



EUROPEAN CENTRAL BANK
EUROSYSTEM

Working Paper Series

Michele Lenza, Claudia Marchini

Forecasting wages with local linear forests

No 3291 / October 2026



Disclaimer: This paper should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the authors and do not necessarily reflect those of the ECB.

Abstract

This study applies a Local Linear Forest (LLF) for wage forecasting in France. The LLF outperforms benchmarks such as Random Walk (RW), Ridge Regressions (RR) and Random Forests (RF). We also show that adding foreign predictors (i.e. measures of real activity, price and wage pressures from Germany and Italy) to French economic variables significantly improves wage predictions in France. The implications of our results are that wage dynamics in France exhibit non-linearity, to a certain extent, and that the better ability of LLFs relative to RFs to fit smooth signals is a valuable feature, especially at times in which wage growth reaches unprecedented levels from an historical perspective.

Keywords: Negotiated Wages, Local Linear Forest, Forecasting.

JEL Codes: C45, C55, E37, J30

Non-Technical Summary

In this paper, we adopt a machine learning technique called Local Linear Forests (LLF) to predict wage growth in France.

Wages are a key ingredient in understanding inflation, since they represent a large share of business costs, and central banks such as the ECB pay close attention to wage trends to avoid the kind of harmful “wage-price spirals” seen in the 1970s.

Traditionally, economists have relied on relatively simple statistical models to forecast wages, but this study tests whether more advanced, flexible methods can do a better job, especially when wage growth reaches unusual levels. Specifically, we compare four forecasting approaches: a simple “no change” prediction (random walk), a standard linear statistical model (ridge regression), a machine learning method called random forests, and local linear forests, which blends the flexibility of machine learning with the strengths of simple linear models.

We test these models using French economic data—such as unemployment, productivity, inflation expectations, and past wage growth—going back to 1985, and also check whether adding similar economic data from Germany and Italy improves the forecasts for France.

The results show that all the models outperform the naive “no change” prediction, confirming that economic indicators do help forecast wages. More importantly, the local linear forest model consistently performs best, especially over longer forecast horizons, showing that wage dynamics in France have some non-linear patterns that simpler models can’t fully capture. The local linear forest also does better than the plain random forest, particularly during the unusual post-financial-crisis period when wage growth fell to historically low levels, suggesting that its ability to track smooth signals is especially valuable in unusual times. Finally, the study finds that including economic indicators from Germany and Italy meaningfully improves the accuracy of French wage forecasts, suggesting that despite differences in labor market institutions across European countries, wage and price dynamics are closely interconnected across the euro area.

Overall, the message of the paper is that policymakers and economists forecasting wages—an important piece of the inflation puzzle—should consider using more flexible, non-linear tools like local linear forests, and should also pay attention to economic developments beyond national borders when assessing domestic wage trends.

1 Introduction

Monetary policy actions notoriously affect the economy with a considerable delay. For this reason, central banks aim at achieving their inflation targets in the “medium term”, placing a relatively lower weight on changes in inflation which are due to temporary factors. This explains why central banks monitor measures of underlying inflation pressures which, by filtering out the effects of temporary shocks from headline inflation, help to track the persistent component of inflation.

Wage dynamics are generally considered among the most relevant factors behind underlying inflation pressures. For example, [Fagandini et al. \(2024\)](#) shows that, because of their large share in input costs, changes in wages are an important factor for the evolution of the euro area HICP inflation excluding energy and food (HICPX inflation), a popular measure of underlying inflation. Relatedly, assessing the wage outlook may be helpful for monetary policy-makers to prevent the occurrence of the destabilizing wage-price spirals which were witnessed in the 1970s in many advanced economies and required very persistent and contractionary monetary policy actions to be reined in (see, for example [Lane, 2022](#)).

For these reasons, monitoring and forecasting wage developments is a key aspect of the economic analysis at central banks and other policy institutions ([Byrne and Zekaite, 2018](#); [Bobeica et al., 2019](#); [Paans and Vlekke, 2020](#); [Bonam et al., 2021](#); [Bank of England, 2024](#); [Gray et al., 2025](#); [Heise et al., 2026](#)). For example, recently, wage trackers that incorporate news on wage bargaining in a timely fashion (see, for example, [Górnicka et al., 2024](#); [Bates et al., 2024](#)) or scrape wage quotes from job-ads (see, for example, [Adrjan and Lydon, 2022](#)) have become rather popular for policy analysis. This paper contributes to the literature on wage forecasting. We focus on negotiated wages in France, but the methods described in the paper can be readily applied to other countries.

