NBER CONFERENCE - AI AND ECONOMIC MEASUREMENT

Measuring the Intensive Margin of Work: Task Shares and Concentration

Pierre Bouquet[a], Yossi Sheffi[a]

[a]Massachusetts Institute of Technology (MIT), Center for Transportation & Logistics, 1 Amherst Street, Cambridge, MA, USA;

ABSTRACT

Are jobs diffuse bundles of activities, or are they concentrated on a small core? This paper develops a statistical framework to measure the intensive margin of work using task-frequency survey responses from O*NET. Using categorical occurrence data for 12,495 tasks in O*NET 29.0, we construct within-occupation estimators of task flows and task shares that sum to one. Under a within-occupation homogeneous task-duration assumption, these shares can be interpreted as time allocations across tasks. We propagate sampling uncertainty from the underlying survey responses to obtain fully specified estimators of task flows and task shares, and provide diagnostic evidence on the robustness of the resulting measures. The framework yields interpretable, budget-share-like task weights that aggregate transparently into standard economic outcomes, including exposure indices, and workforce or wage-bill decompositions.

Empirically, we document a pronounced core-periphery structure of work: on average, the top three tasks account for about 31% of implied labor input across occupations. Linking task shares to U.S. microdata, we show that a small fraction of tasks accounts for a disproportionate share of both employment and wage bill. We then quantify how AI exposure aggregates depend on the accounting margin: tasks exposed to LLM-powered software account for about 52% of the wage bill but only 42% of employment. As external validation, we show that moments constructed from task-share weights yield economically interpretable wage gradients: occupational specialization predicts higher wages. Conditional on work content, this association attenuates with routine intensity. Standard intensive-margin proxies reproduce neither the specialization gradient nor its attenuation.

Pierre Bouquet (pierre.bouquet.ch@gmail.com)

1. Introduction

AI reshapes work by altering the productivity of particular tasks, not by directly shocking occupations. Yet the empirical literature typically measures AI or automation exposure at the occupational level, often treating tasks as binary attributes or relying on coarse occupational averages of task exposure. This conflates jobs that perform the same exposed tasks but devote very different shares of work to them. This results from the intensive margin of work, or how labor input is allocated across tasks within occupations, being weakly measured. As a result, exposure metrics are often difficult to interpret economically.

This paper's core contribution is an interpretable intensive-margin measurement object: an occupational task share constructed from observed task frequency. Rather than producing a unitless intensity score, we recover, for each occupation i and task k, a share $\hat{\pi}_{i,k}$ satisfying $\sum_k \hat{\pi}_{i,k} = 1$. We interpret $\hat{\pi}_{i,k}$ as a proxy implied by observed task occurrences of the share of labor input allocated to task k within occupation i. Under a homogeneous within-occupation task-duration assumption, $\hat{\pi}_{i,k}$ coincides with the task time share. Because $\hat{\pi}_{i,k}$ behaves like a budget share, it maps directly into standard economic aggregates. In particular, any task-level index or exposure score a_k can be aggregated into an occupation-level intensive-margin index $A_i = \sum_k \hat{\pi}_{i,k} \cdot a_k$.

The proposed measurement framework covers more than 11,500 O*NET tasks. O*NET surveys use a skip pattern: they first ask a worker whether they perform a task and, conditional on being performed, measure annual frequency on an ordinal scale. We leverage this structure to construct three linked measurement objects. First, we recover the unconditional distribution of task-frequency categories for each occupation–task pair. Second, we map ordinal frequency categories into annualized counts using a transparent calibration, producing an annualized task "labor flow" measure. Third, we normalize these flows within occupations to obtain proxies of task shares of labor input $\hat{\pi}_{i,k}$, which can be merged with employment, earnings, and automation-exposure scores to produce economically interpretable aggregates.

The paper also provides a disciplined uncertainty framework to enable inference on aggregation and nonlinear functionals (e.g. concentration measures or exposure indices). We propagate survey uncertainty to report point estimates and estimated variances for each measurement object through a covariance completion step. The completion is statistically parsimonious, selecting a minimum-complexity covariance matrix consistent with observed marginal variances. By construction, it avoids imposing spurious correlations not warranted by the available information. Additionally, the completion criterion is strictly convex, and thus the recovered measurements are uniquely defined, ensuring full reproducibility. Finally, we assess conceptual sensitivity to alternative frequency calibrations and to plausible heterogeneity in task duration.

To connect our measurement layer to canonical constructs, we classify tasks under the routine-task taxonomy of Autor, Levy, and Murnane (2003) using the LLM-based protocol of Autor and Thompson (2025). Aggregating these task labels to the occupation level using our labor input proxies, we benchmark the resulting occupation scores against existing indices drawn from Acemoglu and Restrepo (2022). We also show how intensive-margin aggregation changes the composition of measured exposure across task types when tasks are weighted by employment, or labor costs.

The fourth part of the paper presents three stylized facts that illustrate why intensive-margin measurements matter for labor, automation and AI exposure modeling. First, occupations exhibit a pronounced core-periphery structure: the median occupation performs 19 tasks, yet the top three tasks account for about 31% of labor

input. Second, this concentration persists when weighting by employment: while the average task count rises (from 29 to 36), the top-three's share remains near 30%, implying that labor input concentrates in a small set of tasks. Third, intensive-margin weighting changes the composition and incidence of AI exposure across routine and non-routine work. Tasks directly exposed to conversational LLMs represent a relatively small share of aggregate activity (about 15% of labor input and cost) and are concentrated in routine cognitive tasks. In contrast, exposure through LLM-augmented software accounts for about 40% of the labor input and 51% of the total labor cost and is composed primarily of non-routine and interpersonal cognitive tasks. Finally, "not-exposed" work is disproportionately manual and non-routine interpersonal (about 45% of labor input but 34% of the total labor cost).

The final part of the paper provides measurement validation using wages as an external criterion. We document that task-share-based metrics of task allocation are systematically related to earnings: specialization (task-share concentration) is positively associated with wages (about 6% per standard deviation), routine intensity is negatively associated with wages (about 10% per standard deviation), and the specialization-wage gradient is larger in low-routine occupations and attenuates with routine intensity. We further show that commonly used alternative task weighting schemes do not reproduce these specialization patterns, indicating that existing intensive-margin proxies are not interchangeable with task-share weights.

Our work contributes to several strands of economic literature. Canonical models of Autor, Levy, and Murnane (2003); Acemoglu and Autor (2011) conceptualize occupations as bundles of tasks but largely operate on the extensive margin, whether a task is performed, without a principled way to weight tasks by intensity or labor input. A growing literature aggregates task information into occupation-level indices using different weighting schemes: uniform averages of task scores (Autor and Thompson 2025), weights based on O*NET task attributes that are informative about intensity (Webb 2019), and frequency-based proxies that incorporate task intensity (Martin and Monahan 2022; Tomlinson et al. 2025; Bouquet, Bagnoli, and Sheffi 2025). We contribute to this literature by providing an economically interpretable measurement object for the intensive margin. It aggregates directly into standard exposure and productivity statistics. As it is a set of explicit statistical estimators, it allows other researchers to construct exposure indices, concentration measures, and other derived statistics with uncertainty quantification, rather than relying on simple or coarse weighting schemes.

This measurement motivation is reinforced by the literature on AI and automation exposure, including Brynjolfsson and Mitchell (2017); Webb (2019) and Eloundou et al. (2023), who construct task-based exposure metrics under diverse aggregation schemes. Notably, Eloundou et al. (2023) highlights the intensive-margin measurement problem, reporting that exposure estimates can be sensitive to the choice of task weights and difficult to interpret economically. They ultimately adopt uniform or simple 2:1 ratio weights (based on core versus supplemental task classifications) and report exposure as the share of tasks exposed to large language models. Our empirical results document substantial dispersion in within-occupation task intensity and show that accounting for this heterogeneity yields AI exposure measures that are interpretable in economically meaningful units.

The rest of the paper is organized as follows. Section 2 introduces the framework and derives the key equations. Section 3 presents the data and robustness tests. Section 4 presents evidence on task concentration, task-type patterns in LLM exposure when aggregated by labor input and labor cost. Section 5 provides measurement validation using wages as an external criterion and benchmarks task share based moments against

existing intensive-margin metrics. Section 6 concludes.

2. Measurement Framework

This section develops a conceptual framework for estimating a proxy of the distribution of labor inputs across tasks within an occupation. Our measurement challenge is to transform ordinal, qualitative responses from job incumbents' surveys into cardinal estimates of task flow and labor input shares.

2.1. *Recovering the Latent Distribution of Task Frequency*

To characterize the intensive margin, we must first confront the specific structure of the O*NET survey instrument. The data-generating process follows a two-stage skip pattern designed to reduce respondent fatigue. For each task item, incumbents are first asked whether it is part of their job (the extensive margin). If a respondent answers negatively (i.e. that they do not perform the task), the task is explicitly recorded as "not relevant," and the inquiry ends for that item. Only respondents who pass this first stage are asked to report the frequency of execution of the task on a seven-point ordinal scale.

Hence, we introduce the mass point at zero ($r = 0$) to capture these "never perform" answers and construct the unconditional probability vector, $\hat{\mathbf{p}}_{i,k,r}$, the probability that a worker in occupation i selects frequency category r for task k, by scaling the conditional frequency distribution with the probability of a task occurring:

$$\hat{p}_{i,k,r} = \begin{cases} 1 - \hat{R}_{i,k} & \text{if } r = 0 \quad \text{(Extensive Margin: Never)}, \\ \hat{R}_{i,k} \cdot \hat{F}_{i,k,r} & \text{if } r \in \{1, \dots, 7\} \quad \text{(Intensive Margin: Frequency)}, \end{cases} \tag{1}$$

where $\hat{R}_{i,k}$ denotes the proportion of incumbents in occupation i reporting task k as relevant, and $\hat{F}_{i,k,r}$ is the conditional probability of selecting frequency category r. This transformation yields a complete probability mass function, $\sum_{r=0}^{7} \hat{p}_{i,k,r} = 1$, characterizing the latent distribution of task frequency across the entire population of incumbents.

For $r \geq 1$, $\hat{p}_{i,k,r}$ is constructed from two estimated objects, which O*NET reports with distinct sample sizes and standard errors. We treat the sampling errors in $\hat{R}_{i,k}$ and $\hat{F}_{i,k,r}$ as approximately uncorrelated and propagate sampling uncertainty using Goodman's product-variance approximation (Goodman 1960) for tractability. This approximation is necessary because the released aggregates do not identify the sampling covariance between the two estimators. This yields a fully specified estimator of the latent task-frequency distribution. The full derivation is provided in Appendix A.2.

2.2. *Annual Task Flows*

Economic measurement requires task inputs to be expressed on a cardinal scale. Because O*NET reports task frequency in ordered categories, we translate each category into an expected number of occurrences per worker-year. Following Martin and Monahan (2022); Tomlinson et al. (2025), we define an annualization vector $\mathbf{w}$ that assigns point estimates to each frequency category, from "Never" to "Hourly or more":

$$\mathbf{w} = \big[0,\ 1,\ 4,\ 24,\ 104,\ 260,\ 780,\ 2080\big]^\top. \tag{2}$$

This calibration anchors the survey categories to physical time constraints. For instance, "Hourly or more" is mapped to continuous labor input that saturates a full work year (2,080 hours), whereas "Daily" is mapped to one execution per workday (260 occurrences). This conversion inevitably discretizes an underlying continuous latent intensity; however, because the frequency scale spans three orders of magnitude and incumbent responses exhibit substantial dispersion, the resulting intensity measures are driven primarily by structural heterogeneity in the ordinal data rather than the specific local calibration of the weights. Robustness checks supporting this claim are reported in Section 3.3.1.

The expected annual flow (number of occurrences per year) for task k in occupation i, denoted $\hat{\mu}_{i,k} \in [0, 2080]$. It is the expectation of vector $\mathbf{w}$ with respect to the unconditional probability distribution derived in Equation (1):

$$\hat{\mu}_{i,k} = \mathbf{w}^\top \hat{\mathbf{p}}_{i,k} = \sum_{r=0}^{7} w_r \, \hat{p}_{i,k,r}. \tag{3}$$

The estimator $\hat{\mu}_{i,k}$ captures the expected annual flow of task k within occupation i. By converting probabilities into flows, we establish a bridge to decompose the indivisible occupational "job bundle" into divisible units of labor input.

Estimating the precision of $\hat{\mu}_{i,k}$ presents a challenge: the elements of $\hat{\mathbf{p}}_{i,k}$ are structurally dependent (summing to unity), but the survey design only identifies their marginal variances. To recover a valid covariance structure, we adopt a geometric approach. We estimate the latent covariance matrix $\widehat{\Sigma}_{i,k}$ by solving a semidefinite program that matches the observed marginal variances while minimizing the squared Frobenius norm (Aitchison 1982).

Economically, this objective function acts as a parsimonious selection mechanism. Among all covariance matrices consistent with the observed marginal variances and the unit-sum constraint, minimizing the squared Frobenius norm recovers the latent covariance structure that imposes the minimum necessary dependence between categories. Because the optimization is strictly convex over the feasible set, the solution is unique. This yields a fully specified estimator of annual task flows, enabling valid statistical inference. Full details of the optimization problem are provided in Appendix A.3.

2.3. *Task Share of Labor Input*

We construct a normalized measure of task flow, $\hat{\pi}_{i,k}$, defined as a proxy for the fraction of total labor input in occupation i allocated to task k. In theory, the intensive margin of labor input is a function of two primitives: the frequency of execution and the duration of each occurrence. Let $\mu_{i,k}$ denote the expected annual occurrences of task k, and $d_{i,k}$ denote the average duration (in hours) per occurrence. The true task time share is given by:

$$\pi_{i,k} \;=\; \frac{\mu_{i,k} \cdot d_{i,k}}{M_i}, \qquad M_i = \sum_{j \in K_i} \mu_{i,j} \cdot d_{i,j}. \tag{4}$$

Where M_i is the total annual number of hours allocated across tasks in occupation i, and K_i denotes the set of tasks in occupation i.

Ideally, identification would rely on simultaneous observation of both $\mu_{i,k}$ and $d_{i,k}$. However, while the O*NET protocol yields granular data on frequency, it contains no information on task duration. To operationalize Equation (4), we adopt a parsimonious approximation by setting $d_{i,k} = 1$ for all k (Martin and Monahan 2022). Under this homogeneity assumption, variation in task shares is identified exclusively through the frequency margin:

$$\hat{\pi}_{i,k} \;=\; \frac{\hat{\mu}_{i,k}}{\sum_{j \in K_i} \hat{\mu}_{i,j}}. \tag{5}$$

We interpret $\hat{\pi}_{i,k}$ as a frequency-based proxy of the labor input allocation to task k in occupation i. While this abstraction may under-weight low-frequency, high-duration "deep work," it leverages the substantial cross-task dispersion in frequencies (spanning three orders of magnitude in the annualization vector $\mathbf{w}$) as the primary empirical signal of task intensity. We assess the sensitivity of this approximation to duration heterogeneity by conducting robustness checks under alternative duration profiles in Section 3.3.2.

Finally, we treat $\hat{\pi}_{i,k}$ as an estimated object rather than a deterministic occupational characteristic. Treating $\hat{\mu}_{i,k}$, as an asymptotically linear estimator, we propagate measurement uncertainty to $\hat{\pi}_{i,k}$ via the delta method (van der Vaart 2000). This explicitly accounts for the ratio structure in Equation (5), in which the task-specific flow $\hat{\mu}_{i,k}$ enters both the numerator and the denominator. The result is a fully specified estimator of task composition defined by a point estimate and an associated asymptotic variance. Full derivations are provided in Appendix A.4.

For each occupation i and task k, we obtain point estimates and variances for (i) annualized task flow and (ii) within-occupation task shares (labor-input share proxies). Appendix A.5 summarizes modeling assumptions and choices. Appendix A.6 provides an illustrative table of estimated task flows and implied labor shares for an occupation.

3. Data and Measurement Robustness

In this section, we describe our data sources and report robustness checks of the measurement method.

3.1. *Main Data Sources*

Worker Data. To characterize labor supply and wages, we use microdata from the American Community Survey (ACS) of 2024, provided by IPUMS USA (Ruggles et al. 2025). The ACS is an ongoing national survey conducted by the U.S. Census Bureau which samples approximately 3.5 million households annually. We restrict our sample to the working-age civilian population (ages 16–64). We exclude military personnel and unpaid family workers, and further limit the sample to employed workers with reported positive hours and non-zero annual wages. We derive FTE measures and provide the calculations and descriptive statistics in Appendix B.1.

Task Data. We leverage the major release of the O*NET database in 2024 (version 29.0) for this study. O*NET is maintained by the U.S. Department of Labor's Employment and Training Administration and is one of the primary sources of standardized occupational information in the United States, (National Center for O*NET Development 2025). The database organizes occupational information hierarchically: approximately 55,000 job titles are consolidated into around 900 detailed occupational profiles, which are further decomposed into approximately 18,000 unique task statements describing specific work activities. O*NET collects frequency and relevance ratings for a subset of about 12,500 tasks. Descriptive statistics are provided in Appendix B.2.

