Università degli Studi del Molise
Dipartimento di Economia

Economics and Statistics Discussion Paper
No. 101/25

Predicting Regional Unemployment in the EU

Elena Paglialunga
Giuliano Resce
Angela Zanoni

(This page intentionally left blank.)

Predicting Regional Unemployment in the EU

Elena Paglialunga[a], Giuliano Resce[b], Angela Zanoni[cde]

[a] Department of Economics, University Roma Tre

[b] Department of Economics, University of Molise

[c] Department of Economics and Law, Sapienza, University of Rome

[d] Research Institute for Sustainable Economic Growth, National Research Council, Italy

[e] Sustainable Mobility Program, Fondazione Eni Enrico Mattei, Italy

Abstract

This paper predicts regional unemployment in the European Union by applying machine learning techniques to a dataset covering 198 NUTS-2 regions, 2000 to 2019. Tree-based models substantially outperform traditional regression approaches for this task, while accommodating reinforcement effects and spatial spillovers as determinants of regional labor market outcomes. Inflation—particularly energy-related—emerges as a critical predictor, highlighting vulnerabilities to energy shocks and green transition policies. Environmental policy stringency and eco-innovation capacity also prove significant. Our findings demonstrate the potential of machine learning to support proactive, place-sensitive interventions that aim to predict and mitigate the uneven social and economic impacts of structural change across regions.

Keywords: Regional unemployment; Inflation; Environmental policy; Spatial spillovers; Machine learning.

JEL Classification: E24; J64; Q52; R23.

1 Introduction

Unemployment is a central concern for European policymakers, particularly at the regional level, where economic shocks and structural imbalances produce persistent spatial disparities [Rodríguez-Pose, 2013]. These differences not only reflect unequal access to employment opportunities but also challenge the effectiveness of cohesion policy, social protection systems, and the political legitimacy of European integration [Dijkstra et al., 2015]. Anticipating regional unemployment trends is therefore essential for the timely design and targeting of active labor market policies [Boeri and Terrell, 2002], especially in light of recent crises that have revealed the vulnerability of specific regions to economic downturns [Bagliano and Morana, 2025, Iammarino et al., 2019]. These vulnerabilities are further exposed in the context of the green transition, where ambitious climate and energy policies, while crucial for decarbonisation, may introduce challenges linked to uneven sectoral exposures, regional labour market resilience, and the distributive impacts of higher energy and production costs [Vona, 2023].

This paper develops a predictive framework to anticipate short-term unemployment rates across NUTS2 regions in Europe. Framing the task as a prediction policy problem [Kleinberg et al., 2015], it becomes evident the practical value of forecasting for policy. In particular, timely and reliable predictions can help identify regions most at risk of rising unemployment, thereby supporting the targeting of place-based interventions such as cohesion policies, where the effectiveness of action depends critically on anticipating rather than merely reacting to shocks.

Traditional econometric models used in labor economics, such as fixed-effects panel regressions or structural macro models, often prioritize inference over prediction and tend to perform poorly in out-of-sample forecasting tasks [Stock and Watson, 2002]. In contrast, machine learning (ML) techniques are explicitly designed to maximize predictive accuracy and can flexibly capture complex, non-linear relationships and high-dimensional interactions among variables [Athey and Imbens, 2019]. Recent research has demonstrated the value of ML methods across a variety of policy domains, ranging from criminal justice to social welfare. In particular, these techniques are increasingly being applied to address challenges such as poverty targeting [Jean et al., 2016], evaluating the effectiveness of public programs and spending [Andini et al., 2018, Christensen et al., 2024, Di Stefano and Resce, 2025], analyzing education and training systems [Sansone, 2019, Sansone and Zhu, 2023], and identifying instances of financial distress, corruption, and political connections [Antulov-Fantulin et al., 2021, Titl et al., 2024, D'Ecclesiis et al., 2025]. They have also proven useful in the analysis of environmental and health-related issues [Fabra et al., 2022, Girma and Paton, 2024, Carrieri et al., 2021, Caravaggio et al., 2025]. In the context of labor market analysis, recent studies have applied ML to forecast aggregate unemployment in the euro area [Gogas et al., 2022] and in countries like Germany [Katris, 2020]. However, these approaches often overlook subnational disparities—an aspect we aim to address more directly.

We apply a set of ML models to forecast unemployment rates at the NUTS-2 level, using a rich panel dataset that collects a variety of regional socioeconomic indicators. In line with recent literature, we also incorporate environmental variables to capture potential trade-offs between green transition policies and labor market outcomes. We include spatial lags to account for geographic spillovers, following studies on the spatial dimension of unemployment. Additionally, we include inflation-related indicators to reflect the well-established unemployment-inflation relationship. Our level of spatial granularity is particularly valuable, as NUTS-2 regions are a key target of European cohesion policy and the allocation of related funding instruments. Our models has the potential to support proactive policy design by identifying which regions are likely to experience rising unemployment in the near future—providing early-warning signals that can inform the deployment of European Social Fund resources, the design of retraining programs, and the planning of counter-cyclical investment initiatives.

Our findings show that ML models—particularly Random Forest and Gradient Boosting—effectively capture the complex, non-linear dynamics behind regional unemployment in the EU. Key predictors include lagged unemployment and its spatial lag (*i.e.*, the unemployment rates of neighboring regions), highlighting the relevance of geographical concentration effects in shaping unemployment patterns. Environmental policy and environmental performance also emerge as important predictors. Inflation-related variables retain notable influence, while challenging the validity of the Phillips relationship in local contexts.

The remainder of the paper is structured as follows: Section 2 reviews the relevant literature, Section 3 presents our methodology and data, Section 4 discusses the empirical results, and Section 5 concludes with policy implications and directions for future research.

2 Literature Review

2.1 The unemployment-inflation relationship in the green transition era

Unemployment dynamics are strongly shaped by macroeconomic factors such as economic growth and price fluctuations, with uneven consequences that amplify regional disparities. Shocks to real disposable income, whether from energy and food price increases, or broader economic crises, directly affect consumption capacity and firms' labour demand, thereby reinforcing spatial disparities in unemployment outcomes [Rodríguez-Pose, 2013, Iammarino et al., 2019]. These vulnerabilities have been repeatedly exposed during recent crises, from the economic downturn to the pandemic and the current energy price shock, where weak income growth combined with inflationary pressures exacerbate regional labour market fragilities.

Within this broader macroeconomic context, the Phillips Curve provides a classical framework for examining the inflation–unemployment nexus, originally formulated as an inverse relationship between change of monetary wages and unemployment [Phillips, 1958]. Over time, empirical studies have provided mixed evidence regarding the stability of this relationship. Some research suggests that the Phillips Curve has flattened in recent decades due to more stable inflation expectations [Stock and Watson, 2008], while others argue that regional variations and structural breaks must be considered to better understand its empirical validity [Babb and Detmeister, 2017, Hooper et al., 2020, Smith, 2024]. Despite these issues, renewed interest has emerged, especially in light of contemporary economic conditions and recent inflationary episodes observed in the last years [Hazell et al., 2022]. A substantial body of literature is concerned with empirically assessing the occurrence of the curve. One line of research focuses on the re-emergence of empirical support for the Phillips hypothesis, suggesting that inflation-unemployment trade-offs remain relevant [Hazell et al., 2022]. Other studies emphasize the role of stable inflation expectations in reducing the sensitivity of inflation to unemployment [Stock and Watson, 2008]. Further research underscores the heterogeneity in Phillips Curve coefficients across regions [Babb and Detmeister, 2017, Hooper et al., 2020], while additional contributions examine structural breaks that arise due to economic shifts [Smith, 2024]. Despite these findings, key unresolved issues persist, including the potential nonlinearity of the Phillips relationship [Babb and Detmeister, 2017], the challenge of disentangling the direction of causality between inflation and employment [Hazell et al., 2022], and the endogeneity concerns in estimating the relationship [Smith, 2024].

Recent global events have underscored two critical considerations: the imperative to accelerate energy transition for achieving strategic autonomy, and the adverse macroeconomic consequences associated with abrupt shifts in energy mix. Indeed, climate and energy policies designed to reduce carbon dependency and accelerate the shift to renewable sources may entail short-term increases in energy prices, whether through carbon taxation, the phasing out of subsidies, or the costs of renewable integration. While essential for long-term sustainability, these measures may introduce new inflationary pressures that interact with labour markets in non-trivial ways. On the one hand, higher energy and input costs raise firms' production costs, potentially leading to lower output and weaker labour demand. On the other hand, rising consumer prices erode households' real disposable income, reducing consumption and further depressing economic activity. Together, these dynamics undermine the classical Phillips trade-off and can give rise to episodes in which inflation and unemployment may rise simultaneously, underscoring the importance of incorporating both price and income dynamics into the analysis of unemployment. These developments have sparked interest in examining the nexus between climate performance and inflationary pressures [Boneva and Ferrucci, 2022, Moessner, 2022].

In this light, the inflation–unemployment relationship cannot be interpreted solely as a matter of monetary expectations, as it increasingly reflects also how regions absorb the distributive consequences of the energy transition, with uneven effects across space and sectors. This perspective underscores the importance of accounting for climate and energy policy shocks in empirical analysis, and highlights the need for regionally tailored policies that can mitigate labour market vulnerabilities while sustaining progress towards decarbonisation.

2.2 Environmental policies and unemployment

Even though the green transition may not have sizeable effects on aggregate employment in the EU, it certainly introduces complexities to local labor markets, producing both job creation and destruction, and causing labor reallocation [Vandeplas et al., 2022]. While some studies highlight the potential for green jobs to drive employment growth [Vona, 2023, Marin and Vona, 2019], others underscore the risks associated with the decline of traditional energy-intensive industries [OECD, 2012]. Displaced workers often encounter significant barriers to reemployment, as green jobs [Consoli et al., 2016] may demand distinct skill sets, potentially leading to structural unemployment in affected regions. Moreover, the development of new and sustainable technological trajectories can have unintended consequences, particularly for low-skilled workers, who may face adverse employment effects as industries adapt to environmental regulations [Consoli et al., 2016].

