

NOWCASTING THE DISTRIBUTION OF DISPOSABLE INCOME: A MICROSIMULATION FRAMEWORK ENHANCED BY MACHINE LEARNING

**OECD PAPERS ON WELL-BEING
AND INEQUALITIES**

See all papers in this series

WORKING PAPER No.41



Funded by
the European Union

OECD Papers on Well-being and Inequalities

[See the paper series](#)

The OECD Papers on Well-being and Inequalities Series – managed by the OECD Centre on Well-being, Inclusion, Sustainability and Equal Opportunity (WISE) – features working papers on the measurement agenda for well-being, inclusion, sustainability and equal opportunity as well as papers seeking to deepen the understanding of the drivers of these issues, the ways in which they interact and how they evolve. These papers are prepared by OECD staff, external experts or by outside consultants working on OECD projects.

This work is published under the responsibility of the Secretary-General of the OECD and does not necessarily reflect the official views of the OECD Member countries.

This document was produced with the financial assistance of the European Union. The views expressed herein can in no way be taken to reflect the official opinion of the European Union.

Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcomed and may be sent to wellbeing@oecd.org.

The papers are generally available only in their original language – English or French – with a summary in the other.

This document, as well as any data and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

The names of countries and territories and maps used in this joint publication follow the practice of the OECD.

The release of this working paper has been authorised by Romina Boarini, Director of the OECD Centre on Well-being, Inclusion, Sustainability and Equal Opportunity (WISE).

© OECD 2026



Attribution 4.0 International (CC BY 4.0)

This work is made available under the Creative Commons Attribution 4.0 International licence. By using this work, you accept to be bound by the terms of this licence (<https://creativecommons.org/licenses/by/4.0>).

Attribution – you must cite the work.

Translations – you must cite the original work, identify changes to the original and add the following text: *In the event of any discrepancy between the original work and the translation, only the text of original work should be considered valid.*

Adaptations – you must cite the original work and add the following text: *This is an adaptation of an original work by the OECD and the European Commission. The opinions expressed and arguments employed in this adaptation should not be reported as representing the official views of the OECD or of its Member countries or the European Commission.*

Third-party material – the licence does not apply to third-party material in the work. If using such material, you are responsible for obtaining permission from the third party and for any claims of infringement.

You must not use the OECD logo nor the European Commission logo, visual identity or cover image without express permission or suggest the OECD or the European Commission endorse your use of the work.

Any dispute arising under this licence shall be settled by arbitration in accordance with the Permanent Court of Arbitration (PCA) Arbitration Rules 2012. The seat of arbitration shall be Paris (France). The number of arbitrators shall be one.

Acknowledgements

This working paper was authored by Fabrice Murtin, Head of Research, Modelling and Advanced Analytics Unit, and Taketo Muroya, Policy Analyst in the same unit, at the OECD Centre on Well-being, Inclusion, Sustainability and Equal Opportunity (WISE Centre), and by Jules Linden, Research Scholar at the International Institute for Applied Systems Analysis (IIASA). The paper was prepared under the oversight of Romina Boarini, Director of the OECD WISE Centre.

The authors are grateful for comments and suggestions from colleagues at the European Commission, particularly the Joint Research Centre (JRC), as well as from colleagues at the OECD Directorate for Employment, Labour and Social Affairs (ELS), the OECD Centre for Tax Policy and Administration (CTP), and delegates to the OECD Working Party on Social Policy (WPSP). Anne-Lise Faron, at the OECD WISE Centre, prepared the paper for publication.

The OECD WISE Centre gratefully acknowledges the support of the European Commission. The project was funded by the European Union via the Technical Support Instrument, and implemented by the OECD, in cooperation with the European Commission.

Abstract

Timely and accurate distributional income data are essential for monitoring inequality, assessing the adequacy of tax-benefit systems and designing policy responses during periods of rapid economic change. However, income and inequality statistics derived from EU-SILC are typically published with a lag of 12 to 18 months, limiting their usefulness during fast-moving shocks such as the COVID-19 pandemic, the 2022 energy price shock, or a severe recession. This paper presents results from a new nowcasting framework designed to reduce this information gap for EU Member States. By combining machine-learning (ML) forecasts of market income with EUROMOD microsimulation under the latest available tax-benefit rules, the framework produces timely estimates and forecasts of income inequality.

Résumé

Des données actualisées et précises sur la répartition des revenus sont essentielles pour suivre les inégalités, évaluer l'adéquation des systèmes de prélèvements et de prestations sociales, et concevoir des réponses politiques adaptées en période de changements économiques rapides. Toutefois, les statistiques sur les revenus et les inégalités issues de l'EU-SILC sont généralement publiées avec un décalage de 12 à 18 mois, ce qui limite leur utilité lors de chocs évoluant rapidement, tels que la pandémie de COVID-19, le choc des prix de l'énergie de 2022 ou une grave récession. Ce document présente les résultats d'un nouveau cadre de prévision en temps réel (« nowcasting ») conçu pour réduire ce déficit d'information dans les États membres de l'Union européenne. En combinant des prévisions du revenu de marché fondées sur l'apprentissage automatique (machine learning, ML) avec des microsimulations réalisées au moyen d'EUROMOD selon les règles fiscales et de prestations sociales les plus récentes disponibles, ce cadre produit des estimations et des prévisions rapides des inégalités de revenus.

Table of contents

OECD Papers on Well-being and Inequalities	1
Acknowledgements	2
Abstract	3
Résumé	4
1 Introduction	7
2 Methodology	9
2.1. A two-stage approach	9
2.2. Data sources	9
2.3. Implementation framework	10
3 Results	11
3.1. Main results	11
3.2. Model extensions	14
4 Discussion	16
References	18
Annex A. Methodological note on the overall framework	20
A.1. System architecture	20
A.2. Data	21
A.3. Method	23
A.4. Limitations	26
Annex B. Methodological note on the labour market adjustment module	29
B.1. Overview	29
B.2. Data	29
B.3. Method	30
B.4. Output and diagnostics	32
B.5. Limitations	33
Notes	34

Tables

Table 1. Out-of-sample performance by estimation method and forecast horizon	13
Table 2. Out-of-sample performance of the Gini coefficient and poverty rate with and without LMA	15
Table A A.1. EUROMOD variables description	22
Table A A.2. OECD macroeconomic predictors	22
Table A A.3. Income concepts	23
Table A A.4. Diagnostic statistics	24
Table A A.5. Macroeconomic variables and predictors used in related studies	27

Figures

Figure 1. Out-of-sample and forward predictions of Gini coefficients	12
Figure 2. Mean Absolute Error by country (1-year-ahead predictions)	13
Figure A A.1. Average sMAPE of income level by income decile (1-year-ahead predictions)	25
Figure A A.2. Overview of the model pipeline	28
Figure A B.1. Unemployment-rate ratios across countries	32

1 Introduction

Up-to-date income data are essential for monitoring inequality and designing tax-benefit policy during volatile times. However, income and inequality statistics derived from household surveys such as the EU Statistics on Income and Living Conditions (EU-SILC) are typically available with a lag of 12-18 months. During periods of rapid economic change, this delay undermines the timeliness of distributional analysis and complicates the assessment of policy impacts. This working paper describes a nowcasting approach designed to close that gap by using more timely National Accounts data.

The proposed new empirical framework combines machine learning (ML) estimation of market income (the sum of labour and capital income) with EUROMOD¹ microsimulation to nowcast and forecast (up to four years ahead) household incomes and income inequality across EU member states. The rationale for combining ML with microsimulation is that forecasting disposable income requires first forecasting market incomes via ML techniques, then accounting for the impact of taxes and benefits through microsimulation.

Beyond the nowcasting of disposable income inequality under fixed policy settings, this framework also makes it possible to simulate the impact of tax-benefit policy reforms on the disposable income distribution. Forecasting market income components per decile and then running the uprated microdata² through EUROMOD's tax-benefit model provides a tractable approach, while retaining the ability to simulate various tax-benefit policy changes available to adjust the distribution of market income.

The proposed methodology builds upon the literature on nowcasting income inequality, which can be organised into three broad strands:

- **Time-series econometric models** such as VAR or ARIMA are transparent and computationally inexpensive and generally find that labour force structure indicators are among the strongest short-run predictors (Gindelsky, 2018^[1]; Pierdzioch et al., 2022^[2]). However, country-level time series are often short because comparable inequality data are available only annually, constraining estimation precision. Research also consistently finds that naive benchmarks, including a simple autoregressive model or a random walk, are difficult to beat (Castle, Doornik and Hendry, 2024^[3]).
- **Microsimulation nowcasting** is the preferred approach of Eurostat (2020^[4]; 2023^[5]) and of several national statistical offices such as France, Sweden, Canada, UK and Australia (Levy, 2023^[6]). The EUROMOD nowcasting methodology broadly follows two main steps: i) modelling labour market changes, either by explicitly modelling labour market transitions (Navicke, Rastrigina and Sutherland, 2013^[7]) or by re-weighting observations using LFS employment rates by age, gender, and education; and ii) updating non-simulated incomes and simulating social benefits and taxes. Dynamic microsimulation approaches (Sologon et al., 2022^[8]; O'Donoghue and Sologon, 2023^[9]) calibrate microdata to more timely external data, estimating employment transition probabilities via probit models. These models are highly flexible but expensive to maintain and can struggle when the data-generating process changes.
- **Machine learning applied to National Accounts data** is the third way explored. Cornwall and Gindelsky (2025^[10]) use elastic net to nowcast the U.S. Gini and personal income shares, correctly predicting at least 90% of turning points (100% during COVID-19). Mahler, Castañeda Aguilar and Newhouse (2022^[11]) predict welfare distributions globally using a log-logistic parametrisation, finding that simple GDP-scaling nearly matches complex random forest models. Blanchet et al.

(2022^[12]) produce monthly U.S. distributional national accounts by rescaling income components to monthly National Accounts aggregates and incorporating high-frequency administrative data.

