ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Storm, Eduard

Working Paper

On the measurement of tasks: Does expert data get it right?

Ruhr Economic Papers, No. 948

Provided in Cooperation with:
RWI – Leibniz-Institut für Wirtschaftsforschung, Essen

Suggested Citation: Storm, Eduard (2022) : On the measurement of tasks: Does expert data get it right?, Ruhr Economic Papers, No. 948, ISBN 978-3-96973-111-6, RWI - Leibniz-Institut für Wirtschaftsforschung, Essen, https://doi.org/10.4419/96973111

This Version is available at:
https://hdl.handle.net/10419/259006

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

RUHR

ECONOMIC PAPERS

Eduard Storm

On the Measurement of Tasks: Does Expert Data Get It Right?

rwi #948

Imprint

Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Editors

Prof. Dr. Thomas K. Bauer
RUB, Department of Economics, Empirical Economics
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Ludger Linnemann
Technische Universität Dortmund, Department of Business and Economics
Economics – Applied Economics
Phone: +49 (0) 231/7 55-3102, e-mail: : Ludger.Linnemann@tu-dortmund.de

Prof. Dr. Volker Clausen
University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83-3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Ronald Bachmann, Prof. Dr. Manuel Frondel, Prof. Dr. Torsten Schmidt, Prof. Dr. Ansgar Wübker
RWI, Phone: +49 (0) 201/81 49-213, e-mail: presse@rwi-essen.de

Editorial Office

Sabine Weiler
RWI, Phone: +49 (0) 201/81 49-213, e-mail: sabine.weiler@rwi-essen.de

Ruhr Economic Papers #948

Responsible Editor: Ronald Bachmann

All rights reserved. Essen, Germany, 2022

ISSN 1864-4872 (online) – ISBN 978-3-96973-111-6

The working papers published in the series constitute work in progress circulated to stimulate discussion and critical comments. Views expressed represent exclusively the authors' own opinions and do not necessarily reflect those of the editors.

Ruhr Economic Papers #948

Eduard Storm

On the Measurement of Tasks: Does Expert Data Get It Right?

rwi

Bibliografische Informationen der Deutschen Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at **http://dnb.dnb.de**

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

http://dx.doi.org/10.4419/96973111
ISSN 1864-4872 (online)
ISBN 978-3-96973-111-6

Eduard Storm[1]

On the Measurement of Tasks: Does Expert Data Get It Right?

Abstract

Using German survey and expert data on job tasks, this paper explores the presence of omitted-variable bias suspected in conventional task data derived from expert assessment. I show expert task data, which is expressed at the occupation-level, introduces omitted-variable bias in task returns on the order of 24-34%. Motivated by a theoretical framework, I argue this bias results from expert data ignoring workplace heterogeneity rather than fundamental differences on the assessment of tasks between experts and workers. My findings have important implications for the interpretation of conventional task models as task returns expressed at the occupation-level are overestimated. Moreover, a rigorous comparison of the statistical performance of various models offers guidance for future research regarding choice of task data and construction of task measures.

JEL-Codes: C18, J24, J31

Keywords: Expert vs survey task data; workplace heterogeneity; omitted-variable bias

April 2022

1 Eduard Storm, RWI. - I wish to thank Britta Matthes for providing insight on the relative strengths and weaknesses of different task data and BW Bildung und Wissen for explaining the process of gathering information on expert-based data as part of the BERUFENET database. Moreover, I thank Ronald Bachmann, Gökay Demir, and seminar participants at the SOLE 2021 and EEA 2021 for helpful comments. All remaining errors are my own. – All correspondence to: Eduard Storm, RWI, Hohenzollernstr. 1/3, 45128 Essen, Germany, e-mail: eduard.storm@rwi-essen.de

1 Introduction

A growing body of research has adopted the "task-approach" to labor markets (Autor 2013) that models the assignment of worker-specific skills to job tasks. This framework allows a more nuanced evaluation on the role of skills in the production function as worker's skills are derived from comparative advantages in tasks. Most studies employing task data use information at the occupation-level, which is often based on external assessment by labor market experts. While widely used, this expert data may introduce measurement error attributed to (i) aggregated task data and (ii) misperception of experts on the importance of job tasks. The primary interest in the present paper is on the unit of dimension as expert data disregards heterogeneity within occupations.

Indeed, using survey data on job activities of US workers at the workplace, Autor & Handel (2013) contrast variation in tasks at the individual- and occupation-level and find worker-level information on tasks to be informative about wage differences not only between occupations, but also within. Cassidy (2017) and Rohrbach-Schmidt (2019) provide similar evidence in the German context and de La Rica, Gortazar & Lewandowski (2020) in a cross-country setting using PIAAC data, suggesting this finding is not country-specific. Related evidence on dispersion of tasks within occupations can be found in Atalay, Phongthiengtham, Sotelo & Tannenbaum (2018, 2020), Deming & Noray (2019), and Modestino, Shoag & Ballance (2019). Moreover, building upon Autor & Handel (2013), Stinebrickner, Stinebrickner & Sullivan (2019) use a novel longitudinal data set in which two cohorts of students of the Berea College are tracked after graduation. They find that tasks associated with people's skills and information gathering are highly rewarded in labor markets. Combined, these papers suggest substantial heterogeneity embedded in job tasks and echo the well-known difference in the unit of interest between survey and expert data. While the former emphasizes tasks performed at the workplace, the latter describes occupational characteristics (Autor 2013, Dengler, Matthes & Paulus 2014). By focusing on the occupational dimension, expert data implicitly assumes workers within an occupation perform a common set of tasks. These conventional task models therefore ignore workplace heterogeneity, giving rise to omitted-variable bias in estimated task returns.

While previous contributions on the heterogeneity of job tasks are convincing and important, neither of these studies explicitly measures the bias in task returns embedded in conventional task data. This information is important for practitioners, however, who often use task data on the grounds of theory of comparative advantage in tasks. In this paper I fill this gap by rigorously comparing the statistical properties of task models based on survey and expert data, respectively. This comparison allows me to test and quantify the presence

of omitted-variable bias in task returns based on expert data. To this end, I make four contributions to the existing literature.

First, I find worker-level information on tasks is predictive of wage differences in all specifications and thus in line with prior research. Relative to performing manual tasks, I find a 1 pp. increase in abstract task intensity raises wages by 36-53%. Employing a sizable cross-section of more than 27,000 workers in Germany from 2012-18 with self-reported information on job tasks represents an improvement over the existing literature that either uses much smaller samples (Autor & Handel 2013, Rohrbach-Schmidt 2019) or older data (Cassidy 2017). Idiosyncratic differences in tasks are especially pronounced in models conditional on occupational fixed effects (FE), providing direct evidence on task specialization within occupations.

Second, I conduct formal tests of various task models. In this analysis, I compare the statistical performance of wage regressions comprising survey data and, respectively, expert data, provided by (Dengler, Matthes & Paulus 2014, DMP).[1] Overall, baseline results suggest only minor statistical differences between survey- and expert task data. While goodness-of-fitness measures and information criteria favor models based on worker-level variation, expert data has more unique explanatory power. The broad statistical similarity likewise holds true for a comparison of occupation-level expert data with occupation-level task measures derived from survey data. Hence, assessment by labor market experts on the importance of job tasks does not appear to be fundamentally different from worker assessment. Further robustness checks reverse some of the perceived benefits of expert data in baseline specifications, however. Instead, a majority of robustness tests support statistical superiority of individual-level task measures from survey data, especially with respect to its unique explanatory power. The preferred model uses survey data and conditions worker-level tasks with occupational FE. This specification explains about 20% of the wage variation not accounted for in conventional (Mincerian) wage regressions.

Third, I show the omitted-variable bias in task returns estimated with expert data ranges from 24-34%, depending on specification. In the baseline model, this bias is nearly 30% and most sensitive to assumptions in the construction of task measures. I conceptualize this omitted-variable bias in a theoretical framework in which wages are determined by an individual- and occupation-level task dimension. This model accounts for workplace heterogeneity by highlighting the importance of task specialization within occupations. Since the best-performing specification does combine worker-level information on tasks with occupational FE, I view this theory supported by the data. These findings have important

[1] This data is derived from the BERUFENET Database, a free online portal for occupations provided by the German Federal Employment Agency, thus comparable to the O*NET database in the US.

implications for the interpretation of conventional task models. Economists often conceptualize the association between job tasks and wages with a Roy model in which comparative advantage governs occupational choice (Yamaguchi 2012, Cortes 2016, Cavaglia & Etheridge 2020, Boehm, von Gaudecker & Schran 2021). Subsequently, workers receive occupation-level task returns in their chosen occupation. My findings suggest, however, these task returns are substantially inflated due to confounding with underlying workplace heterogeneity.

Fourth, I present methodological guidance for practitioners seeking to work with task data. The robustness checks in this paper identify assumptions underlying the definition of tasks and occupations as key drivers of differences in statistical performance between survey and expert data. Researchers should therefore pay close attention to classification of tasks and occupations. Specifically, the bias in occupation-level task returns estimated with expert data is higher if (i) occupations are defined broadly (e.g., 2-digit level) and (ii) tasks are defined narrowly (e.g., five task groups). Moreover, specifications that use occupation-level task measures derived from aggregated survey responses display the worst statistical performance. The statistical discrepancies are overall negligible, but, compared to other specifications, these task measures lead to substantially larger point estimates for task returns. This finding warrants caution in the practice of linking aggregated task measures from survey data to other data sources.

2 Conceptual Background on Tasks and Wages

In this section, I discuss the role of tasks in the process of wage determination and highlight potential origins of bias in task data. In general, the task approach allows the researcher to study skills based on observations on job tasks. As workers have different levels of skill, they will be differentially compensated depending on their ability to perform tasks on the job. Variation in observed tasks thus allows the researcher to draw conclusions about underlying skill differences.

