The Causal Effect of Productivity on Wages

Stefano Chiesa, Nazar Liubas

January 30, 2025

Abstract

This short paper examines the causal relationship between productivity and wages in Europe, employing robust econometric methodologies to analyze the dynamics at play. Utilizing a fixed-effects cross-country regression model, the study incorporates lagged productivity changes alongside control variables such as inflation, education levels, and market competition to understand their influence on real hourly wage adjustments. The findings reinforce the decoupling of wage and productivity thesis, showing a not statistically significant effect of productivity on wages.
All code and data used in this study are available on GitHub.

1 Introduction

Europe has long struggled with a productivity gap compared to the United States, a trend exacerbated by structural inefficiencies, fragmented markets, and slower adoption of technological innovations. (Draghi 2024) Within Europe, some countries are falling behind at an even more concerning rate, with Italy standing out as a particularly stark example. In 2023 Italian productivity registered -2,5% (ISTAT 2025). Since Italy had a productivity growth close to zero for the last 30 years, this is pointed out as one of the main reasons for its stagnant real wage growth, since other Western European countries are doing better, as can be seen from Figure 1.
This context frames the econometric analysis undertaken to explore the causal relationship between productivity and wages across European countries. The analysis employs a fixed-effects cross-country aggregate-level regression model. Key variables include lagged productivity changes, which capture the impact of past productivity on current real hourly wage changes, alongside control variables addressing time-varying factors like inflation, education levels, and market competition. Country-fixed and time-fixed effects account for structural and global influences, ensuring robust insights into the dynamics at play. The relationship between productivity and wages has been the subject of extensive research in labour economics: according to the most simple version of the Marginal Productivity Theory, a contribution of Classical Economics, real wages are the direct consequence of marginal labour productivity (MPL) (Clark 1908), assuming the presence of perfect competition in both the labour and product markets and full employment. This theory states that in a perfectly competitive economy, real wages are a function of labour's contribution to production and that wages adjust to reflect the value of the marginal product of labour. MPL is the additional output produced by the employment of one more unit of labour, keeping all other factors of production constant. In the Classical framework, firms hire labour up to the point where the wage rate equals the

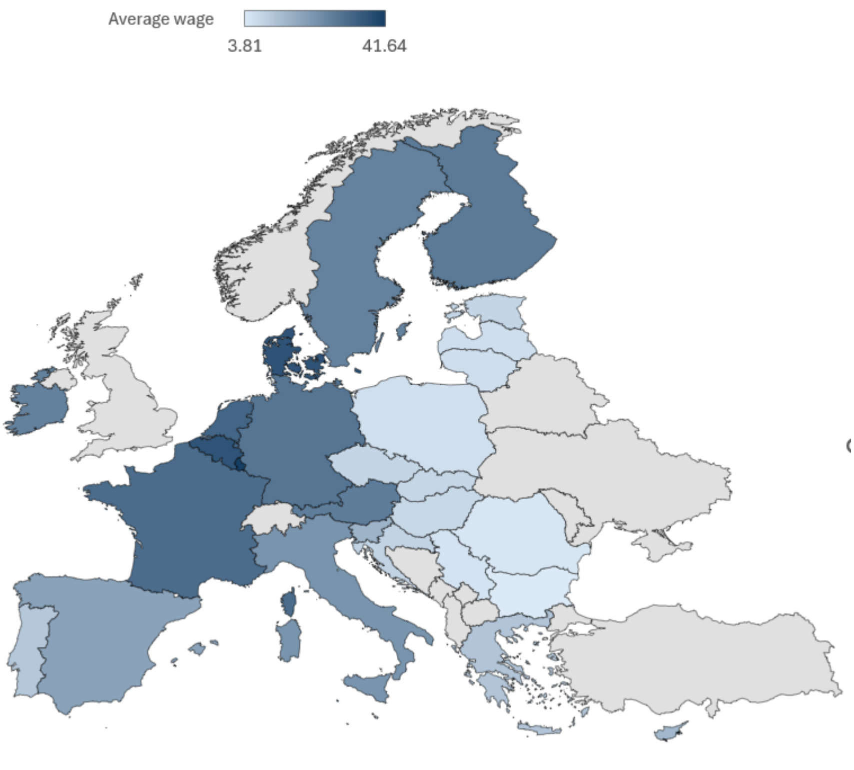

Figure 1: Average real wage over years 2000-2021, Eurostat data

marginal value of the labour's productivity. Workers are paid according to the value they contribute to production.
Mathematically, the marginal product of labour is defined as