The framework presented in this working paper combines the second and third strands. First, it retains the distributional granularity and policy-simulation capability of microsimulations that route uprated income through EUROMOD, thereby capturing the automatic stabiliser effect of the tax-benefit system. Second, it capitalises on the timeliness and strong predictive performance of machine-learning forecasts trained on the *OECD Economic Outlook database* (OECD, 2025^[13]).

2 Methodology

2.1. A two-stage approach

In this setting, the technical parameter used to predict income is called an Up-rating Factor (henceforth URF). It is defined as future income y_{t+h} divided by current income y_t namely $URF_{t+h} = y_{t+h}/y_t$. The system works in two sequential stages, each addressing a distinct modelling challenge. In the first stage, the value of future URFs is learnt from past values thanks to machine learning techniques. This is equivalent to learning future income, which is equal to the URF times current income. In the second stage, URF parameters are injected into the EUROMOD model, which transforms market income (the sum of labour and capital income) into disposable income by simulating taxes and transfers. This stage is not necessary if the up-rated income in stage one is disposable income:

- **Stage 1: ML Forecasting of Up-rating Factors.** Macroeconomic indicators from the OECD Economic Outlook are used to forecast, for each EU country and forecast horizon ($h = 1, 2, 3$, or 4 years ahead): (i) a mean-income up-rating factor (URF) that shifts the overall level of the market income distribution; and (ii) a set of decile-specific URFs that reshape the distribution.³ To do this, six methods/algorithms are supported, namely Ordinary Least Squares (OLS), LASSO, Ridge, Elastic Net, Gradient Boosted Machines (GBM) and Random Forest (RF). As an alternative to (i) the ML-based mean-income URF, the framework can use EUROMOD's standard up-rating procedure, which applies source-specific indices to update non-simulated incomes.⁴
- **Stage 2: EUROMOD Microsimulation.** The URFs are applied to EUROMOD household microdata derived from EU-SILC, which cover EU member states from 2006 to 2024, to rescale income components within each decile. EUROMOD is then run to simulate the tax-benefit system's response (i.e. direct taxes and social security contributions paid, and benefits received at the up-rated income levels), yielding a full estimate of disposable income and the Gini coefficient.

The system supports three prediction modes, selected via a specification parameter that determines which income components are up-rated and whether EUROMOD is required. ML models can be applied to: i) the market income distribution; ii) to the labour and capital distributions separately; iii) to the disposable income distribution, in which case EUROMOD is no longer used since disposable income includes values of taxes and benefits. The third specification is computationally lighter and may be preferable when the processing speed is a priority and capturing the distributional impact of tax-benefit rules is less important. It also constitutes a relevant benchmark to assess the performance of more sophisticated systems.

2.2. Data sources

- **EUROMOD / EU-SILC microdata.** This representative household database, covering EU member states from 2006 to 2024, constitutes the micro foundation of the system. EUROMOD is first run under baseline conditions to compute disposable incomes. Households are ranked into equivalised disposable income deciles using demographic survey weights. Matching the EUROMOD input / EU-SILC data year and macroeconomic variables requires an adjustment, as these datasets

capture incomes from the preceding calendar year. Accordingly, in the baseline preparation step, the macroeconomic variables are aligned to the income reference year.

- **OECD Economic Outlook** (2025_[13]). The predictor set includes GDP growth, the unemployment rate, the wage share of GDP, government and private consumption shares of GDP, central bank and short-term interest rates, inflation rates, unit labour costs, external balances and the first lag of the income deciles' shares. The predictor set is augmented by a set of crisis dummies capturing the global financial crisis (2008-2010), the Euro area sovereign debt crisis (2011-2013), the COVID-19 shock (2020), and the energy price crisis (2022), as well as country fixed effects and regional interaction terms (Nordic, Central, Southern, and Eastern EU country clusters). Squared time trends are also included in the specifications.

2.3. Implementation framework

The framework is implemented through a five-step process (see Annex A for details):

- **Step 1: EUROMOD Data.** EUROMOD is first run for each country and data year via the Stata connector, applying the tax-benefit rules of the income reference year, in accordance with EUROMOD conventions. Individual-level outputs are aggregated to the household level, equivalised using the square root of household size,⁵ and ranked into deciles. Observed Gini coefficients are also calculated and used as targets for the nowcasting exercise.
- **Step 2: ML Upgrading Factor Estimation.** EUROMOD aggregates are merged with OECD macro indicators, and dependent variables (income shares, log shares, or levels by decile) and their lags are constructed. OLS models are estimated using the full predictor set, while regularisation or variable-selection methods are used to control model complexity in the ML algorithms (LASSO, Ridge, Elastic Net, GBM, and RF). Models are estimated recursively.
- **Step 3: Income Nowcasting via EUROMOD.** For each country and base year, EUROMOD microdata is loaded and individuals are assigned to deciles based on the base-year income distribution. URFs from Step 2 are applied in two stages: i) All income components are scaled either by the mean-income URF or by the standard upgrading factors embedded in EUROMOD to update nominal income levels; ii) then, within each decile, income components are multiplied by the decile-specific URF to reshape the distribution. For the market and labour/capital income distributions, the uprated dataset is passed back to EUROMOD, which recomputes tax liabilities and benefits using the tax-benefit rules of the forecast year, where available. The output is a full forecast of equivalised household disposable income, from which the Gini coefficient is computed.
- **Step 4: Diagnostics.** Forecast accuracy is assessed against observed Gini coefficients using two main metrics, namely Mean Absolute Error (MAE) and Directional Accuracy (DA).
- **Step 5: Visualisation and comparison.** Visualisation outputs include scatterplots of predicted Gini with confidence bounds versus observed Gini by horizon. The performance of all six estimation algorithms is compared across all horizons, providing a basis for algorithm selection.

Models are estimated and validated using a pseudo-out-of-sample design that approximates real-time information constraints.⁶ For each cutoff year t^* (income reference year), the model is estimated only on observations dated up to and including t^* . Predictions are produced for horizons from t^*+1 up to t^*+4 . The URF is constructed as the ratio of the ML-predicted income share at the forecast horizon to the observed base-period value, ensuring that no future-dated observations enter model estimation for earlier periods.

For ML algorithms, model-specific tuning and regularisation procedures are applied separately for each combination of forecast horizon and decile, where applicable. These procedures allow predictor selection, shrinkage, or model complexity to adapt across the distribution and as the estimation sample expands, which is particularly important when macro-distributional relationships change over time.

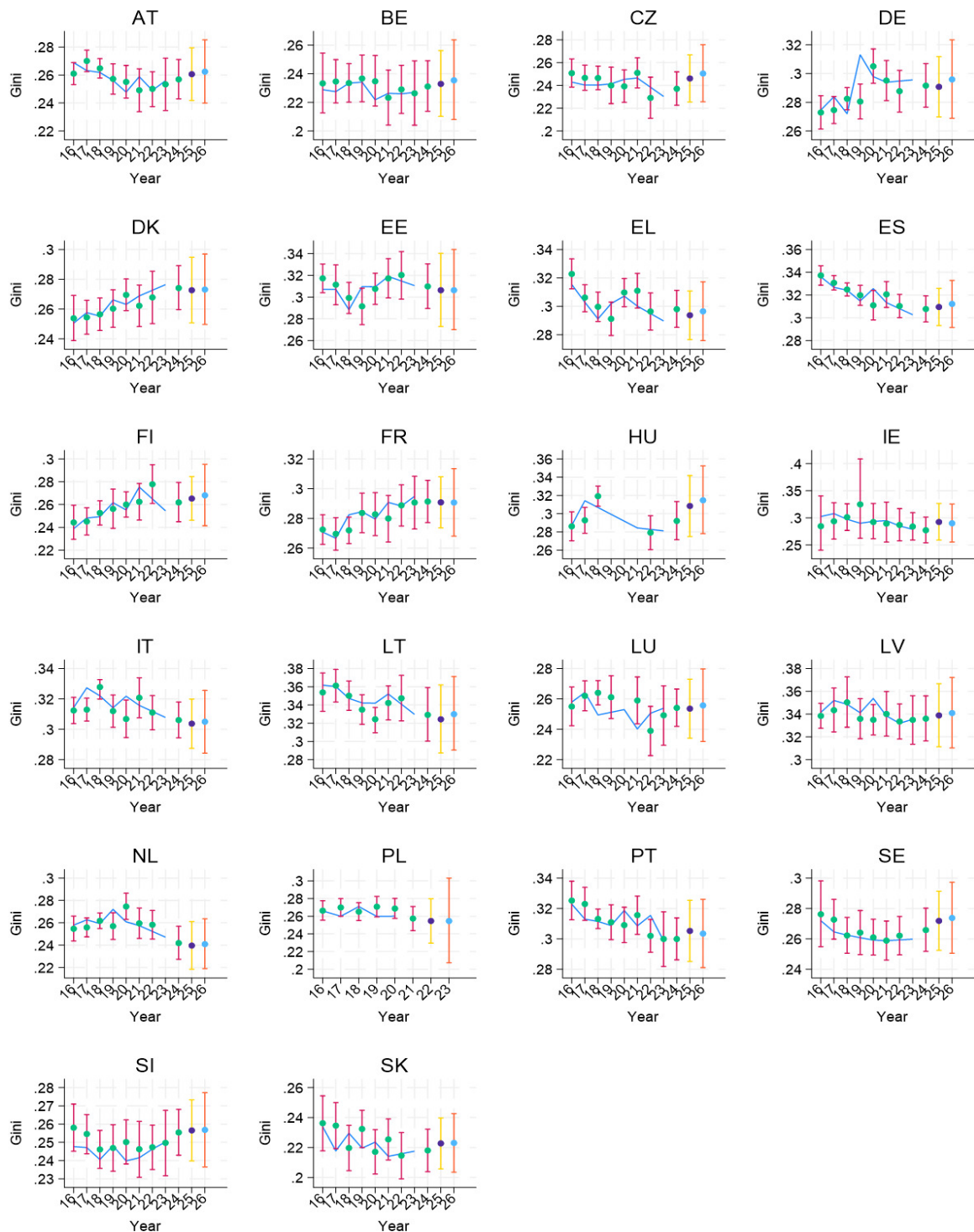
3 Results

3.1. Main results

In the main specification, the disposable income distribution is predicted by applying LASSO-based decile-specific uprating to market income (the sum of labour and capital income) and then re-running EUROMOD, after which Gini coefficients are calculated. The overall income level is uprated using EUROMOD's standard uprating, while LASSO-estimated decile-specific factors allow income growth to vary across the distribution. This specification is preferred because, when combined with ML-based decile-specific uprating, EUROMOD's standard uprating performs better than ML-based uprating for mean income.

