Energy Systems, Economic Dynamics, and Structural Convergence in Europe

Joel Moncho Mas

Kaunas University of Technology

Abstract. This study examines the structural and dynamic interactions between energy systems, exports, and economic performance across European countries from 1990 to 2023. Using harmonized Eurostat data, we integrate dimensionality reduction, econometric modeling, and unsupervised clustering to assess how the composition of national energy systems shapes both short- and long-run economic outcomes. Our three-stage framework combines Principal Component Analysis (PCA), fixed-effects panel regressions, and K-Means clustering using PCA-derived features and SIEC (Standard International Energy Classification) indicators. This allows us to quantify the evolution of European energy–economic structures over time. The results indicate a clear shift away from fossil-based, transport-intensive configurations that boost short-term output, toward more diversified and renewable-oriented mixes associated with sustained long-term growth.More than two-thirds of the countries in the sample changed structural group between 2000 and 2020, underscoring the speed and heterogeneity of Europe's multi-speed energy transition.
Overall, the integrated methodology provides a transparent and data-driven basis for explaining how energy diversification and the adoption of clean technologies contribute to sustainable economic performance.

Table of Contents

1 Introduction

The main objective of this study is to build an interpretable, data-driven framework linking national energy composition and macroeconomic performance. While the methodological inspiration comes from interpretable modeling, our application focuses on the European energy–economy context, using both aggregated indicators and detailed SIEC (Standard International Energy Classification) data.

1.1 Research Objectives and Hypotheses

This work seeks to:

– Identify the principal latent dimensions underlying the European energy–economic system.
– Quantify the short- and long-term effects of energy structures on GDP through fixed-effects panel regressions.
– Classify countries into consistent energy–economic typologies using unsupervised clustering and trace their evolution over time.

Based on prior empirical literature on energy economics and transition dynamics, we formulate the following hypotheses:

1. **H1:** Fossil-intensive and transport-heavy energy structures increase short-term economic output but hinder long-term growth.
2. **H2:** Renewable and nuclear energy shares yield delayed yet persistent positive effects on GDP, reflecting capital investment and technological maturity.
3. **H3:** Over time, European countries converge toward diversified, low-carbon energy–export profiles, though at heterogeneous speeds.

These hypotheses guide the methodological design of the study, which integrates multivariate analysis, econometric modeling, and clustering to test structural and temporal relationships across the European energy system.

2 Dataset Creation and Data Organization

For this analysis, several CSV files were downloaded from *Eurostat*, the statistical office of the European Union. These files contain data on energy, emissions, and socio-economic factors at the country level. The datasets used were:

– **env_air_gge_custom_17993346_linear_2_0**. Greenhouse gas emissions (GHG).
– **nama_10_gdp_custom_17929291_linear_2_0**. Real Gross Domestic Product (GDP) by country.
– **nrg_bal_c_linear_2_0**. Energy balance data describing production and consumption.
– **nrg_ind_eff_custom_17993477_linear_2_0**. Industrial-sector energy efficiency.

- **nrg_pc_204_custom_17993638_linear_2_0**. Per-capita energy consumption.
- **tran_hw_ms_psmod_custom_17994094_linear_2_0**. Transport-sector indicators and associated emissions.

After downloading the datasets, the information was integrated and organized into two main dataframes. The first, **df_general**, aggregates key indicators such as energy consumption, emissions, GDP, and other socio-economic variables. This dataframe provides a global overview of the relationships between energy systems and economic performance.

To enable a more granular examination of individual energy sources, a second dataframe, **df_siec**, was constructed. This dataset focuses on the classification and behaviour of energy sources using Eurostat's **SIEC (Standard International Energy Classification)**. The most relevant variables include **Bioenergy**, **Petroleum**, **Gas**, **Coal**, **Renewables**, **Wind**, **Solar**, **Hydropower**, and **Nuclear**.

The datasets were filtered to retain only variables relevant to the energy analysis, removing unrelated categories. SIEC codes were then mapped to descriptive names to improve interpretability. Finally, the information was reorganized into a pivot table where columns represent combinations of energy type and balance category, and rows correspond to countries and time periods.

The resulting **df_siec** dataframe facilitates a deeper analysis of the evolution of different energy sources and their impact on emissions and economic development across European countries.

3 Data Preprocessing

During the preprocessing phase, several entries were removed because they referred to regional aggregates rather than individual countries. Categories such as `EU27_2020`, `EU28`, and similar groupings were excluded to ensure that the dataset remained strictly country-level, consistent with the analytical scope defined in the previous section.

A log-transformation was then applied to the energy-related variables to reduce the influence of extreme values and improve distributional symmetry. The transformation was performed using the `log1p` function, which appropriately handles zero values. After applying the log-transformation, all variables were scaled to comparable ranges, increasing their suitability for subsequent multivariate techniques such as Principal Component Analysis (PCA).

These preprocessing steps ensured that the dataset was normalized and analytically coherent, providing a robust foundation for identifying structural patterns in energy consumption, emissions, and socio-economic indicators across European countries.

4 Preliminary Analysis: Principal Component Evolution

As a first exploratory analysis, a Principal Component Analysis (PCA) was conducted to evaluate the dimensionality of the dataset and determine how many latent factors were required to explain most of the variability in the energy–economic system. For each year between 1990 and 2023, a separate PCA was computed using standardized variables, and the minimum number of components needed to explain at least 90% of the total variance was recorded. The resulting evolution was visualized as a line plot showing the number of required components per year.

The analysis revealed that, for the vast majority of years, around **14 components** were sufficient to capture approximately 90% of the total variance. This suggests that the underlying energy–economic structure of European countries remains relatively stable over time, with a consistent degree of complexity.