Rodríguez-Pose and Bartalucci [2024] identify unemployment as an indicator of territorial discontent toward the green transition process. The effects of environmental policies on labor markets depend on local factors. Stringent regulations can alter sectoral compositions, leading to job creation in some areas while causing job losses elsewhere, particularly in regions dependent on carbon-intensive industries [Vona, 2023]. The local labor market response is shaped by reconversion and migration costs, wealth distribution, and broader economic conditions. Evidence suggests that the impact of green transition policies varies based on labor mobility, the pace of green innovation and technological adoption [Costantini et al., 2018], existing income inequalities [Napolitano et al., 2022], market concentration and firm size distribution [Zaman and Borsky, 2021], and the evolving demand for green skills [Marin and Vona, 2019]. These factors collectively determine the extent to which a shift to more sustainable productive systems can lead to an increase in unemployment.

2.3 The spatial dimension of unemployment

Since our analysis investigates labour market dynamics at the subnational level, the spatial dimension of unemployment becomes particularly salient. Indeed, economic activity is highly spatially concentrated in most countries, and this structure drives significant disparities in labour market outcomes. Regional unemployment is thus strongly shaped by local structural conditions, including sectoral specialization, innovation capacity, demographic composition, and institutional quality [Iammarino et al., 2019, Kline and Moretti, 2013]. Yet, these factors rarely operate in isolation. Core regions often benefit from agglomeration economies, such as better access to capital, denser labor markets, and knowledge spillovers, while peripheral or less connected areas may face persistent disadvantages due to structural weaknesses, low innovation uptake, or poor infrastructure [Basile, 2009, Moretti, 2012]. As a consequence, unemployment rates vary markedly across cities and regions. These spatial disparities are not merely reflections of individual characteristics but arise from geographically bounded dynamics. As documented in the literature on local labor markets, agglomeration effects, labor mobility, economic specialization, and the spatial sorting of workers and firms tend to reinforce regional divergence in wages and employment outcomes [Greenstone et al., 2010, Kline and Moretti, 2014, Moretti, 2011b,a, 2012]. High-productivity regions attract capital and skilled labor, setting in motion virtuous cycles of employment growth. In contrast, regions with lower innovation capacity or overexposure to declining industries, including brown sectors, face the dual challenge of economic stagnation and outmigration [Iammarino et al., 2019]. Importantly, economic shocks, whether local (e.g., plant closures, automation) or global (e.g., energy price volatility, climate policy reforms), rarely remain confined within administrative borders. They propagate across space via commuting flows, supply chains, and migration patterns, giving rise to significant spillover effects [Moretti, 2011a]. These interdependencies challenge the assumption of independent regional labor markets and highlight the importance of modeling spatial interactions explicitly [Niebuhr, 2003]. Evidence from the Italian context Cracolici et al. [2007] shows that both high- and low-unemployment areas exhibit spatial clustering, confirming that labor market outcomes are subject to spatial autocorrelation. In particular, regions in Southern Italy have persistently high unemployment, often linked to structural disadvantages and weak integration with innovation-intensive northern areas. This pattern is mirrored across many EU countries and is closely tied to the legacy of path-dependent regional development trajectories, lock-in effects, further compounded by differences in innovation adoption, regional variety and firm dynamism [Corradini and Vanino, 2022, Overman and Puga, 2002]. Such disparities have prompted policymakers to adopt place-based development strategies aimed at reducing territorial inequality. In the EU, Cohesion Policy explicitly targets lagging regions, recognizing the need for tailored interventions that address spatially concentrated disadvantages to foster regional resilience. Empirical evidence supports this approach: Caro and Fratesi [2023] found that Cohesion Policy bolsters labour market resilience during economic downturns. These concerns are particularly relevant in the context of the green transition since climate and environmental policies, such as carbon taxation or the phasing out of polluting industries, are expected to have asymmetric effects across space [Coenen et al., 2012]. Regions with high concentrations of brown-sector employment may face significant labor displacement, while innovation hubs and urban centers stand to benefit from green job creation and investment flows [Rodríguez-Pose and Bartalucci, 2024, Vona, 2023].

This emerging research trajectory establishes important foundations for both theoretical frameworks and empirical investigations exploring the interconnections among labor market dynamics, sustainable economic transitions, and price stability.

3 Data and methods

3.1 Data

Our dataset incorporates regional data for EU NUTS 2 regions, supplemented with national data on policy settings, over a 20-year time frame (2000 to 2019). The dataset includes 38 variables (see Table 1 for a detailed description of variables' sources and codification in our dataset). To preserve as much information as possible, breaks in time series are interpolated. When observations for region i still present missing values, the region is dropped to ensure a balanced dataset. This operation leaves us with a panel of 198 regions out of the original 256: Figure 1 shows the regions included in the dataset.

The regional unemployment rate, expressed as a percentage of the labour force has been widely used to measure the societal impacts of socio-technical transitions [Castellanos and Heutel, 2019, Vona, 2021]. The features considered for the prediction task are inspired by the literature review in Section 2. To assess regional economic disparities and potential development constraints related to climate policy implementation, we employ gross domestic product at the NUTS 2 level, expressed in million euros, as our primary indicator of overall economic development

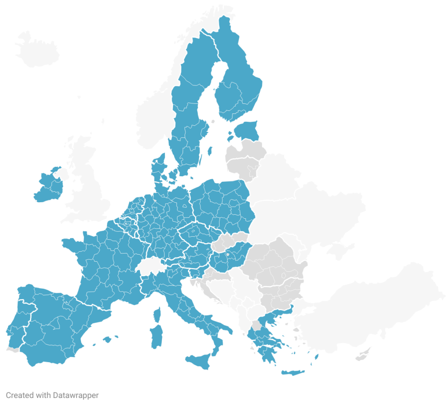

Figure 1: Map of NUTS2 regions included in the panel

[Diemer et al., 2022, Iammarino et al., 2019]. For examining household-level impacts of climate policies, we utilize disposable income data, given its direct connection to consumption patterns and behaviors [Rodríguez-Pose and Bartalucci, 2023]. As an indicator of systemic inclusivity and intersectionality within socio-economic frameworks, we incorporate the gender employment gap across NUTS 2 regions, expressed as a percentage. This approach aligns with established research utilizing gender metrics as proxies for broader systemic inclusivity [Johnson et al., 2020]. The importance of gender representation in climate governance for achieving equitable transitions is well-documented [Kronsell, 2013], particularly given evidence of continued male dominance in green employment sectors [Maldonado et al., 2024]. Considering gender employment balance is even more relevant to our research objectives as Borsekova and Korony [2023] see gender balance as a key dimension of regional labour market efficiency.

For capturing regional innovative capacity, we utilize patent data as a standard indicator [Barbieri et al., 2023]. Specifically, we employ green patent counts to measure eco-innovative potential, which research consistently links to successful environmental transitions [Costantini et al., 2018, Napolitano et al., 2022]. Our patent analysis draws from the OECD REGPAT Database, which compiles patent application data from both the European Patent Office (EPO) and Patent Cooperation Treaty (PCT) frameworks, though we focus exclusively on EPO applications. Using the Spring 2023 dataset version, we extract technological classification information through cooperative patent classification (CPC) codes. Patent applications are allocated to NUTS 2 regions based on inventor location, with our count including applications with priority years spanning 2000-2019. To identify eco-innovations specifically, we filter applications containing CPC code Y02, which designates inventions focused on climate change mitigation and adaptation.

Recognizing that patent metrics alone may incompletely represent regional innovation ecosystems, we incorporate additional knowledge capital indicators from the Regional Innovation Scoreboard (RIS), particularly given research emphasizing R&D support's role in climate preparedness [Cappellano et al., 2024]. From the available RIS metrics, we select three key variables: employment within technology and knowledge-intensive industries, educational participation rates, and tertiary education completion rates for the 25-64 age cohort. While the latter two metrics are already expressed as percentages, we convert the employment data to relative terms by calculating per-capita values using regional population figures.

We also account for regional vulnerability through employment concentration in transition-sensitive sectors. Research consistently identifies dependence on specific productive sectors as a critical factor of vulnerability to the distributional impacts of climate policies [Rodríguez-Pose and Bartalucci, 2023, Vandeplas et al., 2022, Vona, 2021]. To account for this strand of research, our analysis includes employment data for NACE sectors A through F at the NUTS 2 level, measured in thousands of workers.[1]

To control for demographic factors, we include regional median age as an additional variable capturing population structure dynamics. For environmental performance measurement, we utilize greenhouse gas emission indicators, which, while not comprehensively capturing all environmental impacts of human activities, remain widely accepted metrics for evaluating green transition progress and sustainable development achievements [Bianchini et al., 2023]. The EDGAR database provides harmonized emission estimates across multiple pollutants by integrating data from various sources. At the NUTS 2 level, we consider carbon dioxide (CO2), methane (CH4), nitrous oxide (N2O), and fluorinated gases (F-gases), all expressed in kilotons of CO2 equivalent. F-gases are potent greenhouse gases primarily emitted from highly polluting industrial activities such as refrigeration, electronics and semiconductor manufacturing, and some specialized chemical processes, including the production of magnesium and aluminum. [Crippa et al., 2023].

Following recent geopolitical developments that have highlighted both the urgency of energy transition for strategic independence and the potential macroeconomic disruptions from rapid energy mix changes, the relationship between climate performance and inflation has gained renewed attention [Boneva and Ferrucci, 2022, Moessner, 2022]. Inflation dynamics are examined by incorporating the Harmonized Index of Consumer Prices (HICP) and the HICP for energy products to account for the differential impact of energy inflation compared to global inflation [Moessner, 2022]. To capture regional income effects and their relationship with price changes, we include a new variable *rate_disp*, which represents the rate of change in disposable income at the regional level.

