Establishment Size and the Task Content of Jobs: Evidence from 46 Countries*

Micole De Vera[†]
Stone Centre at UCL

Javier Garcia-Brazales[‡]
University of Exeter and LEAP (Bocconi)

July 19, 2024

Abstract

Using a mix of household- and employer-based survey data from 46 countries, we provide novel evidence that workers in larger establishments perform more non-routine analytical tasks, even within narrowly defined occupations. Moreover, workers in larger establishments rely more on the use of information and communication technologies (ICT) to perform these tasks. We also document a 15% raw wage premium that workers in larger establishments enjoy relative to their counterparts in smaller establishments. A mediation analysis shows that our novel empirical facts on the task content of jobs are able to explain between 5–20% of the establishment size wage premium, a similar fraction to what can be explained by selection of workers on education, gender, and age.

JEL codes: J24, J31, L25

Keywords: tasks, occupations, establishment size, cross-country evidence, wage differential

*We thank Manuel Arellano, Jaime Arellano-Bover, and Samuel Bentolila for valuable comments and suggestions. We also thank Andrés Carmona for his research assistance. The data used in this article can be obtained from the websites of the OECD's Survey of Adult Skills (https://www.oecd.org/skills/piaac/) and of the World Bank's STEP Skills Measurement Program (https://microdata.worldbank.org/index.php/collections/step).

[†]micole.devera@ucl.ac.uk.

[‡]Corresponding author: j.garcia-brazales2@exeter.ac.uk

1 Introduction

Empirical studies show that the task composition of jobs can explain a large share of the dispersion of wages across occupations and time (Autor et al., 2003; Acemoglu and Autor, 2011; Autor and Handel, 2013; Goos et al., 2014; Acemoglu and Restrepo, 2022). Most studies are, however, constrained by the fact that direct measures of the task content of jobs are not available in standard data sets. To overcome this limitation, authors typically resort to imputing the task content of occupations by means of alternative data sets such as the Occupation Information Network (O*NET) or the European Working Conditions Survey (EWCS). The use of these general task classifications relies on the implied assumption that the task composition of jobs under the same occupation is homogeneous. This strategy therefore rules out the possibility that heterogeneity in tasks drives within-occupation wage differences.

In this paper, we explore whether the assumption of homogeneity in the task composition of jobs within occupations holds empirically. To do this, we focus on a natural and relevant dimension along which the task content of a given occupation might vary: establishment size. Employer size is meaningful because it is a simple statistic that strongly correlates with total factor productivity and managerial quality both theoretically (Lucas, 1978) and empirically (Bloom and Van Reenen, 2007). Larger employers also tend to use more automation and offshoring (Alekseeva et al., 2021). Given this gradient in how production is organized, the task content of jobs in establishments of different sizes is likely to also differ (Ocampo, 2022). These differences in the task content of jobs have concrete implications on the compensation of workers and may explain part of the so-called establishment size wage premium.

Our main result is that the assumption of homogeneity in the task composition of jobs within occupations does not hold empirically. In particular, we find that, even *within* narrowly defined occupations, there are systematic differences in the task intensities performed by workers employed by establishments of different sizes. We document that the intensity of non-routine analytical tasks of workers in larger establishments is, on average, 5–15% of a standard deviation higher. The size of this gap is comparable in magnitude to the gap present between the countries at the two extremes of the development spectrum, as documented by Caunedo et al. (2023). We also report suggestive evidence in favor of a establishment size gradient in non-routine interpersonal and routine cognitive tasks. Moreover, we find that, to undertake these tasks, workers in larger establishments rely significantly more on information and communication technologies (ICT). We interpret this result as an indication that workers in larger establishments may perform more non-routine analytical, non-routine interpersonal, and routine cognitive tasks through increased use of ICT.

These empirical patterns are robust to using both employer- and employee-based responses and are widely present across the comprehensive set of 46 high-, middle-, and low-income countries that we study after combining two large and representative data sets: the OECD Survey of Adult Skills and the World Bank Skills Measurement Surveys. Furthermore, replicating our analysis separately for each 1-digit ISCO occupation highlights that (1) the patterns are still present when zooming into occupations, which increases the reliability of the within-occupation comparisons that we make between large and small establishments, and (2) while the establishment size gradient in usage of ICT is present in all 1-digit occupations, the gradient for non-routine analytical is concentrated in certain occupations, particularly, professionals and elementary occupations. Finally, we provide evidence that our main finding is not only true for the average worker: the distributions of the intensity of performed tasks in larger establishments are also horizontally shifted relative to smaller establishments.

Though we acknowledge that we cannot rule out that the estimated gaps are partly confounded by selection of workers into firms of different sizes, it is reassuring that our main results hold after controlling for a rich set of observable worker characteristics including education, cognitive and non-cognitive skills, and industry of employment. Moreover, following Oster (2019), we find that selection on unobservables would have to be *at least* as large as selection on observables for our estimated establishment size gradients in task content to be indistinguishable from zero.

This paper provides novel evidence of systematic differences in task intensity by comparable workers in the same occupation according to whether they are employed by a large or a small establishment, and that these differences help explain wage differentials.[1] This result contributes to the recent literature documenting heterogeneity in task content within occupations (e.g., Deming and Kahn, 2018; Stinebrickner et al., 2019), which has emphasized the role of dimensions other than employer size. Atalay et al. (2021) show that the variation in tasks is correlated with city size such that larger cities have higher intensity of analytical and interactive tasks, more technological requirements, and increased task specialization. Additionally, a number of papers report that occupational task content varies across countries (Dicarlo et al., 2016; Lewandowski et al., 2019; De La Rica et al., 2020; Caunedo et al., 2023).

Our results on the establishment size heterogeneity of tasks provide novel foundations to understand employer productivity and its dynamics. Our findings are consistent with the implications of static models that endogenize the employers' decisions on how to allocate tasks for production (Ocampo, 2022; Adenbaum, 2023). For instance, Ocampo (2022) shows that automation may affect the task composition of occupations. Along with finding that workers in larger establishments use more ICT, we also document that they perform more non-routine analytical tasks compared to workers in smaller establishments. As employers grow larger, they likely accumulate more capital, automate, and offshore jobs (e.g., Jaimovich et al., 2023) which leads to changes in the task requirements of production. In particular, jobs may evolve to focus on non-routine analytical tasks and to use more ICT so as to complement the processes that aim to replace routine tasks. Future work exploring this mechanism may provide novel insights to understand the drivers of firm dynamics.

[1] Contemporaneous work by Adenbaum (2023) shows that larger firms hire a wider variety of occupations and make jobs more specialized, which means that workers in the same occupation perform different tasks depending on their firm's size. We provide complementary evidence to show that (1) this pattern is widespread across occupations and countries and (2) differences in task intensity within occupations partly explain wage inequality.

Moreover, the uncovered patterns on task heterogeneity connect naturally with the literature exploring the determinants of within-occupation wage dispersion and provide a plausible novel driver of the establishment size wage premium (ESWP) — the empirical fact that larger establishments tend to pay their workers more for doing the same occupation. We explore the implications of the establishment size gradient in occupational task intensities on wage determination in two steps. First, we document that, on average, workers in larger establishments earn about 15% more than their counterparts in smaller establishments, after controlling for 2-digit occupation codes. Our measured establishment size wage premium is consistent with the existence of a large and economically significant employer-size wage premia found in other studies employing alternative data sets (Velenchik, 1997; Gerlach and Hübler, 1998; Schaffner, 1998; Troske, 1999; Winter-Ebmer, 2001; Dobbelaere, 2004; Söderbom et al., 2005; Lehmer and Möller, 2010; Bloom et al., 2018; Colonnelli et al., 2018; Reed and Thu, 2019; Lochner et al., 2020; Porcher et al., 2023). We extend previous analyses to show that this is not driven exclusively by a few workers in larger establishments that are paid disproportionately more. Rather, the distribution of wages in larger establishments is shifted to the right compared to the distribution of wages in smaller establishments.

Second, we conduct a mediation analysis to provide suggestive evidence on the sources of this establishment size wage premium, including our finding that task composition varies across establishments of differing sizes. A number of explanations for the existence of the ESWP have been proposed (Brown and Medoff, 1989; Oi and Idson, 1999): (i) large employers hire more skilled workers (worker selection); (ii) large employers have worse working conditions (compensating differentials); (iii) large employers have market power and share rents with workers (productivity); (iv) large employers have higher costs of monitoring and pay efficiency wages; and (v) large employers pay higher because of threat of unionization. In this paper, we explore the establishment size gradient in occupational task intensities as a complementary source of the ESWP. We find that this mechanism is able to explain over 10% of the raw ESWP. This proportion ranges from 5 to 20% across the countries in our sample.

Our novel empirical pattern therefore accounts for an economically significant fraction of the ESWP that is comparable to the share explained by the sorting of higher educated individuals into larger establishments.

These results open an exciting research avenue to explore the consequences of the employer size gradient in the task content of jobs on dynamic wage determination. Recent evidence indicates that early-career experience in large firms has dynamic rewards in future worker outcomes (Arellano-Bover, 2024). Our findings suggest a compelling mechanism to rationalize this — the experience in performing non-routine analytical tasks and in using ICT accumulated by younger workers in larger establishments is rewarded with better future earnings prospects (Stinebrickner et al., 2019). We urge future research to probe along these lines.

Outline of the paper. The remainder of this paper proceeds as follows. In Section 2, we describe the main data sets as well as our definition of occupations, and detail the measures of task content used in the analysis. In Section 3, we document novel facts on the heterogeneity of occupational task contents across establishments of differing sizes and discuss potential drivers. In Section 4, we measure the establishment size wage premium and study a number of explanations for its existence, including the establishment size gradient in task intensity. Finally, in Section 5, we conclude with a summary of the findings and a discussion of future directions of work. An appendix contains additional results.

2 Data and measurement

2.1 Data sources

We take advantage of the availability of cross-country harmonized surveys reporting the tasks performed by individuals in their work to construct a rich dataset covering working (not self-employed) individuals aged 16–65 across 46 countries at various stages of economic development. We combine two main data sets.

OECD Survey of Adult Skills. The Survey of Adult Skills is a cross-sectional, cross-country survey conducted under the OECD's Programme for International Assessment of Adult Competencies (PIAAC). This survey aims to measure cognitive skills (literacy, numeracy, and problem solving in technology-rich environments), as well as skills used both at work and in other contexts. It is representative of the country's adult population aged 16–65, with around 5,000 individuals participating in each country.[2] There have been three rounds of data collection (2008–2013, 2012–2016, and 2016–2019). We focus on the surveys collected from the following 30 countries: Belgium, Chile, Cyprus, Czech Republic, Denmark, Ecuador, France, Greece, Hungary, Ireland, Israel, Italy, Japan, Kazakhstan, South Korea, Lithuania, Mexico, Netherlands, New Zealand, Norway, Peru, Poland, Russian Federation, Singapore, Slovak Republic, Slovenia, Spain, Sweden, United Kingdom, and United States.[3]

Full earnings information are not available in the public-use files of New Zealand, Peru, Singapore, and USA. Instead, earnings are only reported in deciles.[4] These countries are still employed in the quantification of the gradient in task intensity by firm size.

World Bank Skills Measurement Surveys. Our second main data source is the World Bank's STEP Skills Measurement Program surveys. They also are cross-sectional surveys that aim at measuring the demand and supply of skills in urban areas of low- and middle-income countries, which allows us to complement the set of high- and middle-income countries available in the OECD's Survey of Adult Skills. There are two types of surveys in the program: household-based and employer-based.

[2]The United States conducted a second round of data collection to get more reliable estimates for certain subgroups. In the graphs presented in this paper, we label the results based on the first round (conducted in 2012) as "United States" and those based on the second round (conducted in 2017) as "USA2".

[3]We exclude Turkey because we cannot construct our measure of non-cognitive skills of the worker. We confirm that the results from specifications that do not control for this variable are similar when including Turkey.

[4]Peru lacks this information as well.

The household-based surveys interview a randomly-selected household member (aged 15 to 64) about their personal education and training history, work status and history, skills used in their jobs, earnings, individual competencies, and non-cognitive traits and abilities (e.g., personality, behavior, risk preferences). Sample sizes varied from 3,000 to 4,000 individuals. We focus on the surveys that contain consistent questions regarding tasks and skills, corresponding to the following 11 countries: Armenia, Bolivia, China (Yunnan Province), Colombia, Georgia, Kenya, Laos, Macedonia, Sri Lanka, Ukraine, and Vietnam.[5]

Additionally, in some countries, firms were also surveyed using an employer-based questionnaire. In this module, an informed respondent from around 300 to 500 firms per country reported the worker composition of the firm, the skills required from workers in different occupations, and the amount of in-firm training provided. We use the employer-based survey of the following nine countries: Albania, Armenia, Azerbaijan, Bosnia-Herzegovina, Georgia, Kenya, Kosovo, Serbia, and Vietnam. Note that four of these countries have also conducted the household-based survey, which allows us to document consistent evidence of within-occupation task heterogeneity both from the employer and the employees' perspective.

Strengths and limitations of data used. The main virtue of these datasets is the availability of information about the tasks performed by individuals in their own work that are comparable across a wide range of countries. The main limitation is that they are cross-sectional. In particular, note that although both STEP and PIAAC were conducted over multiple rounds across years, only one country (the U.S.) was surveyed twice with different sets of respondents. In the absence of a panel, we are limited in the mechanisms that we explore. For instance, we cannot control for additional individual heterogeneity outside the characteristics we observe nor can we speak to the dynamics of task requirements and wages.

[5]Though Ghana has a household-based survey that contains the relevant variables, we exclude it because of the small sample size that remains after sample selection. The Philippines was also a survey country but a different questionnaire was used.

2.2 Measuring establishment size, occupational task content, and wages

Establishment size and the presence of large establishment gaps. Both datasets provide a measure of workplace or establishment size based on the number of employees, reported in bins. We define establishments that have at least 50 employees as large.[6] The establishment size gaps that we measure in the main analyses compare workers in establishments that have at least 50 employees to workers in establishments with less than 50 employees. The qualitative results are robust to the use of alternative cutoffs for the definition of "large".

Occupations. Occupations typically refer to a "set of jobs whose main tasks and duties are characterized by a high degree of similarity".[7] In statistical analyses within economics, occupation classifications also serve as a convenient way to summarize heterogeneity across jobs. Though useful, we show that these occupational classifications may mask heterogeneity in within-occupation task intensity which, in turn, could help to explain within-occupation wage inequality.

We focus on within-occupation variation as defined by the widely-used 2008 version of the International Standard Classification of Occupations (ISCO-08) by the International Labor Organization. The ISCO-08 is a four-level hierarchical, nested classification that is coded using four digits. The first digit refers to one of the 10 "major groups" which typically reflect skill complexity. Each "major group" is divided into several "sub-major groups" (2nd digit), which in turn are divided into one or more "minor groups" (3rd digit), which in turn are divided into one or more "unit groups" (4th digit). There are a total of 43 sub-major groups, 130 minor groups, and 436 unit groups. The most detailed occupational code available in the PIAAC and STEP surveys has 3 digits.

For parsimony, our preferred specifications use 2-digit ISCO-08 codes, which we believe

[6] The OECD classifies micro enterprises as those with fewer than 10 employees, small enterprises as those with 10 to 49 employees, medium-sized enterprises as those with 50 to 249 employees, and large enterprises as those with 250 or more workers. See https://data.oecd.org/entrepreneur/enterprises-by-business-size.htm.

[7] See https://isco-ilo.netlify.app/en/isco-08/.

balances the trade-off between the specificity of the occupations and sample size concerns. Though other datasets often include occupation codes at a more granular level, a number of papers that study job polarization and tasks in the labor market discuss occupations at a similar aggregation to ours (e.g., Spitz-Oener, 2006; Goos et al., 2014; Fonseca et al., 2018). We show that the results are quantitatively similar using 3-digit ISCO-08 codes.

Task content of occupations. We follow the approach in Caunedo et al. (2023) to construct task measures that are comparable to well-established definitions in the literature (e.g., Autor et al., 2003; Acemoglu and Autor, 2011). We distinguish five task components of occupations: non-routine analytical (NRA), non-routine interpersonal (NRI), routine cognitive (RC), routine manual (RM), and non-routine manual (NRM). Non-routine analytical tasks involve reading and thinking creatively. Non-routine interpersonal tasks require interacting with others (e.g., through advising, negotiating, teaching). Routine cognitive tasks require structured repetition of activity planning and time management. Routine manual tasks involve physically demanding activities. Finally, non-routine manual tasks involve manual dexterity. The exact variables employed are listed in Appendix Table A1.

We create an individual index measuring the intensity of a particular task category in two steps. First, we standardize the responses to each task variable to have a within-country mean of zero and standard deviation of unity. Second, to obtain the index for a task category, we average the standardized responses to the task variables and re-standardize the result to again have a within-country mean of zero and a standard deviation of one. By construction, these measures are interpreted as intensities in units of standard deviations relative to the country mean. In Appendix C.1, we consider an alternative construction of task intensity indices using multiple correspondence analysis.

Usage of ICT. A particular focus of our paper is on documenting the intensity with which workers use technologies such as computers and specific software as part of their work.

Though we report results relating to the usage of ICT alongside the task dimensions discussed above, we do not consider it mutually exclusive to those tasks. Rather, *we interpret the use of ICT as a means through which the tasks are performed.* In Table A1, we show the questions in the surveys that are relevant to measure the use of ICT. We create an index in a similar manner as for the above measures of task content.

Wages. STEP's worker-based survey reports log hourly earnings in USD. PIAAC elicits hourly wages (in levels of the domestic currency), but for some countries wages are only reported in bins.[8] To quantify the establishment size wage gap, we focus on log hourly wages in non-self-employment work. We deflate the values to 2018 local currency and use 2018 exchange rates to US dollar to convert earnings to real 2018 USD.

Demographics and additional controls. To increase the comparability of demographic variables across surveys, we first consider the following standard controls: gender, age block (10-year groups starting from age 16 and ending at 65), and three education categories based on ISCED 2008 — (i) primary education or less (ISCED 1); (ii) up to a professional tertiary education degree (ISCED 5), and (iii) bachelor's degree and above (ISCED 5A and beyond).

We further aim to better account for the potential sorting of workers with higher ability or higher non-cognitive skills into larger establishments. In terms of cognition, for STEP countries we standardize, at the country level, the proportion of correct responses over the total number of questions in three different linguistic tests (vocabulary, sentence, and passage). For PIAAC countries, we use the first imputation in both the numeracy and literacy competences, and we verify that the results hold employing item response theory over the ten imputations available in the survey (Khorramdel et al., 2020). In terms of non-cognitive abilities, STEP provides pre-constructed measures for the following traits: openness, stability, agreeableness, and grit. We employ the (standardized) first principal component from these four dimensions.

[8] A complete overview on the availability of the task and wage information for each country can be found in Appendix A.

Though PIAAC is known for being less well-equipped for measuring individual non-cognitive traits, we still use a number of measures that have previously been shown to predict earnings. In particular, we follow Anghel and Balart (2017) in using measures of cultural engagement, social trust, and political efficacy, and we follow Cabrales et al. (2014) in employing a measure of motivation for learning. We combine these four measures by taking their first principal component to proxy for the respondent's non-cognitive skills.

We consider the following sectoral classification for STEP countries: (i) agriculture, fishing, mining; (ii) manufacturing and construction; (iii) commerce; and (iv) other services. For PIAAC, we use more detailed information encompassing twenty-one different industries.

Finally, in additional specifications reported in the Appendix, we account for regional variation in industrial and demographic composition by employing regional fixed effects, which are always available in STEP but are missing for a subset of PIAAC countries (Italy, Norway, and United States). Importantly, the nature of these regions changes across countries, even within the same survey. For instance, among STEP countries, regions refer to metropolitan areas in Colombia while in China (Yunnan) they pertain to census enumeration areas. In PIAAC countries, the geographical information corresponds to OECD TL-2 territorial levels (representing the first administrative tier of subnational government), which are politically defined. Given that this hinders the interpretation of these fixed effects and that they are only available for a subset of countries, we consider the inclusion of regional controls as a robustness check rather than as part of our main specification.

2.3 Summary statistics

Appendix Table A2 reports summary statistics for the 36 countries for which we have a continuous measure of wages. To ensure that our results are not driven by extrapolation, we also impose that each establishment size and 2-digit occupation code cell has at least 5 observations. The number of observations after focusing on working-age individuals that are not self-employed varies from 857 (Greece) to 3,984 (United Kingdom) among PIAAC countries and from 360 (Laos) to 1,289 (Vietnam) among STEP countries. In general, small establishments are more prevalent but there is significant cross-country variation. For instance, in Belgium and the Netherlands 50% of the establishments are large, while the share is around 20% in Ecuador and Greece.

