group–cohort skill distribution using ten equal-width bins on the common skill support. This subsection re-estimates the model using fifteen equal-width bins on the same support, while retaining all remaining model components. This check tests whether the results depend on the discretisation of latent skill.

Figure G2 shows that the graduate–non-graduate mean skill gap falls from 0.848 initial-period standard deviations in 2001–06 to 0.607 in 2013–18, a decline of 0.242. The Professional return falls by 0.045 log points per initial skill standard deviation, while the Routine return rises by 0.260 and the Service return rises by 0.023.

Relative to the 2001–03 average, the model premium is 0.074 log points lower in 2013–18. The changing skill distribution contributes -0.208 log points and choice factors contribute -0.018, while the wage schedule offsets 0.146 and observed composition offsets 0.006. The finer skill grid therefore changes the division of the premium decline between the skill distribution and wage schedule, but preserves the skill distribution as its largest negative component.

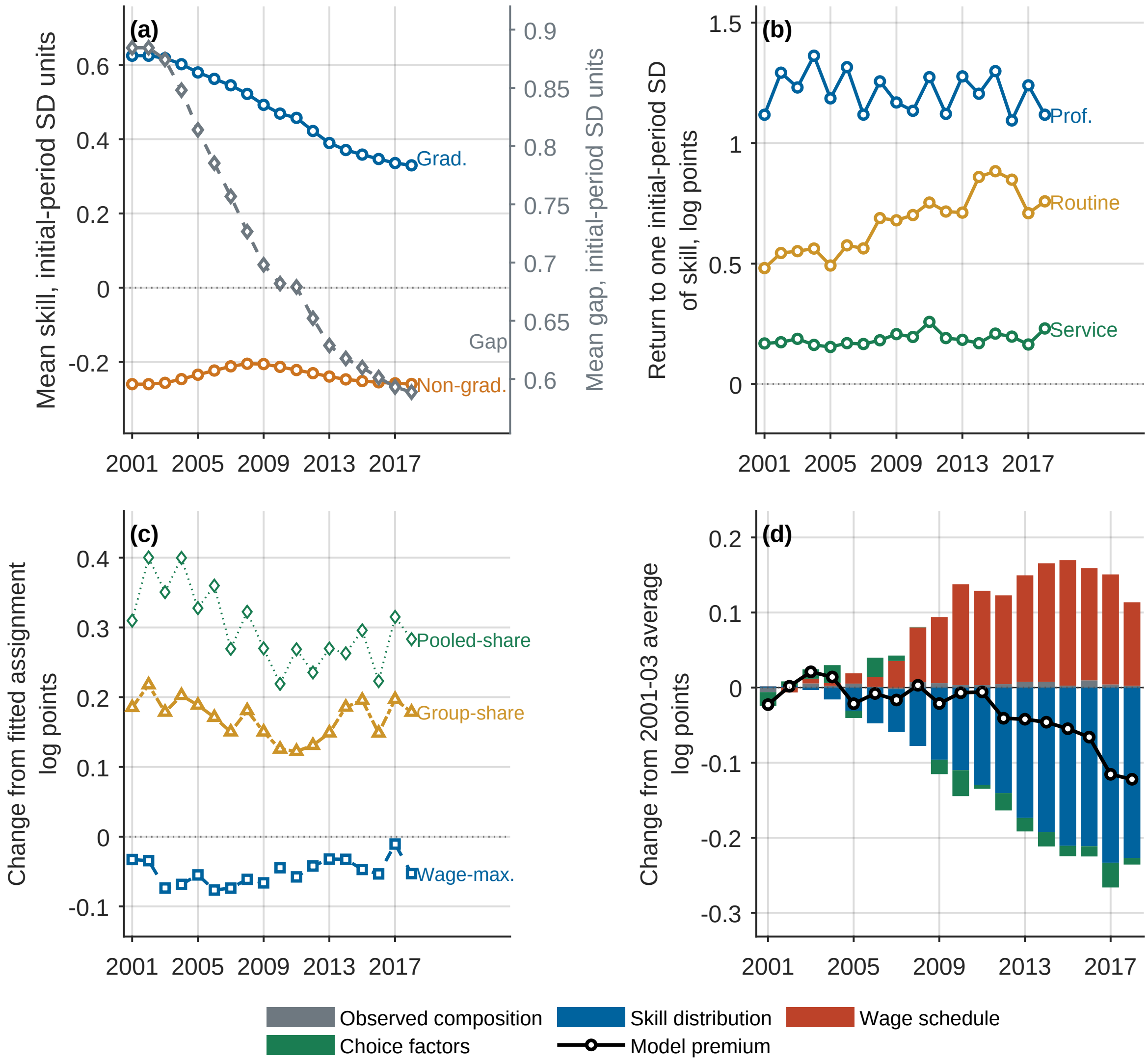


Figure G2: Robustness to fifteen skill bins

Note: Panel (a) reports graduate and non-graduate mean skill in units of the pooled 2001 skill standard deviation; the dashed series on the right axis is the graduate-minus-non-graduate gap. Panel (b) reports the log-wage return to a one-2001-standard-deviation increase in skill by occupation. Panel (c) reports the change in the model graduate premium under wage-maximising, group-share-random and pooled-share-random