The models and the predictors typically used to describe and forecast wages are inspired by the wage Phillips Curve framework. Interestingly, the wage forecasting literature has already explored whether some forms of parametric non-linearity may be helpful for wage forecasting, with mixed results (see, for example [Bobeica et al., 2019](#)). The European Central Bank (ECB) has recently started to exploit Artificial Intelligence (AI) in its monetary policy briefing, including for inflation forecasting (see, for example, [Lenza et al., 2025](#)). One of the main results in [Lenza et al. \(2025\)](#) is that non-linearity characterizes in particular underlying inflation, suggesting that wage dynamics, which are closely related to underlying inflation, may also display non-linear dynamics. Against this background, this paper investigates how a particular machine learning model, the Local Linear Forests (LLFs) of [Bloniarz et al. \(2016\)](#) and [Friedberg et al. \(2021\)](#) can enrich forecasting models and improve predictions for negotiated wages in France, over the last four decades.

[Giannone et al. \(2021\)](#) addresses the question whether economic predictive problems are more likely characterized by sparsity or density, and concludes that the posterior distribution in such problems typically concentrates on a wide set of models that often include many predictors. Hence, linear ridge regression models generally provide very accurate economic predictions. The local linear forest of [Friedberg et al. \(2021\)](#), the variant we use in this paper, is essentially a weighted ridge regression, in which the weights given to past observations when estimating the model are time-varying. The time-variation in the weights implies that the model can potentially capture

non-linear dynamics. The empirical success of the model rests on the ability of the methodology to capture meaningful time-variation in the weights.

The random forest of [Breiman \(2001\)](#) has gained a large popularity in the context of macroeconomic forecasting (see, for example [Medeiros et al., 2021](#); [Lenza et al., 2025](#); [Goulet Coulombe et al., 2022](#)). [Lin and Jeon \(2006\)](#) highlights how the predictions from random forests can be represented as weighted averages of past values of the target variable, where the weights are “state dependent”, in the sense that they depend on the value of the variables used to predict the target variable. For example, [Göbel et al. \(2025\)](#) use of this representation to interpret machine learning predictions, in terms of comparisons with similar historical periods in the past. The Local Linear Forest model of [Friedberg et al. \(2021\)](#) uses a Random Forest to create adaptive, observation-specific weights for each observation, then fits a ridge regression to all the weighted observations and use its outcomes to predict the target variable.

In our forecasting exercise, we compare the accuracy of the LLF predictions to three benchmarks. The first benchmark is the random forest, a special case of the local linear forest in which the slope coefficients of the ridge regression are constrained to be equal to zero. One of the main motivations to model local dynamics in the local linear forest is to better fit smooth signals compared to random forests. This feature also implies that, differently from random forests, the LLF can produce predictions that lie outside the observed historical range of the target variables. The comparison with the random forest should provide some insight whether the ability to better fit smooth signals of the local linear forest is an important feature for predicting wages in France.

The second benchmark is a plain-vanilla ridge regression, which is a special case of the local linear forest in which all observations are equally weighted and, hence, the model dynamics are assumed to be linear. The comparison with this model offers an insight on the relevance of non-linear dynamics.

Finally, we also compare all models with a benchmark of non-forecastability, the random walk model, which predicts future wage growth in France to be equal to the most recently recorded wage growth (See [Atkeson and Ohanian, 2001](#); [Stock and Watson, 2007](#); [Fischer et al., 2009](#); [Faust and Wright, 2013](#); [Banbura et al., 2024](#), for an analysis of the relevance of the RW for inflation forecasting). All the models are estimated on two databases, domestic (including only macroeconomic and financial variables for France) and the full database (which includes wages and other macroeconomic and financial variables for Germany and Italy). Labor markets in Europe are considered to be quite segmented and characterized by considerable cross-country heterogeneity. At the same time, both real and nominal variables remain quite aligned across euro area countries and we aim to assess whether common dynamics are so pervasive that foreign predictors may add valuable information to predict French wages beyond domestic predictors.¹

Our results show that virtually all models add significant information to the random walk predictions. The ridge regression confirms its excellent predictive performance, also for wages in France. At the same time, the LLF generally performs somewhat better than the ridge regression, sug-

¹[De Mol et al. \(2008\)](#) and [Bańbura et al. \(2015\)](#) discuss the link between ridge regressions, bayesian shrinkage and principal components regressions.

gesting that the ability to capture non-linear and, more precisely, local linear dynamics may be important for wage dynamics. The ability to capture smooth signals, especially at times in which wage growth displays historically unprecedented levels, is also a valuable feature, given that the local linear forest outperforms the random forest.

Finally, the predictions of French wages benefit from using foreign variables. This result suggest that despite the labor market segmentation often highlighted for the euro area(see, for example, [Bodnár and Le Roux, 2023](#); [Bates et al., 2025](#)), economic dynamics display common features among countries and the predictors from other euro area countries provide an important signal to predict the evolution of French wages. The structure of the paper is as follows. Section 2 presents the empirical framework adopted in this paper. Section 3 discusses the empirical results and their implications. Section 4 concludes.