3 Establishment size gradient in the task content of jobs

In this section, we document from various perspectives our novel stylized fact that, even within narrowly defined occupation groups, there are significant differences in the task composition of jobs across workers of establishments of different sizes. To quantify such gradient we estimate versions of the following regression:

$$T_i = \beta \times \mathrm{LE}_{j(i)} + X_i' \gamma + \delta^o_{o(i)} + \delta^c_{c(i)} + \varepsilon_i, \tag{1}$$

where T_i is the measure of task content of the job of worker i, $\mathrm{LE}_{j(i)}$ is an indicator of whether the establishment $j(i)$ of individual i has at least 50 employees, X is a vector of individual- and establishment-level characteristics, and δ^o and δ^c are occupation-code and country fixed effects, respectively. We focus on our five main task categories (NRA, NRI, RC, RM, NRM) as well as the use of ICT as outcomes. Our coefficient of interest is β, which captures the average difference in intensity in doing task T between two observably equivalent workers in the same country and occupation who differ in that one is employed by a large establishment and the other by a small one.[9] We report standard errors clustered at the country level.

In Table 1's panels (a) and (b), we report estimates of β in Equation 1 using the pooled samples of PIAAC and STEP countries, respectively. In column (1), we present the establish-

[9]Throughout the paper, in the pooled regressions, we use probability weights, adjusted based on the population of the various countries in 2018 (with the exception of China, for which we use the population of Yunnan – the only province of the country that was surveyed). Intuitively, this weighting approach places more weight on observations from large-population countries.

Table 1: Pooled estimates of establishment size gradient in the task content of jobs

Outcome variable	(1)	(2)	(3) $\hat{\delta}$
Panel (a): PIAAC			
Non-routine analytical	0.155***	0.117***	1.488
	(0.019)	(0.017)	
Non-routine interpersonal	0.074***	0.055**	1.427
	(0.022)	(0.022)	
Routine cognitive	0.008	0.035**	−0.632
	(0.015)	(0.014)	
Routine manual	-0.006	0.011	−0.149
	(0.015)	(0.010)	
Non-routine manual	-0.007	0.001	−0.030
	(0.017)	(0.014)	
Use of ICT	0.156***	0.132***	1.525
	(0.010)	(0.010)	
Sample size	65,151	65,151	
Panel (b): STEP			
Non-routine analytical	0.125***	0.066**	0.395
	(0.032)	(0.023)	
Non-routine interpersonal	-0.015	-0.035	−0.601
	(0.030)	(0.034)	
Routine cognitive	0.169**	0.191***	-4.020
	(0.067)	(0.060)	
Routine manual	-0.020	0.007	−0.068
	(0.034)	(0.033)	
Non-routine manual	-0.048	-0.060	−4.489
	(0.038)	(0.035)	
Use of ICT	0.194***	0.125**	0.696
	(0.057)	(0.043)	
Sample size	8,339	8,339	
Controls:			
Country FE	Yes	Yes	
2-d Occupation FE	Yes	Yes	
Industry FE		Yes	
Individual demographics		Yes	
Individual cognition/noncog.		Yes	

*Notes: Regressions of a given category of task requirement intensity on an indicator of large establishment (at least 50 employees). Each row refers to a specific task category. Additional controls are indicated in the lower part of the table. Individual demographics include education, gender, and age. Regressions are conducted separately for the pooled sample of PIAAC and STEP countries in panel (a) and (b), respectively. Standard errors are reported in parenthesis and clustered at the country level. Column (3) reports the estimated Oster (2019)'s $\hat{\delta}$. *** p<0.01, ** p<0.05, * p<0.1* ment size gradient in task intensity only controlling for occupation and country fixed effects. Qualitatively, we find that workers in larger establishments perform more non-routine analytical tasks and make more intensive use of information and communication technologies. We also find suggestive evidence that workers in larger establishments perform more routine cognitive tasks, a pattern that is more evident in the STEP pooled sample than in the PIAAC one.[10]

We do not find a difference in the intensity with which manual tasks, either routine or non-routine, are performed between workers in larger and smaller establishments. The establishment size gradient in non-routine interpersonal tasks is concentrated among PIAAC countries (we return to this later when we document that the gradients for these tasks appear to be more country-specific). Note that, by controlling for occupation fixed effects at the 2-digit level based on the ISCO-08 classification, we account for the possibility that the occupational structure of large and small establishments differs in a way that could explain these patterns.

In column (2), we build on our baseline results to account for potential confounders of the establishment size gradient. We include industry fixed effects to rule out the possibility that the gaps are driven by larger establishments disproportionally concentrating in industries that use certain tasks more intensively, or have establishments that are more productive. Moreover, we introduce individual controls for education, gender, and age to account for the most salient sources of worker selection into establishments that are typically recorded in standard datasets. Lastly, we take advantage of the availability of measures of cognitive and non-cognitive skills in the surveys we use to show that the task gap barely changes after their inclusion, which reinforces the idea that worker selection cannot fully explain the gradient in task requirements. We find qualitatively similar results, which suggest that our results do not

[10] Apart from the gap in economic development between PIAAC and STEP countries, another potential explanation behind this difference is the fact that the subcomponents of the routine cognitive category may capture different skills in PIAAC and STEP. As can be seen in Table A1, PIAAC focuses on planning-related activities while STEP expands the interest to the actual execution of tasks that require routine cognitive skills (e.g., short, repetitive tasks).

simply arise from differential selection of workers into small and large establishments.[11]

In terms of economic magnitude, focusing on column (2), we find that the average worker in a large establishment performs around 11.7% and 6.6% of a standard deviation more non-routine analytical tasks than a worker in a small establishment in the PIAAC and STEP countries, respectively. We interpret this standard deviation as relative to the country-specific distribution of performed tasks. The average worker in the large establishment also uses 13.2% and 12.5% of a standard deviation more ICT compared to the average worker in smaller establishments in STEP and PIAAC countries, respectively. In terms of routine cognitive tasks, the average large-establishment worker in PIAAC countries performs 3.5% of a standard deviation more than their counterparts in smaller establishments; the difference is larger at 19.1% of a standard deviation in the STEP sample. The size of these gaps is comparable to the one present between the countries at the two extremes of the development spectrum, as measured by GDP per capita (Caunedo et al., 2023).

Robustness to selection on unobservables. The results above show that the gaps in skill use persist after the inclusion of a rich set of observable covariates. Still, the cross-sectional nature of the data limits the amount of unobserved individual heterogeneity that we can control for. To show that the facts that we uncover are unlikely to be driven by selection on unobservables, we follow Oster (2019) who suggests to estimate the value of a parameter, denoted as δ, indicating how much stronger/weaker selection on unobservables would have to be, relative to selection on observables, to render the coefficient of interest indistinguishable from zero.[12] A value of δ of 1 indicates that selection on unobservables would have to be as

[11] In Appendix Table B1, we show the estimates when we introduce the confounders sequentially. Moreover, we include two additional columns where we (1) saturate the regression with the interactions of all our controls with country fixed effects to allow the returns to such controls to vary across countries, and (2) include regional fixed effects to account for spatial differences in the presence of large and small establishments and in tasks and occupations. We find quantitatively similar results so we keep column (2) in Table 1 as our preferred specification for parsimony. In Appendix Table B2, we find the same qualitative patterns employing 3-digit occupation fixed effects instead. In Appendix Table B3, we report the results from the PIAAC sample when we do not discard countries with missing regional information. We do not identify any substantial differences. Appendix Table B4 shows that these findings remain under alternative measures of establishment size.

[12] A more detailed explanation of the methodology is in Appendix D.1.

large as selection on observables to make the coefficient of interest be zero. Since one would typically believe that the included controls are capable of explaining a large fraction of the variation in the outcome, Oster (2019) suggests that a δ of 1 or higher is a rule-of-thumb to be confident that selection on unobservables is not a large issue. The estimated $\hat{\delta}$s for the different establishment size gradients in tasks are reported in column (3) of Table 1. We find that the estimated δs for the tasks for which we identified quantitatively significant gaps comfortably satisfy the proposed rule-of-thumb.[13]

Alternative treatment of occupations. Our main interest is in documenting within-occupation heterogeneity in task composition. While we have demonstrated the robustness of our results to the use of various degrees of specificity in the occupation codes (2-digit and 3-digit), in Table 2 we take an even more flexible approach. In particular, we estimate our preferred specification separately for subsamples defined by 1-digit occupation codes, while still controlling for dummies of 2-digit occupation codes. This approach has the conceptually attractive feature of allowing for more precise within-occupation comparisons of task intensity between establishments of different sizes by considering subsamples that contain more similar occupations.[14] We find that workers in larger establishments across virtually all 1-digit occupation codes perform more non-routine analytical tasks and use more ICT in their work, both within PIAAC and STEP countries, as consistent with the strong effects we found in the pooled results. Not surprisingly, there are differences in the magnitude of the establishment size gradient depending on which 1-digit occupation code we focus on. For instance, among PIAAC countries, the largest gaps in the performance of non-routine analytical tasks and in the use of ICT are seen among services and sales workers. Looking at the undertaking of routine cognitive tasks, we find that that the establishment size gradient in tasks is driven

[13] In column (7) of Table B1, we estimate the corresponding Oster (2019)'s δ where we include industry fixed effects in the baseline regression. The long regression, therefore, adds only individual controls. Relative to selection based on observable individual characteristics only, the estimated δs are generally smaller in magnitude but still indicate that our main results are robust to selection on unobservables.

[14] The main drawback is that the sample size for some 1-digit occupations is limited, particularly among STEP countries, which may affect the precision of our estimates and the extent to which we rely on extrapolation.

Table 2: Establishment size gradient in the task content of jobs, by 1-digit occupation codes

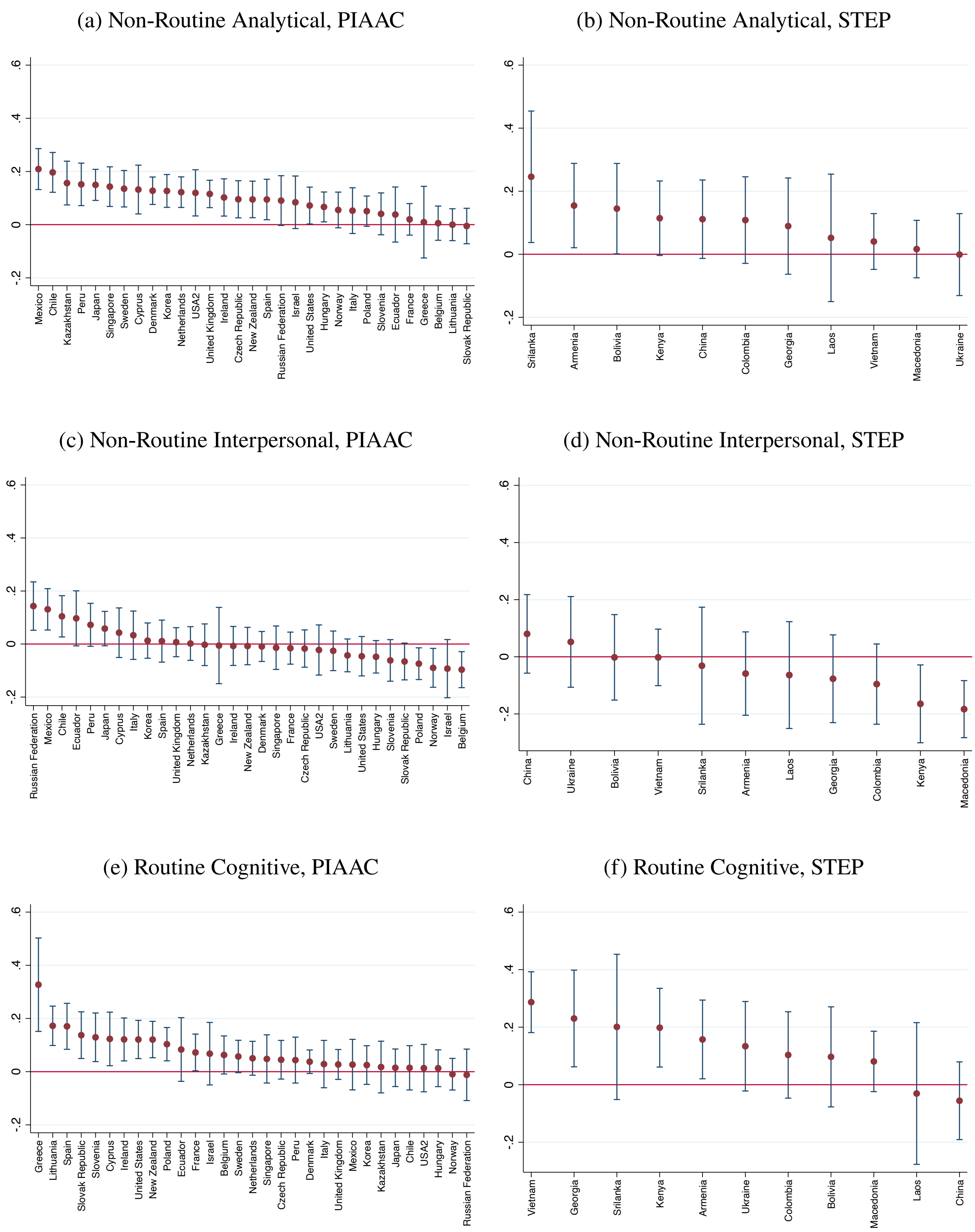
	Task Category						
1-digit ISCO-08 Category	NRA	NRI	RC	RM	NRM	ICT	# Obs.
Panel (a): PIAAC							
Managers	0.087	-0.003	-0.010	-0.020	0.021	0.036	4,157
	(0.052)	(0.042)	(0.024)	(0.047)	(0.060)	(0.033)	
Professionals	0.088***	0.051**	-0.045**	-0.016	-0.082	0.085***	13,472
	(0.026)	(0.024)	(0.020)	(0.040)	(0.063)	(0.014)	
Technicians & associate professionals	0.132***	-0.036	0.021	-0.035	0.059	0.086***	9,601
	(0.027)	(0.025)	(0.025)	(0.038)	(0.038)	(0.027)	
Clerical support workers	0.050	0.040	0.112***	-0.111**	0.011	0.096***	7,503
	(0.044)	(0.035)	(0.040)	(0.050)	(0.040)	(0.025)	
Services & sales workers	0.167***	0.160**	-0.022	0.103**	0.069*	0.223***	12,175
	(0.031)	(0.061)	(0.059)	(0.042)	(0.035)	(0.030)	
Craft & related trade workers	0.076	-0.022	0.087	-0.032	-0.084**	0.190***	6,344
	(0.093)	(0.072)	(0.109)	(0.039)	(0.040)	(0.064)	
Plant & machine operators, & assemblers	0.163*	0.060	0.051	-0.022	0.021	0.179***	5,244
	(0.080)	(0.066)	(0.108)	(0.024)	(0.041)	(0.028)	
Elementary occupations	0.113***	0.131***	0.124***	0.120***	0.027	0.200***	6,132
	(0.023)	(0.043)	(0.039)	(0.042)	(0.056)	(0.038)	
Panel (b): STEP							
Managers	0.150	-0.015	0.369**	-0.170	-0.128	0.260	451
	(0.131)	(0.152)	(0.142)	(0.150)	(0.160)	(0.202)	
Professionals	0.017	0.068	0.074	-0.046	-0.095**	0.177*	2,142
	(0.032)	(0.061)	(0.048)	(0.055)	(0.038)	(0.085)	
Technicians & associate professionals	0.149*	0.271*	0.058	0.150**	-0.027	-0.052	788
	(0.079)	(0.126)	(0.123)	(0.060)	(0.096)	(0.067)	
Clerical support workers	0.035	-0.126	0.060	0.069	0.088	0.198***	895
	(0.067)	(0.098)	(0.101)	(0.038)	(0.131)	(0.038)	
Services & sales workers	0.202***	0.047	0.223***	-0.117	0.009	0.218*	1,847
	(0.062)	(0.080)	(0.065)	(0.075)	(0.084)	(0.106)	
Craft & related trade workers	0.107	-0.214*	0.275**	0.135	0.121	0.093	687
	(0.122)	(0.105)	(0.092)	(0.104)	(0.095)	(0.119)	
Plant & machine operators, & assemblers	0.286**	-0.122	0.194	-0.024	-0.295**	0.139*	588
	(0.108)	(0.068)	(0.111)	(0.075)	(0.120)	(0.075)	
Elementary occupations	0.087	0.204*	0.150	0.042	-0.064***	0.100*	941
	(0.064)	(0.102)	(0.150)	(0.107)	(0.018)	(0.049)	

*Notes: Coefficient of an indicator for large establishment in a regression of task content intensity on indicator of large establishment (at least 50 employees) by 1-digit ISCO-08 occupation codes and the set of controls in Table 1's column (2). We do not report the 1-digit categories corresponding to armed forced occupations and skilled agricultural, forestry and fishery workers due to small sample size. Standard errors are reported in parenthesis and clustered at the country-level. *** p<0.01, ** p<0.05, * p<0.1* by elementary occupations and clerical support workers in the PIAAC sample. Finally, for most 1-digit occupation categories, in both the PIAAC and STEP samples, we find small establishment size differences in the performance of routine and non-routine manual tasks – as we found in our main specification.

Cross-country comparisons. We complement the pooled analysis from Table 1 by exploring the presence of the establishment size gradient separately for each country. We report the results in Figure 1. What we find aligns well with our previous results: (i) the larger reliance on non-routine analytical and the use of information and communication technologies is present in virtually all countries in our sample; and (ii) the higher intensity on non-routine interpersonal and routine cognitive tasks, while prevalent throughout many countries, also features a subset of countries for which the effects are not distinguishable from zero. Appendix Figure B1 shows that the differences in the performance of routine and non-routine manual tasks are mostly indistinguishable from zero and, if any, are negative.

Country-level correlates of the establishment size task gradient. Although the qualitative patterns uncovered above are fairly similar across countries, some quantitative differences arise. There are several reasons why this might be the case, including: (1) differences in labor market institutions, and (2) differences in the relevance of firms with at least 50 employees. We explore whether these differences can be explained by the level of development across countries, focusing on two indicators: log GDP per capita and the proportion of the population that has completed at least tertiary-level education. In Appendix Figures B2 and B3, we plot the establishment size gradients in the task content of jobs against log GDP per capita and the fraction of population with at least tertiary education, respectively. We highlight two empirical patterns. First, the establishment size differences in the use of non-routine analytical tasks is uncorrelated with log GDP per capita and only slightly positively correlated with the proportion of the population that is at least tertiary educated. This suggests that this

Figure 1: Establishment size gradient in the task content of jobs by country, within 2-digit occupations

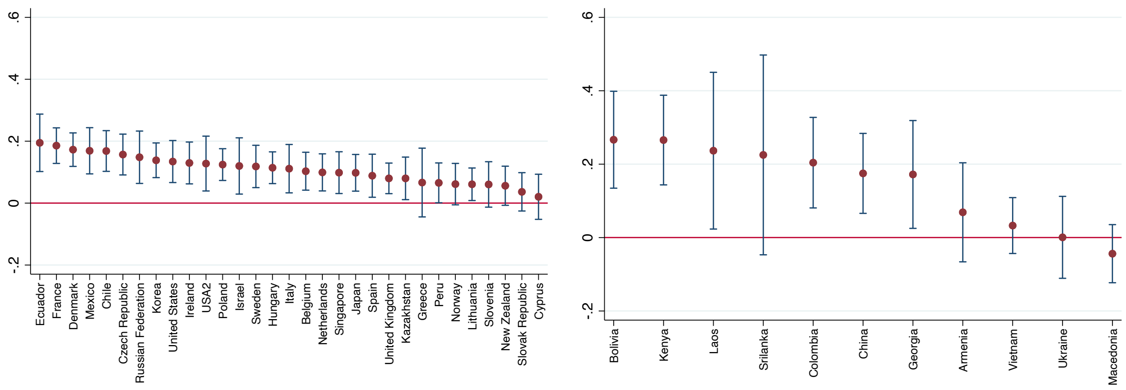


Notes: Coefficient of an indicator for large establishment in a regression of task content intensity on indicator of large establishment (at least 50 employees) and the set of controls in Table 1's column (2). Regressions are done separately for each country. Countries are ordered by decreasing point estimates. "United States" refers to the survey conducted in 2012 and "USA2" to the one conducted in 2017. Reported confidence intervals at 95% confidence level computed using heteroskedasticity-robust standard errors.

Figure 1: Establishment size gradient in the task content of jobs by country, within 2-digit occupations

(g) Use of ICT, PIAAC

(h) Use of ICT, STEP



Notes: Continuation of Figure 1. Coefficient of an indicator for large establishment in a regression of task content intensity on indicator of large establishment (at least 50 employees) and the set of controls in Table 1's column (2). Regressions are done separately for each country. Countries are ordered by decreasing point estimates. "United States" refers to the survey conducted in 2012 and "USA2" to the one conducted in 2017. Reported confidence intervals at 95% confidence level computed using heteroskedasticity-robust standard errors.

pattern is not driven by economic development. In contrast, the establishment size gradient in the performance of routine cognitive and the use of ICT is negatively correlated with economic development—in richer countries and countries with a more educated population, the establishment size differences in the task content of jobs are smaller.