$$MPL = \frac{\Delta Y}{\Delta L}$$

where Y represents output and L represents the quantity of labour.
The wage rate is expressed as:

$$W = P \cdot MPL$$

where W is the wage rate and P is the price of the product being produced. The main problem with this theory is that the concept of perfect competition is just theoretical: unemployment, market competition and labour laws are just some of the variables that might affect wages. The goal of the analysis is to assess the causal effect of productivity on real wages, using a fixed effect cross-country analysis using EU data. Since this paper will not assess the problem of bidirectional causal effect, the analysis will focus on the effect of past productivity change on present real wage change.

2 Literature Analysis

Research generally shows a positive relationship between labour productivity and real wages across various contexts. Workplace innovations that increase productivity lead to higher wages for workers. However, the impact on wages can vary, with profit sharing and stock options potentially leading to lower regular wages. The positive wage effects of employee involvement are amplified in unionized workplaces (Black and Lynch 2000).

A firm-level study from MIT analyses the effect of training on productivity and wages, showing that the productivity premium is higher than the wage premium. The authors argue that this occurs because of imperfect competition (Konings and Vanormelingen 2009). In line with this thesis is an OECD report, that shows with a quantitative analysis that the relationship between productivity and real wages has weakened in recent years, showing a decoupling between the two variables (Schwellnus, Kappeler, and Pionnier 2017). In general, the literature agrees on the positive effect of productivity on wages, in line with the Marginal Productivity Theory. Since a deviation from the assumption of the model can lead to a decoupling of this relationship, the public policy setting, unionisation and competition within a market may change the overall effect of productivity on wages. (Millea 2002) argues that there is bidirectional feedback between wages and productivity, which means that there is also a significant effect of real wages on productivity. They also argue that unionization increases the effect of productivity on real wages.

3 Econometric Methodology

In order to assess the causal effect of productivity and wages, the IV approach was considered but abandoned because data that met the exclusion restriction assumption could not be found. In particular, research and development expenditures, globalization index and patent applications were tried and it was found that their correlation with the dependent variable (wages) is more significant than with productivity. Then a fixed-effects cross-country aggregate-level regression was chosen. The model can be summarised with the following equation:

$$\Delta Y_{it} = \alpha + \beta \mathbf{\Delta P}_{i,t-1} + \theta \mathbf{X}_{it} + \sum_i \gamma_i D_i + \sum_t \delta_t T_t + \epsilon_{it}$$

Where: ΔY_{it} is the hourly real compensation percentage change i at year t, representing real hourly wages. $\mathbf{\Delta P}_{i,t-1}$ is annual lagged productivity percentage change i at year $t-1$, capturing the effect of past productivity on current outcomes. $\mathbf{X}_{it}$ is a vector of time-varying control variables, with coefficient θ, that accounts for additional factors influencing Y_{it}. $\sum_i \gamma_i D_i$ represents country-fixed effects that control unobservable time-invariant characteristics of each country, such as geography or institutional quality. $\sum_t \delta_t T_t$ accounts for time-fixed effects, controlling for global shocks or trends common to all countries in a given year, such as worldwide economic crises or technological changes. ϵ_{it} is the error term, capturing unobserved factors that influence ΔY_{it} and are assumed to be uncorrelated with the included explanatory variables.

Assumptions for Causality of Productivity Coefficient

Several assumptions must hold to interpret the coefficient β on lagged productivity as causal. First, the error term ϵ_{it} must be exogenous, meaning it should not be correlated with lagged productivity change or other control variables. This ensures that unobserved factors influencing the dependent variable Y_{it} do not affect past productivity levels simultaneously.

