▶ Technical annex: World Employment and Social Outlook: September 2024 Update

Estimates of the labour income share

The model estimates a complete panel dataset of the labour income share. To this end, national accounts data from the United Nations Statistics Division (UNSD) and labour income data from the ILO Harmonized Microdata collection are combined. When national accounts data or microdata are not available, the estimates rely on a regression analysis to impute the missing data. The imputation is based on countries that are similar in terms of key economic and labour market variables.

The methodology involves two steps. The first step is to compute the labour income share, adjusted for the labour income of the self-employed, which has been recognized in the economic literature as a crucial element for international comparability. To achieve this,[1] detailed data on status in employment are used to further categorize self-employment into three groups: own-account workers, contributing family workers, and employers. Furthermore, the labour income of each self-employed group relative to employees is estimated based on a regression analysis of the microdata. The resulting estimate corresponds to the share of total income that accrues to labour:

$$\text{Labour income share} = \frac{\text{Labour income}}{\text{Gross domestic product}}$$

The second step imputes a labour income share for each country and year where a computation based on microdata is not possible. The estimates for 2004-2019 are obtained using regressions involving country fixed effects (where the labour income estimates together with microdata from step 1 are available) or region fixed effects (where the labour income estimates together with microdata from step 1 are not available), along with relevant covariates. A separate cross-validation approach is used to select the model that minimizes prediction error in the year 2020 and then again for the year 2021. For the year 2022, model estimates are calculated using the same approach (using the same models) as for the years up to and including 2019. Finally, for 2023 and 2024, since the current UNSD data is only available up to 2022, macroeconomic data, wages,[2] and other labour indicators, along with country fixed effects are used to estimate the values. Additionally, for a group of countries, OECD data on the unadjusted labour income share up to 2024Q1 is available, which is used as model input.

Disaggregation of labour income by gender

From the labour income share estimation procedure, imputed labour income for all workers in the sample is estimated at the micro (individual record) level. Having micro-level imputed labour income enables the production of estimates of labour income of women and men separately, by aggregating all individual records by sex.

[1] See ILO 2019 for a complete description of the imputation methodology.

[2] From the forthcoming ILO Global Wage Report.

When microdata is not available, the estimates rely on regression models to impute the necessary data. The estimates for 2004-2019 are obtained using models with country fixed effects (where at least one microdata-based value is available) or region fixed effects (where no observations are available) along with relevant covariates. A separate cross-validation approach is used to select the model that minimizes prediction error in the year 2020 and then again for the year 2021. For the year 2022, model estimates are computed using the same methodology (using the same models) as those applied for the years up to and including 2019. Finally, for 2023 and 2024, macroeconomic data, wages,[3] and other labour indicators along with country fixed effects are used to estimate the values.

Technology and the labour income share

Empirical strategy

To analyse the relationship between technological innovations and the labour income share, a SVAR (Structural Vector Auto Regression) model is used. This class of models enables the characterization of the joint dynamics of interrelated economic variables. With a set of restrictions, also known as identification schemes, underlying structural economic shocks can be estimated from estimated Vector Auto Regression models. Following a method pioneered by Galí 1999, long-run restrictions are used to identify technology shocks. The key assumption behind this method is that only technological changes will have a long-run impact on productivity. This identification scheme has been widely used in to estimate technology shocks.[4]

The data used ranges from 2003 to 2019,[5] with a sample of 36 countries[6] for which data are available in every single year. The variables of interest are GDP in constant PPP dollars, the unadjusted labour income share (compensation of employees only), employment, and hours worked. The data is sourced from the World Bank (GDP), UNSD (labour income share) and ILOSTAT (employment and hours worked). The unadjusted labour income share is computed by dividing the compensation of employees by the Gross Domestic Product based on the income approach.[7] Data on the adjusted labour income share (which accounts for the labour income of the self-employed) from the ILO modelled estimates is not used, as it requires substantial imputation, and the required data is not generally available for every year. Finally, labour productivity is calculated by dividing GDP by total hours worked.

[3] From the forthcoming ILO Global Wage Report.

[4] As is common across estimations of exogenous economic shocks, there are known limitations to both SVAR models in general (see for instance: Forni and Gambetti 2014), and also to the long-run restriction use for identifying technology shocks (see for instance: Erceg, Gust, Guerrieri 2005).

[5] For the labour income share, the time series used starts at 2004. To identify the technology shocks, the period from 2003 to 2019 is used.

[6] The countries included (predominantly high-income economies) are: Austria, Belgium, Bulgaria, Canada, Switzerland, Cyprus, Germany, Denmark, Spain, Estonia, Finland, France, United Kingdom, Greece, Croatia, Hungary, Ireland, Iceland, Italy, Republic of Korea, Lithuania, Luxembourg, Latvia, Republic of Moldova, Mexico, Mauritius, Netherlands, Norway, Poland, Portugal, Paraguay, Slovenia, Sweden, Turkey, United States, and South Africa.