Differences in the *distribution* of task intensity by establishment size. So far our results document average differences in task composition of occupations between establishments of different sizes. In Appendix C.2, we extend our analysis by also exploring the differences in the distribution of task intensity. For this, we employ distributional regressions in the spirit of Chernozhukov et al. (2013) and show that the large establishment differences in non-routine analytical and in the use of ICT are present at multiple thresholds throughout the support of the distribution. This demonstrates that the mean differences we find are not driven solely by discrepancies in the upper tail of the distribution of tasks performed. Instead, the distributions of the intensity of performed tasks in larger establishments are all shifted to the right compared to those in smaller establishments. For the case of routine and non-routine manual tasks, the insignificant differences are seen throughout the intensity distribution, except for the case of routine manual tasks for large establishments where there is suggestive evidence of a widening of the intensity distribution in larger firms.

When does the gradient arise? We provide two pieces of evidence to suggest that the establishment size gradient in the task content of jobs does not arise from larger employers exclusively assigning more non-routine analytical and usage of ICT to workers with longer tenure in the establishment. Rather, these establishment size gradient is already reflected in the labor demand of the employers and thus appears at the beginning of the job tenure and early in the workers' careers.

First, in Appendix Table B5, we re-estimate our preferred specification conditioning first on workers having been in their current job for a short period of time (up to 2 years) and then additionally on being young (less than 25 years old). This excludes workers that may have adopted more task-intensive work as they progressed in their careers. Among these young workers with short tenure, we find establishment size gradients in non-routine analytical tasks and the use of ICT that are of the same sign and comparable in magnitude to the full sample.

Second, we leverage the availability of employer surveys from the World Bank STEP Skill Surveys program to examine whether the task requirements of firms already differ at the time of hiring.[15] In these surveys, employers answer a limited set of questions on the skill requirements of occupations in the workplace. Based on the questions asked in the survey, we are only able to identify tasks that pertain to the following categories: (1) non-routine analytical and (2) use of ICT. To limit the burden on the survey respondent, STEP only elicits two occupations (randomly selected out of nine categories). Appendix Table B6 reports

[15] We acknowledge that the results in this subsection are based on a small number of low- and middle-income countries so external validity is limited.

estimates of average differences in task requirements between large and small employers, within occupation categories, for the pooled sample of nine countries where the STEP Employer Survey is available. We again find that large employers require more non-routine analytical tasks and use more ICT as early as in the hiring stage. This further suggests that the origins of the gradients arise from systematic differences in how production or technology is organized by employer size rather than from how workers are able to accumulate more specialized tasks over their tenure or career.

Discussion: Sources of the establishment size gradient in task content. The more fundamental question remains: *why* do workers in the same occupation engage in different task intensities depending on their establishment's size? To the best of our knowledge, there is no unified framework that may be used to exhaustively explore the potential drivers of the establishment size gradient, so in this subsection we focus on providing several well-grounded plausible explanations, relate them to our empirical findings, and highlight that further theoretical and empirical analyses will be needed to shed more light on this question.

To make the exposition easier to follow, let us take a concrete example. Our results suggest that an accounting professional in a larger establishment performs non-routine analytical tasks more intensely and uses more ICT than another accounting professional in a smaller establishment, even within the same industry and controlling for the innate skills of the accountants. This is consistent with Adenbaum (2023) who argues that larger, more productive establishments are able to organize production such that workers are more specialized. The accountant in the small establishment may need to perform additional administrative tasks like sorting mail or answering phones, whereas in larger establishments, specific workers are hired to perform these administrative tasks, leaving the accountant to focus on the tasks they specialize in, which tend to be more non-routine analytical and use more ICT. We find some support towards this story in that, for certain occupation categories, we document that workers in larger establishments display higher intensity in some task categories at the expense of engaging less in other task categories. However, our pooled results from Table 1 show that, on average, workers in larger establishments perform more non-routine analytical tasks and use more ICT without significantly decreasing their performance of other task categories.

A second plausible rationale is rooted in the idea that establishment size does not only relate to organizational capacity. Larger establishments are naturally more complex. The increased complexity in larger establishments may lead workers in the same occupation to perform tasks with different intensity to obtain the same goal. The accountant in the small firm will come across the same financial transactions in their day-to-day as the accountant in the large establishment, but the latter is more likely to encounter complex financial transactions in their day-to-day which involve more non-routine analytical work.

To deal with such complexity, larger establishments tend to adopt technologies that complement the scale of their operations (e.g., Alekseeva et al., 2021; Lashkari et al., 2024). While the accountant in the smaller establishment may just need a simple spreadsheet to prepare the employer's financial statements, the accountant in the larger establishment would be required to use more specialized accounting software (which helps the larger establishment to process more complex transactions, facilitates communication with other relevant personnel, and might improve replicability). One can make similar arguments for other occupations. For instance, the most complex equipment a mom-and-pop baker might use is an industrial oven while the baker in a large bread manufacturing plant has to work with complex equipment throughout the whole bread-making process.

Regardless of where the firm size gradient in task content comes from, we have shown that there is heterogeneity in the task content of jobs, even within the same occupation. This heterogeneity may be fundamental in helping us to better understand labor markets. For instance, in the subsequent section, we show that the differences in task content between large and small establishments can explain an economically significant portion of the wage gaps observed between workers in larger and smaller establishments.

4 Establishment size wage premium and the role of individual selection, sectors, and tasks

In this section, we first document in Subsection 4.1 the presence of a establishment size wage premium — both on average and throughout the wage distribution — using the pooled PIAAC and STEP samples separately. We then explore in Subsection 4.2 how much of this raw gap can be linearly explained by various mechanisms, including selection of individuals into occupations and differences in the task composition of occupations.

4.1 Establishment size wage premium: Cross-country evidence

Similar to how we documented the establishment size gap in the task content of jobs in Section 3, in this subsection we explore the establishment size gap in wages. We estimate the following regression:

$$\ln w_i = \beta \times \mathrm{LE}_{j(i)} + X_i'\gamma + \delta^o_{o(i)} + \delta^c_{c(i)} + \varepsilon_i, \tag{2}$$

where $\ln w_i$ is log real hourly wages in 2018 USD of individual i. The rest of the specification is similar to Equation (1). We interpret β as a measure of the ESWP, which is how much more workers with similar observables in establishments with at least 50 employees are paid (in log-points) relative to those in smaller establishments within the same occupation and country.

The estimated ESWP controlling only for country and 2-digit ISCO-08 occupation code fixed effects (Table 3, column (1)) is measured to be about 16.5% and 25.7% for PIAAC and STEP countries, respectively.[16] These results show that the establishment size wage gaps are not perfectly explained by differences in the occupational structure of establishments. In column (2) of Table 3, we saturate the regression with additional controls, including our

[16]Note that $\exp(0.153) - 1 \approx 0.165$.

Table 3: Pooled estimates of the establishment size wage premium

	Mean regressions		Quantile regressions		
	(1)	(2)	(3) p10	(4) p50	(5) p90
PIAAC					
ESWP	0.153***	0.111***	0.117***	0.106***	0.085***
	(0.014)	(0.012)	(0.010)	(0.010)	(0.012)
Sample size	54,782	54,782	54,782	54,782	54,782
STEP					
ESWP	0.229***	0.185***	0.216***	0.143***	0.056***
	(0.028)	(0.025)	(0.034)	(0.022)	(0.030)
Sample size	8,339	8,339	8,339	8,339	8,339
Controls:					
Country FE	Yes	Yes	Yes	Yes	Yes
2-d Occupation FE	Yes	Yes	Yes	Yes	Yes
Tasks		Yes	Yes	Yes	Yes
Industry FE		Yes	Yes	Yes	Yes
Individual demographics		Yes	Yes	Yes	Yes
Individual cognition/noncog.		Yes	Yes	Yes	Yes

*Notes: The first two columns show regressions of log hourly wages (in 2018 USD) on an indicator of large establishment (at least 50 employees) and the set of controls in Table 1's columns (1) and (2), respectively. PIAAC countries for which continuous wage data are not available are excluded. Appendix Table B7 additionally reports the point estimates and standard errors for the various tasks and computer use, which are not reported in the present table for brevity. The last three columns show results from quantile regressions under the specification in column (2). Standard errors are reported in parenthesis and clustered at the country level. *** p<0.01, ** p<0.05, * p<0.1*

proposed mediator, the establishment size gradient in tasks.[17] Though the estimated ESWP falls after the inclusion of these possible mediators, we find that there is still a substantial ESWP left unexplained. In the next subsection, we explore how much of the explained ESWP can be attributed to the different mediators considered.

Expanding on the existing literature, we explore the ESWP beyond the comparison of *average* wages between large and small establishments. In columns (3)–(5) of Table 3, we

[17] In Appendix Table B7, we report how the estimated ESWP changes as the mediators are introduced sequentially.

report the β coefficients in a quantile regression version of Equation (2) at the 10th, 50th, and 90th quantiles, controlling for tasks, industry fixed effects, and individual controls. We find that the worker at the 10th percentile in larger establishments is paid 0.117 and 0.216 log-points more than the worker at the 10th percentile in smaller establishments in PIAAC and STEP, respectively. For the median worker, the difference is about 0.106 and 0.143 log-points and it is 0.085 and 0.056 log-points at the 90th percentile. This suggests that the entire wage distribution of large establishments is shifted to the right compared to smaller establishments, even within occupations.

In Appendix C.3, we provide a richer analysis of the ESWP. We find that the existence of an economically significant ESWP is present when we look at more detailed occupational categories or at individual countries.

4.2 Sources of the establishment size wage premium

There are a number of plausible reasons for the existence of the establishment size wage premium. In this subsection, we explore the role in wage determination of (1) sorting by individual characteristics, (2) industry characteristics, and (3) differential task content of jobs. To quantify their relative importance, we conduct a simple mediation analysis adopting the two-step conditional decomposition developed in Gelbach (2016). A desirable feature of his approach is that the results from the decomposition are independent of the order in which the mediators are introduced in the regression. A limitation, however, is that we require measurement of the key mediators to avoid omitted-variable biases. The decomposition begins with a raw estimate of the ESWP, β^{raw}, from the regression:

$$\ln w_i = \beta^{\text{raw}} \times \text{LE}_{j(i)} + \delta^{o,\text{raw}}_{o(i)} + \delta^{c,\text{raw}}_{c(i)} + \varepsilon^{\text{raw}}_i, \tag{3}$$

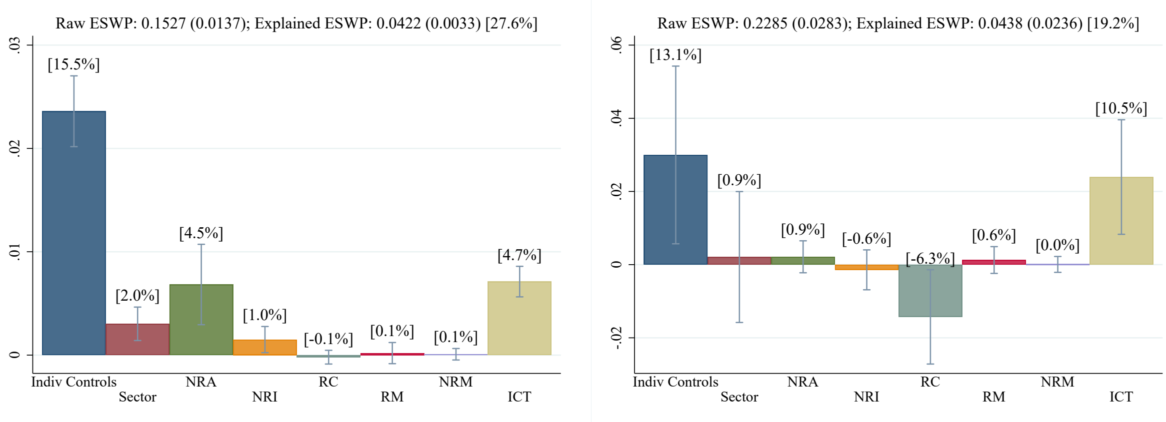
where $\ln w_i$ is log real hourly wages, $\text{LE}_{j(i)}$ is the indicator for worker i being in a large firm, δ^o are occupation fixed effects (2-digit ISCO code), and δ^c are country fixed effects. This raw ESWP estimate coincides with the estimate in column (1) of Table 3. The second step of the decomposition consists of re-estimating the above equation after the inclusion of a set of individual controls X_i that are believed to mediate the LFWP:

$$\ln w_i = \beta^{\text{full}} \times \text{LE}_{j(i)} + X_i'\gamma + \delta^{o,\text{full}}_{o(i)} + \delta^{c,\text{full}}_{c(i)} + \varepsilon^{\text{full}}_i. \tag{4}$$

In our case, the variables incorporated in X are (1) individual characteristics including sex, age and education, (2) sector dummies, and (3) the task content of jobs and usage of ICT reported by workers. In other words, this regression replicates the specification in Table 3's column (2). The difference $\beta^{\text{raw}} - \beta^{\text{full}}$ is interpreted as the part of the ESWP that we are able to explain by controlling for X. Gelbach (2016) uses the formula for the omitted variable bias to apportion the explained part of the ESWP to each of the component variables of X.[18]

Figure 2: Gelbach decomposition of ESWP, pooled

(a) PIAAC (b) STEP



Notes: Pooled PIAAC and STEP samples. Raw ESWP refers to the estimate in Table 3's column (1). Explained ESWP is the difference in the estimate between columns (1) and (2) in that Table. The y-axis is the amount of the ESWP explained by the corresponding component. Numbers in brackets indicate percentages of the raw ESWP. Reported confidence intervals at 95% confidence level. Standard errors are clustered at the country level.

The decomposition results are graphically summarized in Figure 2. We find that the

[18] A more detailed explanation of the methodology is in Appendix D.2.

mediators that we consider are able to explain 27.6% of the raw ESWP in PIAAC and 19.2% in STEP. Individual characteristics (age, sex, and education) explain a significant portion of the ESWP, around 15.5% and 13.1% in PIAAC and STEP, respectively. This suggests that large establishments pay more on average because they hire workers who are older, more educated, and better skilled. This sorting pattern of workers to larger employers has been recently documented by Arellano-Bover (2021). Importantly, though human capital sorting (by occupation or education) explains a large portion, it is unable to fully explain the existence of the ESWP.

The third to eighth bars in both panels of Figure 2 report the fractions of the raw ESWP that are explained by the differences in the tasks performed and in ICT use by the workers documented in Section 3. To help us better interpret the results, in Appendix Table B7, we report the coefficients of the tasks on log wages in the regressions of Table 3. In particular, we document that non-routine analytical, non-routine interpersonal, and the use of ICT have positive returns on wages, whereas routine cognitive and routine manual have negative returns, consistent with existing evidence (Autor and Handel, 2013; Stinebrickner et al., 2019). Something important to notice is that Gelbach (2016)'s decomposition estimates the contribution of each mediator *keeping* the other mediators constant. Hence, while the different task components may be predictive of wages, the variation that explains the ESWP is largely mediated by the variation in the usage of ICT rather than variation in the task content.

Returning to Figure 2, we find that the establishment size gradient in the performance of non-routine analytical tasks explains about 4.5% of the raw establishment size wage premium in PIAAC. Moreover, differences in the use of ICT explain an additional 4.7% of the raw ESWP in PIAAC and 10.5% in STEP. In STEP countries, apart from the large role of ICT, we find that routine cognitive tasks, which are disproportionately undertaken by workers in large establishments but have sizable negative returns on wages, explain -6.3% of the gap.

Overall, we take these results to reflect not only the disproportionately higher intensity with which workers perform various tasks and use ICT in larger establishments, but also the growing importance of computer skills (Alekseeva et al., 2021) in the labor market.[19] Combined, these task components and the use of ICT explain more than 10% of the raw ESWP, a magnitude comparable to that explained jointly by education, age and sex.[20]

The main concern in the performed decomposition analysis is the potential presence of omitted variables. While we do have a rich set of worker controls (including cognitive and non-cognitive skills that should reduce the potential for the presence of unobserved determinants of worker selection into establishments), not observing establishment characteristics, in particular performance, may be a concern. First, the ESWP may be partially driven by differences in establishment productivity—in many models of the labor market, including rent-sharing models or search and matching models, more productive employers pay higher wages to its workers. Unfortunately, we do not observe measures of employer productivity. To partially address this issue, we control for the sector in which the worker works with the aim of accounting for aggregate productivity differences across sectors. We find that the sectoral membership of the worker only partly explains the existence of the ESWP—about 2% of the raw ESWP in PIAAC countries.

A second concern is that the ESWP may be driven by spatial differences in wages. In an attempt to capture within-country spatial differences in wages, we repeat the decomposition including regional fixed effects and report the results in Appendix Figure B5.[21] We find that the regional dummies are able to explain a non-neglible fraction of the ESWP (around 9% and 20% in PIAAC and STEP countries, respectively). Importantly, we show that this does not come at the expense of shifting the importance of the tasks performed, as their importance

[19] A potential concern is that ICT, which explains a large fraction of the ESWP in both samples is itself a mediator of the role of tasks, i.e., after tasks are assigned to workers, ICT use is decided as a function of the tasks. In Appendix Figure B4 we replicate the same analysis excluding ICT as a potential mediator. We find that the fraction of the ESWP that tasks can explain is mostly unaffected. This suggests that ICT use is an independent mechanism in itself.

[20] In STEP countries, routine cognitive tasks explain 6% of the closing of the gap in wages between workers in larger and smaller establishments. This is on top of the differences in the use of ICT explaining 10% of the widening gap in wages between workers in larger and smaller establishments.

[21] We note that in STEP, only urban areas are surveyed which partially alleviates the urban-rural differences we might expect in wages.

in the decompositions remains of about the same size.

Cross-country comparisons. The results of the decomposition exercise by country are graphically summarized in Appendix Figure B6, which focuses on countries for which both the ESWP and the explained portion of the ESWP are statistically significant. We find that the proportion of the raw ESWP explained by the controls that we consider varies between 20% and 40%. In terms of broad patterns, basic individual characteristics such as age, sex, and education consistently explain a significant portion of the raw ESWP (between 10–30%). Sectoral membership is intermittently statistically and economically significant in a handful of countries. In countries where this component accounts for a statistically significant portion, sectors explain around 5–20% of the raw ESWP.

The establishment size gradient in the performance of non-routine analytical tasks and the use of ICT explain, in general, a total of about 5–20% of the raw ESWP. The establishment size gradient in the performance of non-routine analytical tasks explains between 5 and 8% of the raw establishment size wage premium whenever it contributes to a statistically significant share of the ESWP. Among the countries for which the ICT component is statistically significant, the estimates lie mostly between 3 and 9% of the raw ESWP, with a couple of countries where the use of ICT contributes more substantially to the ESWP.[22]

5 Conclusion

In this paper, we document novel stylized facts about the heterogeneity in occupational task intensity across establishments. We find that individuals working in larger establishments report that they perform non-routine analytical tasks more frequently and use ICTs more intensively, even within narrowly-defined occupations. We complement these empirical facts

[22] We also report the version of the decomposition where we additionally include regional fixed effects as mediators by country (Appendix Figure B7). The qualitative results remain and region emerges as a contributor of its own to the ESWP for PIAAC countries.

with demand-side information confirming that larger employers indeed require workers to perform more non-routine analytical and ICT-intensive tasks.

Moreover, we document the existence of an economically significant establishment size wage premium of about 15%. We provide suggestive evidence on the role of task heterogeneity in explaining this ESWP. By controlling for individual characteristics (age, gender, education, cognition, and non-cognition) of the workers, sector, and the task content of jobs, we explain about 28% of the raw ESWP in PIAAC countries and 19% in STEP countries. Differences in the task content of jobs are able to explain over 10% of the raw ESWP. Therefore, accounting for within-occupation heterogeneity in the task content of job enriches our understanding of wage gaps in the labor market.

We consider that our work opens two natural avenues for future research. First, an unresolved question is how these task differences arise in a dynamic economy. In the introduction, we suggested that as employers grow larger, they invest in automation and/or off-shore more work, which transforms the organization of production. These larger establishments engage workers in complementary tasks such as non-routine analytical and routine cognitive tasks. Moreover, these tasks are performed with more ICT. While our results are consistent with this micro-founded mechanism of firm dynamics, it is difficult to establish its consistency with reality in the absence of panel data of employers and tasks.

Second, we leave for future study other implications for the labor market of the establishment size task gradient. We have suggestive evidence of its role in static wage determination but lack exogenous identifying conditions to argue for its causal nature. The implications of our results on dynamic wage determination remain unexplored. More specifically, our results may serve as a nexus between two seemingly parallel strands of the literature. First, a number of studies shows that having experience in certain tasks has different returns in the market: analytical tasks and the use of ICT have been found to have high market returns, especially in recent years (Stinebrickner et al., 2019; Alekseeva et al., 2021). Second, there is evidence that experience in large firms also has higher returns in the market (Arellano-Bover, 2024).

