The Worker's Brain Under the Expert's Hat: Occupation Coding, Tacit Knowledge, and Wages

Emma Du Puy, Gina Markov, and Suresh Naidu[1]

Columbia University

Preliminary and Incomplete. Please do not cite or circulate.

Abstract

Economists and policy makers rely on occupational codings, especially O*NET, for a wide variety of insights into the labor market. Yet it is widely recognized that many jobs are difficult to codify, and require embodied intuitions and past on-the-job experiences to perform successfully. We exploit a little-known feature of the occupational codes in O*NET, where experts and workers are exogenously assigned to code occupational characteristics for the same occupation over time, to construct a measure of occupation-level "tacit knowledge". For a given occupation, switching from a worker rater to an expert rater leads to a significantly different O*NET rating for a variety of skills and task intensities. We document that many prior uses of O*NET by economists significantly differ depending on whether worker or expert ratings are used. Turning to policy, O*NET is extensively used in guest worker visa approvals; we show that switches in coder identity change approval rates for H-1B applications, as experts code jobs as less experience-intensive and thus less specialized, leading to higher rejections. We aggregate the disagreement between experts and workers into a measure of "tacit jobs", and show these have a number of distinctive properties, including less AI exposure and lower skill ratings following a 2010 reform that replaced incumbent worker respondents with analysis. We then focus on comparing the coder differences with ground-truth measures. Workers better understand the injury and experience requirements of the job, while there is no difference between workers and experts in rating the returns to formal education requirements. Finally, we embed these ground truth and O*NET measures together into a structural hedonic model with incomplete information, and study the extent to which tacit information is priced into wages at equilibrium. Counterfactual analysis reveals that tacit information manifests as an upward pressure on wages, by pricing workers' perceived injury risk at work and associated compensating differentials. Construction workers, who have the highest tacit valuation of injury risk, see their wages increase by 3.1% from the inclusion of tacit information. However, compensating differentials for tacit information about disamenities are attenuated by labor market frictions: in our baseline equilibrium with estimated frictions, the average wage is left unchanged by tacit information. Under counterfactual frictionless markets, eliminating the pricing of tacit information would reduce the average wage by 9.5%.

[1]We thank participants at seminars at Princeton and LSE for comments.

1 Introduction

Economists and policy makers rely on detailed occupational codes to understand and regulate the labor market. In the United States, the Occupational Information Network (O*NET) is widely used to operationalize "job content". Each occupation is represented as a long vector of skills, requirements, tasks, and working conditions, recovered from surveys of workers and expert coders. These measures are then taken as stable, invariant occupational characteristics, and are used in policies such as prevailing wages, work visa approvals, and training priorities. Economists use the same variables to define labor markets, to study the role of skills, and to assess the impact of shocks like technology and trade on inequality and employment. Computer science researchers have also increasingly adopted O*NET as benchmarks for a variety of AI tasks. Confidence in these measures is generally high, and papers heavily reliant on O*NET are widely cited.

But much knowledge about jobs is "tacit": difficult to describe and understand outside of the specific context. Foundational writers of the 20th century, such as Friedrich Hayek, Michael Polanyi, and James Scott all argued that much useful knowledge is difficult to write down or even articulate in language. Tacit knowledge at work resists codification, depends on firm or relationship-specific contexts, and is generally accumulated via experience and learning, rather than from formal training. Accurately capturing the tacit characteristics of a job with a standardized survey like O*NET may be quite difficult.

Even if tacit knowledge is prevalent in the labor market, part of the value of competitive markets is that they are efficient aggregators of exactly this kind of dispersed, fundamentally private information (Hayek, 1945). According to this view, economically important knowledge that is not easily described or shared is nonetheless reflected in skill premia and compensating differentials. Yet pervasive labor market frictions may disrupt this information-aggregation property of markets, creating roles for regulations such as health and safety standards and pay scales, and necessitating measurement tools like O*NET for policy makers. Whether these measurement tools are able to accurately encode the private information

that is used by firms and workers to make decisions, and whether this information is priced in wages via competitive equilibrium, are open empirical questions.

To study the nature of tacitness in the workplace, we study the differences between what incumbent workers in an occupation report to O*NET and what "occupational experts", who are often managers or industry representatives at some remove from the job, report. We find these differences are consistent with common definitions of tacit knowledge. In particular, the gap is smaller for the requisite skills that are easily understood and explicit (i.e. an educational requirement or credential) and larger for the necessary skills that can only be understood via close personal experience. In this sense, workers "may know more than they can tell" their supervisors and seniors. (Polanyi, 1958)

We leverage measurement variation in the coders for the same occupation across O*NET updates to quantify the scope of tacit knowledge in the labor market. We first present causal evidence that the values of an occupation's activity, skill, and task content are not invariant to the coder identity. We use a little-known institutional rule in O*NET, where "experts" are used to code an occupation when an insufficiently large sample of workers responds to the survey, to look at the effect of coder identity (worker versus expert) on a large set of O*NET values. According to O*NET, "Experts" tend to be managers, trainers, supervisors, or representatives of professional organizations related to the occupation they code. Experts are generally not currently doing the job being coded, even if they have past experience in it. Qualitative evidence suggests that bureaucratic organizations with layers of management may not be able to faithfully relay information across hierarchical layers. A chemical worker, quoted in Braverman (1974), explains, " It takes skill to be an operator. Maybe you've heard of this job-evaluation program that's been going on. Well, our supervisor thinks there's not much skill in our work. The way he described our jobs for the job-evaluation program, it's like he thinks you could train a bunch of chimps and they could do the job. He thinks we're a bunch of idiots. That has caused unhappy feelings." In this light, contrasting experts' O*NET ratings with current workers' ratings for a given occupation is a broad context in

which to study the prevalence and pricing of tacit information.

Isolating and exploiting this exogenous sampling variation, we show that experts systematically diverge from workers on key dimensions of O*NET. Experts underrate the importance of on-the-job experience, while overrating the importance of education. Experts and workers also differ in their ratings of the social skills and physical hardship on the job. To dig deeper into the specific dimensions of jobs where workers and experts differ, Table 1 previews our approach to quantifying tacit job dimensions for a select set of occupations. The table highlights two aspects of tacit knowledge: knowledge related to the skills required to perform an occupation, and knowledge regarding occupational disamenities (particularly, the physical conditions of the job). These two dimensions are key to the economic understanding of occupations, as they inform the measurement of returns to skills and compensating differentials.

For each O*NET skill variable and O*NET disamenity variable, we compute the difference between the worker and expert values for the same occupation.[2] Table 1 presents the occupations with the largest differences, revealing the nature of tacit information as proxied by our worker-expert difference measure. Skills-related disagreement is most pronounced when it comes to the selection, use and maintenance of occupation-specific material. Experts disagree with workers in the intensity of disamenities that are crucial to occupations; for example, they relatively under-rate the exposure to hazardous equipment for gas plant operators, and the exposure to radiation for magnetic resonance imaging technologists. Both the skill and disamenity dimensions highlight that workers and experts tend to disagree most on situated knowledge, i.e. physical and sensory knowledge that is hard to grasp without personal experience performing the occupation at hand. We also note the prevalence of blue-collar occupations in these high-disagreement skills and disamenities. When we construct an occupation-level measure of tacitness across all O*NET dimensions, we confirm this trend, and we find that some of the least tacit occupations are managerial and professional in

[2]The skill variables are the ones listed in the O*NET Skills module, while the disamenity variables are described in Appendix A.1.

nature—reflecting experts' identity as managers with long tenures.

In the first half of the paper, we show that these coder differences matter for research on the labor market; we show 6 out of 10 different projections of O*NET used in the economics literature significantly differ when expert codings are used instead of workers', including predictions about which occupations are suitable for machine learning. Changes in task intensity, and employment and wages by task-intensity, between 1980 and 2020 are meaningfully different depending on whether workers or expert codings are used. Using the institutional role of O*NET in approving H-1B visas, we show these expert vs worker differences matter for real-world wage setting. When an H-1B eligible occupation is coded by an expert rather than a worker, the importance of experience goes down, and the visa is less likely to qualify as "specialty", and so the probability of approval goes down. Conditional on approval, the expert-coded occupations are associated with both a lower prevailing wage assigned by DOL to the occupation, and a lower wage reported by the firm for similarly employed workers. These decreases imply final H-1B offered wages can be set lower when the associated O*NET information is coded by an expert. They also suggest that it is the relatively skilled jobs within the occupation that are getting rejected.

In the second half of the paper, we take seriously the interpretation of differences between workers and experts as tacit information about skills and amenities, focusing on measures where there is external data: injury risk and returns to education and experience. We show that worker assessments of overall skills, returns to experience, and injury risk more closely track empirical analogues, with no difference in returns to formal education. We then embed these measures into a structural hedonic model with private information. We build on the large mechanism design literature that models decentralized tacit information as private information that needs to be incentivized to be revealed (Maskin, 2008). We incorporate firm and worker private information on amenities and productivity into a two-sided hedonic equilibrium model following (Dupuy and Galichon, 2022). The key novelty is that the surplus of a match can be replaced with the "virtual surplus" adjusted for truthful

information revelation in a post-match optimal contract. The worker and expert values of O*NET enter as proxies for private information in the model as tacit knowledge. Since the equilibrium wages do reflect the O*NET differences, even conditional on our proxies for ground truth, we conclude that the labor market does indeed price tacit information of workers and firms. However, our model also yield estimates of large, and asymmetric, labor market frictions (unpriced but observable heterogeneity due to tastes or search costs), which suggest that the tacit information of workers is less reflected in wages and employment than tacit information of managers/firms. Counterfactual analysis reveals that tacit information manifests mostly as an upward pressure on wages, by pricing workers' perceived injury risk at work and associated compensating differentials. Construction workers, which have the highest tacit valuation of injury risk, see their wages increase by 3.1% from the inclusion of tacit information. However, compensating differentials for tacit information related to disamenities are attenuated by labor market frictions: accounting for frictions, the average wage is left unchanged by tacit information. Under counterfactual frictionless markets, eliminating the pricing of tacit information would reduce wages on average by 9.5%.

In terms of focusing on endogenous construction of administrative categories, a recent paper by Grant (2024) is closest to us. He studies endogenous categorization of trade codes in response to tariff changes. In contrast to us, he studies the optimal construction of product codes, while we study how different measures of the same occupation capture heterogeneous beliefs about a job. In discussing the endogeneity of occupational categories, we build on a large literature in sociology showing the endogenous definition of skill (Attewell, 1990) and expertise (Abbott, 1988). We are also not the first to highlight the omissions and errors in official occupation codes; as early as 1980 the National Research Council conducted a critical review of the Dictionary of Occupational Titles, and O*NET has been reviewed similarly in Hilton and Tippins (2010). Handel (2016) describes a wide variety of limitations in the phrasing of O*NET questions, suggesting that O*NET skill descriptions poorly capture what workers actually do on the job.

Within labor economics, we contribute to an enormous literature using O*NET and occupational categories, beginning with Autor et al. (2003). Acemoglu and Autor (2011) introduce the widely used tasks framework, and operationalize it with O*NET as well as its predecessors, and we discuss numerous subsequent papers using O*NET below. Autor and Handel (2013) use microdata to show that tasks vary considerably within occupations, and predict wages even within O*NET occupations. We are also complementing a large literature on the US H-1B visa program, where scholars have studied the assignment process (Pathak et al., 2025), the effects on wages and innovation (Doran et al., 2022) and political effects (Mayda et al., 2018).

Methodologically, we draw on the literature on two-sided hedonic equilibrium in labor markets (Dupuy and Galichon, 2022; Galichon and Salanié, 2022; Corblet and Dupuy, 2024), which has been used to study classic questions of equilibrium sorting and compensating differentials in the labor market (Mas, 2025; Rosen, 1974; Viscusi, 1981). We augment the widely used logit formulation of the model with ex-post bilateral bargaining under incomplete information within a match. Tacit (or private) information is revealed to the worker and firm after the match, but before the contract is made. The determinants of the number of matches are then predictors of "virtual surplus" rather than the actual surplus.

In this last exercise, we build on venerable themes from economic theory on information revelation in markets. Hayek (1945) presented an early view that market prices were ideal signals aggregating decentralized scarcity. Hurwicz (1972) began the mechanism design approach to this problem investigating the role of information revelation. This literature generally finds that competitive equilibrium prices uniquely implement efficiency in Walrasian environments (Makowski and Ostroy, 2001; Nisan and Segal, 2006). However, in situations of bilateral bargaining (e.g. due to hold-up), Myerson and Satterthwaite (1983) show that there is no ex-post efficient mechanism with budget balance. They show that the constrained-optimal contract with budget-balance maximizes the sum of virtual surpluses, where the true surplus is adjusted by the costs of information revelation. We embed

these classic results from mechanism design in our otherwise standard equilibrium model to quantify the contribution of tacit information to the wage structure. We find that much worker private information is about disamenities (i.e. associated with higher wages and lower employment), while manager/expert private information is about productivities (associated with higher wages and higher employment). In counterfactuals we find, for example, that tacit information held by workers about high-injury occupations keep wages high.

The rest of the paper proceeds as follows. Section 2 presents the institutional details on occupational measurement and the construction of O*NET, including the difference between experts and workers that will be crucial for the paper. Section 3 explains the construction and harmonization of O*NET databases over time, as well as the other sources of data that we use in the analysis. Section 4 presents two-way fixed effect and event-study evidence that the identity of the O*NET coder matters for a variety of measures. We first show which O*NET components differ the most between workers and experts, using high dimensional regression, and find these to be on-the-job education and experience, as well as the social and physical dimensions of the job. We then show that many occupational measures aggregated from O*NET in economics papers change value when the coder changes from worker to expert. Section 5 shows that the change in coder identity changes H-1B application patterns, approval rates, and both prevailing and offered wages. Section 6 constructs a measure of occupational tacit knowledge from the weighted difference in expert and worker O*NET measures, and validates this as a measure of tacit knowledge. In particular we show these tacit jobs are less exposed to AI and machine learning and also more likely to be measured as low-skilled when the coder is changed to a professional occupational analysis. Section 7 focuses on assessing the relative correlations of worker and expert codings with labor market data, focusing on cross-occupational correlations of relevant expert and worker O*NET scores with ground truth measures of job injuries, returns to experience and education. Finally, Section 8 embeds the ground truth and ONET measures into a two-sided matching model with frictions and two-sided ex-post private information, and calculates the counterfactual

wage distribution in an economy where tacit information is eliminated. Section 9 concludes.

2 Institutional Background on O*NET

The first Dictionary of Occupational Titles is published in 1939, as an input for 1930s jobs programs and war manpower allocations. Today it is used for guestworker visa allocations, prevailing wage determination, the regionally-specific content of CETA/Job Corps training programs, job search assistance, and disability compensation.

The original Dictionary of Occupational Titles (DOT, 13,000 occupations) was slated to be officially replaced in the 1990s by O*NET by the 1998 Workforce Investment Act. O*NET was then developed by the North Carolina Employment Security Commission via a Department of Labor grant. O*NET was explicitly designed to shift focus from describing industry-specific manufacturing jobs (as in the DOT) to describing what is needed to do a job in a service-oriented economy. It used 6-digit Standard Occupational Classification titles (around 850 codes), originally developed in 1977, with large revisions in 2000, 2010, 2018.

The O*NET Content Model was developed by industrial psychologists (Holland, 1958; Peterson et al., 1999) based on the premise that some kinds of people are best suited to do certain kinds of jobs. The content model was designed with modules that separately captured worker and job characteristics. Appendix Table A1 presents all the different modules of O*NET, together with the number of elements.

The initial 2001 values of O*NET are taken from the Dictionary of Occupational Titles. When a new occupation is added, it is first coded by either incumbent workers, experts, or an analyst. After that, every year, 1/4 of the occupations are generally updated by either incumbent workers or experts, with only a few exceptions where analysts are used for re-coding. Updating entails surveying incumbent workers wherever possible. O*NET begins by sampling firms from Dun and Bradstreet, and then samples workers within those firms, offering 10$ in compensation for completing the survey. Each sampled worker completes

one module of standardized questionnaires (knowledge, activities, context). However, when finding an adequately representative sample of workers (15 for each module, with unclear regional representation) too difficult, O*NET will collect data from experts, who will complete all the modules and be paid 40$. Importantly, this decision is made after sampling is attempted, and thus means that within-occupation variation in expert vs worker sources is driven by sampling noise, and so quasi-random.

An exception to the expert-worker coding comes from the "skills" reform in 2010 which we discuss more extensively in Appendix A.2. Essentially workers *and* experts were deemed insufficiently reliable for coding specific skills required for the job in 2010, and were replaced with Occupational Analysts subsequently. While our primary identifying variation comes from within-occupation switches between workers and experts, for some of our results we also pool experts and analysts. Appendix Figure A5 shows that differences between worker and analyst versus expert codings are highly correlated, so analysts are much closer to experts than they are to workers in their assessments. In Figure A2 we further show that the effects of the 2010 reform were more pronounced for occupations that were graded by incumbent workers, relative to occupational experts, prior to 2010.

Who are the Experts? The 2024 Office of Management and Budget document on O*NET writes on pp. B-32: "With this method, persons considered experts in the target occupation, rather than job incumbents, are surveyed. Occupation experts are sampled from lists provided by source organizations that can include professional associations, certification organizations, industry associations, and other organizations that can identify qualified experts in a given occupation."[3] For example, O*NET documents that the National Association of Business Economists (NABE) and the American Economic Association (AEA) have helped find expert economists. Other professional associations are composed less of incumbent workers, and more of industry representatives, lobbyists, or prominent employers. For example, the professional associations linked to childcare workers include the Association of

[3]See page B-32 of the Supporting Statement B to the 2024 O*NET OMB application.

Early Learning Leaders, which represents directors, administrators, and service providers of early childhood programs.

The variation in the similarity between incumbent workers and association membership suggests that some experts might be very similar to incumbents in their professional experiences, while others might have a more managerial, educational, or otherwise administrative background. Indeed, to qualify as an expert, O*NET requires that they are "someone who has worked in the occupation for at least 1 year" but also someone who "has 5 years of experience as an incumbent, trainer, or supervision.[4] Since 2016, as part of the metadata published about the data collection process, O*NET reports each expert and worker respondent's years of experience in the occupation. Appendix Figure A4 shows how workers' and experts' experience levels are similar for high-experience professions like engineers and economists, while experts are more experienced than workers in higher turnover occupations like research assistants and childcare workers. Since O*NET requires experts to have at least 5 years of experience as an incumbent, trainer, or supervisor, we consider experts in low-experience occupations as potentially former workers who transitioned to managerial or administrative roles.

In light of this definition of experts, we can view the worker-expert divergence in O*NET ratings as experts' attrition in knowledge about the daily experience of the job; we may also expect that experts will over-emphasize managerial or administrative tasks at the expense of salient on-the-ground experiences, including disamenities. We confirm that disamenities and managerial tasks are indeed two significant dimensions of worker-expert difference in Figures 2a and 2b.

3 Data Construction

O*NET: We download all versions of O*NET from August 2000 to August 2024 from `www.onetcenter.org`, and work with Standard Occupation Codes at the database level.

[4]See page B-2 of the Supporting Statement B to the 2024 O*NET OMB application.

For the bulk of our analysis, we focus on occupations that are balanced over time. Unless otherwise specified, all of our analyses use 6 digit SOC codes. Appendix A.1 presents detailed information on O*NET and the specific modules and variables that we use.

Wages: We use the 2002-2021 American Community Survey to construct occupation (and occupation-year) wages and demographics. We restrict our attention to active workers who are not self-employed, nor institutionalized or part of the military. We further focus on workers aged 17 to 65, with declared usual hours of work equal or above 35h, and who worked for the entire past week. Finally, we drop observations with wages below the minimum wage, or that have been top-coded.

Injuries: We compute non-fatal injury rates from the BLS Survey of Occupational Injuries and Illnesses on an occupation o level, averaged from 2011 to 2020.

H-1B visa applications: We use the universe of H-1B Labor Condition Applications for 2009-2023 obtained from the USCIS website.

ALP American Working Conditions Survey (AWCS): We use the ALP's American Working Conditions Survey, fielded in 2015 and 2018, to measure working conditions on the occupation o level.

4 The Effect of Expert Coding on O*NET Measures

In this section we present estimates of the effect of switching sources from incumbent workers to experts.

Identification: Recall that one fourth of occupations are to be updated each year, and if the sample of workers not large or accessible enough (at least 15 responses per module), the source is replaced with an expert. Thus, within an occupation, switches between worker

and expert sources ought to be exogenous, which we can check by examining pre-switch differences in O*NET values in an event-study design.

Variation in Expert, Analyst, and Worker Ratings: Figure 1 displays the variation over time in the share of occupations rated by workers, experts, and analysts, by the various O*NET element types. Most of the variation in workers to analysts comes from the 2010 skill reform, but rarely, analysts will rate an occupation on the other O*NET elements, including work contexts and activities. This occurs for 94 occupations. In Appendix Figure A5, we show that on the occupation-element pairs where we have at least one worker, expert, and analyst rating, a one unit increase in the difference between experts and workers approximately corresponds to a one unit increase in the difference between analysts and workers. Since analysts are biased in similar directions to experts, we include them in the overall "Expert" category for the main analyses, and exclude them in robustness checks.

On which elements do coders differ? We first examine which specific O*NET elements exhibit the largest changes when the source changes from worker to expert. We focus on work requirements, activities, and contexts as the best proxies for skills and amenities on a job and those most commonly used in the literature (and that are available throughout the post-2010 period). We estimate the heterogeneous effects of a switch from a worker ($Expert = 0$) to an expert rating ($Expert = 1$), using within occupation-element variation over time, and within element-time variation across occupations. Due to the high-dimensionality of work contexts and activities, we select the relevant elements using LASSO and report selection-corrected estimates of β_e in Figure 2, clustering standard errors at the occupation level.

$$y_{oet} = \sum_e \beta_e Expert_{ot} \times \delta_e + \delta_{oe} + \delta_{et} + \epsilon_{oet}$$

Figure 2a reveals how experts and workers systematically disagree about physicality and socialization on the job, corroborating the tacitness of these job dimensions. As Figure 2c

shows, experts also under-value the importance of on-the-job experience, which is a metric that requires close familiarity with the job; on the other hand, they over-value educational requirements, which is a more legible and codifiable measure available even to those who may not be current incumbents. Finally, Figure 2b shows that experts over-emphasize the monitoring, managerial, and training aspects of the job. These biases align with experts' distinct identity in these occupations.

These systematic differences across dimensions of occupational measurement motivate our use of the term "tacit" in describing the information gaps between experts and workers. If the gaps were simply classical measurement error, they would not exhibit these patterns.[5]

4.1 Event Study Evidence on O*NET Indices

Another way to summarize the effect of occupation coding source on measurement is to see if measures used in the economics literature vary with the identity of the respondent. The regular updating of occupations gives a 3-year window in which an occupation is not updated. We construct 10 indices used in the literature and show event study around source change from worker to expert for these indices. Specifically, for occupation o, within the 3 digit SOC code $d(o)$, and database t:

$$Index_{ot} = \sum_{\tau \neq -1, \tau = -3..2} \beta_\tau D_t^\tau \Delta Expert_o + \eta_o + \zeta_{d(o)t} + \varepsilon_{ot}$$

Where $\Delta Expert_o$ is the change in data source from incumbent worker to occupational expert for occupation o within an event window. In a robustness exercise, we show the results for the switch from incumbent worker to either occupational expert or analyst, and show results for an extended set of 15 indices.[6] We cluster the standard errors at the occupation level. We run these event studies at the more detailed O*NET occupation level, in order to avoid

[5] Appendix A.2 further shows that following the 2010 skill reform, analyst down-graded skill values scored by workers for social and cognitive skills, but up-scored the skills related to management.

[6] Some indices rely on O*NET data graded heavily by analysts, and as such do not provide us with enough data for the more restrictive event studies

any bias from any additional cross-walking to SOC occupational code.

Figure 3 shows the results for the main indices. Additionally, we show results when including O*NET data coming from analysts in Figure 4. Figure 3 suggests that experts are biased in non-random ways—in particular, they over-value the non-routine, cognitive aspects of jobs, while under-valuing the routine and manual aspects. This provides further evidence of their managerial bias, placing relatively less weight on physical and sensory aspects of the job. We also show that the identity of the rater has the potential to alter frequently-made conclusions in the literature, like which occupations are exposed to machine learning (Brynjolfsson et al., 2018).[7]

Replicating Occupational Trends: We can replicate general occupational trends from Autor et al. (2003) and Deming (2017), using only O*NET data from incumbent workers or from experts. Results on the evolving prevalence of routine, non-routine analytical and social skill-based tasks in the U.S. labor market are shown in Figure A6, in addition with employment (Figure A7) and wage (Figure A8) trends for high/low social and high/low math occupations. Of interest, the decline in the prevalence of routine tasks from Autor et al. (2003) is only present when using incumbent-based assessment of routine. From Deming (2017), employment shares for various skill groups over time look similar using either incumbent worker or expert data. When looking at the change in wages from 1980 to 2010 for these skill groups, worker O*NET ratings predict that low-social, high-math occupations experience one of the largest declines, while expert O*NET ratings predict that these occupations experience one of largest increases.

