Bönke, Timm; Dany-Knedlik, Geraldine; Pagenhardt, Laura

Article

New DIW Berlin model can nowcast the current income distribution: Inequality likely to slightly increase in 2023

DIW Weekly Report

Provided in Cooperation with:
German Institute for Economic Research (DIW Berlin)

Suggested Citation: Bönke, Timm; Dany-Knedlik, Geraldine; Pagenhardt, Laura (2023) : New DIW Berlin model can nowcast the current income distribution: Inequality likely to slightly increase in 2023, DIW Weekly Report, ISSN 2568-7697, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin, Vol. 13, Iss. 43/44, pp. 283-289, https://doi.org/10.18723/diw_dwr:2023-43-1

This Version is available at:
https://hdl.handle.net/10419/280700

New DIW Berlin model can nowcast the current income distribution; inequality likely to slightly increase in 2023

By Timm Bönke, Geraldine Dany-Knedlik, and Laura Pagenhardt

ABSTRACT

The unequal distribution of labor income in Germany is a hotly debated topic among policymakers and the general public alike. However, the relevant data for calculating the distribution is usually available with a delay of sometimes over two years. Accordingly, previous studies have only been about the past, not the current, distribution. Generally, the current development of the income distribution as well as the effects of current economic policy measures remain oblique. That is where DIW Berlin comes in: A new forecast model is making it possible to make statements about the current state of inequality in the monthly gross labor income of employees. To do so, a number of macroeconomic labor market variables, such as the share of people in short-time work as well as gross wages and salaries in various economic sectors, are combined with household-level data from the Socio-Economic Panel (SOEP). The calculations in this Weekly Report show that the forecast quality of the model is high. Thus, for the first time, insights into a longer period following the coronavirus pandemic can be gained: Following declining inequality up until 2021, primarily between 2010 and 2020, the new DIW Berlin model forecasts an interruption in the trend in 2022 and 2023. Overall, labor income inequality is predicted to have increased slightly over the course of the economic recovery following the pandemic and will be somewhat higher in 2023.

The coronavirus pandemic has affected the development of both the labor market and household labor income over the past few years, with employees being affected to different degrees depending on their sector and qualifications. This has resulted in marked shifts in the labor income distribution. For example, the average monthly gross labor income of employees in the bottom 40 percent of the distribution increased by 4.9 percent in 2020 compared to 2019. The gross labor income of employees in the top ten percent, in contrast, only increased by 1.4 percent. During the coronavirus pandemic, the monthly gross labor incomes converged, mainly due to income-stabilizing measures such as the short-time work allowance.[1]

To date, there is a lack of individual and household data on how the economic normalization following the pandemic will be reflected in the gross income distribution of employees. The microdata from household surveys or administrative data sources, such as labor market data from the social insurance system, that are required for the calculation are usually available with a lag of one and a half years. The lag for more highly aggregated statistics from the Federal Employment Agency or the Structure of Earnings Survey is about one year. Promptly estimating the labor income distribution together with household information and further socio-economic characteristics is not only helpful in estimating the effects of current developments, such as the energy crisis and record inflation, but is also necessary for designing targeted economic policy measures, such as the minimum wage level, tax relief, or the inflation compensation premium.