Our results suggest a plausible mechanism for the bigger dynamic returns to working in larger employers — workers in larger employers gain more experience in performing non-routine analytical tasks and the use of ICT, which are highly valued in the labor market.

References

Acemoglu, Daron and David Autor (2011) "Chapter 12 - Skills, Tasks and Technologies: Implications for Employment and Earnings," 4 of Handbook of Labor Economics, 1043–1171: Elsevier, https://doi.org/10.1016/S0169-7218(11)02410-5. (cited in page 1, 9)

Acemoglu, Daron and Pascual Restrepo (2022) "Tasks, automation, and the rise in U.S. wage inequality," *Econometrica*, 90 (5), 1973–2016, 10.3982/ecta19815. (cited in page 1)

Adenbaum, Jacob (2023) "Endogenous Firm Structure and Worker Specialization," Working Paper. (cited in page 3, 22)

Alekseeva, Liudmila, José Azar, Mireia Giné, Sampsa Samila, and Bledi Taska (2021) "The demand for AI skills in the labor market," *Labour Economics*, 71, 102002, https://doi.org/10.1016/j.labeco.2021.102002. (cited in page 1, 23, 29, 31)

Anghel, Brindusa and Pau Balart (2017) "Non-cognitive skills and individual earnings: new evidence from PIAAC," *SERIEs*, 8 (4), 417–473, `https://link.springer.com/article/10.1007/s13209-017-0165-x`. (cited in page 11)

Arellano-Bover, Jaime (2021) "Who Gets Their First Job at a Large Firm? The Distinct Roles of Education and Skills," *AEA Papers and Proceedings*, 111, 465–69, 10.1257/pandp.20211009. (cited in page 28)

——— (2024) "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size," *Journal of Labor Economics*, 42 (2), 549–589, 10.1086/723500. (cited in page 5, 31)

Atalay, Enghin, Sebastian Sotelo, and Daniel Tannenbaum (2021) "The geography of job tasks," `https://enghinatalay.github.io/city_task.pdf`, Working Paper. (cited in page 3)

Autor, David H. and Michael J. Handel (2013) "Putting Tasks to the Test: Human Capital, Job Tasks, and Wages," *Journal of Labor Economics*, 31 (S1), S59–S96, 10.1086/669332. (cited in page 1, 28)

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, 10.1162/003355303322552801. (cited in page 1, 9)

Barro, Robert J. and Jong Wha Lee (2013) "A new data set of educational attainment in the world, 1950–2010," *Journal of Development Economics*, 104, 184–198, https://doi.org/10.1016/j.jdeveco.2012.10.001. (cited in page A14)

Bloom, Nicholas, Fatih Guvenen, Benjamin S. Smith, Jae Song, and Till von Wachter (2018) "The Disappearing Large-Firm Wage Premium," *AEA Papers and Proceedings*, 108, 317–22, 10.1257/pandp.20181066. (cited in page 4)

Bloom, Nicholas and John Van Reenen (2007) "Measuring and Explaining Management Practices Across Firms and Countries*," *The Quarterly Journal of Economics*, 122 (4), 1351–1408, 10.1162/qjec.2007.122.4.1351. (cited in page 1)

Brown, Charles and James Medoff (1989) "The Employer Size-Wage Effect," *Journal of Political Economy*, 97 (5), 1027–1059, 10.1086/261642. (cited in page 4)

Cabrales, Antonio, Juan José Dolado, and Ricardo Mora (2014) "Dual labour markets and (lack of) on-the-job training: PIAAC evidence from Spain and other EU countries," https://cepr.org/publications/dp10246. (cited in page 11)

Caunedo, Julieta, Elisa Keller, and Yongseok Shin (2023) "Technology and the task content of jobs across the development spectrum," *The World Bank Economic Review*, 37 (3), 479–493. (cited in page 2, 3, 9, 15, A1, A3)

Chernozhukov, Victor, Iván Fernández-Val, and Blaise Melly (2013) "Inference on Counterfactual Distributions," *Econometrica*, 81 (6), 2205–2268, https://doi.org/10.3982/ECTA10582. (cited in page 20, A22)

Colonnelli, Emanuele, Joacim Tåg, Michael Webb, and Stefanie Wolter (2018) "A Cross-Country Comparison of Dynamics in the Large Firm Wage Premium," *AEA Papers and Proceedings*, 108, 323–27, 10.1257/pandp.20181067. (cited in page 4)

De La Rica, Sara, Lucas Gortazar, and Piotr Lewandowski (2020) "Job Tasks and Wages in Developed Countries: Evidence from PIAAC," *Labour Economics*, 65, 101845, https://doi.org/10.1016/j.labeco.2020.101845. (cited in page 3)

Deming, David and Lisa B. Kahn (2018) "Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals," *Journal of Labor Economics*, 36 (S1), S337–S369, 10.1086/694106. (cited in page 3)

Dicarlo, Emanuele, Salvatore Lo Bello, Sebastian Monroy-Taborda, Ana Maria Oviedo, Maria Laura Sanchez Puerta, and Indhira Vanessa Santos (2016) "The skill content of occupations across low and middle income countries: evidence from harmonized data," Working Paper. (cited in page 3)

Dobbelaere, Sabien (2004) "Ownership, firm size and rent sharing in Bulgaria," *Labour Economics*, 11 (2), 165–189, https://doi.org/10.1016/j.labeco.2003.07.001. (cited in page 4)

Fonseca, Tiago, Francisco Lima, and Sonia C. Pereira (2018) "Job polarization, technological change and routinization: Evidence for Portugal," *Labour Economics*, 51, 317–339, https://doi.org/10.1016/j.labeco.2018.02.003. (cited in page 9)

Gelbach, Jonah B. (2016) "When Do Covariates Matter? And Which Ones, and How Much?" *Journal of Labor Economics*, 34 (2), 509–543, 10.1086/683668. (cited in page 26, 27, 28, A31, A33)

Gerlach, Knut and Olaf Hübler (1998) "Firm Size and Wages in Germany – Trends and Impacts of Mobility," *Empirica*, 25 (3), 245–261, 10.1023/A:1006904430572. (cited in page 4)

Giuseppe De Luca, Jan R. Magnus and Franco Peracchi (2019) "Comments on "Unobservable Selection and Coefficient Stability: Theory and Evidence" and "Poorly Measured Confounders are More Useful on the Left Than on the Right"," *Journal of Business & Economic Statistics*, 37 (2), 217–222, 10.1080/07350015.2019.1575743. (cited in page A32)

Goos, Maarten, Alan Manning, and Anna Salomons (2014) "Explaining Job Polarization: Routine-Biased Technological Change and Offshoring," *American Economic Review*, 104 (8), 2509–26, 10.1257/aer.104.8.2509. (cited in page 1, 9)

Jaimovich, Nir, Miao Ben Zhang, and Nicolas Vincent (2023) "Under the hood of the routine share decline," *Economics Letters*, 111437, https://doi.org/10.1016/j.econlet.2023.111437. (cited in page 3)

Khorramdel, Lale, Matthias von Davier, Eugenio Gonzalez, and Kentaro Yamamoto (2020) *Plausible Values: Principles of Item Response Theory and Multiple Imputations*, 27–47, Cham: Springer International Publishing, 10.1007/978-3-030-47515-4_3. (cited in page 10)

Lashkari, Danial, Arthur Bauer, and Jocelyn Boussard (2024) "Information Technology and Returns to Scale," *American Economic Review*, 114 (6), 1769–1815, 10.1257/aer.20220522. (cited in page 23)

Lehmer, Florian and Joachim Möller (2010) "Interrelations between the urban wage premium and firm-size wage differentials: a microdata cohort analysis for Germany," *The Annals of Regional Science*, 45 (1), 31–53, 10.1007/s00168-009-0290-y. (cited in page 4)

Lewandowski, Piotr, Albert Park, Wojciech Hardy, and Yang Du (2019) "Technology, Skills, and Globalization: Explaining International Differences in Routine and Nonroutine Work Using Survey Data," Working Paper. (cited in page 3)

Lochner, Benjamin, Stefan Seth, and Stefanie Wolter (2020) "Decomposing the large firm wage premium in Germany," *Economics Letters*, 194, 109368, https://doi.org/10.1016/j.econlet.2020.109368. (cited in page 4)

Lucas, Robert E. (1978) "On the Size Distribution of Business Firms," *The Bell Journal of Economics*, 9 (2), 508–523, `http://www.jstor.org/stable/3003596`. (cited in page 1)

Ocampo, Sergio (2022) "A Task-Based Theory of Occupations with Multidimensional Heterogeneity," Working Paper. (cited in page 1, 3)

Oi, Walter Y. and Todd L. Idson (1999) "Chapter 33 Firm size and wages," 3B of Handbook of Labor Economics, 2165–2214: Elsevier, https://doi.org/10.1016/S1573-4463(99)30019-5. (cited in page 4)

Oster, Emily (2019) "Unobservable selection and coefficient stability: Theory and evidence," *Journal of Business & Economic Statistics*, 37 (2), 187–204, https://emilyoster.net/wp-content/uploads/UnobservableSelectionandCoefficientStabilityTheoryandEvidence.pdf. (cited in page 2, 13, 15, 16, A6, A31, A32)

Porcher, Charly, Hannah Rubinton, and Clara Santamaría (2023) "JUE insight: The role of establishment size in the city-size earnings premium," *Journal of Urban Economics*, 136, 103556, https://doi.org/10.1016/j.jue.2023.103556. (cited in page 4)

Reed, Tristan and Trang Thu (2019) "The Large-Firm Wage Premium in Developing Economies," https://openknowledge.worldbank.org/handle/10986/32371, Working Paper. (cited in page 4)

Schaffner, Julie Anderson (1998) "Premiums to employment in larger establishments: evidence from Peru," *Journal of Development Economics*, 55 (1), 81–113, https://doi.org/10.1016/S0304-3878(98)00038-8. (cited in page 4)

Spitz-Oener, Alexandra (2006) "Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure," *Journal of Labor Economics*, 24 (2), 235–270, 10.1086/499972. (cited in page 9)

Stinebrickner, Ralph, Todd Stinebrickner, and Paul Sullivan (2019) "Job Tasks, Time Allocation, and Wages," *Journal of Labor Economics*, 37 (2), 399–433, 10.1086/700186. (cited in page 3, 5, 28, 31)

Söderbom, Måns, Francis Teal, and Anthony Wambugu (2005) "Unobserved heterogeneity and the relation between earnings and firm size: evidence from two developing countries," *Economics Letters*, 87 (2), 153–159, https://doi.org/10.1016/j.econlet.2004.09.012. (cited in page 4)

Troske, Kenneth R. (1999) "Evidence on the Employer Size-Wage Premium from Worker-Establishment Matched Data," *The Review of Economics and Statistics*, 81 (1), 15–26, 10.1162/003465399557950. (cited in page 4)

Velenchik, Ann D. (1997) "Government intervention, efficiency wages, and the employer size wage effect in Zimbabwe," *Journal of Development Economics*, 53 (2), 305–338, https://doi.org/10.1016/S0304-3878(97)00019-9. (cited in page 4)

Winter-Ebmer, Rudolf (2001) "Firm Size, Earnings, and Displacement Risk," *Economic Inquiry*, 39 (3), 474–486, https://doi.org/10.1093/ei/39.3.474. (cited in page 4)

A Appendix: Data availability, measurement, and summary statistics

Data availability by country. For the reader's convenience, we report data availability for each of the 46 countries employed in our study:

- PIAAC
  - (Worker-based) tasks + continuous wages: Belgium, Chile, Cyprus, Czech Republic, Denmark, Ecuador, France, Greece, Ireland, Israel, Italy, Japan, Kazakhstan, South Korea, Lithuania, Mexico, Netherlands, Norway, Poland, Russian Federation, Slovak Republic, Slovenia, Spain, Sweden, and United Kingdom.
  - (Worker-based) tasks + bin-based wages (hence excluded from the ESWP analyses): Hungary, New Zealand, Peru, Singapore, and the United States (both rounds).
  - (Worker-based) tasks and no wages: Peru.
- STEP
  - Task intensity based on worker-based survey + wages: Bolivia, China (Yunnan province), Colombia, Georgia, Laos, Macedonia, Sri Lanka, and Ukraine.
  - Task intensity based both on worker- and employer-based surveys + wages: Armenia, Kenya, and Vietnam.
  - Task intensity only based on employer-based surveys (and no wages): Albania, Azerbaijan, Bosnia-Herzegovina, Kosovo, and Serbia.

Construction of task content measures using worker-based surveys. Table A1 summarizes the mapping, following Caunedo et al. (2023), of the questions in PIAAC and STEP to the different task dimension we are interested in: non-routine analytical, non-routine interpersonal, routine cognitive, routine manual, non-routine manual, and the usage of ICT.

Our preferred choice of variable construction standardizes each subcomponent of a task category to have a mean of 0 and a standard deviation of 1 across all the respondents in a given country. Then, all the subcomponents of the category are used to obtained the simple average (i.e., equal weights assigned to each subcomponent). The resulting mean for each task category is once again standardized to have a mean of 0 and a standard deviation of 1 across the respondents in the country.

Construction of task content measures using employer-based surveys. In the employers' questionnaire, a knowledgeable person was asked about the task requirements for two randomly selected occupations. One of them would be chosen among the following ones: manager/professional/technician while the second one would be selected from: clerk/services/sales/crafting/operator/elementary occupation. We are able to obtain a measure of employer-specific demand that closely matches the construction of our non-routine analytical and ICT use measures from the workers' survey, but not for the other task categories. We follow the same procedure as for the worker-based skill intensity measures and construct employer requirements as the standardized value (mean of 0 and standard deviation of 1) of the simple mean of the standardized scores in each of the following questions:

- Non-routine analytical: (a) does the job involve reading? (m_30x_1); (b) does the job involve writing using correct spelling and grammar? (m_30x_2); (c) does the job involve math? (m_30x_3); (d) does the job involve solving problems that take 30 minutes or more to solve? (m_30x_4), and (e) does the job involve speaking other languages? (m_30x_5).[2] (possible answers were yes/no)

- ICT: what is the highest level of computer use involved in this job? (possible responses were: none, straightforward, moderate, complex, and specialized) (m_3_08).

[2] Variable names are based on the Albanian survey. "x" stands for either "a" or "b."

Table A1: Mapping of survey questionnaires to task categories

	STEP surveys		PIAAC surveys	
Task category	Item description	Item nos.	Item description	Item nos.
Non-routine analytical	Type of documents read and frequency	A-4, A-5-(1-6)	Type of documents read and frequency	G_Q01(a-h)
	Think creatively	B-10	Think creatively	F_Q05b
Non-routine interpersonal	Personal relationship	B-5, B-6	Personal relationship	F_Q02a, F_Q02d, FQ_04a, FQ_04b
	Guiding/coaching	B-13	Guiding/coaching	F_Q02b, F_Q02e, F_Q03b
Routine cognitive	Freedom how to decide work	B-14	Planning activities	FQ_03a
	Presence of short, repetitive tasks	B-16	Organizing own time	FQ_03c
	Learning new things	B-17		
Routine manual	Physical demand	B-3	Long physical work	FQ_06b
Non-routine manual	Driving car. truck, three-wheeler	B-7	Use/accuracy hand/fingers	FQ_06c
	Repair/maintain electronic equip.	B-8		
Use of ICT	Used a computer	B-18	Used a computer	G_Q04

Notes: For STEP countries, we diverge from Caunedo et al. (2023) in constructing our measure of routine manual tasks by not including the category of operating heavy machinery, which does not have a clear counterpart in PIAAC. If available, we also employ the question "As a regular part of this work, do you have to read the following...Other?" (A-5-7) as an additional subcomponent in the construction of the non-routine analytical measure for STEP countries.

Table A2: Summary statistics

Variable	New Zealand	United Kingdom	Slovak Republic	Russia	Czech Republic	Mexico	Lithuania	Kazakhstan	Israel
Large Establishment	0.368 (0.482)	0.499 (0.5)	0.434 (0.496)	0.466 (0.499)	0.376 (0.484)	0.318 (0.466)	0.456 (0.498)	0.364 (0.481)	0.369 (0.483)
Log Earnings	2.812 (0.509)	2.772 (0.566)	1.579 (0.672)	0.693 (0.775)	1.703 (0.578)	0.811 (0.874)	2.595 (0.597)	0.288 (0.593)	2.492 (0.809)
	-0.223 - 10.675	0.387 - 6.784	0.209 - 7.617	-3.63 - 5.49	-2.573 - 6.793	-2.051 - 5.01	0.334 - 9.092	-3.185 - 4.362	-0.498 - 8.849
Female	0.576 (0.494)	0.592 (0.491)	0.507 (0.5)	0.637 (0.481)	0.513 (0.5)	0.386 (0.487)	0.608 (0.488)	0.568 (0.495)	0.522 (0.5)
Age (16-24)	0.204 (0.403)	0.128 (0.334)	0.093 (0.291)	0.220 (0.415)	0.147 (0.354)	0.216 (0.412)	0.088 (0.283)	0.106 (0.307)	0.280 (0.449)
Age (25-34)	0.191 (0.393)	0.237 (0.426)	0.254 (0.435)	0.278 (0.448)	0.292 (0.455)	0.267 (0.442)	0.209 (0.407)	0.317 (0.466)	0.328 (0.47)
Age (35-44)	0.223 (0.416)	0.254 (0.435)	0.251 (0.433)	0.200 (0.400)	0.233 (0.423)	0.257 (0.437)	0.234 (0.423)	0.287 (0.452)	0.239 (0.427)
Age (45-54)	0.199 (0.399)	0.229 (0.421)	0.257 (0.437)	0.195 (0.396)	0.188 (0.391)	0.166 (0.372)	0.273 (0.446)	0.208 (0.406)	0.104 (0.305)
Age (above 55)	0.183 (0.387)	0.152 (0.359)	0.145 (0.352)	0.106 (0.308)	0.140 (0.347)	0.094 (0.291)	0.196 (0.397)	0.082 (0.275)	0.048 (0.214)
Education (Primary or less)	0.013 (0.113)	0.078 (0.268)	0.002 (0.041)	0.006 (0.075)	0.001 (0.029)	0.22 (0.415)	0.004 (0.065)	0.003 (0.056)	0.048 (0.214)
Education (up to professional tertiary education degree)	0.578 (0.494)	0.493 (0.5)	0.766 (0.423)	0.235 (0.424)	0.752 (0.432)	0.614 (0.487)	0.605 (.489)	0.534 (0.499)	0.470 (0.499)
Education (bachelor and above)	0.409 (0.492)	0.429 (0.495)	0.232 (0.422)	0.758 (0.428)	0.247 (0.431)	0.166 (0.372)	0.391 (0.488)	0.463 (0.499)	0.481 (0.5)
Observations	2,461	3,984	2,355	1,412	2,412	2,241	2,564	2,265	1,182

Variable	Greece	Ecuador	Chile	Spain	Slovenia	Poland	Norway	Netherlands	Korea
Large Establishment	0.223 (0.416)	0.248 (0.432)	0.361 (0.48)	0.348 (0.476)	0.526 (0.499)	0.375 (0.484)	0.446 (0.497)	0.483 (0.5)	0.378 (0.485)
Log Earnings	2.058 (0.524)	1.117 (0.651)	1.518 (0.783)	2.518 (0.566)	2.223 (0.427)	1.335 (0.604)	3.381 (0.429)	3.057 (0.616)	4.779 (0.733)
	0.324 - 3.892	-0.002 - 4.066	-0.397 - 6.467	0.392 - 9.418	0.967 - 4.277	-1.148 - 10.1	0.495 - 6.559	0.46 - 12.948	2.331 - 8.332
Female	0.567 (0.496)	0.389 (0.488)	0.493 (0.5)	0.486 (0.5)	0.516 (0.5)	0.446 (0.497)	0.518 (0.5)	0.506 (0.5)	0.46 (0.498)
Age (16-24)	0.077 (0.267)	0.207 (0.405)	0.154 (0.362)	0.075 (0.264)	0.042 (0.201)	0.396 (0.489)	0.169 (0.375)	0.173 (0.378)	0.096 (0.294)
Age (25-34)	0.258 (0.438)	0.313 (0.464)	0.284 (0.451)	0.248 (0.432)	0.246 (0.431)	0.313 (0.464)	0.193 (0.395)	0.185 (0.388)	0.265 (0.441)
Age (35-44)	0.324 (0.468)	0.235 (0.424)	0.207 (0.405)	0.305 (0.461)	0.295 (0.456)	0.119 (0.324)	0.246 (0.431)	0.232 (0.422)	0.286 (0.452)
Age (45-54)	0.277 (0.448)	0.156 (0.363)	0.226 (0.418)	0.261 (0.439)	0.329 (0.47)	0.119 (0.313)	0.232 (0.422)	0.249 (0.432)	0.239 (0.427)
Age (above 55)	0.064 (0.245)	0.089 (0.285)	0.129 (0.335)	0.111 (0.314)	0.088 (0.284)	0.062 (0.240)	0.16 (0.367)	0.162 (0.368)	0.115 (0.319)
Education (Primary or less)	0.054 (0.226)	0.304 (0.460)	0.076 (0.266)	0.132 (0.339)	0.004 (0.067)	0.009 (0.095)	0.001 (0.028)	0.05 (0.219)	0.05 (0.217)
Education (up to professional tertiary education degree)	0.534 (0.499)	0.536 (0.499)	0.624 (0.484)	0.424 (0.494)	0.608 (0.488)	0.67 (0.47)	0.555 (0.497)	0.606 (0.489)	0.463 (0.499)
Education (bachelor and above)	0.412 (0.492)	0.160 (0.367)	0.3 (0.458)	0.444 (0.497)	0.388 (0.487)	0.321 (0.467)	0.444 (0.497)	0.344 (0.475)	0.487 (0.5)
Observations	857	1,502	2,149	2,023	1,786	3,821	2,569	2,860	3,008

Notes: All variables except log earnings are binary. Means are reported with standard deviations in parentheses. For log earnings, the min and max are reported on a separate row.