[7]Some AI exposure measures, like that of Eloundou et al. (2024), cannot be examined using our method because they do not make use of the coder-dependent variables, like task-intensity.

5 Expert Coding and H-1B Visa Approval and Wages

As a final piece of evidence that who codes the occupation matters, we present evidence from an institutional use of O*NET. One policy domain that statutorily makes use of O*NET is the guest worker visa approval system. H-1B visas, in particular, are work visas that require the employer to establish "specialty" status among generally college-educated occupations. O*NET is explicitly referenced in regulation governing visa approval for both the Department of Labor's prevailing wage determination and the U.S. Customs and Immigration Service determination of specialty employment status. Employers submit an application with a job title and a SOC code, along with a prevailing wage rate and a tentative wage for the application. The aim of the application is to set a minimum wage rate for the upcoming position. The application is submitted to the Department of Labor, who decides to approve the application or not, depending on whether the job title is inconsistent with the occupation code provided, and whether the prevailing wage rate is correct for the given occupation, commuting zone and year (Title 20 CFR § 655.731(a)).

As evidence for the importance of O*NET, out of 12,000 H-1B USCIS non-precedent appeal decisions since 2005, O*NET was mentioned in 26% of decisions (see Figure A9 for the evolution of O*NET mentions in USCIS appeal decisions over time). One determinant of specialty status explicitly considered by USCIS is the importance of relative experience, which the previous section showed was underestimated by experts vs workers. We thus ask in this section does the exogenous change in source from worker to expert change the visa approval process? Based on the above results, we expect that expert-coded occupations score lower on "experience needed", and so the applications are less likely to qualify as specialty occupations, are more likely to be denied. When applications for these occupations are approved, they are given a lower prevailing wage and a lower actual wage.

For this analysis, we combine the O*NET databases with USCIS Labor Condition Application (LCA) data for over 7M unique H-1B applications made by employers between 2009 and 2023. The LCA data have been used by Kerr and Lincoln (2010) and Glennon (2024),

and we follow much of this work in cleaning our data.[8] For every application we have the application date, firm name, city, (employer-entered) job title and SOC code. We further see whether the application was accepted or rejected, the prevailing wage assigned, the source of the prevailing wage (OES survey, Collective Bargaining Agreement, or private data vendor), and the actual wage rate paid by the employer to all other individuals with similar experience and qualifications for the specific employment in question. Once the LCA is approved, employers are required to pay a wage above the maximum of the prevailing wage, and of the stated wage paid to similar employees. For applications relying on wage data other than a DOL-issued prevailing wage determination, we also see the prevailing wage echelon, which is a number between 1 and 5 indicating the seniority of the position relative to the average occupation, as well as the final offered wage.

We restrict the data to 221 balanced occupational codes according to our assignment, and merge this data at the occupation-year level to our O*NET database, so that each occupation-year in the LCA data is assigned the O*NET values from the immediately preceding database, including the source (worker or expert).

We then estimate the following specification at the LCA level, restricting attention to 5-year windows around a source change (and those with no source changes). We estimate, for occupation o, commuting zone m, and higher-level 3 digit occupation code $d(o)$:

$$Y_{omdt} = \sum_{\tau \neq -1} \beta_\tau D_t^\tau \Delta Expert_o + \delta_{om} + \delta_{mt} + \delta_{d(o)t} + \varepsilon_{ot}$$

Standard errors are clustered at the occupation level. The om fixed effects remove occupation-CZ common effects, while the mt fixed effects remove CZ-year fixed effects, and the $d(o)t$ fixed effects remove common (coarse) occupation shocks. The remaining variation is driven by just the change in reporting source in event year $\tau = 0$ that is common to the fine occupation over time.

[8] Similarly to them, the use of LCA data implies that we track the evolution of the H-1B labor demand, and the lower bounds for these H-1B wages rather than the actual wages offered to workers once visas are accepted.

Figure 5a shows the effect of the source switching from worker to expert on the "required experience" from the Education, Training, and Experience module in O*NET. Figure 5b then shows that the probability of the visa application being accepted goes down by about 2% two years after the change (for a 94% LCA acceptance rate). Figure 5c shows that conditional on being accepted, the prevailing wage goes down by 6% two years after the change (for an average yearly prevailing wage of $50.8k), and that it further results in offered wages going down by 6.% (for an average yearly rate of $62.7k). While we do not see actual acceptances, realized wages, and other labor market outcomes, this pattern of results is consistent with a mechanism where experts in O*NET believe some occupations require less experience than incumbent workers do, and this results in H-1B visa applications being less likely to qualify for "specialty" status. The ones that do qualify seem to be negatively selected, as they are assigned a lower prevailing wage within that occupation, and thus are offered lower wages in the end.[9] Both the effects of a change in O*NET source—when we include rejected LCAs—on the prevailing wage and the offered wage lose in statistical significance, and become smaller in magnitude (resp. 5.3% for the prevailing wage in the second year—significant at the 10% level, and 5.6% for the offered wage in the second year significant at the 10% level). This empirical pattern correlates with this potential selection mechanism.

The wage directly reported by firms in the LCA is also defined as: "the actual wage level paid by the employer to all other individuals with similar experience and qualifications for the specific employment in question" (Doran et al., 2022). Here the deskilling of jobs which follows from a switch to an expert source seems to directly lead firms to compare the LCA position with lower experience and lower pay workers already on their payroll.

[9]The Employment and Training Administration's 2009 Prevailing Wage Determination Policy Guidance explicitly states: "*The NPWHC shall use O*NET information to identify the tasks, work activities, knowledge, and skills generally required for performance in an occupation. A comparison between the particulars of the employer's job offer to the requirements for similar (O*NET) occupations shall be used to determine the appropriate wage level.*"

6 Constructing a Measure of Tacit Jobs

Measuring Tacitness: on what occupations do coders differ? The previous section documented pervasive differences between experts and workers in occupational codings for the same job. In this section, we propose aggregating these differences into an occupation-level measure of "tacitness": the extent to which workers and managers disagree about economically relevant dimensions of the job. Concretely, we define an occupation's tacitness as:

$$Tacitness_o = \sum_e |\beta_e \times w_{eo}|, \tag{1}$$

where β_e is the coefficient of O*NET element e in a high-dimensional LASSO wage regression with Mincerian controls, and $|w_{eo}|$ is the average difference between incumbent workers on the one side, and occupational experts and analysts on the other, for element e and occupation o. Out-of-sample wage prediction for the LASSO regression is shown in Figure A10, and Figure A11 lists the O*NET elements with the largest coefficients in the LASSO regression. Figure A12 shows occupations that are measured to be most and least tacit by that measure. The figure shows a clear shift from hands-on, task-specific work to more abstract and professional roles. Highly tacit occupations also tend to be more blue-collar and healthcare oriented, as opposed to the heavy representation of managers in less tacit jobs. This reflects both experts' identity as managers and the physical nature of tacit knowledge: the divergence between experts and workers are intuitively larger on the physical dimensions involved in being a hand sewer, healthcare worker, or fishing worker without close personal experience. Indeed, in tacit occupations, we see that the O*NET work context where experts and workers disagree most includes the use of equipment and exposure to weather and noises. In contrast, for non-tacit occupations, smaller disagreements occur in very different contexts: social interactions and administrative tasks. This occupation-level view of tacitness complements the element-level regressions in Figure 2. Both approaches also validate our interpretation of experts as managerial in nature, since managers are some

of the least tacit occupations.

Correlates with Occupation Tacitness Figure A13 shows our tacitness measure is positively correlated with a number of measures from the literature, including the "Readability Index" proposed by Autor and Thompson (2024). Occupations that are simpler to read encode more tacit information, by our measure. In our view, this makes sense; "cook" is a simple to describe occupation, but much of the skill and discomfort may be only salient to the worker doing the job, even as the overall linguistic signifier omits it. We use data by Anthropic that classifies prompts sent to Claude into O*NET tasks, which can be aggregated to an occupation measure of AI adoption. With this aggregation, we find a negative correlation between our occupation index of tacitness and AI adoption, as shown in Figure A13. In Figure A14, we show the correlation between occupational tacitness and average occupational demographics taken from the ACS: more tacit occupations feature older workers, more likely to be non-white, women, with a lower level of education, and a lower hourly wage.

Finally, we show that our tacitness measure does not merely capture sampling noise; in other words, we rule out that these are occupations where workers exhibit greater variance in how they answer survey questions related to their jobs. We calculate, within each occupation, the standard deviation of O*NET values across all elements rated by workers and correlate this with our index of tacitness. In Table A2, we find that if anything, tacit occupations exhibit more homogeneous worker responses. We further validate this relationship by taking our tacitness measure to another survey of working conditions: the ALP's American Working Conditions Survey (AWCS). Fielded as a panel in 2015 and 2018, the AWCS includes respondent demographics and SOC codes, and similarly to O*NET, it asks about working conditions like the physical demands of work, hazard exposure, autonomy, and meaning on the job. Table A2 shows that taking the standard deviation of AWCS respondents' answers within occupation, there is no correlation to our occupational tacitness

measure. One advantage of the AWCS over O*NET is that we can see workers' demographic characteristics, and thus probe if our measure of tacitness captures sorting: in more tacit occupations, do workers with different demographics answer working conditions questions differently? Table A3 shows that within occupation, worker demographics are not predictive of their responses. Except for race, worker demographics in more tacit occupations are not differentially predictive of survey responses.

Revisiting the 2010 Skill Reform In Appendix A.2 we discuss the 2010 reform in which O*NET switched the way it was grading data for its Skills module, moving from the common sampling of incumbent workers and occupational experts to only professional analysts, as shown in Figure **??**. This reform led to a general down-grading of skill values in O*NET, but the effects are heterogeneous by both skill dimension (managerial, cognitive, technical, and social) and by occupation tacitness. We analyze the reform within an event study framework, showing that across occupations, analysts down-graded skills related to social and cognitive skills relative to workers, but up-graded skills related to management skills. We first show, in Figure A3 that tacit occupations were relatively more down-graded along non-managerial dimensions of skill, in particular social skills. This is consistent with the larger downgrading of non-managerial skills (and upgrading of managerial skills) for occupations whose pre-reform skills were coded more by incumbent workers vs experts, and our general finding that analysts are closer to experts than to incumbent workers discussed above.

The heterogeneous impacts of the 2010 skill coding reform again present evidence that coder identity matters for measuring skill content, and this effect varies by job and skill dimension. Skills that are more relational and context-specific, like technical and social skills, will be more salient to people that do the job, while skills that are easier to describe to superiors, like managing other people, will be salient to experts.

7 Comparing O*NET Measures to Labor Market Data

To interpret the difference between worker and expert ratings beyond the context of O*NET, we show that workers' ratings are a better signal of labor market realities than experts'. This part provides further evidence against the interpretation of O*NET differences as stemming from measurement error, but rather from worker private information. To compare O*NET ratings to a measure of ground truth in the labor market, we compare O*NET measures to wages, Mincerian returns, and injuries as measured in publicly available data.

We first take an expansive definition of skill, aggregating all the non-amenity elements in O*NET, and we measure the average worker score and average expert score. We then check if the average skill reported by experts and workers differ in how strongly they are correlated with wages. If market wages are pricing skills, then we expect wages to be more strongly correlated with the better measure of skill. Using the American Community Survey data from 2003-2023, we estimate the following regression, where X can include Mincerian controls (age, race, sex, years of education, year FE).

$$\log(w_i) = \alpha + \gamma^w Skill^w_{o(i)} + \gamma^e Skill^e_{o(i)} + X_{it}\Gamma + \epsilon_{iot} \tag{2}$$

Results are in Table 2. When run separately, the coefficient on worker measures of skill are larger than expert measures of skill, and this remains true when controls are included. When both worker and expert measures are included, only the worker measure is positive and significant, while the expert measure flips sign and is smaller in magnitude.

We do not take a stand on why the expert measure becomes negative. One interpretation is non-classical measurement error. Another interpretation is that expert disagreement with workers is itself a disamenity. Either (or both) of these interpretations are consistent with a story in which workers have a better signal of skill on the job than experts.[10]

We can further use O*NET to go beyond a catch-all notion of human capital and look

[10]Note that the negative coefficient on the expert skill measure is inconsistent with pure differential classical measurement error in the two measures.

at different components of skill. We ask how workers' versus experts' O*NET ratings of the importance of experience and education correlate with occupation-level Mincerian returns to experience and education. We restrict attention to workers aged 17-65 and estimate occupation-specific returns to education and returns to experience separately in bivariate regressions using ACS data. Figure 6 shows the corresponding binned scatterplots of the worker versus expert O*NET ratings (on the occupation-database level) against the true returns (on the occupation level). Regression coefficients are reported on the graphs.

As might be expected if workers have a better sense of the tacit skills that come with experience, the returns to experience are more strongly correlated with worker measures than expert measures. Consistent with the notion that workers better perceive the salient dimensions of the job, the differences in worker and expert ratings are biggest for high skill and experience occupations, including engineers, actuaries, and economists. Appendix Figure A15 shows that workers' measures are even more strongly correlated with returns to tenure relative to experts' measures, where average tenure is measured on the occupation level from LinkedIn data (see Figure A15 notes for details). But there is no worker-expert difference in the returns to education, again consistent with the idea that experts can accurately guess the returns to formal, observable credentials, but have less of a sense of the returns to informal, on-the-job learning that is reflected in the returns to experience.

Besides wage returns, we can also look at ground-truth injury risk. We take annual, non-fatal injury rates from BLS Survey of Occupational Injuries and Illnesses, and we average annual occupational injury rates from 2011 to 2020. We then take the average of all O*NET elements associated with environmental and physical conditions on the occupation-database level.[11] The binned scatterplots in Figure 7 show that the worker-based measures of environmental and physical risk are significantly more correlated with true job injury rates than are the expert scores. For occupations that are widely known to be high-risk, like Police Officers, and for occupations that are widely known to be low-risk, like Tax Preparers, experts and

[11]To proxy the environmental and physical conditions in O*NET, we use the elements specified in Measuring Physical Disamenities, Appendix A.1.

workers are close, and correct, in their valuations. However, for above-median injury rate occupations like plant operators and concierges, where only experience can reveal the exact nature of the physicality, workers rate higher disamenity scores than experts.

In sum, the difference between expert and worker ratings of skill, experience, and job risk are consistent with workers having a better signal of these job characteristics than experts. At the same time, the absence of significant differences between workers and experts in the importance of formal education is consistent with the differences reflecting tacit knowledge.

8 Quantifying Tacit Knowledge in Equilibrium

8.1 Theoretical Framework

Set-Up: In this section, we develop a model of two-sided matching in the labor market, building on Dupuy and Galichon (2022), Galichon and Salanié (2022) and Corblet and Dupuy (2024). We study a labor market where workers and firms can form labor contracts and trade one unit of labor. We extend this class of models by allowing for two-sided incomplete information about the benefits and costs of the labor contract. We propose a micro-foundation of this incomplete information, where private information is only revealed to agents in a second step, after they match. Within this second stage, we derive the necessary and sufficient conditions for the existence of a mechanism satisfying ex ante budget balance, incentive compatibility and the participation constraint of the firm and worker that matched. This structure yields a first stage matching game similar to the one considered in usual two-sided matching problems as analyzed by Galichon and Salanié (2022), but where the virtual surplus generated by a match, rather than the actual realized surplus, determines match probability. As such, private information acts as an additional friction in the labor market, which reduces the probability of any worker-firm match.

Building on the results from Section 4 we focus on the returns to education and experience, as well as injuries, as occupational characteristics that workers and firms disagree

about. We assume that, conditional on the empirical measure of the returns to education/experience/injuries, the expert's O*NET measure is a proxy for the firm's private information, while the worker's O*NET measure proxies for the worker's private information.

In this market, a worker $i \in \mathcal{I}$ has an observable type $x(i) \in \mathcal{X}$, a vector of taste shocks $\varepsilon_i = \{\varepsilon_{y_i}\}_y$ distributed T1EV (which we later write as $\varepsilon = \{\varepsilon_y\}_y$ for simplicity), and a vector of utility-relevant tacit knowledge of jobs y, $\xi_w = \{\xi_w^y\}$, which is distributed according to the cdf $F_x(.|Z_y^w)$, with pdf $f_x(.|Z_y^w)$, and support $[\underline{\xi_w^y}, \overline{\xi_w^y}]$. The same holds on the firm side with firms denoted by $j \in \mathcal{J}$, observable characteristics specific to an occupation for which it seeks a worker $y(j) \in \mathcal{Y}$, and then η for the T1EV shocks, and profit-relevant tacit knowledge of jobs y, $\xi_f = \{\xi_f^y\}$, distributed according to the cdf $Q_y(.|Z_y^f)$, with pdf $q_y(.|Z_y^f)$, with support $[\underline{\xi_f^y}, \overline{\xi_f^y}]$.[12]

In mapping to our data, we think of the empirical measures of the occupation-level injuries, returns to education and tenure as variables in y (along with capital/labor ratios). We then use the O*NET worker and expert measures of physical hazards, importance of formal education and experience as parameterizing the distributions of analogous private information vectors.

Timing: While $\{x, y, \varepsilon, \eta\}$ are known, as well as the parametrized distributions of the $\{\xi_w, \xi_y\}$, tacit information is private information to resp. workers and firms. Its presence gives rise to asymmetric information. Furthermore, tacit characteristics are only observed by agents after their initial match. This implies the following timing:

1. **Ex-Ante** $\{\xi_w, \xi_y\}$ are unknown to all, so matching occurs based on $\{x, y, \varepsilon, \eta\}$. Agents match based on their expectations about job utilities, forming pairs or remaining unmatched. We write μ_{xy} for the mass of matches of type (x,y), μ_{x0} for the mass of unemployed workers of type x, and μ_{0y} for the mass of job vacancies of type y.

[12] In this section, we use the words firm and occupation interchangeably. The estimation of the model will use occupation-level data. Within an occupation, the model allows for different firms to exist, but this heterogeneity will be fully unobserved characterized by their specific private information vector ξ_f and their shock η.

2. **Interim** For all agents who have entered a match (x,y), tacit characteristics $\{\xi_w, \xi_y\}$ are realized, but remain unobserved to the other party. Given these, agents see if they can actually enter a work contract. The contract is defined by $t_w(\xi_w, \xi_f)$ and $t_f(\xi_w, \xi_f)$, which are resp. the transfers received by the worker, and paid by the firm, and by $\rho(\xi_w, \xi_f)$, the probability that the unit of labor is traded.[13]

3. **Ex-Post** The mechanism is implemented and types are revealed.

We further assume the following von Neumann Morgenstern worker utility and firm productivity, where A_{xy} and Γ_{xy} are the observed elements of worker amenities and firm productivity, with $\rho(\xi_w, \xi_f)$ the allocation rule indicating whether trade happens, and $t_w(\xi_w, \xi_f)$ and $t_f(\xi_w, \xi_f)$ the transfers received by both agents,[14]

$$
\begin{aligned}
u_{xy}\Big(\xi_w, \xi_f\Big) &= \Big[A_{xy} + \xi_w + t_w(\xi_w, \xi_f)\Big]\rho(\xi_w, \xi_f) + \varepsilon_y \\
v_{xy}\Big(\xi_w, \xi_f\Big) &= \Big[\Gamma_{xy} + \xi_f - t_f(\xi_w, \xi_f)\Big]\rho(\xi_w, \xi_f) + \eta_x
\end{aligned}
$$

Finally, this market is characterized by a mass n_x of workers of type x, and a mass m_y of occupations of type y. We write by $\mathcal{X}_o$ and $\mathcal{Y}_o$ the extended choice sets which include the outside option of no match.

The equilibrium is realized in two stages, first by matching at the ex-ante stage where workers and firms have symmetric expectations about tacit characteristics, second by bargaining once the match is made and as workers and firms learn about their respective private valuations. We study the model using backward induction, focusing first on the second step and equilibrium conditions within a match, and then on the first step and matching given the agents' expected matching outcomes. Appendix B.1 describes the derivations for the second step equilibrium characterization. We specifically derive the conditions under which

[13] Alternatively it can be interpreted as the duration of the labor contract, with $\rho(\xi_w, \xi_f) = 0$ indicating an immediate breach of the contract, and $\rho(\xi_w, \xi_f) = 1$ a contract held for its entire intended duration.

[14] By design, we assume that if trade does not happen, the worker and the firm only receive their random taste shock.

a mechanism exists, such that we can write first stage ex ante expected utilities—the utilities based on which matching happens in the first stage—as:

$$U_{xy} = \mathbb{E}_{\xi}\left[\left(A_{xy} + \xi_w\right)\overline{\rho}_w(\xi_w)\right] + w_{xy} + \varepsilon_y$$
$$V_{xy} = \mathbb{E}_{\xi}\left[\left(\Gamma_{xy} + \xi_f\right)\overline{\rho}_f(\xi_f)\right] - w_{xy} + \eta_x$$

Here, $\mathbb{E}_{\xi}\left[\overline{\rho}_w(\xi_w)\right]$ is the ex ante expectation that trade happens after a worker and a firm match. Appendix B.1 shows that trade will only go through in cases where the virtual surplus generated by the match is positive. Because the virtual surplus corresponds to the expected surplus minus the information rents paid to both agents, private information lowers the probability of trade. w_{xy} corresponds to the expected transfer paid by the firm to the worker if the match happens.

Following the usual derivations from standard discrete choice, and imposing equilibrium conditions, we can then recover the following equilibrium conditions for the first stage matching game, with $\sigma = \sigma_w + \sigma_f$:

$$\mu_{xy} = exp\left(\frac{\tilde{\Phi}(x,y) + \sigma_w log(\mu_{x0}) + \sigma_f log(\mu_{0y})}{\sigma}\right)$$
$$\tilde{w}(x,y) = \frac{\sigma_f}{\sigma}\left(- \sigma_w log(\mu_{x0}) - \mathbb{E}_{\xi}\left[(A_{xy} + \xi_w)\overline{\rho}_w(\xi_w)\right]\right) + \frac{\sigma_w}{\sigma}\left(\mathbb{E}_{\xi}\left[\left(\Gamma_{xy} + \xi_f\right)\overline{\rho}_f(\xi_f)\right] + \sigma_f log(\mu_{0y})\right)$$

We show the derivations for these equations in Appendix B.2. Because a match does not necessarily goes through, or $\mathbb{E}_{\xi}\left[\rho(\xi_w, \xi_f)\right] \leq 1$, $\tilde{\Phi}(x,y)$, which corresponds to the expected surplus from a match conditional on the match leading to trade, is strictly lower than the expected surplus from a match. Compared to a two-sided matching model without private information, here private information lowers the mass of matches μ_{xy}.

8.2 Empirical Implementation

We model the productivity and amenities as flexible functions of worker and occupation characteristics with the following two linear equations[15]:

$$\mathbb{E}_\xi\left[\left(\Gamma_{x'y}+\xi_f\right)\overline{\rho}_f(\xi_f)\right] = \text{return to educ}_y + ONET_y^{educ,expert} + \text{tenure}_y + ONET_y^{exp,expert} + \text{injury rate}_y + ONET_y^{injury,expert} + D_{xy}$$

$$\mathbb{E}_\xi\left[\left(A_{xy}+\xi_w\right)\overline{\rho}_w(\xi_w)\right] = \text{return to educ}_y + ONET_y^{educ,worker} + \text{tenure}_y + ONET_y^{exp,worker} + \text{injury rate}_y + ONET_y^{injury,worker} + D_{xy},$$

We first include occupation y characteristics related to the required education, to tenure, and to injury rates. We include two versions of these characteristics. The first are "ground truth" characteristics (return to educ_y, return to exp_y, injury rate_y) that come either from observed average occupational years of education in the ACS, average tenure from LinkedIn data, or average injury rates from the BLS. Note that these enter the utility functions of both workers and firms, as we assume that both sides observe these characteristics. We include the O*NET values for required education and experience and injury risk as rated by workers in the worker utility function, and as rated by experts in the firm utility function.[16][17] This specification imposes that the worker O*NET values are only visible to workers, while the expert O*NET values are only visible to firms. Put another way, the O*NET values represent the private information that is available to either side, conditional on the "ground truth" of the job, that is visible to both sides.

D_{xy} then includes an additional y characteristic, which is the occupation capital/labor ratio computed from the BEA industry data combined with the ACS industry-occupation variables. It also includes worker-level demographics (female, white, college education, and

[15] For the empirical implementation of the model, we move to 3 digit SOC codes.

[16] O*NET values are taken as averages across all O*NET databases for a given occupation-rater combination.

[17] The O*NET injury data is recoded as a dummy, taking the value of 1 for occupations in the top third of injury risk.

age). Finally, D_{xy} also includes all the possible interactions between any given x characteristic and the complete set of observable y characteristics, including our "ground truth" characteristics. These interactions are necessary to account for potential complementarity in matching.

We estimate the model using maximum likelihood, along the lines of Galichon and Salanié (2022). Because their approach focuses on match data, we augment the likelihood with a term accounting for wages, following Dupuy and Galichon (2022). The estimation is detailed in Appendix B.6.