LLM Task Classification. We classify each O*NET task statement into six mutually exclusive categories, distinguishing between routine and non-routine tasks and further sub-classifying them as manual, cognitive, or interpersonal using a large language model (LLM) as a synthetic task coder. Our taxonomy follows the canonical routine/non-routine framework in Autor, Levy, and Murnane (2003); Acemoglu and Autor (2011). The LLM is applied directly to task text to produce task-level labels using a minimally adapted prompt based on Autor and Thompson (2025). We then aggregate these labels to the occupation level using task shares $\pi_{i,k}$, yielding class-specific task-share measures that sum to one within each occupation.

We validate the LLM-based classifications by comparing the resulting occupation-level class labor input shares to occupation-level work content measures drawn from Acemoglu and Restrepo (2022). Agreement is high: across the matched set of occupations, the average employment-weighted Pearson correlation between our LLM-based measures and the benchmark is $R_W = 0.61$ (unweighted $R_U = 0.56$). The stronger employment-weighted correlations indicate that alignment is greatest in economically central (high-employment) occupations.

Figure 1 reports the class-by-class alignment with the benchmark (weighted R, rank correlation, and the share of zeros), while Appendix C provides the full prompt, model specification, robustness checks, and detailed discussions.

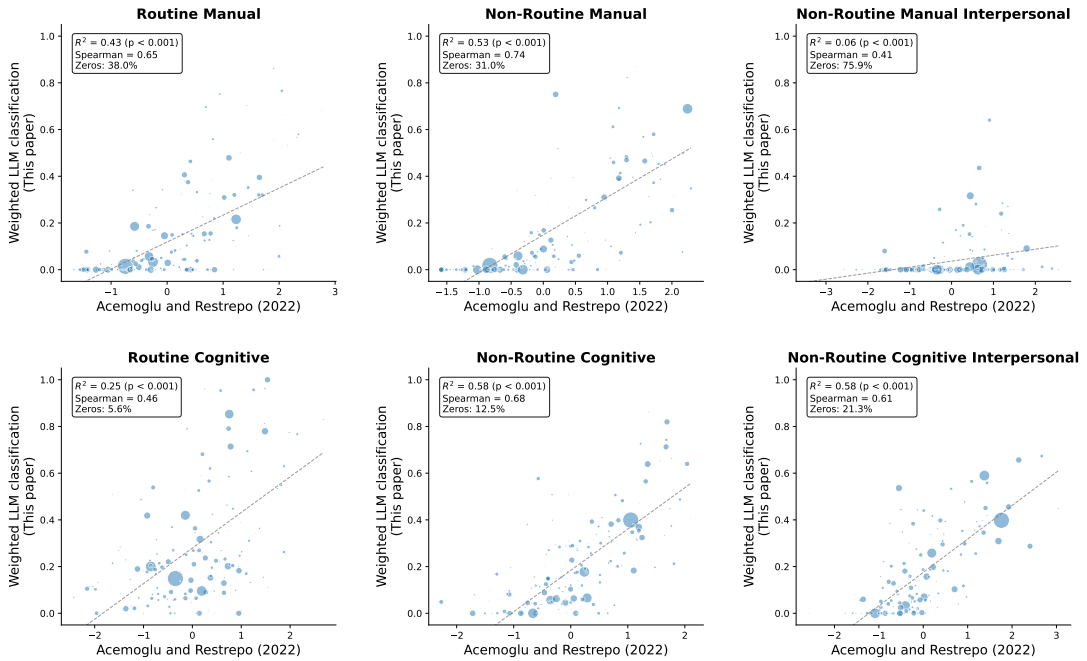


Figure 1. Comparison of labor input weighted LLM task classifications with occupation-level work content drawn from Acemoglu and Restrepo (2022) ($N = 216$). *Notes:* Each panel corresponds to an occupation feature class. Each dot represents an occupation in the Autor and Dorn (2013) taxonomy; marker sizes are adjusted to their employment share. The horizontal axis reports the work content measure, while the vertical axis reports the corresponding task-share–weighted LLM classification. The dashed line shows the weighted least-squares (WLS) fit. Each panel reports the weighted R^2 from the WLS, the unweighted Spearman rank correlation, and the share of occupations with zero labor-input in the work content class.

3.2. *Numerical and Mathematical Robustness of the Estimation Procedure*

Based on the estimators of 12,495 tasks with enough incumbent responses for statistical significance, we perform a series of mathematical diagnostic tests to ensure the stability and validity of the estimated covariance matrices.

First, we verify that all estimated matrices represent mathematically valid covariance structures that satisfy the probability constraints (i.e. that task probabilities sum to one), the mean residuals achieve near machine precision ($\approx 4\times10^{-11}$). This confirms the estimators are valid statistical estimators. Analyzing the covariance structures reveals an average effective rank of 3.7 (1.02). This indicates that, while the O*NET instrument provides eight frequency categories, the latent uncertainty in incumbent responses is concentrated within a three- to four-dimensional subspace on average. This supports that survey responses are internally consistent, and their answers restrict to a few of the frequency categories.

Second, we test for extreme solutions. For each task, we compare the estimated variance to its theoretical maximum and minimum possible values given the observed marginals. We identify 670 tasks (5.36%) in which the estimates lie at the extreme boundaries of the feasible set (normalized scores below 0.05 or above 0.95), suggesting degenerate distributions that may be driven by sparse data.

Third, we evaluate robustness to measurement error. We perturb the input marginal variances by ±5% and re-estimate the covariance matrices. We find that the estimator is very stable to measurement errors and only 48 tasks (0.4%) exhibit statistically significant variance instability under these perturbations.

Fourth, we impose a minimum complexity constraint by excluding occupations with fewer than three valid tasks. This threshold is necessary to filter out occupations for which task information is not sufficiently available. This results in the removal of 2 occupations and 4 tasks (0.03%).

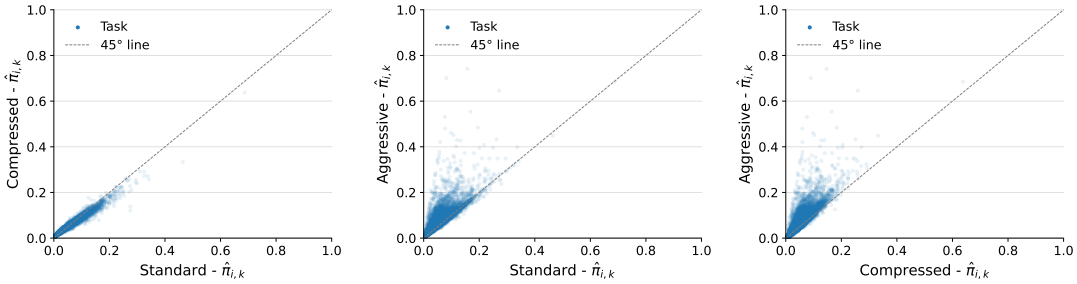
In summary, these quality control procedures result in the exclusion of 720 unique tasks (5.8%). This process removes numerically unstable estimates and sparse job profiles, leaving a robust analytical sample of 11,775 tasks covering 661 occupations. Detailed diagnostic formulas, results, and discussions are provided in Appendix D.2.

3.3. *Assumption Robustness and Economic Interpretation*

3.3.1. Sensitivity to Annualization Weights

The validity of our intensive margin measure depends, in part, on the specification of vector $\mathbf{w}$, which maps ordinal frequency categories to cardinal task shares. To ensure the estimated magnitudes of labor input allocation are not artifacts of our baseline calibration, we test the sensitivity of $\hat{\pi}_{i,k}$ against two alternative weighting regimes: an "Aggressive vector" ($\mathbf{w}_{\text{agg}}$) assuming a continuous 365-day workflow, and a "Compressed vector" ($\mathbf{w}_{\text{comp}}$). The compressed specification is not a simple linear rescaling; it applies a distinct weighting scheme that fundamentally alters the relative importance of tasks by reducing the ratio between the highest and lowest frequency weights from 2080:1 to 100:1. This allows us to verify if our results are robust to a framework where high-frequency tasks are penalized less heavily relative to episodic, low-frequency activities. Detailed weighting schemes for each specification are provided in Appendix D.3.

We re-estimate the covariance completion model for all occupation-task pairs under these specifications, using the intersection of tasks that converged to valid covariance matrices across all three models ($N = 9,821$). Results are presented in Figure 2.



(a) Task intensity from standard vs compressed weights ($R^2 = 0.966$; $\rho = 0.987$)

(b) Task intensity from standard vs aggressive weights ($R^2 = 0.670$; $\rho = 0.907$)

(c) Task intensity from compressed vs aggressive weights ($R^2 = 0.688$; $\rho = 0.904$)

Figure 2. Robustness of Standardized Task Intensities ($N = 9,821$) to Weighting Specifications.

The task shares across specifications exhibit very high ordinal correlation, evidenced by Spearman coefficients ($\rho \geq 0.9$). This supports that the relative hierarchy of tasks

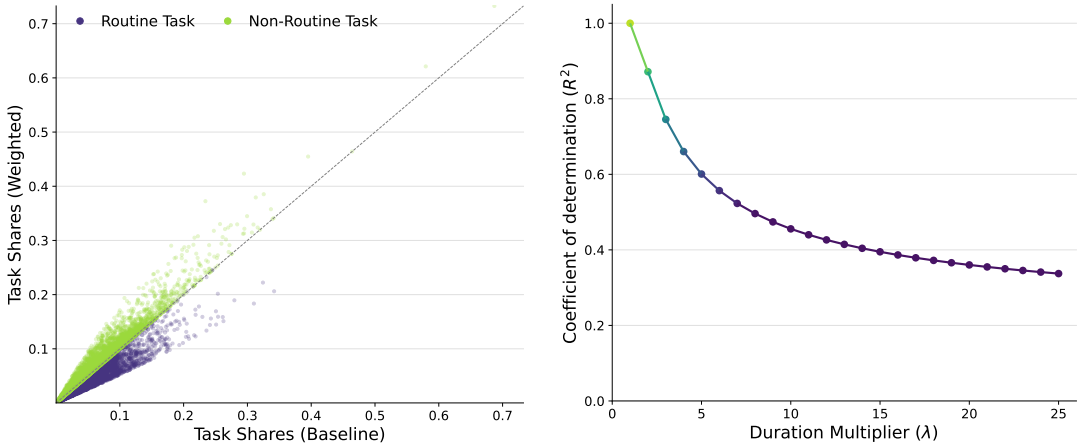
within the occupational portfolio remains stable regardless of the frequency vector's specific calibration.

The cardinal consistency is very high between standard and compressed weights ($R^2 = 0.966$), it moderates when introducing the aggressive specification ($R^2 = 0.670$). As shown in Figure 2 Panels B and C, the aggressive vector tends to inflate shares for middle-frequency tasks relative to the baseline. We interpret this divergence as a mechanical consequence of the non-linear weights assigned to monthly and weekly categories in $\mathbf{w}_{\text{agg}}$, which test the upper bounds of task flow.

3.3.2. Robustness to Duration Heterogeneity

A key challenge in measuring the intensive margin of work is the potential negative correlation between task frequency and duration. In particular, tasks performed less frequently may require more labor time per occurrence than highly repetitive, activities. Our baseline model assumes homogeneous task durations ($d_{i,k} = 1$), but standard economic reasoning implies that non-routine tasks, which require situational judgment and complex problem-solving, are likely to take longer per occurrence than routine tasks (Acemoglu and Autor 2011). If such heterogeneity were severe, a frequency-based estimator could mismeasure the true allocation of time shares.

To investigate the potential bias from unobserved duration heterogeneity, we conduct a "content-weighted" robustness exercise. We assign a duration multiplier $d_k = \lambda$ for all non-routine tasks as defined in Acemoglu and Autor (2011) and captured by our LLM classification, while routine tasks retain $d_k = 1$.



(a) Baseline vs. Content-Weighted ($\lambda = 2$)

(b) R^2 Across Duration Multipliers

Figure 3. Robustness of Task Shares to Duration Assumptions. Panel A compares the baseline task intensity ($\lambda = 1$) against a duration-weighted specification in which non-routine tasks are assigned double weight ($\lambda = 2$). Panel B plots the coefficient of determination (R^2) between the baseline and duration-weighted measures as the multiplier λ increases.

We begin by testing a specification with $\lambda = 2$, meaning that each non-routine occurrence is counted as twice the labor input of a routine task. Despite this adjustment, Figure 3 (Panel A) shows that the baseline frequency measure accounts for the majority of the variation in duration-weighted factor shares ($R^2 = 0.86$). Visually, the point distribution in Panel A reveals a mechanical reallocation: non-routine tasks shift above

the 45° line as their individual weights increase, while routine tasks cluster below the line. This downward shift for routine tasks reflects a mechanical effect: as non-routine activities consume a larger portion of the occupational labor volume (M_i), the relative share of routine tasks decreases mechanically to maintain the unit-sum constraint.

Panel B extends this analysis by varying λ. For plausible duration multipliers ($\lambda \in [1, 3]$), the measure remains robust ($R^2 > 0.71$). Even under the conservative assumption that non-routine tasks require five times the duration of routine tasks ($\lambda = 5$), frequency alone still explains nearly two-thirds of the cross-sectional variation in task shares ($R^2 = 0.56$).

This analysis supports the conclusion that while some duration heterogeneity exists, the primary driver of task shares is the heterogeneity in survey answers and task frequency. This result is numerically coherent: while it is unlikely for task durations to exceed a 1:10 ratio, the annual frequency weights in our model range from 1 to 2,080, making frequency the dominant signal. This evidence suggests that our continuous estimator delivers a strong and robust measure of task-level labor input.

4. The Intensive Margin of Work: Characteristics and Facts

In this section, we investigate the intensive structure of work characteristics and their translation into the U.S. economy. The covariance completion procedure produced $N = 11{,}775$ task share estimates $\hat{\pi}_{i,k}$ for 661 O*NET-SOC occupations. We map O*NET-SOC occupations into 2010 Census occupation codes to harmonize O*NET with IPUMS data. To preserve our statistical methodology, we normalize task shares within each Census occupation so that they sum to one. The crosswalk and aggregation procedure are described in Appendix E.1. This generates 290 Census occupations represented as 8,491 tasks.

Using this data, we document three facts about the intensive margin of work. First, task shares within occupations exhibit a pronounced core-periphery structure. Second, weighting by FTEs and wages shows that this concentration remains significant. Third, mapping tasks into LLM exposure categories reveals sharp differences between exposure measured in labor input versus total labor cost.

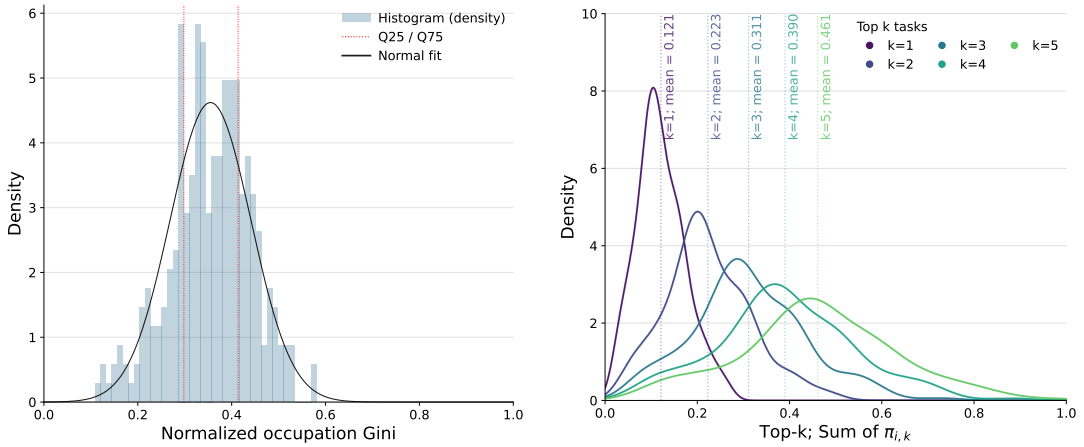
4.1. *Stylized Fact 1: Occupations have a "core-periphery" structure*

A central stylized fact emerging from our analysis is that occupations are not diffuse bundles of evenly weighted activities. Instead, the within-occupation allocation of task shares exhibit a "core-periphery" structure: a small set of tasks accounts for a disproportionate share of total labor input.

Using representative occupations constructed from task shares, we find that occupations comprise, on average, 29 distinct tasks, with a median of 19. This dispersion may be mechanically amplified by the many-to-one mapping from detailed SOC codes to Census occupations, which expands the observed task set at the Census-occupation level. Nevertheless, the cross-occupation distribution of task counts remains wide: the interquartile range spans from 16 tasks at the 25^{th} percentile to 27 tasks at the 75^{th} percentile.

On average, an individual task accounts for 3.4% of the total work, with a median share of 2.1%. Task shares range from below 0.1% to as high as 27%, indicating substantial heterogeneity in how labor is allocated across tasks within occupations. To characterize this within-occupation concentration, we focus on two complementary measures: (i) a normalized task Gini coefficient capturing overall inequality in task shares. The normalized Gini ranges from 0 to 1, where 0 corresponds to a perfectly even task mix and 1 corresponds to complete specialization in a single task, and (ii) the occupation's task shares devoted to the k most intensive tasks (the "top-k share").

Conceptually, the normalized Gini summarizes how unevenly work shares are distributed across the entire task set. By contrast, the top-k share answers the question: "How much of the job is accounted for by its largest tasks?"



(a) Distribution of normalized task Gini coefficients (mean = 0.355, SD = 0.086).

(b) Average cumulative labor share of the top k tasks.

Figure 4. Distribution of task concentration measures across $N = 290$ Census occupations.

