Skill-Based Labour Market Segmentation and Mismatch Dynamics

Peijie Ju

University of Bristol, School of Economics

14 November 2026 · Download the latest version

Abstract

The ability of unemployed workers to fill available vacancies depends partly on how closely their skills match the requirements of those jobs. Yet conventional mismatch measures treat administrative occupation categories as separate search markets. I examine how replacing these boundaries with skill-based groups changes measured mismatch in the UK during 2016–2022. I combine online job advertisements collected by Adzuna, worker employment histories from Understanding Society, and ESCO occupational skill requirements. Over the full sample, the skill-cluster series also has an upward descriptive trend, with increases visible before the pandemic, while the SOC major-group series trends downwards. Around the COVID-19 pandemic, the skill-based occupation clusters reveal larger movements in productivity-weighted mismatch than the administrative SOC occupation groups. Mismatch estimates based on more detailed occupational categories also increase, suggesting that both the level of aggregation and the composition of search markets matter for assessing labour market reallocation. Holding vacancy shares fixed changes the detailed occupational path more than holding jobseeker shares fixed. These findings are consistent with skill-related allocation difficulties extending beyond the immediate pandemic disruption. They make market definition relevant to the UK productivity debate and motivate attention to continuing retraining and worker reallocation.

Keywords: labour market segmentation; skills; mismatch; worker mobility; COVID-19.

1 Introduction

A well-functioning labour market connects people to available jobs in which their skills can match with the occupational requirements. When available workers and vacancies are concentrated in different parts of the economy, recruitment can remain difficult despite unemployment. Such mismatch can reduce job finding and the total productivity (Şahin et al., 2014; Patterson et al., 2016). Measuring it accurately matters for diagnosing whether weak labour market performance reflects a shortage of jobs overall or difficulty reallocating workers towards the jobs that exist. These diagnoses have different implications for job-search assistance, training and policies supporting labour demand.

An essential measurement choice is where to draw the boundaries of a labour market. Conventional mismatch indices use industries or occupational classifications as separate search markets. But a change of occupation need not require a completely new set of skills. Evidence on tasks and worker mobility shows that skills can transfer across occupational boundaries (Gathmann and Schönberg, 2010; Poletaev and Robinson, 2008; Carrillo-Tudela et al., 2016). Consider a fall in hiring in one occupation alongside rising demand in another. If their skill requirements overlap, workers may have a route between them that an occupational boundary treats as a reallocation across markets. Conversely, a broad occupational group may combine jobs requiring quite different skills and conceal an imbalance within that group. Occupational aggregation can therefore hide or create apparent barriers.

This paper takes a new approach to defining labour market segments for the mismatch measurement. I use information on occupational skill to address a limitation of conventional mismatch measures that use administrative occupational categories to define search markets. Broad categories can conceal imbalances between jobs that require different skills, while separate categories can divide jobs between which workers can transfer some of their skills (Gathmann and Schönberg, 2010). I develop two approaches to grouping occupations to calculate a productivity-weighted mismatch. The first groups occupations directly by the similarity of their required skill bundles. The second first groups skill concepts by their co-occurrence in workers' occupational histories, represents each occupation by its shares across these skill domains, and groups occupations using those shares. The two approaches use different information: the first captures similarity in stated job requirements, while the second captures patterns of skill exposure in observed careers. They provide complementary descriptions of occupational relatedness. Both can be applied to other labour markets where occupation–skill links.

I apply these constructions to the UK during 2016–2022, combining online job advertisements collected by Adzuna, monthly worker histories from Understanding Society, and the European Skills, Competences, Qualifications and Occupations (ESCO) supplying the skill descriptions. The UK during 2016–2022 provides a useful setting especially under the reallocation context, because the COVID-19 pandemic disrupted recruitment unevenly across activities (Blundell et al., 2020; Forsythe et al., 2020). The main mismatch analysis compares the direct skill-bundle partition with the nine major groups of the UK's Standard Occupational Classification (SOC), holding the number of search markets constant at nine. Estimates based on three-digit SOC occupations provide a separate assessment of the role of aggregation. Section 4.3 describes the career-based skill domains, and Section 6.2 uses them in a fractional allocation exercise.

The occupation-cluster series shows larger movements around 2020–2021 than the SOC major-group benchmark. Three-digit occupational mismatch also increases during the pandemic, indicating that the stability of the major-group series depends partly on aggregation. The fixed-share comparisons show that the measured path is sensitive to the distribution of vacancies. The skill-domain exercise gives a further comparison using fractional allocations and skill-share weights. As such, it provides a complementary measure of mismatch in which each occupation can contribute to more than one skill domain.

The main findings concern both pandemic-period movements and the direction of change of mismatch. I document larger movements in productivity-weighted mismatch around the COVID-19 pandemic under the nine skill-based occupation clusters than under the nine SOC major groups. The two classifications contain the same number of groups, so this comparison shows that which occupations are grouped together matters for how large the disruption appears. Pooling occupations with different skill requirements can obscure changes in the balance between jobseekers and vacancies. The relatively stable SOC series could therefore give the impression that the pandemic posed little challenge for worker reallocation, even when the skill-based series points to a much larger change.

Over 2016–2022, the two classifications also give different directional accounts: the skill-cluster series has an upward trend, whereas the SOC major-group series trends downwards. The skill-cluster increase is already visible before 2020, becomes more pronounced around 2020–2021 and is only partly reversed thereafter. By the end of 2022, its point estimate remains above its starting level. This pattern matters for how we interpret the recovery. A decline from the pandemic peak can coexist with a longer rise in measured mismatch. The skill-based estimates therefore motivate examining whether workers faced continuing barriers to moving into available jobs and where retraining or job placement support might help.

Two further comparisons help interpret this result. First, detailed occupational mismatch also rises over 2016–2022, showing that broad aggregation can conceal an upward path even within the SOC taxonomy. Second, calculations holding detailed occupational vacancy or jobseeker shares at their 2016 averages show that fixing vacancy composition changes the path more. This points to the distribution of recruitment opportunities as an important part of the measurement.

The paper makes three contributions. First, it provides two implementable ways to construct occupational groups from skills. Research on transferable human capital and skill relatedness motivates looking beyond occupational titles (Lazear, 2009; Lise and Postel-Vinay, 2020; Neffke and Henning, 2013). I bring required-skill similarity and career-path-based skill co-occurrence into a common empirical setting, show how they

can be used to regroup occupational data and to assess their implications for measured mismatch. The contribution lies in these market constructions and their use for measurement. Their application extends beyond mismatch accounting to questions about occupational mobility, changing labour demand and the skill content of potential career transitions.

Second, the paper adds two new facts on UK mismatch on an important time period. The skill-cluster comparison reveals larger pandemic-period movements than the broad SOC benchmark, showing that the definition of labour market segments affects the measured severity of disruption. Its upward path over the full sample is also consistent with skill-related allocation difficulties that predate the pandemic and remain relevant during the recovery. This matters for the UK productivity puzzle. Patterson et al. (2016) find an important role for occupational misallocation, whereas Turrell et al. (2021), using different vacancy data and matching-efficiency estimates, find little support for that account. I identify market definition as another consequential measurement choice: using the same underlying vacancy and worker data, alternative groupings change whether mismatch appears to be declining or accumulating.

Third, the paper sheds light on why the choice of market boundaries matters for skill policy. If an upward path is hidden by broad occupational categories, a policy discussion focused only on temporary pandemic disruption may overlook a continuing allocation problem. Skill-based groupings can help organise an assessment of which occupations share a skill base, where additional training may be needed for a move, and how training provision relates to changing recruitment demand. This motivates sustained attention to reskilling, upskilling and job-placement support as the economy changes.

Section 2 sets out the theoretical framework. Section 3 describes the data, and Section 4 presents the two approaches to labour market segmentation. Section 5 reports the matching-function estimates. Section 6 presents the mismatch estimates for occupation clusters, skill domains and detailed SOC occupations, followed by the fixed-share comparisons and policy implications. Section 7 concludes.

2 Theoretical framework

2.1 Environment and social planner's problem

The framework treats a market, indexed by i, as one element of a mutually exclusive partition of jobs and jobseekers. It can represent a SOC group, a detailed occupation or a skill-based group of occupations. Each worker is assigned to one market in a period, and matching occurs within that market. There is no simultaneous search across markets

or on-the-job search in the benchmark. Workers can change their assigned market across periods. Observed cross-group mobility is therefore informative about the plausibility of a boundary without being inconsistent with the period-by-period accounting.

I use a common matching framework to make the consequences of changing these boundaries transparent. It provides an interpretable counterfactual and connects the exercise to existing mismatch research (Şahin et al., 2014; Patterson et al., 2016). Redefining markets changes which occupational reallocations take place inside a market. It does not estimate workers' complete feasible search sets or remove all barriers to moving between occupations.

At date t, market i has v_{it} vacancies, u_{it} jobseekers and output per worker z_{it}. Matching efficiency $\phi_i > 0$ summarises the market-specific rate at which vacancies and jobseekers produce hires. Hires satisfy

$$h_{it} = \phi_i v_{it}^{\alpha} u_{it}^{1-\alpha}, \qquad 0 < \alpha < 1. \tag{1}$$

The parameter α is the elasticity of hires with respect to vacancies. Aggregate jobseekers and vacancies are $u_t = \sum_i u_{it}$ and $v_t = \sum_i v_{it}$. In the empirical baseline, jobseekers include both unemployment and inactivity; Section 3 discusses this broad definition.

The relevant planner benchmark reallocates the given stock of jobseekers to maximise output generated by new hires, holding vacancies, productivity and matching efficiency fixed:

$$\max_{\{u_{it}^* \geq 0\}} \lambda \sum_i z_{it} \phi_i v_{it}^{\alpha} (u_{it}^*)^{1-\alpha} \quad \text{subject to} \quad \sum_i u_{it}^* = u_t. \tag{2}$$

Here $0 < \lambda \leq 1$ is a common productivity factor for new hires, allowing for time spent in training. It cancels from the proportional mismatch index. This is a conditional allocation exercise for new hiring, not a calculation of the welfare gains from implementing a particular policy.

For active markets, optimality equalises the productivity-weighted marginal contribution of another jobseeker:

$$z_{it} \phi_i \left(\frac{v_{it}}{u_{it}^*} \right)^{\alpha} = \mu_t, \tag{3}$$

where μ_t absorbs the common terms $\lambda(1-\alpha)$. Markets with more vacancies, higher output per worker or greater matching efficiency receive more jobseekers in this benchmark. Defining $x_{it} = z_{it}\phi_i$, the allocation is

$$\frac{u_{it}^*}{u_t} = \frac{v_{it} x_{it}^{1/\alpha}}{\sum_j v_{jt} x_{jt}^{1/\alpha}}. \tag{4}$$

Appendix A gives the derivation. The allocation is feasible for the planner because it ignores the costs of retraining or relocating jobseekers between markets; actual workers need not be able to make these moves immediately.

2.2 Mismatch index

Let $y_t = \lambda \sum_i z_{it} h_{it}$ denote model-implied output associated with hires at the observed allocation and y_t^* the corresponding quantity under equation (4). Output mismatch is

$$\mathscr{Y}_t = 1 - \frac{y_t}{y_t^*} = 1 - \frac{\sum_i x_{it} (v_{it}/v_t)^{\alpha} (u_{it}/u_t)^{1-\alpha}}{\left[\sum_i x_{it}^{1/\alpha} (v_{it}/v_t)\right]^{\alpha}}. \tag{5}$$

An index of 0.05 represents a shortfall of 5% of counterfactual output from new hires, conditional on the inputs and assumptions. It does not represent a 5% loss of UK GDP, aggregate productivity or employment. The index is zero when the observed jobseeker shares coincide with the planner allocation. A common multiplicative rescaling of all x_{it} cancels, whereas changes in their relative values matter.

The productivity-weighted benchmark is useful because the allocation of hires across activities matters alongside their total number. An index targeting the number of hires instead answers a different question. I use only output mismatch in the main occupational comparisons and do not interpret a hires gap relative to an output-maximising allocation as the maximum attainable increase in employment.

Every classification changes the quantities and relative weights entering equation (5). Keeping the number of groups fixed rules out group count as the sole explanation for a difference, but does not hold productivity, matching efficiency or group size constant. The skill-cluster comparison consequently measures the effect of redefining markets, including the associated aggregation of these inputs; it is not a test that changes group membership while holding all parameters fixed.

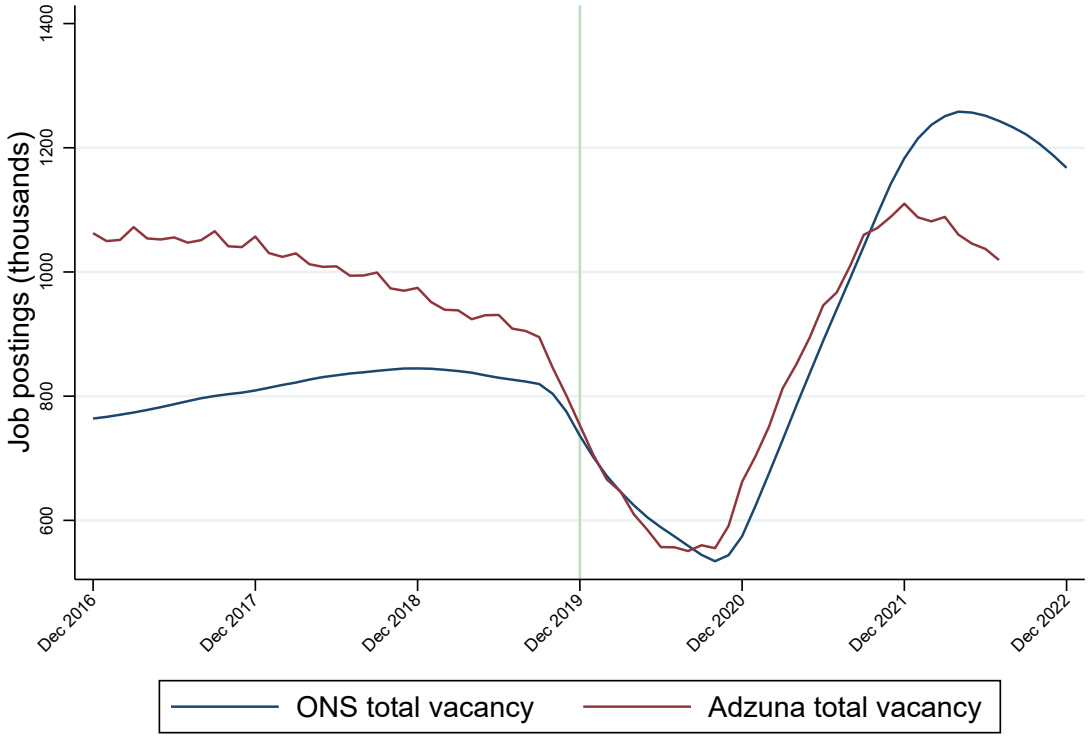
3 Data

The calculation requires vacancies, jobseekers, hires and productivity by sector. I first construct these quantities for three-digit SOC2000 occupations and then aggregate them using the alternative group definitions. Appendix B compares the sources with related studies. The benefit of combining these sources is that detailed occupational quantities can be regrouped under a common skill taxonomy. The longitudinal worker data also allow mobility to be examined alongside the mismatch calculation.

