

Storm, Eduard

Article

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

Journal for Labour Market Research

Provided in Cooperation with:

Institute for Employment Research (IAB)

Suggested Citation: Storm, Eduard (2023) : On the measurement of tasks: Does expert data get it right?, Journal for Labour Market Research, ISSN 2510-5027, Springer, Heidelberg, Vol. 57, Iss. 1, pp. 1-24,
<https://doi.org/10.1186/s12651-023-00332-z>

This Version is available at:

<https://hdl.handle.net/10419/278091>

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.



<https://creativecommons.org/licenses/by/4.0/>

ORIGINAL ARTICLE

Open Access



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

Eduard Storm*

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 26–34%. Motivated by a theoretical framework, I argue this bias results from expert data ignoring individual 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 occupational task returns 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.

Keywords Expert vs survey task data, Individual heterogeneity, Omitted-variable bias

JEL classification C18, J24, J31

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 and 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. Similarly, Autor and Handel (2013) point out that individual job tasks differ within education and demographic groups. Cassidy (2017) and Rohrbach-Schmidt (2019) provide similar evidence in the German context, Storm (2022b) shows differences in task specialization within occupations between natives and foreigners, and de la Rica et al. (2020) in a cross-country setting using PIAAC data, suggesting within-occupation heterogeneity in tasks is not country-specific. Related evidence on dispersion of tasks within occupations can be found in Spitz-Oener (2006), Atalay et al. (2018), Atalay et al. (2020), Deming and Noray (2019), Modestino et al. (2019), and Stinebrickner et al. (2019). These papers highlight rich heterogeneity in tasks that is masked in conventional occupation-level data and rising dispersion of tasks within occupations over time.

Importantly, the existing empirical literature echoes 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

*Correspondence:

Eduard Storm

Eduard.Storm@rwi-essen.de

Department Labor Markets, Education, Population, RWI-Leibniz Institute for Economic Research, Essen, Germany



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

occupational characteristics (Autor 2013, Dengler et al. 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 individual 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 measure 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 and 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 et al. (2014)—henceforth 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 26 to 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 individual 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 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 choices (Boehm et al. 2021; Caviglia and Etheridge 2020; Cortes 2016; Yamaguchi 2012). 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 individual heterogeneity. In this context, survey data has the advantage that it allows the researcher to aggregate individual responses at the occupation-level. Therefore, the researcher can account for task variation at the individual and occupation-level, thereby mitigating omitted variable bias with respect to task information.

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

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. I invite the interested reader to visit the following homepage to explore its information: <http://berufenet.arbeitsagentur.de/berufe/index.jsp>.

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 and 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 to²

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

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

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

This wage equation is identical to Autor and 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, I replace the generic constant α_o with occupation-level activities T_{jo} :

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

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} . Eq. (3) formalizes that data containing worker-level information or, respectively, occupation-level information explain unique parts of the wage variation, as in Autor (2013), and illustrated by the two distinct parameters β_{jo} and λ_{jo} in above model.

Since occupational 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 individual heterogeneity, i.e., specialization in a subset of tasks is compensated differentially across individuals. In my model I conceptualize this insight on individual 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:

$$\ln w_i = \sum_J \beta_{jo} T_{jo} + \mu_i + \epsilon_i \quad (4)$$

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 individual heterogeneity. To this end, I study the potential for omitted-variable bias.

² Note the output price in each occupation is normalized to unity. As pointed out in Autor and 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.

In order to fix ideas, assume the relationship between individual-level and occupation-level tasks follows:

$$T_{ij} = \delta_j T_{jo} + v_i \quad (5)$$

where v_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 individual-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:

$$\ln w_i = \sum_j (\beta_{jo} + \lambda_{jo} \delta_j) T_{jo} + (\epsilon_i + \lambda_{jo} v_i) \quad (6)$$

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

The first assumption (i) is likely not satisfied as workers self-select into occupations based on individual skills (Autor and Handel 2013). The second assumption (ii) captures individual 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

3.1.1 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 2011/2012 (Hall et al. (2020b), doi:10.7803/501.12.1.1.60) and 2017/2018 (Hall et al. (2020a), doi:10.7803/501.18.1.1.10). While interviews took place between October and March, I will refer to the surveys as 2012 and 2018 sample, respectively, for reasons of brevity. 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 Rohrbach-Schmidt and Hall (2018) and Rohrbach-Schmidt and Hall (2013) 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 et al. 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 (). 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 income, allowing 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. The key dependent variable of this paper is log hourly real wages, which I construct as follows. In the first step, I use information on monthly labor income stated by each worker individually. In the second step, I convert this income measure into real monthly income to adjust for purchasing power. To this end, I use data on the German Consumer Price Index (CPI), which is indexed CPI=100 as of 2015.⁴ Third, I calculate hourly wage rate by dividing the real monthly income by individually stated weekly working hours times four (weeks). This way, I account for differences in working hours by, for example, gender and occupations.⁵

3.1.2 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 or must be sufficiently relevant to be included in the database and provide a rich set of occupation-specific information, including common tasks. Overall, the database comprises more than 10,000 narrowly-defined

³ For instance, the PDII data, used in Autor and 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 and Handel (2013) only have 1,333 observations at their disposal. See Rohrbach-Schmidt and Tiemann (2013) for a comprehensive comparison among task data sets. A notable alternative in the German context is described in Matthes et al. (2014) in which the authors collect individual-level task data to be integrated into a wave of the German National Educational Panel Study (NEPS).

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.

⁵ This distinction matters especially in the gender context. The average man in my sample works 38 hours compared to 30 hours for the average woman. Regarding occupations, the average working hours across all (2-digit) occupations is 35.3 hours with a standard deviation of 3.4 hours. Therefore, weekly working hours for most occupations are in the range of 32 - 39 hours, broadly consistent with fulltime-work equivalents.

occupations (Matthes et al. 2008), however, only 3,900 of those occupations contain rich occupational information, such as tasks. This database is therefore conceptually similar to the frequently used O*NET data in the US. In comparison, however, O*NET comprises some 800 occupations that are part of the US Standard Occupational Classification (SOC) (Handel 2016).⁶ Like its American counterpart O*NET, BERUFENET is not solely based on expert assessment but rather the result of a process. To this end, experts use descriptions on vocational training, analysis of vacancies, information from job seekers and employers, and input from various economic associations to describe occupations.⁷

At the core of BERUFENET is the requirement matrix, containing 8,000 skills that are assigned to occupations. This requirement matrix is used for career counseling and therefore continuously updated with monthly checks to identify new requirements and redundancies. In contrast, O*NET is not updated as often, reducing its usefulness for the analysis of task changes over time (Autor 2013). On the flipside, O*NET and its predecessor, the Dictionary of Occupational Titles (DOT), have been used by social scientists for decades (Handel 2016), while the requirement matrix from BERUFENET has only been available since 2008, reducing its usefulness for long-term analysis to date.