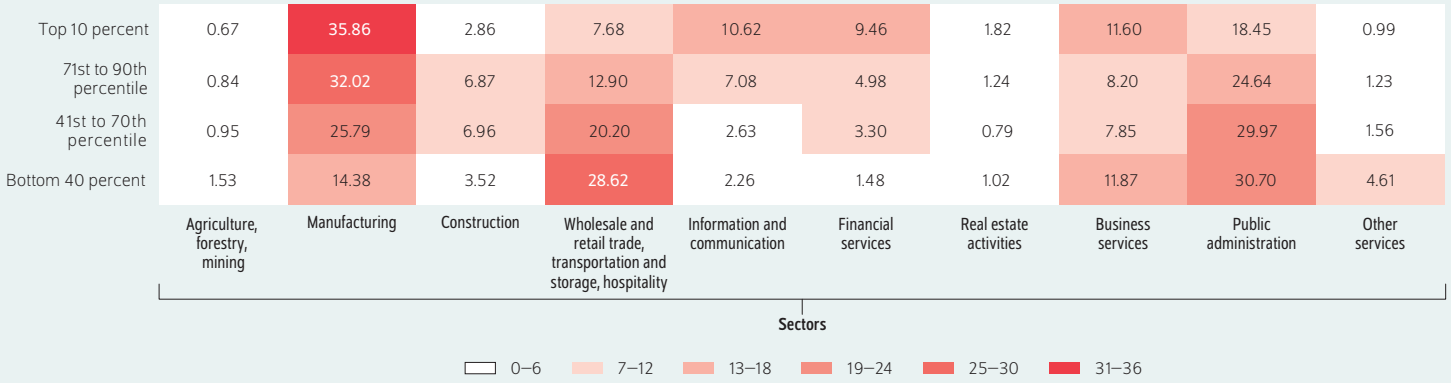
This Weekly Report presents the results of a new nowcast model on the distribution of employees' monthly gross labor income using quarterly and monthly macroeconomic indicators. As the structure of the labor market shifts slowly, it can be assumed that macroeconomic indicators are relevant in explaining the development of individual labor income. This makes it possible to predict the very recent past and the present, i.e., conduct a nowcast, and thus to make statements

1 Cf. Geraldine Dany-Knedlik and Alexander Kriwoluzky, "Income Inequality in Germany Temporarily Sinks During Crises," *DIW Weekly Report*, no. 46 (2011) (available online). Accessed on June 7, 2023. This applies to all other online sources in this report unless stated otherwise).

DOI: https://doi.org/10.18723/diw_dwr:2023-43-1

Figure 1

Distribution of income groups across economic sectors
Shares in percent



Sources: Socio-Economic Panel (SOEP); authors' calculations.

A large share of low-income earners work in wholesale and retail trade, hospitality, and public administration.

about the current state of inequality. As this nowcast is based on microdata, the predicted income development can be used for different analyses on the distribution of GDP, evaluating policy measures, or forecasting tax revenue.

The following sections first illustrate the microfounded data used in the model and present the microfounded labor income distribution that relates to the employment structure in Germany. Subsequently, the nowcast model and forecast quality are presented. Finally, the results of the nowcast and implications for the development of labor income and its distribution in 2021, 2022, and 2023 are shown and discussed.

Household income data is published with a considerable lag

Microeconomic data (microdata) are necessary for estimating the labor income distribution.[2] Potential sources of microdata on labor income in Germany are household surveys and administrative data, such as from the social security system.

Data from the Socio-Economic Panel (SOEP) is used to estimate and forecast the labor income distribution.[3] The SOEP is a representative survey of private households in Germany that has been conducted since 1984. Currently, around 30,000 individuals in around 15,000 households are surveyed each year on different topics, including their current gross monthly income, employment status, and professional situation. Using weighting schemes, this information can be extrapolated to the total population. However, the SOEP data is only available with a significant time lag. Data on the monthly gross labor income for a certain year is generally published in April or May of the following year, thus with a lag of around 16 months. For annual income, the lag is around 28 months. The most recent SOEP data (wave v37) used for this Weekly Report are for 2020. The present analysis is based on the current monthly gross labor income and is limited to employees and civil servants, including those in training and the marginally employed. The self-employed are not included in the analysis.

Sector data have special explanatory power

Several indicators are used to predict the gross labor income of employees from various parts of the distribution. Variables for Germany's GDP are selected using a recently published and carefully validated nowcasting model that contains classic macroeconomic indicators such as industrial output, business climate and consumer confidence indices, and key foreign trade figures.[4] This set of variables is supplemented by macroeconomic labor market indicators, including the number of people in short-term work, gross wages, and salaries. The sector-specific set of indicators serves as the core of the model: SOEP data on the distribution show how people in