8.3 Results

Appendix Table A4 reports the estimated average effects of occupational and worker characteristics on firm productivity and worker amenities, alongside the variance of unobserved taste shocks. We specifically contrast the effects of observable "ground truth" characteristics with tacit occupational requirements measured via O*NET.

Observable physical risks, measured by the BLS injuries survey, generate a compensating differential tradeoff: they significantly decrease worker amenities while increasing firm productivity, suggesting firms adopt riskier processes to achieve higher output. In contrast, tacit injury risks (measured via O*NET) offer no significant productivity gains for firms but further decrease worker amenities. For education, we find that the market relies almost entirely on observable metrics. Observed returns to education significantly raise firm productivity with no significant effect on amenities. The tacit dimension of education (O*NET) has no statistically significant effect on either productivity or amenities. Because formal education is easily observable and verified, tacit educational requirements play little role in the matching equilibrium. Occupations characterized by higher average tenure feature both higher productivity and higher amenities. However, tacit experience requirements act differently. While O*NET experience also raises firm productivity, it strictly decreases worker amenities. This negative amenity effect suggests that tacit experience is highly occupation-specific

and less transferable across the labor market. Consequently, workers view tacit experience requirements as an occupation-specific investment for which they must be compensated.

We finally find that $\sigma_w = 3.01$ and $\sigma_f = 3.51$. The presence of significant idiosyncratic preferences implies substantial labor market frictions.

8.4 Counterfactuals

We use the estimated model to quantify the equilibrium incidence of tacit information in the labor market. Our baseline equilibrium reflects a market shaped by two forces: ex-ante matching frictions (driven by unobserved idiosyncratic taste shocks, $\{\varepsilon, \eta\}$) and ex-post informational frictions (where agents sort based on private, tacit knowledge). To isolate the role of this tacit information, we perform two main counterfactual comparisons.

First, we shut down the tacit information channel entirely, restricting workers and firms to match solely on public information. Comparing this to the baseline isolates the allocative and wage incidence of private information within the current, high-friction market.

Second, we explore the interaction between tacit information and structural market frictions. We establish a "low-friction" baseline by reducing the scale of the unobserved taste shocks ($\{\sigma_w \, \sigma_f\}$) by 80%. We then compare this low-friction environment to an identical one where tacit information is again shut down. This second exercise reveals how the pricing and sorting role of private information changes when standard matching frictions are minimized.

Results for our counterfactual exercises are reported in Table 3 and Table 4, with an illustration of the drivers of wage changes in Figure 8 and Figure 9.

We first isolate the effect of removing private information within the baseline market, which features high ex-ante frictions. As shown in Figure 8, wages adjust predictably to restore equilibrium: occupations that experience a relative decrease in amenities or firm outside options, or an increase in productivity or worker outside options, experience wage increases. For the average worker, tacit information primarily reveals harmful injury risks that depress amenities. Consequently, removing this private information yields a dual positive shock: it

directly improves perceived amenities and raises the worker's outside option (as the overall set of potential occupations is perceived to improve). These joint increases perfectly offset, resulting in a null effect on average wages and a slight increase in total matches.

Wage effects are more significant in the tails of the distribution. At the bottom, Entertainment Attendants see a \$3.61 increase in hourly wages, driven by a combination of decreased amenities (as experience requirements rise) and a drop in the firm's outside option. At the top, Management Occupations experience a \$5.00 decrease in hourly wages, driven by a joint increase in amenities and a decrease in productivity resulting from the removal of high tacit experience requirements.

In a low-friction market, Figure 9 shows that endogenous sorting plays a larger role at equilibrium. Without frictions, workers who experience an increase in their outside option sort into higher-amenity occupations. Because workers are willing to pay for these amenities, this sorting behavior generates a negative correlation between changes in worker outside options and wage changes—a reversal from the high-friction market. Additionally, wage effects are magnified. The average worker still experiences an amenity gain from the removal of tacit injury risks. However, unlike in the baseline market where this was mediated by rising outside options, the low-friction worker outside option actually falls slightly, leading to an average wage decrease of \$0.60. Wage outcomes at the tail of the distribution also shift: the wage gain for Entertainment Attendants is lower tat \$1.70 (compared to \$3.61 under high frictions), while the wage penalty for Management Occupations expands slightly to \$5.35.

While the content of tacit knowledge is highly heterogeneous across occupations, for the average worker, it predominantly signals physical injury risk. Therefore, the presence of private information strictly lowers expected worker utility at work. Crucially, we find that structural frictions in the labor market are large enough to prevent the complete pricing of this tacit information. When frictions are minimized, the wage premium required to compensate workers for tacit requirements becomes significantly larger.

9 Conclusion

In this paper we have explored the differences in views on what a "job" entails between people actually doing the job—incumbent workers—and people at some remove from the job—experts. Exploiting the differential coding of the same job over time by different sources, we document systematic differences in occupational characteristics perceived by workers vs experts, and these differences are on dimensions of the job that are difficult to communicate, what we term "tacit knowledge". We show that these differences in coding are important for both the uses of O*NET in economics, as well as real world outcomes in the H-1B application system.

We then take seriously the idea that O*NET differences between workers and experts encode private information. We embed the O*NET differences into a structural hedonic model with ex-ante frictions and post-matching bargaining under incomplete information.

We examine the implications of tacit information for pricing education, experience, and injury risk. Eliminating tacit information lowers the average wage because this information en. This particularly matters for low-education, high-injury jobs, where the tacit information of workers is ignored by the market, and so we find that labor market inequality increases when the market prices public information rather than tacit information. We also find that labor market frictions exacerbate the mispricing of tacit information; reducing frictions increases the importance of the O*NET variables in explaining labor market outcomes, even conditional on our proxies for the empirical truth.

In sum, our paper examines both the systematic biases in frequently used occupation codings, as well as the wage and employment implications of experts/managers and workers disagreeing about the important features of an occupation. As social scientists, and increasingly, computer scientists, continue to rely on O*NET for measuring and making sense of the future of the labor market, more attention might be paid to how the codes may reflect particular points of view on what is productivity and costly about an occupation.

References

Abbott, A. (1988): *The System of Professions: An Essay on the Division of Expert Labor*, University of Chicago Press.

Acemoglu, D. and D. Autor (2011): "Skills, Tasks and Technologies: Implications for Employment and Earnings," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, Elsevier, vol. 4, 1043–1171.

Attewell, P. (1990): "What is skill?" *Work and occupations*, 17, 422–448.

Autor, D. H. and M. J. Handel (2013): "Putting tasks to the test: Human capital, job tasks, and wages," *Journal of labor Economics*, 31, S59–S96.

Autor, D. H., F. Levy, and R. J. Murnane (2003): "The Skill Content of Recent Technological Change: An Empirical Exploration," *The Quarterly Journal of Economics*, 118, 1279–1333.

Autor, D. H. and N. Thompson (2024): "Expertise," Manuscript, working paper / handout (Aug 29, 2024 version).

Braverman, H. (1974): *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century*, Monthly Review Press.

Brynjolfsson, E., T. Mitchell, and D. Rock (2018): "What can machines learn and what does it mean for occupations and the economy?" in *AEA papers and proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, vol. 108, 43–47.

Corblet, P. and A. Dupuy (2024): "Modelling Monospony on the Labor Market with Separable Matching Models," Tech. rep., IZA Discussion Papers.

Deming, D. J. (2017): "The Growing Importance of Social Skills in the Labor Market," *American Economic Review*, 107, 1593–1640.

Doran, K. B., A. Gelber, A. Isen, and J. Song (2022): "The Effects of High-Skilled Immigration Policy on Firms: Evidence from Visa Lotteries," *Journal of Political Economy*, 130, 2501–2533.

Dupuy, A. and A. Galichon (2022): "A Note on the Estimation of Job Amenities and Labor Productivity," *Quantitative Economics*, 13, 153–177.

Eloundou, T., S. Manning, P. Mishkin, and D. Rock (2024): "GPTs are GPTs: Labor market impact potential of LLMs," *Science*, 384, 1306–1308.

Galichon, A., S. D. Kominers, and S. Weber (2019): "Costly Concessions: An Empirical Framework for Matching with Imperfectly Transferable Utility," *Journal of Political Economy*, 127, 2875.

Galichon, A. and B. Salanié (2022): "Cupid's invisible hand: Social surplus and identification in matching models," *The Review of Economic Studies*, 89, 2600–2629.

Glennon, B. (2024): "How Do Restrictions on High-Skilled Immigration Affect Offshoring? Evidence from the H-1B Program," *Managemnet Science*, 70, 907.

Grant, M. (2024): "Where Do the Data Come From? Endogenous Classification in Administrative Data," Working paper (version dated May 2023; cited in slides as 2024).

Handel, M. J. (2016): "What do people do at work? A profile of US jobs from the survey of workplace Skills, Technology, and Management Practices (STAMP)," *Journal for Labour Market Research*, 49, 177–197.

Hayek, F. A. (1945): "The Use of Knowledge in Society," *The American Economic Review*, 35, 519–530.

Hilton, M. L. and N. T. Tippins (2010): "A database for a changing economy: Review of the Occupational Information Network (O* NET)," .

Holland, J. L. (1958): "A personality inventory employing occupational titles." *Journal of applied psychology*, 42, 336.

Hurwicz, L. (1972): "On informationally decentralized systems," *Decision and organization: A volume in Honor of J. Marschak*.

Kerr, W. R. and W. F. Lincoln (2010): "The Supply Side of Innovation: H-1B Visa Reforms and U.S. Ethnic Invention," *Journal of Labor Economics*, 28, 473.

Makowski, L. and J. M. Ostroy (2001): "Perfect Competition and the Creativity of the Market," *Journal of Economic Literature*, 39, 479–535.

Mas, A. (2025): "Non-Wage Amenities," Working Paper 33643, National Bureau of Economic Research.

Maskin, E. S. (2008): "Mechanism design: How to implement social goals," *American Economic Review*, 98, 567–576.

Mayda, A. M., G. Peri, and W. Steingress (2018): "The Effect of the H-1B Quota on Employment and Selection of Foreign-Born Labor," *European Economic Review*, 108, 105–128.

Myerson, R. B. and M. A. Satterthwaite (1983): "Efficient mechanisms for bilateral trading," *Journal of economic theory*, 29, 265–281.

Nisan, N. and I. Segal (2006): "The communication requirements of efficient allocations and supporting prices," *Journal of Economic Theory*, 129, 192–224.

Pathak, P. A., T. Rees-Jones, and T. Sönmez (2025): "Immigration Lottery Design: Engineered and Coincidental Consequences of H-1B Reforms," *Review of Economics and Statistics*, 107, 1–13.

Peterson, N. G., M. D. Mumford, W. C. Borman, P. R. Jeanneret, and F. E. A. (1999): "An occupational information system for the 21st century: The development of O*NET," *American Psychological Association*.

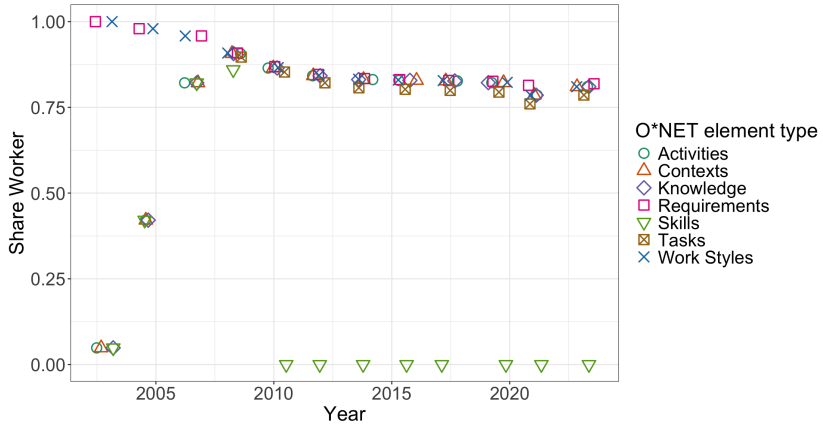
Polanyi, M. (1958): "Personal Knowledge: Towards a Post-Critical Philosophy," .

Rosen, S. (1974): "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 82, 34–55.

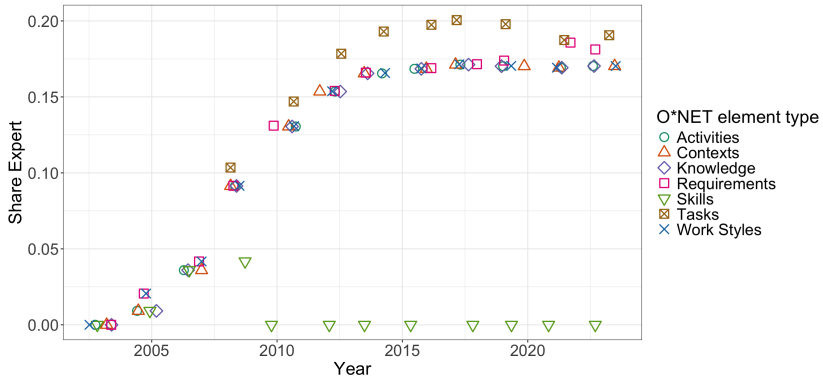
Tsacoumis, S. and C. H. V. Iddekinge (2006): "A Comparison of Incumbent and Analyst Ratings of O*NET Skills," *HumRRO Human Resources Research Organization Report*, FR05-66.

Tsacoumis, S. and S. Willison (2010): "O*NET Analyst Occupational Skill Ratings: Procedures," *HumRRO Human Resources Research Organization Report*, FR-08-70.

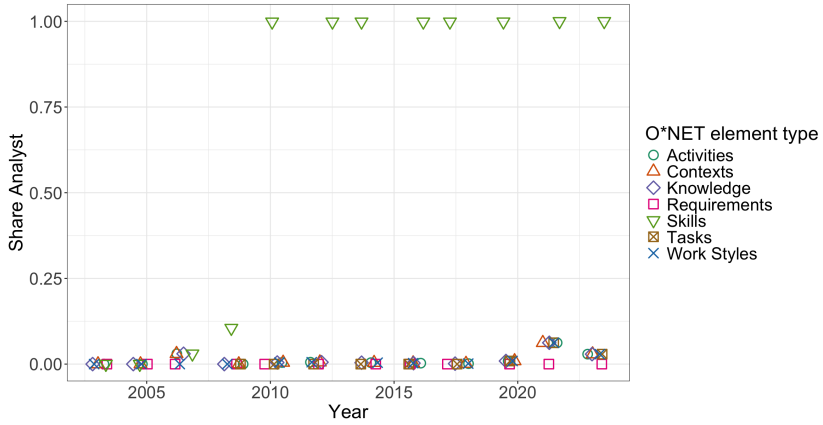
Viscusi, W. K. (1981): "Occupational Safety and Health Regulation: Its Impact and Policy Alternatives," in *Research in Public Policy Analysis and Management*, ed. by J. P. Crecine, JAI Press, vol. 2, 281–299.



(a)



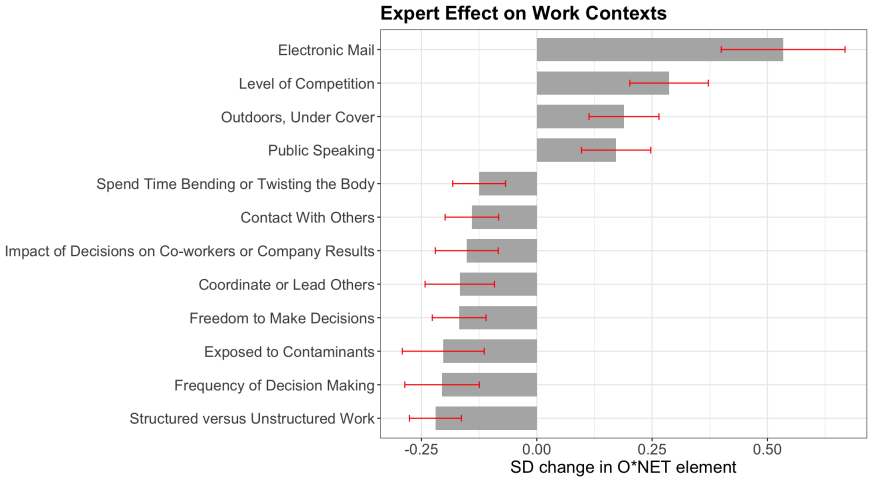
(b)



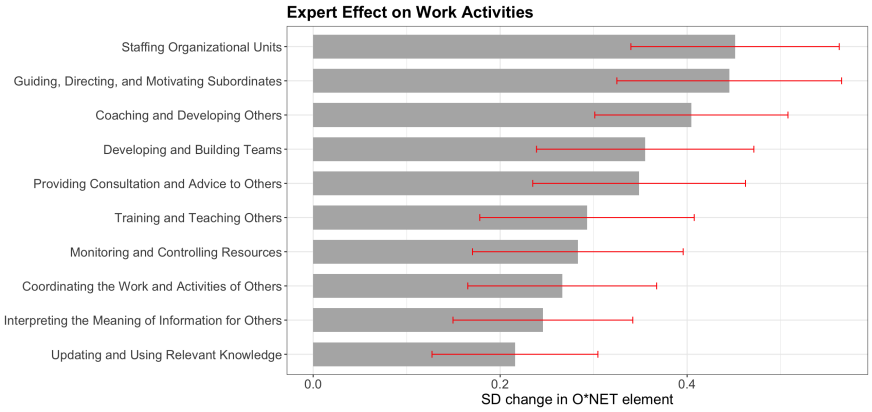
(c)

Figure 1: Share of O*NET Data Graded by Respective Sources

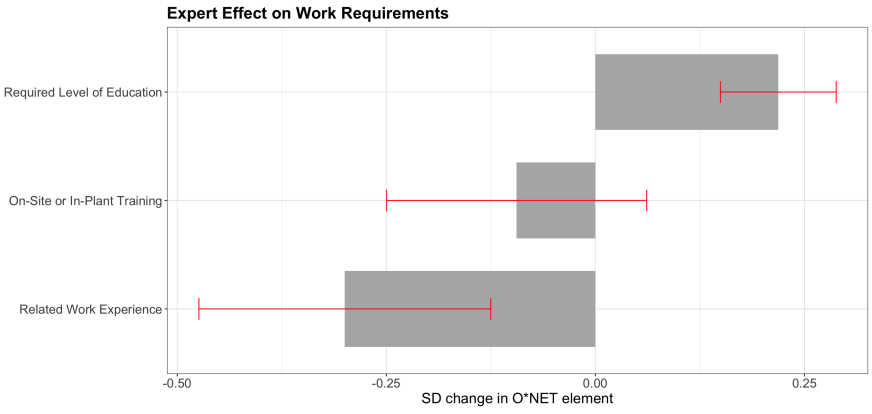
Notes: The figures describe the share of the O*NET data within a given database and a given module, which comes from occupational experts, analysts, and incumbent workers, respectively. The slow increase in the worker and expert shares shows that early O*NET versions were mostly constructed using analysts. The 2010 reform in the collection of the Skills module is visible, with an analyst share jumping to 1 just after 2010.



(a)



(b)



(c)

Figure 2: Expert Effect on Work Contexts, Activities, and Requirements

Notes: The figures plot the coefficients from the O*NET module-specific LASSO regressions used to select the set of O*NET elements on which incumbent workers and occupational experts disagree the most.

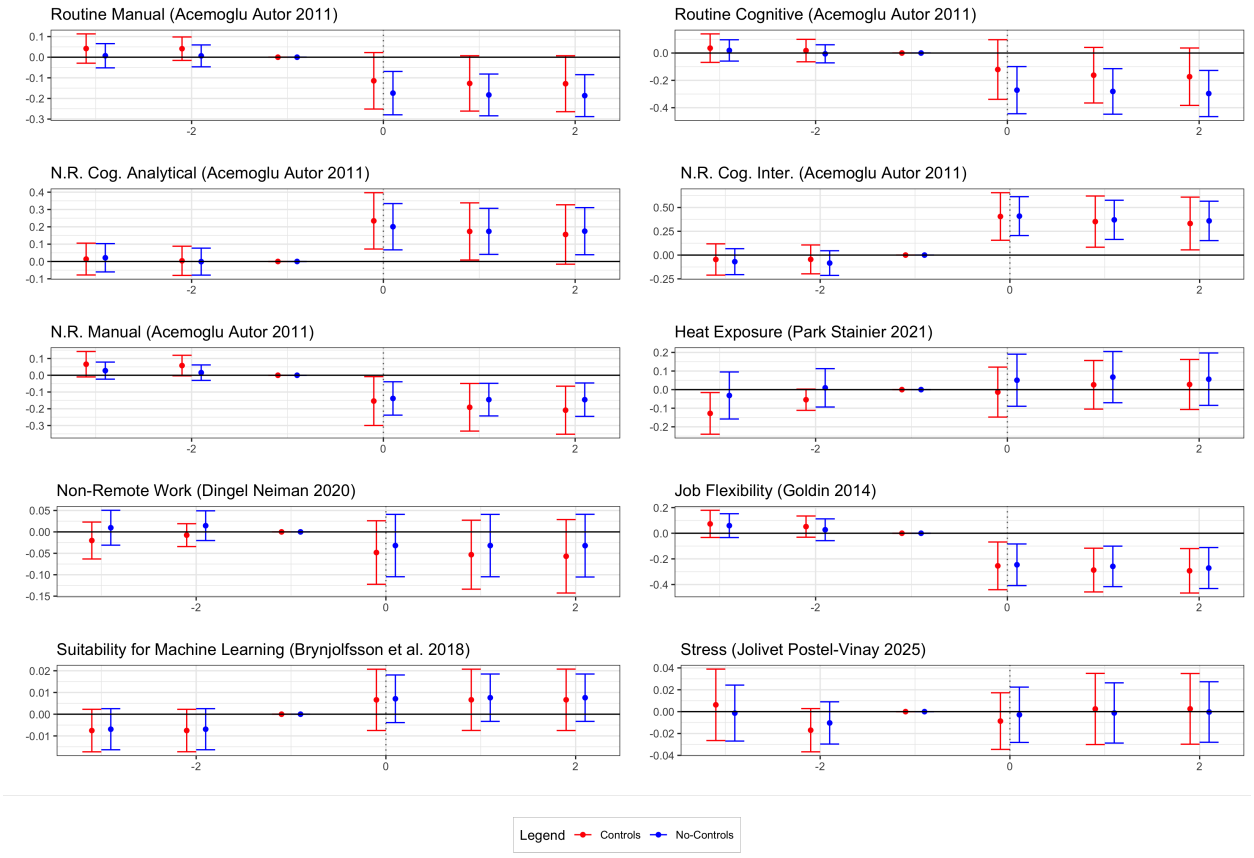


Figure 3: Event study on Literature Indices (Workers to Experts)

Notes: We compute several indices used by the literature, relying either on O*NET data from incumbent workers or from occupational experts. We track the effect of a change in data source from workers to occupational experts. The data is aggregated to 2019 O*NET occupational codes.