In addition, the model assumes no reverse causality; Using lagged productivity change ($\mathbf{\Delta P}_{i,t-1}$), it is assumed that future values of the dependent variable ΔY_{it} do not influence past productivity, thus eliminating concerns of simultaneity. Moreover, all relevant time-varying confounders that could affect both lagged productivity and the dependent variable must be adequately captured by the control variables $\mathbf{X}_{it}$. The inclusion of $(\sum_i \gamma_i D_i)$ and $(\sum_t \delta_t T_t)$ helps address potential omitted variable bias by controlling for time-invariant country-specific characteristics and global shocks, but it does not take into account the time-variant variables. If any important factors are omitted, the coefficient β may suffer from bias. The model assumes that productivity is measured without error, as measurement errors could lead to attenuation bias and an underestimation of β. Furthermore, it is required that the panel data exhibits sufficient variation across countries and periods otherwise all the variance would be captured by the unit fixed effect (Wooldridge 2010).

4 Data Description

The analysis was conducted using Eurostat and World Bank data. Industry-level (NACE Rev. 2) data was poor or not publicly available, hence the analysis was conducted on aggregate data at the country level (Figure 2). It must be clarified that the countries were not selected according to specific criteria, but rather the consequence of the cross-availability of different data. The period is between 2005 and 2019, with some null values (Table 9) The main source of the data was Eurostat, with a contribution from the World Bank and OECD. For additional information on the data, refer to the Appendix (Table 4).

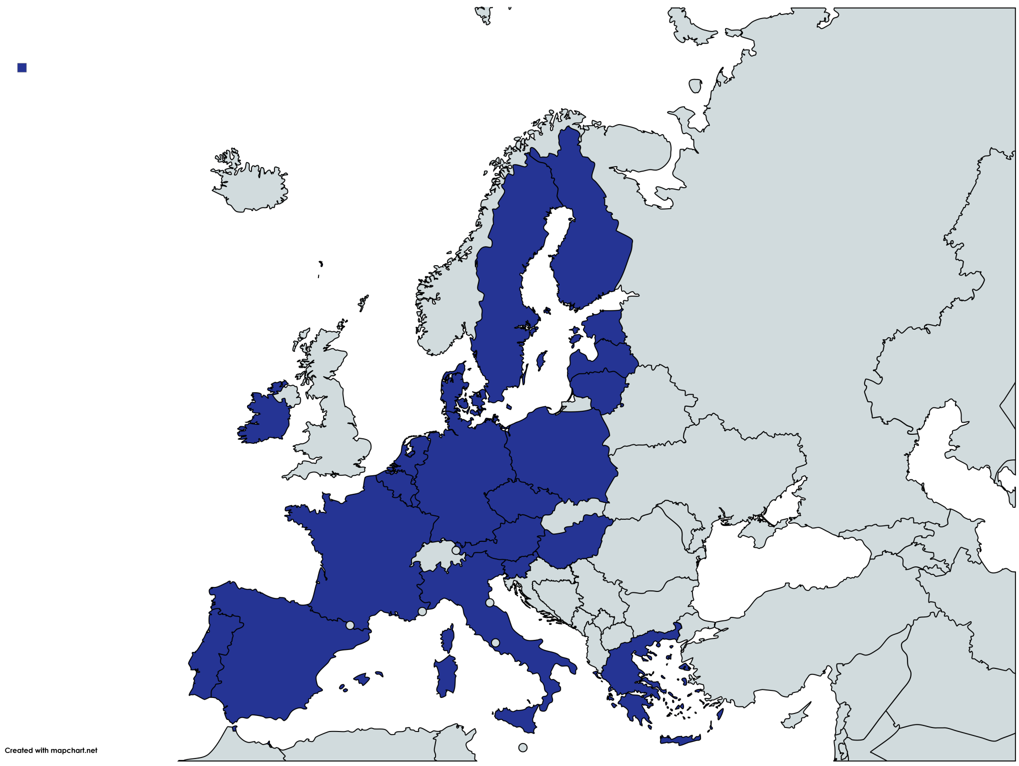

Figure 2: Map of the countries included in the analysis

The idea was to include data about factors that affect the wage equilibrium point in Economics models, such as inflation, level of education, competition in the market, type of contract, state of the economy, training, labour supply and demand. However, some of these variables are not available practically, sometimes are only theoretical concepts. Hence, proxies were selected: for the market competition the Herfindahl Hirschman Index, which has a negative relation with company-side market competition and a positive relationship with market power. The trade union density is a proxy for the influence of centralised bargaining: the higher the percentage of employees unionised, the higher power the unions should have. The unemployment rate is a proxy for the labour supply. For limitations of the proxy or missing variables, refer to Limitations (Section 8).