assignment, relative to fitted assignment. Panel (d) reports the annual Shapley decomposition of the model graduate-premium change, relative to the 2001–03 average, into observed composition, the skill distribution, the wage schedule and choice factors. Each group–cohort skill distribution is approximated using fifteen equal-width bins on the common skill support.
Source: Fifteen-Skill-Bin Model Estimates and Counterfactual Models.

G.3 Common utility scale

The baseline model estimates separate utility scales for graduates and non-graduates. These scales govern how the systematic wage and non-wage components map into occupational choice probabilities. A further check imposes a single common scalar across the two education groups while retaining group-specific skill distributions and the remaining baseline parameters.

Figure G3 shows that the common-scale restriction leaves the principal estimates close to the baseline. The graduate–non-graduate mean skill gap falls from 0.876 initial-period standard deviations in 2001–06 to 0.706 in 2013–18, a decline of 0.169. The Professional return falls by 0.130 log points per initial skill standard deviation and the Routine return rises by 0.105.

Relative to the 2001–03 average, the model premium is 0.076 log points lower in 2013–18. The changing skill distribution contributes −0.119 log points, while the wage schedule offsets 0.054 and observed composition offsets 0.005. Choice factors contribute −0.016 log points. Imposing a common utility scale therefore leaves both the narrowing skill gap and the decomposition of the graduate premium essentially unchanged.