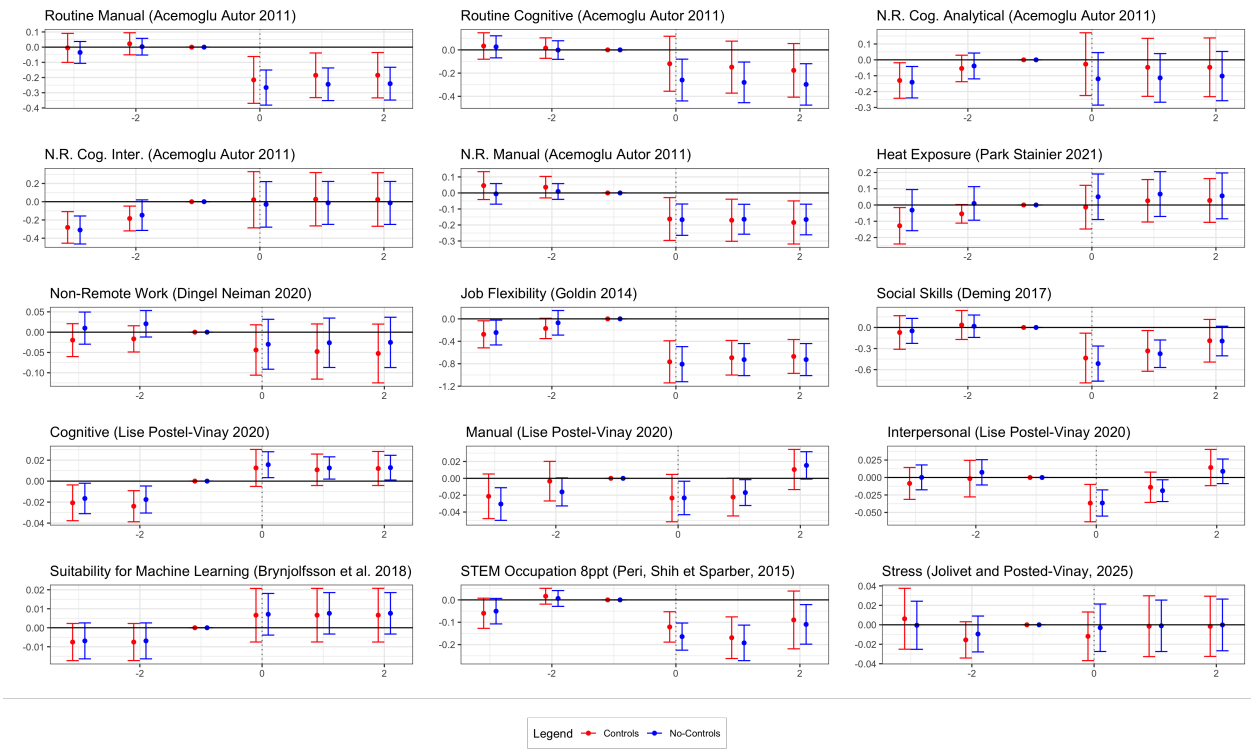
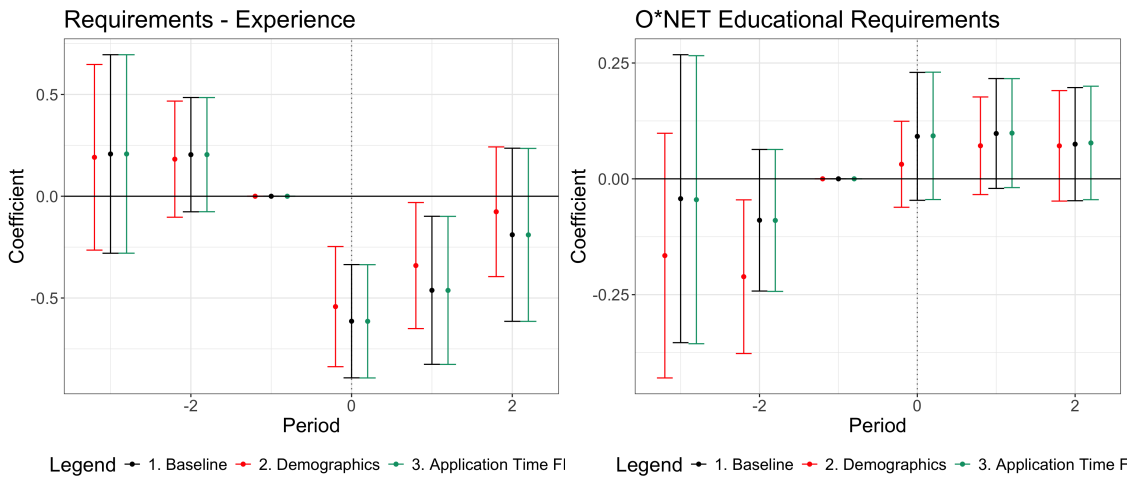
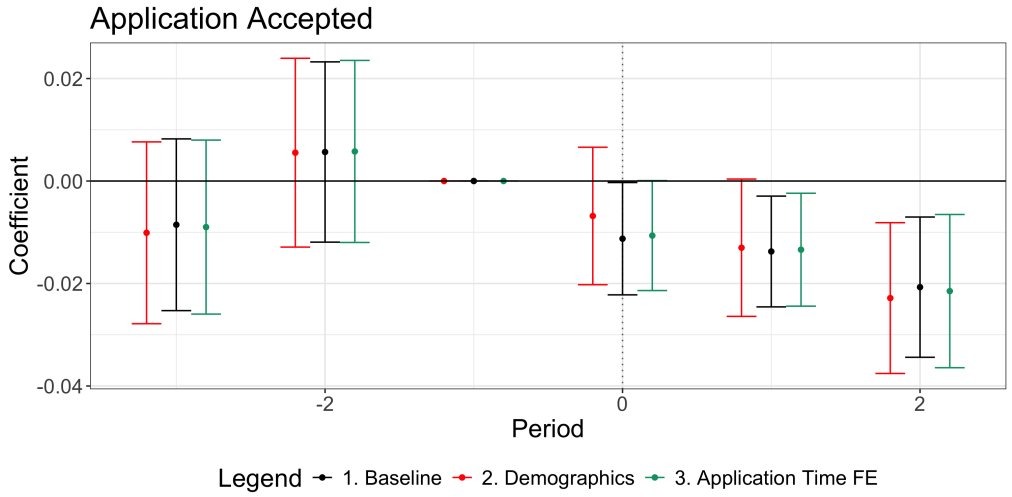


Figure 4: Event study on Literature Indices (Workers to Experts/Analysts)

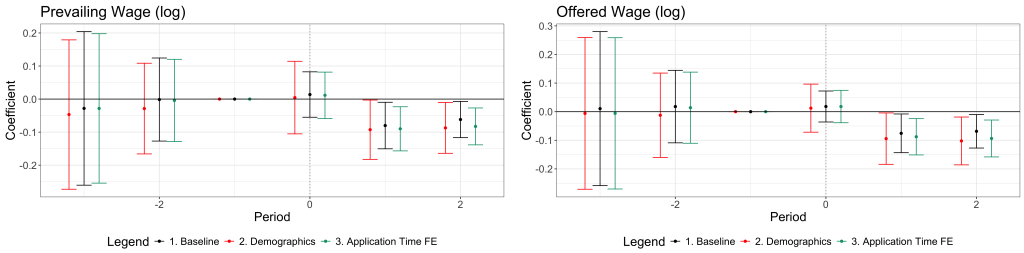
Notes: We compute several indices used by the literature, relying either on O*NET data from incumbent workers, or from occupational experts and analysts. We track the effect of a change in data source from workers to either occupational experts or analysts. The data is aggregated to 2019 O*NET codes.



(a) Required Experience and Education



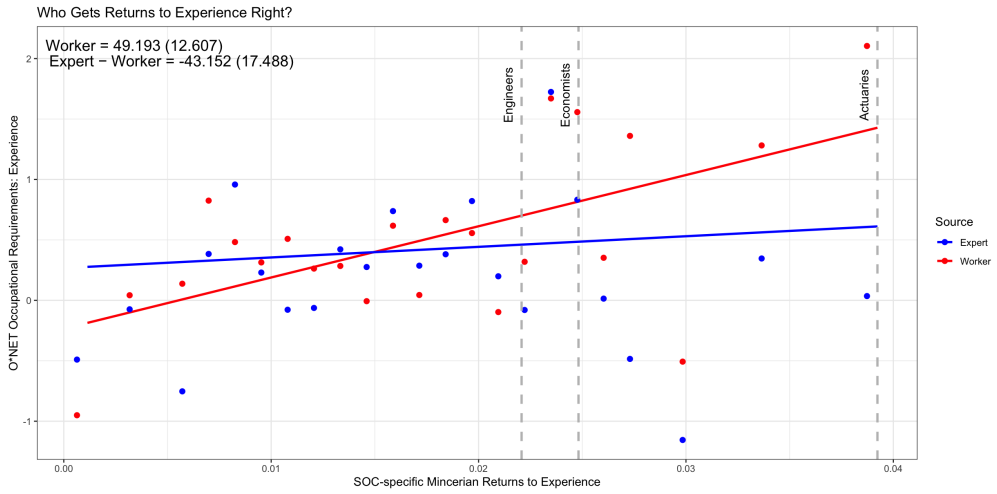
(b) Acceptance Rate



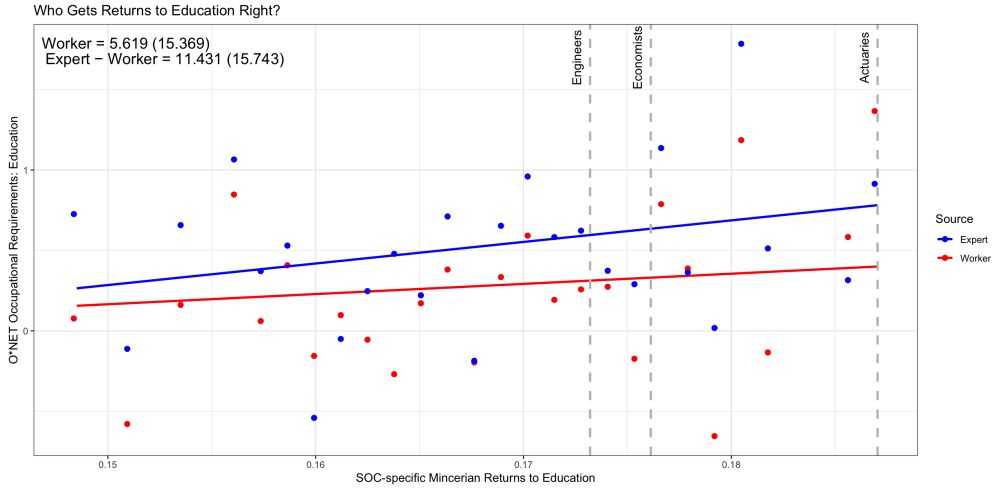
(c) Prevailing and Offered Wages

Figure 5: Effects of Expert Coding

Notes: These figures show the results of the event studies from Section 5. Relying on H-1B LCA data from USCIS, we show the effect of a change in O*NET coding source from incumbent worker to an occupational expert on: (i) the experience and education requirements of the occupations for which H-1B visas are requested by firms, (ii) the probability that these applications are accepted by the DOL, and (iii) the prevailing and offered wages associated with the LCA. The regressions rely on the universe of LCAs from 2009 to 2023, linked to O*NET data using the LCAs' listed 2018 SOC codes and the year in which the LCA was submitted.



(a)



(b)

Figure 6: Expert and Worker O*NET Values vs Labor Market Returns to Experience and Education

Notes: Returns to experience at the occupation level are estimated in the ACS from 2002 to 2022 using the following Mincerian regression:

$$\log(w_{oy}) = male_{oy} + white_{oy} + education_{oy} + \sum_o \beta_o^{exp} experience_{oy} \times \delta_o + \varepsilon_{oy},$$

where demographics are taken as yearly shares, and education and experience are measured in years. Returns to education at the occupation level are estimated using:

$$\log(w_{oy}) = male_{oy} + white_{oy} + experience_{oy} + \sum_o \beta_o^{educ} education_{oy} \times \delta_o + \varepsilon_{oy}.$$

On the binscatters, we visualize occupation-time level O*NET experience (education) requirements against β_o^{exp} (β_o^{educ}); we run this separately on the subset of the O*NET data rated by workers and experts. We report regression coefficients α and γ from:

$$\text{O*NET Requirement}_{ot} = \alpha \text{Returns}_o + \beta \text{expert}_{ot} + \gamma \text{Returns}_o \times \text{expert}_{ot} + \delta_t + \varepsilon_{ot},$$

where standard errors are clustered at the occupation level. t represents database time.

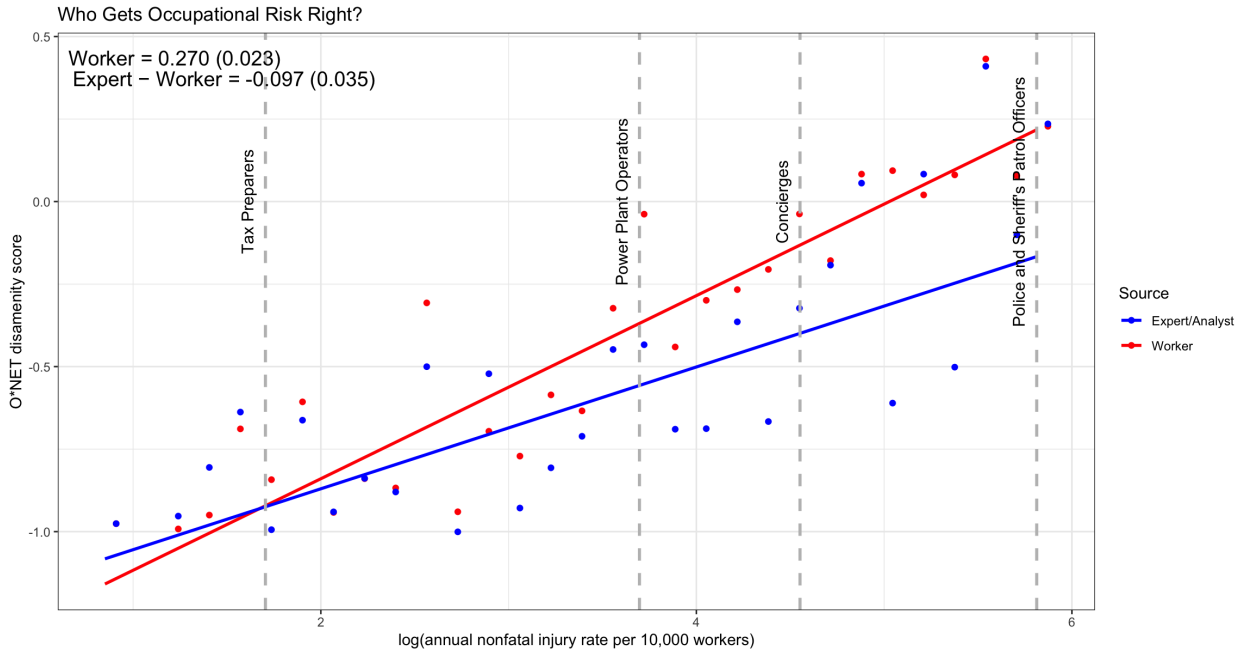


Figure 7: Who is Right: Injuries

Notes: Occupation-level injury rates are averaged from the BLS Survey of Occupational Injuries and Illnesses across 2011 to 2020. On the binscatter, we visualize occupation-time level O*NET disamenity scores against $\log(\text{InjuryRate})_o$; we run this separately on the subset of the O*NET data rated by workers and experts. We report regression coefficients α and γ from:

$$\text{DisamenityScore}_{ot} = \alpha \log(\text{InjuryRate})_o + \beta \text{expert}_{ot} + \gamma \log(\text{InjuryRate})_o \times \text{expert}_{ot} + \delta_t + \epsilon_{ot},$$

where standard errors are clustered at the occupation level.

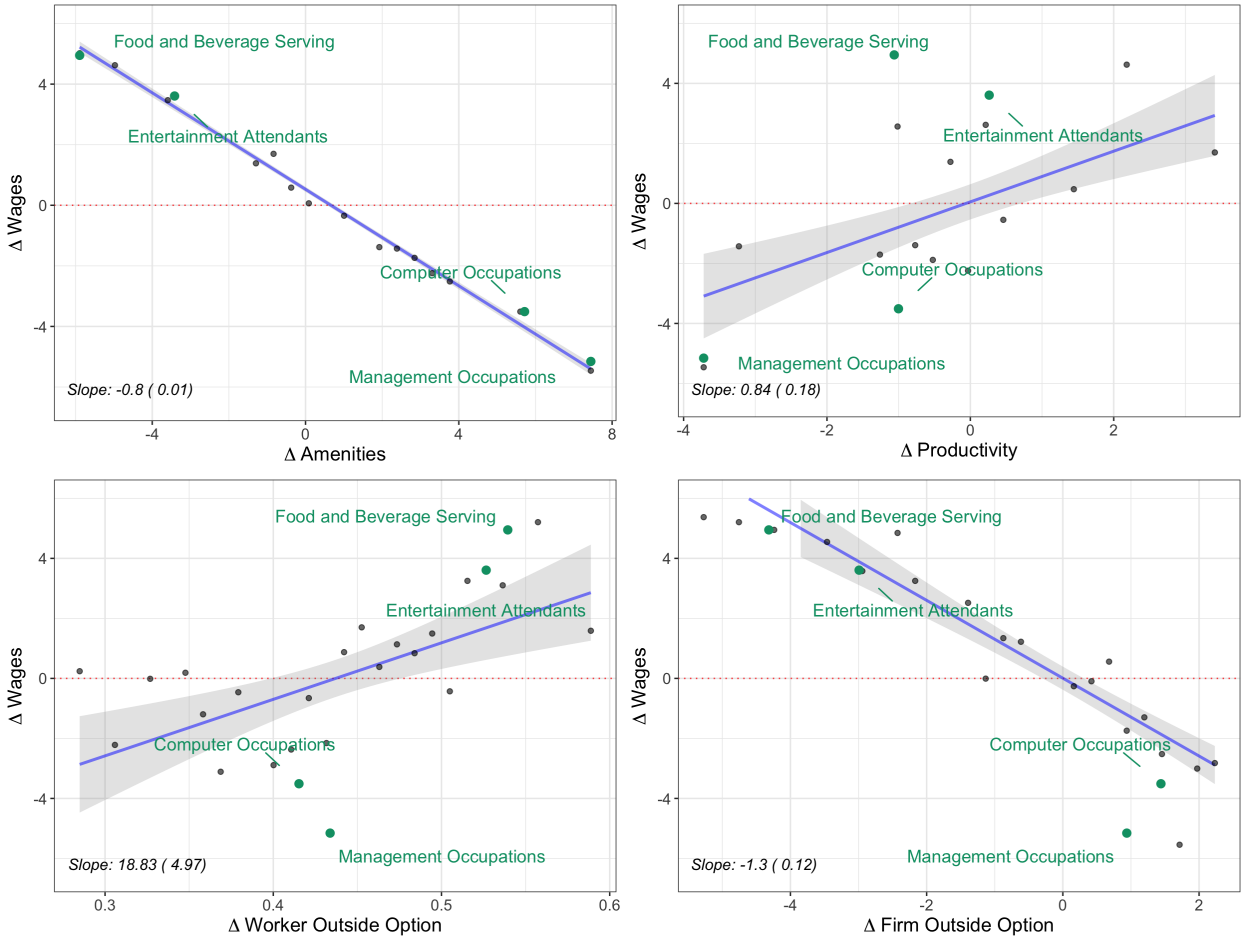


Figure 8: Counterfactuals from Eliminating Private Information

Notes: This figure describes the correlations in changes that arise when moving from a baseline equilibrium, where private information is valued by workers and firms, to a counterfactual equilibrium where private information is no longer used by agents when making matching decisions. This comparison holds the σ_w and σ_f parameters constant at their estimated values. The top-left panel shows the correlation between changes in amenities and changes in wages; the top-right panel shows the correlation between changes in productivity and wages. On the bottom row, the left panel plots the correlation between changes in workers' outside options and changes in wages, while the right panel plots the correlation between changes in firms' outside options and changes in wages. The correlations exhibit the expected signs: occupations that experience larger decreases in amenities or increases in productivity see their wages increase, while occupations that experience an increase in worker outside options or a decrease in firm outside options also see their wages increase.

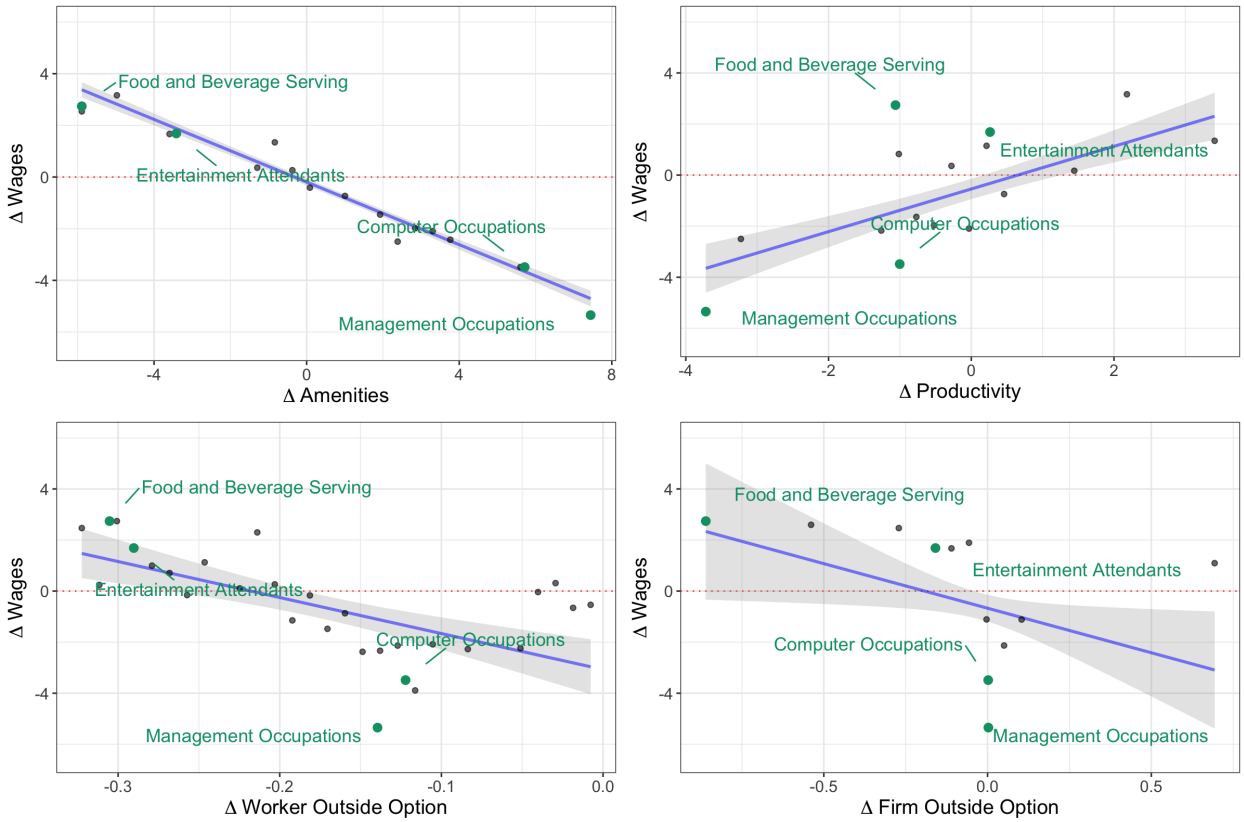


Figure 9: Counterfactuals from Eliminating Private Information in a Market with Low Frictions

Notes: This figure describes the correlations in changes that arise when moving from a counterfactual baseline equilibrium, where private information is valued by workers and firms, to a counterfactual equilibrium where private information is no longer used by agents when making matching decisions. This comparison considers a low-friction market, where the σ_w and σ_f parameters are reduced by 80% relative to their estimated values. The top-left panel shows the correlation between changes in amenities and changes in wages; the top-right panel shows the correlation between changes in productivity and wages. On the bottom row, the left panel plots the correlation between changes in workers' outside options and changes in wages, while the right panel plots the correlation between changes in firms' outside options and changes in wages. All correlations but one exhibit the expected signs: occupations that experience larger decreases in amenities or increases in productivity see their wages increase, and occupations that experience a decrease in firm outside options also see their wages increase. The exception is that occupations experiencing an increase in worker outside options also experience a decrease in wages. This unexpected correlation reflects sorting: workers whose outside options improve sort relatively more into high-amenity occupations, and therefore accept lower wages. In contrast to the market with frictions, workers who gain in outside options sort into higher-amenity occupations rather than receiving higher wages through increased bargaining power.

Table A: *Occupations with 10 Most Disagreed Upon Skills*				
Rank	Occupation	Most Disagreed Skill	Skill Value Worker	Skill Value Expert
1	Microbiologists	Equipment Maintenance	4.13	1.00
2	Anesthesiologists	Equipment Selection	4.48	1.37
3	Management Analysts	Operation and Control	4.18	1.08
4	Athletes and Sports Competitors	Equipment Selection	4.08	1.03
5	Media and Communication Workers, All Other	Equipment Selection	4.24	1.30
6	Terrazzo Workers and Finishers	Installation	4.23	1.29
7	Interior Designers	Installation	3.93	1.00
8	Tire Builders	Equipment Maintenance	4.49	1.57
9	Model Makers, Wood	Equipment Selection	4.31	1.40
10	Rotary Drill Operators, Oil and Gas	Installation	4.03	1.12
Table B: *Occupations with 10 Most Disagreed Upon Disamenities*				
Rank	Occupation	Most Disagreed Disamenity	Disamenity Value Worker	Disamenity Value Expert
1	Healthcare Support Workers, All Other	Wear Common Protective Equipment	1.38	4.83
2	Genetic Counselors	Wear Common Protective Equipment	4.52	1.38
3	Gas Station Operators	Sounds Are Distracting, Uncomfortable	4.66	2.10
4	Gas Plant Operators	Exposed to Hazardous Equipment	3.89	1.40
5	Chemical Equipment Operators	Outdoors, Exposed to Weather	4.07	1.60
6	Petroleum Pump & Refinery Operators	Exposed to High Places	3.96	1.53
7	Magnetic Resonance Imaging Technologists	Exposed to Radiation	4.01	1.63
8	Power Plant Operators	Exposed to High Places	4.13	1.80
9	Coating, Painting, Spraying Machine Operators	Wear Specialized Protective Equipment	3.82	1.50
10	First-Line Supervisors of Landscaping & Groundskeeping Workers	Sounds Are Distracting, Uncomfortable	4.12	1.80

Table 1: Occupations with 10 Highest Skill and Disamenity Disagreement Between Workers and Experts

Notes: In this table we first compute the maximum difference (in absolute value) between average worker and expert skill ratings, within occupation and across skills from the O*NET skill module. We then show the 10 occupations with the largest maximum difference, and the skill on which this maximum disagreement occurs, alongside worker and expert values. We do the same for the maximum difference in ratings for Context and Activities variables that relate to disamenities. We use for this O*NET data for all the databases we clean, using incumbent worker data for workers, and occupational experts and analysts for experts. Values are average within occupation and O*NET element, across databases, prior to being compared within occupation and across rater. The physical disamenities variables selected from the Context and the Activities modules are the ones listed in Appendix A.1.