Preliminary analysis

Variables were checked to have sufficient variation both within (in different years) and between countries. As a result, the market concentration index was excluded from further consideration due to a standard deviation close to zero. To avoid multicollinearity among independent variables, a heatmap of Pearson correlation was created, as one can see in Figure 3. From this, it was concluded that the following sets of variables can not be included in the same model, due to high correlation: {productivity_pp and its lag, productivity_1}, {low_education and middle_education (it is worth mentioning that for every observation 3 values of education sum up to 1, so all of them can't be included simultaneously)}, {percentage change in productivity and percentage change in GDP}.

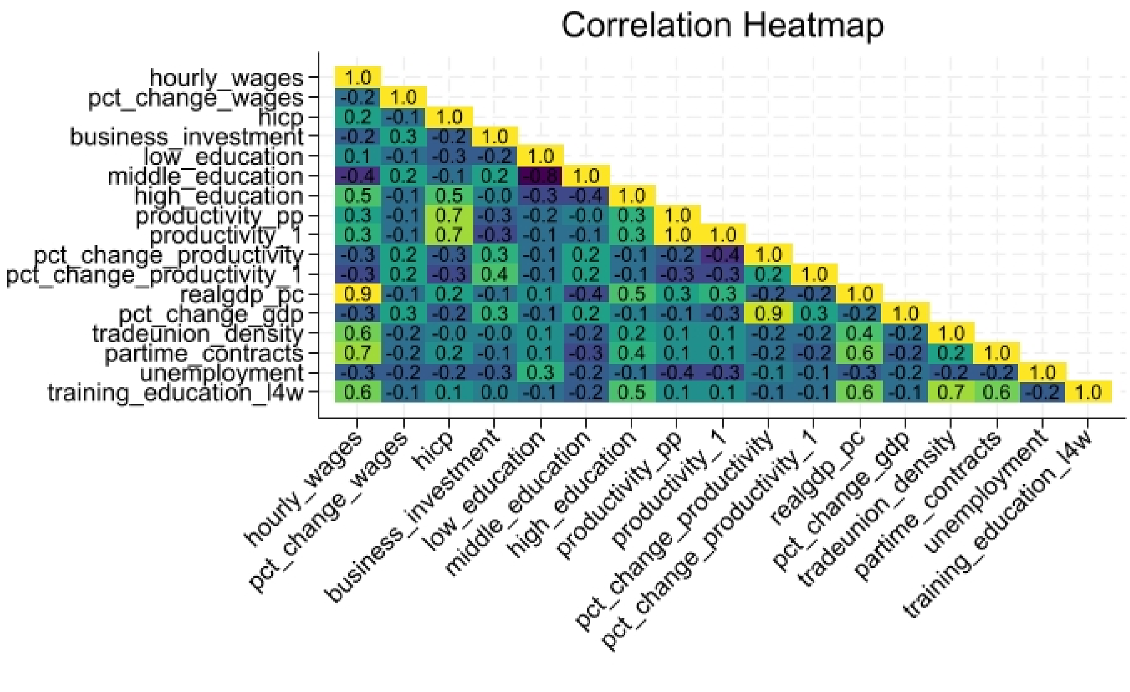

Figure 3: Correlation heatmap of main variables

A modified Wald test for groupwise heteroskedasticity was conducted and it was found that this effect is present in the data (Baum 2000). Therefore, to deal with it, heteroskedasticity-robust standard errors were applied in the regression.

5 Results

Regression results can be seen on the Table 1. Please note that control variables coefficients are not reported for clarity, since they do not have an interpretation in terms of causality. The full table is in the Appendix (Table 5).