To illustrate this idea, I follow Autor & Handel (2013) and let worker i be employed in occupation o in which she receives a wage w in return for performing J tasks. Subsequently, she combines these tasks to produce output according to[2]

$$Y_{io} = exp\left(\alpha_o + \sum_J \lambda_{jo} T_{ij} + \mu_i\right) \tag{1}$$

[2]Note the output price in each occupation is normalized to unity. As pointed out in Autor & Handel (2013), this assumption is not restrictive as a logarithmic change in the price of output can be re-expressed in form of multiplicative change in the exponential term of eq. (1). For instance, think of productivity shifters embodied in the tasks workers perform, possibly reflecting market demand factors and affecting the output price that way.

where T_{ij} denotes task j performed by i and $\lambda_{jo} \geq 0$ represents returns earned for performing task j in o, i.e., task returns are occupation-specific. The parameters α_o and μ_i reflect, respectively, an occupation-specific constant and worker-specific error term. Assuming she is being paid her marginal product, I write her log wage as

$$ln\ w_i = \alpha_o + \sum_J \lambda_{jo} T_{ij} + \mu_i \tag{2}$$

This wage equation is identical to Autor & Handel (2013), implying i's wage is determined by her individual job activities T_{ij}. Next, to conceptualize quality differences in labor, I expedite on the idea that employers hire workers with similar, but not identical, skills. To this end, replace the generic constant α_o with occupation-level activities T_{jo}:

$$ln\ w_i = \sum_J \beta_{jo} T_{jo} + \sum_J \lambda_{jo} T_{ij} + \mu_i \tag{3}$$

where T_{jo} measures occupational skill requirements based on occupation-specific tasks and is described by the average task content among N workers employed in occupation o, i.e., $T_{jo} = \frac{1}{N} \sum_i T_{ij}$. These tasks are compensated with β_{jo} and may differ from individual-level returns λ_{jo}.

Since these skill requirements apply to *all* workers in a given occupation, T_{jo} gives rise to occupational sorting in spirit of ROY (1951). In these Roy-type models, a job is defined as an occupation and workers choose a job that maximizes their expected earnings. Viewing T_{jo} as a representation of occupation-specific skill requirements therefore illustrates occupational sorting resulting from a set of core tasks needed to produce output.

Yet, this framework may be overly restrictive by assuming workers in occupation o perform the same set of tasks. If this were true, all variation in T_{ij} would be entirely absorbed by occupation-level tasks T_{jo}. Otherwise, the implied equivalence will not hold and both task dimensions, T_{ij} *and* T_{jo}, determine i's wage.

The key departure in this model from a Roy-type framework is thus its degree of task specialization. I think of T_{jo} as capturing occupational heterogeneity in spirit of ROY (1951), i.e., occupations compensate tasks differentially. In comparison, I interpret T_{ij} as capturing workplace heterogeneity, i.e., specialization in a subset of tasks is compensated differentially across individuals —in part because they are employed at different firms within the same occupation. This assumption captures idiosyncratic differences in job tasks, but is also motivated by a growing literature documenting firm-specific wage premiums as a contributor to rising wage inequality (Abowd, Kramarz & Margolis 1999, Card, Heining & Kline 2013, Song, Price, Guvenen, Bloom & von Wachter 2019). In my model I conceptualize this insight

on workplace heterogeneity by allowing for individual task specialization *within* occupations.

Omitted-variable Bias & Relationship to Conventional Methods

While T_{jo} is readily available, T_{ij} is usually not available in the data. For this reason, researchers often rely on occupation-level task data that is derived from expert assessment and approximate the relationship between wages and tasks as follows:

$$ln\ w_i = \sum_J \beta_{jo} T_{jo} + \mu_i + \epsilon_i \tag{4}$$

where ϵ_i represents a standard i.i.d. error term. This specification is closely related to Roy-type models by assuming the relationship between wages and tasks is sufficiently described by occupation-level tasks. In this paper, I test whether the assumptions embedded in expert task data lead to biased estimates in β_{jo} by confounding occupation-level task returns with individual task specialization as a result of disregarding workplace heterogeneity. To this end, I study the potential for omitted-variable bias. In order to fix ideas, assume the relationship between individual-level and occupation-level tasks follows:

$$T_{ij} = \delta_j T_{jo} + \nu_i \tag{5}$$

where ν_i is an i.i.d. error term. I interpret δ as task pass-through. This parameter describes the responsiveness of individual activities to variation in occupation-level tasks. The model I propose allows for task specialization within occupations. At one extreme, a value of $\delta = 1$ implies perfect pass-through, i.e. variation at the occupation-level trickles down to the invidividual-level one-by-one. In contrast, a value of $\delta = 0$ implies no task specialization within occupations. Hence, $0 < \delta < 1$ implies imperfect pass-through from task variation at the occupation- to the individual-level. Plugging eq. (5) into (3) yields, after some rearranging, the following wage equation:

$$ln\ w_i = \sum_J (\beta_{jo} + \lambda_{jo}\delta_j) T_{jo} + (\epsilon_i + \lambda_{jo}\nu_i) \tag{6}$$

This model highlights the classic *omitted-variable bias*, implying conventional regressions in spirit of eq. (4) yield biased estimates of (occupation-level) task returns T_{jo} unless (i) $\beta_{jo} = 0$ or (ii) $\delta_j = 0$.

The first assumption (i) is likely not satisfied as workers self-select into occupations based on individual skills (Autor & Handel 2013). The second assumption (ii) captures workplace heterogeneity via task pass-through from the occupation- to the individual-level. Pronounced

task specialization *within* occupations implies workers in said occupation perform a different set of tasks, i.e., $\delta_j > 0$. In this case, assumption (ii) is likewise violated and occupation-level task returns based on conventional wage regressions, such as eq. (6), are biased upwards.

3 Data

3.1 Data Sources

I. Survey Data

The first data source is a series of German employment surveys, assembled by the Federal Institute for Vocational Education (BIBB) and the Federal Institute of Occupational Safety and Health (BAuA), respectively, in 2012 and 2018. This data set establishes a repeated labor force cross-section on qualification and working conditions in Germany, covering 20,000 workers in each wave. See Hall, Siefer & Tiemann (2014) and Hall, Hünefeld & Rohrbach-Schmidt (2020) for data manuals for each of the surveys used in this study.

Three key features make the BIBB/BAuA employment surveys suitable for the present study. First, workers self-report job-related activities. While the primary interest of expert-based data is on the occupational dimension, the unit of interest in survey data is the workplace (Dengler, Matthes & Paulus 2014). Having data at the (aggregated) occupation- and (disaggregated) worker-level thus permits an analysis on the presence of omitted-variable bias as described in section (2). Second, compared to other surveys with task information at the individual level, the BIBB/BAuA data offers a comparably sizable sample.[3] Third, each of the employment surveys provides information on monthly labor income. Income reporting allows me to study the effects of individual variation in tasks on wages. Expert-based data by itself, on the other hand, must be combined with other data sources to infer wage implications. I convert nominal income levels into real terms using CPI=100 as of 2015 and calculate the hourly wage rate using information on weekly hours worked and assuming that each individual works 8 hours per day.[4]

[3]For instance, the PDII data, used in Autor & Handel (2013), has a limited sample size of around 2,500 observations. In order to construct a consistent sample, comprising at least two observations per occupation, Autor & Handel (2013) only have 1,333 observations at their disposal. See Rohrbach-Schmidt & Tiemann (2013) for a comprehensive comparison among task data sets.

[4]The CPI data (FRED 2022) is taken from the Federal Reserve Bank of St. Louis (FRED) and can be downloaded under the following link: `https://fred.stlouisfed.org/series/DEUCPIALLAINMEI`. Accessed 28 March 2022.

II. Expert Data

The second data source is derived from the BERUFENET Database, a free online portal for occupations provided by the German Federal Employment Agency (BA). This database is a popular research tool for people seeking career guidance and exploring job placements. Occupations must offer legally regulated vocational training to be included in the database and provide a rich set of occupation-specific information, including common tasks. Using data compiled by DMP, I gather information on the relative importance of occupation-level tasks. This database is conceptually similar to the frequently used O*NET data in the US.[5]

III. Combined Data

The key variables are tasks performed on the job. DMP use information on occupational requirements from 2011-2013 for their classification. To broadly match this time horizon, I use survey data from 2012 and 2018. I average out task information across all years to enhance statistical precision and merge both data sources via occupational identifiers. Occupations are measured in terms of the 3-digit definition of the official BA Classification of Occupations, issue 2010 (KldB 2010). This classification scheme has a high degree of compatibility with the International Standard Classification of Occupations 2008 (ISCO-08), thus making it comparable with international classifications.

A key disadvantage of the BIBB/BAuA data is its lack of representativeness of the entire workforce. For instance, only workers with a sufficient command of the German language are asked to participate, favoring the native workforce disproportionately. Due to non-random sorting of native and foreign workers into occupations (Peri & Sparber 2009, Storm 2021), occupations and their composition are thus not properly represented.

Despite this shortcoming, I conduct the empirical analysis with a focus on potential biases in expert data. I proceed that way for two reasons. First, this approach is consistent with the model outlined in section (2) and thus anchors the discussion of results on the grounds of theory. In particular, the key implication of this model is that wages are determined by both task dimensions —individual *and* occupational level (see eq. 3). Testing this hypothesis necessarily requires survey data as expert data only contains occupation-level information. Second, exploring biases at the individual-level is notoriously difficult and usually requires experimental evidence. Focusing on potential biases in aggregated data instead is more feasible in practical terms.

[5]Like its American counterpart O*NET, BERUFENET is not solely based on expert assessment but rather the result of a process. Apart from external assessment, job descriptions rely on descriptions on vocational training, analysis of vacancies, information from job seekers and employers, and input from various economic associations. Despite this caveat, BERUFENET contains a substantial element of external assessment on its platform.

Of course, the limitations pertaining to non-representative representation of the workforce, along with common survey concerns such as small sample, warrant caution in the interpretation of empirical results. To gauge the severity of sample issues, I run a number of robustness checks in section (5) with varying sample criteria. Moreover, I discuss some overarching concerns of survey-based task data in the discussion of the results in section (6).

3.2 Task Construction

Initially, I follow Autor, Levy & Murnane (2003) and Spitz-Oener (2006) by pooling activities reported in the surveys into five narrow task categories: (i) Non-Routine (NR) Analytic tasks, (ii) NR Interactive tasks, (iii) Routine Cognitive tasks, (iv) Routine Manual tasks, and (v) NR Manual tasks.

In the second step, I alleviate measurement error from an overly narrow classification (Rohrbach-Schmidt & Tiemann 2013) by adopting the classification proposed in Acemoglu & Autor (2011). This strategy entails subsuming analytic and interactive tasks under "Abstract", involving strong problem-solving skills. Similarly, routine cognitive and routine manual tasks are subsumed under "Routine", characterized by activities following explicit and codifiable rules. Non-Routine manual tasks, on the other hand, are not categorized further and subsequently referred to as "Manual".