We incorporate OECD Environmental Policy Stringency (EPS) indicators to capture national commitment regarding environmental protection. The EPS composite index includes 17 indices; for a detailed descriptio of the index structure, see Figure 2. Higher stringency scores, reflecting supportive institutional and political frameworks, strongly predict both environmental performance outcomes [Frohm et al.] and eco-innovation development [Ghisetti and Pontoni, 2015, Xie et al., 2023]. We apply national EPS scores to all regions within each respective country.

[1] In Rodríguez-Pose and Bartalucci [2023], the so-called "brown" sectors encompass agriculture, forestry, and fishing (A); mining operations (B); petroleum refining (C19); chemical manufacturing (C20); non-metallic mineral production (C23); basic metals (C24); utilities including electricity and waste management (D, E); and transportation (H). Due to data constraints, we retain the NACE sector C entirely. Conversely, because Eurostat incorporates transportation in a broader category including sectors G and I, we exclude it from our analysis. However, we retain construction (NACE F) based on evidence suggesting significant future transformation potential through green job demand [Maldonado et al., 2024].

Table 1: List of Variables

Source	**Name**	**Description**	**Unit of Measure**
EDGAR	ch4	Methane emissions	Kt CO2eq
EDGAR	co2	Carbon dioxide emissions	Kt CO2eq
EDGAR	fgas	Fluorinated gases emissions	Kt CO2eq
EDGAR	n2o	Nitrous oxide emissions	Kt CO2eq
Elaboration on Eurostat data	rate_disp	Rate of change of disposable income	Decimal
EPS	c_co2ets	CO2 Trading Scheme	Integer
EPS	c_co2tax	Carbon dioxide (CO2) Tax	Integer
EPS	c_diesel	Diesel tax	Integer
EPS	c_envpol	Environmental policy stringency index	Integer
EPS	c_lowrd	Low-carbon R&D expenditure	Integer
EPS	c_market	Market-based instruments	Integer
EPS	c_nonmark	Non-market based instruments	Integer
EPS	c_noxlim	Emission limit value NOx	Integer
EPS	c_noxtax	NOx taxation	Integer
EPS	c_pmlim	Emission limit value PM	Integer
EPS	c_renets	Renewable Energy Trading Scheme	Integer
EPS	c_solar	Renewable energy support for solar (Auctions & FITs)	Integer
EPS	c_sox	Emission limit value SOx	Integer
EPS	c_sulflim	Emission limit value for sulphur	Integer
EPS	c_sulftax	Sulphur Oxides (SOx) Tax	Integer
EPS	c_tech	Technology support policies	Integer
EPS	c_wind	Renewable energy support for wind (Auctions & FITs)	Integer
Eurostat	disp	Disposable income	Million purchasing power standards
Eurostat	egap	Gender employment gap	Percentage
Eurostat	gdp	Gross domestic product (GDP)	Million euros
Eurostat	hicp	Harmonised index of consumer prices	Yearly rate of change
Eurostat	hicp_e	Harmonised index of consumer prices for energy products	Yearly rate of change
Eurostat	htec	Employment in technology and knowledge-intensive sectors	Thousands of persons
Eurostat	lllearn	Participation rate in education and training	Percentage
Eurostat	median	Median age of the population	Years
Eurostat	nacea	Employment in Agriculture, forestry and fishing	Thousands of persons
Eurostat	nacebe	Employment in Industry (except construction)	Thousands of persons
Eurostat	nacef	Employment in Construction	Thousands of persons
Eurostat	ted	Tertiary educational attainment, age group 25-64	Percentage
Eurostat	unemp	Unemployment rate	Percentage
REGPAT	ecopat_count	Number of eco-innovative patent applications (EPO)	Integer
REGPAT	pat_count	Number of patent applications (EPO)	Integer

Note: All variables are at the NUTS 2 level, except those from the EPS dataset, the global HICP index, and the energy HICP (which are at the country level). In the model, variables expressed in absolute terms are converted to per-capita values using regional population data or in relative terms using the total number of employed persons in the case of sectoral employment data. The variables that were manipulated in this way carry the label _rel. Employment data by economic activity are classified according to NACE Rev. 2. In Section 4, the variables' code are anticipated by the prefix slag_ when they are the spatial lag of the feature, and/or by lag_ when they are the temporal lag.

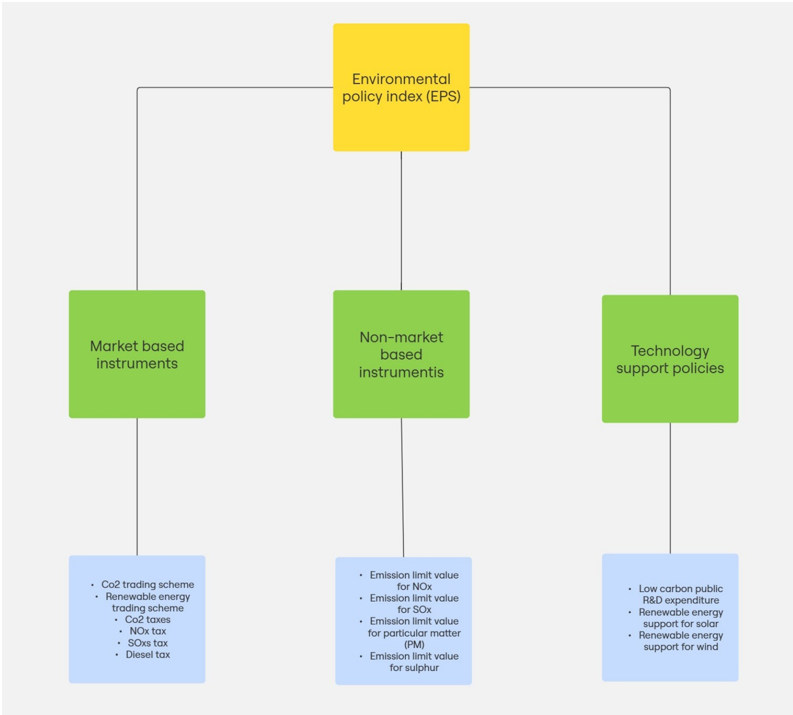

Figure 2: Structure of the OECD Environmental Policy Stringency index

3.2 Methods

We formulate the prediction task as follows, for each region i in year t, based on a set of lagged features $\text{Features}_{i,t-1}$, we aim to determine the function $f(.)$ (a ML model) that best predicts the regional unemployment rate $\text{Unemp}_{i,t}$:

$$\text{Features}_{i,t-1} \xrightarrow{f(.)} \text{Unemp}_{i,t}. \tag{1}$$

Following established approaches in the ML literature, we randomly partition the dataset into 80% for training and 20% for out-of-sample testing. Given the limited number of regional-level observations, different data partitions may yield varying results. To address this limitation, we employ a nested cross-validation strategy, performing 10 different random splits and averaging model performance across repetitions [Resce and Vaquero-Piñeiro, 2022]. Hyperparameter tuning is conducted using a repeated (10 times) five-fold cross-validation procedure.

To prevent data leakage during the training process, we consider only the lagged value (at $t-1$) for all features, including the dependent variable, as suggested by Cerqua et al. [2024]. Finally, we incorporate a spatial lag for all features to account for potential regional interdependencies and agglomeration effects, consistent with the literature reviewed in Section 2. The spatial lag is computed following the standard approach in spatial econometrics [LeSage and Pace, 2009], which calculates the average value of each feature across neighboring regions. This operation doubles the number of features in our model.

To carry out our analysis, we employ four ML algorithms along with a classical multivariate regression model:

Table 2: Models' performances

	GLM	Elastic net	G-boost	Random Forest	NNs
MAPE	0.1191	0.1185	**0.0980**	0.0983	0.8464
RMSE	1.3188	1.3186	**1.1042**	1.1448	9.6891
nRMSE	0.2391	0.2384	**0.2007**	0.2091	1.7599

- **Elastic Net (EN):** A regression technique that balances L1 (LASSO) and L2 (ridge) penalties to improve prediction accuracy and interpretability [Tibshirani, 1996, Zou and Hastie, 2005].
- **Random Forest (RF):** A tree-based ensemble learning method that randomly selects feature subsets to enhance predictive performance [Breiman, 2001].
- **Gradient Boosting Machines (GBM):** An iterative method where each learner corrects the pseudo-residuals of its predecessor [Friedman, 2001].
- **Neural Networks (NNs):** A method where data are processed through hidden layers using a series of weighted connections and activation functions, whose parameters are determined iteratively, to produce a prediction [Bishop, 1995].
- **Multivariate Regression (MR):** An Ordinary Least Squares (OLS) regression approach to use as benchmark.

Performance measures are computed on the test set, with Mean Absolute Percentage Error (MAPE) as the primary evaluation metric due to its unit-independent property [Moreno et al., 2013].

4 Results and discussion

4.1 Prediction performances

The comparison of model performances reveals substantial differences in predictive power. As shown in Table 2, Gradient Boosting Machine emerges as the superior algorithm, demonstrating a robust capacity to capture complex patterns in regional unemployment data. Gradient Boosting follows as the second-best performer, while more traditional models like GLM and Elastic Net show considerably weaker performance. Neural Networks, despite their theoretical capacity to model complex relationships, underperform significantly in this context, with the highest RMSE and MAPE.