This requirement matrix is the foundation for the database provided by DMP. The authors assign requirements to tasks following previous literature and implementing basic rules. I provide more details on this correspondence and differences to existing literature in Section 3.2 below. Using the data compiled by DMP, I gather information on the relative importance of occupation-level tasks and use this information as proxy for expert data on tasks.⁸ The DMP data is especially useful for research on occupational skill requirements and has been widely used ever since its release, for instance in the context of substitution potentials of the digital transformation (Dengler and Matthes 2018), the “greening of jobs” (Janser 2018), labour market entry (Reinhold and Thomsen 2017), and labour market mismatch (Kracke et al. 2018; Kracke and Rodrigues 2020; Storm 2022a). Relatedly, see Christoph

et al. (2020) for a comprehensive overview of relevant occupation-based measures for labour market research, including BERUFENET.

3.1.3 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. This approach moreover avoids a key drawback of the BIBB/BAuA data in the task context, as this data was never intended to operationalise tasks and faced changing survey mode and questionnaires over time (Rohrbach-Schmidt and Tiemann 2013). The data I use in the present study is nearly identical in terms of information on tasks, alleviating this type of measurement error. 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 few disadvantages of the BIBB/BAuA data remain, however. Notably, the data is not representative 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 and Sparber 2009; Storm 2022b), occupations and their composition are thus not representative. However, this data limitation is unlikely to affect my main analysis. In Table 1 I compare employment shares for 2-digit occupation using BIBB/BAuA data, collected in 2011–12 and 2017–18, and data from the German Socioeconomic Panel (SOEP) from the 2011, 2012, 2017, and 2018 surveys. The SOEP data (Liebig et al. (2021), doi:10.5684/soep.core.v36eu) is a representative, multi-cohort household survey with a large sample size that has been running since 1984 and is widely used in labor market research, especially in the German context.⁹ This comparison suggests employment shares in both surveys are broadly similar, alleviating concerns regarding non-random sorting in the BIBB/BAuA data.¹⁰

⁶ Note that O*NET has been updated in recent years. This implies that the number of occupations has changed to some extent. The latest O*NET taxonomy of 2019 comprises 1,016 occupations. However, only 867 of those correspond to jobs included in the SOC. See Gregory et al. (2019) for more details.

⁷ Specifically, occupational descriptions in BERUFENET are derived from labor market experts at the publishing company *BW Bildung und Wissen*: <https://bwverlag.de/> (accessed on 2022/08/12).

⁸ Specifically, I use their data based on a 3-digit occupational definition, using the more recent classification of occupations, issue 2010 (KldB 2010), rather than issue 1988 (KldB 1988). DMP point out some noticeable differences between KldB 2010 and KldB 1988 as the latter differentially maps tasks to occupations, especially for some technical occupations.

⁹ See Goebel et al. (2018) for a description and overview of its applications.

¹⁰ Related to this point, Storm (2022b) shows that, despite underrepresentativeness of foreign workers, the workforce composition by citizenship in BIBB/BAuA broadly matches workforce composition from administrative data from 1992–2018.

Table 1 Workforce composition in German survey data: BIBB/BAuA versus SOEP

(1) KLDB 2010 (2-digit level)	(2) Occ. Share (BIBB/BAuA) (%)	(3) Rank (BIBB/ BAuA)	(4) Occ. Share (SOEP) (%)	(5) Rank (SOEP)
Occupations in business management and organisation	8.0	1	11.3	1
Medical and health care occupations	7.9	2	8.9	2
Occupations in financial services, accounting and tax consultancy	4.8	8	6.8	3
Occupations in education and social work, housekeeping, and the theology	6.0	4	6.3	4
Sales occupations in retail trade	6.7	3	5.8	5
Occupations in law and public administration	5.5	6	5.6	6
Technical occupations in machine-building and automotive industry	4.6	9	5.1	7
Occupations in traffic and logistics (without vehicle driving)	5.1	7	4.3	8
Occupations in computer science, information and communication technology	3.0	14	4.1	9
Occupations in teaching and training	5.7	5	4.0	10
Occupations in mechatronics, energy electronics and electrical engineering	2.9	16	3.8	11
Occupations in purchasing, sales and trading	3.0	13	3.3	12
Occupations in technical R &D, construction, and production planning	2.0	19	3.3	13
Occupations in metal-making and -working, and in metal construction	2.5	18	2.9	14
Occupations in non-medical healthcare, body care, wellness and medical technicians	3.2	11	2.8	15
Drivers and operators of vehicles and transport equipment	3.0	12	2.7	16
Occupations in advertising and marketing, in commercial and editorial media design	1.7	22	2.5	17
Occupations in food-production and -processing	2.8	17	2.1	18
Occupations in building services engineering and technical building services	2.0	20	1.9	19
Occupations in tourism, hotels and restaurants	2.9	15	1.6	20
Occupations in cleaning services	3.6	10	1.6	21
Occupations in mathematics, biology, chemistry and physics	1.2	25	1.3	22
Occupations in construction scheduling, architecture and surveying	1.1	26	1.1	23
Occupations in plastic-making and -processing, and wood-working and -processing	1.3	24	1.0	24
Occupations in building construction above and below ground	1.4	23	1.0	25
Occupations in safety and health protection, security and surveillance	1.9	21	1.0	26
Occupations in paper-making and -processing, printing, and in technical media design	0.9	29	0.8	27
Occupations in interior construction	1.1	27	0.8	28
Occupations in in philology, literature, humanities, social sciences, and economics	0.3	34	0.5	29
Occupations in the performing arts and entertainment	0.6	31	0.4	30
Occupations in textile- and leather-making and -processing	0.4	32	0.4	31
Occupations in gardening and floristry	0.9	28	0.3	32
Occupations in geology, geography and environmental protection	0.2	36	0.2	33
Occupations in production and processing of raw materials, glass and ceramic	0.2	35	0.2	34
Occupations in product design, artisan craftwork, fine arts	0.4	33	0.2	35
Occupations in agriculture, forestry, and farming	0.8	30	0.1	36

N = 27,777

This table display the occupational employment share ("Occ. Share") of workers in the BIBB/BAuA data (column 2) and SOEP data (column 4). Likewise, columns (3) and (5), display the rank of each occupation in the employment share distribution for BIBB/BAuA and, respectively, SOEP. Occupations are defined at the 2-digit level, comprising a total of 36 professions. I use BIBB/BAuA data that has been collected in 2011–12 and 2017–18, and data from the German Socioeconomic Panel (SOEP) from the 2011, 2012, 2017, and 2018 surveys. For the BIBB/BAuA data, see Hall et al. (2020b) and Hall et al. (2020a), respectively. For the SOEP data, see Liebig et al. (2021)

Moreover, BIBB/BAuA contains a rather small number of specific job activities, at least compared to BERUFENET. This limitation naturally makes the

definition of tasks more sensitive to the number and type of underlying job activities. I address this concern in more detail in Sect. 5 by using alternative task definitions.

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 Sect. 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). Therefore, worker-level data naturally captures different parameters than occupation-level data.¹¹ Testing the hypothesis of this model, however, 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.

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 Sect. 5 with varying sample criteria. Moreover, I discuss some overarching concerns of survey-based task data in the discussion of the results in Sect. 6.