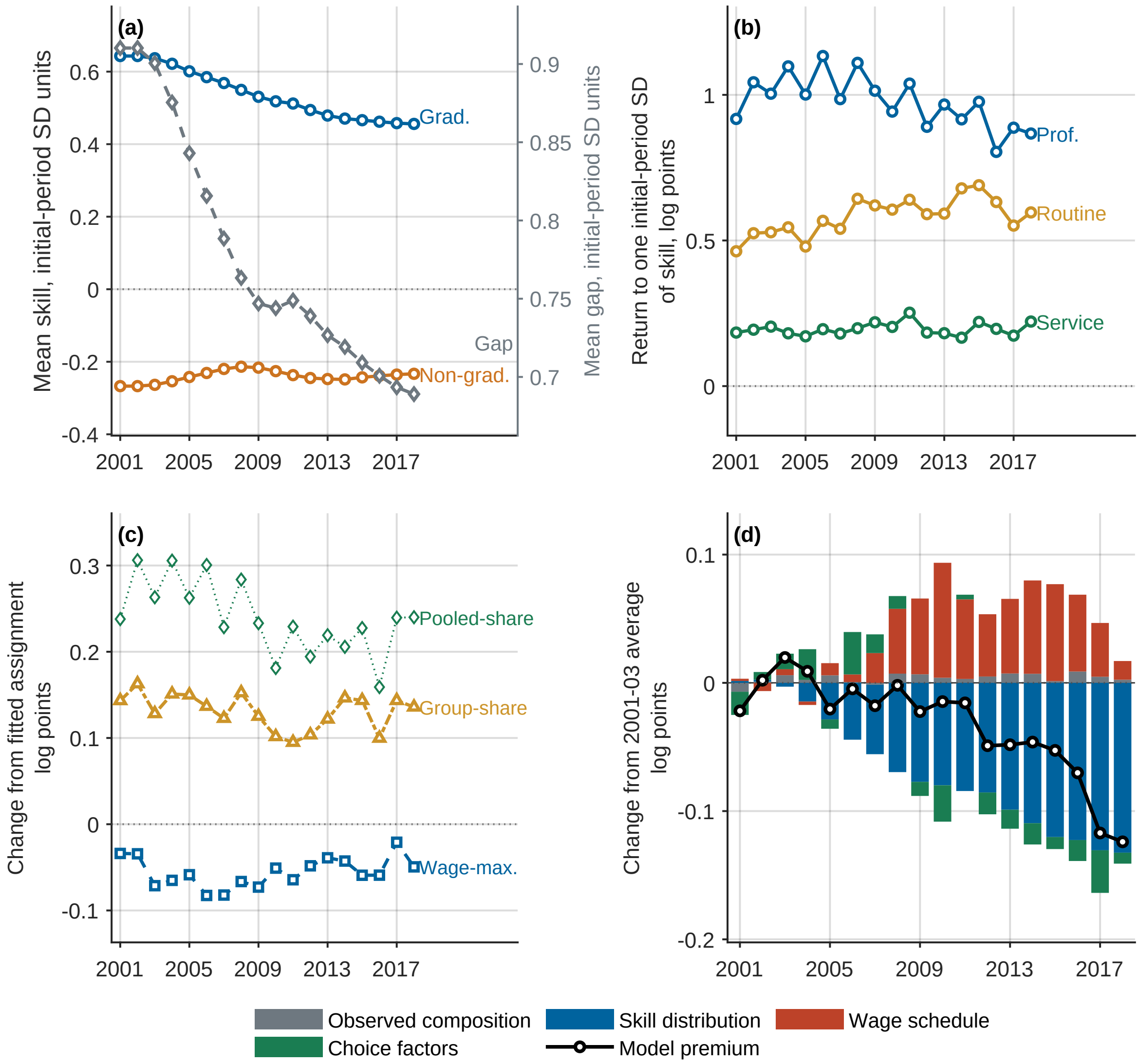


Figure G3: Robustness to a common utility scale

Note: Panel (a) reports graduate and non-graduate mean skill in units of the pooled 2001 skill standard deviation; the dashed series on the right axis is the graduate-minus-non-graduate gap. Panel (b) reports the log-wage return to a one-2001-standard-deviation increase in skill by occupation. Panel (c) reports the change in the model graduate premium under wage-maximising, group-share-random and pooled-share-random assignment, relative to fitted assignment. Panel (d) reports the annual Shapley decomposition of the model graduate-premium change, relative to the 2001–03 average, into observed composition, the skill distribution, the wage schedule and choice factors. A common utility scale is imposed across graduates and non-graduates.
Source: Common-Utility-Scale Model Estimates and Counterfactual Models.

G.4 Separate wage-error variances for observed and imputed wages

The baseline model allows the wage-error variance to vary by occupation and period but imposes a common variance on directly observed and imputed wages. I instead estimate separate occupation–period variances for wages observed in waves 1 and 5 and for the generated wage measurements in waves 2–4. This allows the likelihood to place different weight on the two sources of wage information without changing the wage equation or the choice model.

Figure G4 again shows a narrowing skill gap. Its 2001–06 average is 0.815 initial-period standard deviations, compared with 0.703 in 2013–18, a decline of 0.112. The Professional return falls by 0.021 log points per initial skill standard deviation, while the Routine return rises by 0.029. The return paths are therefore flatter than in the baseline, but their ordering and the direction of the Routine movement are unchanged.

Relative to the 2001–03 average, the model premium is 0.102 log points lower in 2013–18. The skill distribution contributes −0.102 log points and choice factors contribute a further −0.015, while the wage schedule and observed composition offset 0.010 and 0.005, respectively. Allowing the imputed wages to have a different error variance therefore changes the estimated wage-return paths, but not the central role of the changing skill distribution in the premium decline.