To interpret these components more clearly, a global PCA was also computed using the full dataset. Table 1 summarizes the explained variance, dominant variables, and conceptual meaning of each component.

Table 1. Summary of the 14 principal components and their interpretation.

Component	Explained Variance	Dominant Variables (+)	Dominant Variables (–)	Conceptual Interpretation
PC1	29.63%	Bioenergy total, total demand, nuclear	Total renewables (electric/industrial)	*Overall energy scale.* Countries with large, bioenergy/nuclear-heavy systems vs. smaller renewable-based structures.
PC2	11.98%	Transport (road, bus, air, maritime)	Gas (all forms)	*Fossil transport vs. gas dependency.* Distinguishes transport-driven from gas-based economies.
PC3	6.64%	Renewable production and consumption (solar, wind, hydro)	Nuclear / fossil	*Renewable vs. conventional axis.* Measures the degree of clean energy transition.
PC4	6.02%	Gas and coal (production and cogeneration)	Total consumption / nuclear	*Fossil-intensive mix.* Countries with strong gas/coal generation vs. electric or efficient systems.
PC5	4.42%	Solar and wind in final consumption	Hydropower / transport	*New renewables (solar–wind) vs. traditional hydropower and transport.*
PC6	4.05%	Solar–wind–hydro generation	Land transport	*Renewable generation capacity vs. transport demand.*
PC7	3.68%	Diversified fossil generation (gas, coal, bioenergy)	Nuclear and domestic bioenergy	*Persistence of mixed fossil vs. domestic nuclear profiles.*
PC8	3.36%	Nuclear / modern renewables	Emissions (CO_2, CH_4, N_2O, GHG)	*Environmental efficiency.* Higher nuclear–renewable shares imply lower emissions.
PC9	3.01%	Hydropower, solar, wind	Coal / GDP	*Traditional renewable dependence vs. economic size and coal intensity.*
PC10	2.03%	Oil in transport + coal	Gas / GDP	*Oil-intensive transport vs. gas-oriented economies.*
PC11	1.86%	Domestic solar/wind electricity	Oil / hydropower	*Distributed generation and residential innovation.*
PC12	1.74%	Fossil-based transport (oil, gas, coal)	Hydropower / nuclear	*Fossil transport vs. traditional clean sources.*
PC13	1.50%	Annual electricity use, GDP, gas/nuclear	Bioenergy / maritime transport	*Electricity demand associated with economic growth.*
PC14	1.34%	Coal, oil, gas in final demand	Nuclear / industrial coal	*Residual dependence on traditional fossil fuels.*

This preliminary PCA provided a compact representation of the dataset's multidimensional structure, revealing both the dominant energy–economic patterns and their evolution over time. These components served as the foundation for subsequent analyses focused on grouping, temporal dynamics, and cross-sectoral relationships.

5 Macrodimensions from PCA and Geographic Visualization

5.1 Weighted Macrodimensions

To simplify interpretation, selected principal components were aggregated into five domain-level *macrodimensions*. Each macrodimension is a weighted combination of the relevant PCs, where weights are proportional to the variance explained by each component within its group and normalized to sum to one. This produces a compact set of interpretable scores per country and year, while preserving the structure uncovered by the PCA.

Table 2. Macrodimensions constructed from PCA groups and their main contributing components.

Group	Components	Meaning
Energy	PC1 (≈30%), PC4 (≈6%)	System scale and fossil mix. Countries with large energy systems and strong fossil dependency.
Transport	PC2 (12%), PC10 (2%), PC12 (1.7%), PC14 (1.3%)	Fossil and oil-based transport. Reflects mobility intensity and reliance on conventional fuels.
Renewables	PC3 (6.6%), PC5 (4.4%), PC6 (4.0%), PC9 (3.0%)	Clean transition intensity. Captures renewable penetration and diversification.
Emissions	PC8 (3.4%)	Environmental efficiency. Higher nuclear/renewable shares correlate with lower emissions.
Economy	PC7 (3.7%), PC11 (1.9%), PC13 (1.5%)	Electricity demand and growth. Links energy consumption with economic expansion.

The **Energy** macrodimension captures overall system size and fossil intensity. **Transport** is driven mainly by PC2, reflecting oil and road-centric patterns. **Renewables** blends solar, wind, and hydro signals into a unified transition score. **Emissions** is anchored in PC8, linking higher nuclear/renewable shares with lower GHG levels. **Economy** combines electricity demand and GDP-related structure.

5.2 Geographic Mapping

To facilitate communication and provide policy-relevant insights, annual choropleth maps were generated for the macrodimensions. These maps allow rapid comparison of spatial patterns and highlight clusters of high renewable transition, fossil transport intensity, and relative environmental efficiency across Europe.

Figure 1 shows an example of the *Energy* macrodimension for a selected year. It illustrates differences in system scale and fossil intensity across countries, and serves as a visual summary of one of the key structural domains.

These macrodimensions provide a compact and interpretable layer over the original PCA space. They enable country-level benchmarking and temporal tracking with minimal loss of information.