Pseudo-out-of-sample predictions of Gini coefficients for 1-year-ahead horizons, evaluated against observed values before 2025, indicate that the nowcasting model performs satisfactorily. **The level of inequality is generally well captured across EU member states, with predicted Gini coefficients tracking observed trends closely and confidence bounds generally encompassing the observed values (Figure 1).** The model performs particularly well in capturing broad distributional trends and in identifying periods of rising or falling inequality associated with well-identified macroeconomic events such as the global financial crisis and the COVID-19 pandemic.

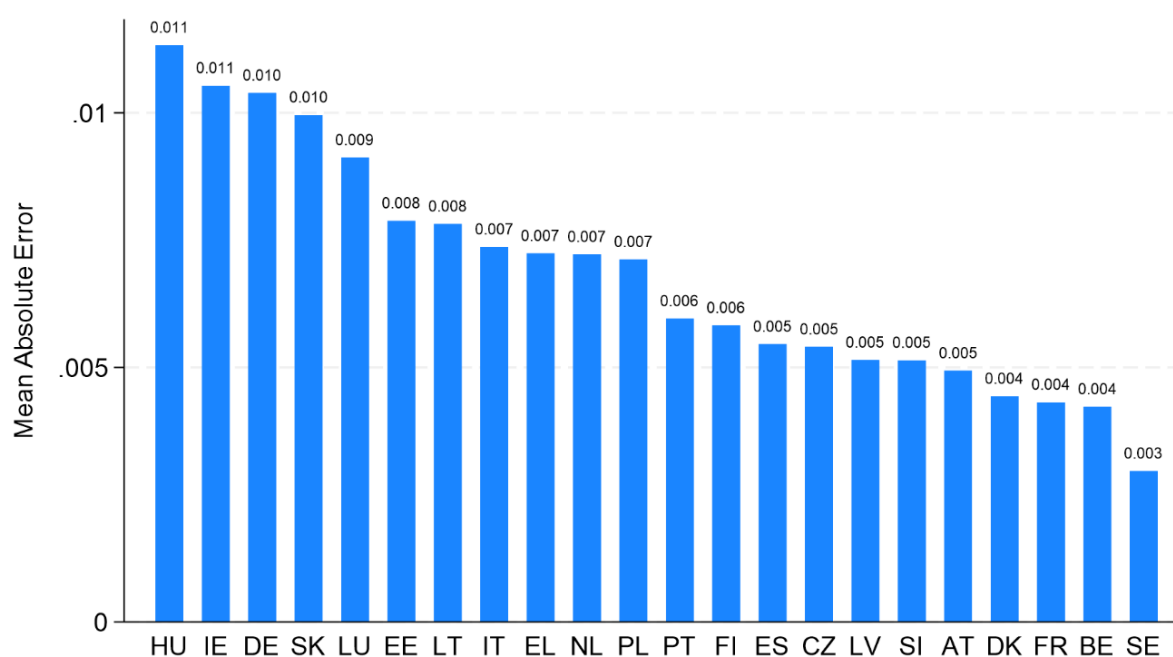
Figure 1. Out-of-sample and forward predictions of Gini coefficients



Note: The blue lines represent the observed Gini coefficients over the years. Each dot indicates a predicted Gini coefficient with its confidence bounds: 1-year-ahead predictions for years before 2025, 2-year-ahead predictions for 2025, and 3-year-ahead predictions for 2026. The LASSO uprating for market incomes and EUROMOD-simulated taxes and benefits are used to predict the disposable income distribution and then Gini coefficients are calculated. Mean-income URF is estimated using the standard EUROMOD uprating factors, while decile-specific URFs are estimated using the ML model. The Gini confidence bounds are derived from 95% confidence intervals (± 1.96 standard errors) around the post-LASSO decile predictions, which are combined to generate lower- and upper-bound income distributions.

However, country-level performance varies (Figure 2). Some member states, typically those with more stable macroeconomic trajectories and with longer and higher-quality EU-SILC series, exhibit very low prediction errors. Others show larger deviations, potentially reflecting: i) shorter or less reliable EU-SILC histories; ii) structural features of the labour market that are not well captured by OECD Economic Outlook predictors; or iii) compositional changes (e.g., large migration flows) that shift the income distribution in ways not reflected in the set of macroeconomic aggregates.

Figure 2. Mean Absolute Error by country (1-year-ahead predictions)



Note: MAE is calculated as the average of absolute differences between predicted and observed Gini coefficients. The results refer to 1-year-ahead predictions pooled over 2016-2024. The LASSO uprating for market income and EUROMOD-simulated taxes and benefits are used to predict the disposable income distribution and then Gini coefficients are calculated. Mean-income URF is estimated using the standard EUROMOD uprating factors, while decile-specific URFs are estimated using the ML model.

Table 1 summarises the performance of the main model configurations (OLS, LASSO, Elastic Net and GBM) across three forecast horizons ($h = 1, 2, 3$ years ahead), using Mean Absolute Error (MAE) and Directional Accuracy (DA) as the evaluation metrics. The LASSO model uprating market income combined with EUROMOD running delivers the best overall performance.

Table 1. Out-of-sample performance by estimation method and forecast horizon

Method	Nowcasting Approach	MAE h=1	MAE h=2	MAE h=3	DA h=1	DA h=2	DA h=3
OLS	Directly uprates disposable income	0.013	0.014	0.032	0.624	0.500	0.466
OLS	Uprates market income, re-runs EUROMOD	0.011	0.012	0.032	0.652	0.583	0.523
LASSO	Uprates market income, re-runs EUROMOD	0.007	0.007	0.008	0.660	0.659	0.725
Elastic Net	Uprates market income, re-runs EUROMOD	0.006	0.008	0.009	0.695	0.629	0.716
GBM	Uprates market income, re-runs EUROMOD	0.007	0.008	0.009	0.645	0.598	0.670

Note: MAE = Mean Absolute Error is the average absolute difference between predicted and observed Gini coefficients. Lower MAE values indicate greater predictive accuracy. DA = Directional Accuracy is the proportion of cases in which the model correctly predicts the direction of change in the Gini coefficient between the forecast origin and the target year. Higher DA values indicate greater directional accuracy. Mean-income URF is estimated using the standard EUROMOD uprating factors, while decile-specific URFs are estimated using the ML model.

The two key findings are as follows:

- **The hybrid ML-EUROMOD approach substantially reduces prediction error at short horizons.** Applying LASSO or Elastic Net to market income before re-running EUROMOD roughly halves the 1-year-ahead MAE relative to the OLS benchmarks (0.006-0.007 versus 0.011-0.013). At the two-year horizon, the gap remains substantial (0.007-0.008 versus 0.012-0.014). This improvement reflects the ability of regularisation to exclude noisy or irrelevant macroeconomic predictors that inflate OLS variance, particularly in the context of relatively short country panels. Similarly, directional accuracy (DA) at the one- and two-year horizons is highest for the LASSO- and Elastic Net-EUROMOD models, reaching around 63-70%. GBM also performs competitively in terms of MAE, with values of 0.007 and 0.008 at the one- and two-year horizons.
- **Performance degrades somewhat at longer horizons, but the ML advantage is maintained.** At the three-year horizon, MAEs are generally somewhat higher, which may reflect the weaker relationship between the base-year income distribution and outcomes several years ahead, together with the limited estimation sample. Notably, the MAEs of LASSO and Elastic Net at three years (0.008-0.009) remain well below those of the OLS benchmarks (0.032), suggesting that the ML approach substantially extends the effective forecast horizon. DA at the three-year horizon is highest for LASSO and Elastic Net, which may partly reflect that longer horizons provide greater scope for directional changes to emerge.

For additional comparison, a simple no-change benchmark, which carries forward the observed value at the forecast origin, yields MAEs of 0.007, 0.009, and 0.010 at the one-, two-, and three-year horizons, respectively. These MAEs are generally higher than those obtained from the models with LASSO- or Elastic Net-based decile uprating. When the model is run without the ML-based decile-specific adjustment, the corresponding MAEs are 0.006, 0.008, and 0.008, while the DAs are 0.660, 0.629, and 0.651, respectively. The generally higher DAs obtained with the decile-specific ML adjustment suggest that this component can improve the model's ability to capture the direction of changes in the Gini coefficient.

3.2. Model extensions

Labour Market Adjustment

The nowcasting framework described above relies on income uprating and tax-benefit simulation but does not explicitly account for changes in employment status. This limitation may be particularly important during periods of substantial labour-market change, as employment transitions can have significant effects on household incomes and income distribution. To address this issue, **this section incorporates the Labour Market Adjustment (LMA) module into the nowcasting framework** and evaluates its contribution to the prediction of distributional outcomes.

The LMA module adjusts the labour-market status of working-age individuals in EUROMOD input microdata to reflect projected employment transitions (see Annex B for a detailed description of the methodology). Transition probabilities are estimated using the Eurostat model (Leulescu et al., 2025^[14]) and calibrated to match aggregate employment-to-unemployment and unemployment-to-employment transition rates observed in the EU Labour Force Survey (EU-LFS). Labour-market transitions are then simulated stochastically, and labour incomes are updated accordingly, with income set to zero for job losers and imputed using predictive mean matching (PMM) for job finders. The adjusted microdata are subsequently passed through the income uprating and EUROMOD re-run procedures described above.