Figure 4 shows that occupations exhibit moderate but systematic task concentration. In Panel A, the mean normalized Gini is 0.355 (s.d. = 0.086), indicating substantial within-occupation inequality in task shares. A test of the null hypothesis of uniform task allocation (normalized Gini = 0) is strongly rejected ($p < 0.01$), implying that task shares are not evenly distributed within occupations. At the same time, concentration is far below the single-task extreme, consistent with a mixed regime in which a small set of core tasks accounts for a disproportionate share of labor input alongside a long periphery of less frequent tasks.

Panel B reinforces this interpretation. On average, the top three tasks account for 31% of total work labor input, while the top five tasks account for 46%. Although these averages point to a stable core-periphery structure, the dispersion in top-k shares is sizable: standard deviations range from 5% for the top task to 17% for the top five tasks, indicating substantial heterogeneity in the strength of task concentration across occupations. Thus, while no single task dominates most jobs, some occupations are organized much more tightly around a small task core than others.

Taken together, these results suggest that even when the feasible task set is large, occupations concentrate a sizable fraction of tasks' shares in a relatively small number of activities, though the degree of such concentration varies meaningfully across occupations. All concentration measures are normalized to account for mechanical dependence on the number of observed tasks; formal definitions are provided in Appendix F.1.

4.2. *Stylized Fact 2: Workers Concentrate in a Few "Vital" Tasks*

Fact 1 established that occupations exhibit a pronounced "core-periphery" structure. A natural concern is that this pattern may be driven by niche or low-FTE occupations, rendering it economically unimportant at the level of workers or the aggregate economy. In Fact 2, we address this concern by mapping occupational task profiles into economy-wide labor-market quantities using IPUMS ACS microdata for 2024.

We construct task FTEs by allocating each occupation's total labor input FTE_i across

its constituent tasks using the estimated task shares $\hat{\pi}_{i,k}$, i.e. $\text{FTE}_{i,k} = \hat{\pi}_{i,k} \cdot \text{FTE}_i$.

We analogously construct task wage bills by allocating each occupation's total annual labor cost WB_i (annual earnings) across tasks using the same intensive-margin shares, $WB_{i,k} = \hat{\pi}_{i,k} \cdot WB_i$. This wage-bill decomposition should be interpreted as a proportional imputation of labor costs, rather than as an estimate of task-specific wage premia. These procedures yield a consistent decomposition of labor input and labor cost into task-level units.

This aggregation reveals two distinct but complementary dimensions of the intensive margin at the macro level. First, the representative worker has a concentrated task portfolio, even when large occupations are weighted more heavily. Second, the aggregate economy relies disproportionately on a small set of high-volume tasks, a "vital few", and these tasks account for an even larger share of the aggregate wage bill.

The Representative unit of labor input. We first assess whether the core-periphery structure identified in Fact 1 remains salient when viewed from the perspective of FTEs rather than occupations. To do so, we compute FTE-weighted averages of the occupation-level concentration measures from Section 4.1, yielding statistics for the representative worker in 2024. Table 1 compares these FTE-weighted moments to their unweighted occupation-level counterparts.

Three patterns emerge. First, task concentration remains substantial for the representative worker. The FTE-weighted top-3 task share is 29.1%, indicating that nearly one-third of a typical worker's labor input is devoted to just three activities. Second, despite this slightly lower top-3 share relative to the unweighted occupation mean, overall inequality in task allocation is higher for the representative worker: the FTE-weighted normalized Gini coefficient rises to 0.378, compared to an unweighted mean of 0.355. Third, large occupations encompass substantially more tasks. The representative worker is associated with 37 distinct tasks on average, compared to 29 for the representative occupation.

Taken together, these results imply that large occupations are simultaneously broader in their definition and more internally hierarchical. While they distribute effort across a larger number of tasks, labor input within those task sets is more unevenly allocated. As a result, the core-periphery structure is not attenuated by FTE weighting; rather, it is sharpened. Task specialization is therefore a defining feature of the typical worker's experience, not merely a property of occupational classifications.

Table 1. Task Concentration: Representative Occupation vs. Representative Worker (2024)

Statistic	Unweighted Mean	FTE-Weighted (2024)
Top-3 Share ($C_{i,3}$)	0.311	0.291
Norm. Gini ($\widetilde{G}_i$)	0.355	0.378
Task Count (K_i)	29.3	36.0

Notes: Unweighted statistics are computed across Census occupations. FTE-weighted statistics weight occupations by IPUMS FTEs counts in 2024. $C_{i,3}$ denotes the cumulative share of labor input allocated to the three most intensive tasks within an occupation.

Macro Concentration: The "Vital Few." We next examine whether the diversity of occupations implies a corresponding diversity of tasks in the aggregate economy. To

do so, we rank all unique tasks by their total economy-wide labor input (measured in FTEs) and, separately, by their total contribution to the aggregate wage bill. We then compute the cumulative shares of aggregate labor input and labor cost accounted for by the top p percent of tasks within their respective distributions.

Table 2 reveals pronounced macro-level concentration in both labor input and labor cost. Of the 8,492 unique tasks observed in 2024, the top 1% (approximately 85 tasks) account for 18.3% of total labor input and 22.6% of the aggregate wage bill, when ranked by their respective metrics. The top 5% account for 44.9% of FTEs and 51.2% of the wage bill, while the top 10% account for 61.0% of FTEs and 66.3% of the wage bill. Overall, a few hundred tasks dominate aggregate labor input. Similarly a few account for an even larger share of aggregate labor cost.

This pattern mirrors the within-occupation core-periphery structure documented in Fact 1: just as tasks' share concentrate on a small subset of tasks within jobs, aggregate economic value concentrates on a small subset of tasks across the economy.

Table 2. Macro Concentration of Task Exposure and Wage Bill: Shares Held by the Top p% of Tasks (N = 8,491, O*NET 29.0, IPUMS 2024)

	Top 1% Share	Top 5% Share	Top 10% Share
Number of unique tasks	85	425	850
Share of total FTEs	18.3%	44.9%	61.0%
Share of total wage bill	22.6%	51.2%	66.3%

Notes: Table reports the cumulative share of total economy-wide task exposure and wage-bill accounted for by the top p percent of tasks, ranked by FTE and wage bill volume. FTEs allocate occupation-level employment across tasks using $\hat{\pi}_{i,k}$, i.e. $FTE_k = \sum_i FTE_i \cdot \hat{\pi}_{i,k}$. Task wage-bills allocate occupation-level wage bills analogously, i.e. $WB_k = \sum_i WB_i \, \hat{\pi}_{i,k}$. Total FTEs N_{FTE} = 83,552,261 (IPUMS 2024) and wage bill: $\$3.13 \cdot 10^{12}$ (1999 constant USD)

4.3. *Stylized Fact 3: The Structure of AI Exposure Across Tasks*

After establishing that aggregate labor input is concentrated in a limited set of economically central tasks (Facts 1 and 2), the analysis now considers how this task concentration maps into exposure to advances in artificial intelligence (AI). Because AI capabilities operate at the task level rather than the occupational level, their aggregate economic relevance depends not only on the number of exposed tasks but also on the volume of labor input and labor cost allocated to those tasks.

This analysis adopts the task-level exposure framework of Eloundou et al. (2023), which classifies tasks according to their potential for LLM–driven reductions in task execution time.

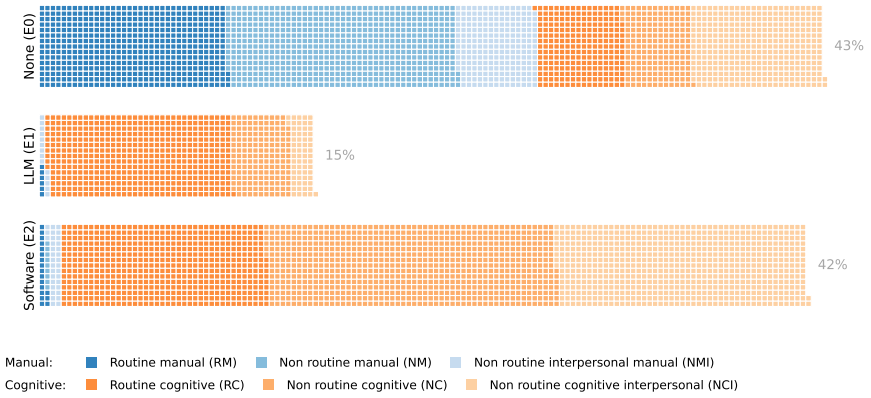
Within this framework, LLMs are defined as general-purpose AI models trained on large corpora of text that can generate language-based outputs and, for certain tasks, substantially reduce the time required for human execution while maintaining output quality. Each task is assigned to one of three mutually exclusive exposure categories:

- **No Exposure–E0:** Tasks for which LLMs cannot meaningfully reduce required labor time while preserving output quality.
- **Direct Exposure–E1:** Tasks for which a standalone conversational LLM could reduce task completion time by at least 50 percent.
- **Software Exposure–E2:** Tasks for which LLMs embedded in software tools could reduce task completion time by at least 50 percent.

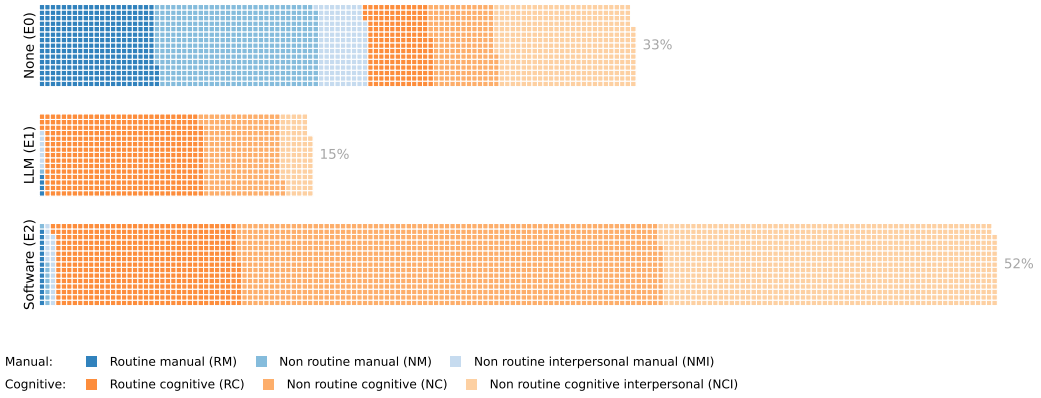
Throughout this analysis, these categories are interpreted as measures of technical exposure potential, the extent to which time savings may be technologically feasible, rather than as estimates of realized adoption or observed productivity effects.

Tasks are classified into routine and non-routine categories using an LLM-based classification procedure, following the methodology introduced in Section 3.1. Appendix C provides details on the classification taxonomy, model specifications, prompt design, results, and robustness checks.

Finally, we aggregate task-level FTEs and task-level wage bills by task type and AI exposure category. These aggregates reveal a pronounced divergence between exposure measured in labor volume and exposure measured in labor cost. The following paragraph reports the resulting estimates.



(a) FTE composition by GPT-4 exposure for a population of $N_{FTE} = 83{,}552{,}262$ workers (IPUMS 2024)



(b) Wage-bill composition by GPT-4 exposure for an aggregate labor cost of $\$3.13 \cdot 10^{12}$ (1999 constant USD).

Figure 5. Waffle plots of FTE and wage-bill composition by GPT-4 exposure category (Eloundou et al. 2023). Color gradients are based on the two broader task categories manual and cognitive. *Notes:* Each waffle represents 0.02% of the total. Shares are computed over the analysis sample (the subset of the 2024 workforce with valid task profiles). Appendix Figure C1 provides an additional visualization based on a Routine/Non-routine task categories.

From Figure 5 and the detailed results in Appendix Table C2, specific patterns stand out. Tasks classified as not exposed (E0) account for 42.9% of aggregate FTEs.

However, these tasks represent a smaller share of labor cost, capturing 33.6% of the aggregate wage bill. Additionally, E0 task labor input is dominated by manual activities (routine and non-routine manual), which account for 63.5% of E0 FTEs. This wedge between volume and value is consistent with the interpretation that the tasks least amenable to LLM time savings are disproportionately concentrated in manual, lower-wage segments of the task distribution.

Tasks with direct exposure to standalone conversational LLMs (E1) comprise 15.0% of FTEs and 15.1% of the wage bill. The composition of E1 is sharply tilted toward cognitive tasks: 97.3% of E1 FTEs are cognitive; more precisely, routine cognitive tasks account for 67.4% of E1 FTEs, while manual tasks account for only 2.7%. Under the E0 classification, technical exposure potential is therefore concentrated in codifiable cognitive activities rather than manual work.

By contrast, tasks classified as exposed to LLMs embedded software tools (E2) account for a larger share of labor cost, capturing 51.2% of the aggregate wage bill, while representing 42.0% of FTEs. The E2 category is overwhelmingly cognitive with 97.1% of its FTEs, and 98.2% of the E2 wage bill in this type of tasks. Its employment composition tilts toward higher-complexity cognitive work, with non-routine cognitive tasks accounting for 70.3% of E2 FTEs (NC and NCI). A particularly pronounced shift concerns non-routine cognitive interpersonal (NCI) activities: NCI tasks account for only 8.35% of E1 FTEs but 32.6% of E2 FTEs.

Three implications follow from this decomposition. First, exposure measured in task counts can be misleading: weighting by FTEs and wage bills yields different exposure shares, so the aggregate relevance of AI exposure depends on where labor input and labor cost are concentrated. Second, the exposure channel matters: under this classification, the E1 category remains economically modest, whereas the E2 category encompasses a majority of the wage bill despite representing a smaller share of FTEs. Interpreted as a measure of technical feasibility, this pattern suggests that the largest concentration of potential economic impact from LLMs may arise through complementary software integration in higher-wage cognitive tasks, rather than through standalone conversational LLMs. Third, exposure differs systematically by task type: E0 is predominantly manual, E1 is concentrated in routine cognitive work, and E2 is concentrated in non-routine cognitive work, including interpersonal activities.

5. Measurement Validation: Wage Gradients from Task Shares Versus Existing Proxies

This section treats wages as an external validation environment for the intensive-margin measurement object recovered from task-frequency data. The goal is not to identify causal wage effects of task structure. Rather, we ask whether moments constructed from the estimated task shares $\hat{\pi}_{i,k}$ behave like economically meaningful within-occupation labor-input allocations.

Wages are an imperfect external criterion, reflecting productivity, rents, sorting, and institutions. We therefore use wage gradients only for validation and do not interpret the regressions causally. We test whether moments of $\hat{\pi}_{i,k}$ organize systematic wage differences beyond observed work content and task breadth. Crucially, $\hat{\pi}_{i,k}$ separates task content (which tasks are performed) from task allocation (how labor input is distributed across those tasks).

We focus on two validation checks. First, we test whether the specialization-wage gradient exhibits economically interpretable heterogeneity with respect to work content: if specialization reflects productive focus primarily in non-routine work, the gradient should be positive in low-routine occupations and attenuate as routine intensity rises. Second, we test non-interchangeability: we examine whether these patterns are specific to labor-input task shares, or whether they also arise under alternative within-occupation weighting schemes commonly used in the task literature (importance weights, relevance weights, and core-supplemental heuristics).

5.1. *Regression Evidence: Wage Association to Specialization and Work Content*

The premise is that occupations differ not only in the tasks they include (work content), but also in how labor input is distributed across those tasks (specialization). While standard task-based approaches typically emphasize content, the present analysis quantifies the intensive margin of specialization and assesses whether it carries independent wage association beyond work content. Addressing this question provides insight into how changes in occupational task structure may affect earnings, even when the set of tasks performed remains fixed.

A priori, the wage effects of occupational specialization are ambiguous. Concentration of task shares in routine tasks may reflect Taylorist fragmentation and deskilling, whereas concentration in non-routine tasks may reflect expertise, skill deepening, and rent generation. We use the interaction between specialization and routine intensity to distinguish empirically between these channels.

Guided by this distinction, we summarize each occupation's task structure using three complementary moments derived from O*NET task shares. First, occupational specialization is measured using the normalized Gini coefficient $\widetilde{G}_i$, introduced in Section 4.2; this coefficient increases as labor input becomes more concentrated in a narrow core of dominant tasks. Second, routine intensity, Rout_i, is defined as the sum of task shares allocated to routine tasks, capturing the content of work. These measures represent distinct features of task structure: specialization reflects the concentration of labor input across tasks, whereas routine intensity reflects the composition of labor input between routine and non-routine activities. Because Spec_i and Rout_i vary only at the occupation level, identification comes from between-occupation variation; stan-

dard errors are clustered by occupation accordingly. Finally, we define task breadth, B_i, as the number of distinct tasks associated with occupation i. Appendix F.1 provides formal definitions and construction details.

We relate worker wages to these occupation-level measures. Let worker l be employed in occupation i and consider the following specification:

$$\ln w_{l,i} = \beta_S \, \mathrm{Spec}_i^Z + \beta_R \, \mathrm{Rout}_i^Z + \beta_B \, \ln(\mathrm{B}_i) + \beta_{SR} \, \mathrm{Spec}_i^Z \cdot \mathrm{Rout}_i^Z + X_{l,i}'\gamma + \varepsilon_{l,i}, \quad (6)$$

where $\ln w_{l,i}$ denotes log hourly wages in 1999 dollars. Spec_i^Z is the standardized occupational specialization and Rout_i^Z is the standardized routine intensity.