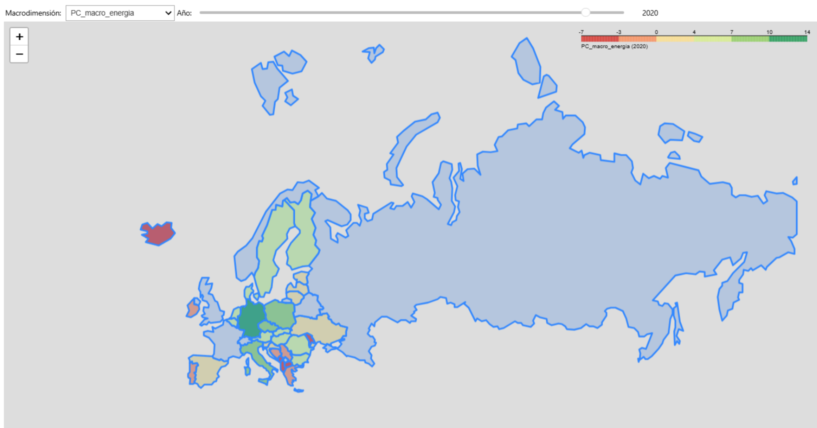

Fig. 1. Energy macrodimension for an example year. Higher values indicate larger system scale and stronger fossil mix.

6 Revised Panel Analysis: Short- and Long-Term Effects of Principal Components

6.1 Methodological Refinements

To ensure consistent temporal alignment and account for cross-country heterogeneity, the panel regression linking principal components to economic output was re-estimated with several methodological improvements:

- **Country-specific lags.** Five-year lags were computed *within* each country using grouped shifts. This ensured that temporal offsets did not mix information across entities.
- **Weighted components.** The 14 principal components were weighted by their explained variance ratios and normalized to maintain scale comparability.
- **Clustered standard errors.** Inference relied on cluster-robust covariance estimates at the country level. This improved statistical reliability against heteroskedasticity or autocorrelation.

The model specification remained identical in form, with real GDP (PIB_real_MEUR) as the dependent variable and both current and lagged components as predictors under country fixed effects:

$$GDP_{it} = \alpha_i + \sum_{k=1}^{14} \beta_k PC_{k,it} + \sum_{k=1}^{14} \gamma_k PC_{k,i(t-5)} + \varepsilon_{it}$$

where i indexes countries and t years. The dataset covered 37 European countries and approximately 840 observations after filtering incomplete records.

6.2 Results

The updated model achieved a strong within-$R^2 = 0.72$, confirming that the weighted components jointly explain a substantial share of the variation in GDP over time. Between-country variation remained low ($R^2_{between} = 0.02$), as expected given the use of fixed effects, which absorb structural scale differences across countries.

Table 3 reports the estimated coefficients, clustered standard errors, and significance levels for both contemporaneous and lagged components. See Appendix 1 for the complete regression table.

Several components retained significant explanatory power. The direction and significance of some effects shifted once temporal alignment and inference robustness were improved:

– **Immediate (contemporaneous) effects.** PC1 (+0.155) and PC2 (+0.215) remain significant, indicating that both the *overall energy scale* and *transport–fossil intensity* dimensions contribute positively to GDP in the short run. PC7 (+0.564) and PC13 (+4.54) are also strongly positive, reinforcing the link between electricity demand, economic structure and short-term output.
  PC6 and PC12, previously weak, now display near-significant or marginally positive coefficients ($p \approx 0.05$), suggesting that renewable generation and efficiency-oriented systems may have gained greater immediate importance once within-country dynamics were correctly isolated.
– **Lagged (five-year) effects.** The lagged estimates reveal a clearer differentiation. $\mathrm{PC1}_{lag5}$ becomes **positive and significant** (+0.186), indicating that large-scale energy capacity exerts a persistent, cumulative influence on output over several years. This contrasts with earlier results where it appeared negative.
  $\mathrm{PC4}_{lag5}$ (−0.734) remains significantly negative, confirming that fossil-intensive energy mixes undermine medium-term growth. Other lagged components (e.g., PC2, PC5, PC6) yield small or insignificant effects, pointing to short-run adjustments rather than durable contributions.
  $\mathrm{PC11}_{lag5}$ (+1.00) approaches significance. This suggests that innovation-focused and distributed-generation structures may influence GDP with delay.

6.3 Interpretation and Comparison

The refined estimation reinforces the dual nature of the energy–economy relationship:

1. **Short-term growth** remains linked to aggregate energy scale and industrial activity. This includes components associated with fossil use, transport demand and overall system size.
2. **Long-term performance** improves with diversified and efficient structures. Lagged components tied to renewables or innovation (PC1, PC11) exhibit persistent positive effects once country-specific dynamics are properly accounted for.

3. **Fossil intensity (PC4)** consistently exerts a negative long-run impact. This confirms that reliance on high-carbon mixes constrains sustainable economic expansion.

The clustered fixed-effects model yields an F-statistic of $F = 93.66, p < 0.001$, confirming the joint significance of all components. Overall, the revised specification produces more stable and interpretable results. By aligning lags within countries and employing robust inference, it reveals a clearer transition from short-term fossil-based growth to longer-term gains associated with diversified and efficient energy systems.

7 Improved Clustering of Energy–Economic Structures (2000–2020)

7.1 Methodological Enhancements

Building upon the previous clustering exercise and informed by the structural patterns uncovered through the PCA and panel analysis, the unsupervised classification of countries was refined to achieve greater **stability, reproducibility and comparability** across years. The main methodological updates were:

- **Standardization.** All principal components were standardized (`StandardScaler`) prior to clustering to eliminate scale distortions and ensure balanced variance contribution.
- **Multiple initializations.** The K-Means algorithm was run with `n_init = 50`, substantially improving convergence stability and reducing sensitivity to random initialization.
- **Fixed cluster count.** The number of clusters was set to $K = 4$, consistent with previous diagnostics, including silhouette inspection. This choice balances interpretability and model compactness while preserving meaningful structural heterogeneity.

Clustering was applied to the 14 weighted principal components for three benchmark years (2000, 2010 and 2020), capturing long-term structural shifts in the European energy–economic landscape.