2 Empirical Framework

We define wage levels as w_t , with growth rate over h quarters defined as $\Delta w_t^h = w_t/w_{t-h} - 1$. x_t is a set of potential predictors for wages, available at the quarterly frequency. p and k are, respectively, the lags of the target variable and the candidate predictors that we include in the model. More formally, we estimate:

$$\Delta w_t^h = m(\Delta w_{t-h}^1, \dots, \Delta w_{t-h-p}^1; x_{t-h}, \dots, x_{t-h-k}) + \epsilon_t$$

and project forward:

$$\hat{\Delta w}_{t+h}^h = m(\Delta w_t^1, \dots, \Delta w_{t-p}^1; x_t, \dots, x_{t-k})$$

where $m(\cdot)$ represents a generic forecasting model. We aim to capture quite general forms of non-linearity by resorting to machine learning techniques and our baseline model is the Local Linear Forest model.

2.1 Database

Our database is available at the quarterly frequency, over the sample 1985Q1 - 2025Q1, and our target variable is a measure of negotiated wages in France. The predictors consist in a set of macroeconomic indicators, including the unemployment rate, GDP, labor productivity, capacity utilization, the output gap, the unemployment gap, inflation and inflation expectations. Given the large commonality across European countries, we test whether foreign variables add predictive power by also considering models that include the same set of predictors for Germany and Italy. In particular, we consider two predictor configurations:

- **Domestic:** this database includes only French variables.
- **Full database:** Domestic database plus negotiated wages in Germany and Italy and (roughly) the same list of predictors for Germany and Italy as for France.

The Appendix [A](#) provides a full description of the database.

2.2 Local Linear Forests

Our baseline model is the local linear forest described in [Friedberg et al. \(2021\)](#). Local linear modelling is a set of non-parametric techniques that estimates complex, nonlinear trends by fitting weighted linear regression models to small, local subsets of data. The local linear forest is a special case of local linear modelling which uses a Random Forest to create adaptive, observation-specific weights for each prediction point and then fits a ridge regression to all the weighted training data.

Formally, defining as z_t the N-dimensional vector of predictors with $t = 1...T$, z_0 as the value of the predictors that we use to predict the target, Δw_t^h as our target variable at horizon h and $\alpha_t(z_0)$ the weights associated to the observations for the predictors at time t , the local linear forest predictions are based on the intercept ($\hat{\mu}(z_0)$) and slope ($\hat{\theta}(z_0)$) coefficients estimated in the following regression:

$$\begin{pmatrix} \hat{\mu}(x_0) \\ \hat{\theta}(x_0) \end{pmatrix} = \arg \min_{\mu, \theta} \left\{ \sum_{t=1}^{T-h} \alpha_t(x_0) \left(\Delta w_t^h - \mu(z_0) - (z_{t-h} - z_0)\theta(z_0) \right)^2 + \lambda \|\theta(z_0)\|_2^2 \right\} \quad (1)$$

The coefficient λ captures the intensity by which the ridge regression enforces the penalty. The larger the penalty coefficient, the stronger the intensity by which the regression coefficients are shrunk toward small values. The ridge regression allows us to handle the potential issue of overfitting, given that we generally estimate large models. Ridge regressions have proved to be particularly well suited, in general, for the prediction of economic variables (see, for example [Giannone et al., 2021](#)). For the specific case of the local linear forest, the risk of overfitting is magnified because the weighting scheme might assign meaningful weights only to a handful of observations. We fix λ to the value of 0.1. We have also run cross-validation to tune the parameters of the local linear forest, but we did not materially improve on our baseline results and, hence, we generally took the default values in the literature for the local linear and random forest. When estimating the local linear regression above, we standardize the predictors, as it is customary for penalized regressions.²

As mentioned above, the weights $\alpha_t(z_0)$ are an outcome of a random forest model ([Breiman, 2001](#)). The random forest is an ensemble method which combines the results from a certain (potentially large) number of non-linear models, called regression trees. A regression tree fits a specific target variable (wage inflation, in our case) by repeatedly splitting the sample of the potential predictors in different sub-samples. Once the final split is achieved, the predicted value of the target variable associated with a specific sub-sample is represented by the sample mean in that sub-sample, called “leaf”. If the random forest included only one tree, the $\alpha_t(z_0)$ ’s would be equal and non-zero only for observations that lie in the same final “leaf” as z_0 .³ Regression trees

²Local linear forests are implemented in the R package `grf`. Precisely, we used the `ll_regression_forest` routine.

³For each specific tree, the sub-sample splits in a regression tree are obtained through a process defined as

are simple models yet they tend to overfit, which makes them bad predicting tools. In the standard random forest applications, many “relatively” uncorrelated regression trees are built to maximize the advantages of combining them, via the following two steps. First, the observations from the original data are bootstrapped with replacement before constructing any new tree. Second, the splits are computed, at each node, only by looking at a randomly selected set of the predictors. The default choice for the size of the latter set, which we take in this paper, is to draw a third of the variables for each split. Finally, we set the number of combined regression trees, i.e. the size of the forest, to the default value of 500.⁴ Hence, in the end, our $\alpha_t(z_0)$ ’s are computed as the average weights across all regression trees corresponding to the value z_0 .