Table A2: Summary statistics (cont.)

Variable	Japan	Italy	Ireland	France	Denmark	Cyprus	Belgium	Vietnam	Ukraine
Large Establishment	0.428 (0.495)	0.376 (0.485)	0.405 (0.491)	0.45 (0.498)	0.458 (0.498)	0.310 (0.463)	0.532 (0.499)	0.416 (0.493)	0.564 (0.496)
Log Earnings	2.638 (0.631)	2.714 (0.516)	3.051 (0.566)	2.782 (0.445)	3.408 (0.439)	2.544 (0.574)	3.122 (0.42)	-6.890 (3.133)	-0.002 (0.612)
	-3.016 - 7.841	0.743 - 5.607	0.179 - 7.333	0.4 - 5.747	0.125 - 6.604	0.683 - 5.836	0.565 - 6.895	-60.477 - -0.61	-4.089 - 3.012
Female	0.489 (0.5)	0.48 (0.5)	0.586 (0.493)	0.499 (0.5)	0.506 (0.5)	0.603 (0.490)	0.496 (0.5)	0.541 (0.499)	0.626 (0.484)
Age (16-24)	0.109 (0.312)	0.051 (0.22)	0.088 (0.283)	0.089 (0.285)	0.119 (0.324)	0.081 (0.274)	0.098 (0.297)	0.117 (0.322)	0.063 (0.244)
Age (25-34)	0.203 (0.402)	0.203 (0.402)	0.266 (0.442)	0.229 (0.42)	0.129 (0.336)	0.325 (0.469)	0.248 (0.432)	0.31 (0.463)	0.263 (0.441)
Age (35-44)	0.268 (0.443)	0.345 (0.475)	0.304 (0.46)	0.263 (0.44)	0.207 (0.405)	0.257 (0.437)	0.245 (0.43)	0.293 (0.455)	0.23 (0.421)
Age (45-54)	0.224 (0.417)	0.278 (0.448)	0.204 (0.403)	0.278 (0.448)	0.231 (0.421)	0.215 (0.411)	0.296 (0.456)	0.202 (0.401)	0.265 (0.442)
Age (above 55)	0.197 (0.397)	0.124 (0.329)	0.139 (0.346)	0.143 (0.35)	0.314 (0.464)	0.121 (0.326)	0.113 (0.317)	0.078 (0.268)	0.179 (0.383)
Education (Primary or less)	0 (0.018)	0.033 (0.179)	0.035 (0.183)	0.029 (0.167)	0.004 (0.06)	0.048 (0.214)	0.02 (0.141)	0.133 (0.34)	0.002 (0.044)
Education (up to professional tertiary education degree)	0.480 (0.5)	0.743 (0.437)	0.487 (0.5)	0.573 (0.495)	0.538 (0.499)	0.386 (0.487)	0.54 (0.499)	0.433 (0.496)	0.453 (0.498)
Education (bachelor and above)	0.520 (0.5)	0.223 (0.417)	0.478 (0.5)	0.399 (0.49)	0.458 (0.498)	0.567 (0.496)	0.44 (0.496)	0.434 (0.496)	0.545 (0.498)
Observations	3,149	1,535	2,227	2,993	3,607	1,608	2,470	1,289	521

Variable	Sri Lanka	Macedonia	Laos	Kenya	Georgia	Colombia	China	Bolivia	Armenia
Large Establishment	0.367 (0.483)	0.421 (0.494)	0.289 (0.454)	0.229 (0.420)	0.387 (0.488)	0.347 (0.476)	0.397 (0.49)	0.270 (0.444)	0.407 (0.492)
Log Earnings	-0.016 (0.949)	0.505 (1.63)	-0.173 (0.766)	0.045 (1.079)	0.122 (0.822)	0.487 (0.838)	0.494 (0.689)	0.618 (2.335)	-6.83 (0.745)
	-2.205 - 3.835	-54,59 - 3.924	-3.488 - 3.958	-4.427 - 4.417	-2.744 - 3.296	-7.503 - 4.517	-2.638 - 3.747	-52.352 - 4.324	-8.545 - 0.690
Female	0.422 (0.495)	0.464 (0.499)	0.408 (0.492)	0.385 (0.487)	0.688 (0.464)	0.430 (0.495)	0.499 (0.5)	0.491 (0.5)	0.664 (0.473)
Age (16-24)	0.098 (0.297)	0.044 (0.205)	0.128 (0.334)	0.248 (0.432)	0.085 (0.280)	0.183 (0.387)	0.067 (0.25)	0.207 (0.406)	0.105 (0.307)
Age (25-34)	0.25 (0.434)	0.275 (0.447)	0.336 (0.473)	0.454 (0.498)	0.213 (0.410)	0.357 (0.48)	0.26 (0.439)	0.345 (0.476)	0.228 (0.42)
Age (35-44)	0.348 (0.477)	0.285 (0.451)	0.303 (0.46)	0.184 (0.388)	0.287 (0.453)	0.235 (0.424)	0.384 (0.487)	0.26 (0.439)	0.221 (0.415)
Age (45-54)	0.199 (0.4)	0.235 (0.424)	0.175 (0.38)	0.083 (0.276)	0.246 (0.431)	0.168 (0.374)	0.235 (0.424)	0.123 (0.328)	0.246 (0.431)
Age (above 55)	0.105 (0.308)	0.162 (0.369)	0.058 (0.235)	0.031 (0.173)	0.169 (0.375)	0.057 (0.231)	0.054 (0.226)	0.065 (0.246)	0.2 (0.401)
Education (Primary or less)	0.098 (0.297)	0.008 (0.087)	0.15 (0.358)	0.271 (0.444)	0 (0)	0.209 (0.407)	0.058 (0.233)	0.091 (0.288)	0.003 (0.056)
Education (up to professional tertiary education degree)	0.715 (0.452)	0.653 (0.476)	0.3489 (0.501)	0.559 (0.497)	0.213 (0.41)	0.442 (0.497)	0.556 (0.497)	0.426 (0.495)	0.263 (0.441)
Education (bachelor and above)	0.888 (0.391)	0.34 (0.474)	0.361 (0.481)	0.171 (0.376)	0.787 (0.41)	0.35 (0.477)	0.386 (0.487)	0.483 (0.5)	0.734 (0.442)
Observations	256	1,325	360	1,131	586	761	868	603	639

Notes: All variables except log earnings are binary. Means are reported with standard deviations in parentheses. For log earnings, the min and max are reported on a separate row.

B Appendix: Additional tables and figures

Table B1: Pooled estimates of establishment size gradient in the task content of jobs

Outcome variable	(1)	(2)	(3)	(4)	(5)	(6)	(7) $\hat{\delta}$
Panel (a): PIAAC							
Non-routine analytical	0.155***	0.159***	0.127***	0.117***	0.114***	0.111***	1.387
	(0.019)	(0.019)	(0.016)	(0.017)	(0.017)	(0.017)	
Non-routine interpersonal	0.074***	0.086***	0.064***	0.055**	0.053**	0.048**	0.796
	(0.022)	(0.023)	(0.022)	(0.022)	(0.022)	(0.022)	
Routine cognitive	0.008	0.006	0.024	0.035**	0.031*	0.039**	−0.581
	(0.015)	(0.015)	(0.015)	(0.014)	(0.016)	(0.015)	
Routine manual	-0.006	-0.000	0.006	0.011	0.019	0.008	−0.172
	(0.015)	(0.014)	(0.011)	(0.010)	(0.012)	(0.011)	
Non-routine manual	-0.007	-0.008	-0.001	0.001	0.006	-0.005	−0.025
	(0.017)	(0.017)	(0.015)	(0.014)	(0.015)	(0.013)	
Use of ICT	0.156***	0.164***	0.140***	0.132***	0.126***	0.121***	1.177
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	
Sample size	65,151	65,151	65,151	65,151	65,151	65,151	
Panel (b): STEP							
Non-routine analytical	0.125***	0.129***	0.069**	0.066**	0.067**	0.081**	0.381
	(0.032)	(0.030)	(0.024)	(0.023)	(0.027)	(0.029)	
Non-routine interpersonal	-0.015	-0.003	-0.033	-0.035	-0.062*	-0.039	−0.292
	(0.030)	(0.033)	(0.037)	(0.034)	(0.033)	(0.028)	
Routine cognitive	0.169**	0.162**	0.188**	0.191***	0.226***	0.194***	−2.359
	(0.067)	(0.066)	(0.064)	(0.060)	(0.052)	(0.053)	
Routine manual	-0.020	-0.019	0.007	0.007	0.011	0.022	−0.066
	(0.034)	(0.035)	(0.032)	(0.033)	(0.022)	(0.028)	
Non-routine manual	-0.048	-0.042	-0.060	-0.060	-0.093**	-0.086**	−2.683
	(0.038)	(0.038)	(0.037)	(0.035)	(0.031)	(0.031)	
Use of ICT	0.194***	0.201***	0.126**	0.125**	0.106**	0.102**	0.644
	(0.057)	(0.056)	(0.043)	(0.043)	(0.042)	(0.043)	
Sample size	8,339	8,339	8,339	8,339	8,339	8,339	
Controls:							
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	
2-d Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes	
Industry FE		Yes	Yes	Yes	Yes	Yes	
Individual demographics			Yes	Yes	Yes	Yes	
Individual cognition/noncog.				Yes	Yes	Yes	
Region FE					Yes	Yes	
Country interactions						Yes	

*Notes: Extension of Table 1 employing different sets of controls. Regressions of a given category of task requirement intensity on an indicator of large establishment (at least 50 employees). Each row refers to a specific skill category. Additional controls are sequentially included across columns and are indicated in the lower part of the table. Individual demographics include education, gender, and age. Regressions are conducted separately for the pooled sample of PIAAC and STEP countries in panel (a) and (b), respectively. Column (7) reports estimates of Oster (2019)'s δ with column (2) as the short regression and column (4) as the long regression. Standard errors are reported in parenthesis and clustered at the country level. *** p<0.01, ** p<0.05, * p<0.1*

Table B2: Pooled estimates of establishment size gradient in the task content of jobs when controlling for 3-digit occupations

Outcome variable	(1)	(2)	(3)	(4)	(5)	(6)
Panel (a): PIAAC						
Non-routine analytical	0.155***	0.157***	0.131***	0.122***	0.117***	0.115***
	(0.020)	(0.021)	(0.019)	(0.019)	(0.020)	(0.020)
Non-routine interpersonal	0.082***	0.092***	0.072***	0.064**	0.060**	0.055**
	(0.023)	(0.025)	(0.024)	(0.024)	(0.024)	(0.024)
Routine cognitive	-0.002	-0.001	-0.014	0.024	0.022	0.020
	(0.023)	(0.024)	(0.024)	(0.023)	(0.023)	(0.025)
Routine manual	0.007	0.014	0.018	0.022	0.029	0.025
	(0.019)	(0.019)	(0.016)	(0.016)	(0.019)	(0.019)
Non-routine manual	0.013	0.016	0.020	0.022	0.029*	0.020
	(0.016)	(0.016)	(0.015)	(0.015)	(0.014)	(0.015)
Use of ICT	0.146***	0.150***	0.129***	0.122***	0.113***	0.103***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)	(0.007)
Sample size	47,905	47,905	47,905	47,905	47,905	47,905
Panel (b): STEP						
Non-routine analytical	0.132***	0.130***	0.074**	0.072***	0.095**	0.106**
	(0.030)	(0.029)	(0.024)	(0.021)	(0.033)	(0.037)
Non-routine interpersonal	-0.006	-0.003	-0.032	-0.034	-0.063	-0.027
	(0.035)	(0.037)	(0.042)	(0.041)	(0.042)	(0.040)
Routine cognitive	0.127*	0.126*	0.157**	0.159**	0.169**	0.152**
	(0.062)	(0.059)	(0.055)	(0.049)	(0.064)	(0.066)
Routine manual	-0.016	-0.016	-0.011	0.012	-0.007	0.006
	(0.038)	(0.037)	(0.033)	(0.034)	(0.020)	(0.022)
Non-routine manual	-0.003	-0.002	-0.031	-0.032	-0.111**	-0.086
	(0.052)	(0.053)	(0.048)	(0.046)	(0.048)	(0.058)
Use of ICT	0.211**	0.218***	0.136**	0.135**	0.105	0.084
	(0.064)	(0.064)	(0.054)	(0.052)	(0.060)	(0.056)
Sample size	4,664	4,664	4,664	4,664	4,664	4,664
Controls:						
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
3-d Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE		Yes	Yes	Yes	Yes	Yes
Individual demographics			Yes	Yes	Yes	Yes
Individual cognition/noncog.				Yes	Yes	Yes
Region FE					Yes	Yes
Country interactions						Yes

*Notes: Replication of Table 1 substituting 2-digit occupation fixed effects by 3-digit ones. The unavailability of this finer degree of information for some countries explains the difference in observations with respect to the main table. *** p<0.01, ** p<0.05, * p<0.1*

Table B3: Pooled estimates of establishment size gradient in the task content of jobs when not discarding countries without region

Outcome variable	(1)	(2)	(3)	(4)	(5)
Non-routine analytical	0.143***	0.145***	0.117***	0.109***	0.105***
	(0.015)	(0.015)	(0.013)	(0.013)	(0.013)
Non-routine interpersonal	0.045*	0.054**	0.036	0.029	0.025
	(0.023)	(0.025)	(0.023)	(0.022)	(0.023)
Routine cognitive	0.015	0.013	0.036	0.046**	0.051***
	(0.016)	(0.016)	(0.018)	(0.017)	(0.016)
Routine manual	-0.034*	-0.030	-0.019	-0.013	-0.021
	(0.018)	(0.019)	(0.016)	(0.015)	(0.013)
Non-routine manual	-0.022	-0.022	-0.013	-0.010	-0.014
	(0.014)	(0.014)	(0.013)	(0.012)	(0.010)
Use of ICT	0.155***	0.162***	0.141***	0.133***	0.128***
	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)
Sample size	73,292	73,292	73,292	73,292	73,292
Controls:					
Country FE	Yes	Yes	Yes	Yes	Yes
2-d Occupation FE	Yes	Yes	Yes	Yes	Yes
Industry FE		Yes	Yes	Yes	Yes
Individual demographics			Yes	Yes	Yes
Individual cognition/noncog.				Yes	Yes
Country interactions					Yes

*Notes: Replication of Table B1 when not discarding PIAAC countries without regional information (Italy, Norway, and United States). This explains the increase in available observations. *** p<0.01, ** p<0.05, * p<0.1*

Table B4: Establishment size gradient in the task content of jobs with more detailed treatment of establishment size

	(1) NRA	(2) NRI	(3) RC	(4) RM	(5) NRM	(6) COMP
Panel (a): PIAAC						
Firm size: 11–50	0.086***	0.061***	0.027	0.065***	-0.013	0.095***
	(0.012)	(0.019)	(0.028)	(0.018)	(0.016)	(0.016)
Firm size: 51–250	0.135***	0.076***	0.045	0.072***	-0.009	0.154***
	(0.016)	(0.020)	(0.032)	(0.020)	(0.014)	(0.013)
Firm size: 250–1,000	0.162***	0.060**	0.093***	0.025	-0.007	0.216***
	(0.028)	(0.027)	(0.026)	(0.019)	(0.015)	(0.019)
Firm size: >1,000	0.272***	0.178***	0.003	0.007	0.003	0.246***
	(0.034)	(0.051)	(0.032)	(0.032)	(0.046)	(0.029)
Observations	65,151	65,151	65,151	65,151	65,151	65,151
R-squared	0.353	0.288	0.160	0.265	0.093	0.399
Panel (b): STEP						
Firm size: 2–5	0.055	0.401***	-0.078	0.145**	0.110	0.036
	(0.072)	(0.047)	(0.107)	(0.050)	(0.078)	(0.065)
Firm size: 6–15	0.087	0.256***	0.115	0.092	-0.012	0.179***
	(0.069)	(0.031)	(0.105)	(0.057)	(0.076)	(0.048)
Firm size: 16–25	0.134*	0.096**	0.160**	0.064	0.027	0.194**
	(0.068)	(0.033)	(0.071)	(0.046)	(0.070)	(0.062)
Firm size: 26–50	0.166**	0.197***	0.299***	0.127**	-0.054	0.273***
	(0.060)	(0.047)	(0.079)	(0.042)	(0.081)	(0.053)
Firm size: 51–200	0.146*	0.198***	0.322***	0.094	-0.013	0.271***
	(0.066)	(0.055)	(0.065)	(0.065)	(0.067)	(0.073)
Firm size: >200	0.220***	0.190***	0.280**	0.119*	-0.070	0.346***
	(0.066)	(0.051)	(0.090)	(0.053)	(0.097)	(0.106)
Observations	7,818	7,818	7,818	7,818	7,818	7,818
R-squared	0.408	0.297	0.175	0.218	0.221	0.492

*Notes: Replication of Table 1's column (2) where the large-establishment indicator has been replaced by multiple indicators for whether the establishment's number of employees falls into specific ranges. These ranges are not consistent between PIAAC and STEP, and the omitted category is having up to 10 employees in PIAAC and being the sole employee in STEP. The sample size for STEP is decreased relative to Table 1 because we have excluded Ukraine from the sample (the elicited firm size categories could not be homogenized with those of the other STEP countries). Standard errors are reported in parenthesis and clustered at the country level. *** p<0.01, ** p<0.05, * p<0.1*

Table B5: Establishment size gradient in the task content of jobs, young workers with short tenure, pooled estimates

	PIAAC		STEP	
Task Category	Short Tenure	+ Age < 25	Short Tenure	+ Age < 25
NRA	0.115***	0.096***	0.070	0.135*
	(0.015)	(0.020)	(0.054)	(0.064)
NRI	0.041	0.084	-0.109	-0.166
	(0.031)	(0.058)	(0.045)	(0.100)
RC	0.029	-0.030	0.168**	0.221***
	(0.031)	(0.086)	(0.059)	(0.046)
RM	-0.010	0.033	0.017	0.219***
	(0.018)	(0.032)	(0.037)	(0.025)
NRM	-0.003	0.128***	-0.025*	0.120
	(0.017)	(0.045)	(0.011)	(0.069)
Use of ICT	0.121***	0.230***	0.183**	0.145
	(0.013)	(0.030)	(0.076)	(0.088)
# of Observations	28,220	5,716	2,656	810

*Notes: Pooled PIAAC and STEP samples. Coefficient of an indicator for large establishment in a regression of task content intensity on indicator of large establishment (at least 50 employees) controlling for the set of controls in Table 1's column (2). The first and third columns restrict the sample to workers with short tenure. In PIAAC there is no direct question about tenure, so we proxy short-tenure by an individual having worked for multiple employers in the last five years. STEP does provide information on the months that the individual has worked for the firm. We are therefore able to define short tenure in a more demanding manner: having worked for the current employer for up to 24 months. The second and fourth columns additionally require the worker to be up to 25 years of age. Standard errors are reported in parenthesis and clustered at the country-level. *** p<0.01, ** p<0.05, * p<0.1*

Table B6: Evidence from the demand side: task requirements

Task category	LE estimate	# Obs.
Non-routine analytical	0.181 (0.035)	8,338
Use of ICT	0.184 (0.023)	8,212

Notes: The table reports the coefficient in a regression of a task measure (non-routine analytical and use of ICT, separately) on an indicator of large employer and fixed effects for sector, country, and the occupation asked at random by the surveyors. In parenthesis, we report the p-values of the test that the effects are null using wild-bootstrapped standard errors clustered at the country level. The estimating sample is obtained pooling the information obtained from all the countries participating in the STEP employer surveys. Further details on the construction of the outcome variables are provided in Appendix A.