Dependent Variable:	Hourly Wage (log)						
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Skill from Workers	0.2729***		0.1814***		0.4036***	0.2827***	
	(0.0288)		(0.0240)		(0.0453)	(0.0352)	
Skill from Experts		0.2171***		0.1176***	-0.2225***	-0.1697***	
		(0.0501)		(0.0374)	(0.0643)	(0.0476)	
Skill from Workers × Years of Educ.							0.0271***
							(0.0035)
Skill from Experts × Years of Educ.							-0.0048
							(0.0046)
Fixed effects							
Mincerian Controls	No	No	Yes	Yes	No	Yes	Yes
Soc 2018	No	No	No	No	No	No	Yes
Fit statistics							
Observations	19,285,970	19,285,970	19,285,970	19,285,970	19,285,970	19,285,970	19,285,970
R^2	0.17	0.07	0.31	0.27	0.19	0.32	0.38

Standard errors clustered by occupation.
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 2: Comparing O*NET Skills to Wages

Notes: For this specific exercise, we take an extensive definition of what corresponds to a skill in O*NET, specifically taking all variables from the Activities, Contexts, Education, Experience and Training, Skills, Knowledge and Work Style modules, except the ones that we identify as (dis)amenities, which are specified in Measuring Physical Disamenities, Appendix A.1. We then sign each skill variable such that an increase in its value increases observed wages in the ACS. This is done by regressing the relevant O*NET variable on observed ACS log wages, controlling for occupation fixed effects. Finally, we take the average within the occupation across all these signed skill variables, respectively for values given by incumbent workers and by analysts and experts. We then track the correlation of these skill variables with observed hourly wages in the ACS.

Occupation	Education (W,E)	Injury (W,E)	Experience (W,E)	ΔA, ΔΓ	Δ O.O. (W,F)	Δ Wage	Δ Count
Average Worker							
Average Worker	0.02, -0.04	0.26, 0.26	0.11, 0.03	0.66, -0.06	0.44, 0.01	0.00	95.1k
Lowest & Highest Wage Occupations in Baseline							
Entertainment Attendants	-0.50, -0.56	0.00, 0.00	-1.13, -0.15	-3.42, 0.26	0.53, -2.99	3.61	-10.4k
Management Occupations	0.91, 1.00	0.00, 0.00	2.46, 2.44	7.44, -3.72	0.43, 0.94	-5.16	1586.0k
Occupations with Highest/Lowest Education Private Information							
Legal Support Workers	2.07, 1.58	0.00, 0.00	0.77, 2.10	2.24, -3.26	0.43, -1.05	-1.43	-23.8k
Production Workers	-0.92, -0.97	0.00, 0.00	-0.42, 0.20	-1.25, -0.22	0.45, -1.02	1.41	-100.1k
Occupations with Highest/Lowest Experience Private Information							
Top Executives	0.91, 1.00	0.00, 0.00	2.46, 2.44	7.44, -3.72	0.40, 1.82	-5.58	348.3k
Home Health and Personal Care Aides	-0.36, -0.71	0.00, 0.00	-1.68, -1.39	-5.11, 2.12	0.47, -1.33	4.40	-245.4k
Occupations with Highest/Lowest Injury Private Information							
Construction Trade Workers	-0.93, -0.67	1.00, 1.00	-0.15, 0.48	0.90, -0.76	0.36, 0.12	-0.56	104.6k
Computer Occupations	1.61, 1.22	0.00, 0.00	1.90, 0.61	5.72, -1.00	0.42, 1.44	-3.51	1,009k
Occupations with Largest Surplus Changes							
Computer Occupations	1.61, 1.22	0.00, 0.00	1.90, 0.61	5.72, -1.00	0.42, 1.44	-3.51	1,009k
Food Preparation and Serving Related Occupations	-1.19, -0.37	0.00, 0.00	-1.95, 0.73	-5.89, -1.06	0.54, -4.32	4.95	-263.4k

Table 3: Removing Private Information in a Market with Frictions: Evolution of Wages & Utilities

Notes: This table describes the effect of moving from a baseline equilibrium where private information is valued by workers and firms, to a counterfactual equilibrium where private information is no longer used by agents when making matching decisions. This comparison holds the σ_w and σ_f parameters constant at their estimated values. The column labeled "Education (W,E)" gives the value of the O*NET education variables evaluated using incumbent data, or expert and analyst data for the occupation shown in the corresponding row. The columns labeled "Injury (W,E)" and "Experience (W,E)" provide similar information for the O*NET experience and injury variables. The columns labeled ΔA and $\Delta\Gamma$ indicate the change in amenities and productivity from removing private information. The column labeled "Δ O.O. (W,F)" indicates the change in worker and firm outside options. The last two columns indicate the change in wages and match masses produced by the removal of private information.

Occupation	Education (W,E)	Injury (W,E)	Experience (W,E)	ΔA, ΔΓ	Δ O.O. (W,F)	Δ Wage	Δ Count
Average Worker							
Average Worker	0.02, -0.04	0.26, 0.26	0.11, 0.03	0.66, -0.06	-0.18, -0.02	-0.59	-680.2k
Lowest & Highest Wage Occupations in Baseline							
Entertainment Attendants	-0.50, -0.56	0.00, 0.00	-1.13, -0.15	-3.42, 0.26	-0.29, -0.16	1.69	-26.2k
Management Occupations	0.91, 1.00	0.00, 0.00	2.46, 2.44	7.44, -3.72	-0.14, 0.00	-5.35	37.5k
Occupations with Highest/Lowest Education Private Information							
Legal Support Workers	2.07, 1.58	0.00, 0.00	0.77, 2.10	2.24, -3.26	-0.14, -0.01	-2.50	-6.3k
Production Workers	-0.92, -0.97	0.00, 0.00	-0.42, 0.20	-1.25, -0.22	-0.21, -0.01	0.36	-11.6k
Occupations with Highest/Lowest Experience Private Information							
Top Executives	0.91, 1.00	0.00, 0.00	2.46, 2.44	7.44, -3.72	-0.11, 0.00	-5.33	10.2k
Home Health and Personal Care Aides	-0.36, -0.71	0.00, 0.00	-1.68, -1.39	-5.11, 2.12	-0.25, -0.00	3.15	-6.0k
Occupations with Highest/Lowest Injury Private Information							
Construction Trade Workers	-0.93, -0.67	1.00, 1.00	-0.15, 0.48	0.90, -0.76	-0.14, 0.00	-1.09	0.7k
Computer Occupations	1.61, 1.22	0.00, 0.00	1.90, 0.61	5.72, -1.00	-0.12, 0.00	-3.49	15.1k
Occupations with Largest Surplus Changes							
Computer Occupations	1.61, 1.22	0.00, 0.00	1.90, 0.61	5.72, -1.00	-0.12, 0.00	-3.49	15.1k
Food and Beverage Serving Workers	-1.19, -0.37	0.00, 0.00	-1.95, 0.73	-5.89, -1.06	-0.31, -0.86	2.74	-636.2k

Table 4: Removing Private Information in a Market without Frictions: Evolution of Wages & Utilities

Notes: This table describes the effect of moving from a counterfactual baseline equilibrium where private information is valued by workers and firms, to a counterfactual equilibrium where private information is no longer used by agents when making matching decisions. This comparison considers a low-friction market, where the σ_w and σ_f parameters are reduced by 80% relative to their estimated values. The column labeled "Education (W,E)" gives the value of the O*NET education variables evaluated using incumbent data, or expert and analyst data for the occupation shown in the corresponding row. The columns labeled "Injury (W,E)" and "Experience (W,E)" provide similar information for the O*NET experience and injury variables. The columns labeled ΔA and $\Delta\Gamma$ indicate the change in amenities and productivity from removing private information. The column labeled "Δ O.O. (W,F)" indicates the change in worker and firm outside options. The last two columns indicate the change in wages and match masses produced by the removal of private information.

Appendix for online publication:

The Worker's Brain Under the Expert's Hat: Occupation Coding, Tacit Knowledge, and Wage-Setting

A Data Details

A.1 Construction of O*NET Data

We build our O*NET data using databases ranging from O*NET 3.0 released in August 2000, until O*NET version 29.0 released in August 2024, for a total of 58 O*NET versions. Across these, we clean the files: "Work Activities", "Work Context", "Education, Training and Experience", "Skills", "Task Ratings" and "Task Statements" (from which we produce our task data), "Knowledge", "Work Styles", "Job Zones", "Tools and Technologies" and "Tools Used" (from which we produce a unique tools dataset), and "Abilities".

O*NET files indicate what the source of the data is. Possible values are: "Legacy Analysts", "Incumbents", "Occupational Experts", "Analysts", "Transition Analysts". Tools and Abilities data is systematically rated by "Analysts" in O*NET, who are trained industrial psychologists. They also rate the "Skills" module starting in 2010. Regarding the other modules, the brunt of the O*NET data comes from surveys, where respondents are either incumbent workers in the occupation surveyed, or occupational experts who have experience related to the occupation surveyed. Additionally, early O*NET databases relied on direct translations of data from the Dictionary of Occupational Titles (DOT) into the relevant occupational codes and O*NET elements. The first O*NET datasets where no occupation is coded using DOT data arrive in 2008. Additionally, when changes in SOC codes (moving from SOC 2000 to SOC 2010, or from SOC 2010 to SOC 2018) lead to the splitting of an occupational code into two, or the merging of several occupations into a unique occupation, an industrial psychologist is tasked with coding the new occupations based on prior O*NET data. These are the "Transition Analysts". After this initial coding, occupations revert to

being filled either by incumbents or occupational experts. Finally, while most new occupations are directly coded using data from incumbents or occupational experts, it happens that new occupations are also first coded by industrial psychologists. These are the 'Analysts'.

Unless specified, our analysis will focus on data coming from incumbent workers or occupational experts. We build from this a dummy variable $DummySource_{ot}$ at the occupation o and time period t, which takes the value of 1 if the data was coded by an occupational expert or analyst, and the value of 0 if it was coded by an incumbent worker.

O*NET data is collected at the level of O*NET occupational codes, which are a strictly finer occupational category than the BLS Standard Occupational Classification (SOC). These codes evolve over times, following the major updates of the SOC system. In order to obtain a panel of comparable occupations, we crosswalk all occupational codes to the 2018 SOC codes. First, O*NET occupational codes are crosswalked into the relevant SOC codes, and we then use the SOC 2000 to SOC 2010 crosswalk, and the SOC 2010 to SOC 2018 crosswalk to obtain a dataset with comparable occupation identifiers. SOC 2018 codes such that they can be matched to several 2010 or 2000 SOC codes are aggregated. As such, all SOC 2000 or 2010 SOC codes are matched at most to one SOC 2018 code. This yields a set of 793 unique occupational codes.[18] When this aggregate results in several observations per 2018 SOC occupation within the same O*NET database, and for the same O*NET variable, we average the $DummySource_{ot}$ variable, and use the average data value. When possible, the analysis is also performed at the O*NET-SOC 2019 level, keeping only occupations that are crosswalked 1-to-1 through all the O*NET databases, in order to prevent any bias to arise from aggregation.

Measuring Physical Disamenities: The exact elements are (from work activities): Handling and Moving Objects, Controlling Machines and Processes,Operating Vehicles, Mechanized Devices, or Equipment, Performing General Physical Activities. (from work

[18]Note that for all O*NET versions ranging from O*NET 13.0 to O*NET 15.0, O*NET occupational codes are only partially based on the 2010 SOC system. Here, we use internal O*NET crosswalks to translate them into O*NET-SOC 2010 codes, and then to 2010 SOC and 2018 SOC codes.

contexts): Frequency of Conflict Situations, Deal With Unpleasant or Angry People, Deal With Physically Aggressive People, Indoors, Not Environmentally Controlled, Outdoors, Exposed to Weather, Outdoors, Under Cover, In an Open Vehicle or Equipment, In an Enclosed Vehicle or Equipment, Physical Proximity, Sounds, Noise Levels Are Distracting or Uncomfortable, Very Hot or Cold Temperatures, Extremely Bright or Inadequate Lighting, Exposed to Contaminants, Cramped Work Space, Awkward Positions, Exposed to Whole Body Vibration, Exposed to Radiation, Exposed to Disease or Infections, Exposed to High Places, Exposed to Hazardous Conditions, Exposed to Hazardous Equipment, Exposed to Minor Burns, Cuts, Bites, or Stings, Spend Time Sitting, Spend Time Standing, Spend Time Climbing Ladders, Scaffolds, or Poles, Spend Time Walking and Running, Spend Time Kneeling, Crouching, Stooping, or Crawling, Spend Time Keeping or Regaining Balance, Spend Time Using Your Hands to Handle, Control, or Feel Objects, Tools, or Controls, Spend Time Bending or Twisting the Body, Spend Time Making Repetitive Motions, Wear Common Protective or Safety Equipment such as Safety Shoes, Glasses, Gloves, Hearing Protection, Hard Hats, or Life Jackets, Wear Specialized Protective or Safety Equipment such as Breathing Apparatus, Safety Harness, Full Protection Suits, or Radiation Protection, Degree of Automation, Pace Determined by Speed of Equipment.

A.2 2010 Skill Reform

The most significant change in O*NET data collection in our period of analysis is the 2010 shift from collecting Skills-related data from incumbents and occupational experts, to having analysts (trained industrial psychologists) grade occupational skills (as illustrated in Figure 1).

O*NET had previously analyzed the relative performance of incumbents and analysts in grading Skills data. In a 2006 report, they compared the grading from both sources for 289 different SOC codes. While they found that there was twice as much variability in skill grading among incumbents than among analysts, they noted that this difference could

represent the diversity of jobs held by incumbents within an occupation code. They also noted that inter-rater reliability for the mean skill rating was comparable across incumbents and analysts. Finally, incumbents were found to give generally higher scores to skills' occupational importance than analysts. The conclusion of the report states: "*We found no clear evidence that one source of raters provides more valid or accurate information than the other. Therefore we suggest that the selection of rater source be made on the basis of other factors, including theoretical and practical considerations*" (Tsacoumis and Iddekinge (2006)).

Following this report, O*NET nonetheless decided to shift the grading of Skills to analysts, using the grading system implemented for the Abilities module. This change was justified by previous industrial psychology literature arguing that "*incumbents are more likely than analysts to deliberately inflate their ratings to influence policy decisions as those associated with compensation and training*" (Tsacoumis and Willison (2010)).

Analysts are provided with O*NET data which should be the sole base for the evaluation. This data contains the following elements: occupational title, definition and job zone; tasks statements (definitions) and data; knowledge statement and data; generalized work activity (GWA) statements and data; work context statements and data. Task data is provided for the task-occupation pairs with a past relevance rating in O*NET over 67%, and a mean importance rating above 3—these are called core tasks—a list of supplementary tasks is also provided if their relevance is above 67%, but their past mean importance is below 3. Knowledge data is presented if importance ratings are above 3—and out of these only the first 10 are shown. GWA information is only presented if: 1) its past mean value is above 3, and that GWA is required to perform the graded skill successfully (see the linkage described in the next paragraph). Work context (WC) data is shown if the past mean value is above 3, and if the skill-WC was identified as relevant.

Skill-GWA/WC linkages: The relevance of specific GWAs and WCs for the grading of an occupation-skill pair was decided in a one time exercise, using data collected from 8

experienced industrial/organizational psychologists. For each pair, the participants made a dichotomous judgment regarding whether the skill was needed to perform each GWA/WC. A majority of judges (5) had to decide the skill was needed.

16 trained occupational analysts are then responsible for the grading the importance and level of the 35 skills for each O*NET occupations, using the provided O*NET data. There is a minimum of 8 raters per occupation. Ratings are reviewed for errors, or common rating errors (leniency or severity). If the standard error of the mean of the importance rating for a skill-occupation pair is large (over .51), it is deemed to have insufficient agreement across raters—and raters are asked to reconsider their ratings.

To study the effect of the 2010 reform, we use the following regression, tracking skill ratings $Skill_{oet}$ for occupation o, skill category e in year t:

$$Skill_{oet} = \sum_{\tau \neq 2009} \delta_\tau 1\{\tau = t\} + \eta_{oe} + \varepsilon_{oet}$$

Standard errors are clustered at the SOC 2018 occupation level. We classify skills into four categories, in order to better highlight heterogeneous effects of the change in grading, and the direction of the analyst biases. The categories are: cognitive skills, social skills, technical skills and management skills.[19] The analysis is done on a balanced set of skills, removing O*NET data pre-2010 that was not graded by either an incumbent worker or occupational expert.[20]

[19]We categorize skills in the following way: **Cognitive Skills**: Active Learning, Active Listening, Complex Problem Solving, Critical Thinking, Learning Strategies, Mathematics, Reading Comprehension, Science, Speaking, Writing, Judgment and Decision Making, Systems Analysis, Systems Evaluation. **Social Skills**: Coordination, Instructing, Negotiation, Persuasion, Service Orientation, Social Perceptiveness. **Technical Skills**: Equipment Maintenance, Equipment Selection, Installation, Operation and Control, Operation Monitoring, Operations Analysis, Operations Monitoring, Programming, Quality Control Analysis, Repairing, Technology Design, Troubleshooting. **Management Skills**: Management of Financial Resources, Management of Material Resources, Management of Personnel Resources, Monitoring, Time Management.

[20]This has the effect of mostly removing O*NET data translated directly from the Dictionary of Occupational Titles.

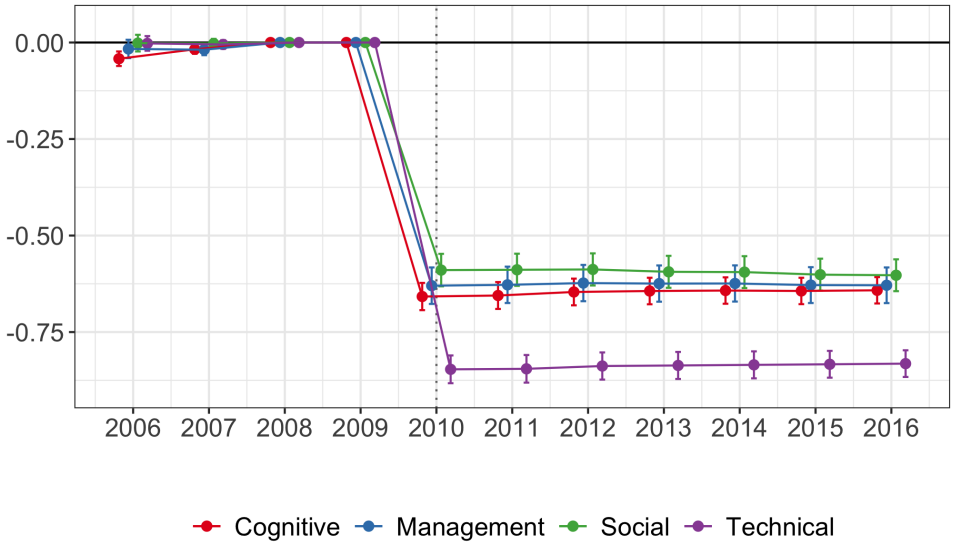


Figure A1: Baseline 2010 Reform Effect, by Skill Category

Notes: We show the event study regression tracking the effect of the 2010 O*NET Skill rating reform. Standard errors are clustered at the SOC 2018 occupation level.

As one could expect, the deskilling was stronger for occupations that were relatively more frequently graded by workers prior to 2010. Here, we compute the share of the grading done by incumbent workers relative to occupational experts during [2007,2009] prior to the reform. We run for this:

$$Skill_{oet} = \beta_o ShareWorker_o + \sum_{\tau \neq 2009} \delta_\tau 1\{\tau = t\} ShareWorker_o + \eta_{oe} + \xi_{et} + \zeta_{a(o)t} + \varepsilon_{oet}$$

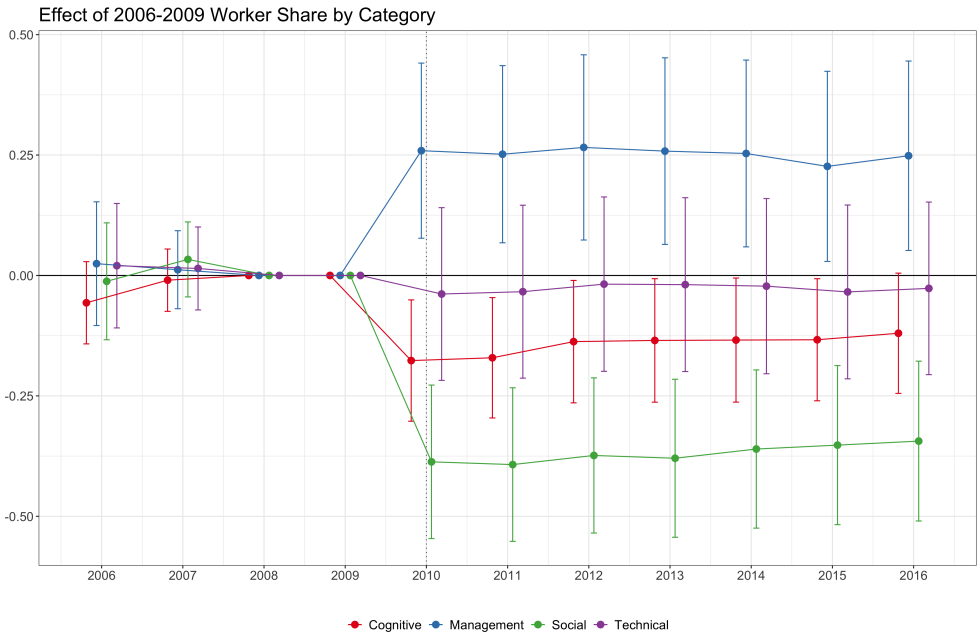


Figure A2: Effect of Pre-Reform Incumbent Share, by Category

Notes: We show the event study regression tracking the relative effect of a larger incumbent worker share over [2007,2009] on the the 2010 O*NET Skill rating reform. Standard errors are clustered at the SOC 2018 occupation level.

Interestingly, we see that analysts do not systematically downgrade worker skill scores. Analysts downgrade skills related to social and cognitive skills, but upgrade skills related to management.

Finally, we perform the same analysis, this time looking at the effect of a larger occupational tacitness on the skill shock. We use the following regression:

$$Skill_{oet} = \beta_o Tacit_o + \sum_{\tau \neq 2009} \delta_\tau 1\{\tau = t\} Tacit_o + \eta_{oe} + \xi_{et} + \zeta_{a(o)t} + \varepsilon_{oet}$$

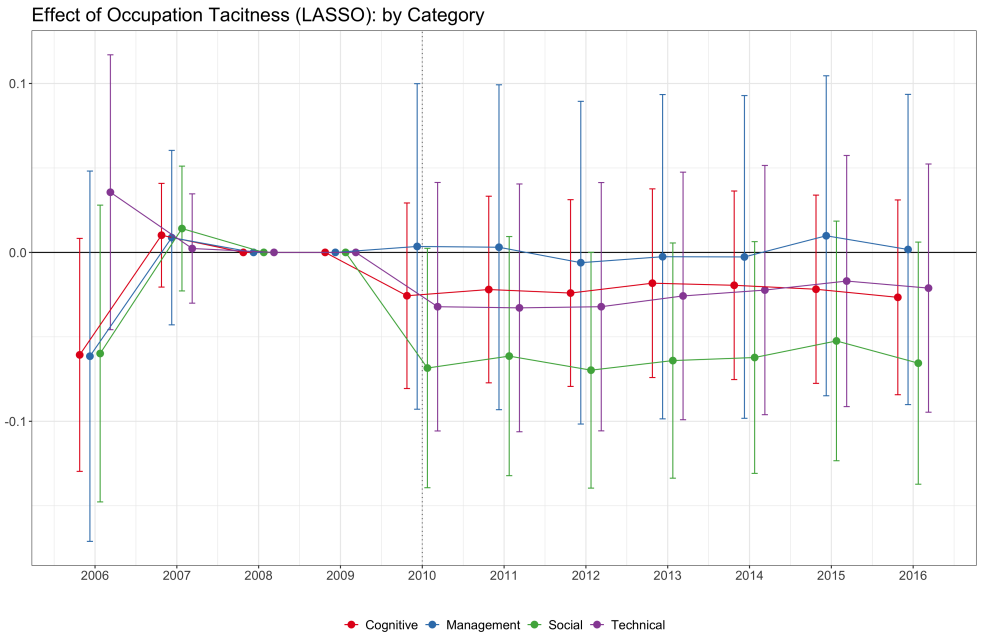


Figure A3: Effect of Tacitness (LASSO), by Category

Notes: We show the event study regression tracking the relative effect of a larger occupational tacitness on the the 2010 O*NET Skill rating reform. Standard errors are clustered at the SOC 2018 occupation level.