2.3 Benchmark models

Our benchmarking models are the Random Forest (RF), the standard Ridge Regression (RR) and the Random Walk (RW).

The Random Forest can be seen as a version of the model in equation 1, with the restriction that $\theta(z_0)$ and λ are equal to zero. In practice, if the RF model turns out to be more accurate than our baseline LLF, we would conclude that the linear local regression is not supported by the data. That would also suggest that the ability of the LLF to better fit smooth signals compared to the random forest, which also entails the ability to extrapolate out of the historical sample, is not a useful feature for the empirical application in this paper.⁵

The Ridge Regression benchmark in our paper is also a special case of equation 1, in which the $\alpha_t(x_0)$ are constant and equal over time. In practice, if the ridge regression turns out to be more accurate than the LLF, we would conclude that the data are unlikely to be characterized by non-linearity.⁶

The random walk is a popular benchmark of non-forecastability, according to which the change in wages expected in the future is equal to the most recent value observed in sample, i.e.:

$$\Delta \hat{w}_{t+h}^h = \Delta w_t^h$$

binary recursive partitioning, an iterative process that splits the data into partitions. The process continues until the splits achieve an improvement in terms of a statistical criterion, such as the mean squared error in the fit for wage inflation (our target variable) or, alternatively, until the splitting process hits a stopping rule.

⁴Probst et al. (2019) discusses the default specification choices for random forests.

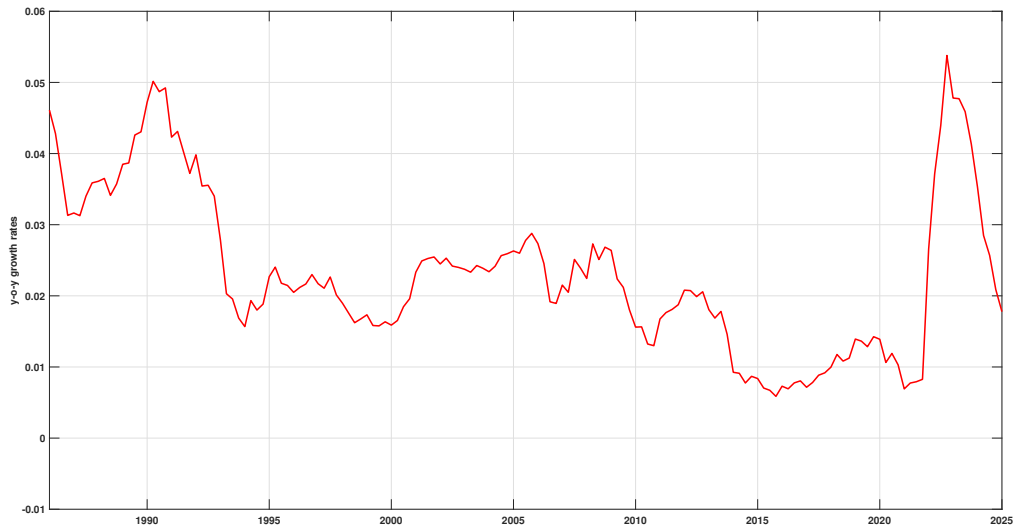
⁵We estimate the random forest by using the R package ranger.

⁶We estimate the ridge regression by using the R package glmnet.

3 Empirical Results

Figure 1 shows the year-on-year growth rates of negotiated wages in France, over the sample 1986Q1 to 2025Q1, which is the target variable in our study.

Figure 1: Negotiated wages in France, y-o-y growth rates



In our forecasting comparison, we take a recursive out-of-sample perspective. Our full sample ranges from 1985Q1 to 2025Q1. Our first estimation sample is 1985Q1-1998Q4. Over this sample, we transform and de-seasonalize the data, estimate the models and produce forecasts for horizons that range from 1 to 5 quarters ahead.⁷ Then, we update the sample by one quarter and repeat the exercise, until the exhaustion of the sample.

Table 1 reports the results of the forecasting evaluation. Each of the panels corresponds to one of the databases described in the data section. The top row in each of the panels reports the model associated with each column. The figures refer to the relative mean squared errors with respect to the random walk. A number smaller than one indicates that the model outperforms the random walk. The lower the figure, the more accurate the model performance. Notice that each panel has two LLF columns: LLF-All is a model in which all variables are always considered at each node in the RF model to determine the weights, while LLF-Small is a model in which we follow the more standard practice to assess splits only on the basis of a randomly chosen third of variables.