Table (1) provides an overview of activities included in these task categories. The reported information moreover offers a comparison between task data derived from BERUFENET (column 3) and the BIBB/BAuA surveys (4). Column (5) displays further descriptions on underlying activities.

[Table (1) here]

For the purpose of task construction I make use of two sections in the survey. In one part, workers report whether they perform specific activities (i) often, (ii) sometimes, or (iii) never. In the baseline analysis, I use a conservative approach, assuming they perform tasks only if they engage in underlying activities "often". This assumption alleviates concerns on measurement error as humans are prone to erroneous self-assessment and may thus overstate the importance of secondary job tasks (Pallier, Wilkinson, Danthiir, Kleitman, Knezevic, Stankov & Roberts 2002).

In another section of the survey, workers provide information on the degree of competencies required in some activities, such as basic math and software applications. Specifically, workers describe whether their job requires (i) professional skills, (ii) basic skills, or (iii) no skills at all. Once more, I opt for a conservative approach by assuming a skill is only required

if it warrants professional knowledge. In Table (1), I highlight which requirements are derived from actual task information (T) compared to those derived from skill levels (S). The latter information is concentrated in routine cognitive activities in order to broadly match the prevalence of various administrative and IT-related duties described in BERUFENET.

In the construction of the individual task content T_{ij} I follow DMP, who themselves apply a common definition introduced by Antonczyk, Fitzenberger & Leuschner (2009). Let A_j denote the number of activities a included in task group j and let A denote the total number of activities a across all j. I then define the individual task content T_{ij} as follows:

$$T_{ij} = \frac{\text{No. of activities } a \text{ performed by } i \text{ in task category } j}{\text{Total no. of activitites a by } i \text{ across all } j\text{'s}} = \frac{\sum_{a=1}^{A_j} d_{iaj}}{A} \tag{7}$$

where $j = 1$ (Abstract), $j = 2$ (Routine), and $j = 3$ (Manual) reflect the three task categories. Hence, for each worker i, I compare the number of activities a belonging to j relative to all activities A. This definition implies $\sum_J T_{ij} = 1$. Intuitively, eq. (7) describes the relative importance of each task category. Pertaining to the empirical implementation, the task vector $\boldsymbol{T_i} = \big(T_{i1}, T_{i2}, ..., T_{iJ}\big)$ is based on a series of dummy variables that, using eq. (7), are subsequently converted into a continuous measure $T_{ij} \in [0, 1]\ \forall j$.

For example, if worker i, Jane, indicates she performs three abstract, one routine, and one manual activity, then her abstract, routine, and manual task content, respectively, is 0.6, 0.2, and 0.2. Therefore, 60% of Jane's overall activities comprise abstract tasks, and 20% each, with respect to routine and manual.

By collecting individual responses of Jane's N_o peers who are likewise employed in occupation o, I compute leave-out-mean (LOM) averages at the occupation-level $\forall j$:

$$T_{jo}^{S} = \frac{1}{N_o} \sum_i T_{ij} \text{ if data source = Survey} \tag{8a}$$

$$T_{jo}^{Exp} = T_{jo} \text{ if data source = BERUFENET} \tag{8b}$$

where T_{jo}^{S} represents occupation-specific averages across individual responses and T_{jo}^{Exp} is taken from DMP, comprising occupation-level task measures assessed by labor market experts. I use LOM averages to alleviate concerns regarding a spurious correlation between individual- and occupation-level task measures derived from survey data. Using eq. (7 - 8b) thus provides me with task measures at the individual- and occupation-level. The primary interest lies in a comparison of models containing individual- and occupation-level tasks, respectively, to gauge the severity of the omitted-variable bias. Moreover, a comparison of models using occupation-level tasks from survey and expert data, respectively, offers insight

into systematic differences in the assessment of job tasks between experts and the average worker.

3.3 Sample Selection & Summary Statistics

To be included in the baseline sample, observations in the survey must meet three criteria. First, individual tasks need to be observed. Second, occupations can be matched to BERUFENET. Third, workers must not be civil servants nor self-employed, thus being subject to social security payments. Applying these restrictions leaves a total sample comprising 27,777 workers. Table (2) provides descriptive statistics on the sample, especially a comparison of the relative importance of tasks based on the BIBB/BAuA surveys (column 2) and BERUFENET (3).

[Table (2) here]

One key difference stands out regarding narrow task definitions. Workers report that one out of four activities are interactive tasks. In comparison, expert data suggests only one out of seven activities are interactive. Within the broader definition of abstract tasks, however, both data sources lead to similar conclusions. Accordingly, abstract tasks represent a bit less than half of all job activities, while two fifths of tasks consist of routine activities instead. For my baseline analysis I use broad tasks, thereby alleviating measurement error resulting from the classification of single activities into broader task groups.

[Table (3) here]

Comparing the relative importance of tasks by occupations, Table (3) illustrates one more difference in both data sets. Survey data offers a more balanced view on the task composition of jobs as many workers report to perform most activities in some capacity. In contrast, expert data has several occupations highly specialized in one particular task category. Exemplary, the abstract task content among the ten most abstract-intensive occupations ranges from 0.98-1 in expert data and 0.65-0.81 in survey data. Overall, both data sets identify similar occupations in terms of their dominant task, however. Abstract-intensive occupations comprise many teaching jobs and scholars, whereas routine-intensive occupations comprise many industrial jobs. On the other hand, manual-intensive occupations include many personal services such as care-taking.

4 Empirical Analysis

The model laid out in section (2) suggests estimation of task returns is prone to omitted-variable bias if task data is derived from external assessment. These data comprise occupation-level information, therefore assuming all workers within an occupation perform a common set of task. This assumption naturally disregards workplace heterogeneity, which is an important reason for wage differences (Card, Heining & Kline 2013). This section analyzes the importance of task specialization within occupations and quantifies the resulting omitted-variable bias in occupation-level returns to tasks.

4.1 Methodology

As a starting point, I first run task regressions in spirit of eq. (5):

$$T_{ij} = \delta T_{jo}^{Exp} + \boldsymbol{\mu X_i} + \eta_r + \theta_s + \nu_i \tag{9}$$

where T_{ij} reflects individual-level tasks as defined in eq. (7). T_{jo}^{Exp} represents occupation-level tasks derived from expert data, per eq. (8b). The vector $\boldsymbol{X_i}$ comprises control variables.[6] Lastly, ν_i denotes an i.i.d. error term.

Of key interest is the coefficient δ, capturing task pass-through, i.e. the extent to which occupation-level variation in tasks trickles down to worker-level variation. Perfect task pass-through implies $\delta = 1$. In contrast, values of $\delta < 1$ imply imperfect pass-through and values of $\delta = 0$ imply no role for task specialization within occupations. To assess the predictive elements embodied in tasks, I subsequently run a series of wage regressions comprising task measures at the individual- and occupation-level. The key regression takes the following form:

$$ln\ w_i = \boldsymbol{\lambda T_i} + \boldsymbol{\beta T_o^k} + \boldsymbol{\mu X_i} + \eta_r + \theta_s + \epsilon_i \tag{10}$$

where w_i is the hourly real wage for individual i and $\boldsymbol{T_o^k}, k = S, Exp$, denotes occupation-level tasks derived from survey and expert data, respectively.

The key coefficients are embedded in the vector λ, capturing individual-level task returns and thus account for workplace heterogeneity. A comparison to β, comprising occupation-level task returns, is informative about the magnitude of the omitted-variable bias. In

[6] I include the following control variables: demographic characteristics (age, age squared; dummies for sex, urban/rural, citizenship (native/ foreign), education dummies (college degree, vocational schooling, no vocational degree), and firm- and occupation-specific variables (firm tenure, firm tenure squared, occupational tenure, occupational tenure squared, firm size indicator). Moreover, η_r and θ_s, respectively, denote 16 regional dummies (state-level) and 34 sectoral dummies (industrial, craft, commerce, services, others).

a similar exercise, I replace $\boldsymbol{T_o^k}$ by up to 139 (3-digit) occupational dummies to test the importance of task specialization *within* occupations in more detail (Autor & Handel 2013). Apart from the choice of task data, all regressions are identical and weighted by survey weights.

To asses the relative importance of task measures across specifications formally, I report (i) Adjusted R^2, (ii) F-test for joint significance of tasks, (iii) incremental R^2 measures, (iv) Akaike (AIC) and Bayesian (BIC) Information Criterion, and (v) Ramsey test. While the first three measures offer insight on the goodness of fit across specifications, the AIC and BIC shed light on model selection resulting from minimized out-of-sample prediction errors. The Ramsey test formally checks whether the model is misspecified and thus may have an omitted-variable problem. While this test is typically used in the context of the linearity assumption in standard regression models, I nonetheless view it indicative on the presence of an omitted-variable problem.

Lastly, three more aspects are worth mentioning. First, by construction, all tasks combined add up to 1. To avoid multicollinearity I thus omit manual tasks, which subsequently serve as reference task. Second, for similar reasons, I omit workers who have not completed any vocational schooling. The reference group therefore consists of workers (i) with no vocational degree and (ii) who perform mainly manual tasks. Since these workers are typically found in lower parts of the wage distribution, I expect positive and sizable task returns. Third, the vector of coefficients $\boldsymbol{\lambda}$ should not be interpreted as task returns in a causal sense as non-random assignment of workers into occupations introduces selection bias (Autor & Handel 2013). OLS results should thus be treated with caution. Nonetheless, Stinebrickner, Stinebrickner & Sullivan (2019) find task returns from OLS and FE specifications to be similar. Their findings therefore suggest OLS regressions provide credible suggestive evidence on task returns.

4.2 Results

I. Task Returns: Survey vs Expert Task Data

Table (4) summarizes results on δ, the task pass-through from variation at the occupation- to the individual-level. The findings are consistent with imperfect task pass-through since $\delta < 1$. In quantitative terms, each 1 pp. increase in expert tasks is associated with an increase in individual-level tasks by 0.31 - 0.39 pp. Hence, only about a third of the occupation-level variation in tasks trickles down to the worker-level.