With a MAPE lower than 10%, Gradient Boosting achieves highly accurate unemployment predictions, demonstrating the practical value of ML approaches for this policy-relevant forecasting task [Lewis, 1982]. This level of accuracy is particularly significant given the complexity of regional unemployment dynamics and the heterogeneous nature of European labor markets. The superior performance of tree-based methods (GBM and Random Forest) over linear approaches confirms

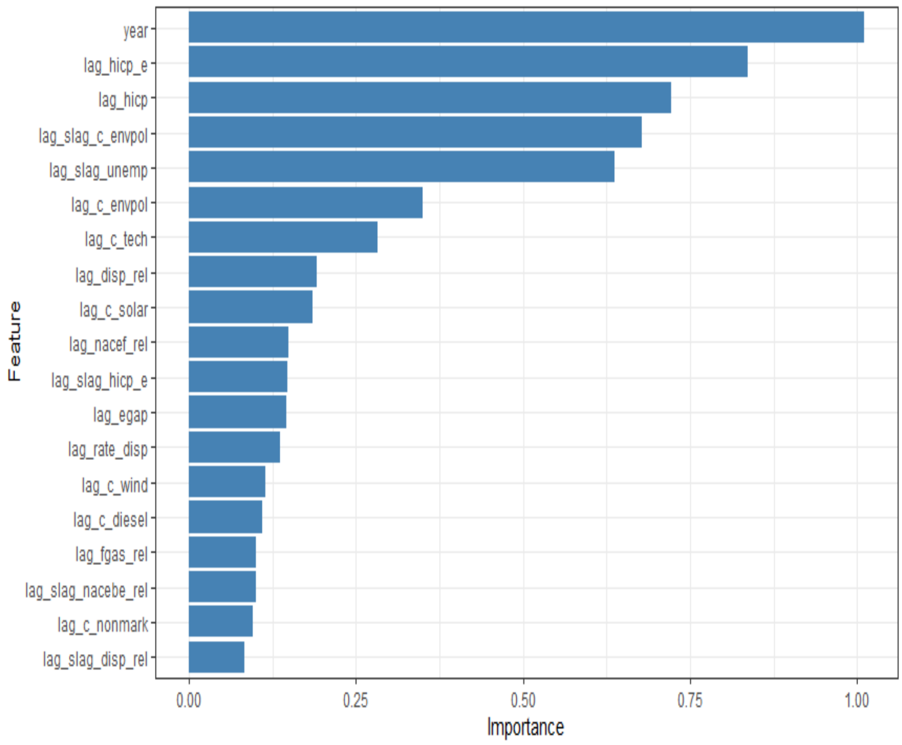

Figure 3: Feature importance plot - first 20 features by importance
Note: The lagged unemployment rate (lag_unemp) constitutes the most influential predictor. Even so, to enhance the interpretability of other substantive predictors, we exclude this dominant variable from the visualization. See Fig. A.1 in Appendix for the case where the lagged unemployed rate is not included among the predictors.

the importance of capturing non-linear relationships and variable interactions in unemployment forecasting [Athey and Imbens, 2019]—a finding that aligns with our theoretical expectation that regional unemployment patterns result from complex interdependencies among economic, demographic, sectoral, and environmental factors. The strong predictive performance of these models validates our methodological approach and suggests that ML techniques can effectively support the proactive policy design emphasized in recent literature on prediction policy problems [Kleinberg et al., 2015]. The ability to forecast regional unemployment with such accuracy provides European policymakers with valuable early-warning capabilities. The excellent predictive performance of tree-based models is confirmed in settings where the lagged unemployment rate is not included among the predictors (see Appendix 5)

4.2 Feature importance

Once the good performance of our models has been established, we turn to the analysis of feature importance in the predictions. The results (Figure 3) offer valuable insights into the key determinants of regional unemployment. Feature importance is reported for the best-performing model, which in our case is the GBM (see Table 2).

In line with our theoretical expectations, the lagged unemployment rate (*lag_unemp*) constitutes the most influential predictor in the GBM model, capturing the majority of explained variance and confirming strong autocorrelation effects that reflect persistent regional unemployment performances. The strong predictive role of the spatial lag of unemployment (*lag_slag_unemp*) further indicates that regional labor markets are deeply interdependent and unemployment outcomes are shaped not only by local structural conditions but also by spillovers from neighboring regions. This result confirms the presence of spatial clustering and agglomeration dynamics in regional unemployment. Note that spatial interdependence appears significant, with the spatial lag of disposable income (*lag_slag_disp_rel*) and of the Regional Harmonized Index of Consumer Prices (*lag_slag_r_hicp*) also featuring among the most important predictors. This validates our methodological choice to incorporate spatial dynamics and corroborates the literature emphasizing the geographic components of unemployment patterns [Iammarino et al., 2019].

The temporal trend (*year*) emerges as the second most important predictor, followed by the harmonized index of consumer prices for energy products (*lag_hicp_e*) and the national HICP (*lag_hicp*). Income-related measures, including disposable income (*lag_disp_rel*), its spatial lag (*lag_slag_disp_rel*), and the rate of change in disposable income (*lag_rate_disp*) emerge as relevant predictors, too. The importance of the time (*year*) component can be partially explained by technological progress and structural change in production systems [Autor et al., 2003, Acemoglu and Restrepo, 2018] and partially by the significant institutional and regulatory reforms affecting the European labour market over the last decades [Bentolila and Ichino, 2008, Boeri and Garibaldi, 2019]. The other factors underscore the central role of income dynamics and inflationary pressures in shaping unemployment outcomes, highlighting the interconnected relevance of economic growth and price stability (two core macroeconomic objectives alongside unemployment) in shaping regional labor market dynamics. Moreover, it supports the claims in Moessner [2022] that energy-specific inflation dynamics and their role in shaping socioeconomic outcomes deserve more in-depth investigation.

Environmental policy stringency (EPS) variables constitute the next relevant block of predictors, with both the national indicator (*lag_c_envpol*) and its spatial lag (*lag_slag_c_envpol*) ranking among the top features. This result highlights that environmental policies, although decided at the national or even supernational level, yield significant effects at the local scale. Notably, the presence of the spatial lag (i.e., the environmental policy stringency of neighboring regions) among the top predictors suggests the presence of relevant country effects in regional economic and labor market outcomes. This is expected, as environmental policies are decided at the national level. Beyond the overall EPS index, policy instruments specifically related to the energy technology mix also appear to play a key role. In particular, *lag_c_tech* captures the EPS associated with technology support policies, including public R&D spending on low-carbon technologies relative to GDP and price support mechanisms for renewable energy (e.g., feed-in tariffs for solar and wind). Among these, the latter type (especially *lag_c_solar*, and to a lesser extent *lag_c_wind*) appear particularly relevant, as both feature among the top predictors. Complementing these, the market-based instrument *lag_c_renets* represents the stringency of Renewable Energy Trading Schemes (e.g., green certificate markets). Together, these results highlight the centrality of regional green technology and innovation capacity, and demonstrate that both market-based regulation and technology support are crucial in shaping regional employment, in line with the findings of Costantini et al. [2018], Marin and Vona [2019], Napolitano et al. [2022].

The number of workers in the industrial sector (relative to regional population) (*lag_nacebe_rel*) and F-gas emissions (*lag_fgas_rel*) also feature among the 20 most important predictors. This suggests that the economic structure of local economic systems, together with their vulnerabilities to the green transition, constitutes a key determinant of regional unemployment dynamics. These findings align with Iammarino et al. [2019], Vandeplas et al. [2022], Rodríguez-Pose and Bartalucci [2024], who identify reliance on labor in "brown" sectors as a key factor of vulnerability to distributive effects of climate policies. In addition, the gender employment gap (*lag_egap*) and its spatial lag also rank among the top 20 features, indicating that in regions (and even more broadly at the macro-regional level) where disparities between male and female employment are wider, labor markets tend to be structurally weaker, reflecting an economic vulnerability but also persistent social inequalities.

4.3 Partial dependence plots

Upon identifying the most important features in predicting regional unemployment, we also investigate the functional impact of each feature on our target variable, relying on partial dependence plots (PDP). This exercise is useful to identify non-linear relationships and to empirically test whether the assumptions that are common to the literature hold when observing real-world data. Indeed, PDPs presented in Figure 4:7 reveal complex non-linear relationships between key features and unemployment rates.

Figure 4 displays the relationships for income- and price-related variables. In the top panel, the rate of change in disposable income (*rate_disp_rel*) shows a markedly negative non-linear relationship with unemployment, though unemployment remains relatively stable at higher growth rates. In contrast, regional disposable income levels (*disp_rel*) exhibit a less clear pattern. This evidence suggests that lower unemployment rates are more likely to be observed in fast-growing regions, while this relationship may not be as automatic when considering static income measures. In sum, stronger income dynamics contribute to more resilient labor market outcomes. At the same time, a positive correlation emerges between inflation-related variables and unemployment growth. This result, confirmed for both the national HICP (*lag_hicp*) and the HICP for energy products (*lag_hicp_e*) (bottom left and bottom right panel), indicates that inflationary pressures linked to general consumption prices and, more markedly, to energy costs, are associated with deteriorating labour market conditions. Rather than indicating a trade-off between inflation and unemployment, the evidence points to a scenario where rising prices erode the purchasing power of households and increase production costs for firms simultaneously, thereby depressing demand and supply. Such dynamics, while challenging the Phillips [1958] hypothesis, highlight the potential for the green transition and energy price shocks to generate stagflation-like outcomes, reinforcing the importance of policies that cushion vulnerable households and firms from the short-term inflationary costs of decarbonisation.