The control vector $X_{l,i}$ includes gender, age, age squared, and fixed effects for education bins and industry bins. All specifications use IPUMS person weights and heteroskedasticity-robust standard errors clustered by occupation. Because specialization may reflect occupational sorting, firm organization, or rent-sharing, we interpret all estimates as descriptive associations rather than causal effects.

Table 3. Wage Associations with Occupational Specialization

	(1)	(2)	(3)	(4)
Specialization	0.0853***	0.0842***	0.0607***	0.0626***
	(0.0156)	(0.0162)	(0.0134)	(0.0137)
Log Breadth		0.00547	0.00480	0.00428
		(0.0344)	(0.0297)	(0.0300)
Routine			-0.108***	-0.102***
			(0.0139)	(0.0142)
Specialization × Routine				-0.0194**
				(0.00984)
Control variables	✓	✓	✓	✓
R^2	0.321	0.321	0.336	0.336
Clusters (occupations)	290	290	290	290
N	838,747	838,747	838,747	838,747

Notes: $N = 838{,}747$, representing a weighted population of 92,161,825 individuals. The dependent variable is log hourly wages in 1999 CPI-adjusted dollars. The standardized normalized task Gini coefficient captures occupational specialization, with higher values indicating greater concentration of labor input in a small set of tasks. Routine intensity is the standardized share of labor input devoted to routine tasks. Task breadth is the log number of distinct tasks mapped to occupation i. All specifications include controls for gender, age, age squared, and fixed effects for education bins and industry bins, and are estimated using IPUMS person weights. Robust standard errors clustered at the occupation level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 3 reports a sequence of specifications designed to distinguish the wage associations of occupational specialization from alternative explanations based on work content and task breadth. Model (1) introduces occupational specialization as the sole

task-based regressor. Specialization is strongly and positively associated with wages: a one standard deviation increase in specialization is associated with an 8.5 percent higher hourly wage ($\hat{\beta}_S = 0.0853$, s.e. $= 0.0156$), and the specification explains 32.1 percent of the variation in log hourly wages.

Model (2) adds controls for task breadth. Task breadth itself has a small and statistically insignificant association with wages, while the specialization coefficient remains essentially unchanged. This suggests that wage differences are not driven by the number of tasks performed within an occupation, but by how labor input is allocated across those tasks.

Model (3) introduces routine intensity in a horse-race specification that jointly includes specialization, routine intensity, and task breadth. Two results stand out. First, routine intensity is strongly negatively associated with wages: a one–standard-deviation increase in routine task share is associated with 10.8 percent lower hourly wage (s.e. $= 0.0139$). Second, although attenuated relative to column (2), occupational specialization remains positively and statistically significantly associated with wages even after controlling for work content ($\hat{\beta}_S = 0.0606$, s.e. $= 0.0134$). Including routine intensity raises explanatory power to $R^2 = 0.336$, while task breadth remains statistically insignificant.

Model (4) adds an interaction between specialization and routine intensity. The interaction term is negative but small and statistically significant at a 5% level ($\hat{\beta}_{SR} = -0.0194$, s.e. $= 0.0098$), providing evidence of meaningful heterogeneity in the wage-specialization gradient by routine intensity in this specification.

Taken together, these results indicate that specialization and routine content capture distinct wage-relevant margins. Specialization reflects *how* labor input is organized across tasks within an occupation, while routine intensity reflects *what* tasks are performed. Even conditional on work content and breadth, specialization carries independent and economically meaningful wage associations.

These associations are robust to alternative earnings definitions and sample restrictions. Appendix G.2 shows that the specialization premium persists when using annual earnings instead of hourly wages, while Appendix G.3 documents similar patterns within a full-time, full-year sample. Across specifications, the magnitude and attenuation pattern of the specialization coefficient closely mirror the baseline results, indicating that the observed associations are not driven by labor supply differences, part-time work, or irregular employment arrangements. Finally, to visualize the independent effects of these mechanisms, we provide conditional partial regression plots for specialization and routine intensity in Appendix G.4.

5.2. *Heterogeneity: Marginal Specialization - Wage Association by Routine Intensity*

Motivated by the intuition that specialization is expected to be more valuable in non-routine occupations and the interaction term results, we examine the implied marginal effect of specialization:

$$\frac{\partial \ln w}{\partial \text{Spec}^Z} = \hat{\beta}_S + \hat{\beta}_{SR} \cdot \text{Rout}^Z, \tag{7}$$

so that a negative $\hat{\beta}_{SR}$ indicates a declining association to specialization as routine intensity rises.

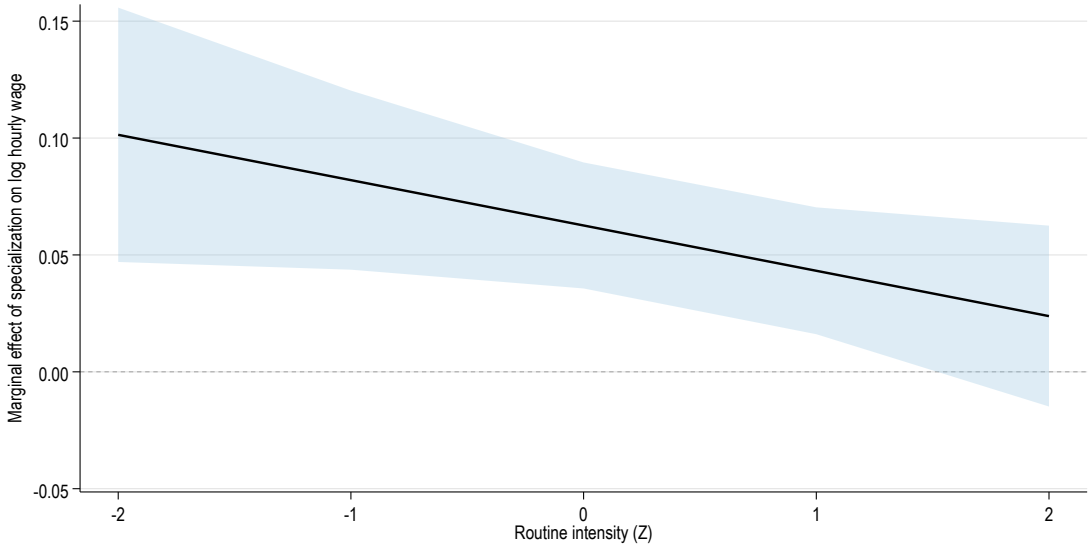


Figure 6. Marginal Specialization; Wage Association by Routine Intensity (95% CI)

Figure 6 plots the implied marginal effects of specialization evaluated at different values of routine intensity. In low-routine occupations, two standard deviations below the mean share of labor input allocated to routine tasks, a one standard deviation increase in specialization is associated with a 10.1% increase in wages (s.e. = 0.028; $p < 0.001$). At the mean level of routine intensity, the implied marginal effect is 6.3% (s.e. = 0.014; $p < 0.001$) and statistically significant. As routine intensity increases, the implied marginal effect attenuates: at one standard deviation above the mean, the estimate falls to 4.3% (s.e. = 0.014; $p = 0.002$), and at two standard deviations above the mean it declines further and becomes statistically indistinguishable from zero ($p = 0.226$). Exact marginal effect estimates are reported in Appendix Table G1.

This pattern is consistent with bundle-based interpretations in which specialization is more valuable when it reflects concentration in complex, discretionary tasks rather than standardized activities. In occupations with substantial non-routine content, specialization may proxy for deeper expertise and greater discretion in task performance, yielding a larger wage premium; in more routine environments, specialization may instead reflect concentration in standardized activities with lower barriers to entry. Importantly, our analysis is descriptive and does not identify causal effects; we use this lens only to interpret why the specialization-wage association varies systematically with routine intensity.

5.3. *Comparison to Alternative Intensive-Margin Proxies*

A common approach in the task literature is to construct within-occupation task weights using proxies such as O*NET importance and relevance ratings, heuristic "core–supplemental" priors, or uniform task weights. These summaries can be useful for describing task salience, but they are not designed to recover an accounting-consistent proxy for how labor input is allocated across tasks.

To assess whether our wage associations are specific to labor-input task shares, we

re-estimate the fully specified wage model using three alternative weighting schemes that preserve within-occupation heterogeneity: (i) O*NET importance weights, (ii) O*NET relevance weights, and (iii) a core–supplemental prior assigning weight 2 to core tasks and 1 to supplemental tasks. Uniform weights are excluded because equal task weights mechanically imply zero specialization score (normalized Gini = 0).

We compare intensive-margin proxies using Pearson and Spearman rank correlations and find that the task shares $\pi_{i,k}$ are positively but only moderately correlated with rating- and taxonomy-based schemes. This indicates that $\pi_{i,k}$ captures a distinct allocation measure compared to other intensive-magin proxies. In addition, the Spearman and Pearson correlations between alternative proxies are moderate indicating that they can materially re-rank occupations and should not be treated as interchangeable measures of the intensive margin. Detailed results are reported in Appendix G.5.

Table 4 shows that the negative association between routine intensity and wages is robust across all weighting schemes. By contrast, the specialization coefficient is sensitive to how tasks are weighted: under task-share ($\pi_{i,k}$) weights, specialization is positive and statistically significant and its interaction with routine intensity is negative. In contrast rating- and heuristic-based specialization becomes negative or statistically indistinguishable from zero and the interaction vanishes.

Table 4. Fully specified wage model (Table 3 - Column 4) under alternative intensive-margin proxy schemes

	Task-share (π)	Importance	Relevance	Core-supp (2:1)
Specialization	0.0626^{***}	-0.0415^{**}	-0.0262^{*}	-0.0234
	(0.0137)	(0.0184)	(0.0155)	(0.0155)
Breadth	0.00428	0.0629	0.0367	0.0313
	(0.0300)	(0.0397)	(0.0351)	(0.0344)
Routine	-0.102^{***}	-0.123^{***}	-0.121^{***}	-0.124^{***}
	(0.0142)	(0.0167)	(0.0176)	(0.0169)
Specialization × Routine	-0.0194^{**}	0.00391	0.00366	0.0126
	(0.00984)	(0.0138)	(0.0133)	(0.0118)
Control variables	✓	✓	✓	✓
R^2	0.336	0.336	0.334	0.334
Clusters (occupations)	290	290	290	290
N	838,747	838,747	838,747	838,747

Notes: Each column reports the fully specified wage regression (Section 5.1–Model 4) from the corresponding proxy-specific table. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

If specialization proxies for depth, its wage gradient should be positive and strongest in non-routine occupations; in highly routine work, specialization may carry a smaller premium or none. This prediction is consistent with evidence that occupations requiring greater expertise tend to pay higher wages and that expertise measures have robust predictive power for earnings (Autor and Thompson 2025). Against this benchmark, the sign reversal under importance/relevance and core-supp weights is more naturally

interpreted as a proxy mismatch: rating- and taxonomy-based schemes capture task salience rather than the within-occupation allocation of labor input.

We replicate the marginal-effect exercise in Section 5.2 under each alternative proxy. The implied marginal effects under rating-based and heuristic schemes are generally statistically insignificant; when significant, they are often negative even in low-routine occupations and exhibit no comparable attenuation pattern. Taken together, these weak and unstable marginal-effect patterns indicate that alternative weighting schemes are not empirically interchangeable with task-share weights for measuring occupational specialization. Detailed results are reported in Appendix G.6.

6. Concluding Remarks

This paper develops an interpretable measurement methodology for the intensive margin of work using publicly available O*NET task-frequency survey responses. Our central output is a within-occupation task share, $\hat{\pi}_{i,k}$, that sums to one within each occupation and can be interpreted as a proxy for the within-occupation allocation of labor input to tasks. Under a homogeneous within-occupation task-duration assumption, $\hat{\pi}_{i,k}$ coincides with task time shares; more generally, it provides a transparent intensity weighting for aggregating task-level attributes into occupation-level indices. Because $\hat{\pi}_{i,k}$ behaves like a budget share, it delivers a disciplined mapping from task content to standard occupation aggregates: for any task attribute a_k, such as exposure, employment or wages.

Methodologically, we translate O*NET's ordinal frequency categories into continuous estimators of task activity and task shares with uncertainty. We recover the unconditional distribution of frequency categories for each occupation-task pair (i, k), convert these probabilities into an annualized task-flow estimator $\hat{\mu}_{i,k}$ using a transparent calibration from ordinal bins to yearly occurrences, and obtain $\hat{\pi}_{i,k}$ by normalizing task flows within occupation. To propagate sampling uncertainty from survey responses into $\hat{\mu}_{i,k}$, $\hat{\pi}_{i,k}$, and downstream statistics, we introduce a covariance completion procedure that respects the simplex geometry of frequency-category probabilities and yields a positive semidefinite covariance matrix consistent with released marginal uncertainties. This delivers a unique set of fully specified, inference-ready measurements of the intensive-margin task weights.

Applying the framework to O*NET v29.0 and IPUMS-ACS microdata highlights several features of work organization and AI exposure. Occupations exhibit pronounced core-periphery structures: a small set of tasks accounts for a substantial share of implied labor input, and concentration is the norm rather than the exception. This concentration matters for economic measurement. Uniform task weights or coarse task groupings can misstate which tasks dominate labor input and compensation, and can shift the implied incidence of task-level capability labels once those labels are aggregated to the occupation level. In the context of AI capabilities, an employment lens versus a wage-bill lens changes both the level and composition of implied exposure, and can alter which occupations and task categories account for the economic stakes of AI. Finally, the intensive-margin objects provide a unified way to characterize work content. In descriptive wage regressions, specialization is positively associated with wages conditional on work content, with the strongest relationships concentrated in non-routine occupations, consistent with the idea that specialization primarily pays off in complex task environments. This specialization metric is not recovered by other standard intensive-margin proxies.

Several limitations motivate natural extensions. Because O*NET does not report task durations, $\hat{\pi}_{i,k}$ is best interpreted as a frequency-based proxy for labor input rather than a direct time-use measure. Hence, heterogeneity in task duration could affect the mapping to time shares and implied concentration. A first priority is therefore to validate and calibrate durations using time-use data (e.g. ATUS). This will enable a better characterization of $\hat{\pi}_{i,k}$. In addition, O*NET task lists are not a full census of workplace activities and update with delay, which may understate emerging tasks. Extending the framework to repeated task-content snapshots would allow us to quantify how work content, task concentration, and $\hat{\pi}$-weighted exposure evolve over time. Finally, while our inference procedure transparently propagates sampling uncertainty, richer within-occupation dependence structures, especially cross-task de-

pendence, could improve uncertainty quantification for downstream aggregates.

Acknowledgments and Disclosure of interest

The authors are grateful to David Autor and Iona Clark for valuable feedback and support. The authors declare that they have no financial or other competing interests related to this research.

Data availability statement

The authors confirm that the data supporting the findings of this study are available. The experiments were conducted in a reproducible manner, and the source code, as well as data are available on a GitHub repository: `https://github.com/Pierre-Bouquet/estimating-intensive-margin-work`.

References

Acemoglu, Daron, and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In *Handbook of Labor Economics*, Vol. 4, 1043–1171. Elsevier.

Acemoglu, Daron, and Pascual Restrepo. 2022. "Tasks, Automation, and the Rise in US Wage Inequality." *Econometrica* 90 (5): 1973–2016.

Aitchison, John. 1982. "The Statistical Analysis of Compositional Data." *Journal of the Royal Statistical Society: Series B (Methodological)* 44 (2): 139–160.

Autor, David, and Neil Thompson. 2025. "Expertise." *Journal of the European Economic Association* 23 (4): 1203–1271.

Autor, David H., and David Dorn. 2013. "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market." *American Economic Review* 103 (5): 1553–1597.

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *The Quarterly Journal of Economics* 118 (4): 1279–1333.

Bouquet, Pierre, Nicolo Bagnoli, and Yossi Sheffi. 2025. *Estimating the Task Content of Work: Workforce Design for AI-Driven Human-Robot Collaboration in Intralogistics*. Research Paper 2025/035. MIT Center for Transportation & Logistics. SSRN Working Paper, `https://ssrn.com/abstract=5841942`.

Boyd, Stephen, and Lieven Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press.

Brynjolfsson, Erik, and Tom Mitchell. 2017. "What Can Machine Learning Do? Workforce Implications." *Science* 358 (6370): 1530–1534.

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock. 2023. "GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models." ArXiv:2303.10130.

Goodman, Leo A. 1960. "On the Exact Variance of Products." *Journal of the American Statistical Association* 55 (292): 708–713.

Martin, Josh, and Ellys Monahan. 2022. *Developing a Method for Measuring Time Spent on Green Tasks*. Technical Report. Office for National Statistics.

MOSEK ApS. 2025. *The MOSEK Python Fusion API Manual*. Documentation for MOSEK Optimization Suite 11.1, Accessed 2026-01-16. `https://docs.mosek.com/11.1/pythonfusion/intro_info.html`.

National Center for O*NET Development. 2025. "O*NET OnLine." Accessed 2026-01-16. `https://www.onetonline.org/`.

Ruggles, Steven, Sarah Flood, Matthew Sobek, Daniel Backman, Grace Cooper, Julia A. Rivera Drew, Stephanie Richards, Renae Rogers, Jonathan Schroeder, and Kari C.W.