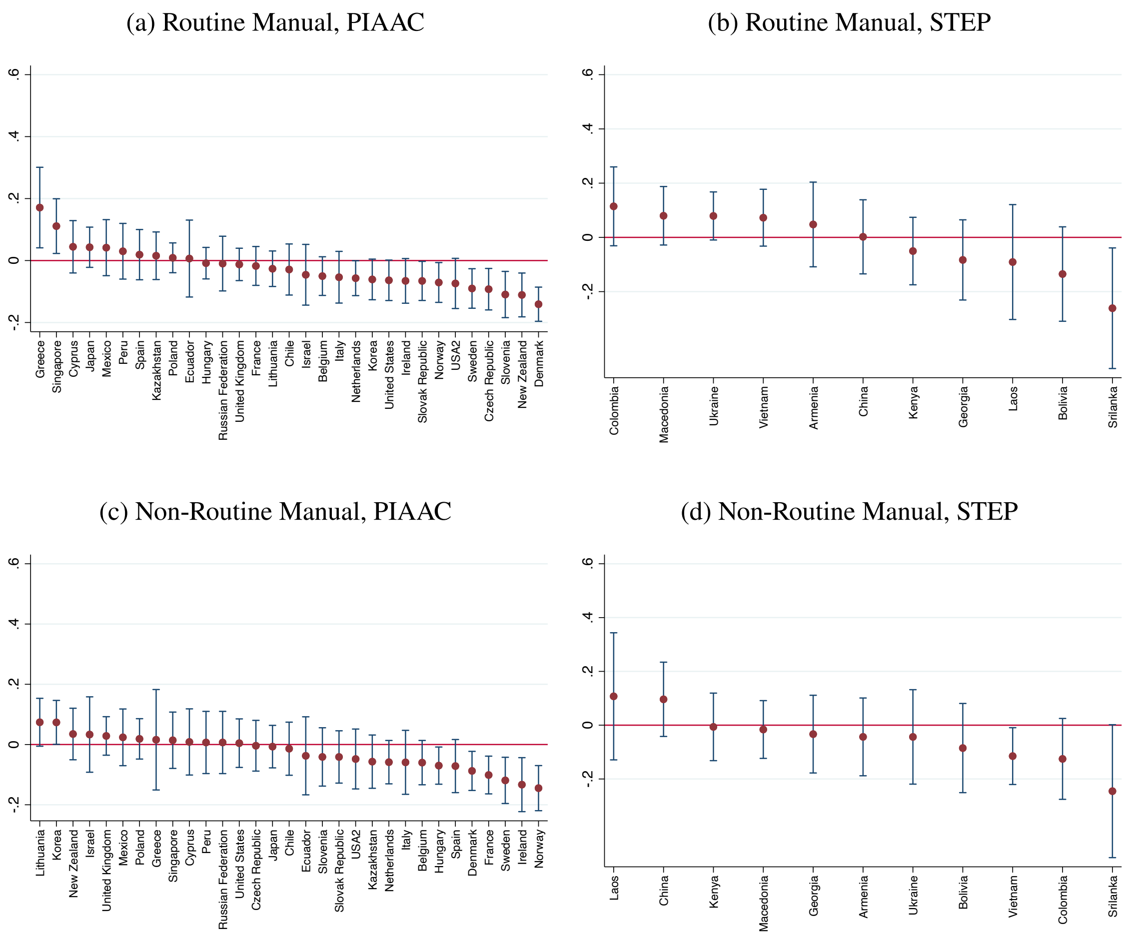
Table B7: Pooled estimates of the establishment size wage premium, explicitly documenting the returns on tasks

(a) Mean regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PIAAC							
ESWP	0.153***	0.132***	0.129***	0.114***	0.111***	0.097***	0.095***
	(0.014)	(0.014)	(0.014)	(0.013)	(0.012)	(0.017)	(0.014)
Non-routine analytical		0.068***	0.066***	0.048***	0.044***		
		(0.010)	(0.009)	(0.008)	(0.007)		
Non-routine interpersonal		0.022**	0.024**	0.021**	0.020**		
		(0.010)	(0.010)	(0.008)	(0.008)		
Routine cognitive		-0.029***	-0.029***	-0.026***	-0.022***		
		(0.006)	(0.006)	(0.008)	(0.007)		
Routine manual		-0.039***	-0.037***	-0.036***	-0.033***		
		(0.007)	(0.007)	(0.006)	(0.006)		
Non-routine manual		-0.025***	-0.025***	-0.017**	-0.016**		
		(0.007)	(0.007)	(0.007)	(0.007)		
Use of ICT		0.052***	0.052***	0.049***	0.045***		
		(0.006)	(0.006)	(0.004)	(0.004)		
Sample size	54,782	54,782	54,782	54,782	54,782	54,782	54,782
STEP							
ESWP	0.229***	0.213***	0.210***	0.185***	0.185***	0.158***	0.139***
	(0.028)	(0.023)	(0.023)	(0.024)	(0.025)	(0.030)	(0.029)
Non-routine analytical		0.039	0.038***	0.016	0.017		
		(0.022)	(0.021)	(0.019)	(0.016)		
Non-routine interpersonal		0.105**	0.107**	0.094**	0.095**		
		(0.040)	(0.040)	(0.034)	(0.035)		
Routine cognitive		-0.090***	-0.089***	-0.084***	-0.085***		
		(0.017)	(0.017)	(0.016)	(0.016)		
Routine manual		-0.057***	-0.057***	-0.062***	-0.062***		
		(0.015)	(0.015)	(0.017)	(0.016)		
Non-routine manual		0.019	0.018	-0.001	-0.001		
		(0.020)	(0.020)	(0.023)	(0.023)		
Use of ICT		0.141***	0.140***	0.123***	0.124***		
		(0.023)	(0.022)	(0.025)	(0.026)		
Sample size	8,339	8,339	8,339	8,339	8,339	8,339	8,339
Controls:							
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2-d Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Tasks		Yes	Yes	Yes	Yes	Yes	Yes
Industry FE			Yes	Yes	Yes	Yes	Yes
Individual demographics				Yes	Yes	Yes	Yes
Individual cognition/noncog.					Yes	Yes	Yes
Region FE						Yes	Yes
Country interactions							Yes

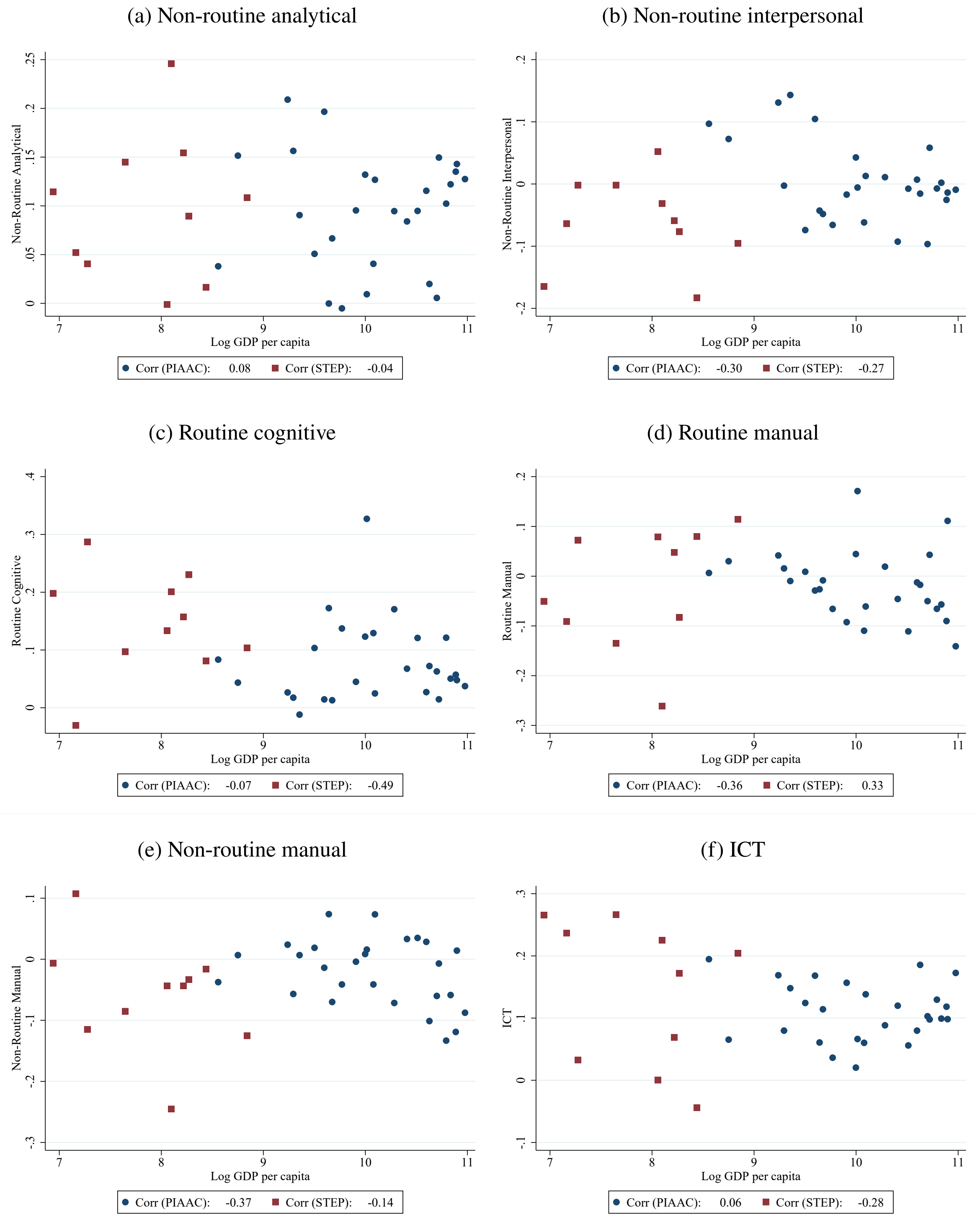
*Notes: Extension of Table 3 where we additionally estimate specifications with alternative sets of controls and we also report the results on the returns of tasks on wages. Column (1) does not report estimates for tasks since tasks are not part of that specification. Columns (6) and (7) do not report them because tasks are interacted with country fixed effects, so the level effect of the task lacks a meaningful interpretation. Standard errors are reported in parenthesis and clustered at the country level. *** p<0.01, ** p<0.05, * p<0.1*

Figure B1: Establishment size gradient in the task content of jobs by country, within 2-digit occupations



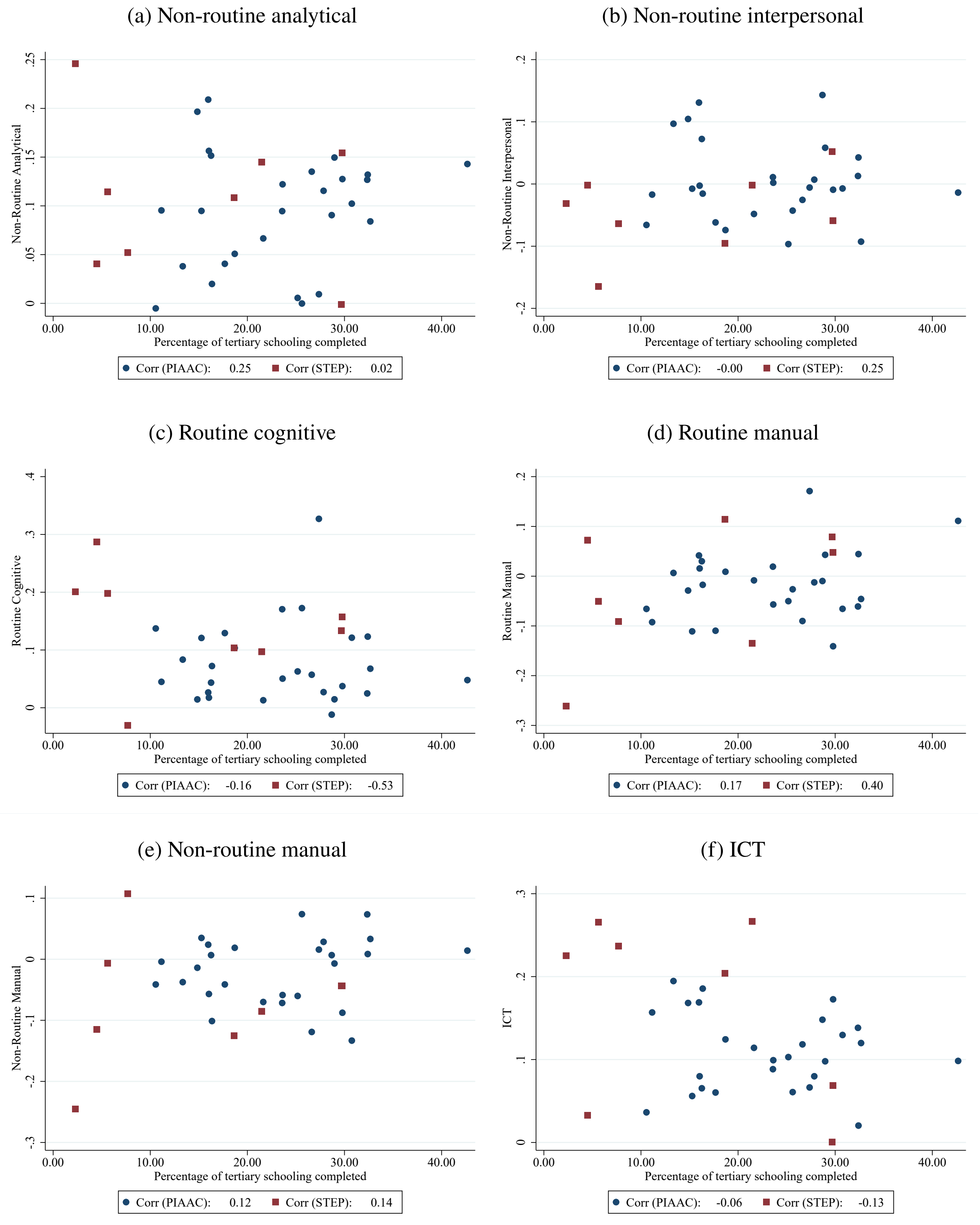
Notes: Continuation of Figure 1. Coefficient of an indicator for large establishment in a regression of task content intensity on indicator of large establishment (at least 50 employees) and the full set of controls as in Table 1's column (2). Regressions are done separately for each country. Countries are ordered by decreasing point estimates. "United States" refers to the survey conducted in 2012 and "USA2" to the one conducted in 2017. Reported confidence intervals at 95% confidence level computed using heteroskedasticity-robust standard errors.

Figure B2: Correlations of the establishment size gradient in task content with log GDP per capita



Notes: Correlations of the establishment size gradient in task content with country-level log GDP per capita. Correlations weighted by estimated precision of estimated establishment size gradients.

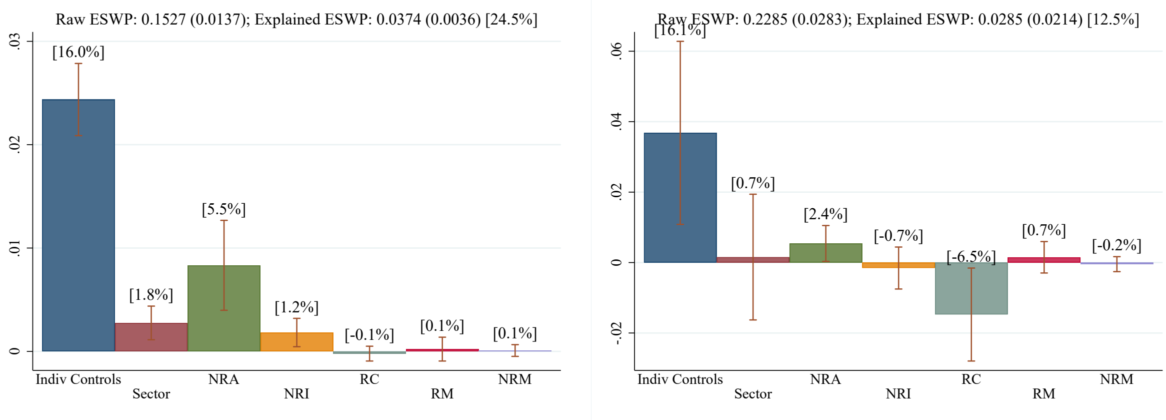
Figure B3: Correlations of the establishment size gradient in task content with fraction of population with at least tertiary education



Notes: Correlations of the establishment size gradient in task content with country-level proportion of population with at least tertiary-level education (as measured using Barro and Lee (2013)'s methodology on 2015 data). Correlations weighted by estimated precision of estimated establishment size gradients.

Figure B4: Gelbach decomposition of ESWP without ICT as mediator, pooled

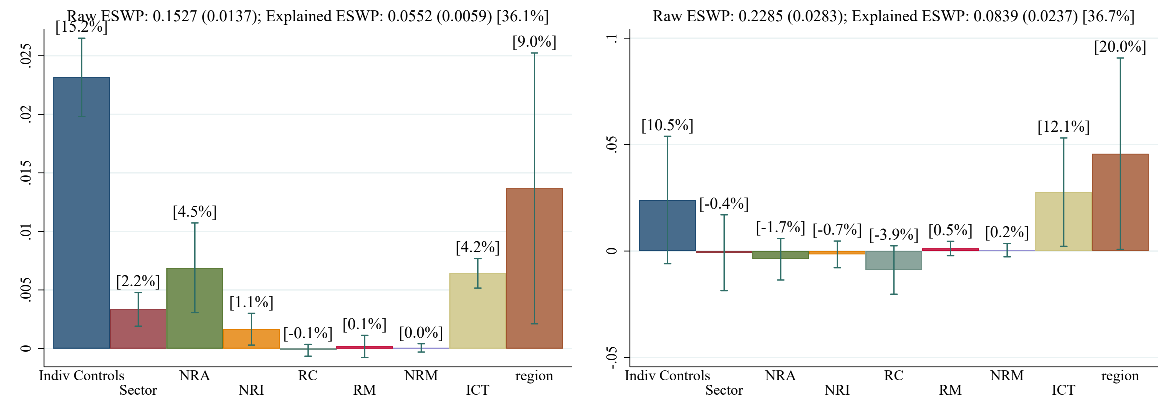
(a) PIAAC (b) STEP



Notes: Pooled PIAAC and STEP samples. Replication of Figure 2 not allowing ICT to be an independent mediator.