The first remark is that all the models generally outperform the random walk. This result validates the choice of variables used in this paper, which are also those regularly monitored by the wage

⁷We choose 1 to 5 quarters ahead because the horizon $h=1$ can be seen as a nowcast of the, say, 1999Q1 wage growth, given that the full set of data for the fourth quarter of 1998 are only released in the course of 1999Q1, in real-time. So, $h=5$, our longest horizon, corresponds to the one-year-ahead horizon.

Table 1: Relative MSE with respect to the RW model

	Full Database				Domestic			
	Ridge	Random Forest	LLF-All	LLF-Small	Ridge	Random Forest	LLF-All	LLF-Small
H=1	0.6815	0.6178	0.6361	0.6075	0.6970	0.7457	0.7501	0.7414
H=2	0.6512	0.6550	0.6199	0.6109	0.7627	0.8664	0.8229	0.7856
H=3	0.6762	0.7692	0.6500	0.6657	0.8085	0.9713	0.9559	0.8649
H=4	0.5959	0.7138	0.6086	0.5594	0.7106	0.8310	0.8110	0.7383
H=5	0.5988	0.6769	0.5973	0.5302	0.6914	0.7571	0.7727	0.6924

Note: A number smaller than one indicates that the model (columns) outperforms the random walk for the specific horizon (rows). Bold font indicates the best models.

experts in the Eurosystem economic analysis. On the other hand, the models in Table 1 are among the most popular for macroeconomic forecasting.

The random forest and, even more, the ridge regression are strong benchmarks, which are not easy to outperform. However, the best performing model is LLF-Small, our baseline model, estimated on the Full Database and especially for the longer forecasting horizons.

These results have three main implications. First, as in [Lenza et al. \(2025\)](#) for inflation, we can conclude that it is important to capture non-linear dynamics to predict wage growth.

Second, the LLF model outperforms the RF, which shows the importance of fitting local linear processes. To visualize this point, figure 2 reports the forecasts from the LLF-Small, RF and RR models for the horizon of $h=3$ and $h=5$.

Comparing the forecasts from the leftmost and middle panels, it is clear how the LLF is especially more accurate than the RF in the low inflation/wage period post-Great Financial crisis. As shown in figure 1, wage growth in that period reached levels which had never been witnessed in the whole sample at our disposal. Hence, the ability to fit smooth signals by capturing local linear dynamics is crucial, especially at times in which wage growth gradually reaches unprecedented levels.

Finally, the Full Database provides, rather uniformly, the best results across models and, hence, considering also variables from Germany and Italy is helpful to predict wages in France. This reflects the fact that economic dynamics in the euro area are quite common across countries and including German and Italian variables helps the models to better extract the common signals in the data (on this point, see [De Mol et al., 2008](#); [Bańbura et al., 2015](#)). Figure 3 reports the forecasts from the LLF-Small, using the domestic and the full databases for the horizon of $h=3$ and $h=5$, with those based on the full database being visibly closer to actual wage growth than those based on the domestic database.