3.1 Vacancies: Adzuna

Adzuna collects online job adverts across the UK, with information on location, pay, job title and occupation. The data cover April 2016–December 2022 and contain an average weekly stock of more than 900,000 adverts. Occupations are mapped from SOC2020 to SOC2000 for the analysis. These are advertisements rather than a direct count of unfilled positions. I use their occupational distribution as a proxy for vacancy shares; the ONS also publishes an online-advert indicator based on Adzuna data (ONS, 2022).

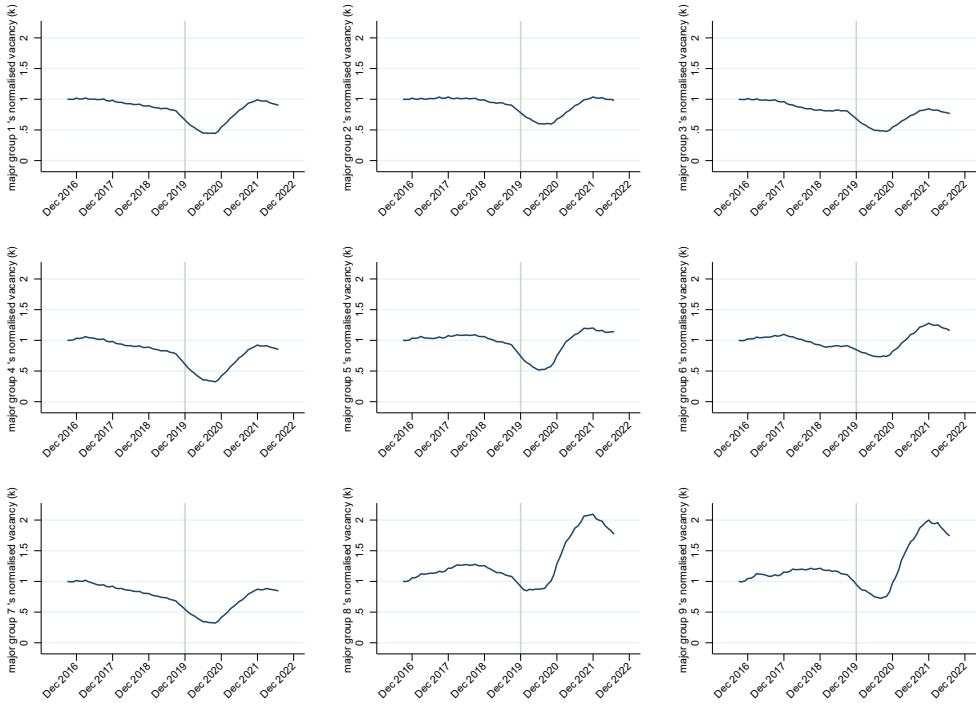
Figure 1: UK job adverts and vacancies, 2016–2022



Note:
[1] Data source: Adzuna, 2016–2022. Data are smoothed using a centred moving average.
[2] The green line marks December 2019 as a pre-pandemic reference date.

Figure 1 compares the stock of Adzuna online advertisements with the ONS Vacancy Survey estimate of unfilled jobs reported by employers. The two series measure different objects: a single advertisement may cover multiple positions, and not all vacancies are advertised online (ONS, 2022). Agreement in their broad direction is a useful aggregate check, but does not validate the occupational shares. Both series fall during the pandemic and recover thereafter. The Adzuna series exceeds its pre-pandemic level by late 2021. The moving average smooths short-term fluctuations, so turning points should be read as broad movements.

Figure 2: Vacancies by SOC2000 major group, 2016–2022



Note:
[1] Data source: Adzuna, 2016–2022. Data are smoothed using a centred moving average, with vacancy levels normalised to set December 2016 vacancy for each group as 1.
[2] The green line marks December 2019 as a pre-pandemic reference date.
[3] Major groups in SOC2000 are listed below: 1 Managers and Senior Officials; 2 Professional occupations; 3 Associate Professional and Technical Occupations; 4 Administrative and secretarial occupations; 5 Skilled trades occupations; 6 Personal Service Occupations; 7 Sales and customer service occupations; 8 Process, plant and machine operatives; 9 Elementary occupations.

Figure 2 shows vacancy indices for the nine SOC major groups. The groups differ in the depth of the decline and the strength of the recovery. Process, plant and machine operative vacancies show a particularly strong rebound.

The variation across groups indicates an uneven recovery in labour demand. It is consistent with the changing opportunities across lower-paid occupations discussed by Joyce et al. (2022).

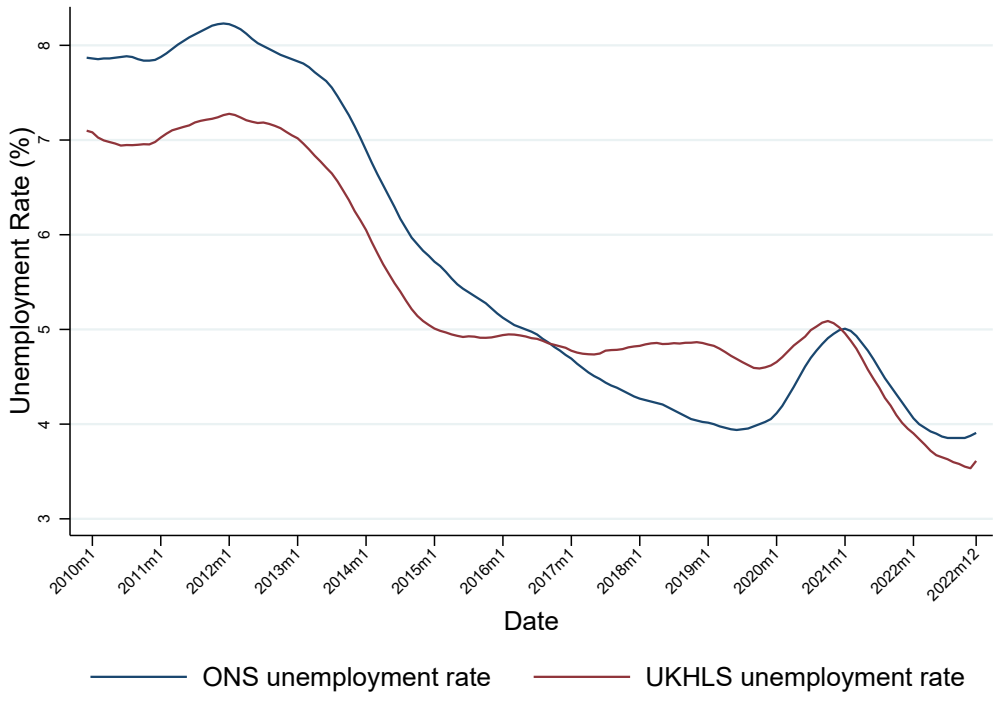
Online adverts provide detailed and frequent observations, but their coverage differs from that of a vacancy survey. An advert can represent several positions; repeated adverts may refer to one position; and recruitment conducted offline is absent. Changes in coverage across occupations or over time can therefore affect both levels and shares.

3.2 Employment, jobseekers and hires: UKHLS

Understanding Society, the UK Household Longitudinal Study (UKHLS), records employment status, occupations and job transitions. I use retrospective employment histories to construct monthly labour force states, following Postel-Vinay and Sepahsalari (2023). Jobseekers comprise unemployment and inactivity. This is a broad pool of potential entrants rather than a claim that all inactive people actively search. Appendix C considers an alternative that weights inactivity by its relative job-finding rate. Hires are monthly inflows into employment from these two states, measured at the three-digit SOC2000 level. This definition excludes job-to-job moves from the matching-function dependent variable.

Figure 3 compares the weighted UKHLS unemployment rate with the ONS Labour Force Survey series over 2010–2022. The smoothed series show similar broad declines, with some differences in level and short-term movement.

Figure 3: UK unemployment rates: UKHLS and ONS, 2010–2022



Note:
[1] Data Source: UKHLS, 2009–2023. Sample size: 448,782 observations from 83,732 individuals aged 16–65.
[2] Weights: Scaled cross-sectional weights were applied in this study.

3.3 Occupational productivity: ONS and UKHLS

I allocate ONS industry gross value added (GVA) to occupations using occupational employment within industries, as in Patterson et al. (2016). For industry k and occupation o, let E_{kot} be employment in occupation o within industry k, E_{kt} industry employment and E_{ot} occupational employment. The proxy is

$$GVA_{ot} = \sum_{k} \frac{E_{kot}}{E_{kt}} GVA_{kt}, \qquad z_{ot} = \frac{GVA_{ot}}{E_{ot}}. \tag{6}$$

This allocates a given industry's GVA in proportion to its occupational employment; differences between occupations reflect their industry composition. It is an accounting proxy for average output per worker, not an estimate of individual marginal productivity or the social value of an occupation. The latter distinction is especially relevant for public services. Descriptive productivity rankings are reported with the other estimated inputs in Appendix B.

4 Alternative labour market segments

The two constructions start from the same occupation–skill links but organise them differently. Direct clustering compares the skill bundles required by jobs. The career-based approach first learns a set of skill domains from observed employment histories and then compares occupations' compositions across those domains. The first supplies the main mismatch partition; the second offers a complementary representation and occupation grouping. Both distinguish occupational similarity from membership of an administrative category.

4.1 Skill data: ESCO

ESCO links occupations to required skill concepts through ISCO-08. I use version 1.2.0, released in May 2024, as a fixed taxonomy and map the profiles to SOC2000. Essential and optional occupation–skill links describe requirements, rather than directly measured individual skills. Because the same taxonomy is applied to 2016–2022, the estimates capture changing allocations under fixed profiles. They cannot measure historical changes in the skills required within an occupation. Tables B.4 and B.5 report the hierarchy and numbers of linked skills.

4.2 Skill-based occupation clusters

The main construction puts occupations requiring similar bundles of skills into the same search market. Its purpose is to allow a labour market boundary to cut across SOC major groups, while retaining mutually exclusive markets suitable for equation (5).

For each occupation o, the vector $s_o = (s_{1o}, \ldots, s_{No})$ records whether each level-3 ESCO skill category is required, with $s_{no} \in \{0, 1\}$. I use k-means to group these vectors into nine clusters. The algorithm groups occupations by proximity to the cluster means in this skill space; every occupation receives one assignment. Euclidean distance is

$$d_{oj} = \left[\sum_{n=1}^{N} (s_{no} - s_{nj})^2 \right]^{1/2}. \tag{7}$$

For binary indicators, squared distance counts the skill categories present in one profile but absent from the other. All included categories receive equal weight; the measure does not incorporate skill proficiency or the time required to acquire a skill.

I choose nine clusters to provide a comparison with the nine SOC major groups at the same level of aggregation by group count. This is a design choice, not a claim that nine is the statistically optimal number of search markets. The scalar-distance fit diagnostic in Appendix B concerns clusters of pairwise distances rather than occupations' skill vectors, so it cannot establish the preferred number of occupation clusters. The level-4 distance heatmap is also reported there as a descriptive illustration at a finer skill resolution.

4.2.1 Comparison with SOC major groups

Table 1 gives short descriptions and examples of the recovered groups; the complete assignments are in Table B.6. These descriptions summarise occupational membership. They should not be read as estimated rankings of the most distinctive skill concepts.

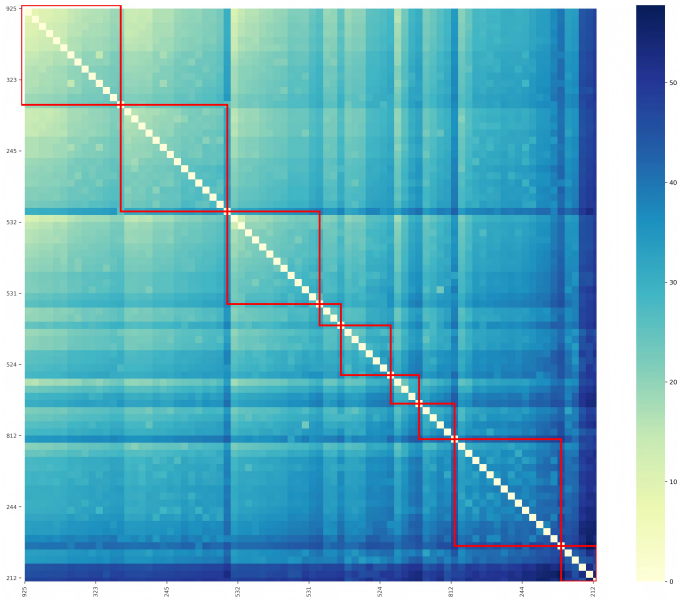
The health cluster, for example, combines health professionals, associate professionals, therapists and personal-service occupations. These jobs span several SOC major groups but share a substantive field of work. Management and administrative occupations, in turn, are distributed across several clusters. Thus the construction both joins occupations separated by SOC and separates occupations that SOC places together. Figure 4 displays the reordered distances, and Figure 5 shows explicitly how SOC groups split across the skill clusters.

The composition comparison is central to the interpretation of the mismatch results. For example, if demand rises for health professionals while falling for another type of professional occupation, the broad professional SOC group can offset the two movements.