Poverty nowcasting

In addition to the Gini coefficient, **the at-risk-of-poverty rate (AROP) is introduced as a nowcasting target**. While the Gini coefficient captures changes in overall income inequality, AROP is more sensitive to income changes around the lower tail of the distribution, where employment transitions are likely to matter most. The poverty rate is measured according to the standard Eurostat AROP definition as the proportion of individuals living in households with equivalised disposable income below 60% of the national median equivalised disposable income.

Table 2 compares the pseudo-out-of-sample predictive performance of the nowcasting framework with and without the LMA module. For the Gini coefficient, directional accuracy (DA) improves at the one- and two-year horizons, while MAE remains unchanged at the one-year horizon and increases slightly at longer horizons. **For the poverty rate, LMA delivers more consistent improvements, particularly in DA**, which increases from 0.617 to 0.674, from 0.629 to 0.667, and from 0.633 to 0.679 at the one-, two-, and three-year horizons, respectively. MAE also improves slightly at the two- and three-year horizons. This pattern is consistent with the purpose of the LMA module: by explicitly modelling movements into and out of employment, it should capture income shocks that are especially relevant for households close to the poverty threshold. Overall, the results suggest that **labour-market adjustment adds more value for predicting poverty dynamics than for predicting changes in aggregate income inequality**.

Table 2. Out-of-sample performance of the Gini coefficient and poverty rate with and without LMA

Target	Prediction Approach	MAE h=1	MAE h=2	MAE h=3	DA h=1	DA h=2	DA h=3
Gini coefficient	Non-LMA & LASSO uprating & EUROMOD re-running	0.007	0.007	0.008	0.660	0.659	0.725
	LMA & LASSO uprating & EUROMOD re-running	0.007	0.008	0.009	0.688	0.674	0.679
Poverty rate	Non-LMA & LASSO uprating & EUROMOD re-running	0.009	0.011	0.012	0.617	0.629	0.633
	LMA & LASSO uprating & EUROMOD re-running	0.009	0.010	0.011	0.674	0.667	0.679

Note: MAE = Mean Absolute Error is the average absolute difference between predicted and observed target indicators. DA = Directional Accuracy is the proportion of cases in which the model correctly predicts the direction of change in the Gini coefficient or poverty rate between the forecast origin and the target year. Results are shown for the LASSO-based decile-specific uprating of market income with EUROMOD re-running, with and without the LMA module.

4 Discussion

The results suggest that combining ML nowcasting and forecasting of market income uprating factors with EUROMOD microsimulation offers a viable route to closing the 12-18-month distributional data gap that currently constrains real-time inequality monitoring in EU member states.

A first, key distinction of the framework is its capacity to **feed uprated market incomes through EUROMOD's tax-benefit model**, setting it apart from approaches based solely on statistical nowcasting. Importantly, the inequality nowcasts **reflect the automatic stabiliser function of the tax and benefit system**, namely the extent to which taxes and transfers cushion the distributional impact of macroeconomic shocks, which is often the policy dimension of greatest interest to governments and analysts during downturns.

The framework is also **policy flexible**. Users can modify EUROMOD's tax-benefit rules to simulate the distributional impact of prospective reforms under the forecasted macroeconomic environment, enabling prospective distributional analysis that would otherwise require waiting for the next survey wave. This opens up new analytical possibilities for ex-ante distributional impact assessments of budget measures, tax reforms or benefit changes.

The LASSO model's good performance also speaks to the value of systematic predictor selection. The regularisation process identifies the macro variables most predictive of distributional change at each horizon and for each decile, producing a **time-varying, decile-specific predictor set that adapts to changing macro-distributional relationships**. This may represent a methodological advance over approaches that apply a fixed, manually chosen set of uprating factors.

The framework has also been extended in two important directions. First, the **Labour Market Adjustment (LMA)** module has been incorporated to simulate transitions into and out of employment. This **addresses the previous limitation that employment status was fixed** in the nowcasting process. Second, **the at-risk-of-poverty rate has been added as an additional nowcasting target** alongside the Gini coefficient. This is important because poverty rates are more sensitive to income changes among lower-income households, where employment transitions are likely to have the effect. The results suggest that LMA adds more value for predicting poverty dynamics than for predicting aggregate income inequality.

The following limitations should be noted, as they define the boundaries of appropriate interpretation and use of the nowcasts:

- **Short estimation sample:** EUROMOD microdata currently covers 2006-2024, providing at most 18-19 annual observations per country. This constrains estimation precision, particularly for the ML algorithms that benefit from larger samples, and limits the statistical power of out-of-sample validation, especially at longer horizons.
- **Static decile assignment:** Decile membership is fixed at the base year. The model does not capture income mobility, namely transitions of households between deciles over the forecast horizon. In periods of large structural labour market change, this assumption may underestimate the true degree of distributional change.
- **Absence of behavioural responses:** The LMA module introduces calibrated employment transitions, but labour supply, savings and consumption responses to income or policy changes

are not modelled. The framework captures first-round tax-benefit effects through EUROMOD and LMA-based employment transitions, but not such broader behavioural responses.

- **Policy-rule availability:** The framework can use tax-benefit rules where they are available in EUROMOD, and users can modify these rules to simulate prospective reforms if they are already known or can be anticipated with reasonable confidence. However, future policy changes are not automatically incorporated unless they are already encoded in the relevant EUROMOD system or manually introduced by users. Uprating indices, used in some countries for a benefit indexation for example, are not automatically updated by the framework, but may be adjusted by users in EUROMOD in the *Uprating Indices* table.

Future research should seek to address these limitations and further develop the framework. Priority areas include exploring more granular macroeconomic predictors, such as sectoral and country-specific indicators; further improving the LMA module; and extending the approach to non-European OECD countries where suitable microdata and tax-benefit simulation tools are available.

References

- Blanchet, T. et al. (2022), “Real-time inequality”, *NBER Working Paper*, Vol. 30229, [12]
<https://doi.org/10.3386/w30229>.
- Castle, J., J. Doornik and D. Hendry (2024), “Forecasting the UK top 1% income share in a shifting world”, *Economica, London School of Economics and Political Science*, Vol. 91/363, pp. 1047-1074, <https://doi.org/10.1111/ecca.12533>. [3]
- Cornwall, G. and M. Gindelsky (2025), “Nowcasting Distributional National Accounts for the United States: A Machine Learning Approach”, *AEA Papers and Proceedings*, Vol. 115, pp. 79-84, <https://doi.org/10.1257/pandp.20251105>. [10]
- Eurostat (2023), *Flash estimates of income inequalities and poverty indicators for 2022 (FE 2022) — Methodological note*. [5]
- Eurostat (2020), *Flash estimates of income inequalities and poverty indicators for 2019 (FE 2019) — Methodological note*. [4]
- Gindelsky, M. (2018), “Modeling and forecasting inequality in the United States”, *Bureau of Economic Analysis Working Paper*, https://www.bea.gov/sites/default/files/papers/WP2018-7_0.pdf. [1]
- Leulescu, A. et al. (2025), “Flash estimates of income distribution: capturing labour market dynamics – 2025 edition”, *Eurostat, Statistical working papers, Publications Office of the European Union*, <https://doi.org/10.2785/5003021>. [14]
- Levy, H. (2023), “Nowcasting and provisional estimates of income inequality using microsimulation techniques”, *OECD Papers on Well-being and Inequalities*, No. 12, OECD Publishing, Paris, <https://doi.org/10.1787/c0b88e7f-en>. [6]
- Mahler, D., R. Castañeda Aguilar and D. Newhouse (2022), “Nowcasting global poverty”, *The World Bank Economic Review*, Vol. 36/4, pp. 835–856, <https://doi.org/10.1093/wber/lhac017>. [11]
- Navicke, J., O. Rastrigina and H. Sutherland (2013), “Nowcasting Indicators of Poverty Risk in the European Union: A Microsimulation Approach”, *Social Indicators Research*, Vol. 119/1, pp. 101-119, <https://doi.org/10.1007/s11205-013-0491-8>. [7]
- O’Donoghue, C. and D. Sologon (2023), “The transformation of public policy analysis in times of crisis — a microsimulation-nowcasting method using big data”, *IZA Discussion Papers*, Vol. 15937, <https://www.iza.org/publications/dp/15937/the-transformation-of-public-policy-analysis-in-times-of-crisis-a-microsimulation-nowcasting-method-using-big-data>. [9]

- OECD (2025), *OECD Economic Outlook, Volume 2025, Issue 1*, OECD Publishing, Paris, [13]
https://www.oecd.org/content/dam/oecd/en/topics/policy-sub-issues/economic-outlook/eo117-june-2025/EO117_Annexes_E.pdf.
- Pierdzioch, C. et al. (2022), “Forecasting changes of economic inequality: A boosting approach”, *The Social Science Journal*, Vol. 59/2, pp. 252-268, [2]
<https://doi.org/10.1016/j.sosci.2019.09.001>.
- Sologon, D. et al. (2022), “The COVID-19 resilience of a continental welfare regime — nowcasting the distributional impact of the crisis”, *The Journal of Economic Inequality*, Vol. 20/4, pp. 777–809, [8]
<https://doi.org/10.1007/s10888-021-09524-4>.

Annex A. Methodological note on the overall framework

This annex provides further technical details on the framework used to nowcast disposable-income distribution. A detailed description of the labour market adjustment methodology is provided in Annex B.