Figure 2: Comparison among LLF, RF and Ridge

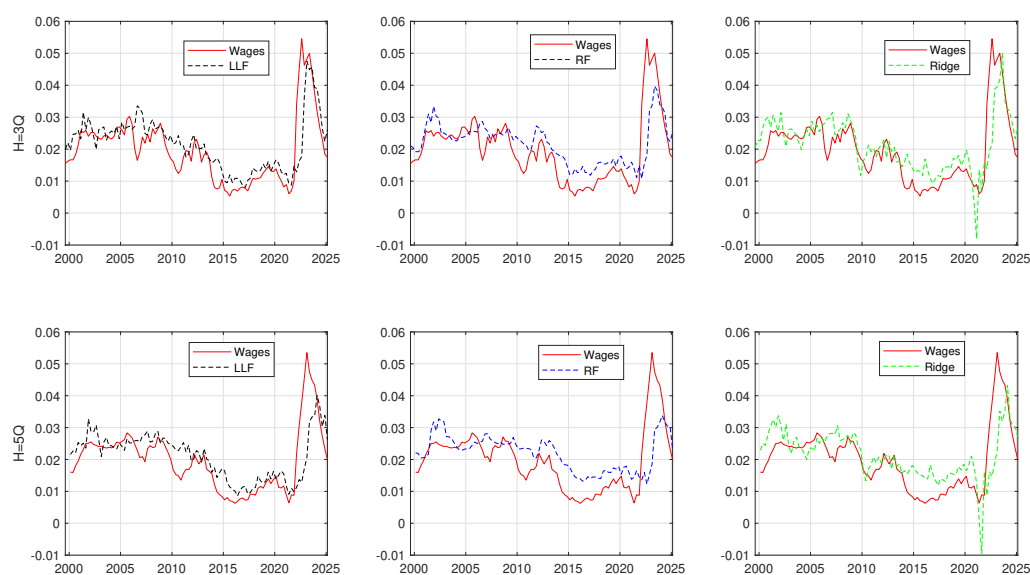
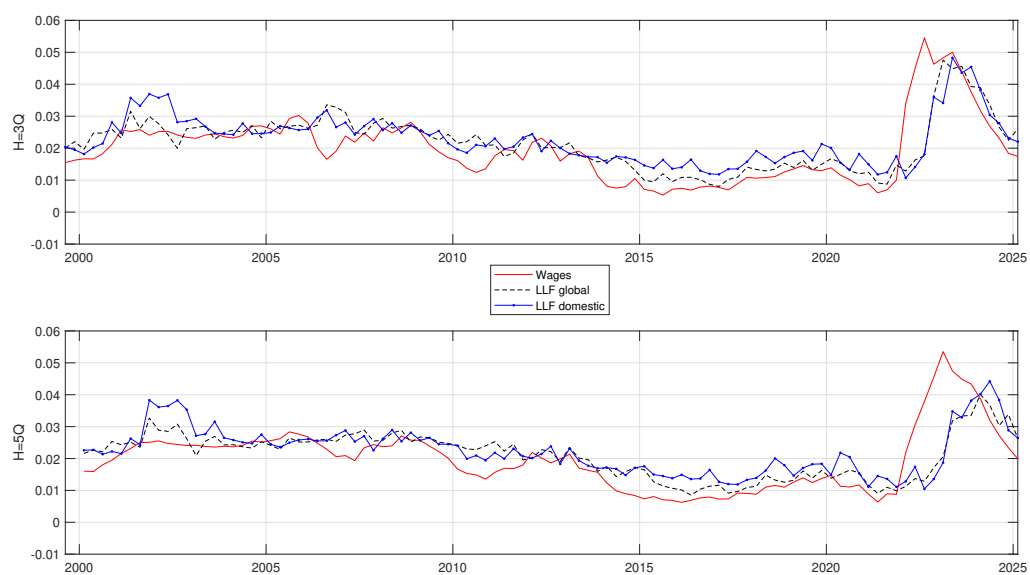


Figure 3: Comparison among full and domestic database



4 Conclusions

LLFs effectively forecast wages, outperforming traditional methods like RW and RF. Their default specification, without having to tune the hyperparameters, yields robust results.

The results also suggest that foreign information is important to predict wages in France and that wage dynamics are characterized by non-linearity, at least to a certain extent.

A Database

The sample in this paper ranges from 1985Q1 to 2025Q1. Table 2 reports the variables used in the model, and describes their allocation in the three databases (column 5). In particular, the domestic database includes all variables labelled with Domestic. The two variables labelled as Foreign Wages are included, together with the Domestic ones in the Foreign Wages database. The variables labelled as Full Database are included, together with the variables in the Foreign Wages database, in the Full Database. FR: France; IT: Italy and DE: Germany.

Table 2: Model Variables and Databases

Definition	Simple Label	Transf.	Country	Database
Indicator of negotiated wage rates, Total, percentage change - transformed into index *	Negotiated wages index	log diff	FR	Domestic
Gross domestic product at market prices, volume, s.w.d.a, Millions of Euro	Real GDP	log diff	FR	Domestic
Unemployment rate, age 15 to 74, total, s.a., percent	Unemployment rate	levels	FR	Domestic
Labour Productivity (per persons), index, s.a	Labour Productivity	log diff	FR	Domestic
Productivity per hours worked (ratio between real GDP and All activities, Employees Hours worked, s.w.d.a)	Productivity per hours	log diff	FR	Domestic
Consumer survey - Price trends over next 12 months - % balances, s.a	Consumer Expectations	levels	FR	Domestic
HICP - Overall index, backdated (2015 = 100)	HICP	log diff	FR	Domestic
HICP - All-items excluding energy and food (2015 = 100)	HICP ex Energy and Food	log diff	FR	Domestic
Gross domestic product at market prices, deflator, s.w.d.a.	GDP Deflator	log diff	FR	Domestic
Euro area capacity utilisation in manufacturing (%), s.a.	Capacity Utilisation	log diff	FR	Domestic
France - Output gap, Percent	WEO (IMF) Output Gap	levels	FR	Domestic
Gap between actual and potential gross domestic product at 2020 reference levels, Pct. of real potential GDP	EC (AMECO) Output gap	levels	FR	Domestic
Unemployment Gap (difference between unemployment rate and NAIRU)	EC (AMECO) Unemployment gap	levels	FR	Domestic
Indicator of negotiated wage rates, Total, percentage change - transformed into index*	Negotiated wages index	log diff	IT	Full Database
Unemployment rate, age 15 to 74, total, s.a., percent	Unemployment rate	levels	IT	Full Database
Gross domestic product at market prices, volume, s.w.d.a, Millions of Euro	Real GDP	log diff	IT	Full Database
Labour Productivity (per persons), index, s.w.d.a	Labour Productivity	log diff	IT	Full Database
Consumer survey - Price trends over next 12 months - % balances, s.a	Consumer Expectations	levels	IT	Full Database
HICP - Overall index, backdated (2015 = 100)	HICP	log diff	IT	Full Database
HICP - All-items excluding energy and food (2015 = 100)	HICP ex Energy and Food	log diff	IT	Full Database
Gross domestic product at market prices, deflator, s.w.d.a.	GDP Deflator	log diff	IT	Full Database
Euro area capacity utilisation in manufacturing (%), s.a.	Capacity Utilisation	log diff	IT	Full Database
Italy - Output gap, Percent	WEO (IMF) Output Gap	levels	IT	Full Database