	ΔY_t
ΔP_{t-1}	0.123
	(0.121)
Observations	196
R-squared	0.489

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1: Fixed Effect Robust Regression with Control Variables

Even though the coefficient β of ΔP_{t-1} is positive, following the economic theory and hypothesis expectations, the coefficient is not statistically significant.
The main problem of this regression is that the control variables have a positive Variance Inflation Factor (VIF) score, showing multicollinearity which may create problems in the estimation of the coefficient β (Table 2). To address it the Principal Component Analysis was applied. It is a widely used unsupervised method to capture information from data reducing its dimensions. The PCA may cause a loss of interpretability of the coefficient in a regression (Vyas and Kumaranayake 2006), however in an FE regression the control variables' coefficients are usually not reported since they are not interesting for the goal of the analysis. The assumption of exogeneity of productivity has been met since the residuals are uncorrelated with the percentage change of past productivity (Table 12, Figure 5). The selected number of components is seven, as this is the threshold above which the explained variance exceeds 90%. This criterion ensures that a substantial portion of the total variance is captured (Table 13).

Variable	**VIF**	**1/VIF**
high_education	23.52	0.042518
training_last_4w	18.15	0.055091
business_impact	14.08	0.071038
middle_education	10.68	0.093674
tradeunion_density	6.54	0.152796
partime_contracts	4.84	0.206558
unemployment	4.46	0.224311
pct_change_wcp	2.39	0.418741
pct_change_gdp	1.37	0.728471
pct_change_t1	1.35	0.740460
Mean VIF	**8.74**	

Table 2: Variance Inflation Factor (VIF) and 1/VIF for each variable.

The PCA revealed the same results (Table 3): the coefficient β is still positive and not statistically significant. The VIF score is now under control (Table 6).

	ΔY_t
ΔP_{t-1}	0.116
	(0.119)
Observations	196
R-squared	0.474

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Fixed Effect Robust Regression with Principal Components

Out of curiosity, the same regression analysis was conducted utilizing present productivity change as the independent variable. The findings were consistent with those previously obtained, yielding a positive coefficient; however, the results exhibited a lack of statistical significance (Table 11).

6 Robustness Check

To evaluate the robustness of the findings the key check involved dividing the dataset into two subgroups based on the real hourly wage: higher and lower than €20/h. The threshold was selected because wages vary across EU countries, with two distinct groups emerging from the data, as can be seen from Figure 4. The idea was to examine whether the relationship between productivity and wages varies across different levels of compensation. This approach assesses whether the conclusions from the aggregate analysis hold across distinct subpopulations, particularly those with lower versus higher real wages. The results are confirmed for both groups: productivity is positive but not statistically significant, as can be shown in the regression tables in the Appendix (Table 7 & Table 8). Spain is ambiguous since its boxplot's confidence intervals are above the threshold. We chose to put it in the lower group.