B Model Derivations and Estimation

B.1 Second Step Equilibrium Characterization

Notation: We introduce the following interim expressions, which denote the conditional expected utility of agents when they know their type, but not the type of the other: $\overline{\rho}_w(\xi_w) = \mathbb{E}_{\xi_f}\left[\rho(\xi_w,\xi_f)|\xi_w\right], \overline{\rho}_f(\xi_f) = \mathbb{E}_{\xi_w}\left[\rho(\xi_w,\xi_f)|\xi_f\right], \overline{t}_w(\xi_w) = \mathbb{E}_{\xi_f}\left[t_w(\xi_w,\xi_f)|\xi_w\right], \overline{t}_f(\xi_f) = \mathbb{E}_{\xi_w}\left[t_f(\xi_w,\xi_f)|\xi_f\right]$. On the worker side, the interim expected utility from reporting a type $\widetilde{\xi}_w$ when one's type is ξ_w is:

$$\widetilde{u}_{xy}(\widetilde{\xi}_w,\xi_w) = \Big(A_{xy} + \xi_w\Big)\overline{\rho}_w(\widetilde{\xi}_w) + \overline{t}_w(\widetilde{\xi}_w) + \varepsilon_y$$

And we write the firm equivalent with $\widetilde{v}_{xy}(\widetilde{\xi}_f,\xi_f)$. We write the interim expected utility from truthful reporting for the worker as $\widetilde{U}_{xy}(\xi_w) = \widetilde{u}_{xy}(\xi_w,\xi_w)$, and use $\widetilde{V}_{xy}(\xi_f)$ on the firm side. Finally, we write the worker ex ante expected utility from truthful reporting as

$U_{xy} = \mathbb{E}_{\xi_w}\left[\widetilde{U}_{xy}(\xi_w)\right]$, and use V_{xy} on the firm side. We assume that $\widetilde{U}_{xy}(\xi_w)$ and $\widetilde{V}_{xy}(\xi_f)$ are differentiable.

Post-Matching Equilibrium: We will focus on direct revelation mechanisms. As such, a mechanism will be defined by the tuple of potential worker and firm actions, the allocation function, and the transfer function:

$$\Omega = \left\{ \left\{ [\underline{\xi_w}, \overline{\xi_w}], [\underline{\xi_f}, \overline{\xi_f}] \right\}, \rho, \left\{ t_f, t_w \right\} \right\}$$

We will require that our mechanism satisfies interim incentive compatibility (IC), interim participation (PC), and ex ante budget balance (BB). We now define these conditions in turn, give sufficient and necessary conditions for IC, and then give necessary conditions for a mechanism with increasing interim expected allocations to satisfy our three conditions. We finally define our chosen allocation rule, and explicit the conditions under which it corresponds to an equilibrium allocation.

Interim Incentive Compatibility (IC): On the worker side (a symmetric result holds on the firm side), a mechanism is said to be interim incentive compatible if:

$$\widetilde{u}_{xy}(\xi_w, \xi_w) \geq \widetilde{u}_{xy}(\widetilde{\xi}_w, \xi_w) \ , \forall \ \xi_w, \widetilde{\xi}_w$$

We show in B.3 that a mechanism is interim IC if and only if the following holds:

- $\overline{\rho}_w(.)$ is non-decreasing in ξ_w.
- $\widetilde{U}_{xy}(\xi_w) = \widetilde{U}_{xy}(\hat{\xi}_w) + \int_{\underline{\xi_w}}^{\xi_w} \overline{\rho}_w(x)dx \ , \forall \ \hat{\xi}_w, \xi_w$

Interim Participation Constraint (PC) We require that the following holds: $\widetilde{U}_{xy}(\xi_w) \geq \varepsilon_y \ , \forall \ \xi_w$ and $\widetilde{V}_{xy}(\xi_f) \geq \eta_x \ , \forall \ \xi_f$.

Ex Ante Bduget Balance (BB) We require that the mechanism is budget balanced at the ex ante stage, or that:

$$\mathbb{E}_{\xi}\left[\rho(\xi)\left\{t_f(\xi_w,\xi_f)-t_w(\xi_w,\xi_f)\right\}\right]\geq 0$$

Characterizing the equilibrium With $\overline{\rho}_w(.)$ and $\overline{\rho}_f(.)$ increasing, we show in B.4 that there exists a transfer scheme $\{t_f,t_w\}$ such that Ω satisfies interim IC, PC and ex ante BB only if:

$$\mathbb{E}_{\xi}\left[\rho(\xi)\left\{\xi_w+\xi_f+A_{xy}+\Gamma_{xy}-\frac{1-F(\xi_w))}{f(\xi_w)}-\frac{1-Q(\xi_f))}{q(\xi_f)}\right\}\right]\geq 0$$

Transfer Scheme: We consider the following transfer schemes, which are shown in B.5 to meet the IC, PC and BB conditions. This scheme maximizes the virtual surplus, and yields ex ante expected utilities that are readily usable in the first stage matching problem. Specifically, we consider, with w_{xy} a constant which we further define at the ex ante matching stage:

$$\overline{t}_w(\xi_w)=w_{xy}+\int_{\underline{\xi}_w}^{\xi_w}\overline{\rho}_w(x)dx-\left(A_{xy}+\xi_w\right)\overline{\rho}_w(\xi_w)-\mathbb{E}_{\xi_w}\left[\int_{\underline{\xi}_w}^{\xi_w}\overline{\rho}_w(x)dx\right]+\mathbb{E}_{\xi_w}\left[\left(A_{xy}+\xi_w\right)\overline{\rho}_w(\xi_w)\right]$$

$$\overline{t}_f(\xi_w)=w_{xy}-\int_{\underline{\xi}_f}^{\xi_f}\overline{\rho}_f(x)dx+\left(\Gamma_{xy}+\xi_f\right)\overline{\rho}_f(\xi_f)+\mathbb{E}_{\xi_f}\left[\int_{\underline{\xi}_f}^{\xi_f}\overline{\rho}_f(x)dx\right]-\mathbb{E}_{\xi_f}\left[\left(\Gamma_{xy}+\xi_f\right)\overline{\rho}_f(\xi_f)\right]$$

Note that the deviation between w_{xy} and $t_f(\xi)$ means that the model explicitly gives room to deviation between the unique expected wage at the ex ante stage, and the realized wage received by workers ex post, and gives rise to wage heterogeneity within the match at the ex post stage. With these, we can also write the expression for the ex ante expected utilities:

$$U_{xy}=\mathbb{E}_{\xi}\left[\left(A_{xy}+\xi_w\right)\overline{\rho}_w(\xi_w)\right]+w_{xy}+\varepsilon_y$$

$$V_{xy}=\mathbb{E}_{\xi}\left[\left(\Gamma_{xy}+\xi_f\right)\overline{\rho}_f(\xi_f)\right]-w_{xy}+\eta_x$$

Equilibrium Allocation Rule: The final step is to define the equilibrium allocation rule. This function has to exist in the $[0,1]$ space, only lead to a transfer when an ex ante budget balance transfer scheme exists, and be such that interim expected allocations are increasing (as assumed when we characterized the equilibrium). We use the following rule:

$$\rho(\xi) = \begin{cases} 1, & \text{if } \xi_w + \xi_f + A_{xy} + \Gamma_{xy} - \frac{1-F(\xi_w))}{f(\xi_w)} - \frac{1-Q(\xi_f))}{q(\xi_f)} \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

ρ is increasing if $\xi_w - \frac{1-F(\xi_w))}{f(\xi_w)}$ is increasing, as well as $\xi_f - \frac{1-Q(\xi_f))}{q(\xi_f)}$. This is the case if the distributions associated to the ξ_w and ξ_f random variables are regular. One sufficient condition for regularity is that the monotone hazard condition is met. An increasing ρ in turn implies that the interim expected allocations are also increasing. Finally, we can note that this allocation rule differs from the first best allocation $\rho^{FB}(\xi)$, and leads to less trade:

$$\rho^{FB}(\xi) = \begin{cases} 1, & \text{if } \xi_w + \xi_f + A_{xy} + \Gamma_{xy} \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

B.2 First Step Equilibrium Characterization

From the standard discrete choice derivations, we obtain the following conditional choice probabilities:

$$\pi_{y|x} = \frac{exp\left(\frac{\mathbb{E}_\xi\left[\left(A_{xy}+\xi_w\right)\overline{\rho}_w(\xi_w)\right]+w_{xy}}{\sigma_w}\right)}{\sum_{y'} exp\left(\frac{\mathbb{E}_\xi\left[\left(A_{xy'}+\xi_w\right)\overline{\rho}_w(\xi_w)\right]+w_{xy'}}{\sigma_w}\right)}$$

This in turn leads to the following mass of matches:

$$\mu_{xy} = exp\left(\frac{\mathbb{E}_\xi\left[\left(A_{xy}+\xi_w\right)\overline{\rho}_w(\xi_w)\right]+w_{xy}-a_x}{\sigma_w}\right)$$

with $a_x = -\sigma_w n_x + \sigma_w log\Bigg(\sum_{y'} exp\Bigg(\frac{\mathbb{E}_\xi\Big[\Big(A_{xy'}+\xi_w\Big)\overline{\rho}_w(\xi_w)\Big]+w_{xy'}}{\sigma_w}\Bigg)\Bigg)$.

On the firm side, we obtain:

$$\mu_{xy} = exp\Bigg(\frac{\mathbb{E}_\xi\Bigg[\Big(\Gamma_{xy}+\xi_f\Big)\overline{\rho}_f(\xi_f)\Bigg] - w_{xy} - b_y}{\sigma_f}\Bigg)$$

with $b_y = -\sigma_f m_y + \sigma_f log\Bigg(\sum_{x'} exp\Bigg(\frac{\mathbb{E}_\xi\Bigg[\Big(\Gamma_{x'y}+\xi_f\Big)\overline{\rho}_f(\xi_f)\Bigg] - w_{x'y}}{\sigma_f}\Bigg)\Bigg)$. Equilibrium implies equality between the two expressions for the mass of matches. This allows us to obtain the following, with $\tilde{\Phi}(x,y) = \mathbb{E}_\xi\big[\rho(\xi_w,\xi_f)\big(A_{xy}+\xi_w+\Gamma_{xy}+\xi_f\big)\big]$:

$$\mu_{xy} = exp\Bigg(\frac{\tilde{\Phi}(x,y) - a_x - b_y}{\sigma}\Bigg) \tag{3}$$

And the following expression for the wage:

$$\tilde{w}(x,y) = \frac{\sigma_f}{\sigma}\Bigg(a_x - \mathbb{E}_\xi\big[\big(A_{xy}+\xi_w\big)\overline{\rho}_w(\xi_w)\big]\Bigg) + \frac{\sigma_w}{\sigma}\Bigg(\mathbb{E}_\xi\Bigg[\Big(\Gamma_{xy}+\xi_f\Big)\overline{\rho}_f(\xi_f)\Bigg] - b_y\Bigg) \tag{4}$$

B.3 Proof of the Interim IC Characterization:

Necessity:

First, note that we can write:

$$\begin{aligned}\widetilde{u}_{xy}(\widetilde{\xi}_w,\xi_w) &= \Big(A_{xy}+\xi_w\Big)\overline{\rho}_w(\widetilde{\xi}_w) + \overline{t}_w(\widetilde{\xi}_w) + \varepsilon_y \\ &= \widetilde{U}_{xy}(\widetilde{\xi}_w) + \overline{\rho}_w(\widetilde{\xi}_w)\Big[\xi_w - \widetilde{\xi}_w\Big]\end{aligned}$$

Assume that $\widetilde{\xi}_w > \xi_w$, then if the mechanism is IC:

$$\begin{aligned}
&\widetilde{U}_{xy}(\xi_w) \geq \widetilde{u}_{xy}(\widetilde{\xi}_w, \xi_w) \\
\Leftrightarrow\ &\widetilde{U}_{xy}(\xi_w) - \widetilde{U}_{xy}(\widetilde{\xi}_w) \geq \overline{\rho}_w(\widetilde{\xi}_w)\left[\xi_w - \widetilde{\xi}_w\right] \\
\Leftrightarrow\ &\overline{\rho}_w(\widetilde{\xi}_w) \geq \frac{\widetilde{U}_{xy}(\widetilde{\xi}_w) - \widetilde{U}_{xy}(\xi_w)}{\widetilde{\xi}_w - \xi_w}
\end{aligned}$$

And we similarly get:

$$\begin{aligned}
&\widetilde{U}_{xy}(\widetilde{\xi}_w) \geq \widetilde{u}_{xy}(\xi_w, \widetilde{\xi}_w) \\
\Leftrightarrow\ &\overline{\rho}_w(\xi_w) \leq \frac{\widetilde{U}_{xy}(\widetilde{\xi}_w) - \widetilde{U}_{xy}(\xi_w)}{\widetilde{\xi}_w - \xi_w}
\end{aligned}$$

Hence, we have:

$$\overline{\rho}_w(\xi_w) \leq \frac{\widetilde{U}_{xy}(\widetilde{\xi}_w) - \widetilde{U}_{xy}(\xi_w)}{\widetilde{\xi}_w - \xi_w} \leq \overline{\rho}_w(\widetilde{\xi}_w)$$

This implies that $\overline{\rho}_w(.)$ is non-decreasing. We can further write:

$$\widetilde{U}_{xy}(\xi_w)' = \lim_{\widetilde{\xi}_w \to \xi_w} \frac{\widetilde{U}_{xy}(\widetilde{\xi}_w) - \widetilde{U}_{xy}(\xi_w)}{\widetilde{\xi}_w - \xi_w} = \overline{\rho}_w(\xi_w)$$

Finally:

$$\begin{aligned}
&\int_{\widetilde{\xi}_w}^{\xi_w} \widetilde{U}_{xy}(x)' dx = \widetilde{U}_{xy}(\xi_w) - \widetilde{U}_{xy}(\widetilde{\xi}_w) = \int_{\widetilde{\xi}_w}^{\xi_w} \overline{\rho}_w(x) dx \\
&\Rightarrow \widetilde{U}_{xy}(\xi_w) = \widetilde{U}_{xy}(\widetilde{\xi}_w) + \int_{\widetilde{\xi}_w}^{\xi_w} \overline{\rho}_w(x) dx
\end{aligned}$$

Sufficiency:

Consider $\xi_w > \widetilde{\xi}_w$, we have:

$$\widetilde{U}_{xy}(\xi_w) - \widetilde{U}_{xy}(\widetilde{\xi}_w) = \int_{\widetilde{\xi}_w}^{\xi_w} \overline{\rho}_w(x) dx \geq \int_{\widetilde{\xi}_w}^{\xi_w} \overline{\rho}_w(\widetilde{\xi}_w) dx = \left(\xi_w - \widetilde{\xi}_w\right) \overline{\rho}_w(\widetilde{\xi}_w)$$

Then, we have that:

$$\widetilde{U}_{xy}(\xi_w) \geq \widetilde{U}_{xy}(\widetilde{\xi}_w) + \Big(\xi_w - \widetilde{\xi}_w\Big)\overline{\rho}_w(\widetilde{\xi}_w) = \widetilde{u}_{xy}(\widetilde{\xi}_w, \xi_w)$$
$$\widetilde{U}_{xy}(\widetilde{\xi}_w) \geq \widetilde{U}_{xy}(\xi_w) + \Big(\widetilde{\xi}_w - \xi_w\Big)\overline{\rho}_w(\xi_w) = \widetilde{u}_{xy}(\xi_w, \widetilde{\xi}_w)$$

B.4 Proof of the Equilibrium Characterization

Proof: We can write the ex ante expected transfer received by the worker as, using the alternative characterization of the incentive constraint:

$$\begin{aligned}\mathbb{E}_{\xi}\Big[t_w(\xi)\rho(\xi)\Big] &= \mathbb{E}_{\xi_w}\Big[\bar{t}_w(\xi_w)\Big] = \mathbb{E}_{\xi_w}\Big[\widetilde{U}_{xy}(\xi_w) - \Big(A_{xy} + \xi_w\Big)\overline{\rho}_w(\xi_w) - \varepsilon_y\Big] \\ &= \mathbb{E}_{\xi_w}\Big[- \Big(A_{xy} + \xi_w\Big)\overline{\rho}_w(\xi_w) + \int_{\underline{\xi}_w}^{\xi_w} \overline{\rho}_w(x)dx\Big] + \widetilde{U}_{xy}(\underline{\xi}_w) - \varepsilon_y\end{aligned}$$

We then integrate by parts the expression $\mathbb{E}_{\xi_w}\Big[\int_{\underline{\xi}_w}^{\xi_w} \overline{\rho}_w(x)dx\Big]$. First, note that using the Leibniz rule, we get:

$$\frac{\partial}{\partial \xi_w}\int_{\underline{\xi}_w}^{\xi_w} \overline{\rho}_w(x)dx = \overline{\rho}_w(\xi_w).$$

Hence, we can write:

$$\begin{aligned}\mathbb{E}_{\xi_w}\Big[\int_{\underline{\xi}_w}^{\xi_w} \overline{\rho}_w(x)dx\Big] &= \int_{\underline{\xi}_w}^{\overline{\xi}_w}\int_{\underline{\xi}_w}^{\xi_w} \overline{\rho}_w(x)dx f(\xi_w)d\xi_w \\ &= \Big[\int_{\underline{\xi}_w}^{\xi_w} \overline{\rho}_w(x)dx F(\xi_w)d\xi_w\Big]_{\underline{\xi}_w}^{\overline{\xi}_w} - \int_{\underline{\xi}_w}^{\overline{\xi}_w} \overline{\rho}_w(x)F(x)dx \\ &= \int_{\underline{\xi}_w}^{\overline{\xi}_w} \overline{\rho}_w(x)\Big[1 - F(x)\Big]dx\end{aligned}$$

We can use this expression into our formula for the ex ante expected worker transfer:

$$\mathbb{E}_{\xi}\Big[t_w(\xi)\rho(\xi)\Big] = \mathbb{E}_{\xi_w}\Big[\overline{\rho}_w(\xi_w)\Big(- A_{xy} - \xi_w + \frac{1 - F(\xi_w)}{f(\xi_w)}\Big)\Big] + \widetilde{U}_{xy}(\underline{\xi}_w) - \varepsilon_y$$

We can obtain on the firm side a symmetric expression:

$$\mathbb{E}_{\xi}\Big[t_f(\xi)\rho(\xi)\Big] = \mathbb{E}_{\xi_f}\left[\overline{\rho}_f(\xi_f)\left(\Gamma_{xy} + \xi_f - \frac{1-Q(\xi_f)}{q(\xi_f)}\right)\right] - \widetilde{V}_{xy}(\underline{\xi}_f) + \eta_x$$

Pooling these together, we obtain:

$$\begin{aligned}
&\mathbb{E}_{\xi}\left[\rho(\xi)\left\{\xi_w + \xi_f + A_{xy} + \Gamma_{xy} - \frac{1-F(\xi_w))}{f(\xi_w)} - \frac{1-Q(\xi_f))}{q(\xi_f)}\right\}\right] = \\
&\mathbb{E}_{\xi}\left[\rho(\xi)\left\{t_f(\xi_w,\xi_f) - t_w(\xi_w,\xi_f)\right\}\right] + \left(\widetilde{U}_{xy}(\underline{\xi}_w) - \varepsilon_y\right) + \left(\widetilde{V}_{xy}(\underline{\xi}_f) - \eta_x\right)
\end{aligned}$$

The first element of the LHS is non-negative because of the ex ante budget balance constraint, and the second element is positive because of the interim participation constraint.

B.5 Proof that Transfers meet IC, PC and BB

Transfer Scheme: We consider the following transfer which ensures that IC, PC and BB hold, maximizes the virtual surplus, and yields ex ante expected utilities that are readily usable in the first stage matching problem. Specifically, we consider the following transfers:

$$\begin{aligned}
\overline{t}_w(\xi_w) &= w_{xy} + \int_{\underline{\xi}_w}^{\xi_w} \overline{\rho}_w(x)dx - \Big(A_{xy} + \xi_w\Big)\overline{\rho}_w(\xi_w) - \mathbb{E}_{\xi_w}\left[\int_{\underline{\xi}_w}^{\xi_w} \overline{\rho}_w(x)dx\right] + \mathbb{E}_{\xi_w}\left[\Big(A_{xy} + \xi_w\Big)\overline{\rho}_w(\xi_w)\right] \\
\overline{t}_f(\xi_w) &= w_{xy} - \int_{\underline{\xi}_f}^{\xi_f} \overline{\rho}_f(x)dx + \Big(\Gamma_{xy} + \xi_f\Big)\overline{\rho}_f(\xi_f) + \mathbb{E}_{\xi_f}\left[\int_{\underline{\xi}_f}^{\xi_f} \overline{\rho}_f(x)dx\right] - \mathbb{E}_{\xi_f}\left[\Big(\Gamma_{xy} + \xi_f\Big)\overline{\rho}_f(\xi_f)\right]
\end{aligned}$$

Note that the deviation between w_{xy} and $t_f(\xi)$ means that the model explicitly gives room to deviation between the unique expected wage at the ex ante stage, and the realized wage received by workers ex post, and gives rise to wage heterogeneity within the match at the ex post stage.

These transfers give rise to the following interim expected utilities:

$$\widetilde{U}_{xy}(\xi_w) = \mathbb{E}_{\xi_w}\left[\left(A_{xy} + \xi_w\right)\overline{\rho}_w(\xi_w)\right] + \left(\int_{\underline{\xi}_w}^{\xi_w} \overline{\rho}_w(x)dx - \mathbb{E}_{\xi_w}\left[\int_{\underline{\xi}_w}^{\xi_w} \overline{\rho}_w(x)dx\right]\right) + w_{xy} + \varepsilon_y$$

$$\widetilde{V}_{xy}(\xi_f) = \mathbb{E}_{\xi_f}\left[\left(\Gamma_{xy} + \xi_f\right)\overline{\rho}_f(\xi_f)\right] + \left(\int_{\underline{\xi}_f}^{\xi_f} \overline{\rho}_f(x)dx - \mathbb{E}_{\xi_f}\left[\int_{\underline{\xi}_f}^{\xi_f} \overline{\rho}_f(x)dx\right]\right) - w_{xy} + \eta_x$$

To show IC, we have assumed that interim expected allocations are increasing. We can further re-express interim expected utilities as follows:

$$\widetilde{U}_{xy}(\xi_w) = \widetilde{U}_{xy}(\underline{\xi}_w) + \int_{\underline{\xi}_w}^{\xi_w} \overline{\rho}_w(x)dx$$

$$\widetilde{V}_{xy}(\xi_f) = \widetilde{V}_{xy}(\underline{\xi}_f) + \int_{\underline{\xi}_f}^{\xi_f} \overline{\rho}_f(x)dx$$

Together, these imply that IC holds.