Table 1: Overview of the nine occupation clusters

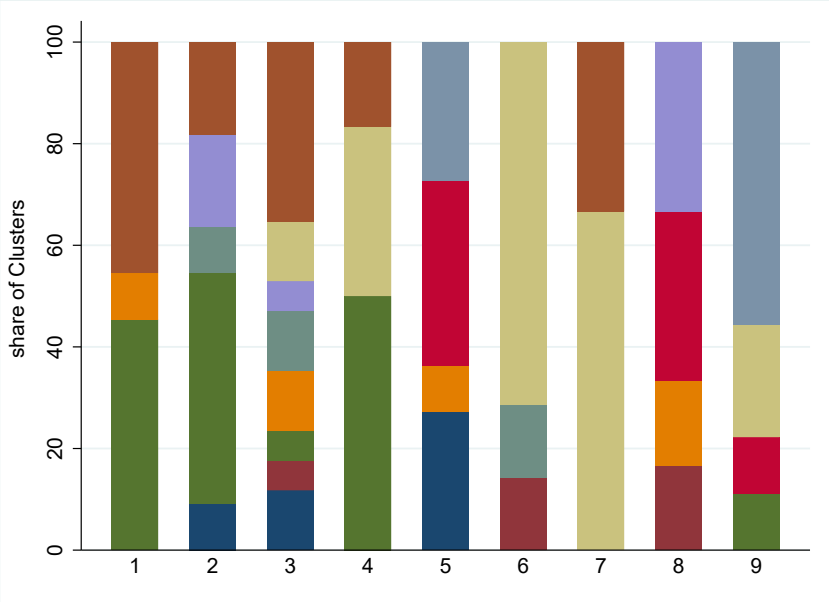
Cluster	Membership-based description	Examples of member occupations
1	Design, Inspection and Craft Trades	Architects, town planners, surveyors; Draughtspersons and building inspectors
2	Recreation, Leisure and Transport Services	Sports and fitness occupations; Leisure and travel service occupations
3	Management, Professional and Administrative Services	Quality and customer care managers; Financial institution and office managers
4	Production, Logistics and Land-Based Operations	Production managers; Transport associate professionals
5	Health Care Practitioners and Support Services	Health professionals; Health associate professionals
6	Construction and Industrial Trades and Operatives	Metal forming, welding and related trades; Vehicle trades
7	Science, Engineering and Technical Operations	Science professionals; Engineering professionals
8	Customer, Personal and Administrative Services	IT service delivery occupations; Social welfare associate professionals
9	Leadership and Public-Facing Professional and Associate Roles	Corporate managers and senior officials; Functional managers

Figure 4: Occupational distances ordered by the nine skill-based clusters



Note:
[1] Data source: ESCO,
[2] Occupations are ordered by the average distance to other occupations in the same cluster.

Figure 5: Composition of SOC major groups by occupation cluster



Note:
SOC major-group codes follow Figure 2.

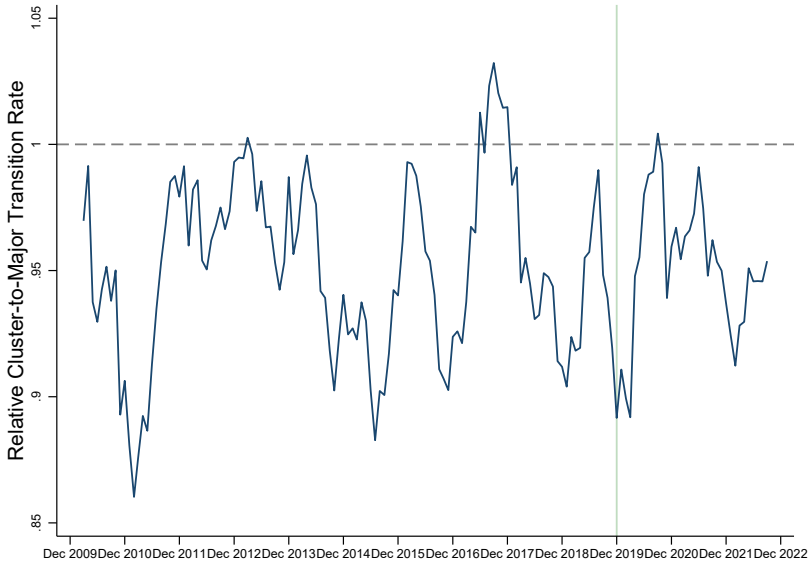
The skill grouping can place these occupations in separate markets. Conversely, changes across SOC categories can be absorbed within a skill cluster. This illustrates why re-defining boundaries can either amplify or attenuate the measured imbalance; it does not attribute the observed aggregate result to a particular cluster without a cluster-level decomposition.

4.2.2 Occupational mobility

Figure 6 compares the rate of observed transitions across skill clusters with the rate across SOC major groups. A ratio below one means that a larger fraction of moves remains within the skill grouping; a ratio above one means the reverse. The figure varies around one rather than showing uniformly fewer crossings under the skill classification. The evidence therefore supports examining alternative boundaries but does not establish that skill clusters are impermeable or universally more accurate search markets. Lower crossing rates could also reflect unequal group sizes or composition.

The main partition uses occupational skill requirements rather than the transition rates shown here. The mobility comparison is a descriptive check on what its boundaries mean. In the career-based construction below, histories enter the grouping itself, so a mobility comparison has a different interpretation.

Figure 6: Cross-group transition rates: occupation clusters relative to SOC major groups



Note:
SOC major-group codes follow Figure 2.
Occupation-cluster codes follow Table B.6.

4.3 Skill domains

The second approach uses the jobs workers have held to identify skills that tend to occur together. For each worker, I combine the skill requirements of all occupations observed in that worker's employment history. Skills can occur together because they are linked to the same occupation or because a worker has held several occupations. These links describe exposure to occupational skill requirements. The construction draws on the idea of skill relatedness in mobility (Neffke and Henning, 2013).

I use the resulting domain shares in two ways. For the mobility comparison, I group whole occupations by the composition of their skill domains. For the mismatch exercise in Section 6.2, I allocate each occupation's quantities across domains in proportion to its skill shares. Table 2 summarises the difference between the direct occupation clusters and the fractional allocation across skill domains.

Career-history skill exposure. I merge the occupations observed at interviews with ESCO skill requirements. The indicator $x_{p,s}$ equals one when worker p has an observed occupation linked to skill s, and zero otherwise. The resulting individual–skill matrix records exposure through occupational requirements. A worker with a single observed occupation can consequently be linked to several skills.

Table 2: Occupation clusters and skill domains

	Occupation clusters	Skill domains
Object grouped	Occupations	Skill concepts
Information	Required skill indicators	Co-occurrence in career exposure vectors
Algorithm	K-means, nine clusters	Leiden on a thresholded skill network
Occupation assignment	One cluster per occupation	Shares across domains
Quantity allocation	Whole occupation	Fractional accounting allocation
Output weight	Output-per-worker proxy	Skill-share sum q_{ct}
Interpretation	Conditional output mismatch	Separate weighted allocation diagnostic

Skill relatedness. For each pair of skills, the Jaccard coefficient is

$$J_{ij} = \frac{A_{ij}}{s_i + s_j - A_{ij}}, \qquad A_{ij} = \sum_p x_{p,i} x_{p,j}, \quad s_i = \sum_p x_{p,i}.$$

It measures the overlap in the sets of workers exposed to the two skills. I construct a weighted, undirected network containing edges with $J_{ij} \geq 0.41$ and use J_{ij} as the edge weight.

Skill communities. I apply Leiden community detection (Traag et al., 2019) to the weighted network. The objective is

$$Q = \frac{1}{2m} \sum_{ij} \left[w_{ij} - \gamma \frac{k_i k_j}{2m} \right] \delta(c_i, c_j),$$

where k_i is weighted node degree and $m = \frac{1}{2} \sum_{ij} w_{ij}$. The calculation uses $\gamma = 1$ and produces nine communities. These communities group skill exposures, whereas the main k-means exercise groups occupations represented by skill vectors. The network threshold, resolution parameter and observed histories influence membership. Nine communities here are an outcome of this construction; they are not automatically equivalent to the nine occupation clusters.

Occupation–domain shares. Let $\mathscr{S}_o^m$ denote the skill entries for occupation o whose identifiers match a domain assignment. The share allocated to domain c is

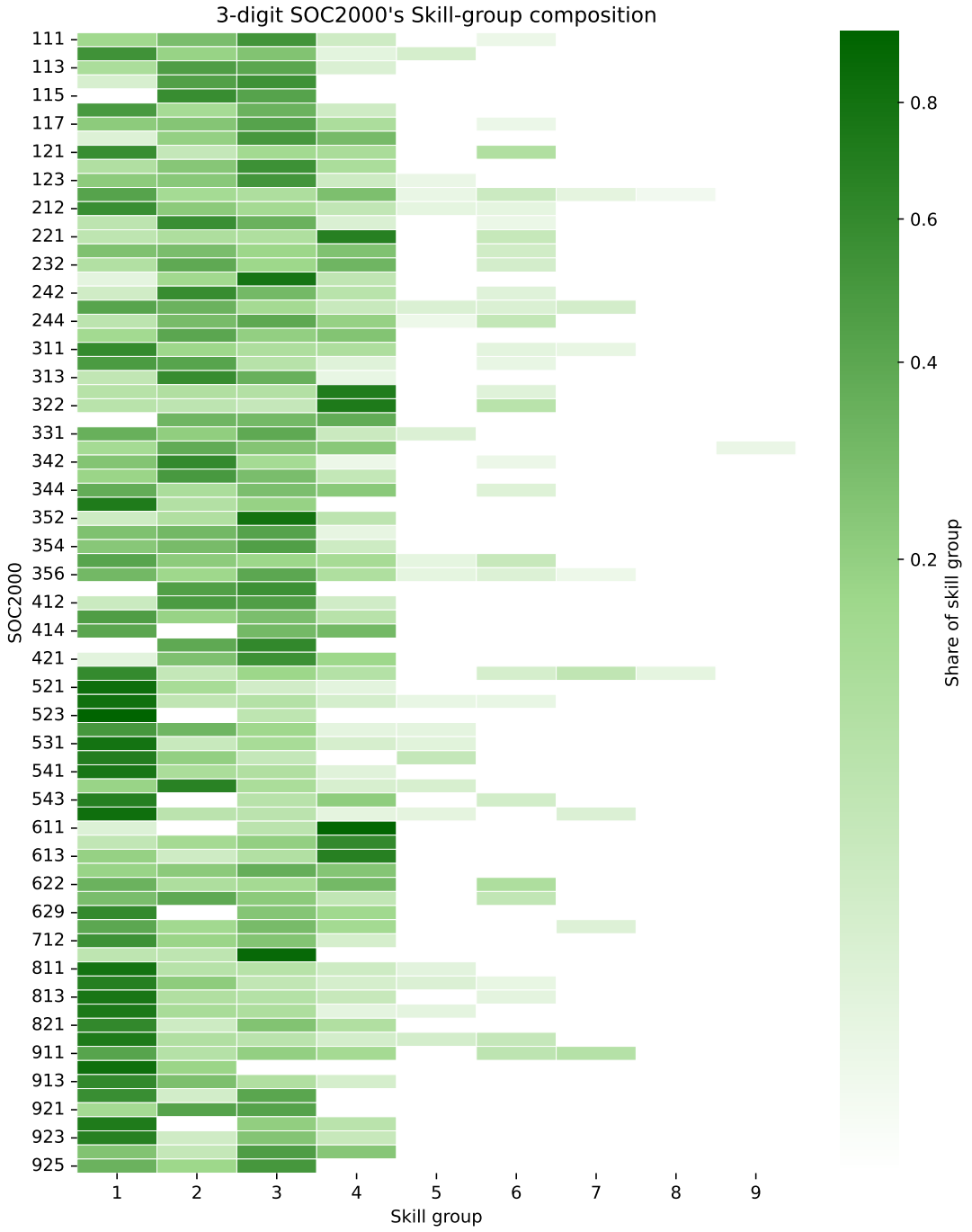
$$n_{o,c} = \sum_{s \in \mathscr{S}_o^m} \mathbf{1}\{c(s) = c\}, \qquad \text{share}_{o,c} = \frac{n_{o,c}}{\sum_{c'} n_{o,c'}}.$$

For example, an occupation with 60 matched skill entries in one domain and 40 in another receives shares 0.6 and 0.4. This is an illustrative calculation, not a reported occupation estimate. Shares sum to one within each mapped occupation. The identifier match retains 12,695 of 45,492 occupation–skill links, so the domain allocation describes the matched subset of skill requirements (approximately 27.9% of the original links). Normalising the retained links to sum to one preserves totals for mapped occupations but does not recover the omitted skill content. The resulting domain shares can therefore be sensitive to identifier coverage. Figure 7 shows the resulting composition matrix. Occupations can contribute to several domains.

Table 3 labels the nine communities using their constituent skill concepts. The labels summarise the concepts in each group; occupation shares in Figure 7 show their distribution across jobs.

Occupational mobility. I apply k-means to occupations' nine domain shares to form nine mutually exclusive occupation groups. Figure 8 compares transitions across these groups with transitions across SOC major groups. This is an in-sample comparison because the same career histories enter the construction of the domain shares.

Figure 7: Skill-domain shares by three-digit SOC2000 occupation

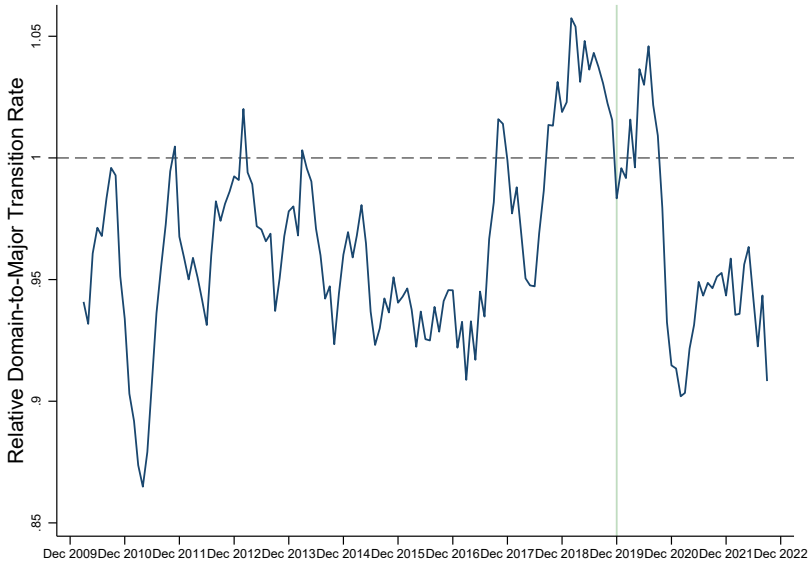


Note: Data sources: ESCO occupation–skill links and skill-domain assignments constructed from UKHLS career histories.