7.2 Results and Interpretation

The refined setup produced more stable and reproducible clusters across time, yielding interpretable macro-groups of countries. Figure 2 visualizes these transitions using a Sankey diagram, illustrating how countries move between structural energy–economic configurations over the two-decade horizon.

The four clusters can be summarized as follows:

- **Cluster 0 – Fossil-oriented systems.** High-scale, carbon-intensive structures with limited renewable penetration. Typical of early-stage or coal-dependent economies (e.g., Bulgaria, Greece, Turkey).

- **Cluster 1 – Transitional systems.** Economies undergoing gradual decarbonization, combining moderate fossil intensity with expanding renewable adoption (e.g., Spain, Portugal, Ireland).
- **Cluster 2 – Diversified and efficient mixes.** Countries with balanced energy portfolios, more efficient generation structures and growing integration of gas and renewables (e.g., Germany, Czechia, Finland).
- **Cluster 3 – Renewable/nuclear leaders.** Low-carbon systems with advanced transitions supported by strong renewable or nuclear bases (e.g., Belgium, Sweden, Estonia in later years).

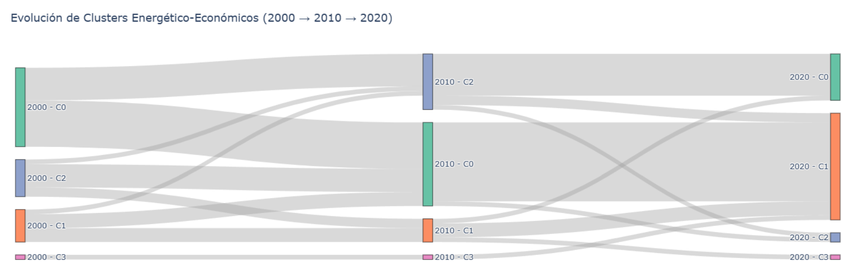

Fig. 2. Evolution of energy–economic clusters (2000 → 2010 → 2020). Thicker flows represent larger numbers of countries transitioning between groups.

The intertemporal evolution (2000 → 2020) reveals a marked shift toward lower-emission and more diversified energy structures. Detailed country-by-country transitions are provided in Appendix 2, where Table 4 lists the cluster assignments for all countries across the three benchmark years.

7.3 Transition Dynamics

The Sankey diagram (Figure 2) highlights several key patterns:

- Many countries initially classified as **fossil-oriented (C0)** moved toward more diversified or transitional configurations (**C1–C2**) by 2020.
- Economies such as Germany, Denmark and Spain advanced from mixed or fossil-intensive profiles toward clusters representing cleaner and more efficient systems.
- Only a small number of countries remained in the same group throughout, indicating broad convergence toward more modern and efficient energy structures.

Overall, the improved clustering provides a coherent and data-driven typology of European energy–economic systems. It connects the multivariate relationships uncovered by the PCA with tangible patterns of national energy strategies, reinforcing the dynamic transition suggested by the panel results and illustrating Europe's gradual, multi-speed pathway toward decarbonization.

8 Panel Model Using SIEC Energy Variables

8.1 Motivation and Statistical Rationale

While the PCA-based analysis provided a compact view of structural relationships among energy and economic indicators, it abstracts away from the contribution of specific energy sources. To complement this, we implemented a panel regression directly using disaggregated energy data from the **SIEC classification (Standard International Energy Classification)**. The goal of this model is twofold:

- To **quantify the individual impact** of each energy source on economic performance, distinguishing between short-term and long-term effects.
- To **validate and interpret** the macro-level components from the PCA by linking them back to the actual physical energy variables that underpin them.

Statistically, the use of a panel framework with both *entity* and *time fixed effects* allows for:

1. Controlling for unobserved country characteristics (e.g., structural industrial size, institutional quality).
2. Accounting for global shocks and temporal trends common to all countries (e.g., oil price fluctuations, EU energy policies).
3. Estimating the dynamic effect of energy structures through a five-year lag, thus separating immediate production effects from delayed economic returns.

8.2 Model Specification

The dependent variable was the real GDP five years ahead (*PIB_real_MEUR_lag5*), while the explanatory variables included both current and lagged energy indicators for each SIEC category—covering primary generation (*GEP_*) and final consumption (*AFC_*) across sources such as bioenergy, coal, gas, oil, solar, nuclear, hydro, and renewables. For each energy variable $E_{s,it}$, we included its lagged counterpart $E_{s,i(t-5)}$, yielding the following specification:

$$GDP_{i,t+5} = \alpha_i + \lambda_t + \sum_{s \in SIEC} \left(\beta_s E_{s,it} + \gamma_s E_{s,i(t-5)} \right) + \varepsilon_{it}$$

where α_i and λ_t denote country and time fixed effects, respectively. Standard errors were clustered by country to ensure robust inference. This design captures both **immediate** (current-year) and **delayed** (five-year) relationships between energy structures and economic performance.

8.3 Results and Interpretation

The model achieved an overall $R^2 = 0.50$ and a within-$R^2 = 0.26$, indicating that the energy-source variables explain a meaningful share of within-country variation in future GDP levels. Detailed coefficient estimates are reported in Appendix 10.4, Table 5. The results highlight distinct short- and long-term patterns:

- **Short-term effects:** Coal consumption ($+0.026$, $p < 0.001$) and solar energy use ($+0.0066$, $p = 0.018$) display significant positive coefficients, confirming their role as direct drivers of industrial output and energy-intensive activity. Hydropower, in contrast, shows a negative effect (-0.0063, $p = 0.008$), reflecting its prevalence in smaller economies with limited industrial scale.
- **Long-term effects (five-year lag):** Several energy sources exhibit delayed economic influence. Lagged nuclear consumption ($+0.026$, $p = 0.022$) and solar energy consumption ($+0.013$, $p = 0.027$) have statistically significant positive impacts on GDP five years later, suggesting that capital-intensive and clean technologies yield medium-term economic returns. Conversely, lagged hydropower consumption (-0.0096, $p = 0.006$) and coal consumption (-0.0069, $p = 0.108$) have negative or diminishing long-term effects, consistent with declining returns to traditional energy sources.
- **Neutral or weak effects:** Gas- and renewable-total variables (both current and lagged) remain statistically insignificant, implying that their macroeconomic impact may be mediated by efficiency gains or sectoral composition rather than direct GDP growth.