[Table (4) here]

Table (5) shows results from wage regressions. As a baseline, columns (1) - (3) display task returns based on specifications that include, respectively, occupation-level survey data, individual-level task data, and occupation-level expert data. All three models reveal significant and positive estimates on task returns. For instance, column (2) indicates performing 1 pp. more abstract tasks, relative to performing manual tasks, raises log wages by 0.53 points at the individual-level. Point estimates are broadly similar based on expert data, yet, are substantially larger when survey-based tasks are aggregated at the occupation-level.

Columns (4) and (5) combine tasks at the individual- and each occupation-level measure. Individual-level variation remains robust and economically meaningful to inclusion of occupational measures derived from survey or expert data. These findings reaffirm previous research, suggesting idiosyncratic factors in the task content are an important component in the process of wage determination (Autor & Handel 2013, Cassidy 2017, Rohrbach-Schmidt 2019). Including task measures at the individual- and occupation-level in a wage regression, however, shrinks all coefficients on task returns compared to specifications with only one task dimension. Hence, part of the effect of tasks on wages is attributed to the omitted task dimension. This observation lends credence to the theoretical wage equation (3), accounting for both task dimensions.

Next, I quantify the magnitude of the omitted-variable bias from the perspective of conventional wage regressions that use expert data with task information at the occupation-level. Exemplary, I examine the case of abstract tasks and collect estimates on the task pass-through (δ), along with estimates of task returns at the individual-level (λ) and occupation-level (β). Plugging results from Table (4) and Table (5), column (5), into the wage equation with presumed omitted-variable bias (eq. 6) yields: $\lambda\delta+\beta = 0.37\times0.36+0.32 = 0.13+0.32 = 0.45$.

This value is very close to the estimate of 0.46 in Table (5), column (3), displaying a conventional wage regression using only expert task data. The fact these two values are almost identical lends credence to the omitted-variable formula derived in the theoretical section of this paper (eq. 6). Consequently, using 0.46 as reference value for occupation-level task returns, 28% ($\frac{\lambda\delta}{\lambda\delta+\beta} = \frac{0.13}{0.46}$) of occupation-level returns to performing abstract tasks in fact reflect workplace heterogeneity. Following similar logic, the omitted-variable bias of occupation-level returns to performing routine tasks amounts to 29%.[7]

[Table (5) here]

[7]This calculation goes as follows: $\lambda\delta + \beta = 0.20 \times 0.31 + 0.14 = 0.06 + 0.14 = 0.20$, which is very close the estimate of 0.21 found in Table (5), column (3). Hence, expert data overstates occupation-level returns associated with routine tasks by 29% (0.06/0.21)

II. Statistical Performance: Survey vs Expert Task Data

This section compares the statistical performance of task models relying on survey and expert data, respectively. Overall, these exercises reveal no uniformly superior model. On the one hand, the information criteria at the bottom of Table (5) point to a prominent role for the idiosyncratic task dimension. Both, AIC and BIC, suggest models comprising individual-level task measures have smaller out-of-sample prediction error relative to conventional occupation-level measures. On the other hand, F-tests on joint significance of tasks indicate all task measures explain statistically significant portions of wage variation. In a similar vein, Adj. R^2 is essentially the same for all specifications. From a statistical point of view, all task measures thus perform quite similar.

This observation appears puzzling at first. On the one hand, the results suggest expert assessment on the importance of job tasks does not fundamentally differ from worker assessment. On the other hand, if individual-level task data were to provide more information on job activities than common occupation-level task data, would we not expect superior statistical performance? This somewhat surprising result is reinforced by the fact that the Ramsey test does not reject the null hypothesis of no omitted-variable bias for specifications containing individual and expert-based tasks, respectively. In contrast, the null hypothesis is rejected in all specifications using occupation-level tasks derived from survey data, however. Potential measurement error in the BIBB/BAuA surveys that can be attributed to non-representative representation of workforce and occupations may thus be amplified by aggregation of this data.

To shed more light on the role of omitted-variable bias in conventional wage regressions, I inspect raw correlations between all variables. To cause sizable omitted-variable bias, omitted task measures must be correlated with (i) wages and (ii) other independent variables. Table (6) shows these conditions are only partially fulfilled. Correlation between tasks and wages is modest and only relevant with respect to abstract tasks. The correlation between tasks and other regressors is likewise modest in most instances. Consequently, none of the regressions systematically over- or under-predict the data, as illustrated in the residual plots in Figure (1).

[Table (6) here]

The only variables that are highly correlated with each other are (i) all task measures and (ii) task measures and occupational characteristics. This observation has three important implications. First, the high correlation between occupation-level tasks from survey and expert data reinforces the view that expert assessment on the importance of job tasks does not

fundamentally differ from the assessment of the average worker in a given occupation. Second, modest correlation between tasks and wages stresses substantial heterogeneity in wage variation and explains similar statistical performance of wage regressions. Third, comparable statistical properties of different task dimensions do not hide the fact that economists must be cautious about the interpretation of task models. The sizable, yet imperfect, correlation between individual and occupation-level tasks suggests a substantial fraction of task returns commonly subsumed under occupational returns in fact mask underlying workplace heterogeneity.

To address heterogeneity in task models more explicitly, consider column (6) in Table (5). In this specification, I account for occupational affiliation via FE. This model has the best statistical properties among all specifications tested. Consistent with theory laid out in section (2), the most convincing task model thus accounts for task specialization within occupations.

The last exercise in this section quantifies the unique contributions from individual-level tasks more rigorously. To this end, I compute incremental R^2 measures. The baseline measure is the squared semipartial correlation associated with each task measure and summarized in Table (7). For reference: Expert task data explains 18.9% of variation that is not accounted for in traditional Mincerian wage regressions (column 3). But how much wage variation remains unexplained in these conventional task models?

According to Table (7), column (5), individual-level differences in tasks explain 5.8% of the unique variation not accounted for by expert data nor any other covariates. These contributions are driven primarily by abstract tasks. Column (6) underlines the most successful model combines individual-level tasks conditional on occupational FE, explaining 20% of total wage variation. This observations reinforces prior findings on workplace heterogeneity in job tasks. Analysis based on a related measure, squared partial correlation, leads to similar conclusions.

Notably, specifications containing occupation-level measures from BIBB/BAuA display the least explanatory power (columns 1 and 4). Combined with above evidence on the omitted-variable bias in specification relying only on occupation-level task measures, their underwhelming statistical performance raises concerns about the validity of linking occupation-level tasks from BIBB/BAuA to other data sources.

[Table (7) here]

5 Robustness

Baseline results present evidence on the omitted-variable bias in occupation-level task returns of around 30%. At the same time, most specifications reveal similar statistical properties of task models using survey and/ or expert data. Naturally, these findings may be influenced by sample properties and assumptions underlying the construction of task measures. To gauge the validity of baseline results, this section thus performs a number of robustness exercises.

5.1 Robustness Tests

The first set of robustness tests addresses restrictions implied by sample selection. In baseline specifications, I make no restrictions on income and employment to preserve statistical precision. Now, I restrict the sample to workers with an hourly wage of at least 5 EUR and a weekly minimum of 15 hours workers. In a separate exercise, I only consider occupations with at least 100 observations to alleviate outlier effects resulting from a small number of workers in an employment spell. These analyses also offer sensitivity checks on the impact of non-random sampling in the surveys.

The second set of robustness tests aims at the definition of occupations. In baseline models, I define occupations at the 3-digit level. However, in some applications such narrow definitions may not be available. I therefore repeat the analysis using a broader classification of occupations at the 2-digit level instead.

The third and final set of robustness tests considers alternative task definitions. To this end, I perform four more robustness checks. One, in baseline specifications I assume workers perform a task only if underlying activities are performed "often". I expand on this definition and assume workers perform a task if underlying activities are performed "often" *or* "sometimes". Two, I use a narrow classification of tasks by splitting abstract tasks into non-routine (NR) analytic and NR interactive and, respectively, routine (R) tasks into R cognitive and R manual. Three, I use "Routine" as omitted task category, thus providing results on relative returns for abstract and *manual* tasks. Lastly, baseline task measures use information on (i) actual tasks performed and (ii) the level of skill required for some routine activities. In a final exercise, I follow prior research and create more traditional task measures, relying on task information (i) only.

For brevity, I restrict the presentation of robustness tests to specifications (i) containing individual-level data from BIBB/BAuA and expert data and (ii) individual-level data conditional on occupational FE. The comparably worse performance of occupation-level tasks from BIBB/BAuA generally carries over to the robustness checks, albeit with some sensitivity to the chosen specification. Nonetheless, I do not consider this exercise adding much

extra insight to baseline results. A full set of robustness tests is available from the author upon request.

5.2 Robustness Results

To start off, Table (9) summarizes robustness checks on task regressions. The key takeaways do not change as $\delta < 1$ in all specifications. This finding reinforces baseline evidence on the imperfect task pass-through from occupation-level measures to the individual level.

Table (10) provides robustness checks on the omitted-variable bias based on a task model containing survey and expert data. The results on task returns are similar to those in baseline specifications. However, note the null hypothesis of the Ramsey test is rejected in half the cases. This findings suggests presence of omitted-variable bias that is not captured by any task measure is sensitive to the specific model.

To quantify the omitted-variable bias associated with expert task data, consider, exemplary, abstract tasks and the baseline estimate of $\beta = 0.46$ (from Table 5, column 3).[8] Applying the omitted-variable bias formula from eq. (6) to each robustness exercise implies an omitted-variable bias on the order of 24%-34%, well in the range of baseline estimates of 28%. Note that the lower bound, 24%, is inferred from column (6), using "Routine" as reference task group, and the upper bound, 34%, is inferred from column (3), defining occupations at the 2-digit level.[9] The largest impact on the omitted-variable bias is therefore attributed to the definition of occupations and tasks.

Overall, these findings lend credence to the omitted-variable bias in task returns estimated from expert data and are further supported by models with (i) occupational FE (Table 11) and (ii) a narrow task classification (Table 12). Notably, the null hypothesis of the Ramsey test cannot be rejected in all but one specification in Tables (11) - (12). Hence, these models reduce the risk of a misspecified model.

The comparison of statistical properties of task models using survey or expert data remains broadly consistent with baseline specifications. If anything, the robustness exercises suggest statistical superiority of survey data. This observation is especially supported by Tables (13) - (14), displaying incremental R^2 measures for each of the above robustness

[8]The estimated return on abstract tasks in the model containing only expert data has remained stable in all robustness specifications, ranging from 0.43-0.48. Therefore, using the estimate from the baseline specification does not fundamentally change the takeaways from these robustness exercises.