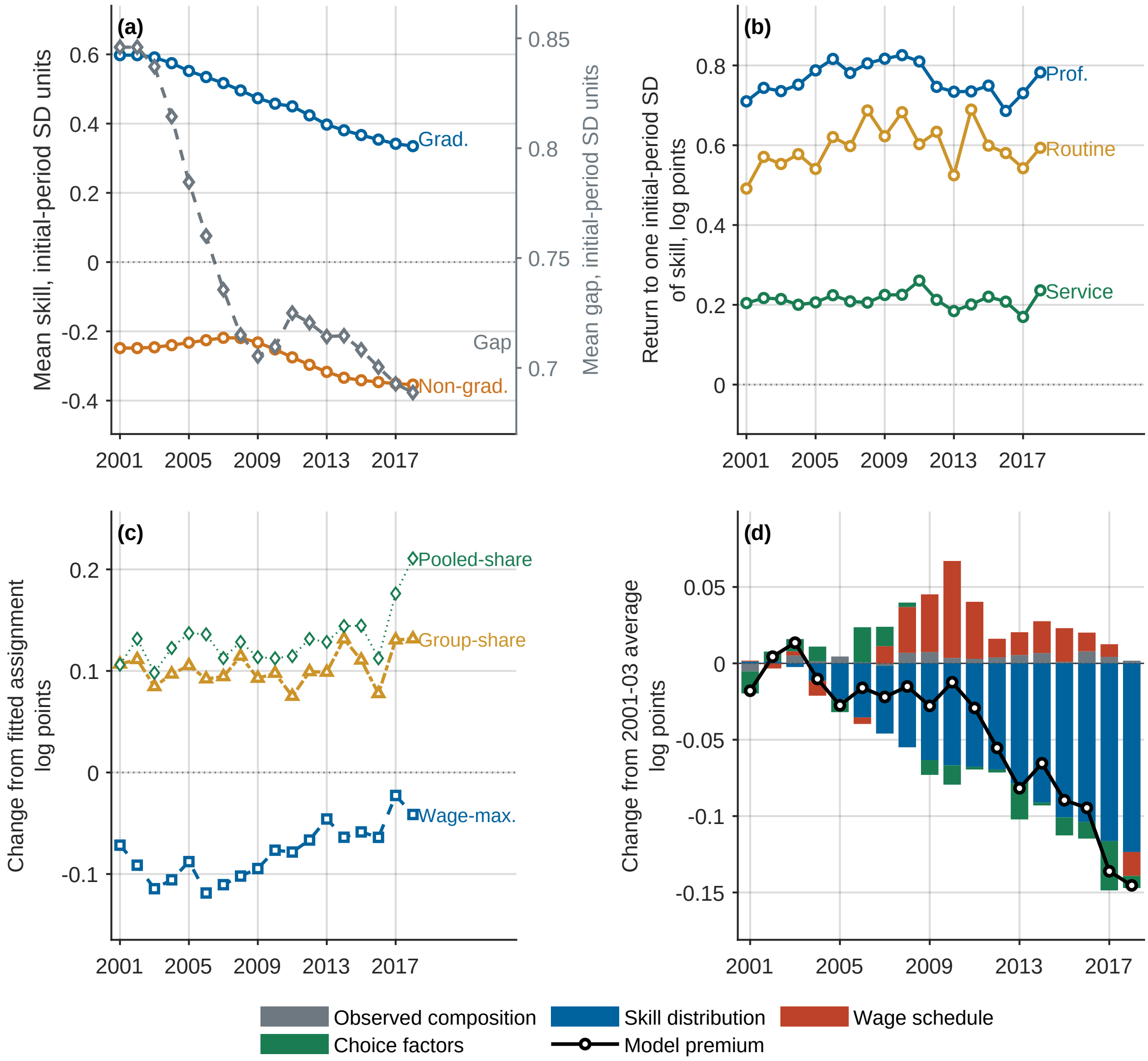


Figure G4: Robustness to separate observed- and imputed-wage error variances

Note: Panel (a) reports graduate and non-graduate mean skill in units of the pooled 2001 skill standard deviation; the dashed series on the right axis is the graduate-minus-non-graduate gap. Panel (b) reports the log-wage return to a one-2001-standard-deviation increase in skill by occupation. Panel (c) reports the change in the model graduate premium under wage-maximising, group-share-random and pooled-share-random assignment, relative to fitted assignment. Panel (d) reports the annual Shapley decomposition of the model graduate-premium change, relative to the 2001–03 average, into observed composition, the skill distribution, the wage schedule and choice factors. The wage-error standard deviations vary by occupation and year and are estimated separately for wages observed in waves 1 and 5 and wages imputed in waves 2–4.
Source: Separate-Variance Model Estimates and Counterfactual Models.

G.5 Three-year parameter blocks

The baseline model allows wage and choice parameters to vary annually. Finally, I reduce this time variation by holding period-specific parameters fixed within three-year blocks. The sample is divided into 2001–03, 2004–06, 2007–09, 2010–12 and 2013–15 blocks; 2019 is combined with 2016–18.