3.2 Task construction

Initially, I follow Autor et al. (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. This is the same task classification as in DMP, enhancing comparability between our task definitions.

In the second step, I alleviate measurement error from an overly narrow classification (Rohrbach-Schmidt and Tiemann 2013) by adopting the classification proposed

in Acemoglu and 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”.

These task groups are often portrayed in the context of complementarity and substitutability of workers with computer capital and robots (Acemoglu and Autor 2011). Abstract tasks are complementary with these technologies as they raise productivity of those working with them. Hence, a greater share of abstract tasks is associated with positive task returns. In contrast, routine tasks are substitutable with these technologies as machines are increasingly able to perform repetitive tasks previously performed by workers, but at a smaller cost. Hence, a greater share of routine tasks is associated with weaker task returns compared to abstract tasks. Manual tasks are the least affected by technological change because they involve personal services and require lots of hand-eye coordination. Jobs with a high share of manual tasks are typically found at the lower parts of the wage distribution, offering low returns.

Table 2 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.

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 et al. 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 2, I highlight which requirements are derived from actual task information (T) compared to those derived from skill levels (S). The task literature, e.g., Spitz-Oener (2006) usually only makes use of actual task information.

¹¹ Similarly, worker-level information derived from surveys introduces more variance. On the one hand, this variation is welcomed to exploit variation in tasks within occupations. On the other hand, this can be worrisome if some of the additional variance reflects measurement error, for instance in from of coding errors of occupational titles. However, the survey administrators are aware of this issue and take careful steps to reduce measurement error from occupational miscoding. To this end, occupational coding is performed by professionals at the data analytics company Kantar Public (KP). First, KP performs automatic encoding based on electronically available directories. Second, titles not identified in the first step, are subsequently manually encoded. For this purpose, KP assigns two separate professionals to encode occupations and thus reduce measurement error. In case of discrepancies, an experienced coder decides which code is appropriate. Dengler et al. (2014) perform a similar manual encoding strategy for assigning individual occupational requirements to tasks, in this case even with three manual encoders. See Rohrbach-Schmidt and Hall (2013) for details on strategies to reduce measurement error in BIBB/BAuA.

Table 2 Task categories and their contents

(1) Task category (Broad)	(2) Task category (Narrow)	(3) Requirements (BERUFENET)	(4) Requirements (BIBB/BAuA)	(5) 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, Developing, Evaluating IT-Development, 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)
	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)
		Routine manual	Cultivation, Farming, Construction, Manufacture, Production, Harvesting	Pack, Ship (T) Operate Machines (T) Process (T)
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

The BIBB/BAuA employment surveys differentiate between tasks workers actually perform at the workplace (T) and varying degrees of skills required for their job (S). In column (4) I highlight this distinction. I use BIBB/BAuA data that has been collected in 2011–12 and 2017–18, and data from BERUFENET, covering the years 2011–13. For the BIBB/BAuA data, see Hall et al. (2020b) and (Hall et al. (2020a), respectively. For the BERUFENET data, see Dengler et al. (2014)

In general, DMP and I follow this literature with respect to task classification, especially Spitz-Oener (2006) who uses similar data as we do. However, given that her classification procedure is about 20 years old, it is somewhat outdated. One key concern applies to the rising prevalence of various administrative and IT-related duties using computers, activities that are concentrated in routine cognitive activities. By virtue of being more recent, the DMP-classification accounts for these important changes at the workplace more effectively. Contrary to the task literature, I therefore augment the routine cognitive category by skill requirements to broadly match the DMP-classification. Another key difference between DMP and the existing literature is the classification of managerial tasks. While Spitz-Oener (2006) assigns these tasks to NR Interactive, DMP assign these tasks to NR Analytic. I follow DMP to maintain greater comparability of survey and expert data.

In the construction of the individual task content T_{ij} I likewise follow DMP, who themselves apply a common definition introduced by Antonczyk et al. (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:

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

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 $\mathbf{T}_i = (T_{i1}, T_{i2}, \dots, T_{ij})$ is based on a series of dummy variables that, using Eq. (7), are subsequently converted into a continuous measure $T_{ij} \in [0, 1] \forall j$. DMP adopt an equivalent strategy to assign occupational skill requirements to tasks. Their ‘‘DMP-task-index’’ (Dengler et al. 2014, p.17) relates the share of single occupational requirements that belong to task j to all occupational requirements. Again, similar approaches in aggregating data enhance the comparability of results based on survey- and expert-based data.

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 $\forall j$:

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

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

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. 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. Note that I do not classify occupations as abstract, routine, or manual occupations, respectively, which implicitly defines the dominant tasks within occupations. Instead, my primary

interest lies in describing the task composition of occupations and workers’ job activities. This way I can compare models containing individual- and occupation-level task compositions, respectively, to gauge the severity of the omitted-variable bias.

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 3 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).

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

Table 3 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					

I use BIBB/BAuA data that has been collected in 2011–12 and 2017–18, and data from BERUFENET, covering the years 2011–13. For the BIBB/BAuA data, see Hall et al. (2020b) and (Hall et al. (2020a), respectively. For the BERUFENET data, see Dengler et al. (2014). Descriptive statistics are weighted by sample weights

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.

Comparing the relative importance of tasks by occupations, Table 4 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. For instance, 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 caretaking.

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 individual heterogeneity, which is an important reason for wage differences (Card et al. 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):

$$T_{ij} = \delta T_{jo}^{Exp} + \mu X_i + \eta_r + \theta_s + v_i \quad (9)$$

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 X_i comprises control variables.¹² Lastly, v_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:

$$\ln w_i = \lambda T_i + \beta T_o^k + \mu X_i + \eta_r + \theta_s + \epsilon_i \quad (10)$$

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

¹² 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). Note that vocational schooling includes workers with initial and advanced vocational education. The latter group comprises, for instance, master craftsman and technicians. Moreover, η_r and θ_s , respectively, denote 16 regional dummies (state-level) and 34 sectoral dummies (comprising various industries within broader industrial, craft, commerce, and service sectors).

Table 4 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 Scientists	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	Housekeeping	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

N = 27,777

Occupations are defined at the 3-digit level, comprising a total of 139 professions. I use BIBB/BAuA data that has been collected in 2011–12 and 2017–18, and data from BERUFENET, covering the years 2011–13. For the BIBB/BAuA data, see Hall et al. (2020b) and (Hall et al. (2020a), respectively. For the BERUFENET data, see Dengler et al. (2014)

The key coefficients are embedded in the vector λ , capturing individual-level task returns and thus account for individual heterogeneity. Of course, this is only an imperfect measure for individual heterogeneity as it does not allow me to control for traits such as “ability”. However, the cross-sectional nature of the data does not allow me to account for some of the unobserved (time-invariant) heterogeneity. A comparison to β , comprising occupation-level task returns, is informative about the magnitude of the omitted-variable bias. In a similar exercise, I replace T_o^k by up to 139 (3-digit) occupational dummies to test the importance of task specialization *within* occupations in more detail (Autor and Handel 2013). Apart

from the choice of task data, all regressions are identical and weighted by survey weights.