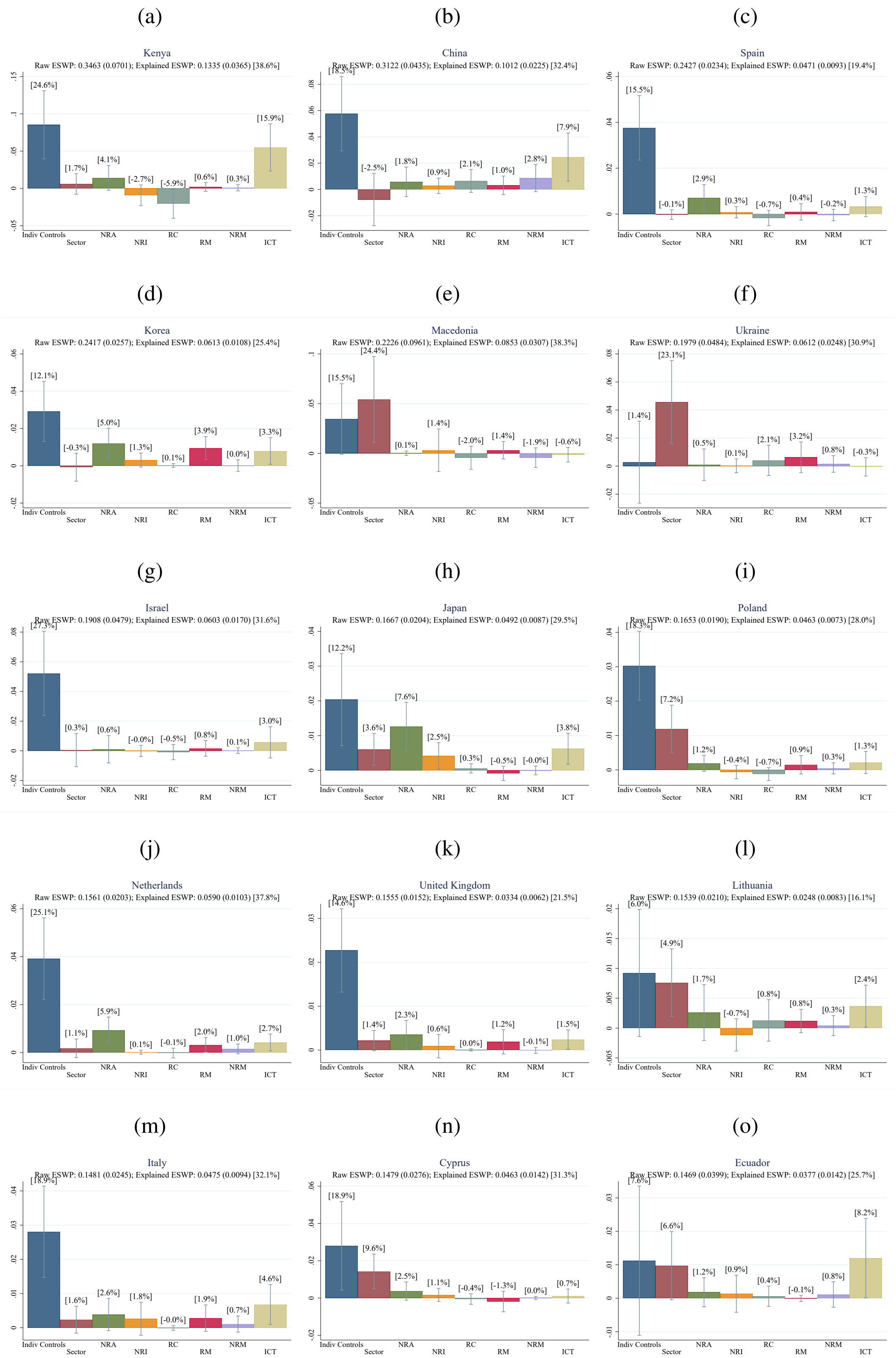
Figure B5: Gelbach decomposition of ESWP with region FE as mediators, pooled

(a) PIAAC (b) STEP



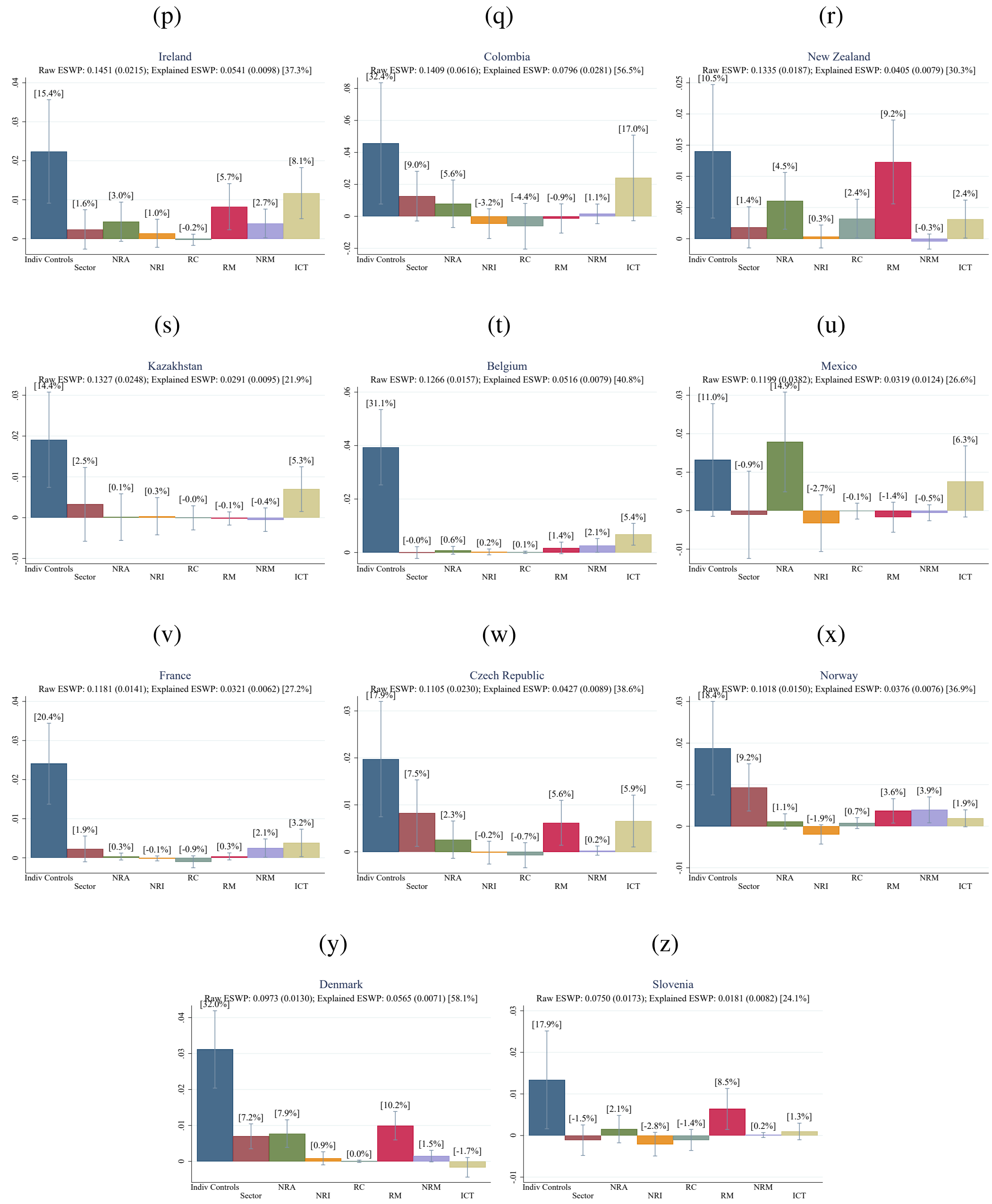
Notes: Pooled PIAAC and STEP samples. We include region as a potential independent mediator. Numbers in brackets indicate percentages of the raw ESWP. Reported confidence intervals at 95% confidence level. Standard errors are clustered at the country level.

Figure B6: Gelbach decomposition of ESWP, by country, no region FE as mediators



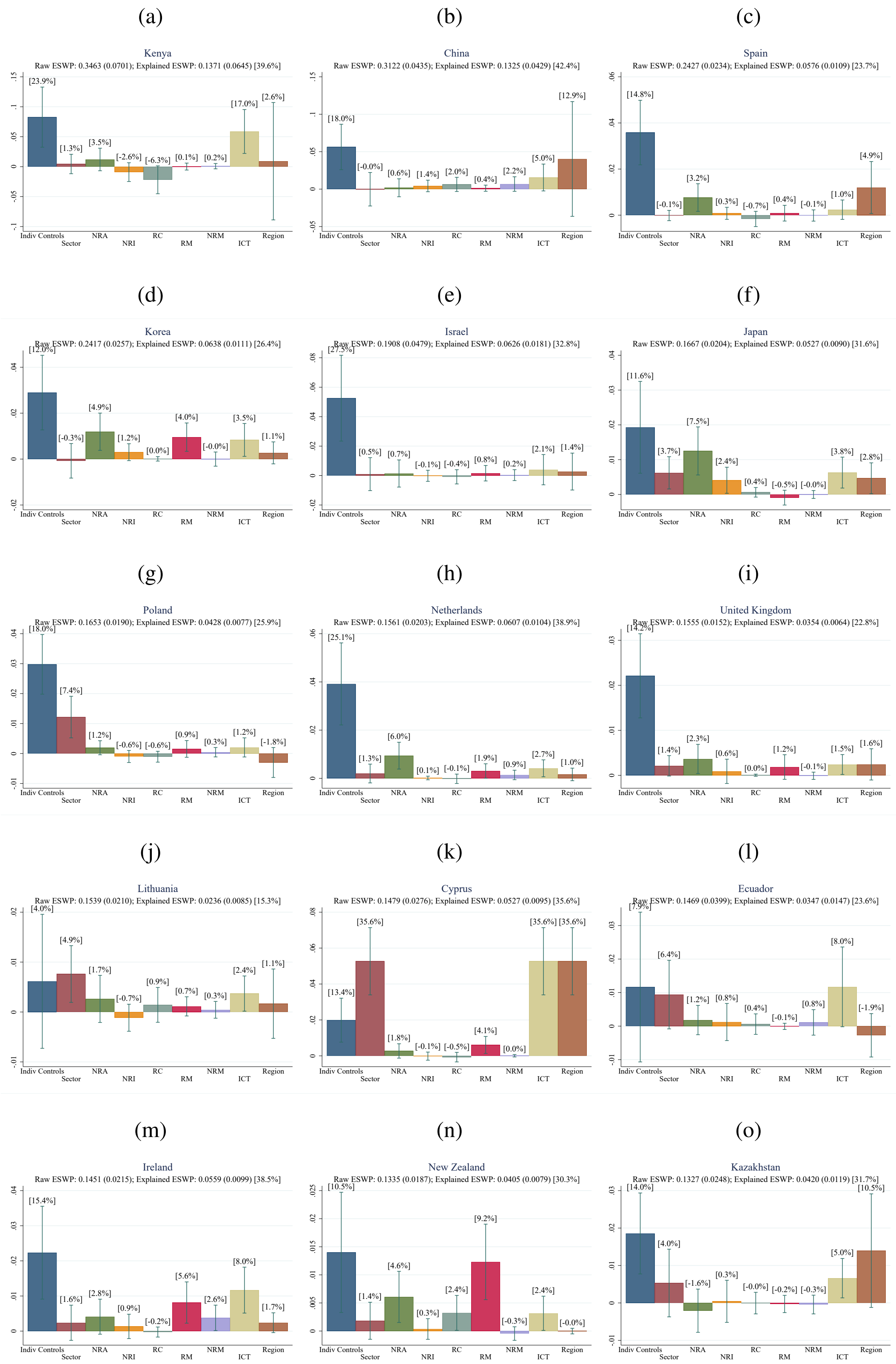
Notes: Only countries for which both the ESWP and the explained portion of the ESWP are statistically significant at the 95% confidence level are reported. Countries ordered by decreasing point estimate of the ESWP. Numbers in brackets indicate percentages of the raw ESWP.

Figure B6: Gelbach decomposition of ESWP, by country, no region FE as mediators (cont.)



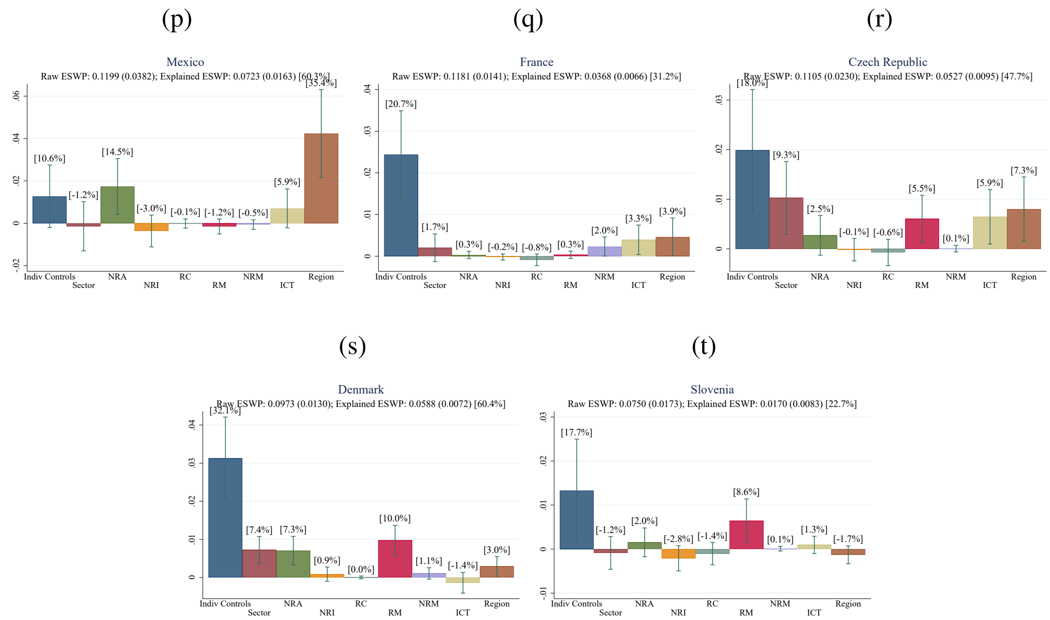
Notes: Only countries for which both the ESWP and the explained portion of the ESWP are statistically significant at the 95% confidence level are reported. Countries ordered by decreasing point estimate of the ESWP. Numbers in brackets indicate percentages of the raw ESWP.

Figure B7: Gelbach decomposition of ESWP with region FE as mediators, by country



Notes: Only countries for which both the ESWP and the explained portion of the ESWP are statistically significant at the 95% confidence level are reported. Countries ordered by decreasing point estimate of the ESWP. Numbers in brackets indicate percentages of the raw ESWP.

Figure B7: Gelbach decomposition of ESWP with region FE as mediators, by country (cont.)



Notes: Only countries for which both the ESWP and the explained portion of the ESWP are statistically significant at the 95% confidence level are reported. Countries ordered by decreasing point estimate of the ESWP. Numbers in brackets indicate percentages of the raw ESWP.

C Appendix: Additional analyses

C.1 Alternative construction of measures: Multiple correspondence analysis (MCA)

We assess the robustness of our qualitative results by performing some of the same analyses using a differently-constructed measure of task content. In particular, we use the same questions detailed in Appendix Table A1 but aggregate them in a different way. For this, we perform a multiple correspondence analysis (MCA) and choose the first dimension extracted, which corresponds to the dimension that explains the largest variance in the data. MCA is an attractive tool in that it can be thought of as the counterpart of principal component analysis for categorical (particularly, ordinal) data, just like the responses to the questions we have. We opt to perform the MCA by country and we standardize the resulting measure within the country. This means that they are again interpreted as standard deviations relative to the country mean.[3]

Figures available upon request summarize the proportion of total variance explained by the first dimension in the MCA. For most of the task dimensions, the MCA measures constructed explain a large proportion of the variation in the responses to the underlying questions. For example, for non-routine analytical tasks, the MCA measure explains around 70–90% of the variation, depending on the country, with this fraction being larger in STEP countries. The fraction is around 60–70% for non-routine interpersonal in PIAAC countries and close to 100% for STEP ones. For routine cognitive it is 60–80% in PIAAC countries but this is reduced to 30–60% in STEP countries. Indeed, the variance explained in STEP countries by the first dimension in the MCA using the questions related to routine cognitive is relatively small — around 20–60%, depending on the country, while the second dimension

[3]Note that we are not able to construct MCA-based measures for routine manual, non-routine manual, and use of ICT since they were originally constructed out of a single variable (in the case of routine manual in STEP we have two variables but the resulting MCA-based variable perfectly correlates with the non-MCA measure).

in the MCA still explains a substantial portion of the variation (20–30%). This suggests that the questions we associate with routine cognitive tasks capture multiple dimensions that a single index could not fully capture.

Table C1: Pooled estimates of establishment size gradient in the task content of jobs, MCA measure

Outcome variable	(1)	(2)	(3)	(4)	(5)	(6)
PIAAC						
Non-routine analytical	0.139***	0.143***	0.107***	0.094***	0.094***	0.089***
	(0.015)	(0.016)	(0.015)	(0.016)	(0.016)	(0.017)
Non-routine interpersonal	0.090***	0.101***	0.075***	0.066***	0.065***	0.059**
	(0.024)	(0.024)	(0.023)	(0.022)	(0.023)	(0.022)
Routine cognitive	0.013	0.012	0.030*	0.041**	0.035**	0.043***
	(0.017)	(0.017)	(0.016)	(0.015)	(0.016)	(0.015)
Sample size	65,151	65,151	65,151	65,151	65,151	65,151
STEP						
Non-routine analytical	0.138***	0.142***	0.078**	0.074**	0.086**	0.091**
	(0.037)	(0.036)	(0.031)	(0.029)	(0.038)	(0.038)
Non-routine interpersonal	-0.017	-0.005	-0.031	-0.032	-0.054	-0.029
	(0.032)	(0.036)	(0.039)	(0.037)	(0.034)	(0.031)
Routine cognitive	0.045	0.037	0.020	0.020	0.006	-0.018
	(0.070)	(0.067)	(0.064)	(0.064)	(0.085)	(0.075)
Sample size	8,339	8,339	8,339	8,339	8,339	8,339
Controls:						
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
2-d Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE		Yes	Yes	Yes	Yes	Yes
Region FE			Yes	Yes	Yes	Yes
Individual demographics				Yes	Yes	Yes
Individual cognition/noncog.					Yes	Yes
Country interactions						Yes

*Notes: Replication of Table B1 where task requirement intensity is our MCA measure. Only those tasks for which an MCA measure can be computed are reported. Standard errors are reported in parentheses and clustered at the country level. *** p<0.01, ** p<0.05, * p<0.1*

In Appendix Table C1, we replicate the results reported in Table 1 using the MCA measures. We find not only qualitatively but also quantitatively similar results to the ones in the main text. The pattern is not repeated for routine cognitive in STEP countries. This is not surprising since, as we argued before, our MCA measure is likely to capture a very specific dimension of routine cognitive tasks that does not really reflect the same aspects as our original measure did.

C.2 Distributional differences in the task content of jobs by establishment size

To complement our finding on the task intensity gradients by establishment size, we run various distribution regressions that model the conditional distribution of the outcome (Chernozhukov et al., 2013).[4] We approximate the probability that the task intensity performed T_i is greater than a particular value $t \in \mathcal{T} \subset \text{Supp}(T_i)$ as a linear probability model:

$$\Pr(T_i > t \mid LE_{j(i)}, o(i), c(i)) = \beta \times \text{LE}_{j(i)} + X_i'\gamma + \delta^o_{o(i)} + \delta^c_{c(i)} + \varepsilon_i, \tag{5}$$

where $LE_{j(i)}$ is the indicator for worker i being in a large establishment, X is a vector of individual- and establishment-level characteristics, δ^o are occupation fixed effects, and δ^c are country fixed effects. We report the estimates for β for a set of support points $\mathcal{T}$. We have also used a probit specification and the qualitative results do not change.

Appendix Table C2 summarizes the estimated coefficients in the distribution regression of the task content of jobs on an indicator of being in a large establishment controlling for the set of controls in column (2) of Table 1. We consider the points on the support $\{-0.75, -0.5, 0, 0.5, 0.75\}$ as thresholds. We find that the coefficient on establishment size for non-routine analytical, routine-cognitive and use of ICT broadly maintain the sign and the significance throughout the five support points. This suggests that our baseline results are not driven by a few workers that perform these tasks more intensively, that is, the difference in means is not because of differences in the tails but because the entire tasks intensity distribution in large establishments is shifted to the right relative to smaller establishments.

[4] A complementary approach is quantile regression. However, since the task content measures inherit the discrete nature of the responses to the survey questions, quantile regressions may not be appropriate while distribution regression remains valid.

Table C2: Distribution regression estimates of establishment size gradient in the task content of jobs, pooled sample

	Support points				
	(1)	(2)	(3)	(4)	(5)
Outcome variable:	−0.75	−0.50	0.00	0.50	0.75
Panel (a): PIAAC					
NRA	0.031***	0.040***	0.049***	0.054***	0.051***
	(0.008)	(0.005)	(0.008)	(0.009)	(0.009)
NRI	0.031***	0.028***	0.023**	0.011	0.007
	(0.008)	(0.009)	(0.010)	(0.010)	(0.008)
RC	–	0.012*	0.014**	0.017***	0.010
	(–)	(0.007)	(0.005)	(0.005)	(0.007)
RM	-0.013***	-0.008*	0.008*	0.018**	0.017***
	(0.004)	(0.004)	(0.004)	(0.007)	(0.006)
NRM	-0.010	-0.006	0.004	0.013*	0.007
	(0.006)	(0.007)	(0.008)	(0.007)	(0.007)
Use of ICT	0.061***	0.061***	0.061***	0.059***	0.030**
	(0.005)	(0.005)	(0.005)	(0.005)	(0.012)
Sample Size	61,151	61,151	61,151	61,151	61,151
Panel (b): STEP					
NRA	0.024***	0.026*	0.021	0.031***	0.024**
	(0.005)	(0.013)	(0.017)	(0.017)	(0.010)
NRI	-0.031*	-0.013	-0.008	-0.006	0.005
	(0.017)	(0.021)	(0.014)	(0.013)	(0.011)
RC	0.072***	0.060**	0.057**	0.063**	0.064**
	(0.017)	(0.024)	(0.022)	(0.023)	(0.023)
RM	-0.013	-0.007	0.002	0.010	0.011
	(0.018)	(0.019)	(0.014)	(0.009)	(0.014)
NRM	–	0.003	-0.022*	-0.022*	-0.016
	(–)	(0.009)	(0.012)	(0.012)	(0.019)
Use of ICT	0.032	0.058**	0.058**	0.058**	0.058**
	(0.021)	(0.020)	(0.020)	(0.020)	(0.020)
Sample Size	8,339	8,339	8,339	8,339	8,339

*Notes: Pooled PIAAC and STEP samples. Coefficient of an indicator for large establishment in a distribution regression (Equation (5)) of task content intensity on indicator of large establishment (at least 50 employees) and the set of controls in column (2) of Table 1 estimated as a linear probability model. "–" indicates cases where no observation is found below the threshold indicated by the relevant column. Standard errors are reported in parenthesis and clustered at the country level. *** p<0.01, ** p<0.05, * p<0.1*

C.3 Additional analyses: Establishment size wage premium

Alternative treatment of occupations. Similarly to our analysis for the gradient in the task content of jobs, Appendix Tables C3 and C4 show that the presence of a sizable ESWP remains when we employ 3-digit occupational fixed effects and when we expand the sample to include the PIAAC countries that lack regional information. Moreover, in Appendix Tables C5 and C6, we again account for occupations more flexibly by estimating Equation 2 conditioning on workers being in specific 1-digit occupation codes, and controlling for 2-digit occupation codes. Both in PIAAC and STEP countries, we find a comparatively large ESWP among workers in service and sales (0.159 and 0.175 log points, respectively), with clerical and support workers, managers, and professionals also commanding a large wage premium.