Figure G5 reports the block-level results. Comparing the average of the first two blocks with the average of the final two, the graduate–non-graduate mean skill gap falls from 0.921 to 0.846 initial-period standard deviations, a decline of 0.075. The Professional return falls by 0.134 log points per initial-period skill standard deviation, while the Routine return rises by 0.037. Aggregating the annual parameters therefore smooths the paths, but preserves the narrowing skill gap and the relative movements in Professional and Routine returns.

Relative to the initial 2001–03 block, the model premium is 0.084 log points lower on average in the final two blocks. The changing skill distribution contributes -0.098 log points and choice factors contribute -0.021. Wage schedules offset 0.030 log points and observed composition offsets 0.005. The block-level specification therefore delivers the same decomposition ordering as the annual model.

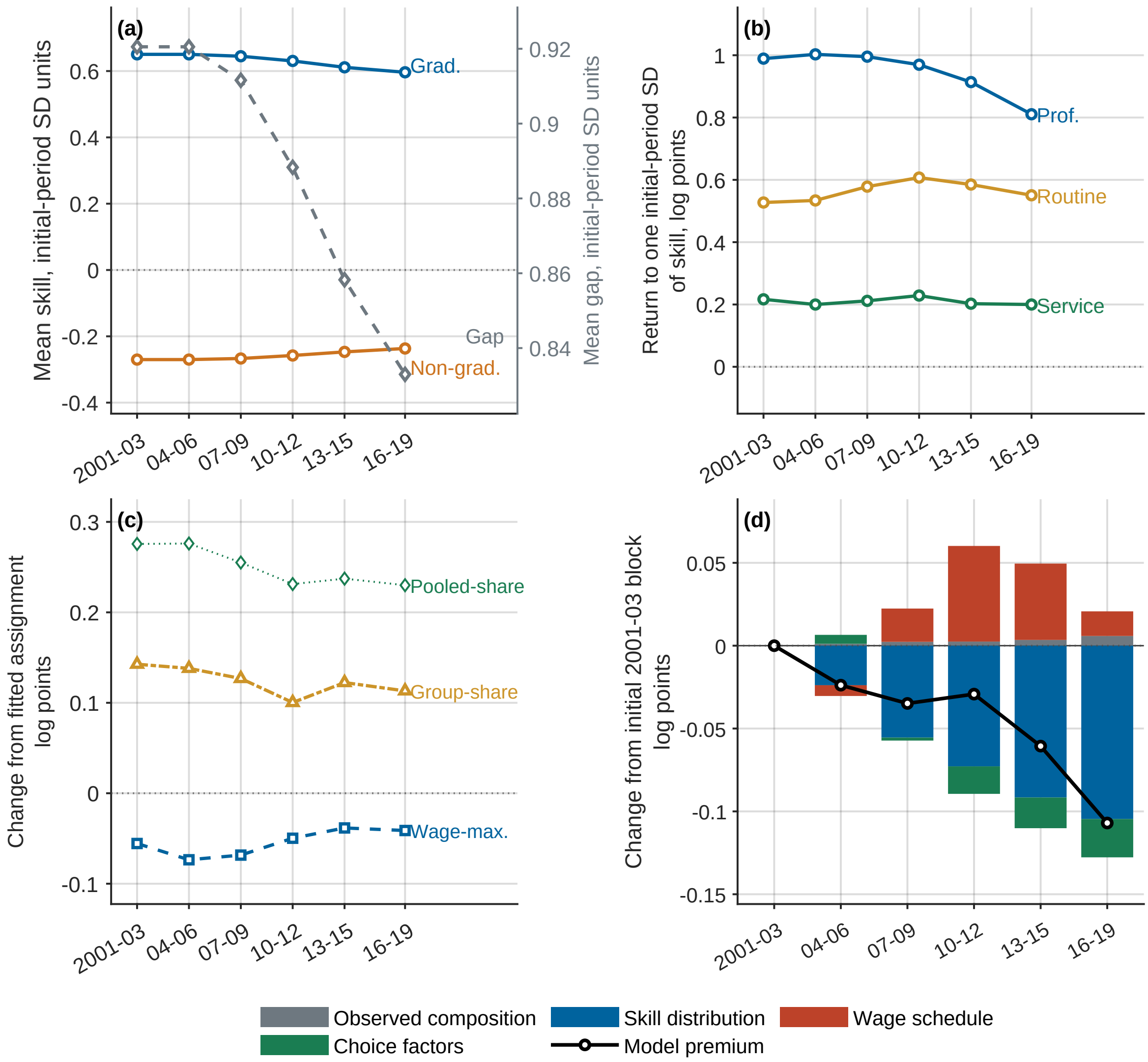


Figure G5: Robustness to three-year parameter blocks

Note: Panel (a) reports graduate and non-graduate mean skill in units of the pooled 2001–03 skill standard deviation; the dashed series on the right axis is the graduate-minus-non-graduate gap. Panel (b) reports the log-wage return to a one-initial-period-standard-deviation increase in skill by occupation. Panel (c) reports the change in the model graduate premium under wage-maximising, group-share-random and pooled-share-random assignment, relative to fitted assignment. Panel (d) reports the Shapley decomposition of the block-level change in the model graduate premium, relative to the initial 2001–03 block, into observed composition, the skill distribution, the wage schedule and choice factors. Parameters are constant within the displayed blocks; the final block covers 2016–19.