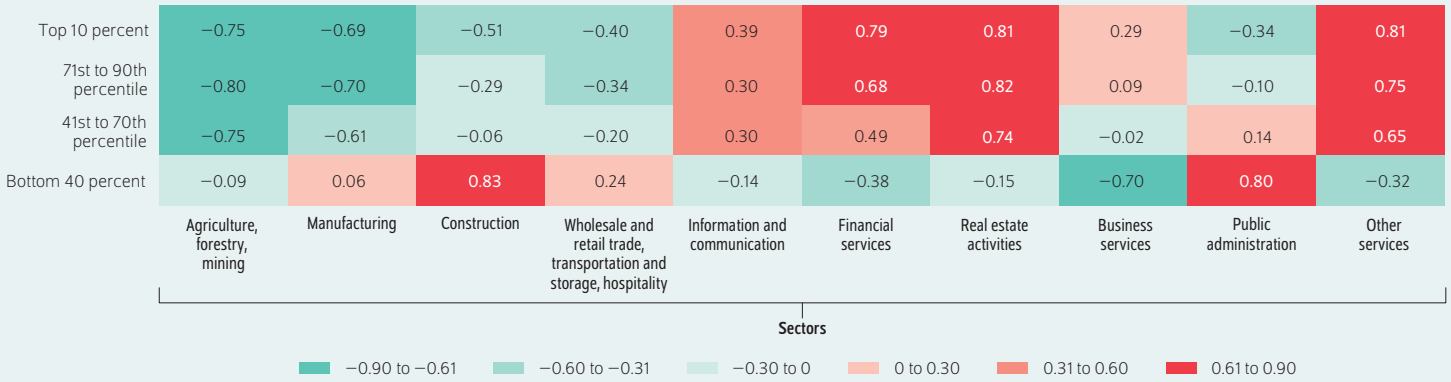
2 In addition, the distributions can be estimated using highly aggregated data, for example statistics tabulated by income group, such as in the Structure of Earnings Survey. This data is available quickly but is limited to certain employment types (for example, full-time employees subject to social security).

3 Alternative microdata such as the sample from the *Integrierten Arbeitsmarktbiografien* (SIAB) at the Institute for Employment Research (*Institut für Arbeitsmarkt- und Berufsforschung*, IAB) from the social security administrative process are available with the same time lag as SOEP data. While these data are computable at higher (for example, daily or monthly) frequencies, they do not provide information on the socioeconomic context, in particular the size of the household, which is necessary for further analysis, such as the evaluation of fiscal policy support measures.

4 Paolo Andreini et al., "Nowcasting German GDP: Foreign factors, financial markets, and model averaging," *International Journal of Forecasting*, no. 39, vol. 1 (2023): 298–313. The authors compare different established nowcasting methods and different sets of variables in order to establish the model with the best forecast quality and the lowest forecast error.

Figure 2

Correlation of the mean income of an income group with employment in an economic sector in the previous year[1]
Correlation of the variables on a scale of −1 to 1[2]



1 Price-adjusted and trend-adjusted data. Green fields imply a negative correlation, while red fields imply a positive one.

2 The closer the value is to 0, the smaller the correlation.

Example: Employment in construction in the previous year has a strong positive correlation with the income in the bottom income group. Thus, the higher the employment, the higher the average income in this income group. In the business services sector, there is a strong negative correlation for the bottom 40 percent of the distribution. Thus, the higher the employment, the lower the average income in this income group.

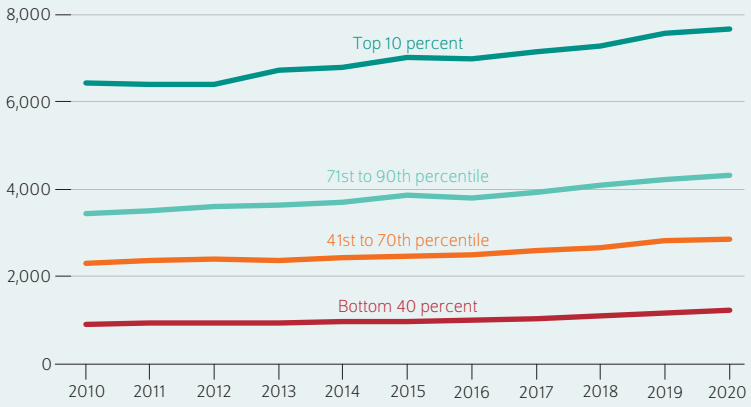
Sources: Socio-Economic Panel (SOEP); Federal Statistical Office; authors' calculations.