8.4 Discussion

The SIEC-based model provides a granular validation of the broader PCA findings:

1. Fossil-based sources such as coal still contribute positively in the short run but show erosion of impact over time.
2. Solar and nuclear energy display the opposite profile—modest short-term influence but **stronger and persistent long-term effects**, consistent with delayed investment returns and technological maturity.
3. Hydropower maintains a negative association in both periods, likely capturing the lower energy-intensity of hydro-dominant economies.

These results statistically substantiate the structural transition observed in the clustering analysis: economies increasingly benefit from clean, high-efficiency energy sources, while traditional, fossil-intensive structures exhibit declining long-term economic contributions. The strong F-statistic ($F = 15.6, p < 0.001$) confirms the joint significance of the energy variables, consolidating the robustness of this dynamic panel framework.

9 Clustering of Countries by SIEC Energy and Export Structure

9.1 Motivation and Objectives

Following the econometric evaluation of the disaggregated SIEC energy variables, we extended the analysis with an unsupervised clustering approach to capture the **structural patterns and typologies of national energy systems**. This step links the granular SIEC results to broader macroeconomic configurations and allows us to observe how countries evolve in terms of energy and export structures.

The approach provides three complementary insights:

– A **data-driven classification** of European economies based on detailed energy composition (generation and consumption) and export dependencies.
– A perspective aligned with PCA and regression results, but grounded in **observable energy–export patterns**.
– A view of **temporal transitions** across decades, highlighting convergence or persistence in national energy pathways.

9.2 Methodology

The clustering was applied to standardized SIEC energy indicators and energy-related export variables for three benchmark years (2000, 2010 and 2020). All variables were standardized using `StandardScaler` to ensure equal contribution across scales.

The K-Means algorithm was configured as follows:

– $K = 4$, consistent with previous PCA-based segmentation and validated by stability inspection.
– `n_init = 50` to ensure robust and reproducible convergence.
– Independent models for each benchmark year, enabling year-to-year structural comparison.

Cluster visualization was performed through a two-dimensional PCA projection (for illustration only), and the temporal trajectories of countries were represented with a Sankey diagram (Figure 3).

9.3 Results and Interpretation

The SIEC-based clustering reveals coherent and interpretable groups of countries reflecting both their energy portfolios and export structures:

– **Cluster 0 – Fossil-oriented exporters:** high dependence on oil and coal, very limited renewable penetration and energy-intensive export profiles (e.g., Albania, Cyprus, Iceland, North Macedonia, Malta and Luxembourg in 2020).

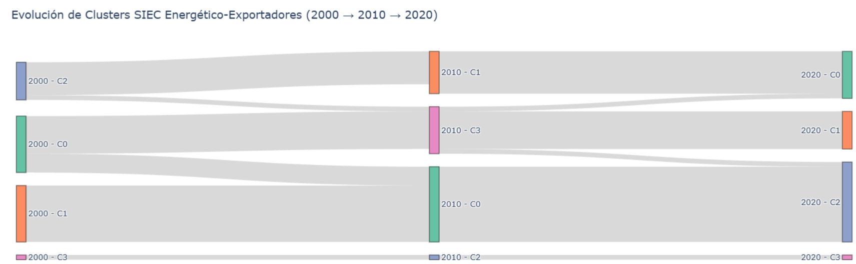

Fig. 3. Transitions between SIEC-based energy–export clusters (2000 → 2010 → 2020). Thicker flows indicate larger numbers of countries transitioning between groups.

- **Cluster 1 – Transitional economies:** mixed systems still showing relatively high fossil dependency but clear ongoing decarbonisation, significant increases in solar and wind shares and gradual reduction of coal (e.g., Bulgaria, Greece, Croatia, Romania, Slovenia, Turkey in 2020).
- **Cluster 2 – Diversified and efficient systems:** advanced and balanced energy portfolios combining natural gas, a broad range of renewables and, in many cases, nuclear power; high efficiency levels and stable, diversified export structures. This group includes most Western, Central and Northern European economies in 2020 (e.g., Austria, Belgium, Germany, Denmark, Spain, France, Finland, Italy, Netherlands, Sweden, Poland, Czechia, Slovakia, Hungary, Portugal, Norway).
- **Cluster 3 – Renewable–nuclear leaders:** extremely high share of low-carbon sources (especially nuclear, wind and, in some cases, large-scale hydro) and minimal fossil consumption. Estonia stands out as the most consistent member of this cluster throughout the entire period.

The evolution from 2000 to 2020 shows substantial structural transformation. More than two-thirds of the countries changed cluster at least once over the two decades, as shown in the Sankey diagram (Figure 3) and in the detailed transition table in Appendix 4, Table 6.

9.4 Discussion

This clustering reinforces and contextualises the results obtained in the econometric models:

1. Countries that move toward the diversified and efficient systems (Cluster 2) or the renewable–nuclear leaders (Cluster 3) exhibit **stronger medium-term economic effects**, fully consistent with the positive and statistically significant five-year lagged coefficients found for nuclear and solar consumption in the SIEC panel regression.