Figure 5 presents partial dependence plots for the most relevant indicators of

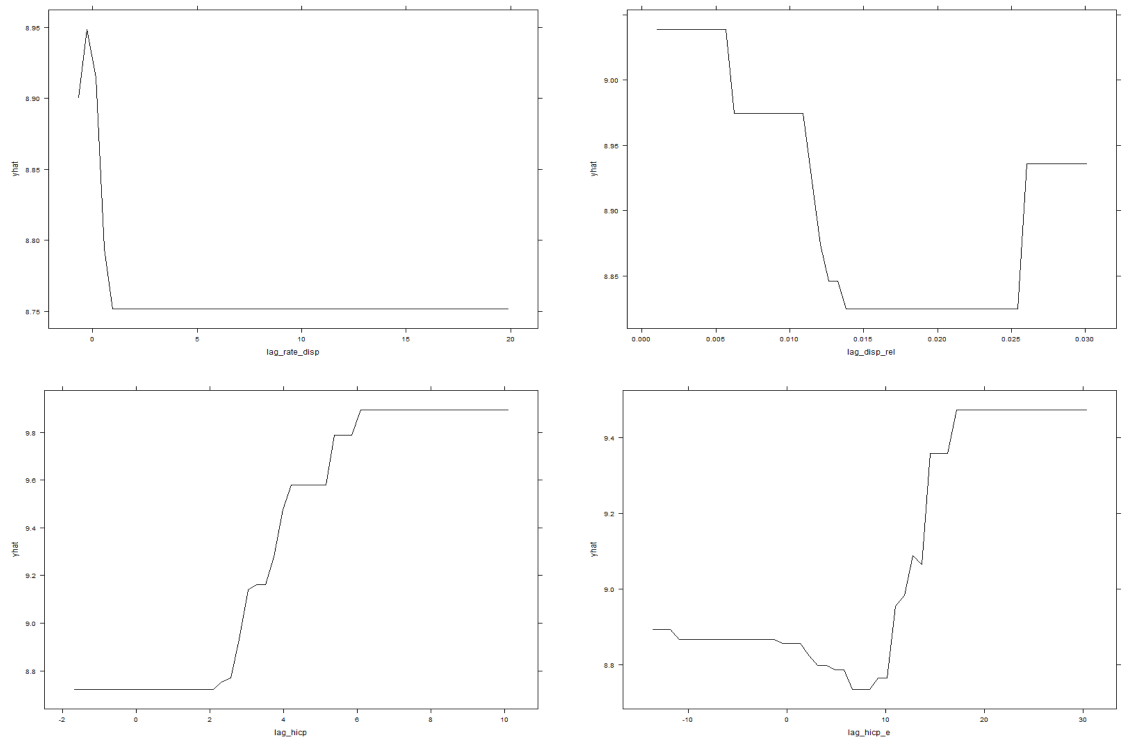

Figure 4: Partial dependence plot on variables of income and prices. Top Left: Rate of change in national income. Top Right: Growth rate of regional income. Bottom Left: HICP. Bottom Right: HICP for energy products.

environmental policy, which reveal important threshold effects and a heterogeneous pattern across instrument types. Starting from the top left (and proceeding clockwise), the overall EPS index (*c_envpol*) shows a U-shaped relationship with unemployment, suggesting that at lower levels greater environmental policy stringency is associated with a lower unemployment rate, whereas at higher levels further increases in stringency correspond to higher unemployment. The technology support policies variable (*c_tech*) displays a pronounced step pattern, with unemployment rates rising sharply at low levels of stringency and stabilizing quickly at higher levels. The specific measure in the form of price support for solar energy (*c_solar*) also shows an overall positive association with unemployment, with relatively lower rates at low levels of stringency and higher rates at higher levels. On the other hand, the stringency of renewable energies markets (*c_renets*), a market-based measure, exhibits a pattern in which unemployment declines sharply at relatively low levels of stringency and then remains low. Overall, this suggests that while technology support policies are essential for fostering innovation and reducing investment risks, they do not necessarily translate into broad employment gains, as innovation activities often rely on high-skilled labor concentrated in high-tech sectors. By contrast, stronger market-based measures (such as renewable energy trading schemes) appear to generate greater benefits for the local labor market, as reflected in lower unemployment rates. This may be due to their fostering effect on the commercialization, adoption, and diffusion of renewable energy technologies and on the related products and services that enable their large-scale economic deployment.

Further nuances appear in the PDPs of variables related to the regional innovation system (Figure 6). Patent applications (*pat_count*) show a generally negative relationship with unemployment rates, though with diminishing returns at higher

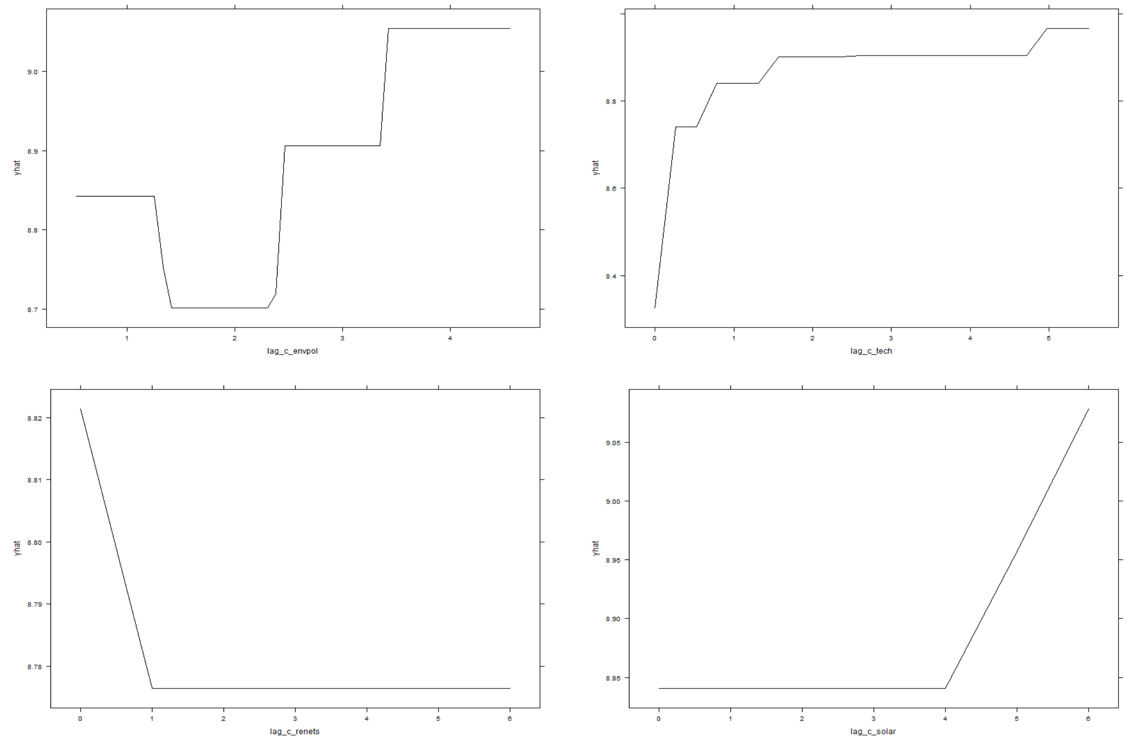

Figure 5: Partial dependence plot on selected EPS indicators. Top Left: EPS composite index. Top Right: Technology support policies. Bottom left: renewable energy trading scheme. Bottom Right: Solar energy support.

patent counts. Green patent applications (*ecopat_count*) exhibit a similar pattern, but with steeper initial reductions in unemployment, suggesting potentially stronger employment benefits from green innovation. This finding supports Napolitano et al. [2022], Costantini et al. [2018], who emphasize the potential of green innovation to drive employment growth. Employment in high-tech sectors (*htec*) also shows a negative association with unemployment, indicating that regions with a higher share of workers in high-tech and knowledge-intensive sectors experience lower unemployment and greater resilience, a hypothesis supported by Marin and Vona [2019]. Finally, tertiary education also shows a negative relationship with unemployment, consistent with the human capital theory that education enhances employment resilience, particularly during periods of economic transition, in line with Consoli et al. [2016], Vandeplas et al. [2022], who highlight the role of skills in differentiating green from non-green jobs. Taken together, the patterns emerging from innovation capacity and technology-related policy variables suggest the possibility of heterogeneous development trajectories – regions endowed with stronger innovation capacity, higher shares of high-tech and knowledge-intensive employment, and a more educated workforce appear to be better equipped to benefit from environmental and technological transitions, as reflected by lower unemployment rates. In contrast, regions with weaker human capital endowments and less advanced sectoral structures are less able to capture the employment gains associated with innovation and may thus face greater adjustment costs.

Figure 7 presents partial dependence plots for emission variables and sectoral employment, revealing complex non-linear relationships that provide important insights into the environment-employment nexus. CO_2 (Top Left) and F-gas (Top Right) emissions display distinctive, yet differentiated, non-monotonic relationships

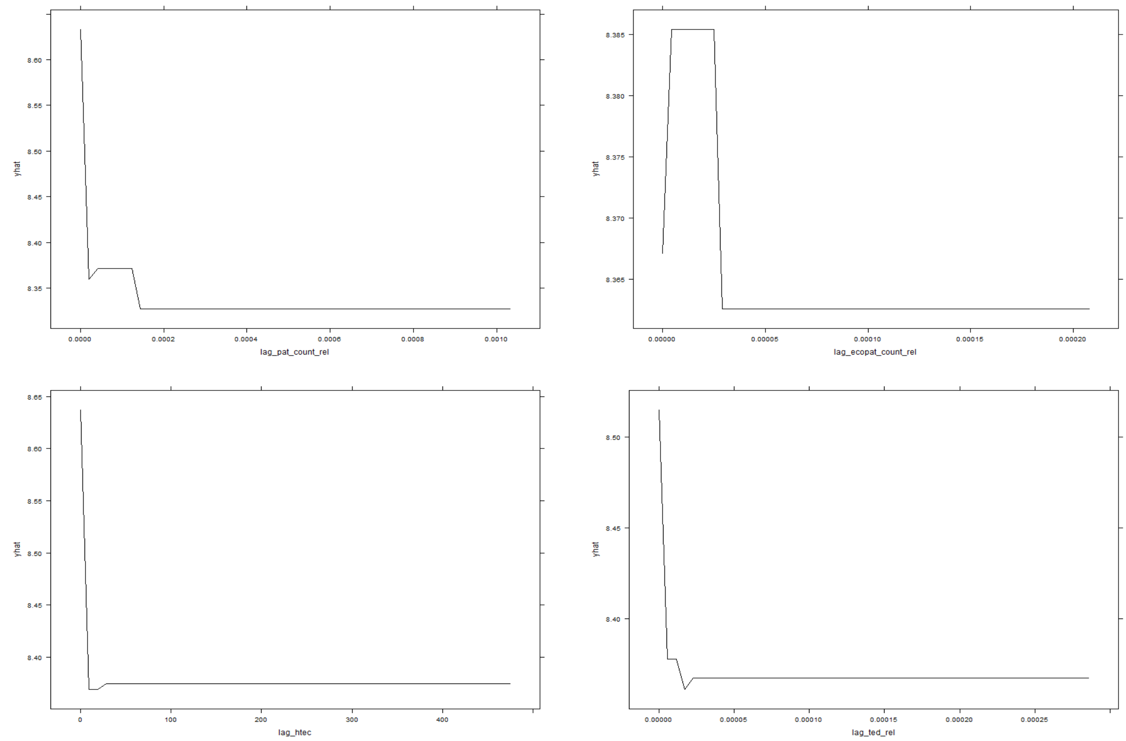

Figure 6: Partial dependence plot on variables of innovation. Top Left: number of patent applications. Top Right: number of green patent applications. Bottom Left: percentage workforce employed in technology and knowledge intensive sectors. Bottom Right: tertiary education attainment.