[9]The calculation for the lower bound goes as follows: $\lambda\delta + \beta = 0.17 \times 0.36 + 0.18 = 0.06 + 0.18 = 0.24$, where 0.36 is the same task pass-through as in the baseline analysis and the remaining numbers taken from Table (10), column (6). Note that the estimate for γ is equal to 0.25 in a specification comprising only expert tasks (omitted in output). Hence, the bias is equal to 24% (= 0.06/0.25). Similarly, the calculation for the upper bound goes as follows: $\lambda\delta + \beta = 0.37 \times 0.42 + 0.26 = 0.16 + 0.26 = 0.42$, where all numbers are taken from Table (9), column (5), and Table (10), column (5). Hence, the bias is equal to 34% (= 0.16/0.46).

checks. Survey data explains more unique wage variation than expert data in five out of seven robustness checks. However, the only two specifications in which expert data explains more unique variation —sample restrictions on income and employment (Table 13, column 1) and exclusion of competencies in task construction (Table 13, column 5) —reveal negligible differences between survey and expert data.[10]

In contrast, several robustness checks indicate that the unique wage variation explained by survey data exceeds the unique wage variation explained by expert data by more than 30%. This finding applies to specifications in which I (i) use a broad definition of occupations at the 2-digit level (Table 13, column 3), (ii) assume workers perform activities underlying tasks "often" *or* "sometimes" (Table 13, column 4), or (iii) adopt a narrow classification of five (rather than three) tasks (Table 14).

Overall, these tests support the conjecture that the omitted-variable bias in task returns estimated with expert data is primarily attributed to missing workplace heterogeneity. This hypothesis is reinforced in explicit robustness checks of task specialization within occupations. Table (15) summarizes the results of various specifications in which individual-level task data is conditioned with occupational FE. In most instances, the unique wage variation explained by survey data is close to the baseline estimates of around 20%.

6 Discussion of Results

The empirical analysis has shown the omitted-variable bias of task returns in expert data ranges from 24-34%, suggesting the importance of occupational characteristics in conventional task models is inflated. Does this mean researchers should no longer use expert data? Not necessarily. While most specifications indeed favor survey data in terms of unique explanatory power in wage regressions, the statistical performance of survey and expert task data is likewise broadly similar in many instances. From a statistical point of view, the choice of task data is thus somewhat arbitrary.

Even though worker-level survey data performs slightly better, some readers may rightfully wonder why task data derived from worker assessment has not performed *substantially* better. After all, are workers not the ones performing these tasks? Throughout this paper, I focus on omitted-variable bias in task returns as key limitation in expert data. However, survey data has its own set of limitation. In the data section (3) I already discuss a limitation

[10]Note that exclusion of competencies has a disproportionate effect on the explanatory power of survey data. This information in the survey is concentrated in routine cognitive activities and I include it in baseline specifications to broadly match the prevalence of various administrative and IT-related duties described in BERUFENET. Omitting these skill competencies leads to substantial differences in the measurement of routine tasks between survey and expert data at the detriment of survey data.

specific to the BIBB/BAuA data: Non-random sampling. By being non-representative of the German workforce nor of the occupational composition, the data naturally introduces measurement error.[11]

An overarching point of criticism is somewhat speculative, but well-grounded in the psychology literature. People are generally prone to *overconfidence bias* by displaying greater confidence in their (subjective) ability than justified by their (objective) performance (Brenner, Koehler, Liberman & Tversky 1996). This erroneous self-assessment is especially common in the cognitive domain (Pallier, Wilkinson, Danthiir, Kleitman, Knezevic, Stankov & Roberts 2002) and may thus induce workers to overstate the importance of abstract tasks in their job. The Dunning-Kruger Effect (Kruger & Dunning 1999) suggests this cognitive bias may be especially pronounced among workers with lesser ability, possibly leading them to overstate the complexity of their job. Combined, this insight from the psychology literature warrants some caution on the credibility of worker's self-reported assessment on job tasks.

Despite these shortcomings in the survey data, however, I argue economists should be careful about the interpretation of task models. Viewing my results through the lens of a Roy model suggests around 30% of (occupation-level) task returns can in fact be attributed to task specialization within occupations. I view this caveat important as many studies use a Roy framework to conceptualize the relationship between variation in (occupation-level) tasks and variation in wages (Yamaguchi 2012, Cortes 2016, Cavaglia & Etheridge 2020, Boehm, von Gaudecker & Schran 2021). In these models, workers choose an occupation as a result of comparative advantage. Crucially, using expert data implicitly assumes comparative advantages are muted within occupations. The results in this study, however, do not support this implicit assumption, implying occupation-level returns are substantially inflated.

I view this insight especially relevant in the context of a growing literature that has attributed rising wage inequality, observed in many countries, to worker and firm heterogeneity (Card, Heining & Kline 2013, Barth, Bryson, Davis & Freeman 2016, Song, Price, Guvenen, Bloom & von Wachter 2019, Dostie, Li, Card & Parent 2020). This research does not find occupational sorting to be the primary reason for trends in wage inequality. Instead, these studies highlight increasing segmentation in the labor market along the firm dimension. Accordingly, high-wage workers are increasingly employed at high-wage firms and vice versa for low-wage workers and low-wage firms. Combined with evidence on within-firm wage inequality, this literature thus stresses workplace heterogeneity as an important factor to understand rising occupation-level wage differences. In general, data availability makes a

[11] These data limitations are potentially worrisome in the context of aggregating worker-level information on tasks at the occupation-level. However, non-representative composition of workers implies occupational averages derived from survey data will not result in an unbiased estimator of the population. Linking survey-based task data at the occupation-level to other data sources may thus introduce measurement error.

detailed analysis on task specialization within firms challenging. In light of the rising degree of specialization at the workplace, I view enhanced task specialization a plausible mechanism for workplace heterogeneity (Becker, Egger, Koch & Muendler 2019, Cortes & Salvatori 2019).

7 Conclusions

This paper compares German survey data, comprising information on self-reported job tasks of more than 27,000 workers, and data derived from an online job platform, comprising expert assessment on the importance of job tasks, to test for the omitted-variable bias suspected in conventional expert data. I show that occupation-level returns on tasks estimated with expert data introduces omitted-variable bias ranging from 24-34%. Motivated by a theoretical framework, I argue this bias is largely attributed to the fact that conventional task data ignores workplace heterogeneity, thereby disregarding task specialization *within* occupations.

The evidence presented in this study reinforces the importance of individual-level variation in tasks as an important element in the process of wage determination (Autor & Handel 2013, Cassidy 2017, Rohrbach-Schmidt 2019). On the one hand, this information is useful in some applications in which heterogeneity is of key interest, e.g., research goals aimed at explaining rising heterogeneity within occupations (Hershbein & Kahn 2018, Atalay, Phongthiengtham, Sotelo & Tannenbaum 2018, 2020, Deming & Noray 2019, Modestino, Shoag & Ballance 2019).

Yet, opting for occupation-level data may be justified in many instances, as survey and expert data display broadly similar statistical properties in various specifications. To minimize measurement error when using expert task data, my findings suggest researchers should strive to adopt a (i) rather broad definition of three tasks (e.g., abstract, routine, manual) and (ii) sufficiently narrow definition of occupations (at least at the 3-digit level).

My results do warrant caution, however, regarding the common practice of aggregating (survey-based) worker-level information at the occupation-level and linking these aggregated task measures to other data sources via occupational identifiers. Specifications containing these task measures show the worst statistical performance and result in substantially larger task returns than other task measures, possibly because they confound common concerns in survey data (e.g., non-random sampling, small sample size) with aggregation bias. In the German context, given that BERUFENET covers the years 2011-13, applications with a recent and short time window may indeed warrant consideration of expert data. For a long-term analysis, however, spanning several decades, BIBB/BAuA is likely still the preferred

choice to account for ongoing trends such as technological change and globalization.

Finally, I recommend researchers consider the limitations of expert task data in regards to interpretation. While I doubt any central *qualitative* conclusions in task models are affected by my results, this paper suggests *quantitative* repercussions. For instance, labor economists often conceptualize returns to skills in a Roy model in which comparative advantage governs workers' occupational choice. Using conventional task data, however, yields occupation-level task returns that are inflated by about 30%. Occupational characteristics thus capture the association between task specialization and wages only partially. Possibly, the missing link is rising firm heterogeneity, documented by a growing literature on assortative sorting between of workers into firms (e.g., Card, Heining & Kline 2013). While data limitations do not permit a detailed analysis on the tasks performed within firms to date, collection of this type of data is a promising direction for future research to better understand the degree of task specialization in the modern workplace.

References

Abowd, J. M., Kramarz, F. & Margolis, D. N. (1999), 'High wage workers and high wage firms', *Econometrica* **67**(2), 251–333.

Acemoglu, D. & Autor, D. (2011), Skills, tasks and technologies: Implications for employment and earnings, Vol. 4 of *Handbook of Labor Economics*, Elsevier, pp. 1043–1171.

Antonczyk, D., Fitzenberger, D. & Leuschner, U. (2009), 'Can a task-based approach explain the recent changes in the german wage structure?', *Jahrbücher für Nationalökonomie und Statistik* (229), 214–238.

Atalay, E., Phongthiengtham, P., Sotelo, S. & Tannenbaum, D. (2018), 'New technologies and the labor market', *Journal of Monetary Economics* **97**, 48–67.

Atalay, E., Phongthiengtham, P., Sotelo, S. & Tannenbaum, D. (2020), 'The evolution of work in the united states', *American Economic Journal: Applied Economics* **12**(2), 1–34.

Autor, D. H. (2013), 'The "task approach" to labor markets: an overview', *Journal for Labour Market Research* **46**(3), 185–199.

Autor, D. H. & Handel, M. J. (2013), 'Putting tasks to the test: Human capital, job tasks, and wages', *Journal of Labor Economics* **31**(S1), S59–S96.

Autor, D. H., Levy, F. & Murnane, R. J. (2003), 'The skill content of recent technological change: An empirical exploration', *The Quarterly Journal of Economics* **118**(4), 1279–1333.

Barth, E., Bryson, A., Davis, J. C. & Freeman, R. (2016), 'It's where you work: Increases

in the dispersion of earnings across establishments and individuals in the United States', *Journal of Labor Economics* **34**(S2), S67–S97.