Williams. 2025. "IPUMS USA: Version 16.0." Minneapolis, MN. Dataset, https://doi.org/10.18128/D010.V16.0.

Tomlinson, Kiran, Sonia Jaffe, Will Wang, Scott Counts, and Siddharth Suri. 2025. "Working with AI: Measuring the Applicability of Generative AI to Occupations." ArXiv:2507.07935.

van der Vaart, Aad W. 2000. *Asymptotic Statistics*. Cambridge University Press.

Webb, Michael. 2019. "The impact of artificial intelligence on the labor market." *Available at SSRN 3482150* .

Appendix A.

A.1. *Glossary of Variables*

- $\hat{p}_{i,k,r} \in [0,1]$: The unconditional probability that a worker in occupation i selects frequency category r for task k.
- $\hat{\mu}_{i,k} \in [0,2080]$: The expected annual flow of task k occurrences for a worker in occupation i.
- $\hat{\pi}_{i,k} \in [0,1]$: The expected labor input allocated to task k within occupation i.

A.2. *Variance of the Unconditional Probability Vector*

The unconditional probability vector $\hat{\mathbf{p}}_{i,k}$ is constructed by scaling the conditional frequency distribution $\hat{\mathbf{F}}_{i,k}$ by the relevance probability $\hat{R}_{i,k}$. We treat the estimators of the extensive margin (relevance) and the intensive margin (frequency) as statistically independent random variables inherent to the survey design. Under this independence assumption, the variance of the unconditional estimators is given by Goodman's product variance formula (Goodman 1960).

- For the mass point at zero ($r = 0$), which represents the complement of the extensive margin:

$$\mathrm{Var}(\hat{p}_{i,k,0}) = \mathrm{Var}(1 - \hat{R}_{i,k}) = \mathrm{Var}(\hat{R}_{i,k}). \tag{A1}$$

- For the active frequency categories ($r \in \{1, \ldots, 7\}$):

$$\hat{p}_{i,k,r} = \hat{R}_{i,k} \cdot \hat{F}_{i,k,r}. \tag{A2}$$

  Applying the exact variance formula for the product of two independent random variables:

$$\mathrm{Var}(\hat{p}_{i,k,r}) = \hat{F}^2_{i,k,r}\,\mathrm{Var}(\hat{R}_{i,k}) + \hat{R}^2_{i,k}\,\mathrm{Var}(\hat{F}_{i,k,r}) + \mathrm{Var}(\hat{R}_{i,k})\,\mathrm{Var}(\hat{F}_{i,k,r}). \tag{A3}$$

where $\hat{R}_{i,k}$ and $\mathrm{Var}(\hat{R}_{i,k})$ denote the point estimate and variance of the relevance estimator for task k in occupation i, and $\hat{F}_{i,k,r}$ and $\mathrm{Var}(\hat{F}_{i,k,r})$ denote the point estimate and variance of the conditional frequency estimator for category r. This calculation yields the vector of marginal variances for the task frequency distribution.

A.3. *Covariance Completion*

Estimating the variance of the annualized task frequency, $\hat{\mu}_{i,k} = \mathbf{w}^\top \hat{\mathbf{p}}_{i,k}$, requires the full covariance matrix of the probability vector, denoted $\hat{\Sigma}_{i,k} \in e\mathbb{S}^8$. Crucially, the elements of $\hat{\mathbf{p}}_{i,k}$ are structurally dependent: because they sum to unity, they are necessarily correlated (i.e. an increase in the probability of one frequency category necessitates a decrease in others). However, the O*NET survey design identifies only the marginal variances (the diagonal elements), leaving these cross-category correlations undefined.

To recover a valid covariance structure, we exploit the geometric properties of the probability simplex. Any valid covariance matrix for a probability vector must satisfy the tangent space constraint $\Sigma\mathbf{1} = \mathbf{0}$ (Aitchison 1982), which explicitly encodes the negative correlation structure required by the unit-sum constraint. We estimate the latent covariance matrix $\hat{\Sigma}_{i,k}$ by solving a convex completion problem that matches the observed marginal variances while minimizing the squared Frobenius norm $\|\Sigma\|_F^2$. This objective function acts as a parsimonious regularizer (Boyd and Vandenberghe 2004), recovering the minimum-energy covariance structure required to satisfy the geometric constraints without imposing spurious correlations.

Let $\mathbf{v}_{i,k}$ denote the vector of marginal variances derived in Appendix A.2, where the r-th element corresponds to $\mathrm{Var}(\hat{p}_{i,k,r})$. We obtain $\hat{\Sigma}_{i,k}$ as the solution to the following semidefinite program (SDP):

$$\begin{aligned} \hat{\Sigma}_{i,k} \;\in\; \arg\min_{\Sigma \in \mathbb{S}^8} \; & \|\Sigma\|_F^2, \\ \text{s.t.} \quad & \Sigma \succeq 0, \\ & \mathrm{diag}(\Sigma) = \mathbf{v}_{i,k}, \\ & \Sigma\mathbf{1} = \mathbf{0}. \end{aligned} \tag{A4}$$

where $\Sigma \succeq 0$ ensures the matrix is positive semidefinite. We solve this optimization using the MOSEK solver, the parameters are available in Appendix D.1 (MOSEK ApS 2025). With the full covariance matrix recovered, the variance of the annual flow estimator follows from the quadratic form:

$$\mathrm{Var}(\hat{\mu}_{i,k}) \;=\; \mathbf{w}^\top \hat{\Sigma}_{i,k} \mathbf{w}. \tag{A5}$$

This scalar variance provides a measure of uncertainty for the annualized flows, serving as the requisite input for characterizing the precision of the standardized labor service estimates.

A.4. *Asymptotic Variance of Standardized Task Intensities*

This section derives the asymptotic variance for the standardized task intensity $\hat{\pi}_{i,k}$. We begin with the general case where tasks may have heterogeneous durations, and then derive the simplified estimator used in the main text.

Let $d_{i,k}$ denote the fixed duration (in hours) per occurrence of task k. The task intensity is defined as the share of total occupational labor input allocated to task k:

$$\hat{\pi}_{i,k} = \frac{\hat{\mu}_{i,k}\, d_{i,k}}{M_i}, \qquad \text{where} \quad M_i = \sum_{j \in K_i} \hat{\mu}_{i,j}\, d_{i,j}. \tag{A6}$$

We assume the flow estimators $\{\hat{\mu}_{i,j}\}_{j \in K_i}$ are independent random variables across distinct tasks. Applying the multivariate delta method, the asymptotic variance is given by the first-order Taylor expansion:

$$\mathrm{Var}(\hat{\pi}_{i,k}) \;\approx\; \sum_{j \in K_i} \left(\frac{\partial\, \hat{\pi}_{i,k}}{\partial\, \hat{\mu}_{i,j}} \right)^2 \mathrm{Var}(\hat{\mu}_{i,j}). \tag{A7}$$

The partial derivatives with respect to the flow vectors are derived via the quotient rule. Note that $\frac{\partial M_i}{\partial \hat{\mu}_{i,j}} = d_{i,j}$.
For the own-task effect ($j = k$):

$$\frac{\partial \hat{\pi}_{i,k}}{\partial \hat{\mu}_{i,k}} = \frac{d_{i,k} M_i - (\hat{\mu}_{i,k} d_{i,k})\, d_{i,k}}{M_i^2} = \frac{d_{i,k}}{M_i}\,(1 - \hat{\pi}_{i,k})\,. \tag{A8}$$

For the cross-task effect ($j \neq k$):

$$\frac{\partial \hat{\pi}_{i,k}}{\partial \hat{\mu}_{i,j}} = -\frac{\hat{\pi}_{i,k}\, d_{i,j}}{M_i}, \qquad j \neq k. \tag{A9}$$

Substituting these gradients into the variance sum yields the general asymptotic variance:

$$\mathrm{Var}(\hat{\pi}_{i,k}) \;\approx\; \frac{1}{M_i^2} \left[d_{i,k}^2 (1 - \hat{\pi}_{i,k})^2\, \mathrm{Var}(\hat{\mu}_{i,k}) \;+\; \hat{\pi}_{i,k}^2 \sum_{j \neq k} d_{i,j}^2\, \mathrm{Var}(\hat{\mu}_{i,j}) \right]. \tag{A10}$$

And making the assumption from Section 2 on $d_{i,j} = 1$ for all $j \in K_i$ yields:

$$\mathrm{Var}(\hat{\pi}_{i,k}) \;\approx\; \frac{1}{M_i^2} \Big[(1 - 2\hat{\pi}_{i,k})\, \mathrm{Var}(\hat{\mu}_{i,k}) + \hat{\pi}_{i,k}^2\, \mathrm{Var}(M_i) \Big]. \tag{A11}$$

A.5. *Modeling Assumptions and Choices*

The empirical strategy relies on a set of structural assumptions and pragmatic modeling choices that ensure tractability, internal consistency, and reproducibility. These assumptions underpin the estimation of task distributions, covariance structures, and FTE allocations.

- **Survey-stage independence.** The task relevance $\hat{RT}_{i,k}$ and conditional frequency $\hat{FT}_{i,k,r}$ are treated as statistically independent. This enables exact variance evaluation for unconditional category probabilities $\hat{p}_{i,k,r}$ through the product-variance formula.
- **Aggregation independence.** Tasks are assumed independent within and across occupations.
- **Normal inference.** Interval estimates rely on normal approximations obtained from delta-method variance propagation, with truncation applied as needed to respect simplex or nonnegativity constraints.
- **Covariance completion.** The latent covariance matrix of the eight-category probability vector is recovered by selecting a positive semidefinite matrix that matches the observed O*NET marginal variances and satisfies the simplex tangent constraint ($\Sigma\mathbf{1} = 0$). Among all feasible completions, we choose the minimum–Frobenius-norm solution, $\arg\min_{\Sigma} \|\Sigma\|_F^2$, which is a strictly convex objective over a convex feasible set and therefore yields a unique solution. This solution selects the "minimum-energy" covariance consistent with the observed marginals and avoids introducing correlation structure beyond what is required by the constraints.
- **Annualization mapping.** Ordinal task-frequency categories are mapped to expected annual occurrence counts using a fixed midpoint vector w, providing a consistent and conservative annualization of task frequencies.
- **Equal per-occurrence duration.** Within an occupation, occurrences of a given task are assumed to have a common expected duration. Under this assumption, annualized occurrence counts proxy relative task shares across tasks.
- **Task–DWA allocation.** The O*NET crosswalk linking tasks to DWAs is treated as fixed. When a task maps to multiple DWAs, its FTE contribution is split evenly across linked DWAs to avoid double counting.
- **Numerical stabilization.** Small ε-regularization and renormalization are applied to near-zero probabilities or standard errors to maintain numerical feasibility. Solutions are accepted when positive semidefiniteness holds within standard solver tolerances.

A.6. *Example Estimated Task Flows and Task Shares for an Occupation*

Table A1. Illustrative estimated task flows and task shares for "Farm Labor Contractors"

Task	$\hat{\mu}_{i,k}$	$\hat{\pi}_{i,k}$ (%)
Supervise the work of contracted employees.	471.29 (116.16)	24.05 (5.40)
Employ foremen to deal directly with workers when recruiting, hiring, instructing, assigning tasks, and enforcing work rules.	354.35 (142.12)	18.08 (6.30)
Pay wages of contracted farm laborers.	342.61 (167.77)	17.48 (7.30)
Provide food, drinking water, and field sanitation facilities to contracted workers.	286.79 (51.34)	14.63 (3.00)
Recruit and hire agricultural workers.	246.38 (90.35)	12.57 (4.40)
Furnish tools for employee use.	142.47 (29.96)	7.27 (1.70)
Direct and transport workers to appropriate work sites.	115.83 (25.31)	5.91 (1.50)

Notes: $\hat{\mu}_{i,k}$ denotes the estimated annualized task flow (occurrences per worker-year) constructed from the unconditional O*NET frequency distribution and the annualization vector $\mathbf{w}$. $\hat{\pi}_{i,k}$ denotes the implied within-occupation task share under the normalization $d_{i,k} = 1$, computed as $\hat{\pi}_{i,k} = \hat{\mu}_{i,k} / \sum_{j \in K_i} \hat{\mu}_{i,j}$. Standard errors are propagated via the delta method as described in Appendix A.4. Each cell reports estimate (SE). For $\hat{\pi}_{i,k}$, both the point estimate and the SE in parentheses are reported in percentage points.

Appendix B.

B.1. *IPUMS Descriptive Statistics*

Our analysis utilizes microdata from the 2024 American Community Survey (ACS) as provided by IPUMS (Ruggles et al. 2025). We restrict the analytical sample to civilian, non-institutionalized individuals aged 16 to 64 who are currently employed as wage and salary workers. We exclude self-employed individuals, unpaid family workers, and those in military occupations to focus on the labor supply of the core workforce. All earnings data are deflated to constant 1999 U.S. dollars using the CPI-U-RS. Hourly wages are constructed by dividing real annual earnings by total annual hours, where annual hours are the product of usual weekly hours and the midpoint of the weeks-worked interval. To mitigate the impact of measurement error common in self-reported earnings and hours data, we trim observations with calculated hourly wages below the 1st and above the 99th percentiles. Consequently, the observation counts for hourly wage statistics in Table B1 are approximately 2% lower than for other demographic variables.

Table B1. 2024 Population Descriptive Statistics

Variable	**Mean (SD)**	**N**
Demographics		
Age	39.88 (12.90)	1,279,118
Female (%)	48.70 (50.00)	1,279,118
HS or less (%)	38.20 (48.60)	1,279,118
Some college (%)	21.80 (41.30)	1,279,118
College or more (%)	40.00 (49.00)	1,279,118
Labor Supply		
Usual weekly hours	39.11 (10.99)	1,279,118
Weeks worked (midpoint)	48.04 (9.22)	1,279,118
Earnings (1999$)		
Annual earnings	36,687 (40,481)	1,279,118
Log annual earnings	10.06 (1.08)	1,279,118
Hourly wage[a]	17.72 (14.24)	1,252,503
Log hourly wage[a]	2.62 (0.71)	1,252,503
Observations (unweighted)		1,279,118
Population (weighted)		139,516,356

Notes: Data from the 2024 IPUMS American Community Survey (ACS). Sample includes employed civilian wage-workers aged 16–64. Means are weighted using personal weights (`perwt`). Binary variables (Female, Education) are reported as percentages.
[a] Hourly wage variables are trimmed at the 1st and 99th percentiles to mitigate measurement error, resulting in a lower observation count for these rows.

Full-time-equivalent labor input. Let ℓ index individuals in the ACS, and let UHRSWORK_ℓ denote usual weekly hours worked and WKSWORK_ℓ denote a midpoint for weeks worked last year (constructed from WKSWORK2 bins). We define annual hours as

$$H_\ell = \text{UHRSWORK}_\ell \cdot \text{WKSWORK}_\ell. \tag{B1}$$

We then define the individual's full-time-equivalent labor input as

$$\text{FTE}_\ell = \frac{H_\ell}{H^{FT}}, \qquad H^{FT} = 2080 \text{ hours } (40 \times 52). \tag{B2}$$

Using the IPUMS person weight PERWT_ℓ, occupation-level FTE labor input is

$$\text{FTE}_i = \sum_{\ell \in i} \text{PERWT}_\ell \cdot \text{FTE}_\ell. \tag{B3}$$

After performing the FTEs change, we find that the population $N = 139,516,356$ represents $N_{FTE} = 128,025,669$ FTEs.

B.2. *O*NET Descriptive Statistics*

We use task-level ratings from version 29.0 of the O*NET database, released in August 2024. The analysis draws on the `Task Ratings` table, which reports standardized assessments of task attributes collected from job incumbents and subject-matter experts.

The Task Ratings table covers 879 distinct occupations defined at the O*NET–SOC level. Across these occupations, the data include 17,639 occupation–task combinations, corresponding to 16,580 unique task statements. Tasks are evaluated along three distinct rating scales (importance, relevance, and frequency), spanning seven task categories.

For each occupation–task–scale combination, the table reports a mean data value along with supporting survey statistics, including the number of respondents (N), standard errors, and confidence interval bounds. In total, the table contains 13,604 nonmissing task ratings, reflecting the fact that not all tasks are rated for all occupations or scales. Metadata fields indicate whether a rating is recommended for suppression, the date of data collection, and the source of the assessment.

Appendix C.

C.1. *LLM Specifications for Task Classification*

We use a LLM to classify each O*NET task statement into one of six mutually exclusive task categories: `RC`, `RM`, `NC`, `NM`, `NCI`, and `NMI` as defined by Autor and Thompson (2025), where:

- **Routine Cognitive – RC:** Tasks involving cognitive processes that are codifiable, i.e. that can be fully specified through a set of ordered instructions. These tasks are procedural, structured, repetitive, and rule-based, and often require precision.
- **Routine Manual – RM:** Tasks that involve physical labor that are codifiable, i.e. can be accomplished by following explicit rules. They rely on predictable, consistent operations with clear steps.
- **Non Routine Cognitive – NC:** Non-codifiable tasks that require analytical skills. They are analytic in that they require cognitive capacity such as judgment, strategic thinking, problem-solving, creativity, intuition, visual processing and/or analysis. They are non-codifiable, meaning that the rules for accomplishing the task are not sufficiently well understood to be specified explicitly to be executed by machines.
- **Non Routine Manual – NM:** Tasks that involve physical work which cannot be fully codified due to the need for on-the-spot adaptation, fine motor processing skills and/or situational judgment. These tasks demand manual dexterity and flexibility in response to changing conditions.
- **Non Routine Cognitive Interpersonal – NCI:** Non-codifiable cognitive tasks that fundamentally involve interpersonal interaction, including influence, leadership, persuasion, or strategic engagement with others.
- **Non Routine Manual Interpersonal – NMI:** Non-codifiable manual tasks centered on care, service, or assistance, requiring physical presence, emotional labor, or direct personal interaction.