Rising employment in construction and in public administration is associated with rising incomes in the bottom income group.

various income groups are distributed across the different sectors of the German economy (Figure 1). For example, many of the low-income earners work in trade and hospitality as well as public administration, while the high-income earners work in the manufacturing sector and financial services. Therefore, it can be assumed that sector-specific labor market data, such as employment and hours worked per capita per sector, have relevant explanatory power for the different sections of the income distribution. Correlational data strengthens these assumptions: They show that employment in different sectors correlates with the income groups to differing degrees (Figure 2). For example, rising employment in construction and the public sector, including health and care services, is associated with rising income in the bottom 40 percent of the income distribution. Relatively low wages are often paid in these sectors, so an increase in employment there shifts the distribution downward and incomes at the upper end are lower on average. In contrast, the income in the higher income groups is more strongly correlated with employment in real estate as well as financial and other services. The comparatively high wages that new employees in these sectors are paid increase the incomes of high-income earners.

Figure 3

Mean nominal income by income groups
In euros per month[1]



1 Monthly income is only surveyed in the Socio-Economic Panel (SOEP) once per year.

Sources: Socio-Economic Panel (SOEP); authors' calculations.

Monthly incomes have increased across all income groups over the past years.

Macroeconomic indicators forecast middle incomes particularly well

To nowcast the income distribution, it is first divided into four sections: the bottom 40 percent, the 41st to 70th percentile, the 71st to 90th percentile, and the top ten percent (Figure 3). This simplification allows for a very good approximation of the distribution, using all available data as well as the most parsimonious model.[5] The four sections reflect the characteristics of the labor income distribution (Box 1). The bottom 40 percent is characterized by the marginally

5 The purpose of parsimonious modeling is to simultaneously achieve the highest possible expressive and predictive power of the model using the fewest number of variables possible, thereby avoiding excessive computational load.

Box 1

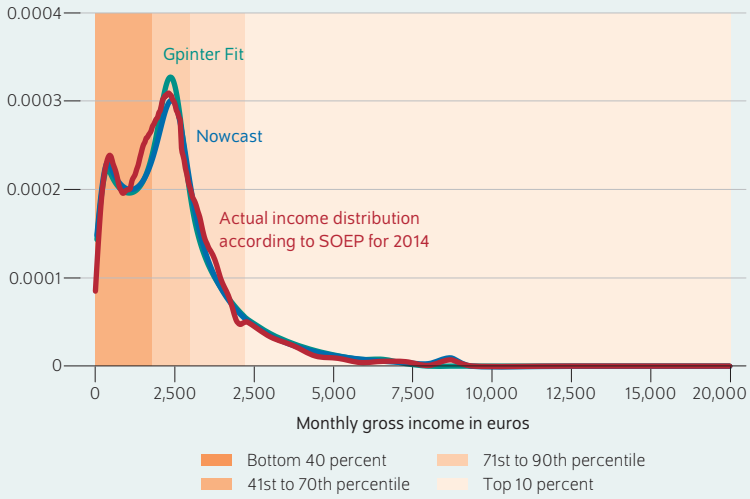
Reconstructing the income distribution

Income distributions, especially labor income distributions, frequently adhere to principles. These principles can be used to reconstruct an income distribution at the individual or household level with little information (for example, across the average income in deciles).

To make parsimonious modeling of nowcast and forecast models possible, the relevant information about the income distribution is captured using as few parameters as possible. To do so, the distribution is divided into four groups: The bottom 40 percent, the bottom middle (41st to 70th percentile), the upper middle (71st to 90th percentile), and the top 10 percent. Assuming a generalized Pareto distribution implemented with the Gpinter R-Package,[1] it is possible to reconstruct the course of the original income distribution (SOEP) using the income limits and average incomes. How well the reconstruction works can be shown by comparing the original distribution with the Gpinter Fit (Figure). The nowcast of the parameters of the income distribution also produces a result that does not differ statistically from the original distribution in the SOEP.