To assess 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, and (iv) Akaike (AIC) and Bayesian (BIC) Information Criteria. 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.

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 λ should not be interpreted as task returns in a causal sense as non-random assignment of workers into occupations introduces selection bias (Autor and Handel 2013). OLS results should thus be treated with caution. Nonetheless, Stinebrickner et al. (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

4.2.1 Task returns: survey vs expert task data

Table 5 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 6 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

Table 5 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

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

This output displays estimates on the task pass-through, conceptually derived from (5) and estimated via (9). I use BIBB/BAuA data that has been collected in 2011–12 and 2017–18, and data from BERUFENET, covering the years 2011–13. For the BIBB/BAuA data, see Hall et al. (2020b, 2020a), respectively. For the BERUFENET data, see Dengler et al. (2014)

research, suggesting idiosyncratic factors in the task content are an important component in the process of wage determination (Autor and Handel 2013; Cassidy 2017; Rohrbach-Schmidt 2019).¹³ 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, while this finding is unsurprising given the correlation between individual- and occupational-level task measures (Table 7), this finding likewise suggests part of the effect of tasks on wages is attributed to the omitted task dimension. For instance, if all workers in an occupation were to perform the same set of tasks, all individual-level variation in tasks would be absorbed by occupation-level tasks, making inclusion of individual-level tasks obsolete. Since this is clearly the not the case, however, my findings lend credence to the theoretical wage Eq. (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. For instance, 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 Tables 5 and 6, column (5), into the wage equation with presumed omitted-variable bias (Eq. 6) yields: $\lambda\delta + \beta = 0.37 \times 0.36 + 0.32 = 0.13 + 0.32 = 0.45$.

This value is very close to the estimate of 0.46 in Table 6, 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

¹³ Note that, unlike the cited studies, the explanatory power in my regressions is smaller, especially in specifications including occupational FE (like in Autor and Handel (2013) and Rohrbach-Schmidt (2019)). I attribute this discrepancy primarily to four reasons. First, I use completely different task measures. For instance, while Autor and Handel (2013) and Rohrbach-Schmidt (2019) use a few generic task measures, I use a broad variety of production-related tasks. Second, above studies standardize their tasks measures using Principal Component analysis, while I follow Antonczyk et al. (2009), a definition that has been adopted widely ever since. Third, Autor and Handel (2013) use 240 (6-digit) occupational dummies, while Rohrbach-Schmidt (2019) uses (5-digit) 198 dummies. In contrast, I use 139 (3-digit) dummies. Fourth, our samples are different, making comparisons generally difficult.

Table 6 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*** (0.06)			0.60*** (0.06)		
Routine (Occ.)	0.42*** (0.06)			0.24*** (0.07)		
Abstract (Ind.)		0.53*** (0.03)		0.36*** (0.04)	0.37*** (0.03)	0.38*** (0.03)
Routine (Ind.)		0.30*** (0.03)		0.19*** (0.04)	0.20*** (0.03)	0.20*** (0.04)
Abstract (Exp., Occ.)			0.46*** (0.03)		0.32*** (0.03)	
Routine (Exp., Occ.)			0.21*** (0.03)		0.14*** (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
Observations	27777	27777	27777	27777	27777	27777

Robust standard errors in parentheses

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

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. I use BIBB/BAuA data that has been collected in 2011–12 and 2017–18, and data from BERUFENET, covering the years 2011–13. For the BIBB/BAuA data, see Hall et al. (2020b) and (Hall et al. (2020a), respectively. For the BERUFENET data, see Dengler et al. (2014)

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 individual heterogeneity. Following similar logic, the omitted-variable bias of occupation-level returns to performing routine tasks amounts to 29%.¹⁴

4.2.2 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 6 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. R² 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

¹⁴ 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 6, column (3). Hence, expert data overstates occupation-level returns associated with routine tasks by 29% (0.06/0.21)

job activities than common occupation-level task data, would we not expect superior statistical performance?

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 7 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 underpredict the data, as illustrated in the residual plots in Fig. (1).

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 individual heterogeneity.

To address heterogeneity in task models more explicitly, consider column (6) in Table 6. 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 Sect. (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 by running several wage regressions on the same set of controls and only changing task measures. Comparing models with survey-based and expert-based task

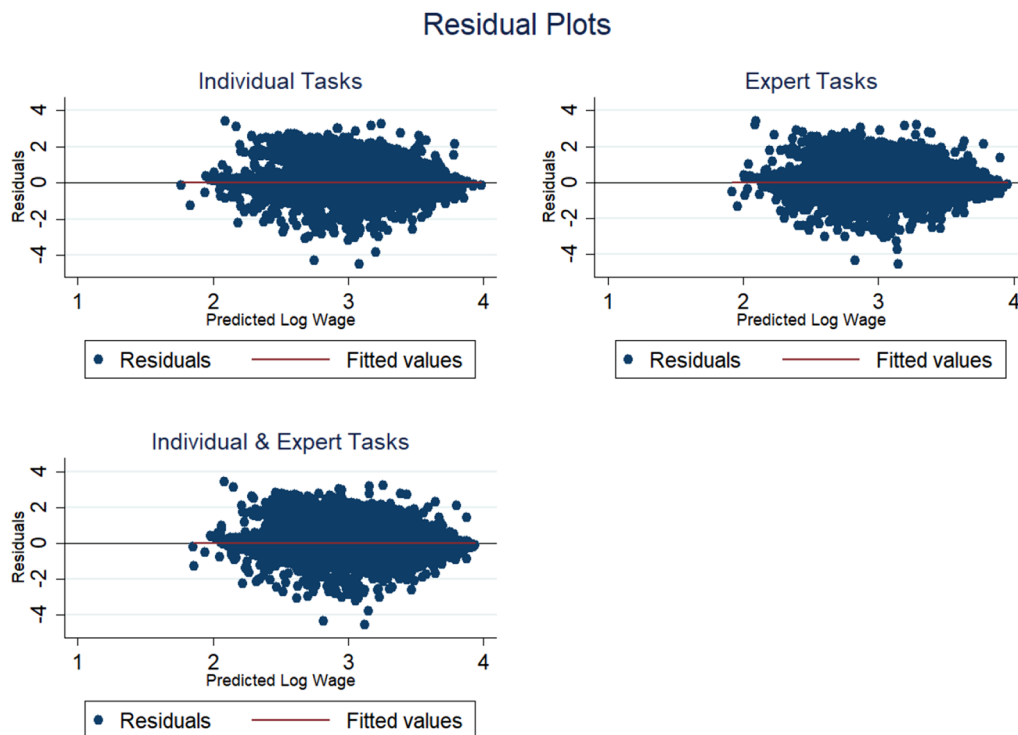


Fig. 1 Residual Plots of Wage Regressions containing Individual- and Expert-based Task Measures. NOTE.-The panels display residual plots of three regression models. The panel in the top left ("Indiv") uses survey data at the individual-level. The panel in the top right ("Exp") uses expert data at the occupation-level. The panel at the bottom ("Indiv & Exp.") combines both task dimensions. I use BIBB/BAuA data that has been collected in 2011–12 and 2017–18, and data from the German Socioeconomic Panel (SOEP) from the 2011, 2012, 2017, and 2018 surveys. For the BIBB/BAuA data, see Hall et al. (2020b) and (Hall et al. (2020a), respectively. For the BERUFENET data, see Dengler et al. (2014)