Classifications are produced at the task level and keyed by their associated `Task_ID`.

C.1.1. Inputs

For each task, the model receives three inputs:

(1) The task identifier (`Task_ID`),
(2) The job title (`Title`) associated with the task,
(3) The task description text (`Task`) as recorded in O*NET.

Tasks are evaluated as they would have been understood prior to the diffusion of personal computing, enterprise software, the internet, and modern robotics using the prompt shown in Appendix C.2. The model is explicitly instructed to classify tasks based on codifiability in principle as of that period and not on subsequent adoption or technological diffusion. The model is instructed to assign exactly one category per task and to return outputs in a strict JSON format consistent with the predefined schema.

C.1.2. Model and inference configuration

All classifications are generated using Google Gemini via the `google.genai` client, with model identifier `gemini-3-flash-preview`. To minimize stochastic variation, inference is performed with a low temperature setting ($T = 0.01$). We enable internal model reasoning using a medium thinking level while constraining outputs to structured JSON by specifying `response_mime_type = application/json` and enforcing a response schema at generation time. The LLM is used solely as a semantic classifier and has no access to external tools, databases, or post-processing logic beyond schema validation.

C.1.3. Output

The model is required to return a JSON, where each object contains exactly two required fields: `task_id` and `classification`. The `classification` field is restricted to the six admissible labels. Formally, each returned item satisfies:

```
{
  "task_id": "<string>",
  "classification": "RC" | "RM" | "NC" | "NM" | "NCI" | "NMI"
}
```

Any response that fails to satisfy this schema is rejected.

The final output is a task-level dataset containing `task_id` and its assigned category label in {`RC`, `RM`, `NC`, `NM`, `NCI`, `NMI`}.

C.2. *LLM Prompt*

```
You will be provided with a task description and the job title (oc-
cupation) it is related to. Carefully read and understand the de-
scription, then classify it into one of the following six categories
based strictly on its characteristics.
==================================================
CATEGORY DEFINITIONS
==================================================

1. Routine Cognitive Tasks (RC)
Definition: Tasks involving cognitive processes that are codifi-
able, i.e. that can be fully specified through a set of ordered
instructions. These tasks are procedural, structured, repetitive,
and rule-based, and often require precision.
Examples: Data entry, basic bookkeeping, standardized clerical work,
repetitive customer service (e.g. bank teller).


2. Routine Manual Tasks (RM)
Definition: Tasks that involve physical labor that are codifiable,
i.e. can be accomplished by following explicit rules. They rely on
predictable, consistent operations with clear steps.
Examples: Picking/sorting, repetitive assembly.


3. Non-Routine Cognitive Analytic Tasks (NC)
Definition: Non-codifiable tasks that require analytical skills.
They are analytic in that they require cognitive capacity such as
judgment, strategic thinking, problem-solving, creativity, intu-
ition, visual processing and/or analysis. They are non-codifiable,
meaning that the rules for accomplishing the task are not suffi-
ciently well understood to be specified explicitly to be executed by
machines.
Examples: Deciphering handwriting on a check, strategic planning,
forming and testing hypotheses, medical diagnoses, legal writing,
complex problem solving, creative design.


4. Non-Routine Manual Tasks (NM)
Definition: Tasks that involve physical work which cannot be fully
codified due to the need for on-the-spot adaptation, fine motor
processing skills and/or situational judgment. These tasks demand
manual dexterity and flexibility in response to changing conditions.
Examples: Skilled craftsmanship, janitorial services, truck driving,
complex repair work.


5. Non-Routine Cognitive Interpersonal Tasks (NCI)
Definition: Non-codifiable cognitive tasks that fundamentally in-
```

volve interpersonal interaction, including influence, leadership, persuasion, or strategic engagement with others.
Examples: Negotiation, team management, consulting, persuading or selling.

6. Non-Routine Manual Interpersonal Tasks (NMI)
Definition: Non-codifiable manual tasks centered on care, service, or assistance, requiring physical presence, emotional labor, or direct personal interaction.
Examples: Caregiving, food service, hospitality, hairstyling, flight attendants.

==========
GUIDANCE FOR BORDERLINE CASES
==========

Use the job title to contextualize the task description when necessary.

==========
INSTRUCTIONS
==========

For the task description provided, carefully analyze its features. Then classify the task as belonging to one of these categories, based on how well the task matches. Output only a single class (e.g. RC, RM, NC, NM, NCI, or NMI).

- Assign exactly one category to each task.
- Do not invent new categories.
- Output a JSON list matching the provided schema exactly.

C.3. *The Distribution of Labor Input Across Classes*

The LLM inference assigns a task-class label to each O*NET task description. Task-level classifications are aggregated to the Census occupation level using task shares, yielding an estimator of the proportion of each task share within an occupation.

Let k represent tasks and i represent Census occupations. Let $\pi_{i,k}$ denote the task share of task k within occupation i, and let $C_k \in \{RC, RM, NC, NM, NCI, NMI\}$ indicate the LLM-assigned class for task k. For each class c, the class-specific task share of occupation i is defined as follows:

$$\nu_{i,c} = \sum_{k \in \mathcal{K}(i)} \pi_{i,k} \mathbb{1}\{C_k = c\}, \tag{C1}$$

where, $\mathcal{K}(i)$ denotes the set of tasks associated with occupation i. In contrast to occupation-level task intensity indices, $\nu_{i,c}$ constitutes a labor-share partition, assigning exactly zero mass to classes with no task share in a class within an occupation.

Class-specific task shares are computed for 66.3% of Dorn occupations (Autor and Dorn 2013) with available task shares, encompassing 80.9% of aggregate employment as measured by labor-share weights (LSWT). Table C1 presents the resulting distribution of labor input across task classes. Labor input is distributed unevenly across task types: cognitive tasks comprise 71.8% of total labor input, whereas manual tasks account for 28.2%.

To assess external validity and evaluate the application of LLMs as synthetic task coders in economic measurement, the LLM-based labor-share measures are compared to occupation-level task intensity indices from the replication kit of Acemoglu and Restrepo (2022), which builds on Acemoglu and Autor (2011) and Autor, Levy, and Murnane (2003). These benchmark indices are derived from occupation-level descriptors intended to capture pre-digital task content. Since the estimator (i) is a labor-share partition that assigns zero mass to classes with no measured allocation and (ii) reflects contemporaneous task execution as encoded in O*NET v29.0, rank- and correlation-based agreement is expected to be lower in sparsely represented categories relative to pre-digital benchmarks.

Figure 1 and Table C1 present the comparison for the matched set of occupations ($N = 216$). Overall agreement is high: across task categories, the weighted Pearson correlation between the LLM-based measures and the benchmark is 0.606, while the unweighted Pearson correlation is 0.547 and the unweighted Spearman rank correlation is 0.575. The higher weighted correlations suggest that alignment is strongest among large occupations, which constitute the majority of aggregate labor input.

Alignment is especially strong for manual task classes and for non-routine cognitive interpersonal work. Routine manual tasks exhibit a weighted correlation of 0.665, non-routine manual tasks 0.757, and non-routine cognitive interpersonal tasks 0.772. Routine cognitive tasks display a lower correlation (weighted $R = 0.450$), which is consistent with two aspects of the comparison. First, benchmark indices are designed to summarize pre-digital occupational descriptors, while the content and organization of cognitive work likely evolved more substantially with the adoption of enterprise software than did many physically grounded manual tasks. Second, the present measures

are derived from task descriptions in O*NET v29.0 and classified within a modern task context. Therefore, to the extent that task content changed between 1980 and 2025, divergence from pre-digital benchmarks is anticipated.

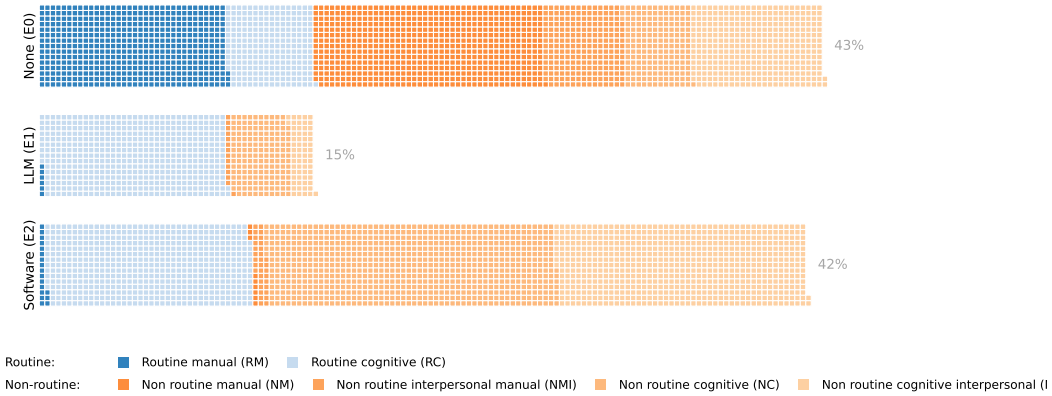
Another source of divergence is mechanical rather than classificatory. Benchmark indices assign continuous intensity scores to nearly all occupations, whereas the LLM-based approach records zero task share when no tasks are measured in the class. Consequently, correlation and rank-based measures are mechanically reduced in classes with a high proportion of zeros, most notably non-routine manual interpersonal (73.6% zeros), which also constitutes a small share of aggregate labor input (3.4%). Overall, these results demonstrate that LLM-based classification aligns closely with established task taxonomies in economically significant classes.

Table C1. Comparison of LLM Classification vs. Acemoglu and Autor (2011)

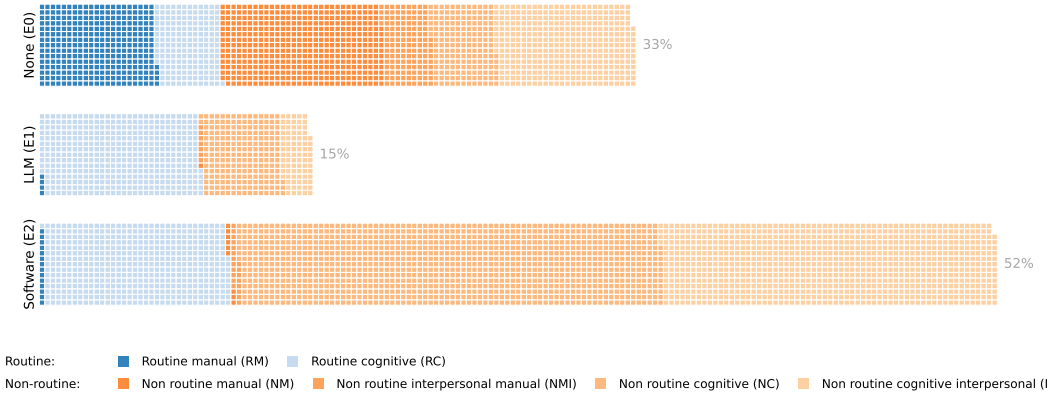
Category	Correlation Metrics			Distribution Metrics	
	Weighted Pearson (R)	**Unweighted Pearson (R)**	**Spearman Rank (ρ)**	**Share of Zeros**	**Global Share (% LSWT)**
Non-Routine Cognitive	0.682	0.572	0.582	8.3%	23.0% (0.28%)
Non-Routine Cognitive Interpersonal	0.772	0.623	0.591	18.5%	21.3% (0.26%)
Routine Cognitive	0.450	0.467	0.419	6.0%	27.5% (0.30%)
Routine Manual	0.665	0.660	0.699	31.9%	13.1% (0.23%)
Non-Routine Manual	0.757	0.666	0.768	28.2%	11.8% (0.22%)
Non-Routine Manual Interpersonal	0.310	0.296	0.390	73.6%	3.4% (0.07%)
Global Average	*0.605*	*0.549*	*0.576*	*27.8%*	–

Notes: $N = 216$ occupations. "Global Share" represents the weighted percentage of the total population (LSWT) falling into each category; values in parentheses denote standard errors reflecting measurement uncertainty from propagated task-level variance estimates ($\text{Var}(\pi)$), not cross-occupation dispersion. Weighted Pearson is calculated using LSWT as weights.

C.4. *Alternative Plot - Routine vs Non-Routine Exposure to LLM Exposure*



(a) Task-FTE composition by GPT-4 exposure for a population of $N = 91{,}923{,}781$ workers (IPUMS 2024) across 8,473 tasks.



(b) Wage-bill composition by GPT-4 exposure for an aggregate wage bill of $\$3.12 \cdot 10^{12}$ (1999 constant USD).

Figure C1. Waffle plots of Task-FTE and wage-bill composition by GPT-4 exposure category. Color gradients are based on the two broader task categories routine and non-routine. *Notes:* Each waffle represents 0.02% of the total. Shares are computed over the analysis sample (the subset of the 2024 workforce with valid task profiles).

C.5. *Detailed LLM Exposure Results*

Table C2. Task Exposure by GPT-4 Category

	Non-Routine				Routine		Share of total (%)
	NC	NCI	NM	NMI	RC	RM	
Panel A. Aggregate Task-FTEs (millions)							
E0	3.071	6.026	10.525	3.698	3.982	8.575	42.9 (0.3)
	(0.075)	(0.134)	(0.168)	(0.079)	(0.085)	(0.165)	
E1	2.708	1.048	0.009	0.232	8.462	0.098	15.0 (0.2)
	(0.081)	(0.043)	(0.002)	(0.021)	(0.140)	(0.008)	
E2	13.250	11.439	0.146	0.555	9.429	0.302	42.0 (0.3)
	(0.218)	(0.229)	(0.007)	(0.020)	(0.253)	(0.029)	
Share of total (%)	22.8 (0.3)	22.2 (0.3)	12.8 (0.2)	5.4 (0.1)	26.2 (0.3)	10.7 (0.2)	100.0 (0.0)
Panel B. Aggregate Wage Bill (billions of 1999 USD)							
E0	112.1 (3.4)	237.5 (6.7)	286.8 (4.1)	92.5 (2.6)	115.1 (2.5)	210.7 (4.2)	33.6 (0.3)
E1	133.9 (4.8)	50.2 (2.8)	0.3 (0.1)	5.1 (0.4)	281.3 (5.1)	2.7 (0.2)	15.1 (0.2)
E2	699.0 (13.8)	553.9 (12.1)	5.5 (0.3)	14.2 (0.5)	324.3 (8.3)	9.3 (0.8)	51.2 (0.4)
Share of total (%)	30.1 (0.4)	26.9 (0.4)	9.3 (0.1)	3.6 (0.1)	23.0 (0.3)	7.1 (0.1)	100.0 (0.0)

Notes: NC, NCI, NM, and NMI denote non-routine task categories, while RC and RM denote routine task categories. GPT-4 exposure categories (E0–E2) follow Eloundou et al. (2023). Entries report aggregate Task-FTEs (Panel A, millions) and aggregate wage bills (Panel B, billions of 1999 USD); standard errors are in parentheses. The *Share of total* column reports the row sum as a percentage of the panel total; the *Share of total* row reports the column sum as a percentage of the panel total. Share standard errors are computed via a delta-method approximation under the assumption that cell estimates are independent.

Appendix D.

D.1. *Solver Settings and MOSEK Parameters*

We implement the semidefinite programming (SDP) estimation using the MOSEK conic interior-point optimizer. To guarantee that the resulting covariance matrices are mathematically robust and strictly positive semi-definite, we enforce a strict tolerance profile. Specifically, we reject inaccurate solutions and minimize the Frobenius norm using the following configuration:

`solver_order` = (MOSEK)
`optimiser_norm` = Frobenius
`accept_inaccurate` = **False**

Table D1. MOSEK/IPM configuration used in all main experiments.

Parameter	Value	Purpose
`MSK_DPAR_INTPNT_CO_TOL_REL_GAP`	10^{-10}	Relative duality gap tolerance
`MSK_DPAR_INTPNT_CO_TOL_PFEAS`	10^{-10}	Primal feasibility tolerance
`MSK_DPAR_INTPNT_CO_TOL_DFEAS`	10^{-10}	Dual feasibility tolerance
`MSK_DPAR_INTPNT_CO_TOL_MU_RED`	10^{-10}	Barrier reduction (mu) tolerance
`MSK_IPAR_NUM_THREADS`	5	Parallelism
`MSK_IPAR_PRESOLVE_USE`	1	Enable presolve
`MSK_IPAR_INTPNT_SCALING`	1	Enable scaling

D.2. *Covariance Model Robustness*

In this section, different robustness checks are applied to evaluate the numerical feasibility and compositional integrity of each covariance matrix $\Sigma_{i,k}$, the covariance matrix of categories of task k in occupation i.

D.2.1. Feasibility and Simplex Geometry

For each matrix $\Sigma_{i,k}$, we verify that (i) the matrix is positive semidefinite (PSD), (ii) the simplex geometry implied by task shares is maintained, and (iii) the completed variances align with observed survey marginals.