Figure

SOEP reconstruction of the income distribution
Density



Sources: Socio-Economic Panel (SOEP); authors' calculations.

The nowcast results in a distribution that barely differs from the actual distribution according to SOEP.

1 For more detailed documentation, see Thomas Blanchet, Juliette Fournier, and Thomas Piketty, "Generalized Pareto Curves: Theory and Applications," *Review of Income and Wealth*, no. 68, vol. 1 (2021): 263–288.

employed, part-time workers, and new workforce entrants, and is more susceptible to fluctuations. Full-time work in industry and comparably stable employment are concentrated in the middle of the distribution. The upper section of the distribution also develops a cyclical momentum that often sets it apart from the other income groups, in part due to more flexible pay arrangements.

A total of 52 indicators are inserted into a dynamic factor model (DFM) together with the respective target variable, i.e., the mean income of a given part of the distribution (Box 2). A pseudo real-time estimation is used to determine how well the model can predict the different incomes over time. The average deviation of the model's predictions from the average income of the respective groups is no more than 5.6 percent (Table 1). Comparing the forecast accuracy across quarters, it is noticeable that the forecasts from the second quarter are significantly better. This is primarily due to data availability: In the first quarter of a given year, for example 2020, SOEP data are available for up to three years before, i.e. 2017. In the second quarter, the data become available for an additional year, i.e. 2018, in this example, thereby significantly reducing the forecast error. Only data up until 2020 are available for predicting the 2023 distribution at the time this Weekly Report was written.

Comparing the forecast quality across the respective income groups, it is noticeable that the middle incomes between the 41st and 90th percentile are predicted most accurately. On average, the model overestimates or underestimates the nominal gross wages and salaries in the middle incomes by only a good two percent. The forecast is somewhat less exact at the margins of the income distribution: The estimated incomes of the low and top earners differ from the actual values by around four percent on average.

Inequality has slightly increased since the coronavirus pandemic

It is clear that the pandemic-related changes to labor income for workers vary by industry and skill level. For example, the development of the distributive data over the past years shows that nominal gross wages and salaries have increased overall since 2010, especially for the bottom 40 percent of the distribution (Table 2). Here, the nominal gross labor income increased by nearly 40 percent from 2010 to 2020, while the middle-income groups increased by 25 percent and the top earners by 19 percent. In 2020 alone, growth of the gross wages of workers in the bottom 40 percent was nearly five percent. In contrast, the average salary of the two middle groups as well as of the top ten percent of the distribution was only around two percent.

Layoffs of some employees, especially marginal ones, explain why average salaries increased in 2020: These employees and their low incomes were not included in the 2020 distribution. If predominantly workers in the lower half of the distribution are being let go, then the average income increases overall, as well as across the four income groups if the salaries

Box 2

Methods of income forecasting

A dynamic factor model (DFM) is used to nowcast the mean income, which is used to calculate the distribution.[1] In this process, many indicators are combined into a few common components (factors):

$$y_t = \Lambda f_t + \epsilon_t$$

y_t is an n-dimensional vector made of indicators, with n describing the number of indicators. f_t is the factor vector and the matrix Λ contains the factor loading; it describes how the indicators relate to the factors. ϵ_t describes the idiosyncratic components.

The number of factors is selected such that the model remains parsimonious on the one hand, while the factors explain as much variance of the dataset as possible on the other hand. In the present report, a model with two factors is estimated. The factors themselves follow a vector-autoregressive process of order 1, which means that a factor value can be influenced by the last prior values of all factors, including its own.

$$f_t = Af_{t-1} + u_t$$

The matrix A contains the autoregressive coefficients, the error u_t is normally distributed.