Table 7 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	27,777	27,777	27,777	27,777

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

I use BIBB/BAuA data that has been collected in 2011–12 and 2017–18, and data from BERUFENET, covering the years 2011–13. For the BIBB/BAuA data, see Hall et al. (2020b) and (Hall et al. (2020a), respectively. For the BERUFENET data, see Dengler et al. (2014)

Table 8 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%

N = 27,777

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. I use BIBB/BAuA data that has been collected in 2011–12 and 2017–18, and data from BERUFENET, covering the years 2011–13. For the BIBB/BAuA data, see Hall et al. (2020b) and (Hall et al. (2020a), respectively. For the BERUFENET data, see Dengler et al. (2014)

Table 9 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%

N = 27,777

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 8 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. I use BIBB/BAuA data that has been collected in 2011–12 and 2017–18, and data from BERUFENET, covering the years 2011–13. For the BIBB/BAuA data, see Hall et al. (2020b) and (Hall et al. (2020a), respectively. For the BERUFENET data, see Dengler et al. (2014)

measures, respectively, thus permits a comparison of unique variation in wages explained by either task measure. The baseline measure is the squared semipartial correlation associated with each task measure and summarized in Table 8. 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 8, 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 individual 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 specifications 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.

While both correlation coefficients are indicative of model quality, by virtue of explaining the unique wage variation associated with specific task measures, their interpretation must be treated with caution. Their validity depends on correct model specification. Hence, if important variables are missing from the model, but have an impact on wages, interpretation derived from comparing correlation coefficients will be misleading. Any

conclusions regarding model quality based on a comparison of correlation coefficients should thus be treated as suggestive evidence.

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 (Table 9).

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 five more

Table 10 Task regressions: robustness

Dependent Variable: T_{ij}	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Abstract (Exp., Occ.)	0.35*** (0.01)	0.36*** (0.01)	0.37*** (0.01)	0.19*** (0.00)	0.38*** (0.01)	0.22*** (0.00)	0.39*** (0.01)
Routine (Exp., Occ.)	0.30*** (0.01)	0.32*** (0.01)	0.29*** (0.01)	0.16*** (0.00)	0.31*** (0.01)	0.21*** (0.01)	0.33*** (0.01)
Manual (Exp.)	0.37*** (0.01)	0.41*** (0.01)	0.39*** (0.01)	0.26*** (0.01)	0.40*** (0.01)	0.33*** (0.01)	0.42*** (0.01)
Wage/hr ≥ 5 & Empl. Hours ≥ 15	✓						
Observations per Occupation ≥ 100		✓					
Occupational Classification: 2-digit			✓				
Activities performed “often” OR “sometimes”				✓			
Task Construction: Exclude competencies					✓		
Task Normalization a la Alda (2013)						✓	
2012 Sample Only							✓
Expert tasks (Occupational)	✓	✓	✓	✓	✓	✓	✓
Adj. R^2	0.25	0.34	0.30	0.35	0.30	0.39	0.24
Observations	26641	25468	27777	28026	27069	28015	13756

Robust standard errors in parentheses

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

This output displays robustness estimates on the task pass-through, conceptually derived from (5) and estimated via (9). Recall that this model regresses individual-level tasks on occupation-level tasks. 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”. I use BIBB/BAuA data that has been collected in 2011–12 and 2017–18, and data from BERUFENET, covering the years 2011–13. For the BIBB/BAuA data, see Hall et al. (2020b) and (Hall et al. (2020a), respectively. For the BERUFENET data, see Dengler et al. (2014)

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 account for the fact that task categories differ by number of tasks, which may confound my results. To this end, I adopt the method proposed in Alda (2013) to normalize the number of activities across categories. This normalization method weights tasks by the frequency with which workers perform them, i.e. “often,” “sometimes,” or “never.” Four, I restrict the sample to the 2012 survey only to align the time horizon of the survey data more closely to the time horizon in DMP. Lastly, baseline task measures use information on (i) actual tasks performed and (ii) the level of skill required for some routine

activities. In this exercise, I thus follow prior research and create more traditional task measures, relying on task information (i) only.¹⁵

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 10 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 11 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

¹⁵ I performed a few more related robustness tests such as removing individual tasks groups due to concerns over high measurement error or re-classifying certain tasks. These concerns may apply, for instance, to tasks associated with IT-Development, Programming and use of email/ internet, and calculation/ accounting. However, none of these robustness checks change my findings substantially. I therefore do not include them in this paper, but these robustness checks are available from the author upon request.

Table 11 Robustness tests on task measures as wage predictors: survey vs expert data

Dependent Variable: log hourly real wage	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Abstract (Ind.)	0.37*** (0.04)	0.34*** (0.03)	0.42*** (0.03)	0.68*** (0.06)	0.38*** (0.03)	0.62*** (0.05)	0.37*** (0.05)
Routine (Ind.)	0.19*** (0.04)	0.18*** (0.03)	0.22*** (0.03)	0.40*** (0.06)	0.16** (0.03)	0.44*** (0.05)	0.22*** (0.04)
Abstract (Exp., Occ.)	0.33*** (0.03)	0.35*** (0.03)	0.26*** (0.03)	0.29*** (0.03)	0.31*** (0.03)	0.27*** (0.03)	0.35*** (0.04)
Routine (Exp., Occ.)	0.16*** (0.03)	0.19*** (0.03)	0.14*** (0.03)	0.10*** (0.03)	0.15*** (0.03)	0.08** (0.03)	0.10** (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					✓		
Task Normalization a la Alda (2013)						✓	
2012 Sample Only							✓
Expert tasks (Occupational)	✓	✓	✓	✓	✓	✓	✓
Adj. R^2	0.20	0.20	0.19	0.20	0.20	0.22	0.19
Observations	26641	25468	27777	28026	27069	27777	13756

Robust standard errors in parentheses

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

This output provides robustness tests on the baseline estimates of the model described in Eq. (10) and displayed in Table 6. Recall that this model regresses log wages on individual- and occupation-level tasks. For brevity, I only focus on the model of interest comprising individual- and expert data on tasks, i.e. column (5) of Table 6. I use BIBB/BAuA data that has been collected in 2011–12 and 2017–18, and data from BERUFENET, covering the years 2011–13. For the BIBB/BAuA data, see Hall et al. (2020b) and Hall et al. (2020a), respectively. For the BERUFENET data, see Dengler et al. (2014)

those in baseline specifications. To quantify the omitted-variable bias associated with expert task data, consider, for instance, abstract tasks and the baseline estimate of $\beta = 0.46$ (from Table 6, column 3).¹⁶ Applying the omitted-variable bias formula from Eq. (6) to each robustness exercise implies an omitted-variable bias on the order of 26%–34%, well in the range of baseline estimates of 28%. Note that the lower bound, 26%, is inferred from column (2), restricting the sample to occupations with at least 100 observations, and the upper bound, 34%, is inferred from column (3), defining occupations at the 2-digit level.¹⁷ The largest impact on the omitted-variable bias is therefore attributed to the definition of occupations.