Table 3: Skill-domain labels and example concepts

Domain	Label	Description	Example skill concepts
1	Industrial, engineering and creative production	Industrial processes, engineering and creative production.	chemical engineering and processes; building materials industry; music and performing arts
2	Education and qualifications	Education, teaching and general-qualification concepts.	education science; teacher training; generic programmes and qualifications
3	Management, law and administration	Management, administration, marketing and legal knowledge.	management and administration; marketing and advertising; law
4	Health and human development	Health, language acquisition and related human-development concepts.	philosophy and ethics; language acquisition; medical diagnostic and treatment technology
5	Extraction and energy	Extraction, fuels and industrial materials.	mining and extraction; adhesives; alternative fuels
6	Religion and natural sciences	Religious knowledge and selected natural-science concepts.	religion and theology; forestry
7	Horticulture and agriculture	Horticulture and plant production.	0812 - horticulture
8	Applied science and construction	Biological-science and construction concepts.	biological and related sciences not elsewhere classified; constructing
9	Creative and performing arts	Artistic creation, performance and entertainment.	arts not elsewhere classified; performing and entertaining

Figure 8: Cross-group transitions: domain-composition groups relative to SOC major groups



Note:
SOC major-group codes follow Figure 2.
The numerator is the transition rate across nine occupation groups obtained by applying k-means to occupations' skill-domain share vectors. These groups differ from the nine Leiden communities of skills. The denominator is the transition rate across SOC major groups. The plotted ratio is smoothed using a centred seven-observation moving average. The numerator is labelled "Domain-to-Major" in the graph.

5 Matching-function estimates

5.1 Matching elasticity

Mismatch requires more than a grouping: each grouping also needs vacancies, jobseekers, productivity and matching efficiency on the same basis. I distinguish the construction of the underlying data from the estimates used to weight the allocation benchmark. For a partition $\mathscr{G}$ with occupations $o \in g$, the additive quantities are $V_{gt} = \sum_{o \in g} V_{ot}$, $U_{gt} = \sum_{o \in g} U_{ot}$ and $H_{gt} = \sum_{o \in g} H_{ot}$. An internally consistent group output-per-worker proxy requires summing allocated GVA and dividing by group employment, rather than summing occupational productivity ratios.

I estimate the matching relationship over April 2016–December 2020, using aggregate data and two- and three-digit occupational panels:

$$\ln\left(\frac{h_{it}}{u_{it}}\right) = \theta' \mathbf{T}_t + \ln \phi_i + \alpha \ln\left(\frac{v_{it}}{u_{it}}\right) + \varepsilon_{it}. \tag{8}$$

The time controls $\mathbf{T}_t$ are a quartic trend or quarterly seasonal indicators, depending on the specification. The panel regressions include occupation fixed effects; the aggregate

regression uses OLS. Table 4 reports the estimates and sample coverage.

These coefficients describe a conditional matching relationship. Vacancy posting and the distribution of non-employment can respond to unobserved hiring conditions; measurement error in the common denominator u_{it} can also affect both sides. Fixed effects and time controls need not remove these sources of endogeneity. I therefore use the estimates to describe and calibrate the allocation exercise rather than interpret α as an identified causal response to an exogenous vacancy change.

Table 4: Matching elasticity estimates using aggregate and two- and three-digit occupational data

	Elasticity parameter (α)		Cross-sections	Observations
	(1)	(2)		
Aggregate	0.561***	0.563***	1	56
	(0.161)	(0.131)		
2-digit SOC2000	0.432***	0.395***	25	1160
	(0.103)	(0.098)		
3-digit SOC2000	0.471***	0.441***	80	2123
	(0.063)	(0.062)		

Note:
[1] Column (1) shows results from the matching regression with a quartic time trend, while Column (2) presents results from the matching regression with seasonal dummies.
[2] *** denotes statistical significance at the 1% level. Reported standard errors are in parentheses. Occupational regressions include occupation fixed effects. The estimation window is shorter than the period displayed in the mismatch figures.
[3] The estimation window is April 2016–December 2020; December 2019 is absent from the matching inputs.

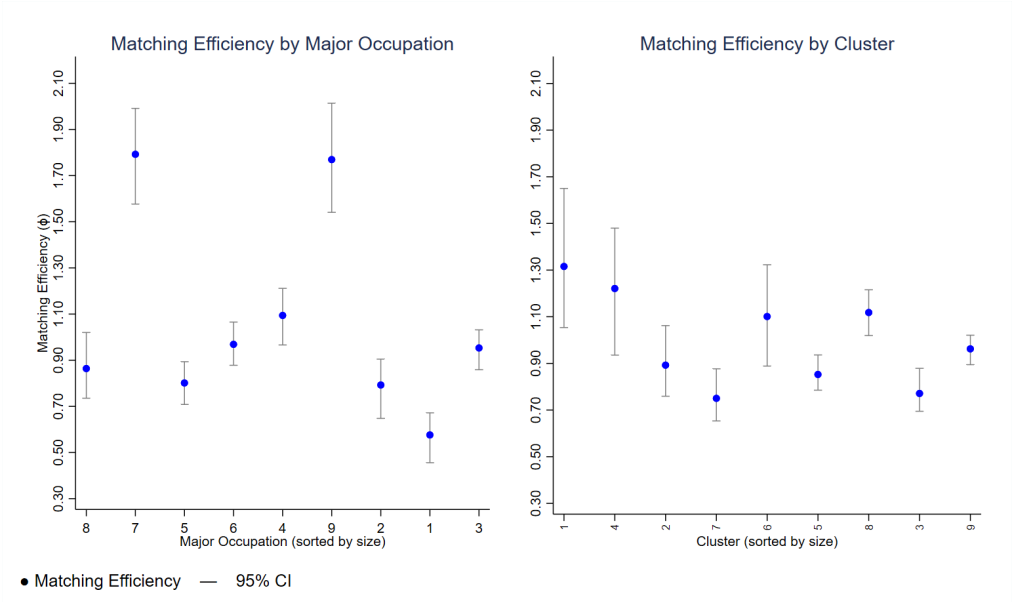
The aggregate elasticity estimates are approximately 0.56, while the detailed occupational estimates are 0.441–0.471. I use a common $\alpha = 0.5$ across classifications to keep the technology parameter fixed when comparing market definitions and to retain comparability with Patterson et al. (2016). The estimates provide a benchmark for this calibration, rather than a separate elasticity used for every partition. This choice does not establish robustness to alternative elasticities: equation (4) shows that α also affects how strongly relative productivity and efficiency tilt the planner allocation.

5.2 Occupation-cluster matching efficiency and productivity

Matching efficiency is a group-level object. A Cobb–Douglas function fitted to an aggregate of occupations does not generally have an efficiency equal to an unweighted average of the occupational efficiencies. Likewise, group output per worker depends on employment weights. Changing segmentation therefore requires quantities and weights to refer to the same groups; it is not sufficient simply to relabel the SOC effects.

Figures 9 and 10 report the group-level efficiency estimates and productivity summaries for the two nine-group classifications. The cluster efficiency estimates are less dispersed in the figure, while productivity remains heterogeneous. On the displayed productivity scale, SOC major-group values range from 66 to 163 (mean 127, standard deviation 32), and cluster values from 69 to 184 (mean 133, standard deviation 36). The difference in group averages reflects how the same occupations are combined and weighted; the unweighted mean across groups need not be invariant to the partition.

Figure 9: Matching efficiency: SOC major groups and occupation clusters



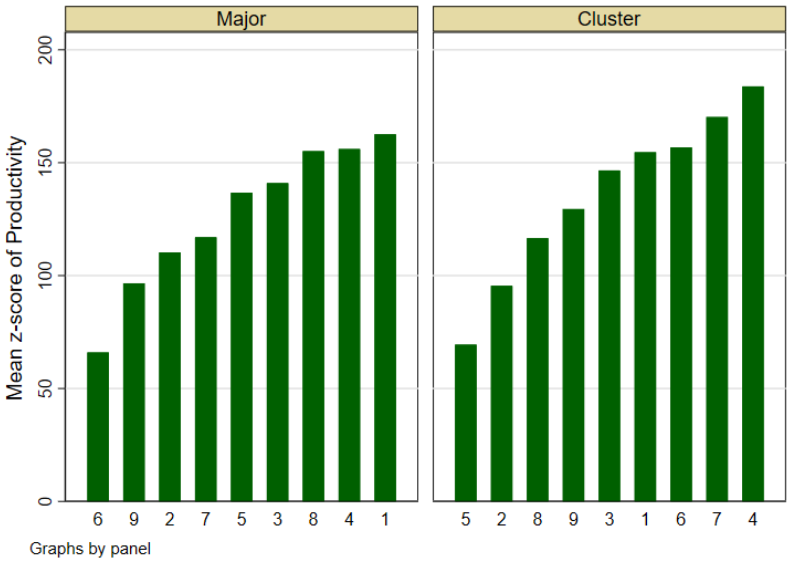
Note:
SOC major-group codes follow Figure 2.
Occupation-cluster codes follow Table B.6.

The efficiency estimation window ends in 2020, whereas the mismatch series extend through 2022. The later values are consequently conditional on the estimated matching structure, rather than fresh estimates of monthly efficiency throughout the recovery. Appendix B documents the occupational dispersion and imprecision in these estimates, and Appendix C reports alternative jobseeker definitions and estimation windows. These qualifications matter when assessing changes in relative $z_{it}\phi_i$, which enter both observed and counterfactual output.

5.3 Skill-domain matching efficiency

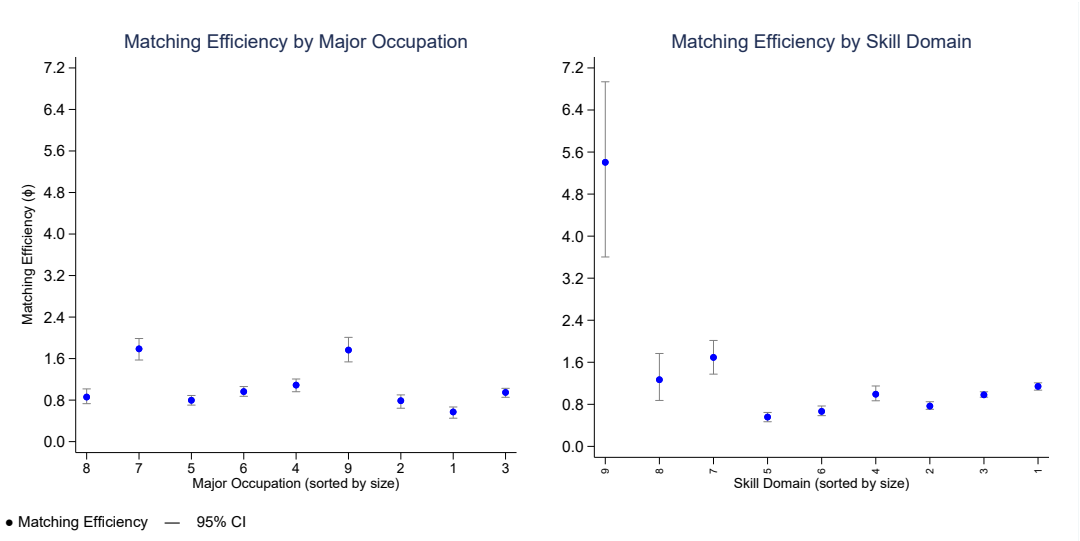
Several domains contribute small shares to the occupation–domain matrix. Figure 11 reports the matching-efficiency estimates and their resampling intervals.

Figure 10: Productivity: SOC major groups and occupation clusters



Note:
SOC major-group codes follow Figure 2.
Occupation-cluster codes follow Table B.6.

Figure 11: Matching efficiency: SOC major groups and skill domains



Note:
SOC major-group codes follow Figure 2.
Skill-domain codes follow Table 3.

Figure 11 compares matching-efficiency estimates for SOC major groups and skill domains. Most lie between about 0.5 and 1.8. Domain 9 has a substantially higher estimate and a wide interval, so its point estimate carries little precision.

6 Sectoral mismatch

I first compare the nine occupation clusters with the nine SOC major groups. I then report the allocation index across skill domains. The final comparisons use three-digit SOC occupations to examine the role of aggregation and changes in vacancy and jobseeker shares.

6.1 Mismatch across skill-based occupation clusters

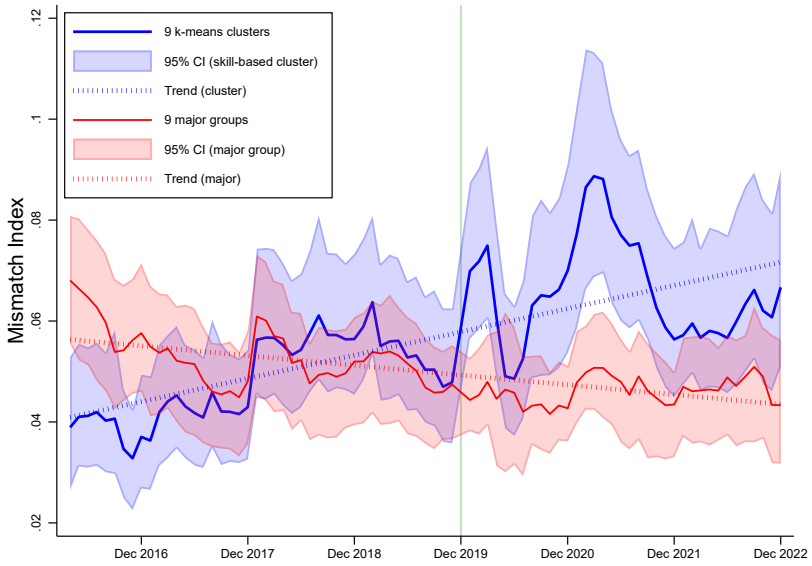
The main comparison asks how market definition affects both the size of measured mismatch movements during the pandemic and the direction of change over 2016–2022. Figure 12 applies the same output-mismatch formula to nine skill-based occupation clusters and nine SOC major groups. Blue denotes the skill clusters and red the SOC benchmark. The two findings concern different features of the same series: their pandemic-period fluctuations and their fitted sample trends.