with unemployment, a pattern that is known to researchers Vona [2021]. On the one hand, CO2 suggests the existence of a "functional" emissions threshold that sustains a minimal economic development: regions with low or with very low emissions may experience higher unemployment due to limited industrial activity, while regions characterised by high level of CO2 emissions are more likely to exhibit lower unemployment rates. This pattern reflects the potential tension between economic activity requirements and environmental sustainability, aligning with Castellanos and Heutel [2019]'s analysis of the complex trade-offs inherent in environmental transitions and labor market outcomes. On the other hand, F-gas emissions exhibit a different pattern, showing a positive relationship with unemployment that grows exponentially at low emission levels before stabilizing at high unemployment values. This distinctive trajectory indicates that regions featuring industrial production systems characterized by high F-gas emissions, typically associated with older, less sustainable technologies [Crippa et al., 2023], are likely to be less resilient in terms of labor market outcomes.

Finally, the relative number of workers employed in industrial (Bottom Left) and Construction (Bottom Right) further exhibit distinctive patterns that reveal important structural differences across economic activities. While higher unemployment rates are associated with higher employment in construction activities[2], employment in the industrial sector shows a strong negative relationship with unemployment. This divergent pattern suggests that manufacturing may operate as a resilient employment anchor with greater opportunities for integrating environmental constraints with job stability, especially if in presence of strong (green) innovative

[2] A similar pattern is observed for employment in the agricultural sector.

capacity. By contrast, regions relatively more specialized in construction activites, in addition to be particularly exposed to the impacts of climate change, risk facing additional adjustment costs due to the green transition. The fact that the building sector is expected to be directly included in the EU ETS2 carbon market (alongside road transport and other activities) further highlights the need for targeted policy interventions.

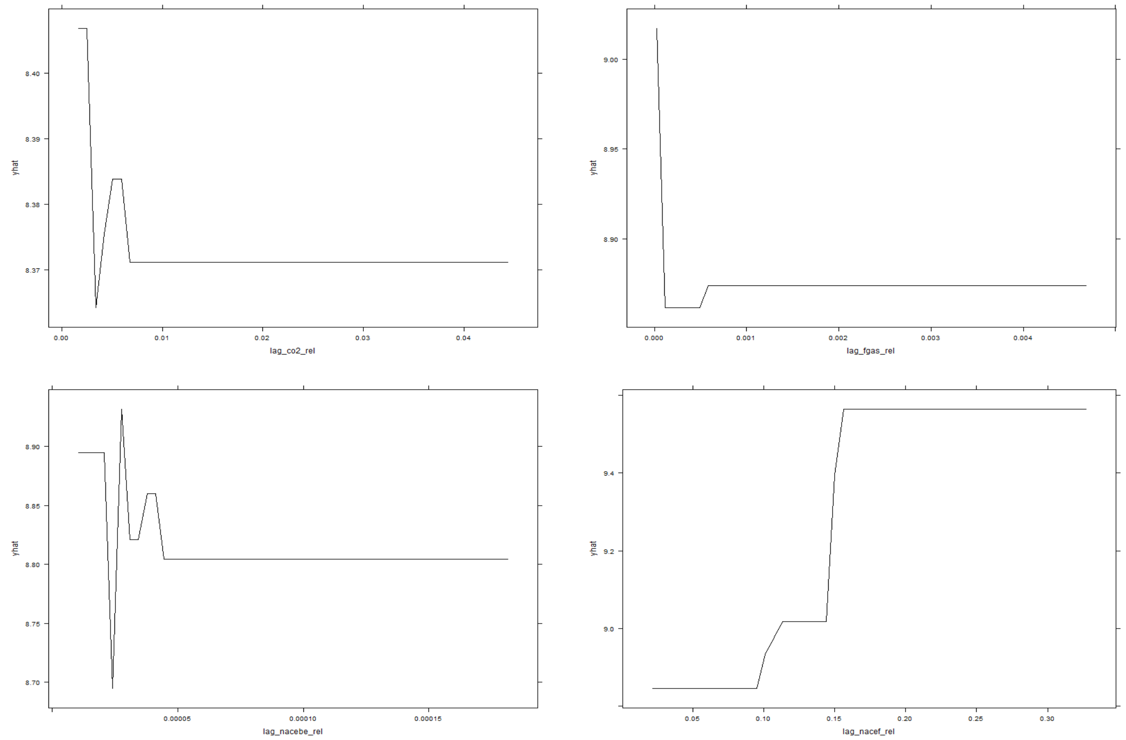

Figure 7: Partial dependence plot on variables of emissions and sectoral composition. Top-left: CO2 emissions. Top-right: F-gas emissions. Bottom-left: share of employment in industry relative to the number of employed persons in the region. Bottom-right: share of employment in construction relative to the number of employed persons in the region.

5 Conclusions and policy implications

This paper set out to predict regional unemployment patterns in the EU by leveraging machine learning methods and a high-dimensional panel of socioeconomic, environmental, and policy-related variables. Our results demonstrate that tree-based models, particularly Gradient Boosting and Random Forest, substantially outperform traditional linear approaches, achieving remarkably accurate forecasts of unemployment dynamics at the NUTS-2 level. This confirms the usefulness of ML techniques for capturing the non-linearities, spatial dependencies, and complex interactions that shape labour market outcomes. The superior performance of tree-based methods over linear approaches confirms the importance of capturing non-linear relationships and variable interactions in unemployment forecasting, aligning with recent applications of ML in economic policy analysis.

By providing highly accurate and spatially granular forecasts, our model offers policymakers an effective early-warning tool to anticipate unemployment risks across regions. Such predictive capacity can support the timely design of place-based labour market interventions, guide the allocation of cohesion funds, and help mitigate the uneven social costs of the green transition through proactive rather than reactive policy responses. Furthermore, our results contribute to several streams of literature by providing data-driven insights on key determinants of regional unemployment. First, we identify a strong link between inflation and unemployment, consistent with decades of economic research. Energy-related inflation appears particularly relevant, underscoring the urgency of considering the direct labor market effects of energy supply shocks. This finding carries important implications as concerns about energy security and affordability emerge in relation to the transition away from fossil fuels. Second, the high predictive importance of spatial lag variables confirms the relevance of agglomeration effects and regional spillovers emphasized in the literature. This spatial interdependence has important implications for European cohesion policy, suggesting that regional interventions must account for cross-border spillover effects and the interconnected nature of regional labor markets. Third, the high predictive importance of environmental policy variables validates concerns raised in the green transition literature about the heterogeneous labor market impacts of climate policies.

Our partial dependence analysis reveals several patterns that merit deeper investigation. Most notably, the relationship between inflation metrics and unemployment contrasts with the traditional Phillips curve hypothesis. The positive correlation between inflation and unemployment growth challenges the conventional inverse relationship, echoing recent critiques regarding the stability and regional applicability of the Phillips curve. This finding is particularly relevant given renewed interest in inflation-unemployment trade-offs following recent inflationary episodes and energy price volatility.

The non-monotonic relationship with emission variables also warrants further investigation. The pattern observed for CO2 emissions suggests critical threshold effects: regions with very low emissions may experience higher unemployment due to limited industrial activity, while regions exceeding optimal emission thresholds face elevated unemployment risks from transition pressures. Environmental policy stringency indicators reveal that different policies produce varying employment effects. Technological support measures, in particular, may not generate the employment gains associated with other market-based instruments. Overall, our results confirms that regions with stronger human capital endowments and environmental protection frameworks demonstrate greater resilience to unemployment shocks. This could raise concerns of equity as the green transition unfolds.

This study exemplifies how ML approaches can advance policy-relevant economic research while providing an ideal testing ground for economic theories. The policy implications are substantial: ML methods can predict unemployment rates with high accuracy, providing policymakers with powerful tools for targeted interventions. As European regions navigate the challenges of green transition and macroeconomic instability, such approaches can provide the granular, predictive insights necessary for effective and spatially-targeted policy interventions, enabling proactive rather than reactive policy responses in regions at risk of rising unemployment.

References

D. Acemoglu and P. Restrepo. Artificial intelligence, automation, and work. In *The economics of artificial intelligence: An agenda*, pages 197–236. University of Chicago Press, 2018.

M. Andini, E. Ciani, G. de Blasio, A. D'Ignazio, and V. Salvestrini. Targeting with machine learning: An application to a tax rebate program in italy. *Journal of Economic Behavior & Organization*, 156:86–102, 2018.

N. Antulov-Fantulin, R. Lagravinese, and G. Resce. Predicting bankruptcy of local government: A machine learning approach. *Journal of Economic Behavior & Organization*, 183:681–699, 2021.