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 12) and (ii) a narrow task classification (Table 13). 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 14, 15, displaying incremental R^2 measures for each of the above robustness 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 14, column 1) and exclusion of competencies in task construction (Table 14, column 5)—reveal negligible differences between survey and expert data.¹⁸

¹⁶ 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.

¹⁷ The calculation for the lower bound goes as follows: $\lambda\delta + \beta = 0.36 \times 0.34 + 0.35 = 0.12 + 0.35 = 0.47$, where where all numbers are taken from Table 11, column (2). Hence, the bias is equal to 26% ($= 0.12/0.46$). 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 10, column (5), and Table 11, column (5). Hence, the bias is equal to 34% ($= 0.16/0.46$).

¹⁸ 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.

Table 12 Robustness tests on task measures as wage predictors: within-occupation task specialization

Dependent variable: log hourly real wage	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Abstract (Ind.)	0.34*** (0.03)	0.38*** (0.04)	0.41*** (0.03)	0.68*** (0.06)	0.32*** (0.03)	0.64*** (0.05)	0.40*** (0.05)
Routine (Ind.)	0.17*** (0.03)	0.19*** (0.04)	0.21*** (0.03)	0.38*** (0.07)	0.09** (0.03)	0.42*** (0.06)	0.22*** (0.05)
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					✓		
Task Normalization a la Alda (2013)						✓	
2012 Sample Only							✓
Occupation Dummies (Survey data)	✓	✓	✓	✓	✓	✓	✓
Adj. R^2	0.22	0.21	0.20	0.22	0.22	0.22	0.21
Observations	26,641	25,468	27,777	28,026	27,069	27,777	13,756

Robust standard errors in parentheses

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

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 6). I use BIBB/BAuA data that has been collected in 2011–12 and 2017–18, and data from BERUFENET, covering the years 2011–13. For the BIBB/BAuA data, see Hall et al. (2020b) and (Hall et al. (2020a), respectively. For the BERUFENET data, see Dengler et al. (2014)

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 14, column 3), (ii) assume workers perform activities underlying tasks “often” or “sometimes” (Table 14, column 4), or (iii) adopt a narrow classification of five (rather than three) tasks (Table 15).

Overall, these tests support the conjecture that the omitted-variable bias in task returns estimated with expert data is primarily attributed to missing individual heterogeneity.¹⁹ This hypothesis is reinforced in explicit robustness checks of task specialization within occupations. Table 16 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 26–34%, 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, even though both data providers—BIBB/BAuA regarding survey data and the German Employment Agency regarding expert data—go to great lengths to reduce measurement errors (e.g., coding errors), I cannot rule out that differences in the results are affected by ways of collecting and aggregating data. Moreover, the survey data I am using has its own set of

¹⁹ Note the number of observations differs among several specifications in Tab. 14 due to varying assumptions on sample selection in the data. In an alternative proceeding, I create a fixed sample by imposing minimum hourly wage of 5 EUR, a weekly minimum of 15 hours workers, and at least 100 observations per 3-digit occupation. These restrictions reduce the number of observations by 13%, down to 24,140. Unlike some of the robustness exercises in this section, however, this proceeding allow me to test the robustness checks on a common sample (as the remaining assumptions do not affect sample size). The overarching takeaway remains robust as individual-level variation in tasks comprises important and unique variation in wages on the order of 4.9 - 7.2%, in line with the baseline approach to my robustness tests. This comparison suggests my baseline results are not too sensitive on sample selection. The results of this alternative proceeding for the robustness tests are available from the author upon request.

Table 13 Task measures as wage predictors: narrow task definitions

Dependent variable: log hourly real wage	(1)	(2)
NRA (Ind.)	0.40*** (0.04)	0.42*** (0.04)
NRI (Ind.)	0.28*** (0.04)	0.31*** (0.04)
RC (Ind.)	0.28*** (0.04)	0.29*** (0.04)
RM (Ind.)	− 0.03 (0.04)	0.01 (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
Observations	27,777	27,777

Robust standard errors in parentheses

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

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). I use BIBB/BAuA data that has been collected in 2011–12 and 2017–18, and data from BERUFENET, covering the years 2011–13. For the BIBB/BAuA data, see Hall et al. (2020b) and (Hall et al. (2020a)), respectively. For the BERUFENET data, see Dengler et al. (2014)

unique limitations, especially pertaining to non-random sampling. By being non-representative of the German workforce nor of the occupational composition, the data naturally introduces measurement error.²⁰

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 et al. 1996). This erroneous self-assessment is especially common in the cognitive domain (Pallier et al. 2002) and

may thus induce workers to overstate the importance of abstract tasks in their job. The Dunning-Kruger Effect (Kruger and 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 (individual-level) wages, e.g., Yamaguchi (2012), Cortes (2016), Cavaglia and Etheridge (2020). 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 (Barth et al. 2016; Card et al. 2013; Dostie et al. 2020; Song et al. 2019). 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 individual heterogeneity as an important factor to understand rising occupation-level wage differences. In general, data availability makes a 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 individual heterogeneity (Becker et al. 2019; Cortes and 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

²⁰ 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.

Table 14 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)	(7)
Abstract (Ind.)	5.7%	5.9%	7.4%	6.2%	4.4%	5.8%	4.1%
Routine (Ind.)	1.8%	1.6%	2.2%	1.9%	0.2%	3.6%	1.9%
Abstract (Exp.)	5.6%	4.4%	2.9%	4.4%	3.2%	3.7%	5.7%
Routine (Exp.)	2.3%	1.6%	1.6%	1.5%	1.8%	1.0%	1.3%
Total (Occ.)	7.9%	6.0%	4.5%	6.0%	4.9%	4.7%	7.1%
Total (Ind.)	7.5%	7.4%	9.6%	8.0%	4.6%	9.4%	6.0%
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					✓		
Task normalization a la Alda (2013)						✓	
2012 Sample only							✓
Observations	26,641	25,468	27,777	28,026	27,069	27,777	13,756

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 8 for reference. I use BIBB/BAuA data that has been collected in 2011–12 and 2017–18, and data from BERUFENET, covering the years 2011–13. For the BIBB/BAuA data, see Hall et al. (2020b) and (Hall et al. (2020a), respectively. For the BERUFENET data, see Dengler et al. (2014)

Table 15 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	27,777	27,777

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). I use BIBB/BAuA data that has been collected in 2011–12 and 2017–18, and data from BERUFENET, covering the years 2011–13. For the BIBB/BAuA data, see Hall et al. (2020b) and (Hall et al. (2020a), respectively. For the BERUFENET data, see Dengler et al. (2014)

suspected in conventional expert data. I show that occupation-level returns on tasks estimated with expert data introduces omitted-variable bias ranging from 26–34%. Motivated by a theoretical framework, I argue this bias is largely attributed to the fact that conventional task data ignores individual 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 and Handel 2013; Cassidy 2017; Rohrbach-Schmidt 2019). This information is especially useful in applications with a focus on heterogeneity, e.g., research goals aimed at explaining rising heterogeneity within occupations (Atalay et al. 2018, 2020; Deming and Noray 2019; Hershbein and Kahn 2018; Modestino et al. 2019). More broadly speaking, survey data has a key advantage over conventional expert data as it allows the researcher to measure both, individual-level variation and, via aggregation by occupations, occupation-level variation.