Diagnostics definitions. Let $\mathbf{1}$ be an all-ones vector of length $n = 8$ (frequency categories), and $v = (\mathrm{se}_p)^2$ the vector of design-based marginal variances. For each completed covariance Σ, the following metrics are computed:

$$\text{Tangent residual:} \quad r_\infty = \max\big(\|\Sigma\mathbf{1}\|_\infty, \|\mathbf{1}^\top\Sigma\|_\infty\big), \tag{D1}$$

$$\text{PSD check:} \quad \lambda_{\min}(\Sigma) \geq -10^{-8}, \tag{D2}$$

$$\text{Diagonal deviation:} \quad d_\infty = \|\mathrm{diag}(\Sigma) - v\|_\infty, \tag{D3}$$

$$\text{Uncertainty magnitude:} \quad \mathrm{trace}(\Sigma), \tag{D4}$$

$$\text{Effective rank:} \quad r_{\text{eff}} = \exp\Big(-\sum_j p_j \log p_j\Big), \qquad p_j = \lambda_j / \sum_\ell \lambda_\ell, \tag{D5}$$

In this context, λ_j denotes the eigenvalues of Σ, which are clipped at zero to enforce positive semidefiniteness. The entropy-based effective rank $r_{\text{eff}} \in [0, 7]$ quantifies the dimensionality of task-level uncertainty.

Results and discussion. Table D2 summarizes the convergence diagnostics for the full sample of 12,495 tasks. The optimization successfully recovered a mathematically valid covariance structure for 100% of tasks. The tangent residual ($\approx 4 \times 10^{-11}$) is near machine precision, confirming that the estimated matrices reside strictly on the tangent space of the simplex. This ensures that the structural constraint–that probabilities must sum to one–is preserved perfectly in the variance domain. The diagonal deviation reflects the geometric tension between the stochastic survey estimates and the strict requirement that a covariance matrix be positive semi-definite. The maximum observed deviation (9.01×10^{-6}) corresponds to a standard error approximation discrepancy of roughly $\sqrt{10^{-6}} \approx 0.001$. This magnitude is negligible relative to standard survey sampling error, indicating that the estimator maintains high fidelity to the input data while correcting for geometric infeasibilities. Finally, the effective rank (mean = 3.76) highlights the structure of survey uncertainty. While the probability space has 8 dimensions, the uncertainty typically occupies a subspace of fewer than 4 dimensions. This suggests that measurement error is not random white noise distributed across all categories, but is instead structured reflecting respondent selecting only a few frequency bins (e.g. distinguishing "Weekly" from "Monthly"). This supports the hypothesis that respondent assessments are coherent, with disagreement constrained to specific, localized shifts rather than broad divergence.

Table D2. Covariance Completion Feasibility Diagnostics

Metric	All Tasks ($N = 12{,}495$)	
	Value	Tasks Removed
PSD Percentage	100 %	0
Max tangent residual	4.29×10^{-11}	0
Max diagonal deviation	9.01×10^{-06}	0
Trace (mean [SD])	0.021 (0.009)	n/a
Effective rank (mean [SD])	3.76 (1.02)	n/a

D.2.2. Functional Variance Bounds and Position

The variance of an annual-occurrence functional is benchmarked against its feasible range for each task.

Diagnostics definitions. Let $w \in \mathbb{R}^8$ denote the fixed annual-occurrence weights over the eight O*NET frequency categories, let $v = \mathrm{Var}(p_{i,k})$ be the vector of design-based marginal variances, and let $\hat{\Sigma}$ denote the completed task-level covariance matrix. The realized functional variance associated with task (i, k) is given by $w^\top \hat{\Sigma}\, w$.

To assess whether this value is geometrically plausible given the marginal variances and the simplex constraint, we compute the feasible lower and upper bounds through the following convex programs, which impose the same constraints used during covariance completion—positive semidefiniteness, fixed diagonal entries, and tangency to the simplex:

$$\underline{V} = \min_{\Sigma \succeq 0} \; w^\top \Sigma\, w \quad \text{s.t. } \mathrm{diag}(\Sigma) = v,\; \Sigma \mathbf{1} = 0, \tag{D6}$$

$$\overline{V} = \max_{\Sigma \succeq 0} \; w^\top \Sigma\, w \quad \text{s.t. } \mathrm{diag}(\Sigma) = v,\; \Sigma \mathbf{1} = 0. \tag{D7}$$

The relative position of the completed functional variance within its task-specific feasible interval is then summarized by the normalized statistic

$$\mathrm{pos} \;=\; \frac{w^\top \hat{\Sigma}\, w \;-\; \underline{V}}{\overline{V} \;-\; \underline{V}} \;\in\; [0, 1]. \tag{D8}$$

Tasks whose normalized position lies near the boundaries of the feasible set–specifically, $\mathrm{pos} < 0.05$ or $\mathrm{pos} > 0.95$–are flagged as extreme and removed from downstream analysis. This robustness filter excludes covariance structures that, while feasible, imply implausibly tight or excessively diffuse uncertainty relative to the geometric constraints of the probability simplex.

Results and discussion. Table D3 summarizes the position of the estimated functional variances relative to their theoretically feasible bounds. Across the full sample

of 12,495 tasks, the mean normalized position is 0.39 (SD = 0.13). This statistic indicates that for the typical task, the covariance structure recovered by the minimum Frobenius norm objective lies near the center of the feasible set, rather than clustering at the extremes. This centrality suggests that the estimator is generally "conservative," avoiding the imposition of maximal or minimal correlations unless driven by the data. However, the analysis identifies a subset of 670 tasks (5.36%) where the normalized position falls outside the $[0.05, 0.95]$ interval. These boundary solutions represent "degenerate" cases where the observed marginal variances are so restrictive that they force the functional variance to the geometric limits of the probability simplex (e.g. requiring perfect correlation or extreme anti-correlation to satisfy $\Sigma\mathbf{1} = 0$). Because such boundary solutions often imply unstable or implausible uncertainty structures driven by sparse survey data, we exclude these tasks to ensure that downstream inference is robust to numerical edge cases.

Table D3. Variance Position and Outlier Summary

	All Tasks (N = 12,495)	
Metric	Value	Tasks Removed
Mean normalized position	0.39 (0.13)	n/a
Outlier (position < 0.05 / > 0.95)	5.36%	670

D.2.3. Sensitivity To Measurement Errors

We assess the stability of task-level functional variances to modest perturbations of the marginal targets.

Diagnostics definitions. Let $v = (\mathrm{se}_p)^2$ denote the design-based marginal variances for the eight frequency categories and $w \in \mathbb{R}^8$ the annual-occurrence weights. For each task, with completed covariance $\hat{\Sigma}$, the baseline functional variance is

$$V_0 \;=\; w^\top \hat{\Sigma}\, w. \tag{D9}$$

We form perturbed diagonal targets $v^{(+)} = (1+\tau)\, v$ and $v^{(-)} = (1-\tau)\, v$ with $\tau = 0.05$, and recompute two completions by solving the same SDP used at baseline (PSD, fixed diagonal, simplex tangency $\Sigma\mathbf{1} = 0$). Let $V_+ = w^\top \hat{\Sigma}^{(+)} w$ and $V_- = w^\top \hat{\Sigma}^{(-)} w$. We then compute

$$\Delta_+ = V_+ - V_0, \qquad \Delta_- = V_- - V_0, \qquad \delta_\pm^{\mathrm{rel}} = \frac{|\Delta_\pm|}{\max(10^{-9},\, |V_0|)}. \tag{D10}$$

To diagnose linearity and symmetry, we use the symmetrized relative change

$$\mathrm{symm_rel} \;=\; \frac{\left|\Delta_+ + \Delta_-\right|}{\max(10^{-9},\, \max\{|\Delta_+|, |\Delta_-|\})}, \tag{D11}$$

Tasks are flagged if (i) symm_rel > 0.1 (nonlinear/asymmetric response), or if (ii) the directional relative changes leave the expected linear band $\delta_{\pm}^{\text{rel}} \notin [0.9\tau, 1.1\tau] = [0.045, 0.055]$.

Results and discussion. Table D4 summarizes the stability of the functional variance estimator under $\pm 5\%$ perturbations of the input marginals.

The estimator exhibits high local stability for the vast majority of the sample. The mean symmetrized relative change is one order of magnitude lower than the perturbation (2.80×10^{-3} against $\tau = 5\%$ (5×10^{-2}). This suggests that the estimator's response is dominated by a stable linear pass-through, with non-linear residuals remaining negligible for the typical task.

However, we detect numerical instability in a small subset of the data. Approximately 0.38% of tasks exhibit high nonlinearity (symm_rel > 0.1), and a similar fraction fail the directional bounds test, where the output variance shifts by more (or less) than the expected $\approx 5\%$ linear pass-through. In total, 52 unique tasks (0.42%) are flagged as sensitive. These cases likely correspond to covariance structures that are "brittle"–statistically feasible but poised on a manifold where small changes in constraints force large reconfigurations of the off-diagonal elements. To ensure robustness, these unstable tasks are excluded from the final analytical sample.

Table D4. Sensitivity Analysis (for $\pm 5\%$ perturbation)

	All Tasks (N = 12,495)	
Metric	Value	Tasks Removed
Symmetric relative change	$2.80 \cdot 10^{-3}$ ($1.93 \cdot 10^{-2}$)	n/a
High sensitivity ($> 10\%$)	0.38%	48
Δ rel. (minus) out of range	0.19%	24
Δ rel. (plus) out of range	0.20%	25
Total unique tasks affected	0.42%	52

D.3. *Weighting Specifications Robustness*

To map the ordinal frequency scales from O*NET to cardinal annualized occurrences, we employ three distinct weighting vectors ($\mathbf{w}$). These specifications test the sensitivity of our results to assumptions regarding the work calendar (e.g. 260 vs. 365 days) and the relative weight of high-frequency tasks. Table D5 details the mapping for each of the $R = 8$ possible frequency categories in our estimation framework.

- **Baseline ($\mathbf{w_{base}}$):** Assumes a standard work year (52 weeks, 5 days/week, 8 hours/day). The maximum frequency (Index 7) corresponds to continuous hourly execution ($52 \times 40 = 2080$).
- **Aggressive ($\mathbf{w_{agg}}$):** Assumes a continuous flow or "24/7" availability model (365 days). This tests the upper bound of task volume.
- **Compressed ($\mathbf{w_{comp}}$):** Applies a non-linear transformation that significantly dampens the ratio between high-frequency and low-frequency tasks. By reducing the max-to-min ratio from 2080:1 to 100:1, this specification tests whether the identification of task intensity is driven solely by the dominance of the top frequency category.

Table D5. Cardinalization of Frequency Categories under Alternative Specifications

Index	Frequency	Baseline ($\mathbf{w}_{\text{base}}$)	Aggressive ($\mathbf{w}_{\text{agg}}$)	Compressed ($\mathbf{w}_{\text{comp}}$)
0	Never	0	0	0
1	Yearly	1	1	1
2	Quarterly	4	6	2
3	Monthly	24	52	5
4	Weekly	104	156	10
5	Daily	260	365	25
6	Multiple Daily	780	1,095	50
7	Hourly	2,080	2,920	100
Max-to-Min Ratio		2080:1	2920:1	100:1

Notes: The Baseline vector assumes a standard 260-day work year. The Aggressive vector assumes a 365-day year. The Compressed vector imposes an arbitrary log-like scale to test robustness to heavy-tailed distributions.

Appendix E.

E.1. *Crosswalks*

Table E1. Sample Construction and Occupational Mapping Coverage

Step	Occupation codes	# occ.	# tasks	% Labor Input	% Labor Cost
(1) Baseline	O*NET SOC	879	17,639	n/a	n/a
(2) Filter valid profiles	O*NET SOC	663	12,495	n/a	n/a
(3) Covariance completion	O*NET SOC	661	11,775	n/a	n/a
(4) Mapping SOC to Census	Census 2010	290	8,492	65.3	60.8
(5) Census 2010 to Dorn 1990	Dorn 1990	218	8,492	45.7	41.5

Notes: Rows summarize the progressive restriction of O*NET task profiles and subsequent crosswalks from O*NET SOC to Census 2010 occupation codes.

- The analysis starts from the O*NET 29.0 `Task Ratings` data, comprising 879 unique O*NET-SOC occupations and 17,639 unique task statements.
- We restrict attention to statistically informative task profiles by dropping tasks that lack incumbent responses or yield non-significant/incalculable estimates. This results in 663 well-defined occupational profiles spanning 12,495 tasks.
- We then characterize within-occupation task-share objects and apply our covariance-completion and conditioning checks, removing ill-conditioned profiles. This yields 661 occupations characterized by 11,775 tasks.
- Next, we map O*NET-SOC occupations to the Census 2010 occupational taxonomy. Tasks are grouped to Census occupations and the resulting task shares are renormalized within each Census code. The resulting sample covers 65.3% of aggregate labor input (FTEs) and 60.8% of aggregate labor cost (earnings).
- For comparability with benchmarks that use the Dorn (1990) classification, we further crosswalk Census 2010 occupations to Dorn 1990 codes. This yields 218 Dorn occupations, while leaving the task universe unchanged at 8,492 tasks, and captures 45.7% of labor input and 41.5% of labor cost.

Appendix F.

F.1. *Task Bundling and Specialization*

This appendix defines the task concentration measures used in the main text and reports additional descriptive statistics.

F.1.1. Core task share (Top-n share).

Let $\mathcal{K}_{i,\text{top}n}$ denote the set of the n tasks with the highest task shares in occupation i, where tasks are ranked in descending order of $\hat{\pi}_{i,k}$. The Top-n share is defined as

$$C_{i,n} = \sum_{k \in \mathcal{K}_{i,\text{top}n}} \hat{\pi}_{i,k}. \tag{F1}$$

This measure captures the extent to which an occupation's labor input is concentrated in a small core set of tasks.

F.1.2. Task Gini and normalized Gini.

We measure inequality in labor input allocation using an occupation-level Gini coefficient, defined as the relative mean absolute difference of task shares:

$$G_i \;=\; \frac{1}{2K_i} \sum_{k=1}^{K_i} \sum_{j=1}^{K_i} \left|\pi_{i,k} - \pi_{i,j}\right|, \tag{F2}$$

where K_i is the number of distinct tasks in occupation i and $\pi_{i,k}$ denotes the share of total labor input allocated to task k, with $\sum_{k=1}^{K_i} \pi_{i,k} = 1$. This statistic captures the degree of concentration of labor input across tasks within an occupation.

Because the maximum attainable Gini mechanically depends on the number of tasks–reaching $G_i^{\max} = (K_i - 1)/K_i$ under complete task concentration–we report a normalized Gini index:

$$\widetilde{G}_i \;\equiv\; \frac{G_i}{(K_i - 1)/K_i}. \tag{F3}$$

Substituting the expression for G_i and using the probability normalization yields the simplified closed-form:

$$\widetilde{G}_i \;=\; \frac{\sum_{k=1}^{K_i} \sum_{j=1}^{K_i} \left|\pi_{i,k} - \pi_{i,j}\right|}{2\,(K_i - 1)}. \tag{F4}$$

The normalized Gini satisfies $\widetilde{G}_i \in [0, 1]$ for all occupations with $K_i > 1$, mapping a uniform task allocation to 0 and a single-task bottleneck to 1, and is therefore directly comparable across occupations with differing task counts. Higher values of $\widetilde{G}_i$ indicate greater task specialization, while lower values reflect more even multitasking across task content.

F.1.3. Work Content

We model work content as the aggregate labor input allocated to specific functional categories, following the taxonomy of Autor, Levy, and Murnane (2003). Specifically, the "Routine" content of an occupation is measured as the sum of task shares classified as either routine-manual or routine-cognitive:

$$\text{Rout}_i = \sum_{k \in \mathcal{R}_i} \hat{\pi}_{i,k}, \tag{F5}$$

where $\mathcal{R}_i$ is the set of tasks classified as routine in occupation i, and $\hat{\pi}_{i,k}$ is the estimated task share for task k in occupation i. Rout_i is strictly bounded between 0 and 1 and represents the total share of occupational labor input dedicated to routine work.

F.1.4. Task Breadth

To distinguish between the composition of work and its variety, we define task breadth as the total number of distinct tasks performed within an occupation:

$$\text{B}_i = \ln |\mathcal{K}_i|, \tag{F6}$$

where $\mathcal{K}_i$ is the set of all tasks mapped to occupation i.

F.2. *Specialization descriptive statistics*

The statistics in Table F1 reinforce the main-text point that occupations typically have a small set of dominant tasks. The median top-3 share is 0.324, while the top-10 share is 0.786 at the median, indicating that a modest subset of tasks accounts for most labor input. The raw index G_i mechanically decline with the number of observed tasks, which varies substantially across Census occupations; the normalized indices address this comparability issue while preserving substantial dispersion across occupations.

Table F1. Appendix: Descriptive Statistics of Task Concentration Metrics ($J = 329$)

Metric	**Mean**	**Std. Dev.**	**Min**	**P25**	**Median**	**P75**	**Max**
Tasks per Job (K_i)	24.28	35.34	3.00	14.00	18.00	25.00	589.00
Top 1 Share ($C_{i,1}$)	0.146	0.082	0.008	0.097	0.121	0.182	0.684
Top 3 Share ($C_{i,3}$)	0.357	0.152	0.021	0.256	0.324	0.428	1.000
Top 10 Share ($C_{i,10}$)	0.765	0.189	0.061	0.648	0.786	0.918	1.000
Gini Coefficient (G_i)	0.350	0.090	0.076	0.288	0.352	0.412	0.579
Normalized Gini ($\widetilde{G}_i$)	0.372	0.095	0.089	0.307	0.372	0.438	0.631

Notes: Metrics are computed from within-occupation task shares $\{\hat{\pi}_{i,k}\}_{k=1}^{K_i}$.