Source: Three-Year-Block Model Estimates and Counterfactual Models.

G.6 Alternative Age Windows

The baseline sample includes workers aged 25–30 at wave 1. This subsection re-estimates the model for two alternative wave-1 age windows: ages 23–27 and 28–32. The alternative samples contain 8,353 and 14,020 workers, respectively. Each exercise otherwise retains the baseline two-observed-wage Mundlak specification, occupational classification and sample period. The figures follow the layout of Appendix F: panel (a) reports the estimated skill distributions through their group means and the graduate–non-graduate gap, panel (b) reports occupation-specific skill returns, panel (c) reports the effect of alternative occupational-assignment regimes on the model graduate premium, and panel (d) reports the Shapley decomposition of the model premium.

G.6.1 Ages 23–27

Figure G6 reports the youngest age-window check. Between 2001–06 and 2013–18, the graduate–non-graduate mean skill gap falls from 0.793 to 0.578 initial-period standard deviations, a decline of 0.216. Professional occupations retain the highest return to skill and Service occupations the lowest. The Professional return falls by 0.123 log points per initial skill standard deviation, while the Routine return rises by 0.253.

Relative to the 2001–03 average, the model premium is 0.008 log points lower in 2013–18. The decomposition attributes −0.097 log points to the changing skill distribution and −0.014 to choice factors. Wage schedules offset these forces by 0.097 log points, while observed composition contributes 0.006. The skill-distribution channel therefore remains the largest negative contribution in the youngest sample, although the other channels jointly offset most of it over the final period.