A.1. System architecture

The system works in two stages:

- **Stage 1 – ML Forecasting:** Macroeconomic indicators from the OECD Economic Outlook are used to forecast decile-level income and average income uprating factors (URFs) for each EU country and forecast horizon. Six algorithms are supported: OLS, LASSO, Ridge, Elastic Net, GBM, and RF. For average income uprating, the system also allows the use of the standard uprating factors embedded in EUROMOD.
- **Stage 2 – EUROMOD Microsimulation:** The URFs are applied to EUROMOD household microdata to scale gross market incomes. EUROMOD is re-run to simulate the tax-benefit system response, yielding a forecast of equivalised household disposable income. The Gini coefficient and poverty rate are subsequently calculated from the resulting disposable income distribution.

Income growth is modelled separately for each income decile and for average income. Average income and decile income shares are uprated with separate URFs. Users can choose to uprate average incomes using ML-computed URFs or to use standard EUROMOD uprating. The system can be run for various income concepts: (i) disposable income, (ii) labour and capital income and (iii) market income.

- (i) When *disposable income* is predicted, decile-level shares of equivalised household disposable income are used as the dependent variables. Average equivalised household disposable income is used to uprate the overall income level. The URFs are applied directly to equivalised household disposable income, and EUROMOD is not re-run following the uprating.
- (ii) When *labour and capital income* are predicted, two sets of URFs are computed, one for labour income and one for capital income. Log decile shares of labour income and decile-level capital income levels are used as dependent variables, respectively. Average labour and capital incomes are used to uprate the mean income to capture the growth of these income sources. The URFs are then applied to individual labour and capital incomes, and EUROMOD is re-run following the uprating to compute disposable income.
- (iii) When *market income* is predicted, the procedure is similar to that under option (ii), except that only one URF is computed for each decile-country-horizon combination rather than two. Decile individual market income (labour + capital) shares are used as dependent variable. Average market income is used to uprate the mean income to capture the growth rate of these income sources. The URFs are then applied to individual market incomes, and EUROMOD is re-run following the uprating to compute disposable income.

Stage 1 – ML Upating Factor Estimation

For each income decile d ($d = 1 \dots 10$) and country c ($c = 1 \dots N$), the model predicts a value for each decile's income share and for average income. The dependent variable is the (log of the) income share or average income and the independent variables are macroeconomic variables and the h -year lag of income.

$$\hat{y}_{share_{dc,t}} = \alpha_{dh} + \beta_{dh} y_{share_{dc,t-h}} + \sum_{j=1}^K \gamma_j x_{jct} + \sum_{r=1}^{N-1} \delta_r D_{rc} + \varepsilon$$

where $y_{share_{dc,t-h}}$ is the observed income share for income decile d ($d = 1 \dots 10$) and country c ($c = 1 \dots N$) at time $t - h$, ($h = 1 \dots 4$), x_{jct} are macroeconomic variables, D_c are a set of country-dummy variables and ε is a prediction error. The model is estimated on a pooled panel of country-year observations.

The model then computes the ratio of the income share (or level) in the target year t to its base-period value at $t - h$. This ratio constitutes the Upating Factor:

$$URF_{dc,t,h} = \hat{y}_{share_{dc,t}} / y_{share_{dc,t-h}}$$

where y_{share_t} denotes the ML-predicted value in the target year t (also denoted F_t) and $y_{share_{t-h}}$ is the lagged observed base-period value at $t-h$ (also denoted L_t). A separate URF for mean income is optionally estimated to help maintain consistency between the income level and EUROMOD's income thresholds.

For average income $\bar{y}$, the model computes the URF as the ratio of predicted average income in the target year t to its observed base-period value at $t-h$.

$$URF_{\bar{y},t,h} = \hat{y}_{\bar{y},t} / y_{\bar{y},t-h}$$

Upated income is therefore found by

$$y_t = y_{t-h} * URF_{dc,t,h} * URF_{\bar{y},t,h}$$

The first factor adjusts the shape of the distribution, while the second factor shifts the level of the entire distribution (preserving relative shares). Together they ensure that both the level of income and its distribution are updated simultaneously. For average-income upating, the system also allows the use of the standard upating factors embedded in EUROMOD.

Our system also allows using the Labour Market Adjustment (LMA) module to simulate transitions into and out of employment (see Annex B for a detailed description of the methodology). When the Labour market adjustment (LMA) module is applied, the ML-based upating factor (URF) for average income is computed using the employed individuals only, which avoids double accounting for the impact of employment loss on average incomes and more closely reflects wage growth.

Stage 2 – EUROMOD Microsimulation

The URFs are applied to microdata and income components are scaled by average income and adjusted for changes in the decile income share for each decile, and the updated microdata is passed to EUROMOD to re-compute tax liabilities and benefit entitlements using the tax-benefit rules of the forecast year, where available. The result is a full estimate of equivalised household disposable income. As the system is implemented in Stata, it utilises the EUROMOD Stata connector to call EUROMOD from within the system.

A.2. Data

EUROMOD Household Microdata

EUROMOD operates on representative household microdata derived from EU-SILC. The data covers EU member states from 2006 to 2024 in the current configuration. Our system runs EUROMOD twice. First,

EUROMOD is run to compute income lags and baseline disposable incomes to assess pseudo out-of-sample prediction performance. Second, EUROMOD is re-run after incomes are uprated to account for the interaction of tax and benefit policy with income levels.

Careful alignment of EUROMOD data years, system years, and macroeconomic variables is required because EUROMOD income data refer to the preceding calendar year. For example, the 2016 dataset contains incomes for 2015. Accordingly, in the baseline run, the system year and macroeconomic variables are aligned with the income reference year (data year – 1). In the nowcasting run, the system year is shifted forward according to the forecast horizon (data year – 1 + h).

Table A A.1. EUROMOD variables description

Variable	Description	Role
yem	Employee income	Labour income – uprated by decile (and average)
yse	Self-employment income	Labour income – uprated by decile (and average)
iyi	Investment income	Capital income – uprated by decile (and average)
ypr	Property income	Capital income – uprated by decile (and average)
ils_dispy	Equalised disposable income	Primary outcome for Gini and poverty rate computation
dwt	Demographic / survey weight	Used in all aggregations, Gini and poverty rate

Note: Household income is equalised using the square root of household size.

OECD Economic Outlook (macroeconomic predictors)

Macroeconomic predictors are drawn from the *OECD Economic Outlook* (OECD, 2025_[13]). The predictor set is relatively short compared to previous studies (see Table A A.5 later in this annex) and includes:

Table A A.2. OECD macroeconomic predictors

Variable (OECD code)	Description
GDP_ANNPCT	GDP growth (annual %)
UNR	Unemployment rate
sh_WSSS	Compensation of employees (Share of GDP)
sh_CG / sh_CP / sh_IT	Govt consumption / Private consumption / Investment (Share of GDP)
IRCB / IRS	Central bank / short-term interest rates
GDP_deflat_ANNPCT	GDP deflator growth (annual %)
ULC	Unit labour costs
CGV, CPV, ITV, ET, FDDV, MGSV, TDDV_ANNPCT	Growth (annual %) of Govt consumption / Private consumption / Investment / Employment / Final domestic expenditure / Imports / Total domestic expenditure
CBGDPR/BSIIQ/NTRQ/ FBGSQ	Current account balance / Balance of primary income / Balance of secondary income / Net exports of goods and services (% of GDP)
crisis2008	Dummy: global financial crisis (2008–2010)
eurocrisis	Dummy: Euro area sovereign debt crisis (2011–13)
covid2020	Dummy: COVID-19 shock (2020)
energy2022	Dummy: energy price crisis (2022)
dumCtry*	Country fixed-effect dummies
Region interactions	Region (Nordic, Central, Southern, Eastern EU countries) × lags and sh_CG
time, time2	Linear, quadratic time trends

Income concepts

The system supports four income concepts, controlled by the parameter *def*. This choice determines which EUROMOD income components are uprated, what functional form is applied to the dependent variable, and whether EUROMOD is subsequently re-run.

Table A A.3. Income concepts

Code	Concept	Components uprated	Dependent variable	EUROMOD re-running
disp	Disposable income	ils_dispy	Income share by decile	No
mkt	Market income	yem + yse + yiy + ypr	Income share by decile	Yes
lab	Labour income	yem + yse	Log income share by decile	Yes (joint with cap)
cap	Capital income	yiy + ypr	Income level by decile	Yes (joint with lab)

When *def* = *lab*, capital income (*cap*) is estimated simultaneously, and both components are uprated before EUROMOD re-simulation. The *disp* specification is computationally lighter but does not capture the tax-benefit system's response to income changes; the *mkt* and *lab/cap* specifications route through EUROMOD and are preferred when policy-sensitivity is required.

A.3. Method

Model pipeline

The pipeline comprises five sequential steps (see Figure A A.2 later in this annex for an overview). Steps 1 and 2 are estimation steps (run once per configuration); Steps 3-5 are nowcasting and validation steps.

Step 1 – EUROMOD data preparation

For each country and data year, EUROMOD is run via the Stata connector, applying the tax-benefit rules of the *preceding* year (system year = data year – 1), consistent with EU-SILC income reference conventions. Individual-level outputs are aggregated to household level and equivalised. Households are ranked into deciles on equivalised disposable income (weighted by *dwt*). Decile income totals and conditional averages (conditional on positive income) are computed.

Output: EUROMOD_AGG_y_allyearscountries.dta — panel at country–year–decile level.

Step 2 – ML uprating factor estimation

This is the core estimation step. EUROMOD aggregates are merged with OECD macro indicators; dependent variables (income shares, log shares, or levels by decile) and their lags are constructed. The model is estimated recursively and URFs are saved for use in Step 3.

Output: UPRF_share_{def}_inc_{algo}.dta

Step 3 – Income nowcasting via EUROMOD

For each country and base year, EUROMOD microdata is loaded. When the Labour Market Adjustment (LMA) module is used, the corresponding LMA microdata are loaded first. Individuals are assigned to deciles based on the base-year income distribution. URFs from Step 2 are then applied in two stages:

- **Average-income uprating:** when ML-based mean-income uprating is selected, the relevant income measure or components are scaled by the corresponding mean-income URF, updating the overall nominal income level. Alternatively, the standard auto-uprating for average incomes embedded in EUROMOD is also available.