Continued on next page

Table 2 – *Continued from previous page*

Definition	Simple Label	Transf.	Country	Database
Gap between actual and potential gross domestic product at 2020 reference levels, Pct. of real potential GDP	EC (AMECO) Output gap	levels	IT	Full Database
Unemployment Gap (difference between unemployment rate and NAIRU)	EC (AMECO) Unemployment gap	levels	IT	Full Database
Basic pay rates, overall economy, excluding ancillary benefits, excluding one-off payments, on a monthly basis, Germany -2020=100	Negotiated wages index	log diff	DE	Full Database
Gross domestic product at market prices, volume, s.w.d.a, Millions of Euro	Real GDP	log diff	DE	Full Database
Unemployment rate, age 15 to 74, total, s.a., percent	Unemployment rate	levels	DE	Full Database
Labour Productivity (per persons), index, s.w.d.a	Labour Productivity	log diff	DE	Full Database
Consumer survey - Price trends over next 12 months - % balances, s.a	Consumer Expectations	levels	DE	Full Database
HICP - Overall index, backdated (2015 = 100)	HICP	log diff	DE	Full Database
HICP - All-items excluding energy and food (2015 = 100)	HICP ex Energy and Food	log diff	DE	Full Database
Gross domestic product at market prices, deflator, s.w.d.a.	GDP Deflator	log diff	DE	Full Database
Euro area capacity utilisation in manufacturing (%), s.a.	Capacity Utilisation	log diff	DE	Full Database
Germany - Output gap, Percent	WEO (IMF) Output Gap	levels	DE	Full Database

References

- ADRIAN, P. AND R. LYDON (2022): “Wage Growth in Europe: Evidence From Job Ads,” *Economic Letters* 7/EL/22, Central Bank of Ireland.
- ATKESON, A. AND L. E. OHANIAN (2001): “Are Phillips curves useful for forecasting inflation?” *Quarterly Review*, 25, 2–11.
- BAÑBURA, M., D. GIANNONE, AND M. LENZA (2015): “Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections,” *International Journal of forecasting*, 31, 739–756.
- BANBURA, M., M. LENZA, AND J. PAREDES (2024): “Forecasting inflation in the US and in the euro area,” *Handbook of Macroeconomic Forecasting*, forthcoming.
- BANK OF ENGLAND (2024): “How Does the Monetary Policy Committee Forecast Wage Growth?” Bank Overground, accessed: 2024.
- BATES, C., K. BODNÁR, V. BOTELHO, AND F. ROUSSEAU (2025): “Real wage catch-up in the euro area,” *Economic Bulletin Boxes*, 5, None.
- BATES, C., V. BOTELHO, S. HOLTON, M. ROCA I LLEVADOT, AND M. STANISLAO (2024): “The ECB wage tracker: your guide to euro area wage developments,” *ECB Blog*.
- BLONIARZ, A., A. TALWALKAR, B. YU, AND C. WU (2016): “Supervised neighborhoods for distributed nonparametric regression,” in *Artificial Intelligence and Statistics*, PMLR, 1450–1459.
- BOBEICA, E., G. KOESTER, E. LIS, C. NICKEL, AND M. PORQUEDDU (2019): “Understanding low wage growth in the euro area and European countries,” Occasional Paper Series 232, European Central Bank.
- BODNÁR, K. AND J. LE ROUX (2023): “Wage dynamics across euro area countries since the start of the pandemic,” *Economic Bulletin Boxes*, 8, None.
- BONAM, D., J. DE HAAN, AND D. VAN LIMBERGEN (2021): “Time-varying wage Phillips curves in the euro area with a new measure for labor market slack,” *Economic Modelling*, 96, 157–171.
- BREIMAN, L. (2001): “Random forests,” *Machine learning*, 45, 5–32.
- BYRNE, D. AND Z. ZEKAITE (2018): “Missing wage growth in the euro area: is the wage Philips curve non-linear?” *Economic Letters* 9/EL/18, Central Bank of Ireland.
- DE MOL, C., D. GIANNONE, AND L. REICHLIN (2008): “Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?” *Journal of Econometrics*, 146, 318–328.
- FAGANDINI, B., E. GONÇALVES, I. RUBENE, O. KOUVAVAS, K. BODNÁR, AND G. KOESTER (2024): “Decomposing HICPX inflation into energy-sensitive and wage-sensitive items,” *Economic Bulletin Boxes*, 3, None.
- FAUST, J. AND J. H. WRIGHT (2013): “Chapter 1 - Forecasting Inflation,” in *Handbook of Economic Forecasting*, ed. by G. Elliott and A. Timmermann, Elsevier, vol. 2 of *Handbook of Economic Forecasting*, 2–56.
- FISCHER, B., M. LENZA, H. PILL, AND L. REICHLIN (2009): “Monetary analysis and monetary policy in the euro area 1999-2006,” *Journal of International Money and Finance*, 28, 1138–1164.
- FRIEDBERG, R., J. TIBSHIRANI, S. ATHEY, AND S. WAGER (2021): “Local Linear Forests,” *Journal of Computational and Graphical Statistics*, 30, 503–517.
- GIANNONE, D., M. LENZA, AND G. E. PRIMICERI (2021): “Economic Predictions With Big Data: The Illusion of Sparsity,” *Econometrica*, 89, 2409–2437.
- GÖBEL, M., P. GOULET-COULOMBE, AND K. KLIBER (2025): “Dual Interpretation of Machine Learning Forecasts,” Working Papers 265, Oesterreichische Nationalbank (Austrian Central Bank).
- GOULET COULOMBE, P., M. LEROUX, D. STEVANOVIC, AND S. SURPRENANT (2022): “How is machine learning useful for macroeconomic forecasting?” *Journal of Applied Econometrics*, 37, 920–964.
- GRAY, N., F. LATTIMORE, K. MCLOUGHLIN, AND C. WINDSOR (2025): “An AI-powered Tool for Central Bank Business Liaisons: Quantitative Indicators and On-demand Insights from Firms,” .