The first finding is that the skill-cluster series exhibits larger movements around the pandemic. Its more pronounced rises and reversals around 2020–2021 contrast with the smaller movements in the SOC major-group series. The comparison shows that broad occupational categories can mute the measured size of a reallocation imbalance during a major labour market disruption, even when the number of groups is held at nine.

The second finding concerns the direction of change over the full sample. The SOC major-group index falls early in the sample and then fluctuates around a comparatively stable level, giving a downward fitted trend overall. The skill-cluster index has an upward fitted trend: its point estimate starts below the SOC index and finishes above it. This directional contrast complements the larger pandemic-period movements. Changing which occupations share a market affects both how large the disruption appears and whether measured mismatch seems to be easing or building up over time.

The timing is also informative. The skill-cluster series increases before 2020, rises more sharply around 2020–2021 and then partly reverses. By late 2022, its point estimate remains above both its initial level and the level near the pre-pandemic reference date. The rise is uneven rather than monotonic, and some movements are imprecisely estimated. Nevertheless, the displayed path is consistent with allocation difficulties extending beyond the immediate pandemic disruption. It should not be summarised solely as a tem-

Figure 12: Output mismatch: nine occupation clusters and nine SOC major groups, 2016–2022



Note:
[1] The blue line reports the nine skill-based occupation clusters; the red line reports the nine SOC major groups. Shaded areas are the plotted 95% resampling intervals. Dotted lines are descriptive fitted trends over the displayed sample; the skill-cluster trend slopes upwards and the SOC trend downwards. They are not tests of trend significance or estimates of a pandemic effect.
[2] Each index is a gap relative to counterfactual output from new hires under its own grouping. The levels are not percentages of national GDP.
[3] Data source: UKHLS, Adzuna, ONS, ESCO.
SOC major-group codes follow Figure 2.
Occupation-cluster codes follow Table B.6.

porary increase that disappears with reopening, nor extrapolated beyond the 2016–2022 observation window.

Holding the number of groups at nine makes the contrast informative about market composition. Changes hidden within a broad SOC group can become changes between skill clusters, while movements between SOC groups can become internal to a cluster. Grouping also changes group sizes, productivity and matching-efficiency weights. These ingredients together generate the alternative measured paths; the comparison does not isolate group membership while holding all other inputs fixed.

For the UK productivity debate, the relevant message is that a broad occupational measure can indicate an easing allocation problem while a skill-motivated measure points in the opposite direction. Misallocation of new employment towards less productive uses is one possible link between mismatch and weak productivity (Patterson et al., 2016). The sensitivity documented by Turrell et al. (2021) to vacancy data and matching efficiency already makes this link a measurement question.

Figure 12 shows why market boundaries belong in that assessment as well. The result

is a difference in the measured sample path under two economically interpretable classifications. Its policy relevance is that a diagnosis based only on broad SOC groups may miss a continuing reason to examine skill transfer, retraining and worker reallocation.

6.2 A skill-domain allocation exercise

An occupation can draw on more than one domain. I therefore split its measured employment, potential jobseekers, adverts and hires in proportion to its matched skill counts. In the 0.6/0.4 example in Section 4.3, 100 occupational adverts contribute 60 and 40 advert-equivalents to the two domains. These are accounting units, not separately observed domain vacancies or two independent groups of workers. Formally:

$$E_{ct} = \sum_o \text{share}_{o,c} E_{ot}, \quad U_{ct} = \sum_o \text{share}_{o,c} U_{ot},$$
$$V_{ct} = \sum_o \text{share}_{o,c} V_{ot}, \quad H_{ct} = \sum_o \text{share}_{o,c} H_{ot}.$$

Here U_{ot} includes unemployment and inactivity. The calculation preserves totals over mapped occupations when the shares sum to one.

For a domain-level output-per-worker measure, allocating employment and output consistently gives

$$z^E_{ct} = \frac{\sum_o \text{share}_{o,c} E_{ot} z_{ot}}{\sum_o \text{share}_{o,c} E_{ot}}, \qquad E_{ct} > 0.$$

The domain exercise shown below uses a different weight,

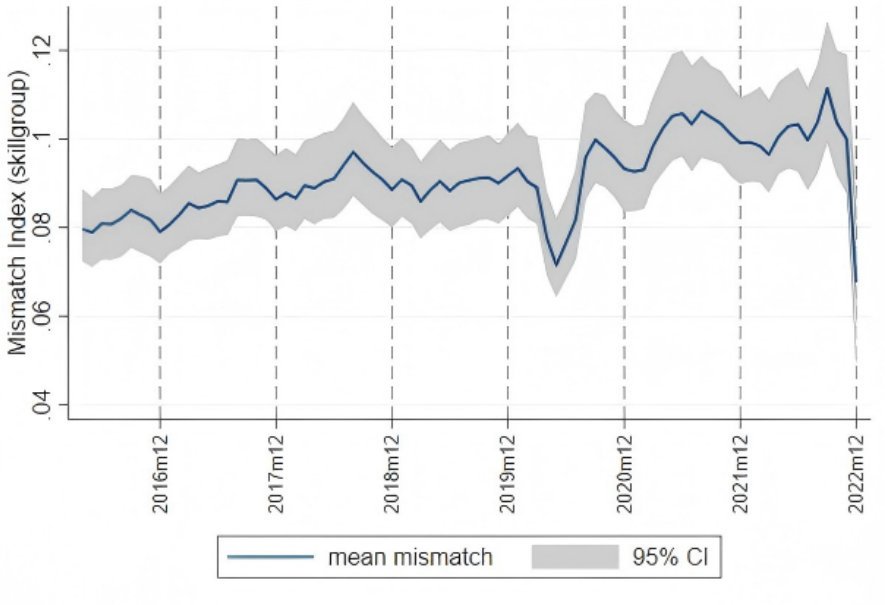
$$q_{ct} = \sum_o \text{share}_{o,c} z_{ot}.$$

This weight aggregates occupational productivity by skill share. Its scale depends on the number and composition of occupations contributing to a domain; it is not domain output per worker. The 2021–2022 calculations use the corresponding monthly weights from 2020.

I apply the allocation formula using $x_{ct} = q_{ct}\phi_c$. The resulting index describes the gap relative to the allocation that maximises this weighted matching criterion. It is a complementary weighting exercise, with a different interpretation from the occupational output index based on GVA per worker. Domain shares describe accounting allocations across skill groups, not directly observed choices of a single search market by each worker.

Figure 13 shows a decline in the domain allocation index around early 2020, followed by a rebound. Its timing differs from the occupation-cluster series. The fractional map-

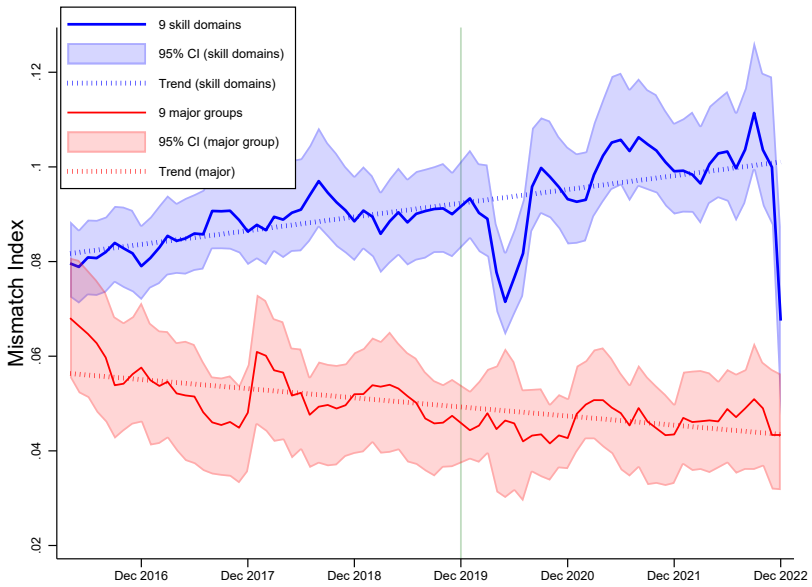
Figure 13: Skill-domain allocation index, 2016–2022



Note:
[1] Data source: UKHLS, Adzuna, ONS, ESCO.
Skill-domain codes follow Table 3.

ping and domain weights both enter this comparison.

Figure 14: Skill-domain allocation index and SOC major-group benchmark, 2016–2022



Note:
[1] Data source: UKHLS, Adzuna, ONS, ESCO.
SOC major-group codes follow Figure 2.
Skill-domain codes follow Table 3.

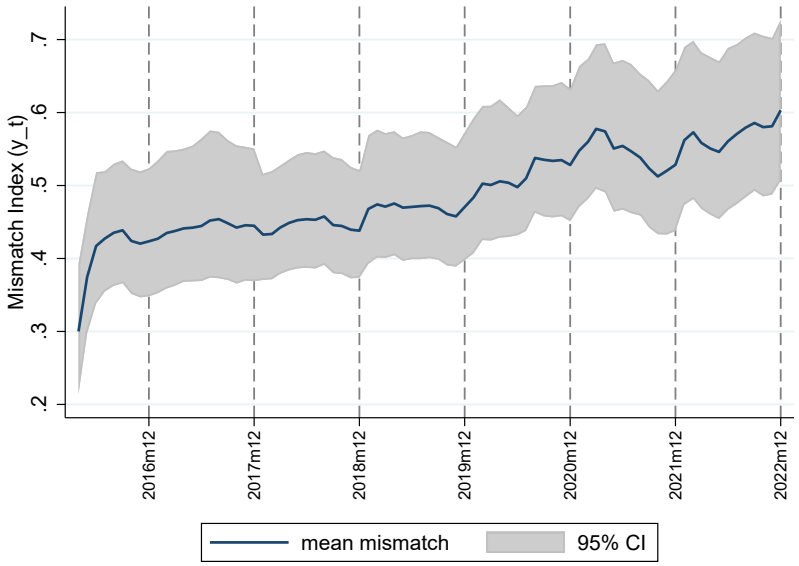
Figure 14 places the domain index alongside the SOC major-group benchmark. The domain series has a higher displayed level and a different time profile. Both constructions use nine groups, but differ in their allocation rule and productivity weights. The wide band near the end of the sample also makes the final movement imprecise.

The domain index describes how the allocation of vacancies and potential jobseekers across skill domains changes over time. Its level depends on the fractional allocation rule and on q_{ct}. The comparison with the SOC benchmark therefore combines differences in market definition, how occupational quantities are allocated and how domains are weighted. The transition rates in Figure 8 describe mobility across mutually exclusive occupation groups constructed from domain shares. They provide information about a different grouping from the fractional allocation used here.

6.3 Three-digit occupational mismatch

The first check retains the conventional SOC taxonomy and increases its detail. Figure 15 reports equation (5) for 80 three-digit occupations. This asks whether the flat major-group path is also present before occupations are combined into broad categories.

Figure 15: Three-digit occupational output mismatch, 2016–2022



Note:
[1] Occupation classification: 3-digit SOC2000, 80 groups.
[2] The line reports the mean estimated index and shading shows the resampling bands. The denominator is counterfactual output from new hires.
[3] Data source: UKHLS, Adzuna, ONS.

The detailed occupational series also has an upward path over 2016–2022, including

increases before 2020. It reaches a local high around early 2021, eases, and then rises again towards the end of the sample. The final point remains above its pre-pandemic level, although the wide intervals limit precision. Thus the evidence of a longer sample increase is not confined to the skill-cluster construction. The contrast with broad SOC groups shows how aggregation can suppress that direction of change. It complements the equal-count comparison: one changes the degree of aggregation within SOC, while the other changes composition at a common group count. Differences in weights and precision remain, so the index levels are not interchangeable measures of national GDP loss.

6.4 Fixed vacancy and jobseeker shares

The second check holds one side of the market distribution fixed. Define $s^u_{it} = u_{it}/u_t$ and $s^v_{it} = v_{it}/v_t$, with $\bar{s}^u_{i,2016}$ and $\bar{s}^v_{i,2016}$ their mean shares in 2016. I recalculate the detailed occupational index first replacing s^u_{it} by $\bar{s}^u_{i,2016}$, and then replacing s^v_{it} by $\bar{s}^v_{i,2016}$. Each counterfactual changes both the observed-allocation term and, where relevant, the planner normalisation. The exercise asks how the path changes when the distribution of jobseekers or recruitment opportunities is held at its starting-year composition.

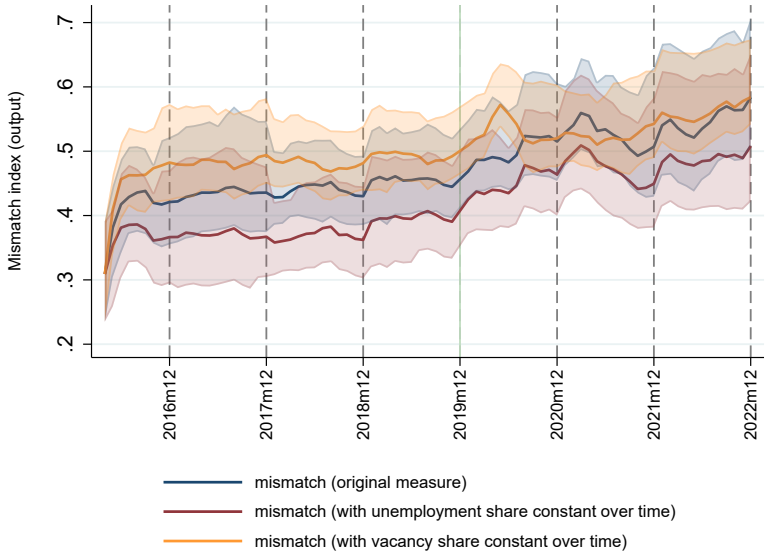
Table 5: Covariance of baseline and fixed-share output mismatch indices

Covariance Matrix	Mismatch (original)	Mismatch (constant job-seeker share)	Mismatch (constant vacancy share)
Mismatch (original)	0.27	0.26	0.19
Mismatch (constant job-seeker share)	0.26	0.27	0.17
Mismatch (constant vacancy share)	0.19	0.17	0.17

Note:
[1] Occupation classification: 3-digit SOC2000, 80 groups.
[2] Data source: UKHLS, Adzuna, ONS.
[3] The numbers displayed in the table represent 100 times the covariance values.
[4] The occupational job-seeker/vacancy share utilised in the calculation is the mean in 2016 for 3-digit SOC2000 occupation.