S. Athey and G. W. Imbens. Machine learning methods that economists should know about. *Annual Review of Economics*, 11(1):685–725, 2019.

D. H. Autor, F. Levy, and R. J. Murnane. The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics*, 118(4): 1279–1333, 2003.

N. R. Babb and A. K. Detmeister. Has the us phillips curve flattened? a semi-structural exploration. *Finance and Economics Discussion Series 2017-065*, 2017.

F. C. Bagliano and C. Morana. Eurozone economic integration: Historical developments and new challenges ahead. *European Economic Review*, page 105023, 2025.

N. Barbieri, D. Consoli, L. Napolitano, F. Perruchas, E. Pugliese, and A. Sbardella. Regional technological capabilities and green opportunities in europe. *Journal of Technology Transfer*, 48:749–778, 4 2023. ISSN 15737047. doi: 10.1007/S10961-022-09952-Y/FIGURES/11. URL https://link.springer.com/article/10.1007/s10961-022-09952-y.

R. Basile. Productivity polarization across regions in europe: The role of nonlinearities and spatial dependence. *International Regional Science Review*, 32(1): 92–115, 2009.

S. Bentolila and A. Ichino. Unemployment and consumption near and far away from the mediterranean. *Journal of Population Economics*, 21(2):255–280, 2008.

S. Bianchini, G. Damioli, and C. Ghisetti. The environmental effects of the "twin" green and digital transition in european regions. *Environmental and Resource Economics*, 84:877–918, 4 2023. ISSN 15731502. doi: 10.1007/S10640-022-00741-7/TABLES/8. URL https://link.springer.com/article/10.1007/s10640-022-00741-7.

C. M. Bishop. *Neural networks for pattern recognition.* Oxford university press, 1995.

T. Boeri and P. Garibaldi. A tale of comprehensive labor market reforms: evidence from the italian jobs act. *Labour Economics*, 59:33–48, 2019.

T. Boeri and K. Terrell. Institutional determinants of labor reallocation in transition. *Journal of economic perspectives*, 16(1):51–76, 2002.

L. Boneva and G. Ferrucci. Inflation and climate change: the role of climate variables in inflation forecasting and macro modelling. 4 2022. URL https://www.inspiregreenfinance.org/publications/.

K. Borsekova and S. Korony. Resilience and vulnerability of regional labour markets: principal component analysis of labour market efficiency in the eu. *Regional Studies*, 57(12):2373–2390, 2023.

L. Breiman. Random forests. *Machine Learning*, 45:5–32, 10 2001. ISSN 08856125. doi: 10.1023/A:1010933404324/METRICS. URL https://link.springer.com/article/10.1023/A:1010933404324.

F. Cappellano, A. M. Santos, and N. F. Dotti. Regional ri ventures to tackle climate change: A new geography of challenge-oriented innovation landscape. *Papers in Regional Science*, 103:100052, 10 2024. ISSN 1056-8190. doi: 10.1016/J.PIRS.2024.100052.

N. Caravaggio, G. Resce, and C. Vaquero-Piñeiro. Predicting policy funding allocation with machine learning. *Socio-Economic Planning Sciences*, 98:102175, 2025.

P. D. Caro and U. Fratesi. The role of cohesion policy for sustaining the resilience of european regional labour markets during different crises. *Regional Studies*, 57 (12):2426–2442, 2023.

V. Carrieri, R. Lagravinese, and G. Resce. Predicting vaccine hesitancy from area-level indicators: A machine learning approach. *Health Economics*, 30(12):3248–3256, 2021.

K. A. Castellanos and G. Heutel. Unemployment, labor mobility, and climate policy. Technical report, National Bureau of Economic Research, 2019.

A. Cerqua, M. Letta, and G. Pinto. On the (mis) use of machine learning with panel data. *arXiv preprint arXiv:2411.09218*, 2024.

P. Christensen, P. Francisco, E. Myers, H. Shao, and M. Souza. Energy efficiency can deliver for climate policy: Evidence from machine learning-based targeting. *Journal of Public Economics*, 234:105098, 2024.

L. Coenen, P. Benneworth, and B. Truffer. Toward a spatial perspective on sustainability transitions. *Research policy*, 41(6):968–979, 2012.

D. Consoli, G. Marin, A. Marzucchi, and F. Vona. Do green jobs differ from non-green jobs in terms of skills and human capital? *Research Policy*, 45(5):1046–1060, 2016.

C. Corradini and E. Vanino. Path dependency, regional variety and the dynamics of new firm creation in rooted and pioneering industries. *Journal of Economic Geography*, 22(3):631–651, 2022.

V. Costantini, F. Crespi, and E. Paglialunga. The employment impact of private and public actions for energy efficiency: Evidence from european industries. *Energy Policy*, 119:250–267, 2018.

M. F. Cracolici, M. Cuffaro, and P. Nijkamp. Geographical distribution of unemployment: An analysis of provincial differences in italy. *Growth and Change*, 38 (4):649–670, 2007.

M. Crippa, D. Guizzardi, F. Pagani, M. Banja, M. Muntean, E. Schaaf, W. Becker, F. Monforti-Ferrario, R. Quadrelli, A. Risquez Martin, et al. Ghg emissions of all world countries. *Publications Office of the European Union, Luxembourg*, 10: 953322, 2023.

R. Di Stefano and G. Resce. The determinants of missed funding: Predicting the paradox of increased need and reduced allocation. *Journal of Economic Behavior & Organization*, 231:106910, 2025.

A. Diemer, S. Iammarino, A. Rodríguez-Pose, and M. Storper. The regional development trap in europe. *Economic Geography*, 98:487–509, 10 2022. ISSN 19448287. doi: 10.1080/00130095.2022.2080655. URL https://www.tandfonline.com/doi/abs/10.1080/00130095.2022.2080655.

L. Dijkstra, E. Garcilazo, and P. McCann. The effects of the global financial crisis on european regions and cities. *Journal of Economic Geography*, 15(5):935–949, 2015.

E. A. D'Ecclesiis, E. Levi, and F. Patriarca. Exploring the multifaceted relationship between environmental attitudes and political voting. *Ecological Economics*, 230: 108518, 2025.

N. Fabra, A. Lacuesta, and M. Souza. The implicit cost of carbon abatement during the covid-19 pandemic. *European Economic Review*, 147:104165, 2022.

J. H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232, 2001. ISSN 00905364, 21688966. URL http://www.jstor.org/stable/2699986.

E. Frohm, F. M. D'arcangelo, T. Kruse, and M. Pisu. Environmental policy stringency and co2 emissions: Evidence from cross-country sector-level data. doi: 10.1787/53ddcef7-en. URL https://dx.doi.org/10.1787/53ddcef7-en.

C. Ghisetti and F. Pontoni. Investigating policy and rd effects on environmental innovation: A meta-analysis. *Ecological Economics*, 118:57–66, 10 2015. ISSN 0921-8009. doi: 10.1016/J.ECOLECON.2015.07.009.

S. Girma and D. Paton. Using double-debiased machine learning to estimate the impact of covid-19 vaccination on mortality and staff absences in elderly care homes. *European Economic Review*, 170:104882, 2024.

P. Gogas, T. Papadimitriou, and E. Sofianos. Forecasting unemployment in the euro area with machine learning. *Journal of Forecasting*, 41(3):551–566, 2022.

M. Greenstone, R. Hornbeck, and E. Moretti. Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of political economy*, 118(3):536–598, 2010.

J. Hazell, J. Herreño, E. Nakamura, and J. Steinsson. The slope of the phillips curve: evidence from us states. *The Quarterly Journal of Economics*, 137(3):1299–1344, 2022.

P. Hooper, F. S. Mishkin, and A. Sufi. Prospects for inflation in a high pressure economy: Is the phillips curve dead or is it just hibernating? *Research in Economics*, 74(1):26–62, 2020.

S. Iammarino, A. Rodriguez-Pose, and M. Storper. Regional inequality in europe: evidence, theory and policy implications. *Journal of Economic Geography*, 19: 273–298, 3 2019. ISSN 1468-2702. doi: 10.1093/JEG/LBY021. URL https://dx.doi.org/10.1093/jeg/lby021.

N. Jean, M. Burke, M. Xie, W. M. Alampay Davis, D. B. Lobell, and S. Ermon. Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301):790–794, 2016.

O. W. Johnson, J. Y. C. Han, A. L. Knight, S. Mortensen, M. T. Aung, M. Boyland, and B. P. Resurrección. Intersectionality and energy transitions: A review of gender, social equity and low-carbon energy. *Energy Research Social Science*, 70: 101774, 12 2020. ISSN 2214-6296. doi: 10.1016/J.ERSS.2020.101774.

C. Katris. Prediction of unemployment rates with time series and machine learning techniques. *Computational Economics*, 55(2):673–706, 2020.

J. Kleinberg, J. Ludwig, S. Mullainathan, and Z. Obermeyer. Prediction policy problems. *American Economic Review*, 105(5):491–495, 2015.

P. Kline and E. Moretti. Place based policies with unemployment. *American Economic Review*, 103(3):238–243, 2013.

P. Kline and E. Moretti. Local economic development, agglomeration economies, and the big push: 100 years of evidence from the tennessee valley authority. *The Quarterly journal of economics*, 129(1):275–331, 2014.

A. Kronsell. Gender and transition in climate governance. *Environmental Innovation and Societal Transitions*, 7:1–15, 6 2013. ISSN 2210-4224. doi: 10.1016/J.EIST.2012.12.003.

J. LeSage and R. K. Pace. *Introduction to spatial econometrics*. Chapman and Hall/CRC, 2009.

C. D. Lewis. International and business forecasting methods, 1982.

J. E. Maldonado, A. Vandeplas, I. V. Vigani, and A. Turrini. Assessing green job dynamics in the eu: A comparison of alternative methodologies, 7 2024.

G. Marin and F. Vona. Climate policies and skill-biased employment dynamics: Evidence from eu countries. *Journal of Environmental Economics and Management*, 98:102253, 2019.