Yet, opting for expert data may be justified in many instances. On the one hand, survey and expert data display broadly similar statistical properties in various specifications. On the other hand, researchers may require narrowly-defined occupations. Sample size limitations may force researchers to aggregate occupations at broader levels using survey data instead. To minimize measurement error when using expert task data, my

Table 16 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)	(7)
Abstract (Ind.)	13.8%	13.5%	14.3%	13.2%	10.5%	12.7%	11.7%
Routine (Ind.)	5.4%	4.5%	5.2%	4.8%	0.7%	8.4%	5.3%
Total (Ind.)	19.1%	18.0%	19.4%	18.0%	11.2%	21.1%	17.0%
Occupation dummies	✓	✓	✓	✓	✓	✓	✓
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					✓		
Task normalization a la Alda (2013)						✓	
2012 Sample only							✓
Observations	26641	25468	27777	28026	27069	27777	13756

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 8 for reference. I use BIBB/BAuA data that has been collected in 2011–12 and 2017–18, and data from BERUFENET, covering the years 2011–13. For the BIBB/BAuA data, see Hall et al. (2020b) and (Hall et al. (2020a), respectively. For the BERUFENET data, see Dengler et al. (2014)

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. This limitation of expert data may soon be alleviated, however, as Dengler et al. (2014) will supplement their already publicly available data (covering 2011–2013) with updated three-year intervals for 2016 and 2019 as part of a larger dataset (Christoph et al. 2020, p. 61).

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 et al., 2013). Data limitations complicate a detailed analysis on the tasks performed within firms. Notable exceptions are Friedrich (2021), who uses a German firm-level panel to study the link between firms’ task composition and training decisions, and Cortes and Salvatori (2019) who uses British firm-level data to study occupational specialization within firms. Further efforts in collection of this type of data and especially matching this data to the worker-level is a promising direction for future research to better understand the degree of task specialization in the modern workplace.

Abbreviations

DMP	Dengler, Matthes & Paulus (2014)
BIBB	Federal Institute for Vocational Education
BAuA	Federal Institute of Occupational Safety and Health
KIdB	BA classification of occupations
NR	Non-routine
LOM	Leave-out-mean
AIC	Akaike Information Criteria
BIC	Bayesian Information Criteria
FE	Fixed effects

Acknowledgements

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. I also thank two anonymous referees for great suggestions that have improved the paper significantly. 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.

Author contributions

The author read and approved the manuscript.

Funding

Not applicable.

Availability of data and materials

The BERUFENET data is freely available and can be downloaded from the FDZ at the IAB: <https://fdz.iab.de/187/section.aspx/Publikation/k141204302>. The BIBB/BAuA are available for scientific research via the BIBB Research Data Centre, see: <https://www.bibb.de/en/15182.php>. Access to the SUF is provided by GESIS for a small provision fee. All program codes necessary to replicate the study will be made available upon publication.

Declarations**Ethics approval and consent to participate**

Not applicable.

Consent for publication

The author consents that the text and any pictures published in the article will be freely available on the internet and may be seen by the general public. The pictures and text may also appear on other websites or in print, may be translated into other languages or used for commercial purposes.

Competing interests

Not applicable.