For the PC, we need:

$$\widetilde{U}_{xy}(\underline{\xi}_w) \geq \varepsilon_y \Leftrightarrow w_{xy} \geq \mathbb{E}_{\xi_w}\left[\int_{\underline{\xi}_w}^{\xi_w} \overline{\rho}_w(x)dx\right] - \mathbb{E}_{\xi_w}\left[\left(A_{xy} + \xi_w\right)\overline{\rho}_w(\xi_w)\right]$$

$$\widetilde{V}_{xy}(\underline{\xi}_f) \geq \eta_x \Leftrightarrow w_{xy} \leq \mathbb{E}_{\xi_f}\left[\left(\Gamma_{xy} + \xi_f\right)\overline{\rho}_f(\xi_f)\right] - \mathbb{E}_{\xi_f}\left[\int_{\underline{\xi}_f}^{\xi_f} \overline{\rho}_f(x)dx\right]$$

These together imply that we need a constant w_{xy} such that:

$$\mathbb{E}_{\xi_w}\left[\int_{\underline{\xi}_w}^{\xi_w} \overline{\rho}_w(x)dx\right] - \mathbb{E}_{\xi_w}\left[\left(A_{xy} + \xi_w\right)\overline{\rho}_w(\xi_w)\right] \leq w_{xy} \leq \mathbb{E}_{\xi_f}\left[\left(\Gamma_{xy} + \xi_f\right)\overline{\rho}_f(\xi_f)\right] - \mathbb{E}_{\xi_f}\left[\int_{\underline{\xi}_f}^{\xi_f} \overline{\rho}_f(x)dx\right]$$

Such a wage can exist if the following holds, which is just the equilibrium characterization condition that the expected virtual surplus is positive:

$$\mathbb{E}_{\xi}\left[\rho(\xi)\left\{\xi_w + \xi_f + A_{xy} + \Gamma_{xy} - \frac{1 - F(\xi_w))}{f(\xi_w)} - \frac{1 - Q(\xi_f))}{q(\xi_f)}\right\}\right] \geq 0$$

Finally, we can check the ex ante BB condition, which holds by design of the transfers:

$$\begin{aligned}\mathbb{E}_\xi\Big[t_f(\xi_w,\xi_f)-t_w(\xi_w,\xi_f)\Big] &= \mathbb{E}_\xi\left[\rho(\xi)\left\{\xi_w+\xi_f+A_{xy}+\Gamma_{xy}-\frac{1-F(\xi_w))}{f(\xi_w)}-\frac{1-Q(\xi_f))}{q(\xi_f)}\right\}\right] \\ &\quad - \mathbb{E}_\xi\left[\rho(\xi)\left\{\xi_w+\xi_f+A_{xy}+\Gamma_{xy}-\frac{1-F(\xi_w))}{f(\xi_w)}-\frac{1-Q(\xi_f))}{q(\xi_f)}\right\}\right] \\ &= 0\end{aligned}$$

B.6 Model Estimation

Following Galichon and Salanié (2022), we note that the sample log likelihood associated to the match data can be expressed as, with Θ the vector of parameters:

$$logL_m(\Theta) = \sum_{x\in\mathcal{X},y\in\mathcal{Y}} \hat{\mu}_{xy} log(\frac{\mu^\Theta_{xy}}{H^\Theta}) + \sum_{x\in\mathcal{X}} \hat{\mu}_{x0} log(\frac{\mu^\Theta_{x0}}{H^\Theta}) + \sum_{y\in\mathcal{Y}} \hat{\mu}_{0y} log(\frac{\mu^\Theta_{0y}}{H^\Theta}).$$

Here, $\{\hat{\mu}_{xy},\hat{\mu}_{xo},\hat{\mu}_{yo}\}$ are the counts of matches, unemployed workers, and vacancies observed in the sample. $\{\mu^\Theta_{xy},\mu^\Theta_{x0},\mu^\Theta_{y0}\}$ are the simulated counts given the parameter vector Θ. We detail next how these are computed. Finally, H^Θ is the simulated mass of agents, or:

$$H^\Theta = \sum_{x\in\mathcal{X},y\in\mathcal{Y}} \mu^\Theta_{xy} + \sum_{x\in\mathcal{X}} \mu^\Theta_{x0} + \sum_{y\in\mathcal{Y}} \mu^\Theta_{y0}.$$

Next, following Dupuy and Galichon (2022), we assume that the observed average wage for a pair $\{x,y\}$ is a noisy measure of the true unobserved transfer, such that:

$$\hat{w}_{xy} = w_{x,y} + \epsilon_{xy},$$

where measurement error ϵ_{xy} follows a $\mathcal{N}\big(0,s^2\big)$ distribution and is independent of (x,y). This allows us to write the sample log likelihood function related to the wage data as:

$$logL_w(\Theta) = \sum_{x\in\mathcal{X},y\in\mathcal{Y}} \hat{\mu}_{xy}\left(-\frac{1}{2}log(2\pi s^2) - \frac{(\hat{w}_{xy}-w^\Theta_{xy})^2}{2s^2}\right)$$

Finally, the log likelihood of the sample is then:

$$logL(\Theta) = logL_m(\Theta) + logL_w(\Theta)$$

The estimation algorithm procedure relies on an outer loop where the log likelihood is evaluated given a guess Θ, and an inner loop where we solve for the matching equilibrium implied by these parameters. For this inner loop, we use a coordinate descent algorithm, which relies on the following marginal equations which have to be met at equilibrium:

$$n_x = \mu_{x0} + \sum_{y \in \mathcal{Y}} \mu_{xy}$$
$$m_y = \mu_{0y} + \sum_{x \in \mathcal{X}} \mu_{xy}$$

Substituting for the equilibrium expression for matched masses, we get:

$$n_x = \mu_{x0} + \sum_{y \in \mathcal{Y}} exp\left(\frac{\tilde{\Phi}(x,y) + \sigma_w log(\mu_{x0}) + \sigma_f log(\mu_{0y})}{\sigma} \right)$$
$$m_y = \mu_{0y} + \sum_{x \in \mathcal{X}} exp\left(\frac{\tilde{\Phi}(x,y) + \sigma_w log(\mu_{x0}) + \sigma_f log(\mu_{0y})}{\sigma} \right)$$

Galichon et al. (2019) and Galichon and Salanié (2022) show that this system of equations can be solved efficiently via coordinate descent, for any given parameter vector Θ which parametrizes the expected surplus $\tilde{\Phi}(x,y)$. Solving these equations yields $\{\mu^{\Theta}_{x0}, \mu^{\Theta}_{y0}\}$, from which we can readily recover $\{\mu^{\Theta}_{xy}, w^{\Theta}_{xy}\}$. As such, the estimation algorithm is the following:

- Guess a parameter vector Θ,
- Solve for the equilibrium using the coordinate descent algorithm, recover $\{\mu^{\Theta}_{x0}, \mu^{\Theta}_{y0}, \mu^{\Theta}_{xy}, w^{\Theta}_{xy}\}$,
- Evaluate the likelihood function $logL(\Theta)$,

The minimization is performed using a quasi-Newton method.

C Additional Figures and Tables

Domain	Module (Sub-domain)	Number of Elements
Worker Characteristics	Abilities	52
	Occupational Interests	6
	Work Values	6
	Work Styles	16
Worker Requirements	Basic & Cross-Functional Skills	35
	Knowledge	33
	Requirements (Education, Experience, etc.)	5
Occupational Requirements	Generalized Work Activities	41
	Work Context	57
Occupation-Specific Information	Tasks	*Varies by Occupation*
	Tools & Technology	*Varies by Occupation*

Table A1: List of O*NET Modules

Notes: The table describes the overall structure of the O*NET data, by listing the overall data categories, the data modules within each category, and then finally the number of elements (individual variables) for which data is collected within each module.

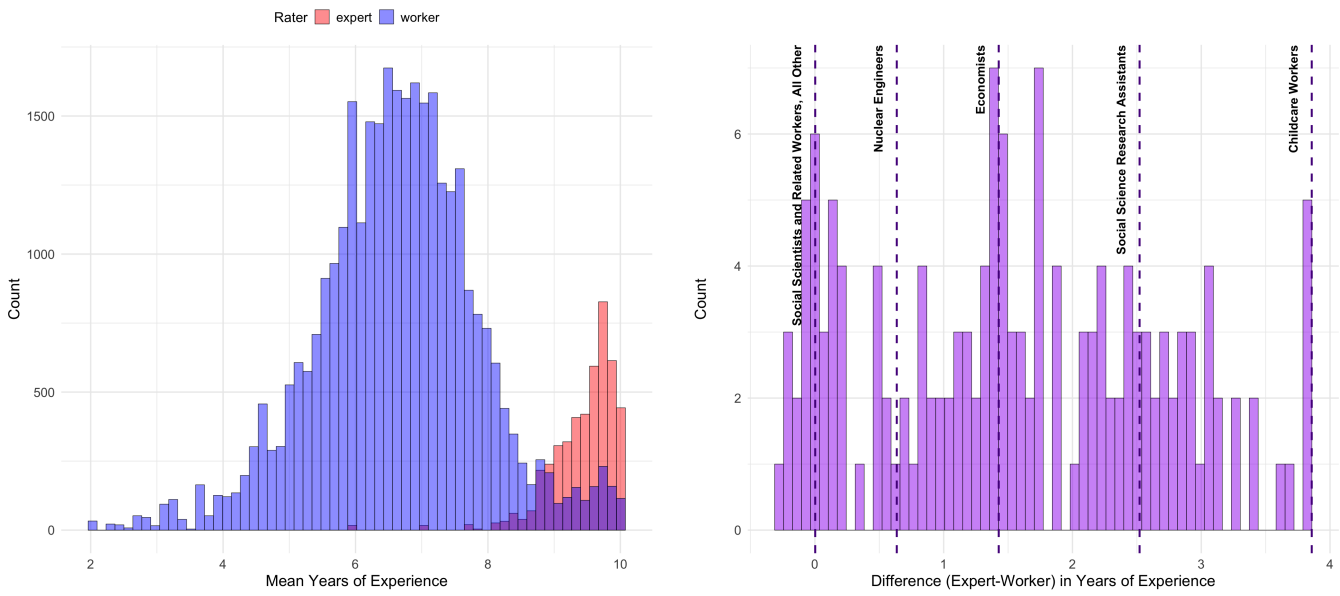


Figure A4: Distribution of Incumbent and Occupational Expert Experience

Notes: Each survey respondent (expert or incumbent worker) participating in O*NET data collection is asked about their years of experience in the occupation. According to O*NET, "Incumbents indicated in the background questionnaire how long they had been in their current job." On the other hand, "Occupational experts indicated in the background questionnaire how much experience they have performing work in this occupation." Valid responses to both questions range from < 1 year to 10 years or more. We plot the distribution of the mean years of experience for each occupation, with occupational means taken separately by expert and worker respondents.

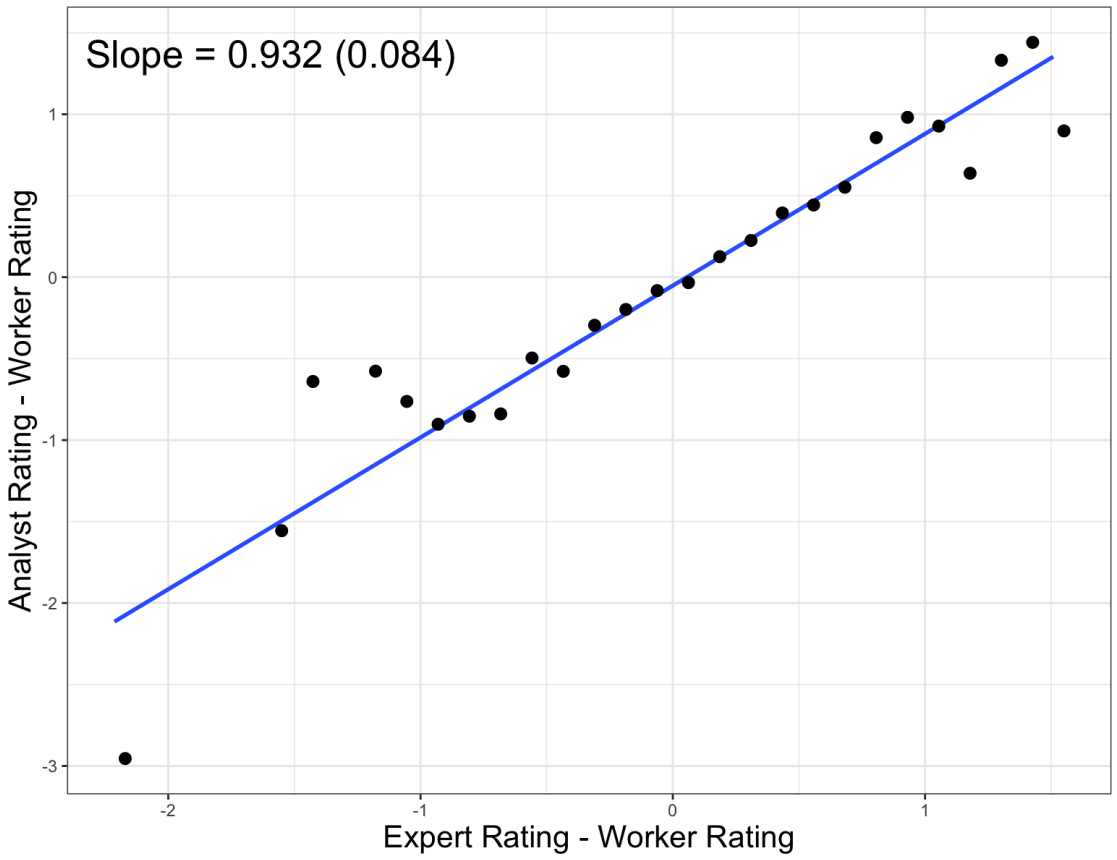


Figure A5: Relationship between Worker, Expert, and Analyst Ratings

Notes: We consider the set of occupation o-element e pairs that have been rated at least once by an expert, worker, and analyst; this includes 25 unique occupations. We calculate the average difference in the expert-worker rating, and the average difference in the analyst-worker rating, and run the following regression: $(\text{analyst} - \text{worker})_{oe} = \beta(\text{expert} - \text{worker})_{oe} + \varepsilon_{oe}$. We plot the differences in a binscatter, and we report $\beta = 0.932$ (0.084) with standard errors clustered at the occupation level. The coefficient is robust to the inclusion of occupation and element fixed effects.

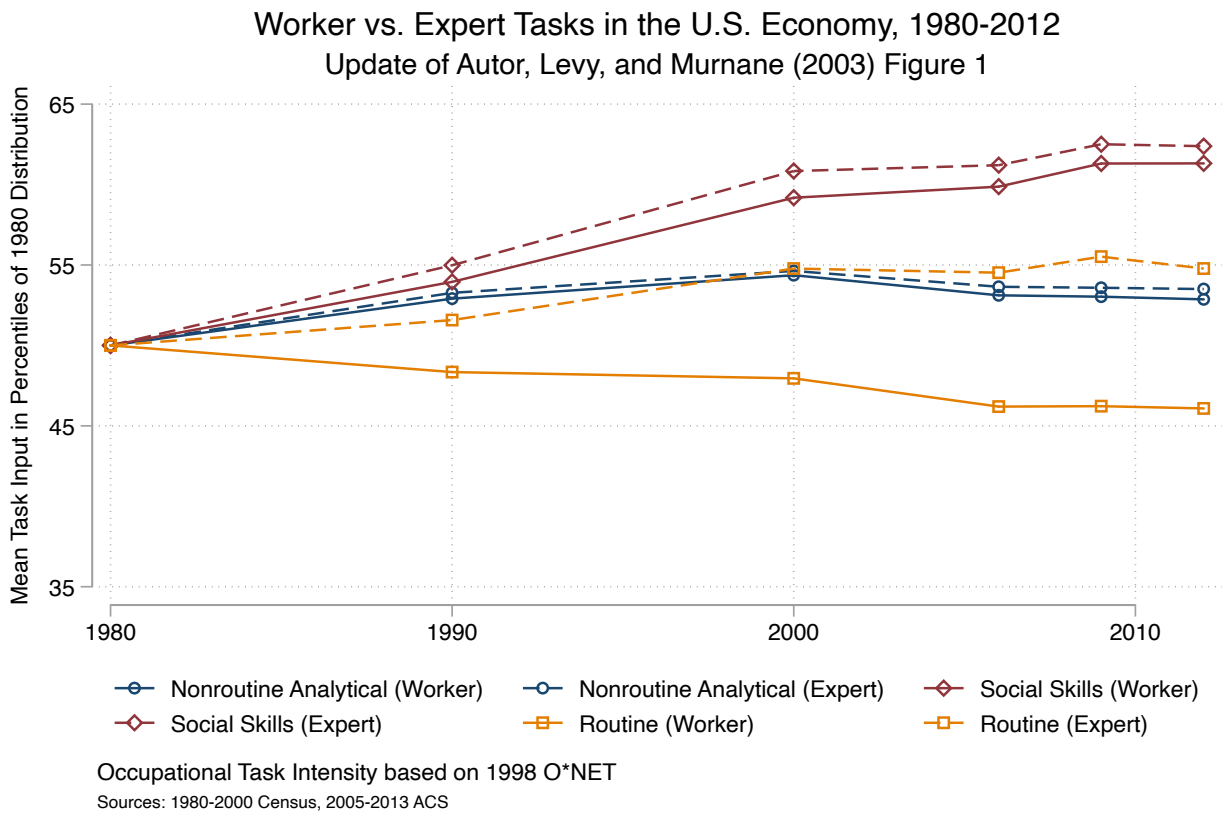


Figure A6: Replication of Deming (2017): Evolution of Mean Task Input

Notes: Replicating Figure 3 of *The Growing Importance of Social Skills in the Labor Market*, David Deming, Quarterly Journal of Economics, 2017, using O*NET data coming strictly from incumbent workers, analysts or occupational experts. O*NET task measures as defined by Deming (2017) are paired with data from IPUMS 1980–2000 censuses and the 2005–2012 American Community Survey samples. The data is aggregated to industry-education-sex cells by year, and a cell is assigned to its rank in the 1980 distribution of the task input (this rank is held constant across time). The figure then tracks the evolution of the relevance of each task input over time, as the mean percentile of that task in a given year, weighted by its employment (by construction this value is 50 in 1980). An increase in the percentile value of a task implies that cells with a high task input have an increase in their employment over time.

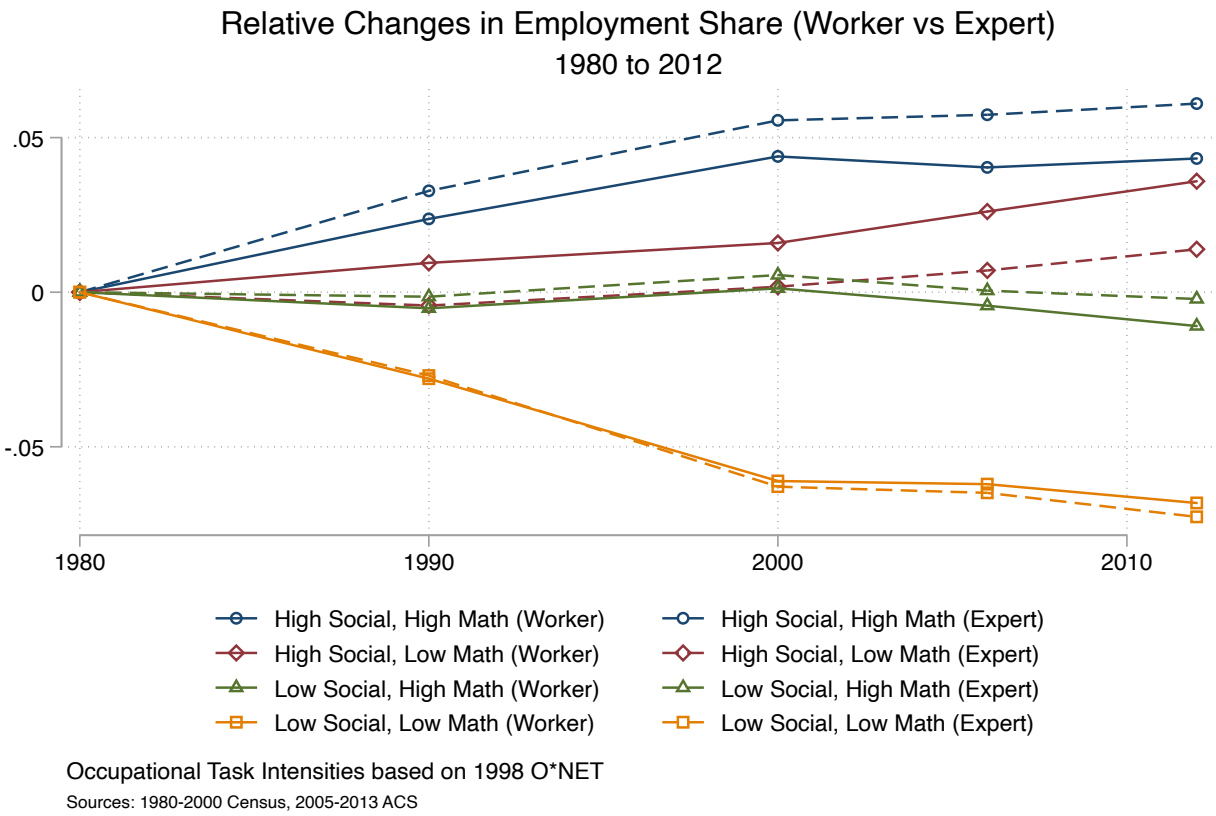


Figure A7: Replication of Deming (2017): Evolution of Employment Shares

Notes: Replicating Figure 4 of *The Growing Importance of Social Skills in the Labor Market*, David Deming, Quarterly Journal of Economics, 2017, using O*NET data coming strictly from incumbent workers, analysts or occupational experts. Here, each line plots 100 times the change in employment share (relative to 1980), for occupations that are above/below the 50th percentile in nonroutine analytical and social skill intensity, as measured in O*NET.

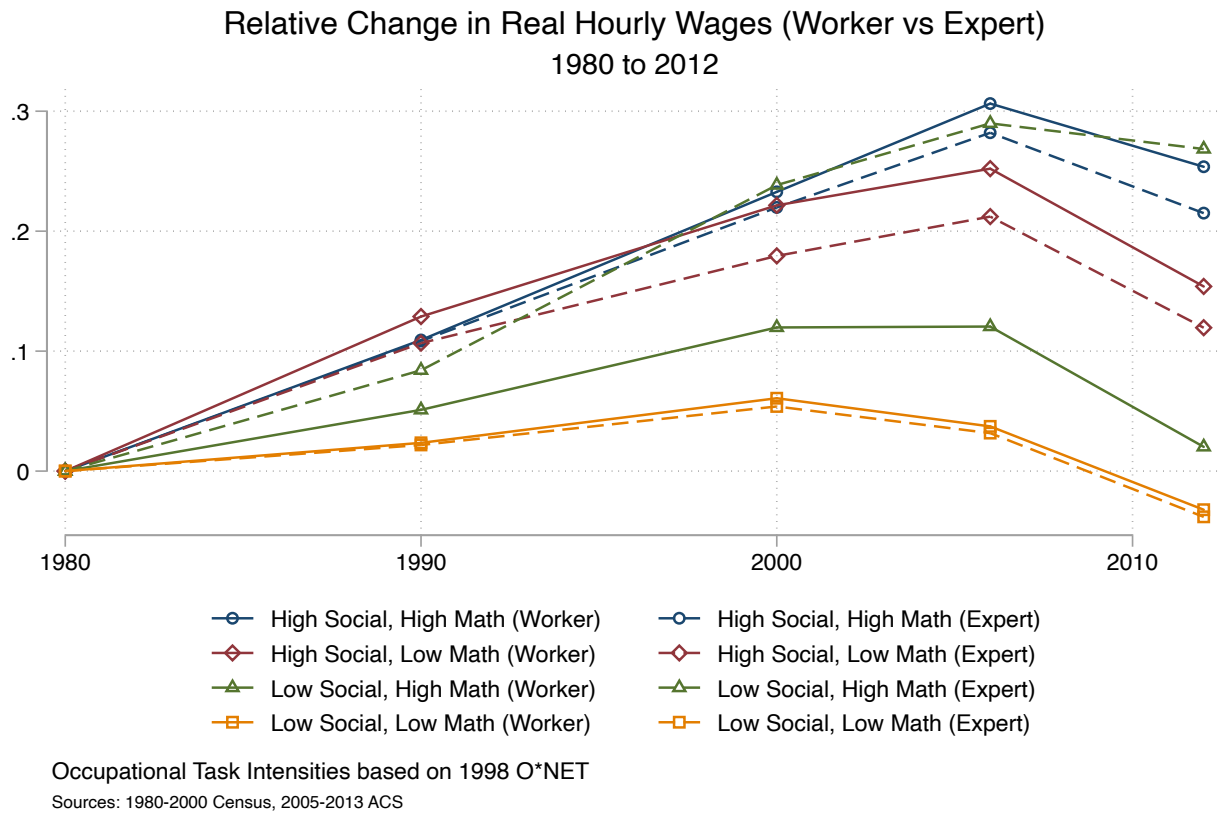


Figure A8: Replication of Deming (2017): Evolution of Real Hourly Wages

Notes: Replicating Figure 5 of *The Growing Importance of Social Skills in the Labor Market*, David Deming, Quarterly Journal of Economics, 2017, using O*NET data coming strictly from incumbent workers, analysts or occupational experts. Each line plots the percent change in wage (relative to 1980) for occupations that are above/below the 50th percentile in nonroutine analytical and social skill task intensity as measured in O*NET.

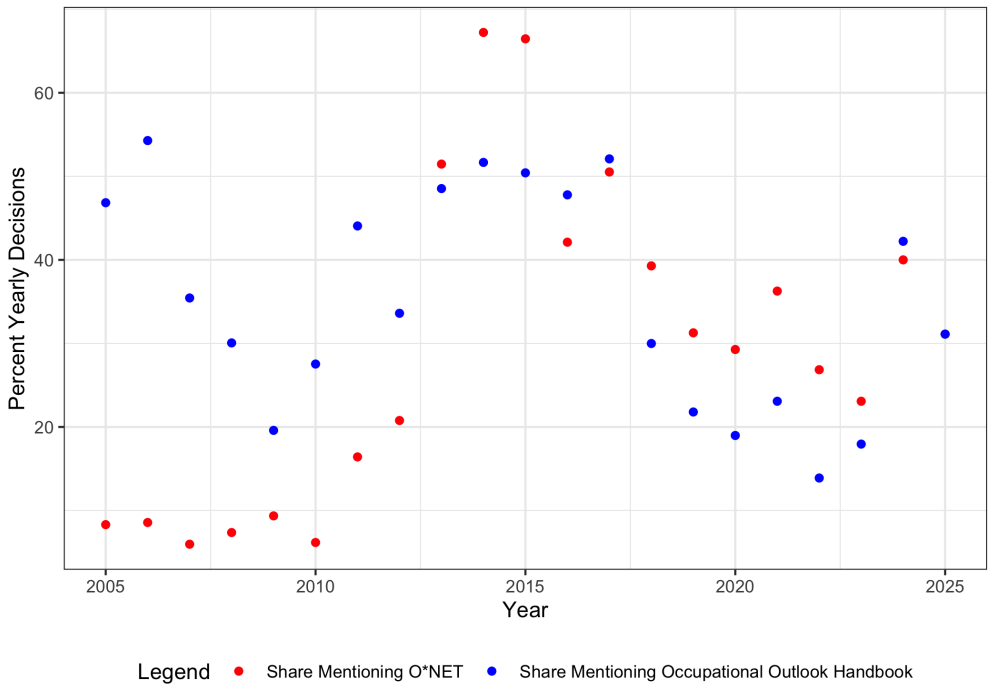


Figure A9: Percent of U.S. Citizenship and Immigration Service H1-B Visa Appeal Decisions mentioning O*NET

Notes: Fraction of USCIS non-precedent appeal decisions in a year making a direction mention of O*NET. For this figure, we download the universe of USCIS Non-Precedent decisions from 2005 to 2025 from the USCIS website, obtaining about 16k decisions in total. We then process the files, and check whether decisions make an explicit mention of O*NET. We also track whether decisions mention the Occupational Outlook Handbook from the BLS, which is the alternative source of occupation data used in the H-1B visa allocation process.