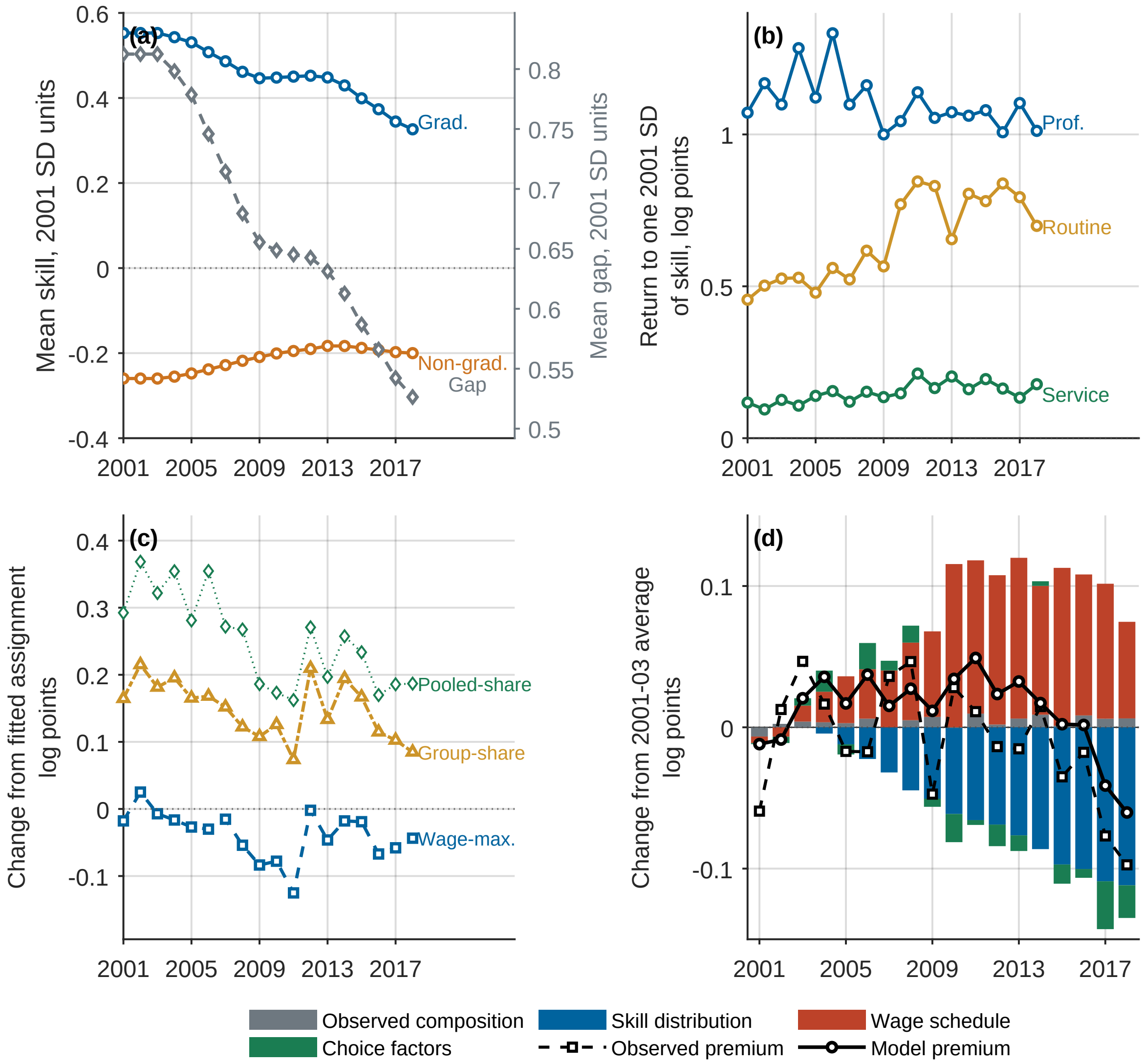


Figure G6: Age-window robustness: ages 23–27 at wave 1

Note: Panel (a) reports graduate and non-graduate mean skill in units of the pooled 2001 skill standard deviation; the dashed series on the right axis is the graduate-minus-non-graduate gap. Panel (b) reports the log-wage return to a one-2001-standard-deviation increase in skill by occupation. Panel (c) reports the change in the model graduate premium under wage-maximising, group-share-random and pooled-share-random assignment, relative to fitted assignment. Panel (d) reports the annual Shapley decomposition of the model graduate-premium change into observed composition, the skill distribution, the wage schedule and choice factors. The solid black line reports the model-premium change and the dashed black line reports the observed graduate-premium change, each relative to its corresponding 2001–03 average. The sample contains workers aged 23–27 at wave 1. *Source:* Age-Sample Model Estimates and Counterfactual Models.

G.6.2 Ages 28–32

Figure G7 reports the older age-window check. The mean skill gap falls from 0.879 initial-period standard deviations in 2001–06 to 0.694 in 2013–18, a decline of 0.185. Professional, Routine and Service occupations preserve the same ordering of skill returns. The Professional return falls by 0.034 log points per initial skill standard deviation, while the Routine return rises by 0.093.

Relative to the 2001–03 average, the model premium is 0.086 log points lower in 2013–18. The changing skill distribution contributes −0.205 log points, while wage schedules offset 0.133 and observed composition offsets 0.006. Choice factors contribute −0.020 log points. Across both alternative age windows, the graduate–non-graduate skill gap narrows and the skill-distribution component supplies the largest negative contribution to the change in the model premium.