Becker, S. O., Egger, H., Koch, M. & Muendler, M.-A. (2019), 'Tasks, occupations, and wage inequality in an open economy'.

Boehm, M., von Gaudecker, H.-M. & Schran, F. (2021), 'Occupation growth, skill prices, and wage inequality'.

Brenner, L. A., Koehler, D. J., Liberman, V. & Tversky, A. (1996), 'Overconfidence in probability and frequency judgments: A critical examination', *Organizational Behavior and Human Decision Processes* **65**(3), 212–219.

Card, D., Heining, J. & Kline, P. (2013), 'Workplace heterogeneity and the rise of West German wage inequality*', *The Quarterly Journal of Economics* **128**(3), 967–1015.

Cassidy, H. (2017), 'Task variation within occupations', *Industrial Relations: A Journal of Economy and Society* **56**(3), 393–410.

Cavaglia, C. & Etheridge, B. (2020), 'Job polarization and the declining quality of knowledge workers: Evidence from the UK and Germany', *Labour Economics* **66**, 101884.

Cortes, G. M. (2016), 'Where have the middle-wage workers gone? a study of polarization using panel data', *Journal of Labor Economics* **34**(1), 63–105.

Cortes, G. M. & Salvatori, A. (2019), 'Delving into the demand side: Changes in workplace specialization and job polarization', *Labour Economics* **57**, 164–176.

de La Rica, S., Gortazar, L. & Lewandowski, P. (2020), 'Job tasks and wages in developed countries: Evidence from piaac', *Labour Economics* **65**, 101845.

Deming, D. & Noray, K. (2019), 'STEM careers and the changing skill requirements of work', *Working Paper* .

Dengler, K., Matthes, B. & Paulus, W. (2014), 'Occupational tasks in the german labour market: An alternative measurement on the basis of an expert database', *FDZ Methodenreport 201412* .

Dostie, B., Li, J., Card, D. & Parent, D. (2020), 'Employer policies and the immigrant-native earnings gap', *NBER Working Paper No. 27096* .

FRED (2022), 'Consumer price index: All items for germany'.
URL: *https://fred.stlouisfed.org/series/DEUCPIALLAINMEI*

Hall, A., Hünefeld, L. & Rohrbach-Schmidt, D. (2020), 'BIBB/BAuA-Erwerbstätigenbefragung 2018 - Arbeit und Beruf im Wandel, Erwerb und Verwertung beruflicher Qualifikationen'.

Hall, A., Siefer, A. & Tiemann, M. (2014), 'BIBB/BAuA-Erwerbstätigenbefragung 2012 - Arbeit und Beruf im Wandel, Erwerb und Verwertung beruflicher Qualifikationen'.

Hershbein, B. & Kahn, L. B. (2018), 'Do recessions accelerate routine-biased technological

change? evidence from vacancy postings', *American Economic Review* **108**(7), 1737–1772.

Kruger, J. & Dunning, D. (1999), 'Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments', *Journal of Personality and Social Psychology* **77**(6), 1121–1134.

Modestino, A. S., Shoag, D. & Ballance, J. (2019), 'Upskilling: Do employers demand greater skill when workers are plentiful?', *The Review of Economics and Statistics* **4**, 1–46.

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L. & Roberts, R. D. (2002), 'The role of individual differences in the accuracy of confidence judgments', *The Journal of general psychology* **129**(3), 257–299.

Peri, G. & Sparber, C. (2009), 'Task specialization, immigration, and wages', *American Economic Journal: Applied Economics* **1**(3), 135–169.

Rohrbach-Schmidt, D. (2019), 'Putting tasks to the test: The case of germany', *Social Inclusion* **7**(3), 122–135.

Rohrbach-Schmidt, D. & Tiemann, M. (2013), 'Changes in workplace tasks in Germany—evaluating skill and task measures', *Journal for Labour Market Research* **46**(3), 215–237.

Roy, A. D. (1951), 'Some thoughts on the distribution of earnings', *Oxford Economic Papers* **3**(2), 135–146.

Song, J., Price, D. J., Guvenen, F., Bloom, N. & von Wachter, T. (2019), 'Firming up inequality', *The Quarterly Journal of Economics* **134**(1), 1–50.

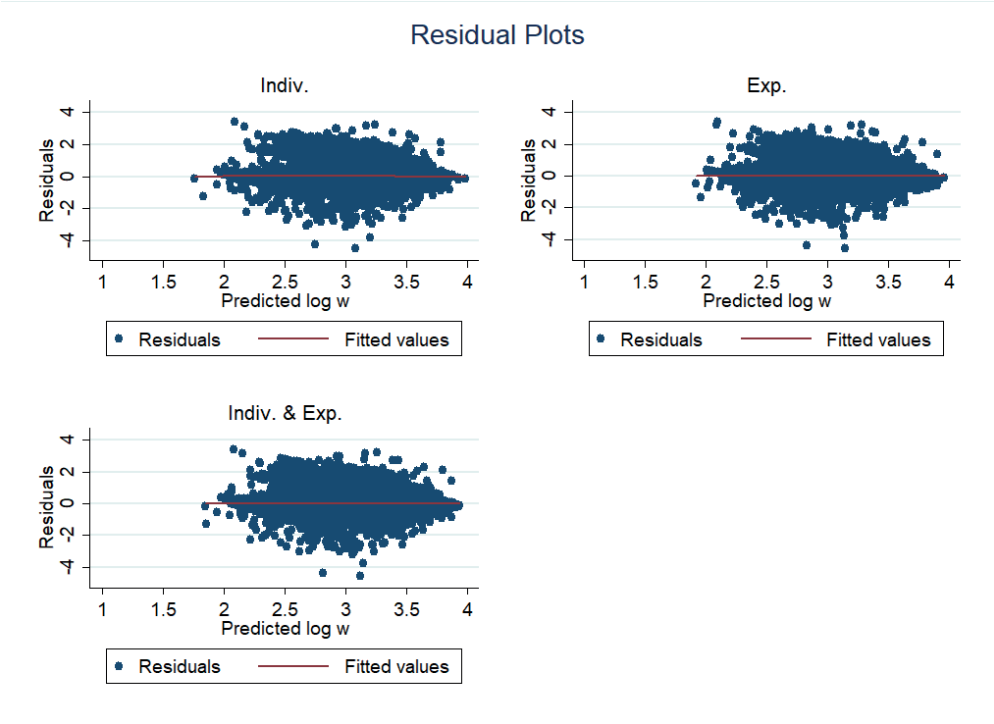
Spitz-Oener, A. (2006), 'Technical change, job tasks, and rising educational demands: Looking outside the wage structure', *Journal of Labor Economics* **24**(2), 235–270.

Stinebrickner, R., Stinebrickner, T. & Sullivan, P. (2019), 'Job tasks, time allocation, and wages', *Journal of Labor Economics* **37**(2), 399–433.

Storm, E. (2021), 'Task specialization and the native-foreign wage gap: Evidence from worker-level data', *Ruhr Economic Papers* (928).

Yamaguchi, S. (2012), 'Tasks and heterogeneous human capital', *Journal of Labor Economics* **30**(1), 1–53.

A Figures



NOTE. —The panels display residual plots of three regression models. The panel in the top left ("Indiv") uses survey data at the individual-level. The panel in the top right ("Exp") uses expert data at the occupation-level. The panel at the bottom ("Indiv & Exp.") combines both task dimensions.

Figure 1: Residual Plots of Wage Regressions containing Individual- and Expert-based Task Measures

B Tables

Table 1: Task Categories and their Contents

(1)	(2)	(3)	(4)	(5)
Task Category (Broad)	**Task Category (Narrow)**	**Requirements (BERUFENET)**	**Requirements (BIBB/BAuA)**	**Task Content**
Abstract	Non-Routine Analytic	Management, Planning & Supervision, Fields of Competencies, Economy, Leadership, Network Certifications, Monitoring, Music, Singing, Ballet, Musical Instruments, Optics, Applying Laws Design, Design (Art), Analysis, Control, Therapy, Programming	Research, Analyze (T) Plan, Construct (T) Design, Create, Evaluate (T) Work out Rules/ Regulations (T) Employ or Manage Staff (T)	Gathering Information, Investigating, Researching Organizing, Making Plans, Decision Making Constructing, Devloping, Evaluating Working with Computers, Programming Managing Personnel, Leading, Employing
	Non-Routine Interactive	Commerce, Counselling, Service, Support, Training, Marketing, Advertisting	Consult, Inform (T) Negotiate, Represent Interests (T) Teach, Train (T) Sell, Purchase, Acquire Customers (T) Advertise, Entertain, Present (T)	Consulting, Advising, Negotiating , Lobbying Teaching, Training, Educating Purchasing, Procuring, Selling Marketing, PR, Presenting
Routine	Routine Cognitive	Technology, Metrics, Administration, Graphics, Network Technology, Network Protocols Operating Systems, Certificates, Languages, Knowledge of Goods & Products, Competencies, Sensor Technology, Electronics, Mechanics, Mechanotrics, Hydraulics, Processing, Revision, Test, Inspection, Measurement, Monitoring, Procedures, Diagnostics	Correct Texts/ Data (T) Measure Length/ Height/ Temperature (T) Apply Languages (S) Calculate, Accounting (S) Application User Programs (S) IT Administration (T)	Use of Email, Internet Measuring, Evaluating Frequent Use of Foreign Languages Frequent Calculating/ Applying Basic Math and Statistics Frequent Use of Software database, Computer Programs Administration of database, Networks, IT-Systems
	Routine Manual	Cultivation, Farming, Construction, Manufacture, Production, Harvesting	Pack, Ship (T) Operate Machines (T) Process (T)	Planting, Storing, Transporting, Stocking, Posting Operating, Controling, Equipping Producing, Manufacturing Goods
Non-Routine	Non-Routine Manual	Dancing, Refurbishing, Service, Therapy (Manual Focus), Special/ Custom/ Bespoke Productions, Handicraft Businesses (Bakery, Carpentry, etc.)	Clean (T) Guard (T) Caretake (T) Repair, Renovate (T) Host (T)	Cleaning, Recycling Guarding Caretaking, Healing Repairing, Renovating, Restoring, Refurbishing Preparing Food, Serving

NOTE. —The BIBB/BAuA employment surveys differentiate between tasks workers actually perform at the workplace (T) and varying degrees of skills requreid for their job (S). In column (4) I highlight this distinction.