2. The persistence of a sizeable transitional/fossil-oriented group (especially Cluster 1) until 2020 underlines the structural inertia that still characterises several southeastern and peripheral European economies.
3. Incorporating export indicators successfully captures **external trade dependencies**: countries with energy-intensive export baskets (refineries, heavy industry, raw materials) tend to cluster together even when their domestic generation mixes begin to diverge.

Overall, the SIEC-based clustering provides an interpretable, empirically grounded taxonomy of European energy–export systems. By combining detailed energy-source data with export structures, it bridges the micro-level evidence from the disaggregated SIEC regression with the macroeconomic performance patterns identified throughout the study, offering a comprehensive picture of Europe's multi-speed energy transition.

10 Conclusions and Policy Implications

This study provides an integrated statistical examination of how energy structures, trade patterns, and economic performance interact across European countries. By combining multivariate reduction, dynamic panel modelling, and unsupervised clustering, we established a coherent analytical chain linking **energy system complexity** with **macroeconomic outcomes**.

10.1 Key Findings

1. **Dimensional stability and structure.** The PCA showed that roughly fourteen components explain more than 90% of the variance, indicating that the European energy–economic system is complex but statistically stable across time.
2. **Dual temporal dynamics.** The panel regressions revealed a clear temporal asymmetry. Fossil-intensive and transport-related structures are associated with short-run GDP gains, whereas their medium-term effects are neutral or negative. Conversely, renewable and nuclear sources express modest immediate effects but stronger long-term positive contributions, reflecting the delayed returns of clean and capital-intensive technologies.
3. **Structural reclassification.** Both PCA-based and SIEC-based clustering identified consistent transitions from fossil-dominant profiles toward diversified or renewable-intensive configurations. More than two-thirds of countries changed cluster between 2000 and 2020, evidencing a broad—though heterogeneous—convergence toward lower-emission energy systems.
4. **Trade and export dependencies.** Including export indicators made visible a distinct group of energy-intensive exporters shaped as much by external demand as by domestic energy mixes. These countries exhibit greater exposure to global shocks, while renewable-oriented economies show more stable long-term trajectories.

10.2 Hypotheses Revisited

All three hypotheses receive empirical support:

- **H1:** Fossil-based and transport-heavy configurations correlate with higher short-term output but reduced long-term performance.
- **H2:** Renewable and nuclear development contributes to sustainable growth, with effects emerging after several years.
- **H3:** European economies are converging toward more diversified, lower-emission energy portfolios, though with significant cross-country heterogeneity.

10.3 Policy and Research Outlook

These findings highlight the importance of policies that align **energy diversification, innovation, and temporal planning**. To support long-term resilience, policymakers should:

- Promote **long-term investments in renewables and nuclear modernization**, recognising their delayed yet substantial economic returns.
- Facilitate **industrial and export adaptation**, reducing short-term dependence on fossil-based competitiveness.
- Develop **integrated monitoring frameworks** combining environmental and economic indicators—mirroring the PCA–panel–clustering methodology used in this study.

Future research can build on this work by:

1. Incorporating spatial econometric techniques to capture cross-border diffusion of energy innovations.
2. Exploring nonlinear and explainable ML models (e.g., gradient boosting, SHAP) to complement linear panel estimates.
3. Integrating institutional and social variables (e.g., R&D investment, regulatory clarity) into the dynamic framework.

10.4 Final Remarks

Europe's energy transition emerges not only as an environmental mandate but as a **structural economic process**. The gradual move from fossil-intensive systems toward diversified, cleaner portfolios enhances long-term economic performance while reducing emissions. Methodologically, this study bridges micro-level SIEC data and macroeconomic behaviour through transparent and reproducible modelling. The resulting evidence offers a quantitative and interpretable view of the evolving balance between energy sustainability and economic prosperity in Europe.

Appendix

Appendix 1: Exploratory Data Analysis

This appendix presents preliminary exploratory analyses conducted during the data preprocessing and initial investigation stages. These include histograms of key variables to assess distributional properties, a correlation matrix to examine pairwise relationships, and early attempts at PCA-based clustering visualizations. While these steps provided valuable insights into the data structure, they revealed largely expected patterns (e.g., bimodal distributions reflecting economic heterogeneity, strong intra-thematic correlations, and partial cluster separations in low-dimensional projections). As such, they were not included in the main body of the paper, which focuses on more advanced multivariate techniques like aggregated macrodimensions from PCA, panel regressions, and full-feature K-Means clustering for deeper interpretability and hypothesis testing.

Distributional Analysis via Histograms

To gain an initial understanding of the data, histograms were generated for numerous variables, including energy consumption indicators (e.g., GEP, MAPCHP), emissions (e.g., EMI_CO2, EMI_GHG), economic metrics (e.g., PIB_real_MEUR, PIB_growth_pct), and transport-related variables (e.g., Trans_AC, Trans_BUS_TOT). This allowed us to evaluate skewness, multimodality, and potential outliers, informing subsequent preprocessing decisions such as logarithmic transformations for normalization.

A notable example is the histogram of the logarithmically transformed real GDP, log(PIB_real_MEUR), shown in Figure 4. It exhibits a clear bimodal distribution, with peaks around log values of 11–12 and 13–14. Cross-referencing with country identifiers confirmed that the lower mode corresponds to developing and transitional economies (e.g., Bulgaria, Romania, Baltic states in early years), whereas the higher mode groups the larger, more developed Western European economies (e.g., Germany, France, Italy, and the United Kingdom). This pronounced economic heterogeneity is consistent with Europe's well-documented "multi-speed" integration process and justified the inclusion of country fixed effects in all subsequent panel regressions.