The fixed-jobseeker-share series follows the observed index more closely than the fixed-vacancy-share series. In the rounded covariance table, the baseline variance is 0.27 and its covariance with the fixed-jobseeker series is 0.26, compared with 0.19 for the fixed-vacancy series; all entries are scaled by 100. These values imply correlations of approximately 0.96 and 0.89, respectively. Thus both counterfactual paths remain positively associated with the baseline, but fixing vacancy shares removes more of its variation in this comparison.

Figure 16: Output mismatch under fixed occupational job-seeker or vacancy shares, UK, 2016–2022



Note:
[1] Data source: UKHLS, Adzuna, ONS.
[2] Occupation classification: 3-digit SOC2000, 80 groups.
[3] The occupational job-seeker/vacancy share utilised in the calculation is the mean in 2016 for 3-digit SOC2000 occupation.
[4] Lines report the mean estimated indices; shading shows the resampling bands. The denominator is counterfactual output from new hires.

Economically, this is consistent with changing recruitment opportunities altering where the planner would place jobseekers. It does not establish that vacancy changes caused the measured mismatch: vacancies can respond to the same shocks as worker supply, efficiency and productivity. Nor is Table 5 an additive variance decomposition. The index is nonlinear, so these covariances cannot be converted directly into percentages attributable to supply and demand. A decomposition of the difference between the two nine-group indices would require a further exercise that switches their inputs systematically; the present comparison concerns three-digit occupational dynamics.

6.5 Discussion and policy implications

The comparisons establish four linked results. Nine skill-based occupation clusters reveal larger pandemic-period movements than nine SOC major groups. Replacing SOC groups with skill clusters also changes the direction of the fitted sample trend. Detailed SOC estimates rise as well, revealing changes suppressed by broad aggregation. Holding detailed vacancy shares fixed changes the time profile more than holding jobseeker shares fixed. Together, these findings make the definition of markets and the distribution of recruit-

ment central to interpreting the persistence of measured mismatch. They do not identify a uniquely correct partition or the source of every skill-related allocation difficulty.

The main uncertainties follow directly from measurement. Online adverts may change coverage; the survey-based occupational stocks and flows are imprecise for small groups; the output proxy uses industry composition; and fixed matching parameters can miss changes in recruitment efficiency during the recovery. The resampling intervals describe the uncertainty represented by the existing resampling procedure, not all uncertainty about the taxonomy, coverage or model assumptions. Appendix D reports public search interest as background.

The upward skill-cluster path motivates a policy perspective that extends beyond temporary pandemic adjustment. Within the period observed, the allocation problem begins before the pandemic and is not fully reversed when its largest fluctuations subside. This gives a reason to assess continuing provision for reskilling, upskilling and occupational transitions alongside measures intended to manage short-lived disruption. A falling broad SOC index alone is insufficient evidence that the need for such assessment has diminished.

The two grouping approaches offer different information for that task. Required-skill clusters identify occupations with similar stated skill bundles. Career-based domains describe how skill requirements are combined in the careers observed in the data and give a second basis for comparing occupations. Together with occupation-specific requirements, these representations can help frame questions about which existing skills a worker could reuse and what additional training a move might require. They can also connect discussions of training provision to the composition of recruitment demand. The fixed-share results make that demand side relevant, without establishing that vacancies are an exogenous cause of mismatch.

This is a motivation for policy design and evaluation, not an evaluation of a programme. An increasing allocation index cannot distinguish inadequate training from other obstacles to moving into jobs, and the planner benchmark abstracts from the costs of doing so. Programme evidence is therefore necessary to assess which interventions improve outcomes. For example, Card et al. (2018) show that average active labour market programme effects can take time to emerge and differ across programme types and participants, with larger medium-term gains for programmes emphasising human capital accumulation. For training policy, the appropriate next step is to relate the skill-based diagnosis to the costs, feasibility and subsequent employment and earnings outcomes of particular transitions.

7 Conclusion

This paper develops two ways to group occupations using skills: direct clustering of required-skill bundles and grouping through skill domains constructed from workers' career histories. They offer complementary representations of occupational relatedness and can be applied beyond the UK setting when the required crosswalks and histories are available. Their value is to make market boundaries an explicit part of the analysis of worker allocation.

The UK application shows that those boundaries affect both pandemic-period movements and the direction of measured mismatch. Nine skill-based occupation clusters reveal larger movements in productivity-weighted mismatch around the COVID-19 pandemic than the nine SOC major groups. Over 2016–2022, the skill clusters also give an upward descriptive trend, while the SOC groups give a downward one. The skill-cluster increase begins before the pandemic and is only partly reversed after its peak. Detailed occupational mismatch also rises, and fixed-share calculations show that its time profile is particularly sensitive to the composition of vacancies. These results challenge the use of a stable or falling broad occupational index as sufficient evidence that skill-related allocation difficulties are easing. The fractional skill-domain exercise gives a separate view of mismatch when occupations contribute to several domains.

The findings are relevant to the UK's continuing productivity difficulties because the allocation of new employment across more and less productive activities is one channel through which mismatch can matter. They show that assessing this channel requires attention to market definition as well as data and matching-efficiency estimates.

For policy, the results motivate sustained attention to retraining, skill development and support for worker reallocation. The two occupational representations provide ways to organise information about shared skills and possible transitions as recruitment demand changes. Turning that information into effective policy requires evidence on the additional skills workers need, the costs of acquiring them and the outcomes of specific interventions. The contribution here is to improve the diagnosis on which those decisions depend.

A Derivation of the mismatch index

Write $x_{it} = z_{it}\phi_i$ and $D_t = \sum_i v_{it} x_{it}^{1/\alpha}$. Differentiating the objective in equation (2) gives $\lambda(1-\alpha)x_{it}(v_{it}/u_{it}^*)^{\alpha}$ equal across active markets. Hence u_{it}^* is proportional to $v_{it}x_{it}^{1/\alpha}$. Imposing $\sum_i u_{it}^* = u_t$ yields equation (4).

Substitution into the matching function gives

$$y_t^* = \lambda \sum_i x_{it} v_{it}^{\alpha} \left(\frac{u_t v_{it} x_{it}^{1/\alpha}}{D_t} \right)^{1-\alpha} = \lambda u_t^{1-\alpha} D_t^{\alpha},$$
$$y_t = \lambda v_t^{\alpha} u_t^{1-\alpha} \sum_i x_{it} (v_{it}/v_t)^{\alpha} (u_{it}/u_t)^{1-\alpha}.$$

Since $D_t = v_t \sum_i x_{it}^{1/\alpha}(v_{it}/v_t)$, their ratio gives equation (5). The nonnegative gap follows from maximising the same weighted objective over an allocation set containing the observed allocation, subject to consistent positive inputs.

This benchmark is deliberately limited to the allocation problem used in the empirical index. A fully dynamic planner requires assumptions about continuation values, destruction, entry and the evolution of productivity. Those additional objects are not needed to calculate the conditional index reported here. Maximising total hires instead would set the productivity weight to a common value; the hires-maximising and output-maximising allocations need not coincide.

B Additional data and occupational estimates

B.1 Data coverage and productivity summaries

Table B.1 locates the data relative to earlier studies. Detailed online adverts permit regrouping by occupation, while worker histories supply non-employment-to-employment flows and mobility information. Coverage and sampling differ from administrative sources and the Labour Force Survey, so more detail does not automatically imply smaller measurement error.

Table B.1: Data sources in related mismatch studies

Study	Period	Vacancies	Jobseekers	Productivity	Hires for matching-function estimation
Patterson et al.	2007–2012	Jobcentre Plus	Jobseeker's Allowance claimant counts	ONS industry GVA; LFS occupation information	Claimant-count outflows (NOMIS); JCP vacancy outflows
Turrell et al.	2008–2017	Reed	LFS	ONS industry GVA; LFS occupation information	LFS flows
Pizzinelli and Shibata	US: 2005–2020 UK: 2002–2020	US: Job Openings; UK: Indeed, ONS vacancy survey	US: Current Population Survey; UK: quarterly LFS	Not applicable	US: Labour Turnover Survey; UK: LFS
This paper	2016–2022	Adzuna	UKHLS	ONS industry GVA; UKHLS occupation information	UKHLS flows

The productivity figures and ranking describe the industry-based proxy in equation (6). In particular, low measured values for health, education and care occupations should not be interpreted as low social returns or low individual ability. The construction attributes industry output per worker to occupations through industry employment composition.

B.2 Occupational matching-efficiency estimates

The original occupational estimates are retained in Table B.3 and Figures B.3–B.5. Dispersion increases at finer occupational levels, with several large and imprecise effects. Differences in group composition, sample size and the definition of hires matter for com-

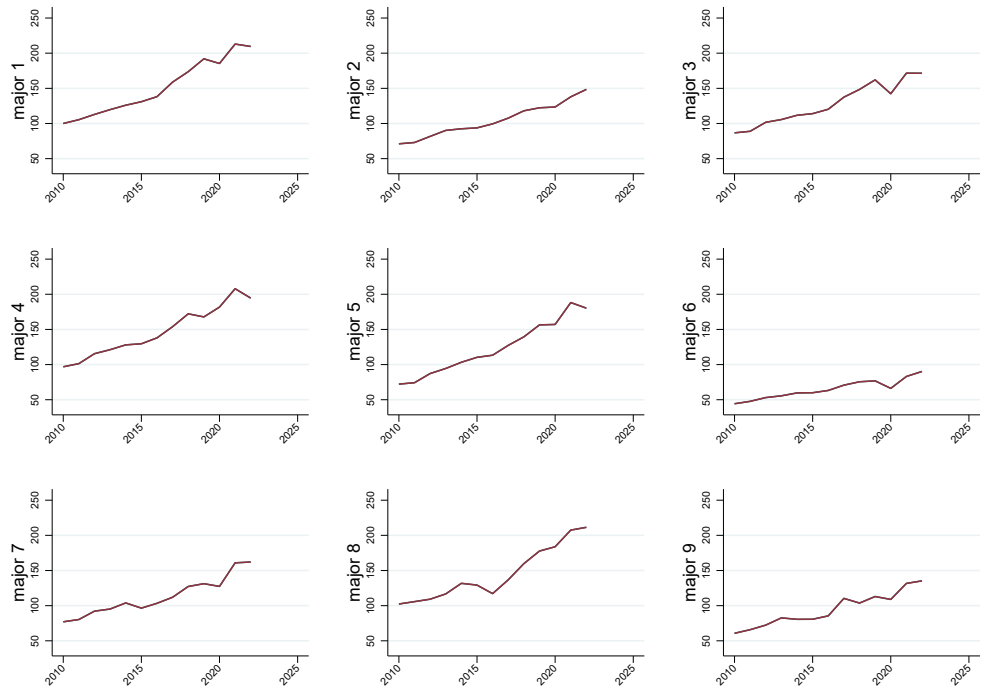
Table B.2: Occupational productivity rankings

Ten 3-digit SOC2000 with lowest productivity	
612	Childcare and related personal services
244	Public service professionals
231	Teaching professionals
611	Healthcare and related personal services
344	Sports and fitness occupations
322	Therapists
321	Health associate professionals
245	Librarians and related professionals
221	Health professionals
925	Elementary sales occupations

Ten 3-digit SOC2000 with highest productivity	
812	Plant and machine operatives
712	Sales related occupations
354	Sales and related associate professionals
814	Construction operatives
243	Architects, town planners, surveyors
123	Managers and proprietors in other service industries
412	Administrative occupations: Finance
353	Business and finance associate professionals
822	Mobile machine drivers and operatives
112	Production managers

Ten occupations with the largest relative productivity increase, 2010–2020	
524	Electrical trades
813	Assemblers and routine operatives
114	Quality and customer care managers
112	Production managers
913	Elementary process plant occupations
121	Managers in farming, horticulture, forestry and fishing
212	Engineering professionals
924	Elementary security occupations
511	Agricultural trades
925	Elementary sales occupations

Figure B.1: Occupational productivity by 1-digit SOC2000 in the UK, 2010-2022



Note:
[1] Occupational GVA per worker is shown for SOC2000 major groups over 2010–2022 on a normalised index scale. The common scaling allows relative comparisons across groups.
SOC major-group codes follow Figure 2.

parisons across classifications and studies. UKHLS hires exclude job-to-job transitions in this calculation.

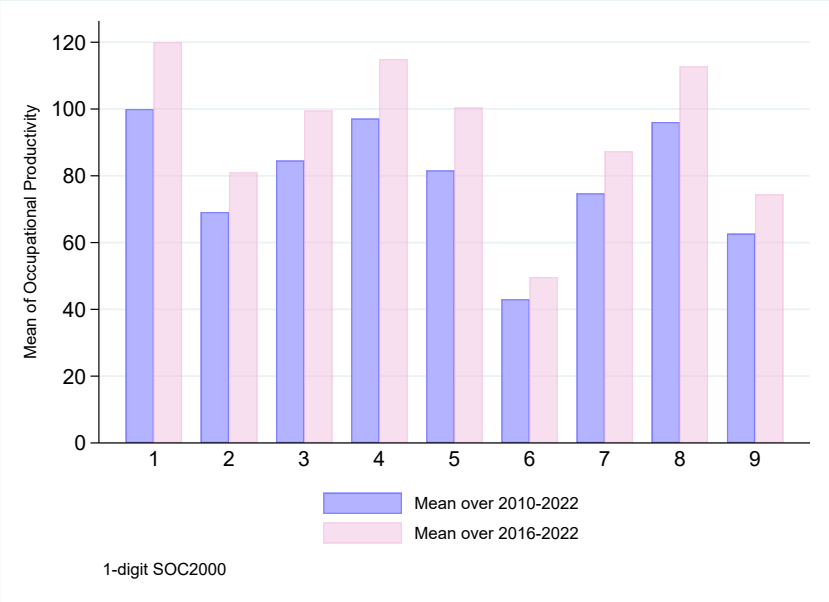
Table B.3: Matching efficiency estimates.