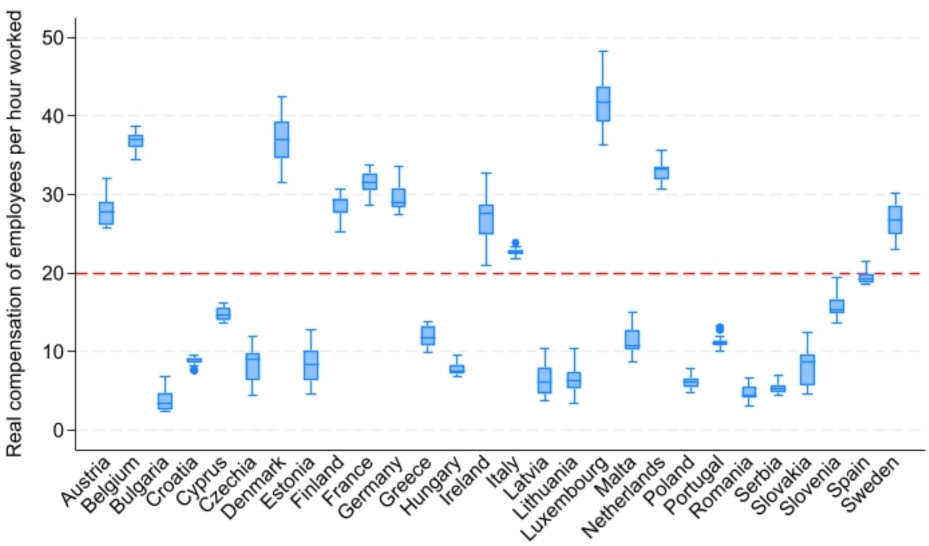

Figure 4: Hourly wages distributions among countries

7 Discussion

The initial hypothesis was a positive causal relationship between productivity and wages; however, the findings indicate that the coefficient for lagged productivity, while positive, is not statistically significant. This result goes against the straightforward application of the Marginal Productivity Theory within the framework of European economies. The results align with existing literature suggesting that the correlation between productivity and wages is not as strong as traditionally presumed. This observation highlights a notable weakening of the wage-productivity link observed over the past few decades. Consequently, the observed decoupling of productivity and wages may significantly influence the causal relationship between these two crucial economic variables.

8 Limitations

The challenge of identifying a valid instrument for the instrumental variable approach necessitated the use of a fixed-effects model. While this model is robust, it may still be susceptible to omitted variable bias if critical time-varying factors are not adequately captured in the control variables. For instance, the model does not have a variable related to job demand, such as job vacancies, due to data availability constraints. Although unemployment can provide some insights since it represents the other side of the equilibrium point, it may not be entirely sufficient. Additionally, the proxy variables employed have their limitations. The HHI serves as a proxy for market competition: concentration is a crucial variable for evaluating market power, but the regulatory environment can influence competitive dynamics. Trade union density acts as a proxy for the impact of unions on the wage equilibrium through negotiations. However, having access to panel data on industry-specific minimum wages would yield a more accurate assessment of unions' influence on salary increases. To address these limitations, future research could consider incorporating additional control variables or more effective proxies for unobserved factors that may affect both productivity and wages. Furthermore, the absence of granular data at the industry level restricts a more nuanced exploration of sector-specific dynamics. Given that productivity growth varies considerably across sectors, particularly in technology-intensive industries, disaggregating the analysis by sector could lead to deeper insights into the relationship between productivity and wages.

9 Conclusion

This short paper aimed to explore the causal relationship between productivity and wages within European economies. By employing a fixed-effects regression model, we sought to quantify the effect of past productivity on current real wages, focusing on a cross-country aggregate analysis utilizing Eurostat, World Bank data and OECD data. Our analysis revealed that, although the coefficient for lagged productivity is positive, it lacks statistical significance. This finding challenges the classical Marginal Productivity Theory, which posits a direct and strong link between productivity growth and wage increases. The observed weakening of the productivity-wage connection, especially in recent decades, aligns with existing literature suggesting that other factors, such as market structures, union power, and technological diffusion, may have a more substantial influence on wage dynamics than productivity alone. The model has some limitations, such as the lack of detailed industry-level data, which may have constrained the results of our analysis. Nevertheless, the results remain consistent with a robustness check, even when segmenting the data by different real wage thresholds. The decoupling of productivity and wages underscores the need for further investigation into the factors that shape this relationship, including labour market policies, industrial structures, and technological changes. Future research could benefit from disaggregating data by industry, using microdata or incorporating more granular measures of market competition and union influence to provide deeper insights into the productivity-wage relationship.

10 Transparency Statement

This paper is the work of two Data Science for Economics students, as a project for the exam of Causal Inference and Policy Evaluation, Università degli Studi di Milano. Grammarly, a large language model AI, was utilized to improve the quality of the grammar.

References

Baum, Christopher F (Oct. 2000). *XTTEST3: Stata module to compute Modified Wald statistic for groupwise heteroskedasticity.* Statistical Software Components, Boston College Department of Economics. URL: https://ideas.repec.org/c/boc/bocode/s414801.html.

Black, Sandra E. and Lisa M. Lynch (2000). "What's Driving the New Economy?: The Benefits of Workplace Innovation". In: *IO: Productivity.* URL: https://api.semanticscholar.org/CorpusID:11092486.

Clark, John Bates (1908). *The Distribution of Wealth: A Theory of Wages, Interest and Profits.* New York: Macmillan.

Draghi, Mario (2024). *The Future of European Competitiveness – A Competitiveness Strategy for Europe.* European Commission. URL: https://commission.europa.eu/topics/strengthening-european-competitiveness/eu-competitiveness-looking-ahead_en.