- GÓRNICKA, L., G. KOESTER, D. RADOWSKI, E. GAUTIER, M. I. PEINADO, A. STIGLBAUER, D. WITTEKOPF, S. PUENTE, C. F. DUARTE, AND F. MARTINS (2024): “A forward-looking tracker of negotiated wages in the euro area,” Occasional Paper Series 338, European Central Bank.
- HEISE, S., J. PEARCE, AND J. P. WEBER (2026): “Wage growth and labor market tightness,” *Journal of Monetary Economics*, 158, 103903.
- LANE, P. (2022): “Inflation Diagnostics,” *ECB Blog*.
- LENZA, M., I. MOUTACHAKER, AND J. PAREDES (2025): “Density forecasts of inflation: A quantile regression forest approach,” *European Economic Review*, 178.
- LIN, Y. AND Y. JEON (2006): “Random Forests and Adaptive Nearest Neighbors,” *Journal of the American Statistical Association*, 101, 578–590.
- MEDEIROS, M. C., G. F. R. VASCONCELOS, ÁLVARO VEIGA, AND E. ZILBERMAN (2021): “Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods,” *Journal of Business & Economic Statistics*, 39, 98–119.
- PAANS, J. AND M. VLEKKE (2020): “Forecasting Wage Growth and Price Inflation in the Netherlands with a BVAR Model,” Cpb background document, CPB Netherlands Bureau for Economic Policy Analysis.
- PROBST, P., M. N. WRIGHT, AND A.-L. BOULESTEIX (2019): “Hyperparameters and tuning strategies for random forest,” *WIREs Data Mining and Knowledge Discovery*, 9, e1301.
- STOCK, J. H. AND M. W. WATSON (2007): “Why Has U.S. Inflation Become Harder to Forecast?” *Journal of Money, Credit and Banking*, 39, 3–33.

Acknowledgements

The authors thank Clara Augustin for comments. All remaining errors are our own.

The views expressed are those of the authors and do not necessarily reflect those of the European Central Bank (ECB), the Eurosystem and the CEPR.

Michele Lenza

European Central Bank, Frankfurt am Main, Germany; Centre for Economic Policy Research, London, United Kingdom;
email: michele.lenza@ecb.europa.eu

Claudia Marchini

European Central Bank, Frankfurt am Main, Germany; email: claudia.marchini@ecb.europa.eu

© European Central Bank, 2026

Postal address 60640 Frankfurt am Main, Germany

Telephone +49 69 1344 0

Website www.ecb.europa.eu

All rights reserved. Any reproduction, publication and reprint in the form of a different publication, whether printed or produced electronically, in whole or in part, is permitted only with the explicit written authorisation of the ECB or the authors.

This paper can be downloaded without charge from www.ecb.europa.eu, from the [Social Science Research Network electronic library](#) or from [RePEc: Research Papers in Economics](#). Information on all of the papers published in the ECB Working Paper Series can be found on the [ECB's website](#).

PDF

ISBN 978-92-899-8027-2

ISSN 1725-2806

doi:10.2866/3631049

QB-01-26-229-EN-N