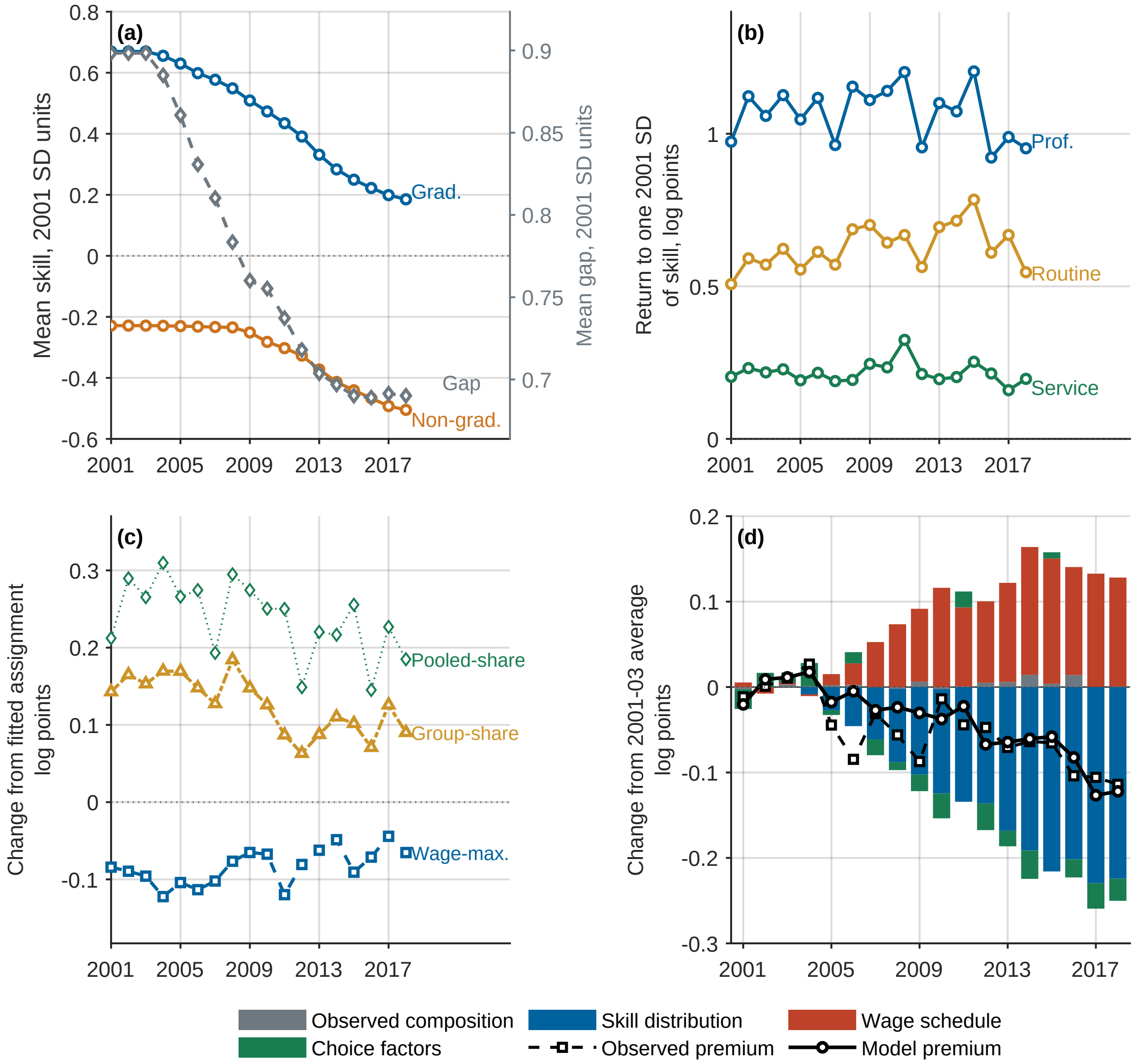


Figure G7: Age-window robustness: ages 28–32 at wave 1

Note: Panel (a) reports graduate and non-graduate mean skill in units of the pooled 2001 skill standard deviation; the dashed series on the right axis is the graduate-minus-non-graduate gap. Panel (b) reports the log-wage return to a one-2001-standard-deviation increase in skill by occupation. Panel (c) reports the change in the model graduate premium under wage-maximising, group-share-random and pooled-share-random assignment, relative to fitted assignment. Panel (d) reports the annual Shapley decomposition of the model graduate-premium change into observed composition, the skill distribution, the wage schedule and choice factors. The solid black line reports the model-premium change and the dashed black line reports the observed graduate-premium change, each relative to its corresponding 2001–03 average. The sample contains workers aged 28–32 at wave 1.
Source: Age-Sample Model Estimates and Counterfactual Models.