Table 2: Descriptive Statistics

(1)		(2)		(3)	
Socio-Economic		**Tasks (BIBB/BAuA)**		**Tasks (BERUFENET)**	
Log Wage	3.08	NR Analytic	0.21	NR Analytic	0.29
Female (% of total workforce)	0.53	NR Interactive	0.24	NR Interactive	0.15
Age	44.71	Routine Cognitive	0.29	Routine Cognitive	0.29
College Degree	0.21	Routine Manual	0.11	Routine Manual	0.09
Vocational Degree	0.73	NR Manual	0.15	NR Manual	0.18
Dropouts	0.06				
Hours worked (Weekly)	34.10	Abstract	0.45	Abstract	0.44
Tenure (Firm, in Years)	12.95	Routine	0.41	Routine	0.38
Tenure (Occup., in Years)	24.42	(NR) Manual	0.15	(NR) Manual	0.18
N = 27,777					

NOTE. —Descriptive statistics are weighted by sample weights.

Table 3: Top 10 Occupations in Abstract, Routine, and Manual Task Intensity: Survey vs Expert Data

	BIBB/BAuA Employment Surveys			BERUFENET	
		Abstract			
1	Glass Designing	0.81	1	Veterinary Medicine	1.00
2	Presenter and Hosts	0.68	2	Teachers (University)	1.00
3	Lobbyist	0.68	3	Musicians	1.00
4	Teachers (extra-curricular in driving and sports)	0.68	4	Social Scienstists	0.99
5	Teachers (extra-curricular for adults)	0.68	5	Teachers (extra-curricular for adults)	0.98
6	Economists	0.67	6	Humanities Scholars	0.98
7	Retail Jobs (Antiques, Arts, Books, Music)	0.66	7	Dentistry	0.98
8	Event Management	0.65	8	Teacher (General)	0.97
9	Public Relations Work	0.65	9	Public Relations Work	0.95
10	Management	0.65	10	Linguist and Literary Scholars	0.98
		Routine			
1	Musical Instrument Making	0.73	1	Metal Production	0.95
2	Ceramic Production	0.72	2	Printmaking	0.89
3	Printmaking	0.67	3	Metal Finishing	0.88
4	Metaldesigning	0.67	4	Plastics Manufacturing	0.87
5	Metalworking	0.67	5	Ceramic Production	0.87
6	Cartography	0.64	6	Glassmaking	0.87
7	Textile Manufacturing	0.63	7	Precision Engineering	0.82
8	Construction Plant Operators	0.61	8	Production of Beverages	0.81
9	Warehousing and Logistics	0.60	9	Mechatronics	0.80
10	Mineral Processing	0.59	10	Chemists	0.80
		Manual			
1	Cleaning	0.68	1	Motorists	1.00
2	Animal Care	0.63	2	Building Engineering	0.87
3	Groom	0.62	3	Civil Engineering	0.86
4	Traffic Control	0.45	4	Traffic Control	0.82
5	Housekeepipng	0.43	5	Body Care	0.81
6	Building Technology	0.41	6	Groom	0.80
7	Eldercare	0.38	7	Animal Care	0.75
8	Arts and Crafts	0.36	8	Cleaning	0.74
9	Food Preparation	0.33	9	Floristry	0.72
10	Catering	0.32	10	Drywall Installation	0.68

NOTE. —Occupations are defined at the 3-digit level, comprising a total of 139 professions.

Table 4: Task Regressions

	(1)	(2)	(3)
Dependent Variable:	Abstract (Ind.)	Routine (Ind.)	Manual (Ind.)
Abstract (Exp., Occ.)	0.36***		
	(0.01)		
Routine (Exp., Occ.)		0.31***	
		(0.01)	
Manual (Exp.)			0.39***
			(0.01)
Controls	✓	✓	✓
Adj. R^2	0.25	0.22	0.32
Observations	27777	27777	27777

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

NOTE. —This output displays estimates on the task pass-through, conceptually derived from (5) and estimated via (9).

Table 5: Task Measures as Wage Predictors: Survey vs Expert Data

Dependent Variable: Log Hourly Real Wage	(1)	(2)	(3)	(4)	(5)	(6)
Abstract (Occ.)	0.94***			0.60***		
	(0.06)			(0.06)		
Routine (Occ.)	0.42***			0.24***		
	(0.06)			(0.07)		
Abstract (Ind.)		0.53***		0.36***	0.37***	0.38***
		(0.03)		(0.04)	(0.03)	(0.03)
Routine (Ind.)		0.30***		0.19***	0.20***	0.20***
		(0.03)		(0.04)	(0.03)	(0.04)
Abstract (Exp., Occ.)			0.46***		0.32***	
			(0.03)		(0.03)	
Routine (Exp., Occ.)			0.21***		0.14***	
			(0.03)		(0.03)	
Survey tasks (Occupational)	✓			✓		
Survey tasks (Individual)		✓		✓	✓	✓
Expert tasks (Occupational)			✓		✓	
Occupation Dummies						✓
F (Task Measures, Occ.)	150.79		192.44	52.42	82.90	
F-pval (Task Measures, Occ.)	(0.00)		(0.00)	(0.00)	(0.00)	
F (Task Measures, Ind.)		145.90		57.48	64.90	66.32
F-pval (Task Measures, Ind.)		(0.00)		(0.00)	(0.00)	(0.00)
Adj. R^2	0.19	0.19	0.19	0.20	0.20	0.22
AIC	49008.01	48969.22	48992.12	48790.82	48730.88	48219.27
BIC	49501.93	49463.13	49494.27	49301.21	49241.26	49832.73
Ramsey-Test	2.57	1.35	1.10	2.92	2.14	0.59
Ramsey-Test (pval)	(0.05)	(0.26)	(0.35)	(0.03)	(0.09)	(0.62)
Observations	27777	27777	27777	27777	27777	27777

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

NOTE. —This output is based on (10). The first two rows display coefficients based on occupational averages derived from individual responses in the employment surveys ("Occ."). Point estimates corresponding to those individual responses are displayed in the third and fourth row ("Ind."). Lastly, the last two rows display coefficients based on occupational averages derived from the Expert database ("Exp."). All specifications include controls for gender, age, age squared, a dummy for living in an urban area, education dummies, occupational tenure, firm tenure, squared tenure for each dimension of experience, and a categorical variable reflecting firm size. Each regression is weighted by sample weights. The omitted task category is "Manual", task returns are thus relative to performing manual tasks.

Table 6: Correlation between Individual-level Tasks & other Covariates

	(1) Abstract (Ind.)	(2) Routine (Ind.)	(3) Abstract (Exp.)	(4) Routine (Exp.)
Abstract (Ind.)		-0.66***	0.47***	-0.28***
Routine (Ind.)	-0.66***		-0.21***	0.40***
Abstract (Exp., Occ.)	0.47***	-0.21***		-0.60***
Routine (Exp., Occ.)	-0.28***	0.40***	-0.60***	
Abstract (Occ., Survey)	0.53***	-0.29***	0.88***	-0.54***
Routine (Occ., Survey)	-0.27***	0.56***	-0.37***	0.71***
Log Hourly Real Wage	0.17***	0.02***	0.18***	0.01*
Occupation	0.33***	-0.29***	0.61***	-0.59***
College	0.25***	-0.05***	0.36***	-0.19***
Voca. Schooling	-0.18***	0.05***	-0.30***	0.17***
Female	0.11***	-0.16***	0.23***	-0.22***
Age	-0.02*	0.01	0.01	-0.02***
Age (Sq.)	-0.02**	0.01	0.01	-0.02***
Urban	0.06***	-0.02***	0.10***	-0.07***
Foreign	-0.00	-0.02*	-0.02**	-0.00
Firm Size	-0.05***	-0.00	-0.03***	-0.02***
Occ. Tenure	-0.07***	0.02***	-0.08***	0.03***
Occ. Tenure (Sq.)	0.09***	-0.03***	0.10***	-0.05***
Firm Tenure	0.01	0.06***	-0.02**	0.09***
Firm Tenure	0.01	-0.06***	0.04***	-0.11***
State	-0.03***	0.01	-0.01*	-0.00
Industry	0.16***	-0.27***	0.28***	-0.43***
Observations	27777	27777	27777	27777

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Unique Variation Explained by Task Measures

Model:	(1)	(2)	(3)	(4)	(5)	(6)
Abstract (Occ.)	12.4%			3.6%		
Routine (Occ.)	5.4%			1.4%		
Abstract (Ind.)		12.7%		3.2%	4.1%	13.5%
Routine (Ind.)		6.0%		1.1%	1.7%	6.4%
Abstract (Exp.)			13.8%		5.2%	
Routine (Exp.)			5.0%		2.1%	
Total (Occ.)	17.9%		18.9%	5.0%	7.3%	
Total (Ind.)		18.6%		4.3%	5.8%	19.9%

NOTE. —The displayed values represent the unique variation in log wages associated with the task measure of interest, expressed relative to the R-squared of the full model. Results are based on computing the squared semipartial correlation between log wages and the task measure of interest. Models (1)-(3) correspond to specifications including occupation-level tasks from Survey data ("(Occ.)"), individual-level tasks ("(Ind.)"), and occupation-level tasks from Expert data ("(Exp.)"), respectively. Models (4) and (5) combine individual-level tasks with occupation-level tasks from Survey and Expert data, respectively. Lastly, model (6) includes individual-level tasks and occupational FE. The two bottom rows summarize the variance in low wages associated with task measures of interest, which has not been explained by all other covariates (including other task dimensions). All specifications include controls for gender, age, age squared, a dummy for living in an urban area, education dummies, occupational tenure, firm tenure, squared tenure for each dimension of experience, and indicator for firm size.

Table 8: Incremental R-squared of Task Measures

Model:	(1)	(2)	(3)	(4)	(5)	(6)
Abstract (Occ.)	14.4%			4.2%		
Routine (Occ.)	6.5%			1.7%		
Abstract (Ind.)		14.5%		3.8%	4.8%	15.5%
Routine (Ind.)		7.0%		1.3%	2.0%	7.5%
Abstract (Exp.)			15.8%		6.2%	
Routine (Exp.)			6.0%		2.5%	
Total (Occ.)	20.9%		21.7%	5.9%	8.7%	
Total (Ind.)		21.5%		5.1%	6.9%	23.0%

NOTE. —The displayed values represent the percentage drop-off in R-squared after removing task measures and are relative to the R-squared of the full model. Results are based on computing the squared partial correlation between log wages and the task measure of interest. The model description for specifications (1)-(6) along with controls included is the same as in Table (7) described above. The two bottom rows summarize the importance of different dimensions of task measures by adding up the decrease in R-squared after removing individual- and occupation-level tasks, respectively, from the model.