- **Decile-specific uprating:** income is further adjusted using the decile-specific URF associated with the individual's base-year decile, reshaping the income distribution.

For *def = disp*: uprated disposable income is taken directly without re-running EUROMOD. For *def = mkt* or *lab*: individual income components (*yem*, *yse*, *yi*, *ypr*) are scaled and the uprated dataset replaces the EUROMOD input file. EUROMOD is then re-run using the tax-benefit rules of the forecast year, where available. The output is a full forecasted household disposable income distribution, from which the Gini and poverty rate are computed.

Output: Forecasted_Gini_{algo}.dta

Step 4 – Diagnostics

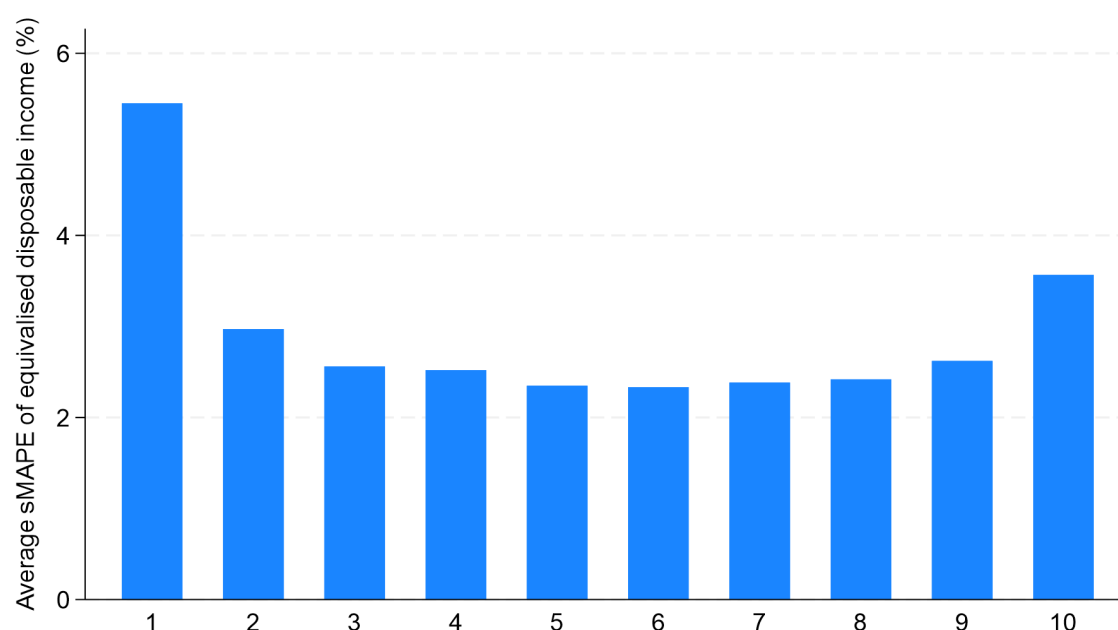
Forecast performance is assessed using several complementary diagnostics. For the final distributional indicators, mean absolute error (MAE) and directional accuracy (DA) are computed across recursive out-of-sample predictions by forecast horizon. MAE measures the absolute difference between predicted and target Gini coefficients or poverty rates, while DA measures whether the model correctly predicts the direction of change between the forecast origin and the target year. At the income-distribution level, sMAPE is calculated for equivalised disposable income by decile and forecast horizon to assess where prediction errors are concentrated across the distribution.

Table A A.4. Diagnostic statistics

Measurement	Description
MAE	Mean Absolute Error – average absolute difference between predicted and target values.
DA	Directional Accuracy – proportion of cases in which the model correctly predicts the direction of change in the Gini coefficient or poverty rate between the forecast origin and the target year.
sMAPE	symmetric Mean Absolute Percentage Error – average relative forecast error in percent, using both predicted and target values in the denominator.

Figure A A.1 shows the average sMAPE of equivalised disposable income by income decile for the 1-year-ahead forecast. Errors are relatively small in the middle deciles but larger at the bottom and top of the distribution, suggesting that poverty-rate and inequality nowcasting may be particularly sensitive to prediction errors in these parts of the income distribution.

Figure A A.1. Average sMAPE of income level by income decile (1-year-ahead predictions)



Note: sMAPE (=symmetric Mean Absolute Percentage Error) is calculated as the absolute difference between predicted and target equivalised disposable income, divided by the average of their absolute values, and multiplied by 100. The results refer to 1-year-ahead predictions pooled over 2016-2024. The LASSO-based uprating of market income is combined with LMA and EUROMOD-simulated taxes and benefits.

Step 5 – Visualisation and comparison

Forecasted Gini coefficients and poverty rates are merged with observed ones. Visualisation outputs include scatter plots of predicted indicators with confidence bounds versus observed ones by horizon and a regression of observed on predicted indicators (with year fixed effects).

For the Gini coefficient, confidence bounds are derived by combining the lower 95% confidence limits for deciles 1-5 with the upper limits for deciles 6-10 to construct the upper-Gini distribution, and vice versa for the lower-Gini distribution. For the poverty rate, the same lower- and upper-bound income distributions are used, with the poverty threshold recalculated in each case as 60% of the weighted median equivalised disposable income. These bounds reflect uncertainty in the decile-level predictions rather than conventional confidence intervals estimated directly for the Gini coefficient or poverty rate.

Machine Learning algorithms

The system supports five ML-based estimation approaches: Post-LASSO, Ridge, Elastic Net, Gradient Boosting, and Random Forest. For Post-LASSO, Ridge, and Elastic Net, regularisation is performed separately for each combination of forecast horizon and income decile, allowing the selected predictor set and shrinkage intensity to vary across the distribution and forecast horizon. Gradient Boosting and Random Forest are implemented to capture potential non-linear relationships and interactions among predictors.

The models draw from a common underlying set of candidate predictors, with model-specific procedures used to control complexity and select or weight predictors. Model selection and tuning are repeated at each recursive cutoff where applicable, allowing the specification to adapt as the estimation sample expands. Users can also modify model specifications and tuning settings where alternative assumptions or additional country-specific knowledge are considered relevant.

Pseudo-out-of-sample estimation framework

The system is validated using a pseudo-out-of-sample (OOS) design. For each cutoff year t^* (income reference year) and forecast horizon h :

- **Estimation sample:** all country-year observations with income reference year $\leq t^*$.
- **Prediction target:** the income shares (or log shares, or levels) for year $t^* + h$.
- **Upating factor:** predicted value for $t^* + h$ divided by the lagged base value at t^* .

This design ensures that future-dated observations are never used to estimate models for earlier periods, approximating the real-time information constraint. It also enables horizon-specific assessment ($h = 1, 2, 3, 4$ years ahead) and allows the regularised models to adapt across cutoff years as the relevance of macroeconomic predictors evolves. However, OECD macroeconomic data are taken from the latest database vintage rather than archived real-time vintages, so the results may benefit from subsequent revisions to historical data. In addition, EUROMOD policy-system information used for longer forecast horizons may not have been available at the historical cutoff date.

EUROMOD integration

System and data year alignment

A critical implementation detail governs the year alignment when EUROMOD is re-run in the nowcasting step (*def = mkt* or *lab*). The system year is set equal to the forecast year ($= \text{data year} - 1 + h$), not to data year - 1 as in the data preparation step. This activates EUROMOD's built-in mean-income uprating factors and the tax-benefit rules of the forecast year, where available, in combination with the ML-predicted decile-specific URFs. When the ML mean-income uprating is active (*sc_uprate_averagelIncome = 1*), EUROMOD's built-in uprating function must be switched off to avoid double uprating.

Labour Market Adjustment (LMA)

The system includes provision for the Labour Market Adjustment (LMA) module, which adjusts employment levels before income uprating (see Annex B for a detailed description of the methodology). When it is active (*sc_LMA_was_run = 1*), the LMA module is applied for all country-years before running the nowcasting step. In addition, when the ML mean-income uprating is active (*sc_uprate_averagelIncome = 1*), conditional positive-income measures are used during URF estimation. This isolates changes in income levels from changes in employment composition and avoids conflating the two effects.

A.4. Limitations

Users should be aware of the following constraints on the scope and reliability of the nowcasts:

- **Short estimation sample:** EUROMOD microdata currently covers 2006-2024, limiting estimation precision and OOS validation power, particularly at longer horizons.
- **Static decile assignment:** decile membership is fixed at the base year. The model does not capture income mobility (transitions between deciles over the forecast horizon).
- **No behavioural responses:** labour supply, savings, and consumption responses to income changes are not modelled. Only the mechanical tax-benefit response through EUROMOD and LMA-based employment transitions are captured.
- **Macro predictor uncertainty:** OECD Economic Outlook projections carry their own forecast errors, which propagate into the inequality nowcasts, particularly at longer horizons.