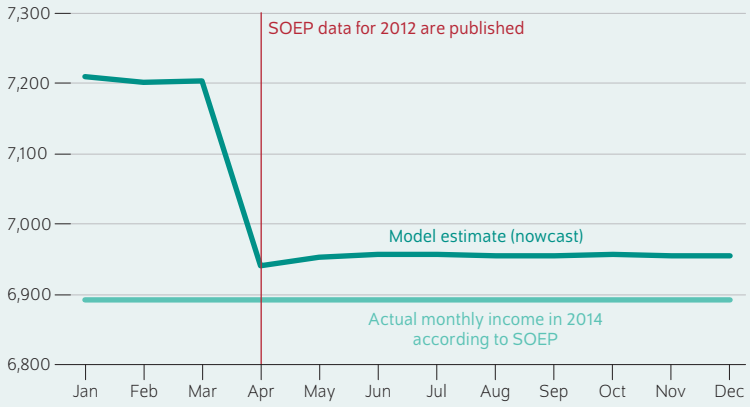
To be able to estimate the model, all contained variables must be stationary, meaning free from trends. Depending on the characteristics of the respective variables, different methods, such as first differences and growth rates, are used. For example, the average incomes are included in growth rates in the model. The DFM is estimated using the expectation-maximization approach. Thanks to the use of a Kalman filter, this approach allows for gaps in the series used and so-called "ragged edges," time series that begin and end at different times. In this way, it is also possible to use time series of different frequencies (like the quarterly and annual series used here) in the same model.

To determine the forecast accuracy of the model, pseudo-real time data is used. A dataset that simulates the data situation at the end of each respective month is created for every month since January 2010 and a nowcast for the respective year is estimated. Comparing the nowcast with the realized income makes it possible to determine the model's forecast quality over time. For example, the dataset in December 2014 included data on industrial output until October and GDP until the third quarter of 2014. This dataset is used to estimate a nowcast for average incomes in 2014, which can then be compared to the realized incomes in 2014, which are now known. The average forecast quality of the model improves the more data is available. This is particularly evident for estimates in the second quarter, when new SOEP data are usually published (Figure).

Figure

Development of the forecast quality over the course of an example year (2014)

Monthly income of the top 10 percent of the distribution in euros



Source: Socio-Economic Panel (SOEP); authors' calculations.

With the publication of new SOEP data, even for the previous year, the estimate for the current income (nowcast) improves significantly.

1 Following the model of Marta Bańbura and Michele Modugno, "Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data," *Journal of Applied Econometrics*, no. 29, vol. 1 (2014): 133–160.

of the remaining employees stay the same. As the groups are defined relative to each other, the poorest individuals in each percentile descend to the next lowest percentile. The larger increase in the average gross salary of the bottom 40 percent, especially in 2020, is likely due to the increase in short-time work by people who were in higher income groups before the pandemic in addition to the decrease in marginal employment.

Overall, income inequality measured via the monthly real gross income of employees from 2010 to 2020 decreased significantly. The inequality measures calculated from the distribution (Figure 4) also show this: For example, the Gini coefficient declined from 0.39 to 0.37 between 2010 and 2020 despite a temporary increase in 2013. The ratio between the top and low earners (P90/10) as well as between the middle and bottom end of the distribution (P50/10) has become smaller. Inequality lessened significantly once again, especially in 2020.

While inequality is likely to have continued its decline throughout 2021, the model predicts an interruption to this downward trend in 2022 and 2023, as real incomes in this time are likely to have increased, especially at the upper end of the distribution. Overall, inequality during the

Table 1

Forecast quality of the model[1] according to income groups and quarters
Mean absolute error in percent

	Bottom 40 percent	41st to 70th percentile	71st to 90th percentile	Top 10 percent
Total	4.21	2.31	2.10	3.86
1st quarter	4.77	3.36	2.62	5.56
2nd quarter	4.08	1.98	1.95	3.35
3rd quarter	3.99	1.93	1.90	3.21
4th quarter	4.01	1.99	1.93	3.33

1 Using a pseudo-real time estimation, it was determined how well the model can predict the different incomes over time. To do so, the average deviation of the forecasts of the nowcast models from the actual average incomes of the respective groups was determined.