SOC digit	mean of ϕ_i	sd of ϕ_i	min of ϕ_i	max of ϕ_i
1	1.04	0.47	0.54	1.94
2	1.19	0.83	0.41	4.42
3	4.50	8.78	0.27	52.54

B.3 Skill taxonomy and clustering diagnostics

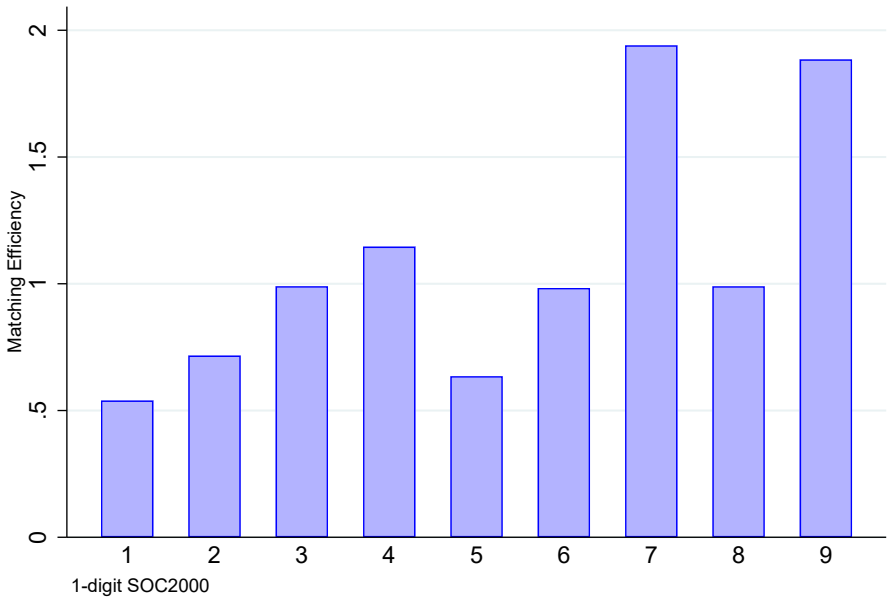
Tables B.4 and B.5 describe the fixed skill taxonomy. Figure B.6 uses individual skill concepts (level 4), whereas the main occupation clustering uses level-3 indicators. The level-4 distances range from 0 to 41, with a mean of 16.26 and standard deviation of 5.16. This finer-resolution illustration should not be confused with the input vectors used for

Figure B.2: Mean productivity by SOC2000 major group



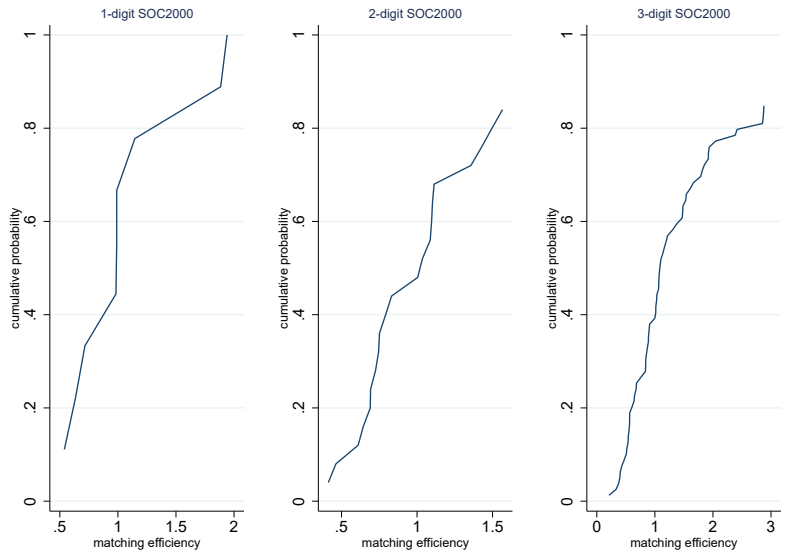
Note:
[1] The lilac and pink bars report mean occupational productivity for SOC2000 major groups over 2010–2022 and 2016–2022, respectively, on a normalised index scale.
SOC major-group codes follow Figure 2.

Figure B.3: Estimated matching efficiency for SOC2000 major groups



Note:
SOC major-group codes follow Figure 2.

Figure B.4: Distributions of matching efficiency by occupational aggregation

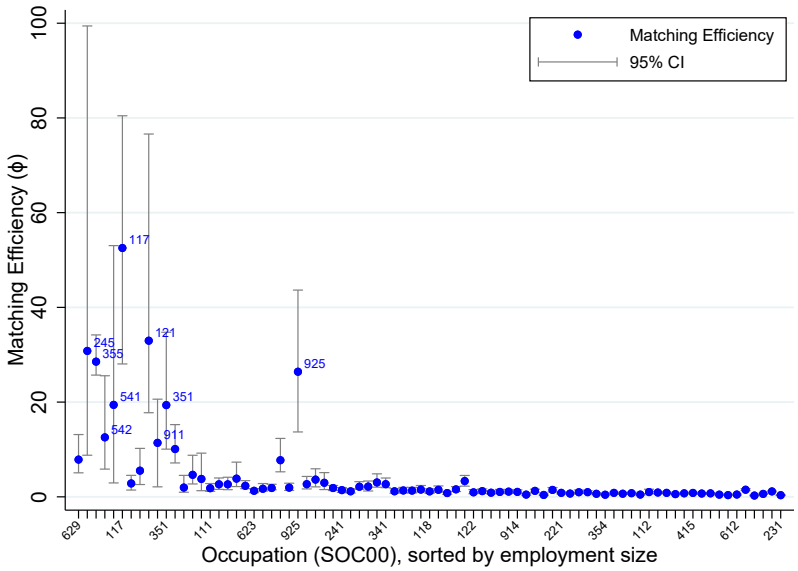


Note:
[1] Data source: UKHLS, Adzuna.
[2] Occupation classification: SOC2000 major groups, two-digit groups and three-digit occupations. The finer-level distributions are truncated as described below.
[3] For the 2-digit and 3-digit SOC2000 classifications, matching efficiencies are highly concentrated at lower levels. To improve visualisation, only values up to the 85th percentile are plotted.

the main partition.

The elbow-style diagnostic groups scalar pairwise distances. Its fit can describe that distribution, but it is not a model-selection criterion for k-means on occupation-by-skill vectors. The main choice of nine occupation clusters is justified by equal-count comparison with SOC, not by a claim of optimality from this graph.

Figure B.5: Matching efficiency by three-digit SOC2000 occupation



Note:
[1] Data source: UKHLS, Adzuna. The figure reports occupational matching efficiency estimates.
[2] Occupation classification: 3-digit SOC2000, 80 groups.
[3] Occupations with large confidence intervals: 117 Protective service officers; 121 Managers in farming, horticulture, forestry and fishing; 245 Librarians and related professionals; 351 Transport associate professionals; 355 Conservation associate professionals; 541 Textiles and garments trades; 542 Printing trades; 911 Elementary agricultural occupations; 925 Elementary sales occupations.

Table B.4: The skill hierarchy in ESCO v1.2.0, May 2024

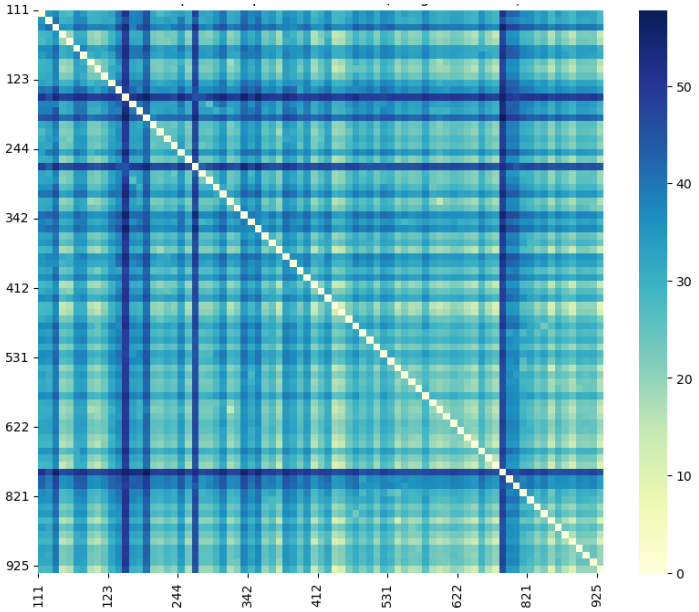
Disaggregation level	No. of skill groups	Examples
level 0	4	knowledge language skills and knowledge skills transversal skills
level 1	28	arts and humanities communication, collaboration and creativity information skills life skills and competences
level 2	156	humanities advising and consulting monitoring, inspecting and testing applying health-related skills and competences
level 3	452	philosophy and ethics advising on workplace health and safety issues inspecting food safety and quality testing vehicles
level 4	13,939	data ethics Albanian advise customers on appropriate pet care assess trim of vessels analyse milled grains follow hygienic practices in fishery operations

Note:
[1] This paper uses ESCO v1.2.0, released in May 2024, as a fixed occupational skill classification.
[2] The second column reports the number of concepts at each level; the third gives examples.

Table B.5: The number of skills associated with each occupation across different levels of skill disaggregation

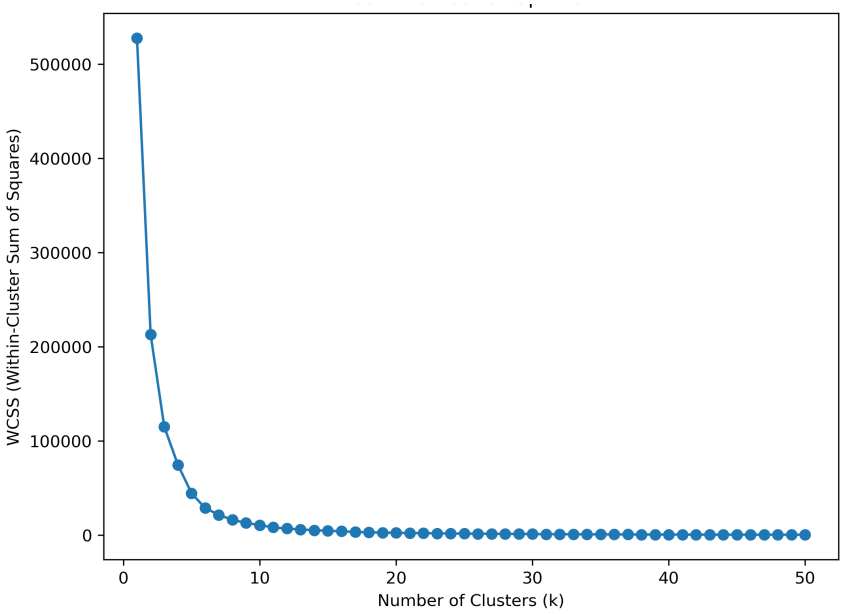
skill disaggregation level	No. of associated skills with one occupation			
	range	mean	sd	75th percentile
level 1	[4,25]	13	3.2	15
level 2	[9,79]	41	15.3	51
level 3	[9,201]	81	41.5	105
level 4	[15,1145]	284	218.4	424

Figure B.6: Occupational Euclidean distance (3-digit SOC2000)



Note:
[1] Data source: ESCO,
[2] Occupations are ordered by SOC2000.

Figure B.7: K-means fit to the distribution of pairwise distances



Note:
Each observation in this diagnostic is a scalar pairwise distance. The graph plots within-cluster sum of squares against the number of groups fitted to those distance values. The occupation partition uses nine groups for comparability with SOC major groups.

B.4 Complete occupation-cluster assignments

Table B.6: Cluster assignments of occupations with SOC2000 codes

Cluster 1 – Design, Inspection and Craft Trades	
243	Architects, town planners, surveyors
312	Draughtspersons and building inspectors
342	Design associate professionals
522	Metal machining, fitting and instrument making trades
524	Electrical trades
541	Textiles and garments trades
549	Skilled trades N.E.C.
Cluster 2 – Recreation, Leisure and Transport Services	
344	Sports and fitness occupations
621	Leisure and travel service occupations
821	Transport drivers and operatives
Cluster 3 – Management, Professional and Administrative Services	
114	Quality and customer care managers
115	Financial institution and office managers
118	Health and social services managers
121	Managers in farming, horticulture, forestry and fishing
122	Managers and proprietors in hospitality and leisure services
213	Information and communication technology professionals
232	Research professionals
241	Legal professionals

242	Business and statistical professionals
245	Librarians and related professionals
352	Legal associate professionals
411	Administrative occupations: Government and related organisations
412	Administrative occupations: Finance
421	Secretarial and related occupations
924	Elementary security occupations
Cluster 4 – Production, Logistics and Land-Based Operations	
112	Production managers
351	Transport associate professionals
355	Conservation associate professionals
511	Agricultural trades
812	Plant and machine operatives
Cluster 5 – Health Care Practitioners and Support Services	
221	Health professionals
321	Health associate professionals
322	Therapists
611	Healthcare and related personal services
Cluster 6 – Construction and Industrial Trades and Operatives	
521	Metal forming, welding and related trades
523	Vehicle trades
531	Construction trades
532	Building trades
542	Printing trades
814	Construction operatives
822	Mobile machine drivers and operatives
911	Elementary agricultural occupations
912	Elementary construction occupations
913	Elementary process plant occupations
914	Elementary goods storage occupations
922	Elementary personal services occupations
923	Elementary cleaning occupations
Cluster 7 – Science, Engineering and Technical Operations	
211	Science professionals
212	Engineering professionals
311	Science and engineering technicians
811	Process operatives
813	Assemblers and routine operatives
Cluster 8 – Customer, Personal and Administrative Services	
313	IT service delivery occupations
323	Social welfare associate professionals
414	Administrative occupations: Communications
415	Administrative occupations: General
543	Food preparation trades
612	Childcare and related personal services
613	Animal care services
622	Hairdressers and related occupations
623	Housekeeping occupations
629	Personal services occupations N.E.C.
712	Sales related occupations

721	Customer service occupations
921	Elementary administration occupations
925	Elementary sales occupations
Cluster 9 – Leadership and Public-Facing Professional and Associate Roles	
111	Corporate managers and senior officials
113	Functional managers
116	Managers in distribution, storage and retailing
117	Protective service officers
123	Managers and proprietors in other service industries
231	Teaching professionals
244	Public service professionals
331	Protective service occupations
341	Artistic and literary occupations
343	Media associate professionals
353	Business and finance associate professionals
354	Sales and related associate professionals
356	Public service and other associate professionals
413	Administrative occupations: Records
711	Sales assistants and retail cashiers

C Additional matching-function estimates

C.1 Alternative measure of jobseekers

The baseline counts unemployment and inactivity equally. The alternative weights inactivity by its job-finding rate relative to unemployment. Table C.1 reports the resulting elasticities for April 2016–December 2020. The estimates are higher than the baseline in Table 4; the weighting also reduces the effective occupational sample.