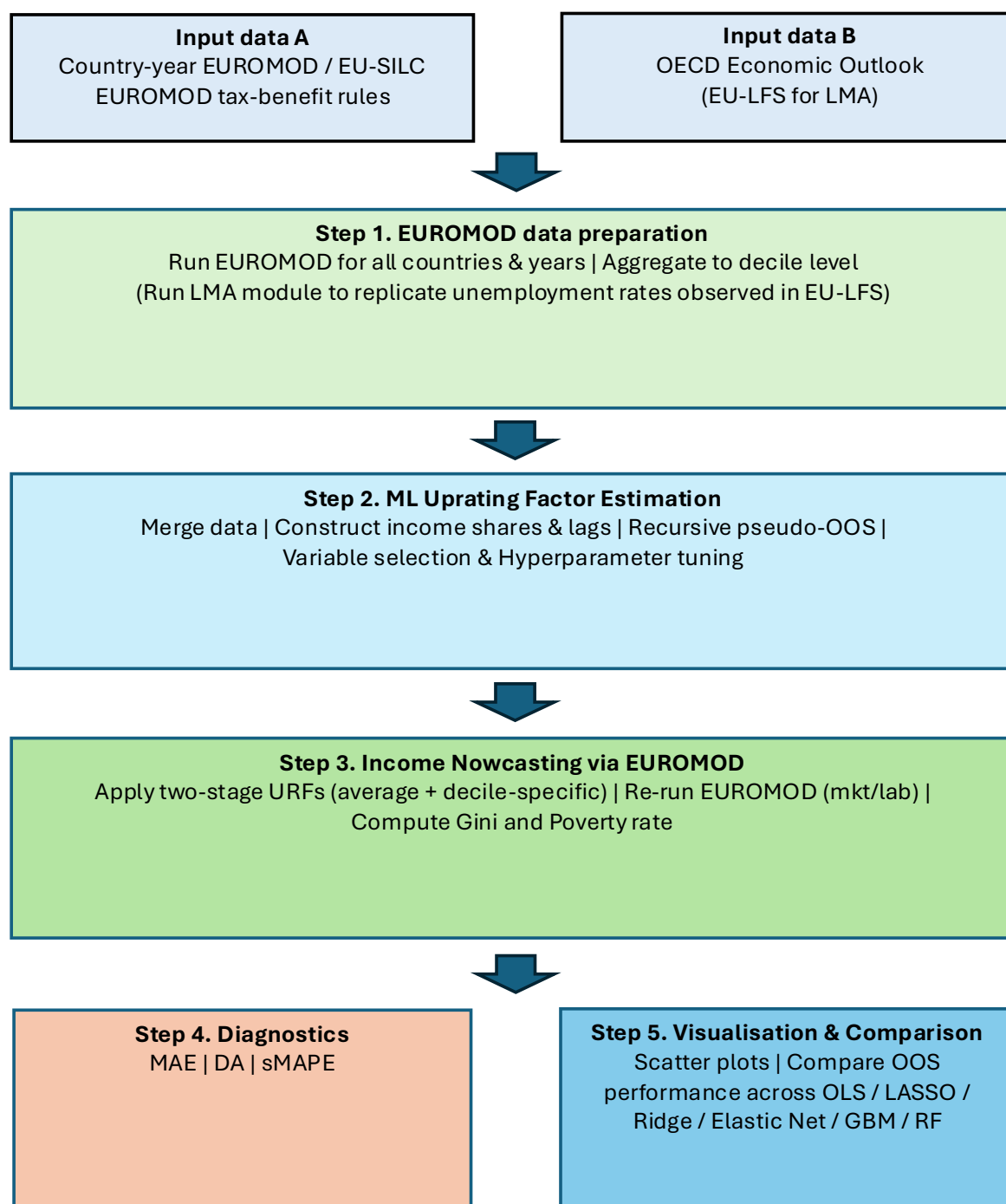
- **Micro data uncertainty:** For certain country-years, year-on-year variations in the disposable income Gini coefficients or poverty rates are very significant, potentially due to changes in sampling design or data collection error. Users should identify such changes.
- **Country coverage:** not all EU member states have EUROMOD microdata for all years. Countries with missing data are excluded from the relevant forecast years.

Table A A.5. Macroeconomic variables and predictors used in related studies

Predictor	Eurostat (2020)	Cornwall and Gindelsky (2025)	Gindelsky (2018)	Castle et al. (2024)	Mahler et al. (2025)
Coverage	EU	USA	USA	UK	Global
GDP (at market prices)	X		X***		
GNI (gross national income)	X				
Inflation / HICP	X	X	X	X	
Activity rate	X	X			
Part-time employment rate	X				
Temporary employment rate	X				
Unemployment	X		X		
Long-term unemployment	X				
Total employment	X				
Compensation of employees	X				
HH savings rate	X				
HH final consumption expenditure	X				
Wages and salaries	X	X*			
Social Benefits	X	X**			
Taxes	X				
GDI (gross disposable income)	X				
Assets		X*			
Rental income		X*			
Case-Shiller Index		X			
Federal Funds rate		X			
Labour force participation rate		X	X		
Mortgage rate		X			
Treasury 5-year rate		X			
Unemployment rate		X			
Education achievement			X		
Employment by skill			X		
Services as share of GDP			X		
Government expenditure as share of GDP			X		
GDP per capita				X	X
Real wage growth				X	
Average tax burden				X	
KOF Globalisation index				X	
Global State of Democracy index				X	
Under-5 mortality rate					X
Life expectancy					X
Rural population					X
Regional dummies					X

Notes: * As share of personal income. ** Medicare, Medicaid, Tax credits, Transfers, Unemployment. *** Real GDP.

Figure A A.2. Overview of the model pipeline



Annex B. Methodological note on the labour market adjustment module

This annex provides further technical details on the Labour Market Adjustment (LMA) module of the nowcasting system. The module simulates individual transitions into and out of employment, and the associated changes in labour income, before income uprating is applied.

B.1. Overview

The LMA module adjusts the employment status of working-age individuals in the EUROMOD microdata so that simulated labour-market flows reproduce macroeconomic transition rates observed in the EU Labour Force Survey (EU-LFS) and then reassigns labour income consistent with the new status. It is applied to each country and base-year dataset prior to income uprating and can be iterated over successive one-year steps (“horizons”) to build an accumulated multi-year adjustment.

The procedure consists of four steps:

- **Transition probabilities.** Individual probabilities of job loss (employment → unemployment) and job finding (unemployment → employment) are computed from logit coefficients estimated externally by Eurostat, evaluated on EUROMOD individual characteristics.
- **Macro calibration.** These individual probabilities are rescaled by a country-year-specific factor so that their weighted mean matches the macroeconomic separation and finding rates derived from the EU-LFS.
- **Stochastic transition assignment.** Each individual is assigned a random number drawn from a uniform distribution. Transitions are assigned by comparing the draw to the transition probability, conditional on the previous labour-market status.
- **Income reassignment.** Labour income is set to zero for new job losers and imputed by predictive mean matching for new job finders.

The scaling matches the macroeconomic target on the weighted mean of individual probabilities, so the simulated transition rate matches the target in expectation rather than exactly in any given draw. The module is run under a fixed random seed for reproducibility.

B.2. Data

Macroeconomic transition target

Macroeconomic separation and finding rates are constructed from EU-LFS data over 2009-2025.⁷ The separation rate is the ratio of short-term unemployment to lagged employment, and the finding rate is one minus the ratio of long-term unemployment to lagged unemployment:

$$s(t) = \frac{STU_t}{E_{t-1}}$$

$$f(t) = 1 - \frac{LTU_t}{U_{t-1}}$$

where STU_t and LTU_t denote short-term and long-term unemployment respectively, and E_{t-1} and U_{t-1} are lagged employment and unemployment.

Because the unemployment level in the EUROMOD microdata generally does not coincide with the LFS level, the separation target is calibrated to the EUROMOD level before use. For each country-year, the system computes the ratio of the EUROMOD to the LFS unemployment rate and rescales the separation rate by this ratio, bounding the result to the unit interval:

$$\lambda(t) = \frac{UR_{EUROMOD}(t)}{UR_{LFS}(t)}$$

$$s_{adj}(t) = \min(\lambda(t) \cdot s(t), 1)$$

The finding rate is not adjusted and is equal to the LFS value, $f_{adj}(t) = f(t)$. The individual-level transitions are therefore calibrated against the macroeconomic target values $s_{adj}(t)$ and $f(t)$.

Transition probabilities

The module relies on externally estimated transition probabilities to select which individuals transition into or out of employment. Individual transition probabilities are based on logit coefficients taken from the Eurostat publication “Flash estimates of income distribution: capturing labour market dynamics” (Leulescu et al., 2025^[14]). For each individual, the module computes two probabilities using country-specific coefficients, one for job loss and one for job finding. User-estimated equations can be substituted for the Eurostat coefficients, and the Eurostat coefficients are the default of the system.

The job-loss equation is evaluated on characteristics available in EUROMOD: age and age squared; citizenship; education; contract type (permanent contract or not); occupation; an employee dummy; sector (NACE Rev.2); gender; student status; degree of urbanisation; and work experience in years. The job-finding equation is evaluated on age and age squared, citizenship, unemployment duration, education, sex, student status, and degree of urbanisation.

B.3. Method

Step 1. Individual-level transition probabilities

For each individual i , the linear predictors for job loss and job finding are formed from the Eurostat coefficients and the corresponding EUROMOD covariates, and converted to probabilities through the logistic (inverse-logit) function $\Lambda(\cdot)$:

$$p_{loss}(i) = \Lambda(x_i \beta_{loss})$$

$$p_{find}(i) = \Lambda(x_i \beta_{find})$$

Two incompatibilities between the EUROMOD data and the Eurostat specification are handled and are discussed in Section B.5: EUROMOD data do not provide information on permanent contracts, and the module therefore constructs a proxy for permanent employment, and the longest unemployment-duration band is not identified in the EUROMOD data as it exceeds 1 year and EUROMOD datasets are annual.

Step 2. Calibration to macroeconomic targets

The module calibrates the probabilities in Step 1 to the macroeconomic targets using multiplicative scaling. The scaling constant for job loss is the separation target divided by the weighted mean of the job-loss

probability computed in Step 1 (weighted using EUROMOD household weights), taken over individuals whose previous status is employed or self-employed. Similarly, the constant for job finding uses the finding target and the weighted mean of the job finding probability over individuals previously unemployed.

$$c_{loss}(T) = \frac{s_{adj}(T)}{mean_w[p_{loss} | prev. emp.]}$$

$$c_{find}(T) = \frac{f(T)}{mean_w[p_{find} | prev. unemp.]}$$

where $mean_w[\cdot]$ denotes the weighted mean over the relevant previous-status group. As EUROMOD income refers to the previous year while labour-market status is observed in the survey year, the LFS transition rates are aligned with the income reference year. At forecast horizon h from income reference year t , the relevant target year is $T=t+h$. The calibrated probabilities are then capped at one:

$$\pi_{loss}(i, T) = \min(c_{loss}(T) \cdot p_{loss}(i), 1)$$

$$\pi_{find}(i, T) = \min(c_{find}(T) \cdot p_{find}(i), 1)$$

By construction, the (uncapped) weighted mean of the calibrated probability over the previous-status group equals the target. Therefore, the simulated transition rate approximately reproduces the macroeconomic target (in expectation).