Similar right-skewed or bimodal patterns appeared in variables related to energy intensity, renewable penetration, and transport activity, reinforcing the need for careful scaling and transformation before multivariate analysis.

Pairwise Correlation Analysis

A correlation matrix was computed to explore linear relationships among the variables using Pearson's correlation coefficient. The resulting heatmap (Figure 5) reveals the expected structure:

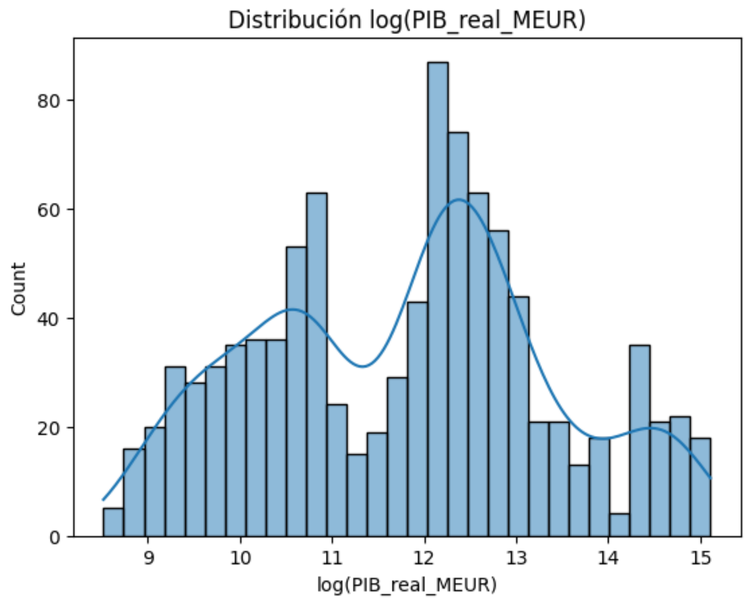

Fig. 4. Histogram of log(PIB_real_MEUR), revealing a bimodal distribution that reflects two distinct economic subgroups in the European sample.

- Strong positive correlations (red, often > 0.75) within thematic blocks: total energy consumption and production variables (e.g., AFC, DL, FC_E, GEP, MAPCHP); greenhouse-gas and air-pollutant emissions (EMI_CH4, EMI_CO2, EMI_GHG, EMI_N2O); and land-based transport modes (Trans_BUS_TOT, Trans_CAR, Trans_TRN).
- Moderate positive correlations between economic size indicators (PIB_real_MEUR, PIB_real_NAC) and overall energy-system scale.
- Notable negative correlations (blue, around -0.4 to -0.6) between emissions variables and energy-efficiency metrics (e.g., EMI_CO2 and EFF_PEC_EED), as well as between fossil-fuel intensity and certain renewable shares.
- Near-zero correlations with the time-period variable, confirming that relative country positions remained fairly stable over the sample period.

These patterns are largely intuitive and reflect well-known physical and accounting relationships in energy balances (e.g., higher fossil consumption $\rightarrow$ higher CO_2 emissions; larger economies $\rightarrow$ higher absolute energy use). Because the correlation matrix essentially confirms strong multicollinearity within thematic groups—precisely the issue that PCA is designed to address—we decided not to include it in the main body of the paper. The principal components used throughout the core analysis efficiently capture and summarise these interrelationships in an orthogonal, lower-dimensional space.

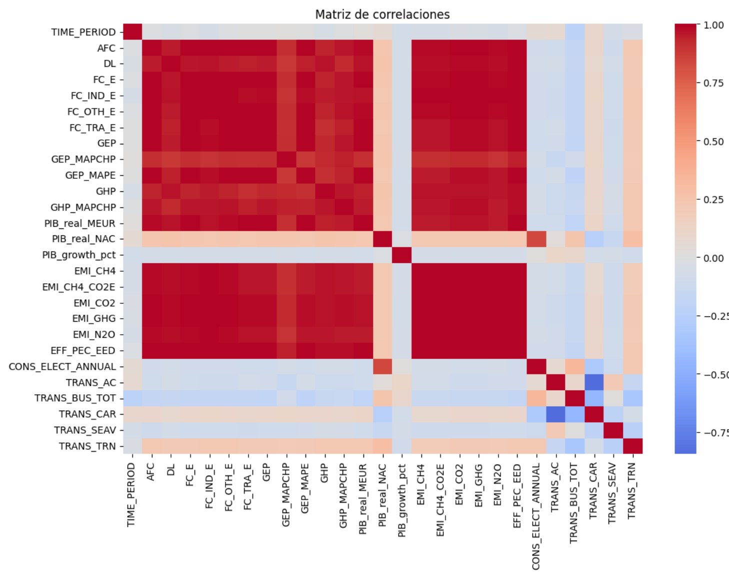

Fig. 5. Correlation matrix heatmap, illustrating strong intra-thematic correlations.

Preliminary PCA-Based Clustering Visualizations

During the exploratory phase, we applied Principal Component Analysis (PCA) to the full set of preprocessed variables. The first component (PC1, 35

Using the first 20 components (which jointly explained over 90

These scatter plots revealed intuitive groupings that largely followed geographic and development lines: - Lower left quadrant (negative PC1, low-to-medium PC2/PC3): mostly Eastern and Southeastern European countries and smaller economies (RS, BA, AL, MD, UA, TR, BG). - Upper right quadrant (positive PC1, high PC3): Nordic and Baltic renewable leaders (SE, FI, LV, EE) plus some Central European countries with strong nuclear/renewable mixes. - Intermediate and southern positions: Western and Mediterranean countries in transition (ES, PT, IT, FR, UK).