Appendices

Table 9: Task Regressions: Robustness

Dependent Variable: T_{ij}	(1)	(2)	(3)	(4)	(5)
Abstract (Exp., Occ.)	0.35***	0.36***	0.37***	0.19***	0.38***
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)
Routine (Exp., Occ.)	0.30***	0.32***	0.29***	0.16***	0.31***
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)
Manual (Exp.)	0.37***	0.41***	0.39***	0.26***	0.40***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Wage/hr $\geq$ 5 & Empl. Hours $\geq$ 15	✓				
Observations per Occupation $\geq$ 100		✓			
Occupational Classification: 2-digit			✓		
Activities performed "often" OR "sometimes"				✓	
Task Construction: Exclude competencies					✓
Expert tasks (Occupational)	✓	✓	✓	✓	✓
Adj. R^2	0.25	0.34	0.30	0.35	0.30
Observations	26641	25468	27777	28026	27069

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

NOTE. —The first two rows display coefficients based on occupational averages derived from individual responses in the employment surveys ("(Occ.)"). Point estimates corresponding to those individual responses are displayed in the third and fourth row ("(Ind.)"). Lastly, the last two rows display coefficients based on occupational averages derived from the Expert database ("(Exp.)"). All specifications include controls for gender, age, age squared, a dummy for living in an urban area, education dummies, occupational tenure, firm tenure, squared tenure for each dimension of experience, and a categorical variable reflecting firm size. Each regression is weighted by the product of sample weight and occupation-specific workforce to account for size effects. The omitted task category is "Manual".

Table 10: Robustness Tests on Task Measures as Wage Predictors: Survey vs Expert Data

Dependent Variable: Log Hourly Real Wage	(1)	(2)	(3)	(4)	(5)	(6)
Abstract (Ind.)	0.37***	0.34***	0.42***	0.68***	0.38***	0.17***
	(0.04)	(0.03)	(0.03)	(0.06)	(0.03)	(0.02)
Routine (Ind.)	0.19***	0.18***	0.22***	0.40***	0.16**	
	(0.04)	(0.03)	(0.03)	(0.06)	(0.03)	
Manual (Ind.)						-0.20***
						(0.03)
Abstract (Exp., Occ.)	0.33***	0.35***	0.26***	0.29***	0.31***	0.18***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
Routine (Exp., Occ.)	0.16***	0.19***	0.14***	0.10***	0.15***	
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	
Manual (Exp., Occ.)						-0.14***
						(0.03)
Wage/hr $\geq$ 5 & Empl. Hours $\geq$ 15	✓					
Observations per Occupation $\geq$ 100		✓				
Occupational Classification: 2-digit			✓			
Activities performed "often" OR "sometimes"				✓		
Task Construction: Exclude competencies					✓	
Reference Task Group: Routine						✓
Expert tasks (Occupational)	✓	✓	✓	✓	✓	✓
Adj. R^2	0.20	0.20	0.19	0.20	0.20	0.22
Ramsey Test	4.30	1.32	1.78	6.02	1.97	2.15
Ramsey Test (pval)	(0.00)	(0.27)	(0.15)	(0.00)	(0.12)	(0.09)
Observations	26641	25468	27777	28026	27069	27777

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

NOTE. —This output provides robustness tests on the baseline estimates of the model described in eq. (10) and displayed in Table 5. For brevity, I only focus on the model of interest comprising individual- and expert data on tasks, i.e. column (5) of Table 5.

Table 11: Robustness Tests on Task Measures as Wage Predictors: Within-Occupation Task Specialization

Dependent Variable: Log Hourly Real Wage	(1)	(2)	(3)	(4)	(5)	(6)
Abstract (Ind.)	0.34***	0.38***	0.41***	0.68***	0.32***	0.18***
	(0.03)	(0.04)	(0.03)	(0.06)	(0.03)	(0.03)
Routine (Ind.)	0.17***	0.19***	0.21***	0.38***	0.09**	
	(0.03)	(0.04)	(0.03)	(0.07)	(0.03)	
Manual (Ind.)						-0.20***
						(0.04)
Wage/hr $\geq$ 5 & Empl. Hours $\geq$ 15	✓					
Observations per Occupation $\geq$ 100		✓				
Occupational Classification: 2-digit			✓			
Activities performed "often" OR "sometimes"				✓		
Task Construction: Exclude competencies					✓	
Reference Task Group: Routine						✓
Occupation Dummies (Survey data)	✓	✓	✓	✓	✓	✓
Adj. R^2	0.22	0.21	0.20	0.22	0.22	0.22
Ramsey Test	1.54	0.63	0.83	2.72	0.55	0.58
Ramsey Test (pval)	(0.20)	(0.60)	(0.48)	(0.04)	(0.65)	(0.63)
Observations	26641	25468	27777	28026	27069	27777

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

NOTE. —This output provides robustness exercises on task specialization within occupations, based on a model comprising individual-level tasks and occupational FE (see column 6 in Table 5).

Table 12: Task Measures as Wage Predictors: Narrow Task Definitions

Dependent Variable: Log Hourly Real Wage	(1)	(2)
NRA (Ind.)	0.40***	0.42***
	(0.04)	(0.04)
NRI (Ind.)	0.28***	0.31***
	(0.04)	(0.04)
RC (Ind.)	0.28***	0.29***
	(0.04)	(0.04)
RM (Ind.)	-0.03	0.01
	(0.04)	(0.04)
NRA (Exp.)	0.31***	
	(0.03)	
NRI (Exp.)	0.12***	
	(0.04)	
RC (Exp.)	0.08**	
	(0.03)	
RM (Exp.)	-0.02	
	(0.04)	
Controls	✓	✓
Occupation Dummies		✓
Adj. R^2	0.20	0.22
Ramsey Test	1.38	0.56
Ramsey Test (pval)	(0.25)	(0.64)
Observations	27777	27777

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

NOTE. —This output is based on eq. (10), but with a narrow definition of tasks instead. "NRA" represents Non-Routine Analytic Tasks, "NRI" reflects Non-Routine Interactive, "RC" is short for Routine Cognitive. Lastly, "RM" represents Routine Manual tasks. For brevity, this robustness exercise is restricted to a model comprising individual survey as well as expert data (column 1) and a model based on individual-level tasks and occupational FE (column 2).

Table 13: Robustness Tests on Unique Variation Explained by Task Measures: Survey vs Expert Data

Model: Individual-level & Expert-based Task Data	(1)	(2)	(3)	(4)	(5)	(6)
Abstract (Ind.)	5.7%	5.9%	7.4%	6.2%	4.4%	1.2%
Routine (Ind.)	1.8%	1.6%	2.2%	1.9%	0.2%	
Manual (Ind.)						1.9%
Abstract (Exp.)	5.6%	4.4%	2.9%	4.4%	3.2%	0.6%
Routine (Exp.)	2.3%	1.6%	1.6%	1.5%	1.8%	
Manual (Exp.)						1.5%
Total (Occ.)	7.9%	6.0%	4.5%	6.0%	4.9%	2.1%
Total (Ind.)	7.5%	7.4%	9.6%	8.0%	4.6%	3.1%
Wage/hr $\geq$ 5 & Empl. Hours $\geq$ 15	✓					
Observations per Occupation $\geq$ 100		✓				
Occupational Classification: 2-digit			✓			
Activities performed "often" OR "sometimes"				✓		
Task Construction: Exclude competencies					✓	
Reference Task Group: Routine						✓
Observations	26641	25468	27777	28026	27069	27777

NOTE. —The displayed values represent the unique variation in log wages associated with the task measure of interest, expressed relative to the R-squared of the full model. The output provides robustness checks on the baseline model comprising individual-level and expert-based task measures, see Table (7) for reference.

Table 14: Robustness Tests on Unique Variation Explained by Task Measures: Narrow Definition of Tasks

Model: Individual-level & Expert-based Task Data	(1)	(2)
NRA (Ind.)	4.4%	6.9%
NRI (Ind.)	0.6%	1.8%
RC (Ind.)	2.2%	6.2%
RM (Ind.)	0.1%	0.0%
NRA (Exp.)	2.7%	
NRI (Exp.)	0.5%	
RC (Exp.)	0.9%	
RM (Exp.)	0.3%	
Total (Occ.)	4.4%	
Total (Ind.)	7.3%	15.0%
Expert Tasks (Occupational)	✓	
Occupation Dummies		✓
Observations	27777	27777

NOTE. —The displayed values represent the unique variation in log wages associated with the task measure of interest, expressed relative to the R-squared of the full model. These specifications are based on a narrow definition of tasks. "NRA" represents Non-Routine Analytic Tasks, "NRI" reflects Non-Routine Interactive, "RC" is short for Routine Cognitive. Lastly, "RM" represents Routine Manual tasks. For brevity, this robustness exercise is restricted to a model comprising individual survey as well as expert data (column 1) and a model based on individual-level tasks and occupational FE (column 2).

Table 15: Robustness Tests on Unique Variation Explained by Task Measures: Within-Occupation Task Specialization

Model: Individual-level Tasks & Occupational FE	(1)	(2)	(3)	(4)	(5)	(6)
Abstract (Ind.)	13.8%	13.5%	14.3%	13.2%	10.5%	1.7%
Routine (Ind.)	5.4%	4.5%	5.2%	4.8%	0.7%	
Manual (Ind.)						4.8%
Total (Ind.)	19.1%	18.0%	19.4%	18.0%	11.2%	6.5%
Occupation Dummies	✓	✓	✓	✓	✓	✓
Hourly Wage ≥ 5	✓					
Observations per Occupation ≥ 100		✓				
Occupational Classification: 2-digit			✓			
Activities performed "often" OR "sometimes"				✓		
Task Construction: Exclude competencies					✓	
Reference Task Group: Routine						✓
Observations	26641	25468	27777	28026	27069	27777

NOTE. —The displayed values represent the unique variation in log wages associated with the task measure of interest, expressed relative to the R-squared of the full model. The output provides robustness checks on a model comprising individual-level and occupational FE, thereby accounting for task specialization within occupations. See Table (7) for reference.