Table C3: Pooled estimates of the establishment size wage premium when using 3-digit occupations (mean regressions)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel (a): PIAAC							
ESWP	0.158***	0.138***	0.135***	0.126***	0.123***	0.106***	0.101***
	(0.011)	(0.010)	(0.010)	(0.009)	(0.009)	(0.011)	(0.009)
Sample size	42,945	42,945	42,945	42,945	42,945	42,945	42,945
Panel (b): STEP							
ESWP	0.242***	0.219***	0.226***	0.205***	0.206**	0.146*	0.075**
	(0.060)	(0.054)	(0.061)	(0.062)	(0.064)	(0.073)	(0.032)
Sample size	4,664	4,664	4,664	4,664	4,664	4,664	4,664
Controls:							
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3-d Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Tasks		Yes	Yes	Yes	Yes	Yes	Yes
Industry FE			Yes	Yes	Yes	Yes	Yes
Individual demographics				Yes	Yes	Yes	Yes
Individual cognition/noncog.					Yes	Yes	Yes
Region FE						Yes	Yes
Country interactions							Yes

*Notes: Replication of Table B7 substituting 2-digit occupation fixed effects by 3-digit ones. The unavailability of this finer degree of information for some countries explains the difference in observations with respect to the baseline table. *** p<0.01, ** p<0.05, * p<0.1*

Table C4: Pooled estimates of the establishment size wage premium when not discarding countries without region (mean regressions)

	(1)	(2)	(3)	(4)	(5)	(6)
ESWP	0.153***	0.133***	0.129***	0.114***	0.111***	0.109***
	(0.013)	(0.013)	(0.013)	(0.012)	(0.011)	(0.009)
Sample size	58,885	58,885	58,885	58,885	58,885	58,885
Controls:						
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
2-d Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes
Tasks		Yes	Yes	Yes	Yes	Yes
Industry FE			Yes	Yes	Yes	Yes
Individual demographics				Yes	Yes	Yes
Individual cognition/noncog.					Yes	Yes
Country interactions						Yes

*Notes: PIAAC pooled sample. Replication of Table B7 when not discarding PIAAC countries without regional information (Italy and Norway). This explains the increase in available observations. *** p<0.01, ** p<0.05, * p<0.1*

Cross-country comparisons: Average ESWP. We find that the pooled estimates are reflective of fairly universal patterns at the country level. In Figure C1, we report the estimates of the ESWP by country. Most of the country-specific ESWP estimates under our main specification lie between 0.05 and 0.20 log points, corresponding to approximately 5–22% average real hourly wage differences between workers in larger establishments relative to those in smaller establishments. The estimated ESWP remains unchanged after additionally adjusting for region fixed effects.

Cross-country comparisons: ESWP over the distribution. In Appendix Figure C2, we report the differences in the distribution of wages across establishment size by country. We find that in almost all of the countries we study, the wage distribution of workers in larger establishments is shifted to the right relative to the wage distribution of workers in smaller establishments. These results echo the qualitative results we obtained using the pooled data.

For privacy reasons, not all countries in PIAAC report information on hourly wages in the public use files. In such cases, only the decile of the wage distribution the person is located

Table C5: Establishment size wage premium, by occupation, PIAAC pooled sample

1-digit ISCO-08 Category	ESWP	# Obs.
Managers	0.131***	3,561
	(0.034)	
Professionals	0.107***	11,408
	(0.015)	
Technicians & associate professionals	0.061*	7,811
	(0.033)	
Clerical support workers	0.114***	6,484
	(0.023)	
Services & sales workers	0.159***	10,410
	(0.016)	
Craft & related trade workers	0.075*	5,414
	(0.038)	
Plant & machine operators, & assemblers	0.147***	4,364
	(0.027)	
Elementary occupations	0.027	4,900
	(0.045)	

*Notes: PIAAC pooled sample. Coefficient of an indicator for large establishment in a regression of log real hourly wages on indicator of large establishment (at least 50 employees) by 1-digit ISCO-08 occupation codes, controlling for the set of controls in Table 3's column (2). We do not report the 1-digit categories corresponding to armed forced occupations and skilled agricultural, forestry and fishery workers due to small sample size. Standard errors are reported in parenthesis and clustered at the country level. *** p<0.01, ** p<0.05, * p<0.1*

in is provided. We utilize this information and estimate linear probability models where the outcome is an indicator taking the value of 1 if the worker is *at least* in a certain wage decile. Appendix Figure C3 reports the coefficient of this linear probability model controlling for the set of controls in column (2) of Table 3. The results complement what we learn from the quantile regressions: not only are workers in larger establishments more likely to have wages in the last decile, but these workers are also more likely to have wages that are at least above the second and fifth decile. This is further evidence towards the wage distribution of large establishments being shifted to the right compared to smaller establishments, even within narrowly-defined occupation groups.

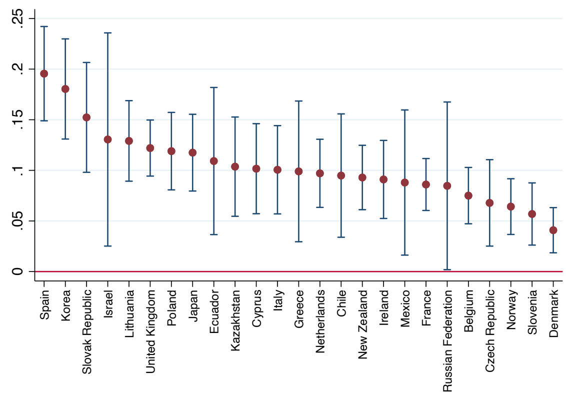
Table C6: Establishment size wage premium, by occupation, STEP pooled sample

1-digit ISCO-08 Category	ESWP	# Obs.
Managers	1.140	451
	(0.894)	
Professionals	0.149**	2,142
	(0.061)	
Technicians & associate professionals	-0.513	788
	(0.448)	
Clerical support workers	0.165**	895
	(0.069)	
Services & sales workers	0.175**	1,847
	(0.059)	
Craft & related trade workers	0.116	687
	(0.079)	
Plant & machine operators, & assemblers	-0.029	588
	(0.090)	
Elementary occupations	0.128	941
	(0.127)	

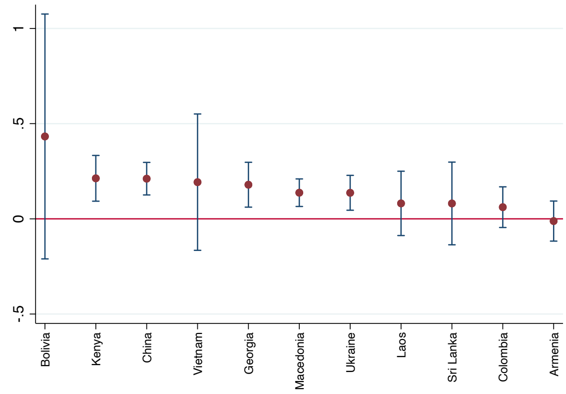
*Notes: STEP pooled sample. Coefficient of an indicator for large establishment in a regression of log real hourly wages on indicator of large establishment (at least 50 employees) by 1-digit ISCO-08 occupation codes, controlling for the set of controls in Table 3's column (2). We do not report the 1-digit categories corresponding to armed forced occupations and skilled agricultural, forestry and fishery workers due to small sample size. Standard errors are reported in parenthesis and clustered at the country level. *** p<0.01, ** p<0.05, * p<0.1*

Figure C1: Estimated average establishment size wage premium by country

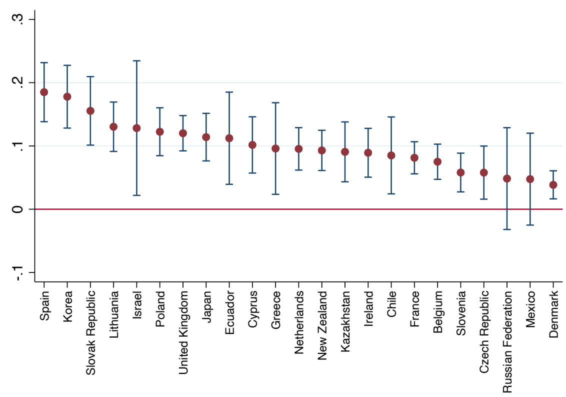
(a) PIAAC, without region FE



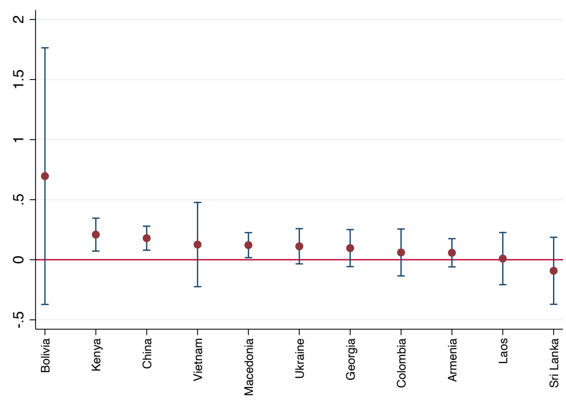
(b) STEP, without region FE



(c) PIAAC, with region FE

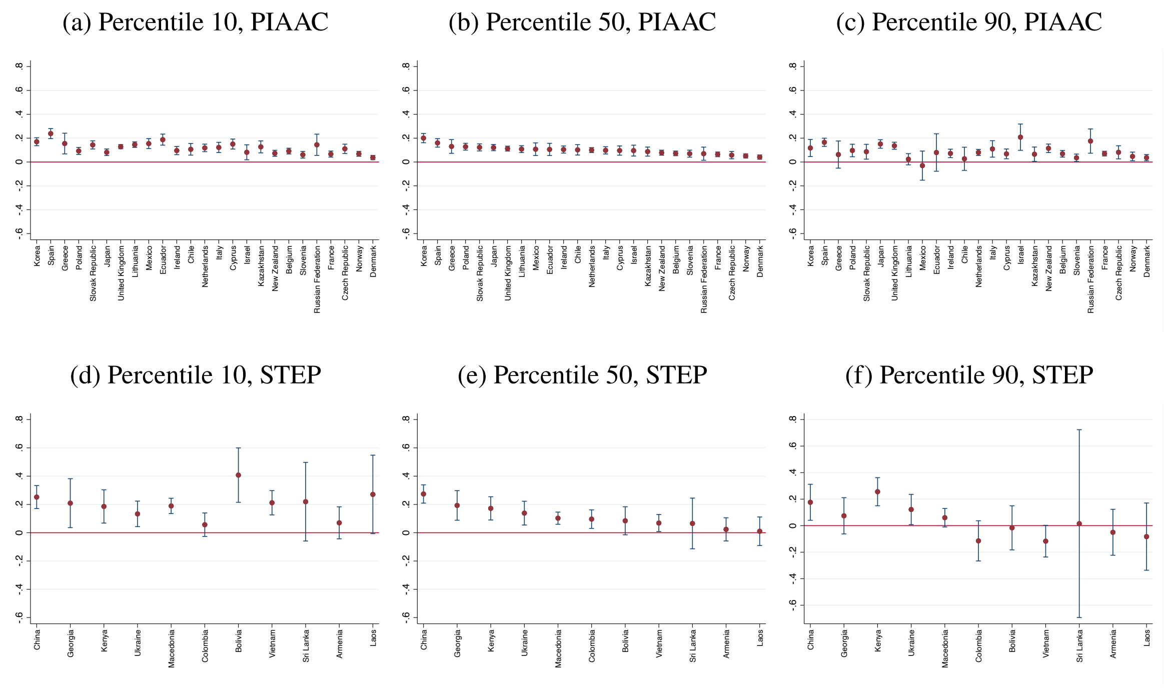


(d) STEP, with region FE



Notes: Regressions of log wages on an indicator of large establishment (at least 50 employees), by country. In subfigures (a) and (b) the controls are the same as in Table 3's column (2). For subfigures (c) and (d), we additionally control for region fixed effects. The estimating sample for subfigures (a) and (b) is the set of countries for which a continuous measure of wages is available. Subfigures (c) and (d) further require that information on the region is available, which decreases the number of countries available for PIAAC (Italy and Norway are unavailable). Reported confidence intervals at 95% confidence level computed using heteroskedasticity-robust standard errors.

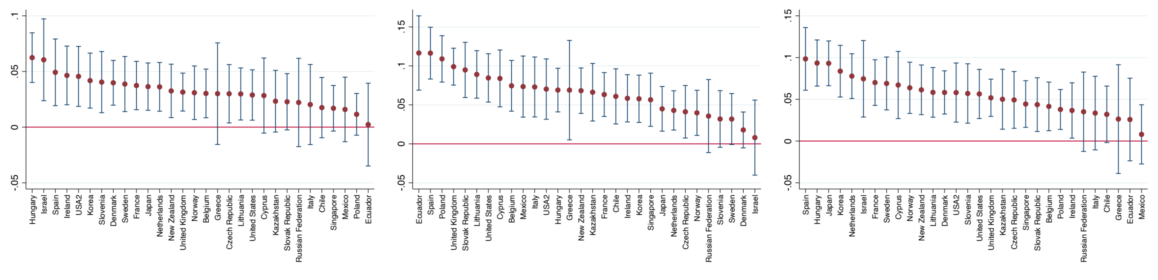
Figure C2: Distribution of wages in large establishments by country, percentiles 10, 50 and 90



Notes: Coefficient of an indicator for large establishment in a quantile regression of task content intensity on indicator of large establishment (at least 50 employees) and the set of controls as in Table 3's column (2). Countries ordered by decreasing point estimates in the median regression. Regressions estimated for each country separately. Reported confidence intervals at 95% confidence level computed using heteroskedasticity-robust standard errors.

Figure C3: ESWP based on wage deciles in PIAAC

(a) ≥ Decile 2 (b) ≥ Decile 5 (c) ≥ Decile 10



Notes: Coefficient of an indicator of large establishment (at least 50 employees) in a linear regression of an indicator of being at least in a certain wage decile on the indicator of large establishment and the set of controls as in Table 3's column (2). Countries ordered by decreasing point estimates. Regressions done for each country separately. Reported confidence intervals at 95% confidence level computed using heteroskedasticity-robust standard errors.

D Appendix: Summaries of Oster (2019) and Gelbach (2016)

D.1 Summary of Oster (2019)

We first set up the problem then provide intuition behind the methodology. Suppose that the true data generating process is the following:

$$Y = X_0\beta + X_1\gamma_1 + X_2\gamma_2 + \varepsilon, \tag{6}$$

where β is the coefficient of interest (the establishment size gradient, in our case), X_1 is a set of observable covariates, and X_2 is a set of unobserved covariates that are possibly inducing misspecification.

Given the observed (Y, X_0, X_1), we can compute the following quantities:

- β_r which is the coefficient of X_0 in the short (restricted) regression of Y on just X_0
- β_u which is the coefficient of X_0 in the long (unrestricted) regression of Y on (X_0, X_1); this also serves as our estimator of the LFWP

The main question is whether we can say anything about the biases of our estimators (specifically, β_u) relative to the population parameter β which we cannot feasibly estimate because X_2 is unobserved. Oster (2019) argues that the bias is related to the following quantities:

- $\beta_r - \beta_u$, difference in the coefficients from the short and long regressions: this is the conventional idea of coefficient stability where if the estimate of the coefficient of X_0 does not change significantly after the addition of observed controls X_1, then the estimates are "robust" to selection. The idea is that "robustness" to controlling for observed controls is somehow informative of "robustness" to unobservables. In general, this is not sufficient if we do not know the exact relationship between X_0 and the unobservables X_2. This is a quantity that has a finite sample counterpart.

- $R^2_u - R^2_r$, difference in the R^2 of the long and short regressions: coefficient stability by taking the difference in the coefficients from the short and long regressions is more informative if the observed controls are able to explain a large part of the variation in the outcome. This is a quantity with a finite sample counterpart.

- $R^2_{\max} - R^2_u$, theoretical maximum explained variance of the outcome: as mentioned previously, coefficient stability is more informative if the observed controls explain a "large" part of the variation in the outcome. What is considered "large" depends on how much of the variation in the outcome is actually explainable by all the factors; that is, how large $R^2_{\max}$, which is the R^2 from the theoretical regression of Y on (X_0, X_1, X_2), can be. Note that while R^2_u has a finite sample counterpart, $R^2_{\max}$ does not.

Let $W_1 \equiv X_1\gamma_1$ and $W_2 \equiv X_2\gamma_2$, and define δ such that

$$\delta \frac{\mathrm{Cov}(W_1, X_0)}{\mathrm{Var}(W_1)} = \frac{\mathrm{Cov}(W_2, X_0)}{\mathrm{Var}(W_2)},$$

where the LHS and RHS contain the quantities obtained from regressing X_0 on X_1 and X_2 separately, respectively. The quantity δ describes the proportionality of the selection between the observables and unobservables.

Under some assumptions, Oster (2019) provides expressions for $(\beta_r - \beta_u, R^2_u - R^2_r, R^2_{\max} - R^2_u)$ which clarify the relationship between the unknown objects $(\beta_u - \beta, \delta, R^2_{\max})$, i.e., the bias, the proportionality of selection, and the theoretical max R^2.[5] Given two of the three unknown quantities, one can solve for the remaining unknown quantity using the derived expressions. Thus, one way to assess robustness of our estimates to selection on unobservables is to find the δ such that the bias implied by the selection on unobservables drives $\beta = 0$ for a given $R^2_{\max}$ (which in practice is set to $1.3 \times R^2_u$ as a rule-of-thumb). We report such estimates $\hat{\delta}$ in Table 1.

[5] Giuseppe De Luca and Peracchi (2019) provide a more general misspecification framework that nests the results of Oster (2019).

D.2 Summary of Gelbach (2016)

Suppose that the data is generated by the model in Equation (4), the "full" model. To ease exposition, we write the model in matrix form:

$$Y = Z\beta^{\text{full}} + X\gamma + \varepsilon^{\text{full}}, \tag{7}$$

where Y is the vector of log wages $\log w_i$; Z is the matrix whose columns include the column of indicators for establishment size, dummies for occupations, and dummies for countries; and X is the matrix whose columns are the mediators we are interested in. If $(\hat{\beta}^{\text{full}\prime}, \hat{\gamma}')'$ denotes the OLS estimates of $(\beta^{\text{full}\prime}, \gamma')'$, then

$$Y = Z\hat{\beta}^{\text{full}} + X\hat{\gamma} + \hat{\varepsilon}^{\text{full}}, \tag{8}$$

where $\hat{\varepsilon}^{\text{full}}$ are the OLS residuals. Pre-multiplying both sides by $(Z'Z)^{-1}Z'$, we get

$$(Z'Z)^{-1}Z'Y = \hat{\beta}^{\text{full}} + (Z'Z)^{-1}Z'X\hat{\gamma}. \tag{9}$$

However, the left hand size is the OLS estimate of the "raw" model defined in Equation (3). Thus, we have that[6]

$$\hat{\beta}^{\text{raw}} - \hat{\beta}^{\text{full}} = (Z'Z)^{-1}Z'X\hat{\gamma}, \tag{10}$$

which characterizes the difference in the two estimates and provides us a natural decomposition. In particular, if we let X_k denote the column corresponding to the k^{th} mediator in X, and $\hat{\gamma}_k$ as the estimated coefficient of the same mediator in the "full" regression, then

$$\hat{\delta}_k = (Z'Z)^{-1}Z'X_k\hat{\gamma}_k \tag{11}$$

[6]Note that the probability limit of $(Z'Z)^{-1}Z'X\hat{\gamma}$ corresponds to the population omitted variable bias in β when excluding X in the full regression.

provides a natural estimate of the contribution of the k^{th} mediator in $\hat{\beta}^{\text{raw}} - \hat{\beta}^{\text{full}}$. Intuitively, the contribution of an individual mediator X_k ($\hat{\delta}_k$) is an estimate of the establishment size gap in X_k (obtained by regressing X_k on Z, as in $(Z'Z)^{-1}Z'X_k$) scaled by its effect on wages (γ_k).