ISTAT (2025). *Misure di produttività – Anni 1995-2023.* Istituto Nazionale di Statistica (ISTAT). URL: https://www.istat.it/wp-content/uploads/2025/01/REPORT_PRODUTTIVITA_2023.pdf.

Konings, Jozef and Stijn Vanormelingen (2009). "The Impact of Training on Productivity and Wages: Firm-Level Evidence". In: *Review of Economics and Statistics* 97, pp. 485–497. URL: https://api.semanticscholar.org/CorpusID:17221615.

Millea, Meghan (2002). "Disentangling the wage-productivity relationship: Evidence from select OECD member countries". In: *International Advances in Economic Research* 8, pp. 314–323. URL: https://api.semanticscholar.org/CorpusID:153354082.

Schwellnus, C., A. Kappeler, and P. Pionnier (2017). "Decoupling of wages from productivity: Macro-level facts". In: *OECD Economics Department Working Papers* 1373. URL: https://doi.org/10.1787/d4764493-en.

Vyas, Seema and Lilani Kumaranayake (2006). "Constructing socio-economic status indices: how to use principal components analysis". In: *Health Policy and Planning* 21.6, pp. 459–468. URL: https://api.semanticscholar.org/CorpusID:52581.

Wooldridge, Jefrey M. (2010). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press. URL: http://www.jstor.org/stable/j.ctt5hhcfr (visited on 01/27/2025).

11 Appendix

Table 4: Description of Dataset Variables and Sources

Column Name	**Description**	**Source**
geo	Country identifier	Eurostat
year	Year of observation	Eurostat
hourly_wages	Compensation of employees per hour worked, then adjusted for inflation with HICP	Eurostat
hicp	Annual HICP then indexed (2020=100) and transformed in % change	Eurostat
business_investment	Business investment to GDP ratio	Eurostat
low_education	Percentage of population aged 15-64 with primary education	Eurostat
middle_education	Percentage of population aged 15-64 with secondary education	Eurostat
high_education	Percentage of population aged 15-64 with tertiary education	Eurostat
hh_index	Herfindahl-Hirschman market concentration index by country	World Bank
partime_contracts	Percentage of part-time employees	Eurostat
productivity_pp	Real Labour Productiv ity per person (2020=100)	Eurostat
productivity_1	Lagged $t-1$ productivity	Elaboration
productivity_2	Lagged $t-1$ productivity	Elaboration
realgdp_pc	Real GDP per capita, then transformed in % change	Eurostat
tradeunion_density	Trade union density (percentage of employees)	OECD
training_education_l4w	Participation in training and education (last 4 weeks, annualised, 15-64)	Eurostat
unemployment	Unemployment rate	Eurostat

VARIABLES	pct_change_wages
pct_change_productivity_1	0.123
	(0.121)
pct_change_gdp	0.264*
	(0.137)
tradeunion_density	0.256*
	(0.140)
pct_change_hicp	-1.702***
	(0.462)
business_investment	-0.0821
	(0.0935)
middle_education	-0.128
	(0.127)
high_education	0.101
	(0.143)
partime_contracts	0.375
	(0.392)
training_education_l4w	-0.0164
	(0.180)
unemployment	-0.858***
	(0.132)
Constant	0.706
	(10.76)
Observations	196
R-squared	0.489

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Fixed Effect Robust Regression with Control Variables, Complete

Variable	**VIF**	**1/VIF**
pct_change_1	1.22	0.817379
pc3	1.11	0.904025
pc2	1.10	0.909960
pc4	1.03	0.972546
pc6	1.01	0.988527
pc7	1.01	0.993906
pc1	1.00	0.995452
pc5	1.00	0.997017
Mean VIF	**1.06**	

Table 6: Variance Inflation Factor (VIF) and 1/VIF for each variable.