Table C.1: Matching elasticity under the alternative job-seeker definition

	Elasticity α		Cross-sections		Observations	
	(1) Quartic	(2) Seasonal	(1)	(2)	(1)	(2)
Aggregate	0.788***	0.789***	1	1	56	56
	(0.0974)	(0.0868)				
2-digit SOC2000	0.805***	0.812***	25	25	792	792
	(0.0309)	(0.0308)				
3-digit SOC2000	0.780***	0.780***	74	74	1,088	1,088
	(0.0280)	(0.0283)				

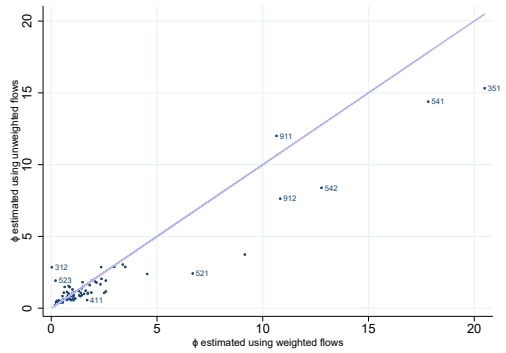
Note:
[1] Column (1) includes a quartic time trend; column (2) includes quarterly seasonal indicators. Occupational regressions include occupation fixed effects.
[2] Standard errors are in parentheses. *** denotes significance at the 1% level. Standard errors use the conventional OLS or within-estimator variance formula.
[3] The estimation window is April 2016–December 2020, excluding December 2019. Observations enter each model when the job-finding-rate weight and logarithms are defined.

Figure C.1 compares occupational matching efficiency under the two jobseeker definitions. Most estimates lie near the equality line, with larger differences for a small number of occupations. Table C.1 reports the effective sample for the weighted regressions.

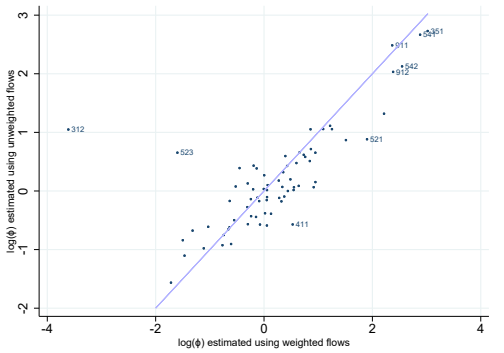
C.2 Estimates over different periods

Figure C.2 compares full-window estimates for April 2016–December 2020 with estimates for April 2016–June 2018 and July 2018–December 2020. The plots show how occupational matching-efficiency estimates vary across these two subperiods.

Figure C.1: Matching efficiency under alternative jobseeker definitions



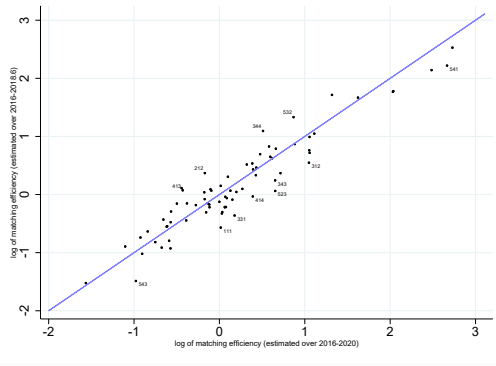
Matching-efficiency estimates



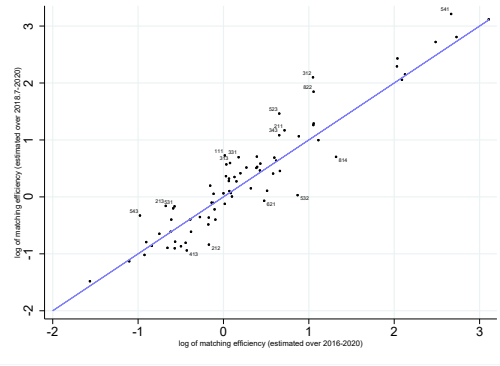
Log matching-efficiency estimates

Note:
[1] Data source: UKHLS, Adzuna. Occupation classification: three-digit SOC2000. Effective estimation samples are reported in Tables 4 and C.1.
[2] The x-axis is the matching efficiency estimations using flows from unemployment and inactive people weighted by job finding rates. The y-axis is the estimations using the unweighted data.
[3] This figure labels selected occupations based on their matching efficiency. These occupations are included either because they exhibit a high matching efficiency (greater than 10) or because the ratio of matching efficiency between the two measures (weighted by unemployment and inactivity versus unweighted) exceeds 2. They are listed below:
occupation 312: "Draughtspersons and building inspectors"; occupation 351: "Transport associate professionals"; occupation 411: "Administrative occupations: Government and related organisations"; occupation 521: "Metal forming, welding and related trades"; occupation 523: "Vehicle trades"; occupation 541: "Textiles and garments trades"; occupation 542: "Printing trades"; occupation 911: "Elementary agricultural occupations"; occupation 912: "Elementary construction occupations";
[4] The solid diagonal is the 45-degree equality line.

Figure C.2: Matching efficiency across estimation periods



2016.4-2020.12 vs 2016.4-2018.6



2016.4-2020.12 vs 2018.7-2020.12

Note:
[1] Data source: UKHLS, Adzuna. Occupation classification: three-digit SOC2000; baseline sample of 80 groups and 2,123 observations.
[2] The x-axis reports log matching efficiency for April 2016–December 2020. The left and right y-axes report log estimates for April 2016–June 2018 and July 2018–December 2020, respectively.
[3] This figure labels selected occupations based on their matching efficiency. These occupations are included because the ratio of matching efficiency between the two measures exceeds 1.5. They are listed below:
occupation 111: "Corporate managers and senior officials"; occupation 211: "Science professionals"; occupation 212: "Engineering professionals"; occupation 213: "Information and communication technology professionals"; occupation 312: "Draughtspersons and building inspectors"; occupation 313: "IT service delivery occupations"; occupation 331: "Protective service occupations"; occupation 343: "Media associate professionals"; occupation 344: "Sports and fitness occupations"; occupation 413: "Administrative occupations: Records"; occupation 414: "Administrative occupations: Communications"; occupation 523: "Vehicle trades"; occupation 531: "Construction trades"; occupation 532: "Building trades"; occupation 541: "Textiles and garments trades"; occupation 543: "Food preparation trades"; occupation 621: "Leisure and travel service occupations"; occupation 814: "Construction operatives"; occupation 822: "Mobile machine drivers and operatives".
[4] The solid diagonal is the 45-degree equality line.

D Google Trends: labour shortages and skill acquisition

This appendix reports UK search interest in labour shortages, reskilling and upskilling over 2016–2022. The series describe public attention to these terms during the study period.

D.1 Data and interpretation

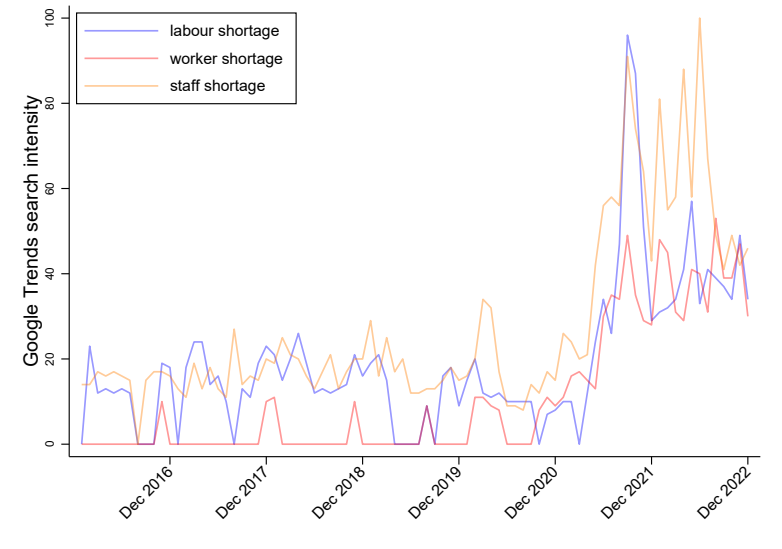
Google Trends scales search interest from 0 to 100 within each query set and period. Each series is interpreted as a change in attention to the selected terms. Separately normalised query sets have different scales.

The first set of queries concerns labour, worker and staff shortages. The second concerns reskilling and upskilling. These terms capture different aspects of public discussion about recruitment and skill acquisition.

D.2 Labour shortage searches

Figure D.1 shows increased attention to shortage-related terms during parts of the reopening period. The movements reflect public interest, which can respond to recruitment conditions as well as news and policy debate.

Figure D.1: Google Trends search intensity for labour shortages in the UK, 2016–2022

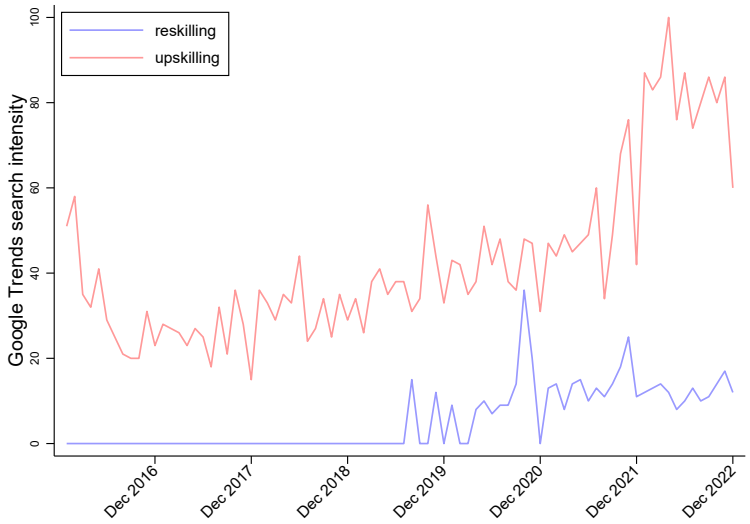


Note:
[1] Data source: Google Trends;
[2] Weekly Google Trends indices, normalised to a maximum of 100 within each query set. The series measure relative search interest.

D.3 Reskilling and upskilling searches

Figure D.2 shows searches for reskilling and upskilling. Search interest in these terms increases during the sample, indicating greater attention to skill acquisition and occupational change.

Figure D.2: Google Trends search intensity for reskilling/upskilling in the UK, 2016–2022



Note:
[1] Data source: Google Trends;
[2] Weekly Google Trends indices, normalised to a maximum of 100 within each query set. The series measure relative search interest.

References

Adams-Prassl, A., Boneva, T., Golin, M. and Rauh, C. (2020). Furloughing. *Fiscal Studies*, 41(3), 591–622.

Blundell, R., Costa Dias, M., Joyce, R. and Xu, X. (2020). Covid-19 and inequalities. *Fiscal Studies*, 41(2), 291–319.

Card, D., Kluve, J. and Weber, A. (2018). What works? A meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association*, 16(3), 894–931. https://doi.org/10.1093/jeea/jvx028.

Carrillo-Tudela, C., Hobijn, B., She, P. and Visschers, L. (2016). The extent and cyclicality of career changes: Evidence for the U.K. *European Economic Review*, 84, 18–41.

Diamond, P. (2011). Unemployment, vacancies, wages. *American Economic Review*, 101(4), 1045–1072.

Elsby, M. W. L., Hobijn, B. and Şahin, A. (2013). Unemployment dynamics in the OECD. *Review of Economics and Statistics*, 95(2), 530–548.

Forsythe, E., Kahn, L. B., Lange, F. and Wiczer, D. (2020). Labor demand in the time of COVID-19: Evidence from vacancy postings and UI claims. *Journal of Public Economics*, 189, 104238.

Forsythe, E., Kahn, L. B., Lange, F. and Wiczer, D. (2022). Where have all the workers gone? Recalls, retirements, and reallocation in the COVID recovery. *Labour Economics*, 78, 102251.

Gathmann, C. and Schönberg, U. (2010). How general is human capital? A task-based approach. *Journal of Labor Economics*, 28(1), 1–49.

Handel, M. J. (2016). The O*NET content model: Strengths and limitations. *Journal for Labour Market Research*, 49(2), 157–176.

Joyce, R., Postel-Vinay, F., Spittal, P. and Xu, X. (2022). *Job opportunities after the pandemic*. Institute for Fiscal Studies, April.

Lazear, E. P. (2009). Firm-specific human capital: A skill-weights approach. *Journal of Political Economy*, 117(5), 914–940.

Lise, J. and Postel-Vinay, F. (2020). Multidimensional skills, sorting, and human capital accumulation. *American Economic Review*, 110(8), 2328–2376.

Neffke, F. and Henning, M. (2013). Skill relatedness and firm diversification. *Strategic Management Journal*, 34(3), 297–316.

Patterson, C., Şahin, A., Topa, G. and Violante, G. L. (2016). Working hard in the wrong place: A mismatch-based explanation to the UK productivity puzzle. *European Economic Review*, 84, 42–56.

Pizzinelli, C. and Shibata, I. (2023). Has COVID-19 induced labor market mismatch? Evidence from the US and the UK. *Labour Economics*, 81, 102329.

Poletaev, M. and Robinson, C. (2008). Human capital specificity: Evidence from the Dictionary of Occupational Titles and displaced worker surveys, 1984–2000. *Journal of Labor Economics*, 26(3), 387–420.

Postel-Vinay, F. and Sepahsalari, A. (2023). Labour mobility and earnings in the UK, 1992–2017. *The Economic Journal*, 133(656), 3071–3098.

Şahin, A., Song, J., Topa, G. and Violante, G. L. (2014). Mismatch unemployment. *American Economic Review*, 104(11), 3529–3564.

Traag, V. A., Waltman, L. and van Eck, N. J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports*, 9, 5233. https://doi.org/10.1038/s41598-019-41695-z.

Turrell, A., Speigner, B., Copple, D., Djumalieva, J. and Thurgood, J. (2021). Is the UK's productivity puzzle mostly driven by occupational mismatch? An analysis using big data on job vacancies. *Labour Economics*, 71, 102013.

Walter, S., Eum, W. and Lee, J.-D. (2024). The European skill space: A cross-country analysis of path-dependent capability development. *Baltic Journal of Economics*, 24(1), 1–28.

Office for National Statistics (2022). *Comparison of labour market data sources*, Annex 6: the Vacancy Survey compared with Adzuna. Methodological publication, revised 27 April 2022. https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/comparisonoflabourmarketdatasources.