Source: Authors' calculations.

Table 2

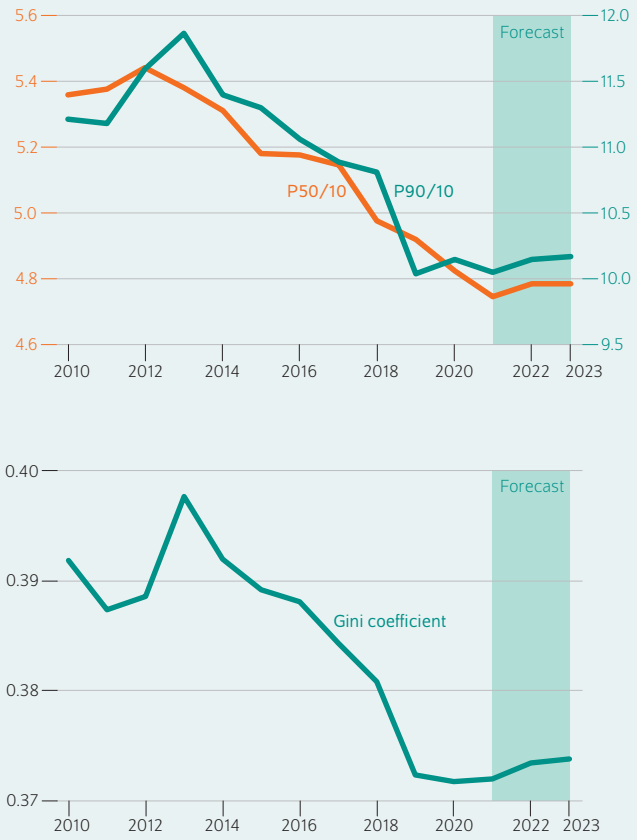
Development and forecast of nominal gross wages and salaries by income group

												Forecast		
	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Nominal income (in euros)														
Bottom 40 percent	891	944	951	943	956	986	1,016	1,038	1,099	1,179	1,236	1,325	1,404	1,440
41st to 70th percentile	2,290	2,359	2,389	2,362	2,420	2,463	2,509	2,586	2,654	2,822	2,860	3,021	3,225	3,309
71st to 90th percentile	3,438	3,513	3,598	3,629	3,705	3,853	3,812	3,944	4,104	4,238	4,314	4,573	4,888	5,018
Top 10 percent	6,436	6,409	6,406	6,720	6,795	7,038	6,987	7,150	7,298	7,572	7,678	8,240	8,804	9,051
Change in percent														
Bottom 40 percent		5.97	0.73	−0.87	1.43	3.09	3.03	2.25	5.80	7.27	4.87	7.24	5.96	2.55
41st to 70th percentile		3.00	1.26	−1.10	2.44	1.77	1.88	3.06	2.62	6.35	1.34	5.63	6.76	2.59
71st to 90th percentile		2.17	2.43	0.84	2.10	4.00	−1.06	3.45	4.05	3.28	1.80	5.98	6.91	2.64
Top 10 percent		−0.41	−0.05	4.90	1.12	3.57	−0.73	2.34	2.07	3.76	1.40	7.32	6.85	2.81

Sources: Socio-Economic Panel (SOEP); authors' calculations.

Figure 4

Measure of income inequality
Percentile ratio[1] and Gini coefficient



1 The P90/P10 ratio compares a person from the top end of the income distribution (only ten percent of people have a high income) with a person from the bottom end of the distribution (only ten percent of people have a low income). Accordingly, the P50/P10 ratio compares a person from the middle of the income distribution with a person from the bottom section of the distribution.

2 The Gini coefficient is between 0 and 1. The higher the value, the greater the inequality.

Sources: Socio-Economic Panel (SOEP); authors' calculations.