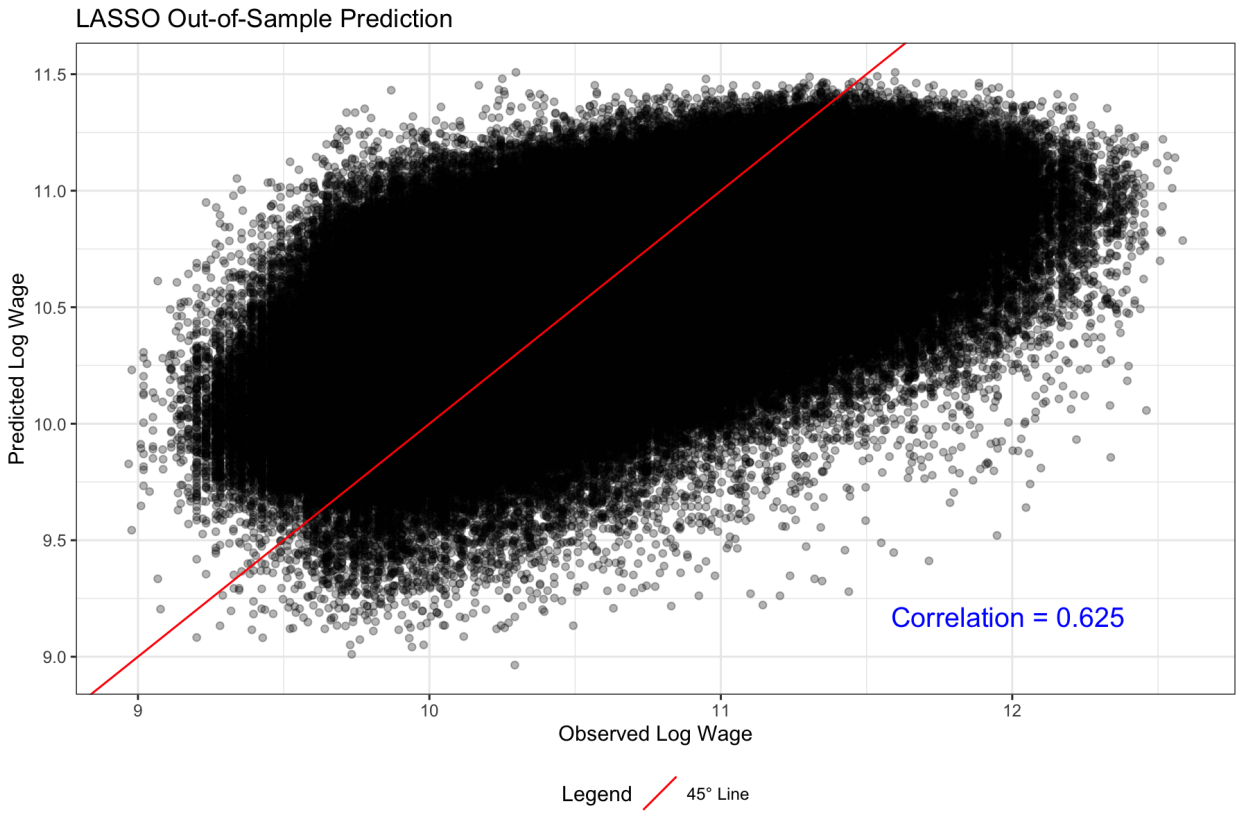


Figure A10: LASSO for Tacit Measure: Out-of-Sample Wage Prediction

Notes: In this figure we show the correlation between observed ACS wages, and wages predicted using the high-dimensional LASSO from which we recover the coefficients used to compute our occupation-level measure of tacitness used for Table 1.

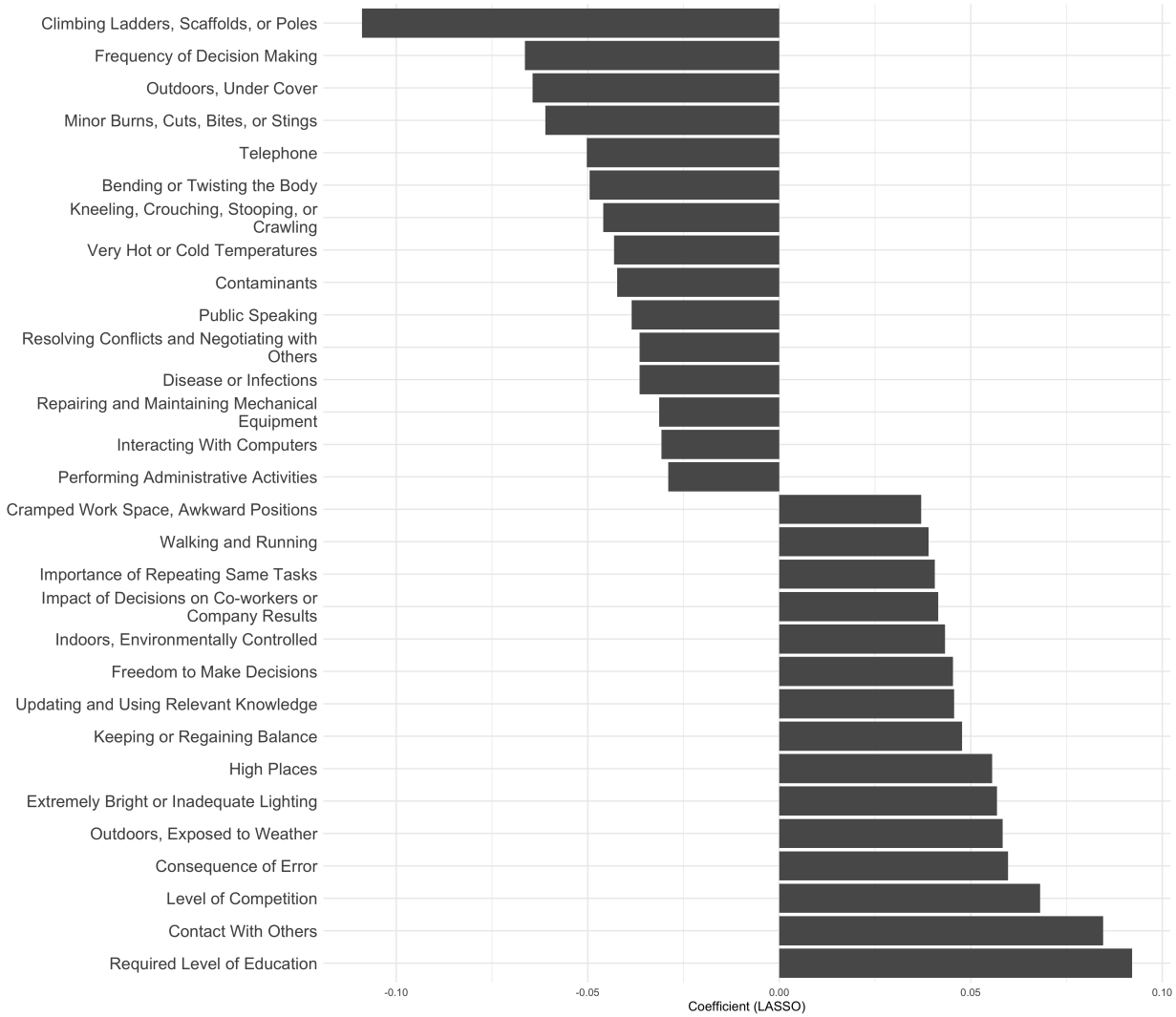


Figure A11: Highest and Lowest Wage Coefficients in High-Dimensional LASSO

Notes: In this figure we show the O*NET elements with the largest coefficients in the high-dimensional LASSO in the ACS, including Mincerian controls, from which we recover the coefficients used to compute our occupation-level measure of tacitness for Table 1.

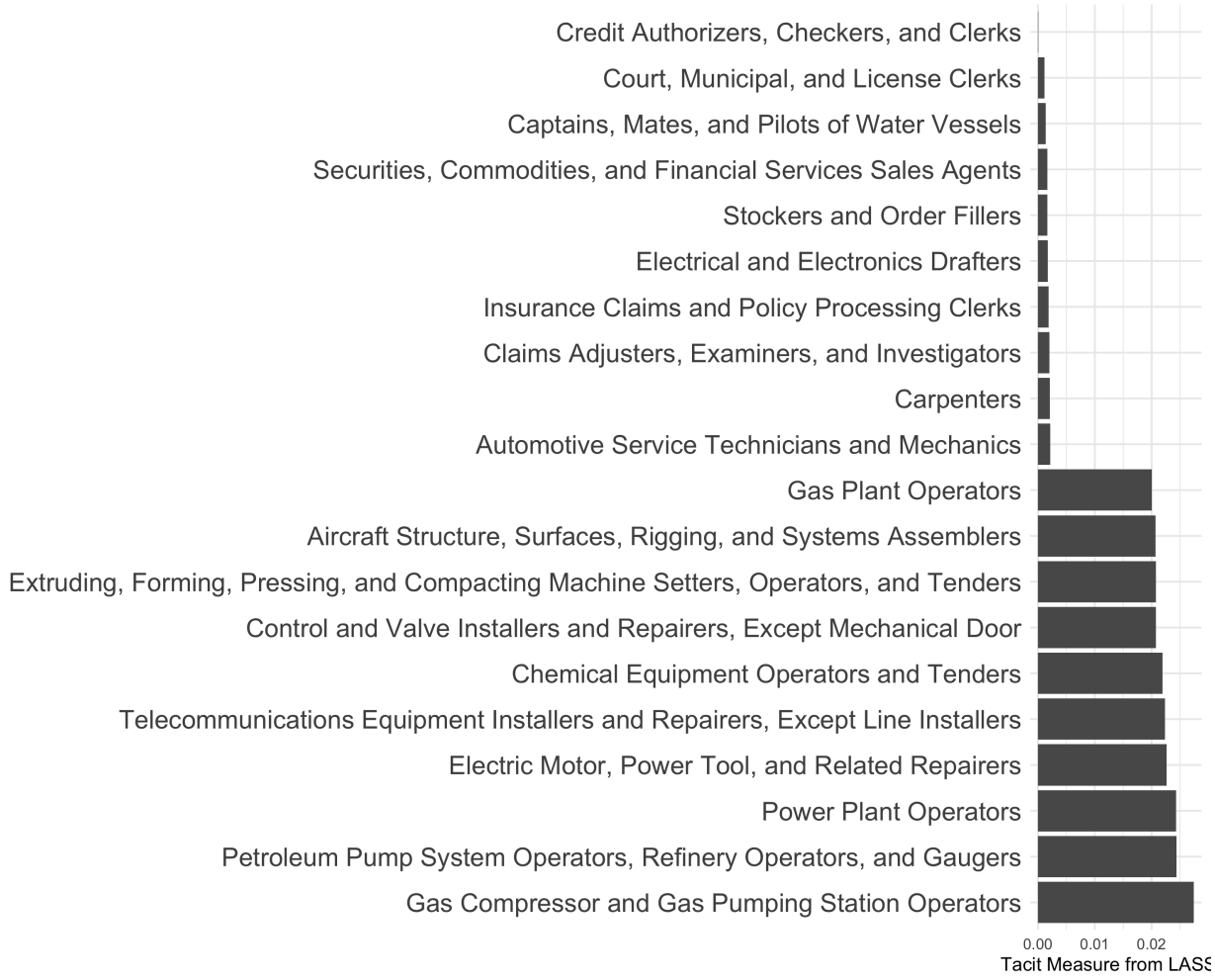


Figure A12: Highest and Lowest Tacit Occupations using High-Dimensional LASSO

Notes: In this figure we show the occupations that are most tacit using the tacitness index developed using the high-dimensional LASSO discussed in Table 1.

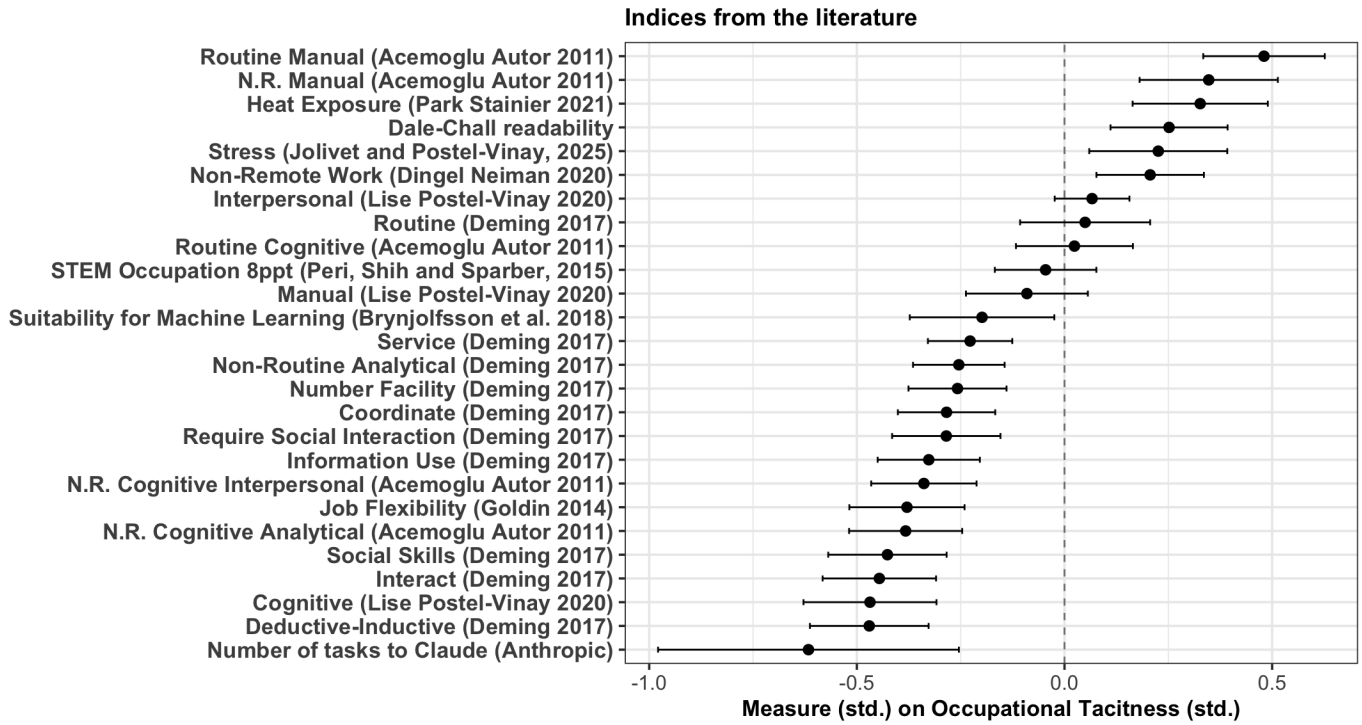


Figure A13: Correlations with Tacitness: Indices from the Literature

Notes: Bivariate regressions of the standardized LASSO-based tacit measure on standardized occupation-level indices from the literature.

	$\sigma(response)$: Within O*NET	$\sigma(response)$: Within AWCS
Intercept	0.978***	0.975***
	(0.021)	(0.020)
tacit	-0.027**	-0.015
	(0.008)	(0.018)
n	0.000**	0.000
	(0.000)	(0.000)
Fit statistics		
Observations	196	91
R^2	0.107	0.012

Heteroskedasticity-robust standard errors.
*+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001*

Table A2: Correlations with Tacitness: Noise

Notes: Tacit is the standardized LASSO-based measure. n represents the number of survey observations within each occupation cell. The categories of ALP AWCS questions used are jobatt_1, jobatt_2, jobatt_3, pacedepends, supatt, meanwork, jobphys, taskautonomy, team_tasks, exposure, physical conditions, jobskill, interrupt.

	AWCS Question Response	
female	-0.026	(0.035)
nonwhite	-0.022	(0.034)
tenure	-0.002	(0.002)
age	-0.002	(0.001)
college	-0.032	(0.035)
female × tacit	0.006	(0.034)
nonwhite × tacit	0.085*	(0.043)
tenure × tacit	0.002	(0.002)
age × tacit	-0.001	(0.001)
college × tacit	0.011	(0.032)
Fit statistics		
Observations	39,073	
Adjusted R^2	0.218	
Fixed effects		
Occupation × AWCS Question	Yes	
Survey Year	Yes	

Standard errors clustered by respondent and occupation.
*+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001*

Table A3: Correlations with Tacitness: Within-Occupation Demographic Differences

Notes: Tacit is the standardized LASSO-based measure. Tenure measures the number of years spent at the current job.

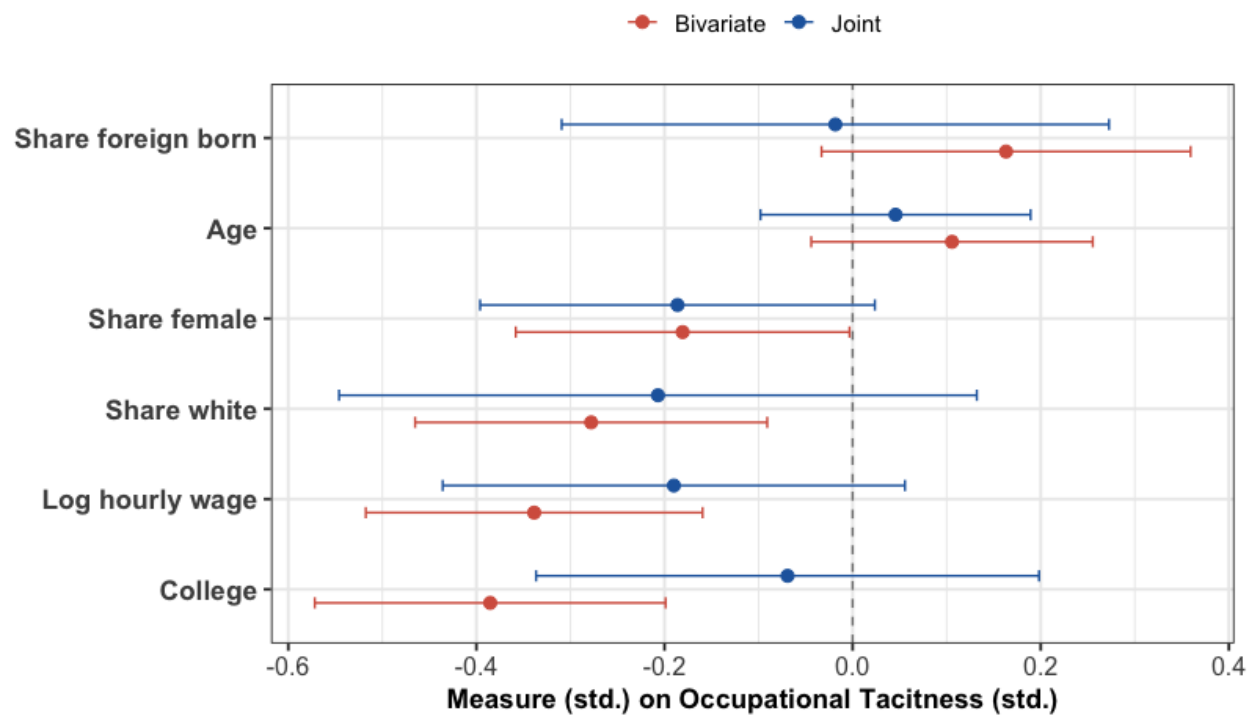


Figure A14: Correlations with Tacitness: Demographics

Notes: This figure plots the correlations between standardized occupation demographics (weighted aggregates to the 2018 SOC level from the ACS 2011-2019) against our standardized LASSO-based measure of occupational tacitness. We make the restrictions to the ACS data described in Section 3.

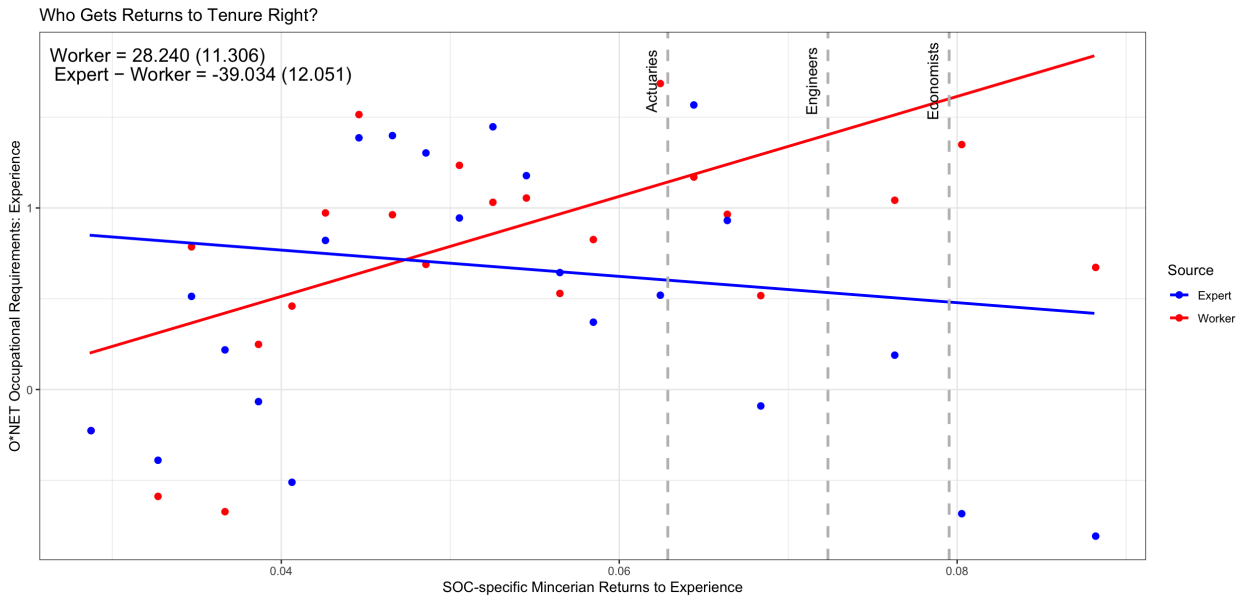


Figure A15: Who is Right: Tenure

Notes: Returns to tenure on the occupation level are estimated from the ACS (education, demographics) and LinkedIn data (occupational tenure) using the following Mincerian regression:

$$\log(w_{oy}) = male_{oy} + white_{oy} + education_{oy} + \sum_{o} \beta_o^{tenure}\, tenure_{oy} \times \delta_o + \varepsilon_{oy},$$

where demographics are taken as yearly shares, education is in years, and tenure is in quarter years. Tenure in the LinkedIn data is constructed from 2007 to 2025 by calculating the total number of days an individual spent in each occupation according to their profile, and then averaging across profiles on the occupation level. On the binscatter, we visualize occupation-time level O*NET experience requirements against β_o^{tenure}; we run this separately on the subset of the O*NET data rated by workers and experts. We report regression coefficients α and γ from:

$$\text{O*NET Experience}_{ot} = \alpha\beta_o^{tenure} + \beta\text{expert}_{ot} + \gamma\beta_o^{tenure} \times \text{expert}_{ot} + \delta_t + \varepsilon_{ot},$$

where standard errors are clustered on the occupation level. t represents database time.

Variables	Firm Premium/Penalty (γ)	Worker Premium/Penalty (α)
	Occupational Characteristics	
OSHA Injuries	3.57 (1.44)	-5.71 (1.21)
O*NET Injuries (expert, worker)	-0.10 (1.35)	-1.68 (0.37)
Returns to Education	2.60 (0.56)	-0.38 (0.42)
O*NET Education (expert, worker)	0.15 (0.08)	0.11 (0.08)
Occupational Tenure	2.18 (0.69)	1.58 (0.58)
O*NET Experience (expert, worker)	1.18 (0.23)	-2.85 (0.25)
Capital/Labor Ratio	0.33 (0.61)	-1.39 (0.47)
	Worker Characteristics	
White	0.85 (0.51)	-1.31 (0.49)
Sex	-2.92 (0.63)	-3.56 (0.67)
Age	0.21 (0.53)	-6.84 (0.68)
College	1.39 (0.22)	-4.43 (0.49)
	Scale Parameters	
Scale for Worker Taste Shocks (σ_w)		3.01 (0.56)
Scale for Firm Taste Shocks (σ_f)		3.51 (0.35)
Std. deviation wage error		4.89 (0.20)
R^2		0.39
N		2,304
Unique Occupations		72
Unique Demographic Combinations		32
Average Cell Size (observation count)		11,690
Observed Within Cell Std. Deviation in Wages		7.74

Table A4: Estimated Average Effects in the Hedonic Model

Notes: Estimated parameters from the hedonic structural model described in Section 8. When showing average effects, we pair them with the standard errors of these average effects, accounting for the observed standard deviations in the averages used for their computation.