[7] Gross domestic product measured following the income approach (sometimes referred to as gross domestic income) should, by definition, be identical to the gross domestic product measured by other approaches (production approach or the expenditure approach). In practice they can differ due to measurement error, however these discrepancies are typically small.

In a first step, technology shocks are estimated for each individual country, using a bivariate SVAR model with long-run restrictions of labour productivity and total hours worked (both in natural logarithms).[8] In a second step, local projections are used to produce the average responses of the variables of interest to the estimated shocks.[9] [10] These cumulative impulse response functions (cIRFs, plotted in Figure 3 and Figure 4) represent the response of the variable of interest to a technology shock. For the cIRFs, the size of the shock is assumed to be one standard deviation of the estimate. This is the common practice in the literature, and choosing another scale for the shock would only shift the effect on all variables of interest by the same degree.

As a robustness check, the exercise is repeated using quarterly data (2003Q1 to 2019Q4). ILOSTAT data is used for employment and hours worked, while OECD data is used for GDP (in constant national currency) and the unadjusted labour income share. In this case, the number of countries with the required data is reduced to 25. While the use of annual data is suitable for the long-run identification scheme, using quarterly data provides a useful robustness check. This is because the estimated technology shocks are less likely to contain dynamic system responses to economic shocks than if they were estimated with annual data. The results of the quarterly specification[11] are broadly similar. A technology shock is estimated to increase productivity and output on impact, and this effect persists four years (16 quarters) after the initial shock. Labour input measures decrease on impact and recover afterwards. Finally, the labour income share declines on impact and remains lower four years later.[12]

Discussion of the results

The main result of the empirical model is that, for the sample countries, the estimated technology shocks from 2003 to 2019 produce, on average, a decline in the labour income share that persists in the medium run (4 years).[13] However, this finding does not imply that technological progress in general, or the development of AI in particular, will necessarily have this effect.

It is useful to consider these findings in the light of the theoretical model developed by Acemoglu and Restrepo 2018. In that model, the effects of technological changes on the labour income share can be drastically different according to the type of innovation and the parametrization of the model. For instance, developments in automation will reduce the labour share, whereas the creation of new tasks (that require relatively more labour) will have the opposite effects. Finally, the sign of the

[8] We use differences of variables guided by unit root tests. Unit root tests showed different results across countries in a few instances. We nonetheless use the specification suggested by the majority of the results. We use a model of one lag, as this is the lag length most often selected by statistical selection criteria. Adding more lags to the model does not sizeably alter the results.

[9] The key advantage of the local projection method in this case is that we can derive valid confidence intervals for the panel. For more information see: Jordà 2023 and Jordà 2005.

[10] The local projections are estimated using the 2005-2019 sample, to ensure that the same data is being used for each variable.

[11] We use 4 lags, given the dispersion of statistical criteria for lag selection and for comparability with the benchmark specification.

[12] Noteworthy differences are that in the 2nd and 3rd year the labour income share decline is not significant and estimates for labour input measures after the initial negative impact are closer to zero and even mildly positive one and two years after impact.

[13] In fact, in both the quarterly and yearly specifications the decline of the labour income share persists for longer. Nonetheless the error increases as the horizon expands, as can be expected given that at longer horizons fewer observations are available to estimate the effects.

effects of factor-augmenting technological changes (of either capital or labour) on the labour income share will depend on the elasticity of substitution between capital and labour.

In this context, the results suggest that, for the period of observation, the average effect is consistent with that of automation.[14] This evidence is useful as a reference, given that the technological progress experienced during the 2003-2019 period has been strongly linked to the development of digital technologies. Nonetheless, since both the magnitude of the impact of AI on productivity and the type of innovations driving it are highly uncertain, the analysis does not aim to simulate its effects.

Finally, although this setup is considered well suited for studying the question at hand, further work would be beneficial. For instance, testing for the informational sufficiency of the SVAR and augmenting the information set if necessary (as suggested by Forni and Gambetti 2014) would strengthen the robustness of the findings. Similarly, adapting the empirical strategy to specifically target automation shocks (as in Bergholt et al., 2022), rather than the more general target of technology shocks, would provide an interesting basis for comparison. These exercises are beyond the scope of the current brief and are left for future work.

Estimates of the share of youth not in employment education or training

The target variable of the model is the share of youth (aged 15 to 24) not in employment, education, or training (NEET):

$$\text{NEET share} = \frac{\text{Youth not in employment, education or training}}{\text{Youth population}}$$

The estimation procedure uses cross-validation to select the regression models with the best pseudo out-of-sample performance. The NEET model estimates all demographic groups jointly, using the appropriate categorical variable as a control in the regression, because the groups are interdependent and data availability is roughly uniform across breakdowns. The model incorporates information on unemployment, labour force participation, and school enrolment rates into the regressions (used alongside other variables to reflect economic and demographic factors). There is one regression model for countries with at least one data point and a second model for countries with no available data.

[14] Nonetheless the empirical specification does not allow us to separately identify the effects according to the theoretical typology, hence we cannot rule out factor augmenting technological change as a driver coupled with the appropriate parametrisation.