Received: 28 March 2022 Accepted: 7 January 2023

Published online: 13 February 2023

References

- Acemoglu, D., Autor, D.: Skills, Tasks and Technologies: Implications for Employment and Earnings. *Handbook of Labor Economics*, pp. 1043–1171. Elsevier, Cham (2011)
- Alda, H.: Tätigkeitsschwerpunkte und ihre Auswirkungen auf Erwerbstätige: Heft-Nr. 138, Bundesinstitut für Berufsbildung, Bonn (2013)
- Antonczyk, D., Fitzenberger, D., Leuschner, U.: Can a task-based approach explain the recent changes in the German wage structure? *Jahrbücher für Nationalökonomie und Statistik* **229**, 214–238 (2009)
- Atalay, E., Phongthientham, P., Sotelo, S., Tannenbaum, D.: New technologies and the labor market. *J. Monet. Econ.* **97**, 48–67 (2018)
- Atalay, E., Phongthientham, P., Sotelo, S., Tannenbaum, D.: The evolution of work in the United States. *Am. Econ. J. Appl. Econ.* **12**(2), 1–34 (2020)
- Autor, D.H.: The “task approach” to labor markets: an overview. *J. Labour Mark. Res.* **46**(3), 185–199 (2013)
- Autor, D.H., Handel, M.J.: Putting tasks to the test: human capital, job tasks, and wages. *J. Law Econ.* **31**(S1), S59–S96 (2013)
- Autor, D.H., Levy, F., Murnane, R.J.: The skill content of recent technological change: an empirical exploration. *Q. J. Econ.* **118**(4), 1279–1333 (2003)
- Barth, E., Bryson, A., Davis, J.C., Freeman, R.: It’s where you work: increases in the dispersion of earnings across establishments and individuals in the United States. *J. Law Econ.* **34**(S2), S67–S97 (2016)
- Becker, S. O., Egger, H., Koch, M., Muendler, M.-A.: ‘Tasks, occupations, and wage inequality in an open economy’ (2019)
- Boehm, M., von Gaudecker, H.-M., Schran, F.: Occupation growth, skill prices, and wage inequality (2021)
- Brenner, L.A., Koehler, D.J., Liberman, V., Tversky, A.: Overconfidence in probability and frequency judgments: a critical examination. *Organ. Behav. Hum. Decis. Process.* **65**(3), 212–219 (1996)
- Card, D., Heining, J., Kline, P.: Workplace heterogeneity and the rise of West German wage inequality*. *Q. J. Econ.* **128**(3), 967–1015 (2013)
- Cassidy, H.: Task variation within occupations. *Ind. Relat. J. Econ. Soc.* **56**(3), 393–410 (2017)
- Cavaglia, C., Etheridge, B.: Job polarization and the declining quality of knowledge workers: evidence from the UK and Germany. *Labour Econ.* **66**, 101884 (2020)
- Christoph, B., Matthes, B., Ebner, C.: Occupation-based measures—an overview and discussion. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* **72**(S1), 41–78 (2020)
- Cortes, G.M.: Where have the middle-wage workers gone? A study of polarization using panel data. *J. Law Econ.* **34**(1), 63–105 (2016)
- Cortes, G.M., Salvatori, A.: Delving into the demand side: changes in workplace specialization and job polarization. *Labour Econ.* **57**, 164–176 (2019)
- de la Rica, S., Gortazar, L., Lewandowski, P.: Job tasks and wages in developed countries: evidence from piacc. *Labour Econ.* **65**, 101845 (2020)
- Deming, D., Noray, K.: ‘STEM careers and the changing skill requirements of work’, Working Paper (2019)
- Dengler, K., Matthes, B.: The impacts of digital transformation on the labour market: substitution potentials of occupations in Germany. *Technol. Forecast. Soc. Chang.* **137**, 304–316 (2018)
- Dengler, K., Matthes, B., Paulus, W.: ‘Occupational tasks in the German labour market: An alternative measurement on the basis of an expert database’, FDZ Methodenreport 201412 (2014)
- Dostie, B., Li, J., Card, D., Parent, D.: Employer policies and the immigrant-native earnings gap. NBER Working Paper No. 27096 (2020)
- FRED: Consumer price index: all items for Germany (2022). <https://fred.stlouisfed.org/series/DEUCPIALLAINMEI>
- Friedrich, A.: Task composition and vocational education and training—a firm level perspective. *J. Vocat. Educ. Train.* 1–24 (2021)
- Goebel, J., Grabka, M.M., Liebig, S., Kroh, M., Richter, D., Schröder, C., Schupp, J.: The German socio-economic panel (SOEP). *Jahrbücher für Nationalökonomie und Statistik* **239**(2), 345–360 (2018)
- Gregory, C., Lewis, P., Frugoli, P., Nallin, A.: Updating the O*NET®-SOC taxonomy: incorporating the 2018 SOC structure: report (2019)
- Hall, A., Hünefeld, L., Rohrbach-Schmidt, D.: ‘BIBB/BAuA employment survey of the working population on qualification and working conditions in Germany 2018. SUF_1.0’, Research Data Center at BIBB (ed.); GESIS Cologne (data access); Bonn: Federal Institute for Vocational Education and Training (2020)
- Hall, A., Siefer, A., Tiemann, M.: ‘BIBB/BAuA employment survey of the working population on qualification and working conditions in Germany 2012. SUF_6.0’, Research Data Center at BIBB (ed.); GESIS Cologne (data access); Bonn: Federal Institute for Vocational Education and Training (2020)
- Handel, M.J.: The O*NET content model: strengths and limitations. *J. Labour Mark. Res.* **49**(2), 157–176 (2016)
- Hershbein, B., Kahn, L.B.: Do recessions accelerate routine-biased technological change? Evidence from vacancy postings. *Am. Econ. Rev.* **108**(7), 1737–1772 (2018)
- Janser, M.: The greening of jobs in Germany: first evidence from a text mining based index and employment register data, IAB-Discussion Paper (14) (2018)
- Kracke, N., Reichelt, M., Vicari, B.: Wage losses due to overqualification: the role of formal degrees and occupational skills. *Soc. Indic. Res.* **139**(3), 1085–1108 (2018)
- Kracke, N., Rodrigues, M.: A task-based indicator for labour market mismatch. *Soc. Indic. Res.* **149**(2), 399–421 (2020)
- Kruger, J., Dunning, D.: Unskilled and unaware of it: How difficulties in recognizing one’s own incompetence lead to inflated self-assessments. *J. Pers. Soc. Psychol.* **77**(6), 1121–1134 (1999)
- Liebig, S., Goebel, J., Schröder, C., Grabka, M., Richter, D., Schupp, J., Bartels, C., Fedorets, A., Franken, A., Jacobsen, J., Kara, S., Krause, P., Kröger, H., Metzger, M., Nebelin, J., Schacht, D., Schmelzer, P., Schmitt, C., Schnitzlein, D., Siegers, R., Wenzig, K., Zimmermann, S., Gericke, M., Griese, F., König, J., Liebau, E., Petrenz, M., Steinhauer, H. W., Deutsches Institut für Wirtschaftsforschung: Sozio-oekonomisches panel, daten der jahre 1984–2019 (soep-core, v36, eu edition) (2021)
- Matthes, B., Burkert, C., Biersack, W.: Berufssegmente: Eine empirisch fundierte neuabgrenzung vergleichbarer beruflicher einheiten, IAB Discussion Paper (35) (2008)
- Matthes, B., Christoph, B., Janik, F., Ruland, M.: Collecting information on job tasks—an instrument to measure tasks required at the workplace in a multi-topic survey. *J. Labour Mark. Res.* **47**(4), 273–297 (2014)
- Modestino, A.S., Shoag, D., Ballance, J.: Upskilling: do employers demand greater skill when workers are plentiful? *Rev. Econ. Stat.* **4**, 1–46 (2019)

- Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., Roberts, R.D.: The role of individual differences in the accuracy of confidence judgments. *J. Gen. Psychol.* **129**(3), 257–299 (2002)
- Peri, G., Sparber, C.: Task specialization, immigration, and wages. *Am. Econ. J. Appl. Econ.* **1**(3), 135–169 (2009)
- Reinhold, M., Thomsen, S.: The changing situation of labor market entrants in Germany. *J. Labour Mark. Res.* **50**(1), 161–174 (2017)
- Rohrbach-Schmidt, D.: Putting tasks to the test: the case of Germany. *Soc. Incl.* **7**(3), 122–135 (2019)
- Rohrbach-Schmidt, D., Hall, A.: BIBB/BAuA employment survey 2012, BIBB-FDZ Data and Methodological Reports No. 1/2013. Version 6.0. Bonn: BIBB. ISSN 2190-300X (2013)
- Rohrbach-Schmidt, D., Hall, A.: BIBB/BAuA employment survey 2018, BIBB-FDZ Data and Methodological Reports No 1/2020. Version 1.0. Bonn: BIBB ISSN 2190-300X (2018)
- Rohrbach-Schmidt, D., Tiemann, M.: Changes in workplace tasks in Germany—evaluating skill and task measures. *J. Labour Mark. Res.* **46**(3), 215–237 (2013)
- Roy, A.D.: Some thoughts on the distribution of earnings. *Oxf. Econ. Pap.* **3**(2), 135–146 (1951)
- Song, J., Price, D.J., Guvenen, F., Bloom, N., von Wachter, T.: Firming up inequality. *Q. J. Econ.* **134**(1), 1–50 (2019)
- Spitz-Oener, A.: Technical change, job tasks, and rising educational demands: looking outside the wage structure. *J. Law Econ.* **24**(2), 235–270 (2006)
- Stinebrickner, R., Stinebrickner, T., Sullivan, P.: Job tasks, time allocation, and wages. *J. Law Econ.* **37**(2), 399–433 (2019)
- Storm, E.: Task-based learning and skill (mis)matching: mimeo (2022)
- Storm, E.: Task specialization and the native-foreign wage gap. *Labour* **36**(2), 167–195 (2022)
- Yamaguchi, S.: Tasks and heterogeneous human capital. *J. Law Econ.* **30**(1), 1–53 (2012)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