Income inequality decreased significantly at the beginning of the coronavirus pandemic.

post-pandemic recovery has probably remained at a constant level or even increased again slightly compared to 2020. This is not unusual during an economic upswing, from which higher earners generally benefit more than low earners, whereas the opposite effect is usually observed in times of crisis.[6] According to the calculations, high inflation also led to a decline in real terms among low-income earners in 2022, while top earners are still likely to have benefited from slight increases in real income.

Conclusion: Nowcast of inequality can improve the targeted nature of policy measures

A nowcast of the changes in the distribution of gross wages and salaries of workers is important in order to design targeted economic policy measures such as the minimum wage or the inflation compensation premium. However, because important labor income data are only available with a delay of at least 16 months, they can only evaluate past, not current, economic policy decisions. This Weekly Report shows that macroeconomic indicators are suitable for and can be used to predict the distribution of gross wages quickly. For example, it is possible to evaluate the current effects of the inflation compensation bonus, to identify vulnerable households, and to provide targeted support if needed.

The nowcasts indicate a slight increase in labor income inequality in the years following the pandemic. This increase, beginning in 2021, is likely to intensify the impact of the energy crisis, which is already primarily affecting poorer households.

6 Cf. Fatih Guvenen, Serdar Ozkan, and Jae Song, "The Nature of Countercyclical Income Risk," *Journal of Political Economy* (2014): 621–660.

The nowcast of current labor income inequality can be an important component of evaluating current economic policy measures as well as the design of future measures. In this context, it is essential to pay attention to the forecast quality, for example by using upper and lower forecast bounds, in order to consider the uncertainty of a point-by-point forecast when interpreting the forecast distribution and the corresponding inequality. Moreover, the nowcast of the distribution can also be used to simulate distributive effects of economic development or to better estimate tax revenue and transfer payments. DIW Berlin plans to extrapolate the labor income distribution into the future, conditional on the business cycle forecast, thus improving DIW Berlin's economic forecasts regarding the labor market in Germany and fiscal policy. Moreover, it is planned to add the implications of the labor income distribution to the national account systems as a sustainability indicator.

Timm Bönke is Co-Head of the Forecasting and Economic Policy Division in the Macroeconomics Department at DIW Berlin | tboenke@diw.de

Geraldine Dany-Knedlik is Co-Head of the Forecasting and Economic Policy Division in the Macroeconomics Department at DIW Berlin | gdanyknedlik@diw.de

Laura Pagenhardt is a Research Associate in the Macroeconomics Department at DIW Berlin | lpagenhardt@diw.de

JEL: E17, D31, E66

Keywords: Nowcast, income distribution, inequality

LEGAL AND EDITORIAL DETAILS

DIW Berlin — Deutsches Institut für Wirtschaftsforschung e. V.
Mohrenstraße 58, 10117 Berlin
www.diw.de
Phone: +49 30 897 89–0 Fax: –200
Volume 13 November 1, 2023

Publishers
Prof. Dr. Tomaso Duso; Sabine Fiedler; Prof. Marcel Fratzscher, Ph.D.; Prof. Dr. Peter Haan; Prof. Dr. Claudia Kemfert; Prof. Dr. Alexander S. Kritikos; Prof. Dr. Alexander Kriwoluzky; Prof. Dr. Lukas Menkhoff; Prof. Karsten Neuhoff, Ph.D.; Prof. Dr. Carsten Schröder; Prof. Dr. Katharina Wrohlich

Editors-in-chief
Prof. Dr. Pio Baake; Claudia Cohnen-Beck; Sebastian Kollmann; Kristina van Deuverden

Reviewer
Dr. Charlotte Bartels

Editorial staff
Rebecca Buhner; Dr. Hella Engerer; Ulrike Fokken; Petra Jasper; Sandra Tubik

Layout
Roman Wilhelm, Stefanie Reeg, Eva Kretschmer, DIW Berlin

Cover design
© imageBROKER / Steffen Diemer

Composition
Satz-Rechen-Zentrum Hartmann + Heenemann GmbH & Co. KG, Berlin

ISSN 2568-7697

Subscribe to our DIW and/or Weekly Report Newsletter at
www.diw.de/newsletter_en