Appendix G.

G.1. *Marginal Effect Results*

Table G1. Implied Marginal Effect of Occupational Focus at Different Routine Intensities

Rout^Z	$\partial \ln w / \partial \text{Spec}^Z$	SE	p-value
-2	0.1014	0.0276	0.000
-1	0.0820	0.0195	0.000
0	0.0626	0.0137	0.000
1	0.0432	0.0138	0.002
2	0.0238	0.0197	0.226

Notes: Marginal effects are computed from the interaction model (Section 5.2) using $\partial \ln w / \partial \text{Spec}^Z = \hat{\beta}_S + \hat{\beta}_{SR}\text{Rout}^Z$. Standard errors are computed using the delta method based on the estimated covariance matrix of $(\hat{\beta}_S, \hat{\beta}_{SR})$.

G.2. *Robustness – Annual Earnings*

Table G2. Occupation Specialization and Annual Earnings

	(1)	(2)	(3)	(4)
Specialization	0.0868***	0.0857***	0.0620***	0.0639***
	(0.0160)	(0.0167)	(0.0138)	(0.0140)
Volume		0.00552	0.00481	0.00429
		(0.0349)	(0.0302)	(0.0305)
Routine			-0.111***	-0.105***
			(0.0143)	(0.0147)
Specialization × Routine				-0.0195*
				(0.0100)
Control variables	✓	✓	✓	✓
R^2	0.647	0.647	0.655	0.655
Clusters (occupations)	290	290	290	290
N (workers)	838,747	838,747	838,747	838,747

Notes: The dependent variable is log annual earnings in 1999 CPI-adjusted dollars. Occupational specialization is the standardized normalized task Gini coefficient. Routine intensity is the standardized share of labor input devoted to routine tasks. Task volume is the log number of distinct tasks mapped to occupation j. All specifications include controls for gender, age, age squared, education, industry fixed effects, log usual hours, and weeks worked fixed effects, and are estimated using IPUMS person weights. Robust standard errors clustered at the occupation level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

In Table G2, we re-examine the returns to specialization using log annual earnings as the dependent variable. This robustness check serves a dual purpose. First, it mitigates potential measurement error inherent in derived hourly wages. Second, and more critically, it allows us to distinguish between the price of specialized labor and the quantity of labor supplied—or, distinctly, the difference between the total return to a job and the "cost of work" (effort and hours) required to sustain it.

A potential concern with the baseline hourly wage results is that specialized occupations might simply be more demanding, requiring workers to supply significantly more hours (the intensive margin) to achieve proficiency. If the specialization premium were entirely explained by longer work weeks, it would represent a compensation for the disutility of labor—a higher "cost of work"—rather than a productivity premium.

Column (1) documents a robust positive association between occupational specialization and annual earnings ($\beta = 0.0868$, $p < 0.01$). Crucially, because all specifications control for log usual hours and weeks worked, this coefficient captures the shift in earnings net of labor supply. This implies that the specialization premium is not an artifact of specialized workers simply working longer hours; rather, earnings are systematically higher in occupations with more concentrated task structures, even conditional on labor supply. Columns (2) through (4) reinforce that this price premium is distinct from other task-based attributes.

In Column (2), we introduce Task Volume to test the hypothesis that specialization is merely a proxy for job simplicity (i.e. doing fewer things). The coefficient on Task Volume is statistically indistinguishable from zero ($\beta = 0.00552$), while the specialization coefficient remains stable at 0.0857. This result is striking: it suggests that the labor market does not penalize "narrow" jobs for their lack of breadth, nor does

it reward "complex" jobs (high task volume) simply for their multifariousness. The premium is driven by the concentration of effort, not the raw count of duties.

Column (3) conditions on Routine Intensity. While the magnitude of the specialization coefficient attenuates (from 0.0857 to 0.0620), it remains economically significant and precisely estimated ($p < 0.01$). Furthermore, Routine Intensity itself is associated with a significant decrease in annual earnings ($\beta = -0.111$, $p < 0.01$).

Column (4) introduces an interaction between specialization and routine intensity to explore whether the returns to concentration are moderated by the nature of the tasks performed. We find a modest negative interaction ($\beta = -0.0195$), significant at the 10% level. Comparison with the baseline specifications reveals a consistent stability in the specialization premium across various controls. Collectively, these results suggest that the specialization premium is consistent with higher returns to concentrated task structures, rather than compensation for increased labor supply or a mechanical reflection of routine task content.

G.3. *Robustness – Full-Time Full-Year Workers*

In Table G3, we restrict the sample to full-time, full-year (FTFY) workers—defined as those who worked at least 50 weeks and usually worked at least 35 hours per week. This restriction reduces the sample size from 838,747 to 625,822 but ensures that our estimates are not driven by part-time labor, gig work, or workers with tenuous labor market attachment.

Table G3. Wage Associations with Occupational Specialization (Full-Time Full-Year Workers)

	(1)	(2)	(3)	(4)
Specialization	0.0913***	0.0906***	0.0629***	0.0633***
	(0.0151)	(0.0155)	(0.0129)	(0.0131)
Volume		0.00309	0.00592	0.00617
		(0.0342)	(0.0289)	(0.0291)
Routine			-0.111***	-0.105***
			(0.0142)	(0.0147)
Specialization × Routine				-0.0172*
				(0.00991)
Control variables	✓	✓	✓	✓
R^2	0.347	0.347	0.366	0.366
Clusters (occupations)	290	290	290	290
N (workers)	625,822	625,822	625,822	625,822

Notes: The dependent variable is log hourly wages in 1999 CPI-adjusted dollars. The sample is restricted to full-time full-year workers (at least 50 weeks worked and 35 usual weekly hours). Occupational specialization is the standardized normalized task Gini coefficient. Routine intensity is the standardized share of labor input devoted to routine tasks. Task volume is the log number of distinct tasks mapped to occupation j. All specifications include controls for gender, age, age squared, education, and industry fixed effects, and are estimated using IPUMS person weights. Robust standard errors clustered at the occupation level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

The results strongly reinforce the baseline findings. In Column (1), the coefficient on specialization is 0.0913, indicating that the returns to occupational focus remain robust among core labor market participants. This effectively rules out the concern that the specialization premium is an artifact of high-wage consulting or irregular employment arrangements.

Columns (2) through (4) follow the familiar pattern: Task Volume remains statistically insignificant, and the introduction of Routine Intensity in Column (3) attenuates the specialization coefficient to 0.0629 but does not eliminate it. Notably, the interaction term in Column (4) is -0.0172 and is statistically significant at the 10% level, suggesting a slight moderation of the specialization premium in highly routine occupations. Overall, these results support the conclusion that the association between occupational specialization and wages is robust across standard employment relationships.

G.4. *Conditional Partial Regressions*

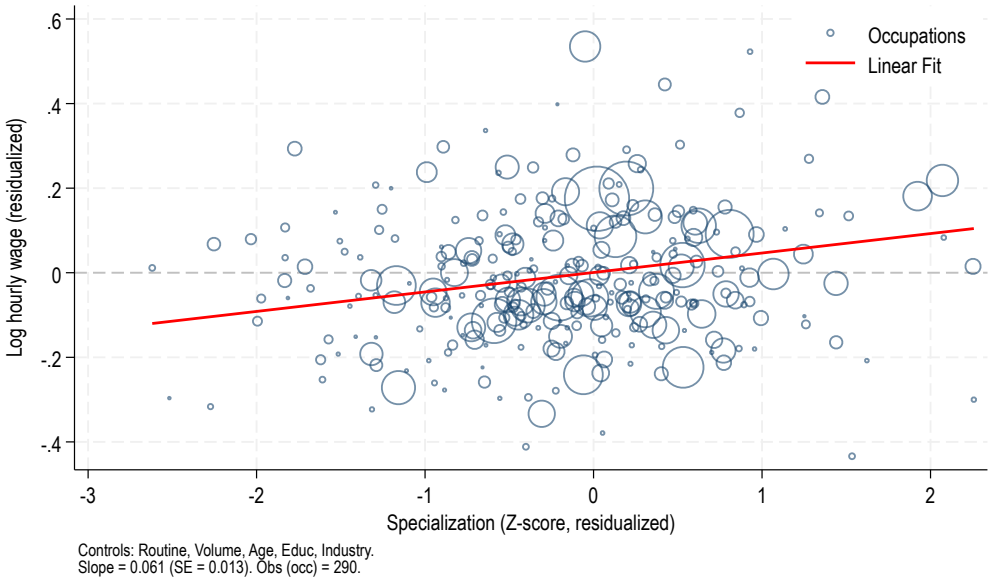


Figure G1. Conditional Partial Regression of Log Hourly Wages on Occupational Specialization.

Figure G1 presents a conditional partial regression plot of log hourly wages on occupational specialization (Spec_Z), net of routine intensity, breadth, age, education, gender, and industry. Each circle represents one of the 290 Census occupations, with marker size proportional to its employment share. The solid red line shows the weighted least-squares linear fit, yielding a statistically significant positive slope of 0.061 ($s.e. = 0.013$), consistent with a wage premium for jobs with more concentrated task structures.

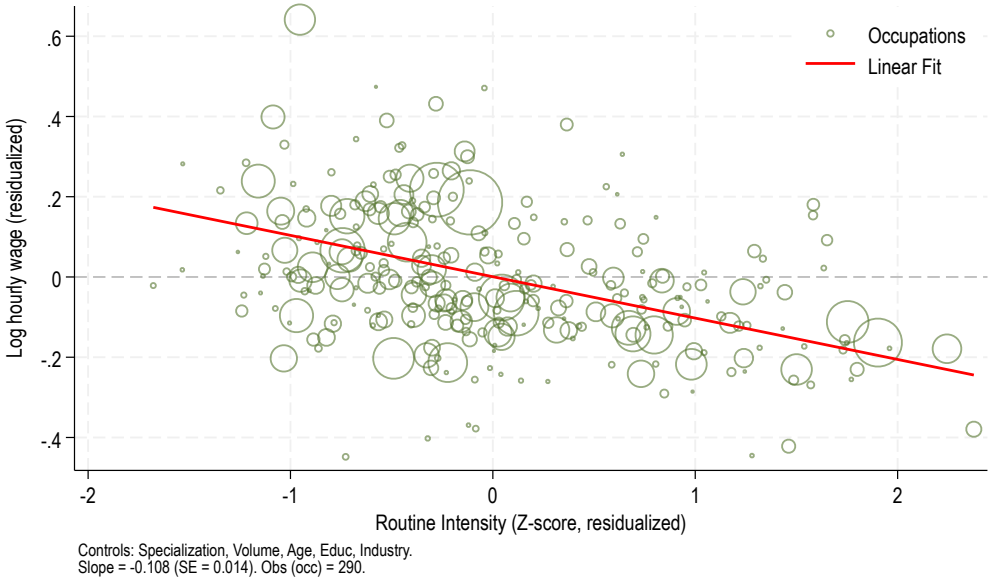


Figure G2. Conditional Partial Regression of Log Hourly Wages on Routine Intensity.

Figure G2 presents a conditional partial regression plot of log hourly wages on routine intensity (Routine_Z), net of occupational specialization, task volume, and worker demographics. Each circle represents one of the 290 Census occupations, with marker size proportional to its employment share. The fitted line exhibits a negative slope of -0.108 ($s.e. = 0.014$), indicating that, holding specialization and other factors constant, a one-standard-deviation increase in an occupation's routine task share is associated with approximately a 10.8% reduction in hourly wages.

G.5. *Correlation of Different Intensive Margin Proxies*

Figure G3 and Figure G4 report pairwise correlations across intensive-margin proxy schemes. Across both correlation measures, all correlations are positive, indicating that the different schemes tend to rank occupations in the same direction on average. However, the strength of alignment varies by proxy pair. The correlation between $\pi_{i,k}$ and the rating-based schemes is moderate—approximately 0.48–0.56 with importance and 0.40–0.46 with relevance—while its correlation with the core-task indicator ("coreweight") is weaker, around 0.31–0.38.

Correlations among the rating/heuristic schemes are higher. In particular, relevance and coreweight are strongly correlated (about 0.71–0.81). This pattern is consistent with the O*NET task classification rules: coreweight is a binary indicator equal to 1 if and only if a task is classified as Core, which requires (i) relevance $\geq 67\%$ and (ii) mean importance ≥ 3.0 on the 1–5 scale. A task is classified as Supplemental otherwise.

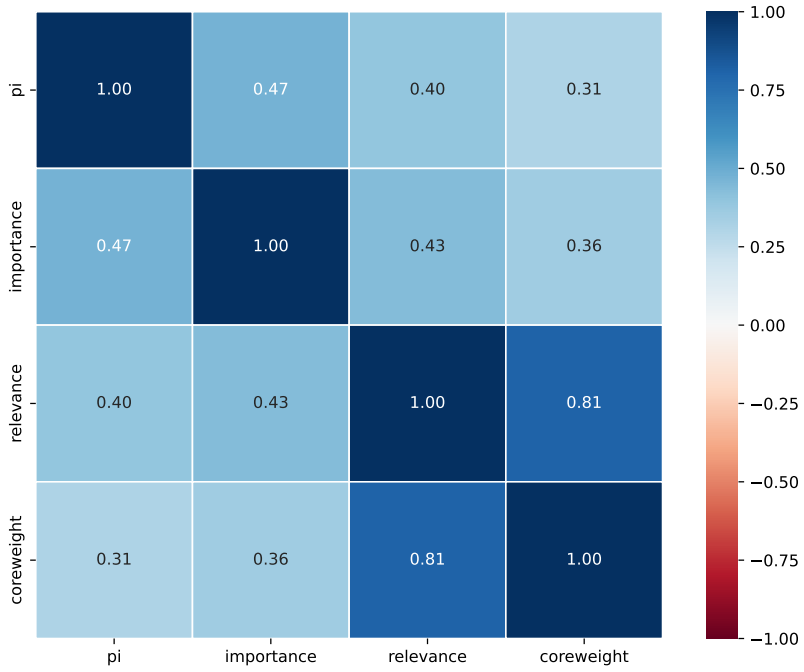


Figure G3. Pearson Correlation

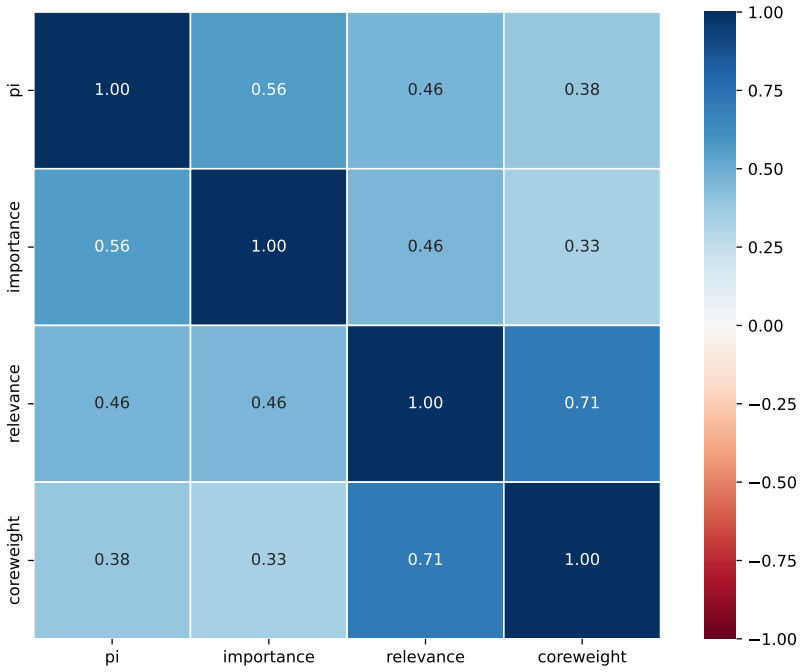


Figure G4. Spearman (Rank) Correlation

G.6. *Marginal Effect of Specialization for Different Intensive Margin Proxies*

Table G4. Marginal effect of specialization on log hourly wage across routine intensity, by intensive-margin proxy

Routine intensity R^Z	Task-share (π)	Importance	Relevance	Core–supp. prior (2:1)
-2	0.101***	-0.049	-0.034	-0.049*
	(0.028)	(0.035)	(0.030)	(0.027)
-1	0.082***	-0.045*	-0.030	-0.036*
	(0.019)	(0.024)	(0.020)	(0.019)
0	0.063***	-0.041**	-0.026*	-0.023
	(0.014)	(0.018)	(0.016)	(0.015)
1	0.043***	-0.038*	-0.023	-0.011
	(0.014)	(0.022)	(0.021)	(0.020)
2	0.024	-0.034	-0.019	0.002
	(0.020)	(0.031)	(0.032)	(0.029)

Notes: Entries are average marginal effects from `margins, dydx(Spec_Z) at(Routine_Z = r)` evaluated at $r \in \{-2, -1, 0, 1, 2\}$ for the fully specified wage model. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Uniform task weights are excluded because uniform bundles imply a constant within-occupation concentration measure (normalized Gini = 0), making specialization mechanically uninformative.