R. Moessner. Evidence on climate policy, carbon dioxide emissions and inflation. *International Journal of Global Warming*, 28:136–151, 2022. ISSN 17582091. doi: 10.1504/IJGW.2022.126059.

J. J. M. Moreno, A. P. Pol, A. S. Abad, and B. C. Blasco. Using the r-mape index as a resistant measure of forecast accuracy. *Psicothema*, 25(4):500–506, 2013.

E. Moretti. Local labor markets. *Handbook of Labor Economics*, 4B:1237–1313, 2011a.

E. Moretti. Local multipliers. *American Economic Review: Papers and Proceedings*, 100(2):1–7, 2011b.

E. Moretti. *The new geography of jobs.* Houghton Mifflin Harcourt, 2012.

L. Napolitano, A. Sbardella, D. Consoli, N. Barbieri, and F. Perruchas. Green innovation and income inequality: A complex system analysis. *Structural Change and Economic Dynamics*, 63:224–240, 2022. URL https://doi.org/10.1016/j.strueco.2022.09.007.

A. Niebuhr. Spatial interaction and regional unemployment in europe. *European Journal of Spatial Development*, 1(5):26–26, 2003.

OECD. The jobs potential of a shift towards a low-carbon economy: Final report for the eu commission, dg employment, 2012.

H. G. Overman and D. Puga. Unemployment clusters across europe's regions and countries. *Economic policy*, 17(34):115–148, 2002.

A. W. Phillips. The relation between unemployment and the rate of change of money wage rates in the united kingdom, 1861-1957. *Economica*, 25(100):283–299, 1958.

G. Resce and C. Vaquero-Piñeiro. Predicting agri-food quality across space: A machine learning model for the acknowledgment of geographical indications. *Food Policy*, 112:102345, 2022.

A. Rodríguez-Pose. Do institutions matter for regional development? *Regional studies*, 47(7):1034–1047, 2013.

A. Rodríguez-Pose and F. Bartalucci. The green transition and its potential territorial discontents. *Cambridge Journal of Regions, Economy and Society*, 17(2): 339–358, 2024.

A. Rodríguez-Pose and F. Bartalucci. The green transition and its potential territorial discontents. *Cambridge Journal of Regions, Economy and Society*, 17 (2):339–358, 11 2023. ISSN 1752-1378. doi: 10.1093/cjres/rsad039. URL https://doi.org/10.1093/cjres/rsad039.

D. Sansone. Beyond early warning indicators: high school dropout and machine learning. *Oxford bulletin of economics and statistics*, 81(2):456–485, 2019.

D. Sansone and A. Zhu. Using machine learning to create an early warning system for welfare recipients. *Oxford Bulletin of Economics and Statistics*, 85(5):959–992, 2023.

C. L. Smith. The phillips curve: A relation between real activity and inflation? *Journal of Macroeconomics*, 79:103543, 2024.

J. H. Stock and M. W. Watson. Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97 (460):1167–1179, 2002.

J. H. Stock and M. W. Watson. Phillips curve inflation forecasts. *Journal of Monetary Economics*, 44:99–116, 2008.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58:267–288, 1 1996. ISSN 1369-7412. doi: 10.1111/J.2517-6161.1996.TB02080.X. URL https://dx.doi.org/10.1111/j.2517-6161.1996.tb02080.x.

V. Titl, D. Mazrekaj, and F. Schiltz. Identifying politically connected firms: A machine learning approach. *Oxford Bulletin of Economics and Statistics*, 86(1): 137–155, 2024.

A. Vandeplas, I. Vanyolos, M. Vigani, L. Vogel, et al. *The possible implications of the green transition for the EU labour market.* Publications Office of the European Union, 2022.

F. Vona. Managing the distributional effects of environmental and climate policies: The narrow path for a triple dividend, 2021.

F. Vona. Managing the distributional effects of climate policies: A narrow path to a just transition. *Ecological Economics*, 205:107689, 2023. ISSN 0921-8009. doi: https://doi.org/10.1016/j.ecolecon.2022.107689. URL https://www.sciencedirect.com/science/article/pii/S0921800922003500.

P. Xie, Y. Xu, X. Tan, and Q. Tan. How does environmental policy stringency influence green innovation for environmental managements? *Journal of Environmental Management*, 338:117766, 7 2023. ISSN 0301-4797. doi: 10.1016/J.JENVMAN.2023.117766.

B. Zaman and S. Borsky. The impact of supply structure on solar home system installations in rural off-grid areas. *Environmental Innovation and Societal Transitions*, 40:625–644, 2021.

H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67:301–320, 4 2005. ISSN 1369-7412. doi: 10.1111/J.1467-9868.2005.00503.X. URL https://dx.doi.org/10.1111/j.1467-9868.2005.00503.x.

Appendix

Prediction performance without lagged unemployment rate

To assess the robustness of our predictive framework, we replicate the fitting and prediction procedure described in Section 3, this time excluding the lagged unemployment rate from the set of predictors.

Table A.1 reports the prediction performance across models. The results demonstrate considerable robustness to the exclusion of the lagged dependent variable. Random Forest achieves the lowest mean absolute percentage error (MAPE) at 9.74%, closely followed by gradient boosting at 9.75%. Gradient boosting delivers superior performance in terms of root mean squared error (RMSE) and normalized RMSE (nRMSE), with values of 1.0709 and 0.1995, respectively. Notably, all tree-based ensemble methods maintain MAPE values below the 10% threshold, indicating strong out-of-sample predictive accuracy.

Table A.1: Prediction performance without lagged unemployment rate

Metric	GLM	Elastic Net	G-Boost	Random Forest	Neural Network
MAPE	0.1190	0.1182	0.0975	**0.0974**	0.8460
RMSE	1.2925	1.2952	**1.0709**	1.1211	9.5273
nRMSE	0.2444	0.2440	**0.1995**	0.2099	1.7868

Figure A.1 displays the feature importance rankings for the Random Forest model estimated without the lagged unemployment rate. The ordering of covariates remains substantively consistent with that reported in Figure 3, suggesting that the key predictive features—such as labor market structure, demographic composition, and regional economic indicators—retain their relative importance. This stability further supports the interpretation that our models identify meaningful economic relationships rather than relying heavily on autocorrelation patterns in the outcome variable.

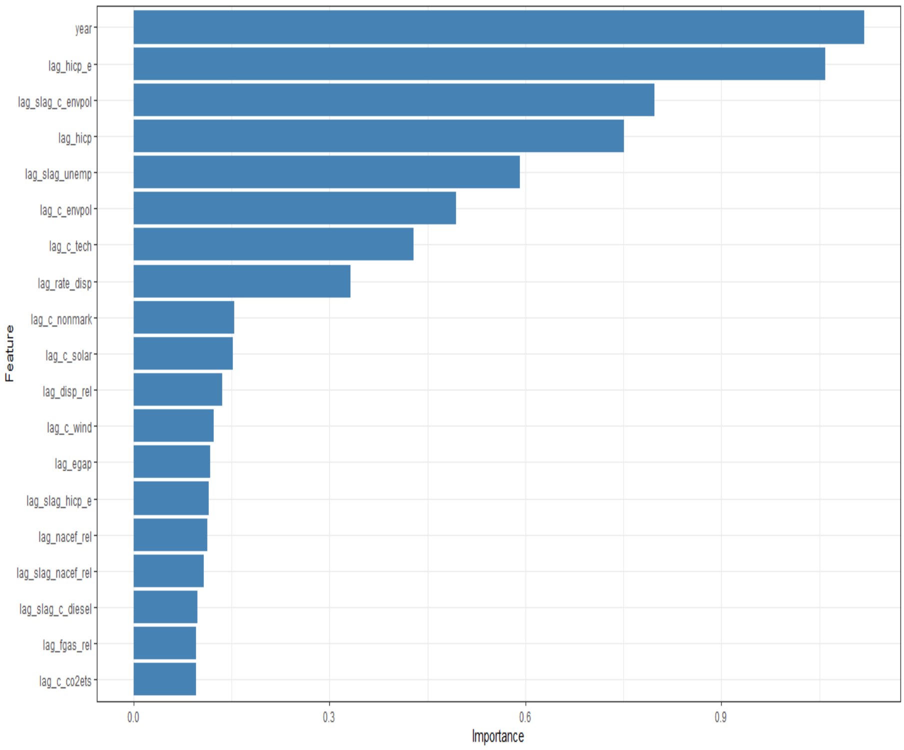

Figure A.1: Feature importance plot - first 20 features by importance

Aknowledgments

Elena Paglialunga acknowledges support from the RE-ADJUST (REgional Assessment of Distributive impacts of a JUst and Sustainable Transition) project, funded through the PRIN 2022 PNRR call on the National Recovery and Resilience Plan (PNRR), Grant Assignment Decree No. 1378 adopted on 01/09/2023 by the Italian Ministry of University and Research (MUR) (Project Code P2022NRAP2, CUP H53D23009510001), financed under Mission 4, Component C2, Investment 1.1 "Fondo per il Programma Nazionale di Ricerca e Progetti di Rilevante Interesse Nazionale (PRIN)", and funded by the European Union – NextGenerationEU.

Giuliano Resce acknowledges support from the THEMIA (THE Mood of Infrastructure's Accessibility) project, funded through the Cascading Call on the National Recovery and Resilience Plan (PNRR), led by the University of Bari "Aldo Moro" (Rectoral Decree No. 424, February 2, 2024), within Spoke 7 "Territorial Sustainability" of the GRINS – Growing Resilient, INclusive and Sustainable project (Project Code PE00000018, CUP H93C22000650001), financed under Mission 4 "Education and Research" – Component 2 "From Research to Business" – Investment 1.3, and funded by the European Union – NextGenerationEU.

We thank participants of the 7th International Conference on Advanced Research Methods and Analytics (2-4 July 2025, Rome) for their valuable comments and suggestions.