	ΔY_t
ΔP_{t-1}	0.232
	(0.342)
Observations	79
R-squared	0.596

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: FE PCA Regression of Countries with Real Wage $\leq$ €20/h

VARIABLES	ΔP_{t-1}
ΔP_{t-1}	0.232
	(0.342)
Observations	79
R-squared	0.596

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 8: FE PCA Regression of Countries with Real Wage > €20/h

Year	Frequency
2005	2
2008	2
2009	18
2010	20
2011	18
2012	18
2013	18
2014	18
2015	17
2016	20
2017	16
2018	16
2019	13
Total	196

Table 9: Yearly Frequency Distribution

Country	Frequency
Austria	11
Belgium	11
Czechia	10
Denmark	11
Estonia	11
Finland	11
France	5
Germany	11
Greece	3
Hungary	4
Ireland	11
Italy	11
Latvia	10
Lithuania	11
Luxembourg	11
Netherlands	11
Poland	7
Portugal	5
Slovenia	7
Spain	11
Sweden	13
Total	196

Table 10: Country Frequency Distribution

	ΔY_t
ΔP_t	0.255
	(0.187)
Observations	196
R-squared	0.477

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 11: FE PCA Regression with Present Productivity

	ΔP_{t-1}	ϵ
ΔP_{t-1}	1.0000	0.0000
ϵ	0.0000	1.0000

Table 12: Correlation between lagged productivity change (ΔP_{t-1}) and residuals (ϵ)

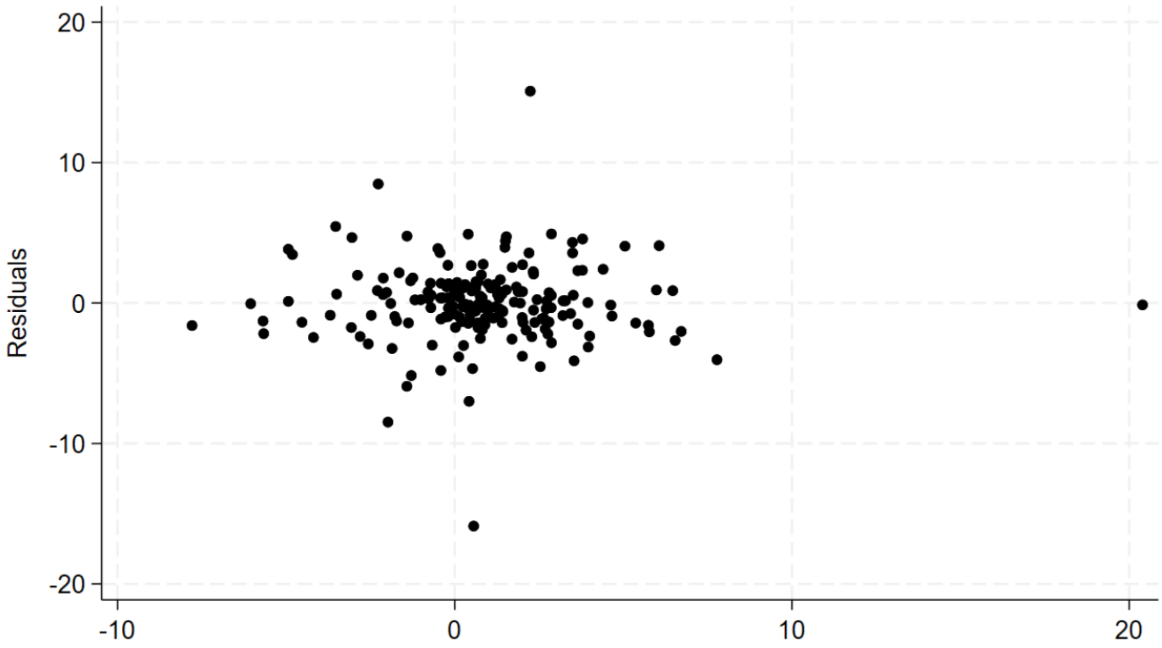

Figure 5: Correlation between residuals and ΔP_{t-1}

Component	**Variance**	**Difference**	**Proportion**	**Cumulative**
Comp1	1.67999	0.276258	0.1867	0.1867
Comp2	1.40373	0.19183	0.1560	0.3426
Comp3	1.2119	0.0804015	0.1347	0.4773
Comp4	1.1315	0.116469	0.1257	0.6030
Comp5	1.01503	0.0129887	0.1128	0.7158
Comp6	1.00204	0.0014865	0.1113	0.8271
Comp7	1.00055	-	0.1112	0.9383

Table 13: Component Variance and Proportions