Step 3. Assigning individuals to transition

Transitions are assigned using Monte Carlo simulations. Each individual draws an independent uniform random number $u(i)$ per horizon. Depending on the previous labour-market status, the system assigns:

- an employment-to-unemployment transition where the individual was employed or self-employed and $u(i) < \pi_{loss}(i, T)$;
- an unemployment-to-employment transition where the individual was unemployed and $u(i) < \pi_{find}(i, T)$.

The simulated status is saved separately for each horizon. For multi-year runs, the procedure is cumulative: for $h>1$, the status simulated at $h-1$ becomes the previous status used to calibrate and simulate horizon h .

Step 4. Income assignment for transitioning individuals

Labour income is updated to reflect the simulated transitions. For individuals moving from employment into unemployment, employment income, self-employment income, and hours are set to zero. For individuals moving from unemployment into employment, employment income and hours are not observed and must be imputed. The module imputes employment income and hours worked using Predictive Mean Matching (PMM), where newly employed individuals (Recipient) receive the income and hours worked of another individual (Donor) with the most comparable characteristics.

- **Donors** are individuals in employment, self-employment, or farming, aged 16-65, with positive employment income and non-missing hours.
- **Recipients** are those moving from unemployed to employed in the current horizon.

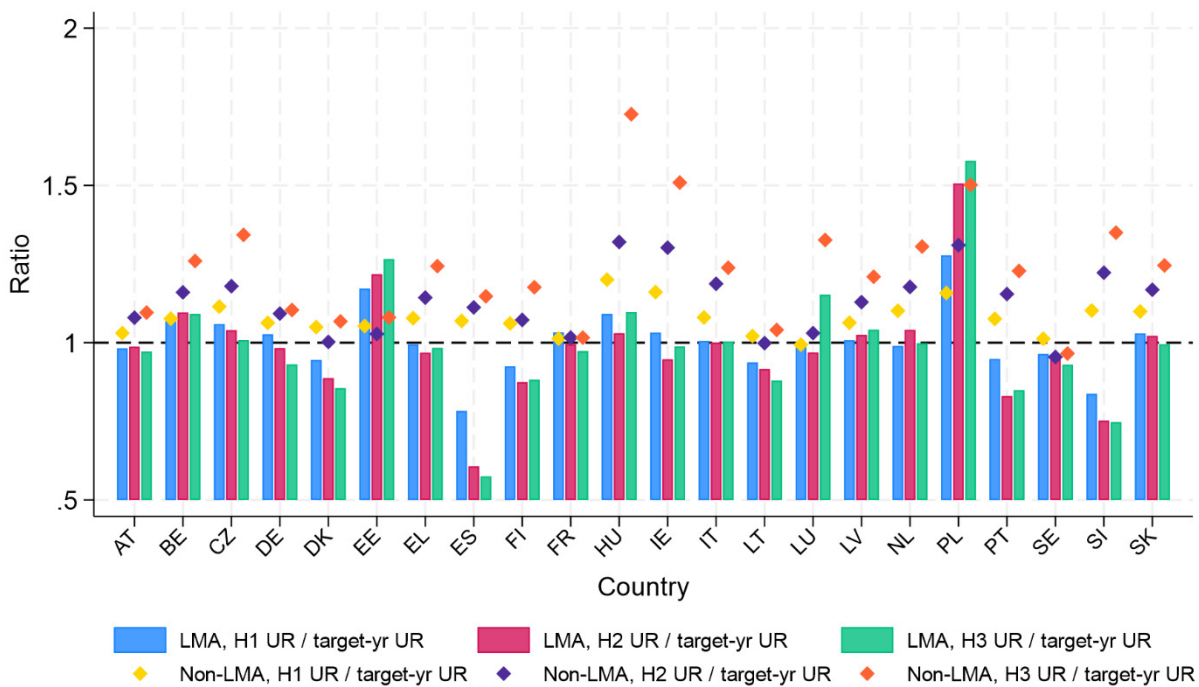
The imputation is performed jointly on log employment income and on hours, by monotone PMM with five nearest neighbours, conditioning on age and age squared, education, sex, work experience, sector, occupation, and degree of urbanisation, weighted by the design weight and run under a fixed seed. Imputed log income is exponentiated back to levels and assigned to recipients. Self-employment income is not imputed for the newly employed. The module therefore assumes that all newly employed individuals are employees and no one transitions from unemployment into self-employment.

B.4. Output and diagnostics

For each country, base year, and horizon, the module saves an updated EUROMOD microdata file containing the simulated labour-market status, the original and adjusted labour income. These files are then passed back to the system, where market incomes are updated, and EUROMOD is run to simulate disposable income and the updated disposable income distribution. When LMA is combined with ML-based mean-income uprating, conditional positive-income measures are used in the estimation of the URFs, helping to separate income-level changes from changes in employment composition.

Figure A B.1 provides a diagnostic check of how the LMA affects simulated labour-market status. It compares, for each country and forecast horizon, the unemployment rate in the nowcasted dataset with the unemployment rate observed in the corresponding EUROMOD target-year dataset. A ratio close to one indicates that the simulated unemployment rate is close to the target-year EUROMOD unemployment rate. The figure shows that the LMA ratios are generally closer to one than the non-LMA ratios, especially at longer forecast horizons where the non-LMA tends to deviate more strongly from the target-year unemployment rate. This suggests that the LMA module improves the alignment of simulated labour-market status with observed labour-market conditions, although some country-specific deviations remain.

Figure A B.1. Unemployment-rate ratios across countries



Note: The ratio is calculated as the unemployment rate in the nowcasted dataset divided by the unemployment rate in the corresponding EUROMOD target-year dataset, by country and forecast horizon. Bars show ratios with LMA for horizons 1-3, while markers show the corresponding ratios without LMA. A value of one indicates exact alignment with the target-year EUROMOD unemployment rate.

B.5. Limitations

Users should be aware of the following limitations, several of which arise from incompatibilities between the EUROMOD input data and the Eurostat transition equations:

- **Permanent-contract proxy.** EUROMOD contains no variable for permanent-contract status or tenure. The indicator entering the job-loss equation is a proxy, set to one for civil servants and for employees who worked the full year, and zero otherwise. The Eurostat coefficient on permanent contract is therefore applied to an approximate, rather than an observed, contract status. A correlation coefficient of above 0.85 is found between this proxy and the permanent-contract variable in EU-SILC on a pooled sample of EU-countries and years. This provides the module with some confidence in this proxy as EUROMOD micro datasets are constructed from EU-SILC data.
- **Unemployment-duration top-coding.** Unemployment duration in EUROMOD is top-coded at twelve months. The job-finding equation includes a band for unemployment of 24 months or more, but this regressor is identically zero in the EUROMOD data. The duration profile of job-finding probabilities is truncated relative to the estimated specification.
- **Approximate calibration.** The multiplicative scaling matches the target on the weighted mean of individual probabilities. The realised transition rate in any given simulation draw differs from the target through sampling variation, and capping calibrated probabilities at one introduces a downward deviation when the scaling constant is large.
- **Static individual characteristics.** Covariates entering the transition equations are held at base-year values across horizons. Characteristics that would evolve with a simulated spell, such as accumulated unemployment duration or age, are not updated.
- **No behavioural response.** Transitions are driven by calibrated statistical probabilities, not by an economic model of labour supply or demand. Wages assigned to job finders are imputed from observed earners and reflect no selection into employment beyond the PMM conditioning variables.
- **External coefficient dependence.** Results inherit the assumptions and country coverage of the Eurostat “Flash estimates” coefficients. Where the Eurostat specification differs materially from local labour-market structure, the calibration corrects the mean transition rate but not the distribution of transitions across individuals.

Notes

¹ EUROMOD is the European Union's tax-benefit microsimulation model, which enables researchers and policy analysts to assess, in a comparable manner, the effects of taxes and benefits on household incomes across EU countries. For details, see: <https://euromod-web.jrc.ec.europa.eu/>.

² Up-rating factors are the adjustment rates used to bring monetary variables from one year to another, so that all values are expressed in consistent terms.

³ The two types of up-rating serve distinct purposes. The first (mean-income URF) updates the level of market income so that EUROMOD's eligibility rules remain consistent with the forecast year, while the second (the decile-specific URF) changes the distribution. Together they ensure that both the level and distribution of income are updated.

⁴ For example, EUROMOD up-rates employment income using the growth of average hourly wages (in some cases by sector or gender), while other income sources are up-rated using the corresponding source-specific indices. For details, see EUROMOD Conventions: <https://euromod-web.jrc.ec.europa.eu/resources/modelling-conventions#section-3672>.

⁵ The square-root equivalence scale is used for consistency with OECD income-distribution analysis and provides a simple way of accounting for household economies of scale. The same scale is applied to both observed and predicted distributional indicators used in the validation exercise. Eurostat's EU-SILC indicators use the modified OECD scale; both are established approaches, although estimates based on the two scales are not directly comparable.

⁶ However, the OECD macroeconomic indicators and forecasts are taken from the latest available database rather than from archived real-time vintages. The exercise therefore prevents future-dated observations from entering the estimation but may still benefit from subsequent revisions to historical data. In addition, EUROMOD policy-system information used for longer forecast horizons may not have been available at the historical cutoff date.

⁷ [Eurostat \(LFS\)](#): "Employment indicators (lfsi_emp_a)" and "Unemployment by duration (lfsa_ugad)"