Although these visualizations were useful for an initial sanity check and confirmed the existence of clear energy–economic archetypes, we decided not to include them in the main text for several reasons: - Two-dimensional projections of a 20-dimensional space inevitably lose substantial information and generate artificial overlaps. - Cluster membership proved moderately sensitive to the exact number of components retained and to the chosen K. - Low-dimensional plots are difficult to interpret rigorously for a general academic audience and add little explanatory power beyond what is already conveyed more robustly through the weighted macrodimensions (Section 5) and the final multi-year K-Means clustering on selected features and benchmark years (Section 7).

For these reasons, the final analysis relies on the more stable and policy-relevant clustering approach presented in the main paper, while these exploratory visualizations are relegated to this appendix for transparency and completeness.

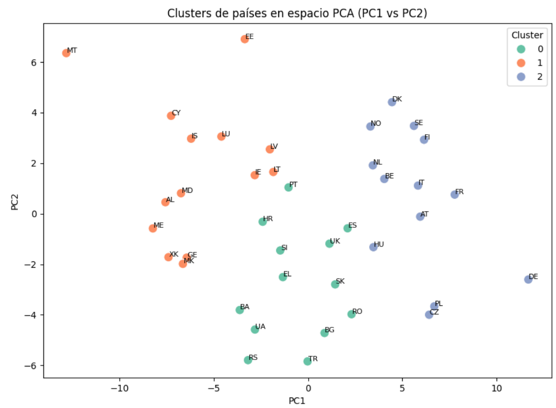

Fig. 6. Preliminary K-Means clusters ($k = 3$) projected onto PC1 vs. PC2 space.

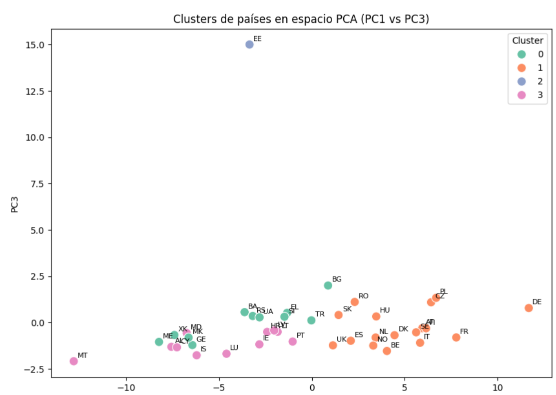

Fig. 7. Preliminary K-Means clusters ($k = 4$) projected onto PC1 vs. PC3 space.

Appendix 2: Main Panel Regression Results (Current and Five-Year Lagged Principal Components)

Table 3. Panel regression of GDP on current and lagged principal components (cluster-robust standard errors).

Component	**Coef.**	**Std. Err.**	**Significance**
PC1	0.155	0.055	**
PC2	0.215	0.095	*
PC3	-0.094	0.274	n.s.
PC4	-0.281	0.283	n.s.
PC5	0.028	0.206	n.s.
PC6	0.701	0.366	†
PC7	0.564	0.213	**
PC8	-0.869	0.842	n.s.
PC9	-0.033	0.368	n.s.
PC10	-0.414	0.784	n.s.
PC11	-0.129	0.559	n.s.
PC12	0.820	0.442	†
PC13	4.544	1.080	***
PC14	-0.679	1.206	n.s.
PC1_{lag5}	0.186	0.044	***
PC2_{lag5}	-0.146	0.087	†
PC3_{lag5}	-0.187	0.185	n.s.
PC4_{lag5}	-0.734	0.163	***
PC5_{lag5}	-0.280	0.187	n.s.
PC6_{lag5}	-0.533	0.359	n.s.
PC7_{lag5}	0.262	0.440	n.s.
PC8_{lag5}	-0.120	0.697	n.s.
PC9_{lag5}	-0.355	0.369	n.s.
PC10_{lag5}	-0.630	0.757	n.s.
PC11_{lag5}	1.004	0.540	†
PC12_{lag5}	-0.344	0.553	n.s.
PC13_{lag5}	-0.460	0.944	n.s.
PC14_{lag5}	0.923	0.866	n.s.
Within R^2	0.72		
Between R^2	0.02		
Overall R^2	0.03		
F-statistic (robust)	93.66, $p < 0.001$		

Notes: † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. n.s. = not significant.

Appendix 3: Cluster Membership by Country (2000–2020)

Table 4. Cluster assignments per country for the benchmark years 2000, 2010 and 2020.

Country	2000	2010	2020
AL	0	2	0
BE	1	1	3
BG	0	0	2
CY	0	0	1
CZ	2	2	0
DE	0	2	2
DK	2	0	1
EE	3	3	1
EL	0	0	1
ES	1	0	1
FI	2	0	1
FR	2	0	1
HR	0	0	1
HU	2	1	0
IE	0	0	1
IS	0	2	0
LT	0	0	1
LU	1	0	1
LV	0	0	1
MK	0	2	0
MT	1	0	1
NL	2	0	1
NO	2	0	1
PL	0	0	1
PT	0	0	1
RO	0	2	1
RS	0	2	0
SE	2	1	1
SI	0	0	1
SK	1	2	1
TR	0	2	0

Appendix 4: Dynamic Panel Regression Using SIEC Energy Variables

¡Perfecto! Aquí tienes el código LaTeX listo para copiar-pegar directamente en tu documento, con exactamente el mismo estilo bonito y profesional que la tabla anterior de componentes principales (alineación perfecta, significancia con símbolos, subíndices lag5 elegantes, notas claras, etc.). latex

Table 5. PanelOLS estimation results for SIEC-based dynamic model (current and five-year lagged energy variables).

Variable	**Coef.**	**Std. Err.**	**Signif.**
AFC_Bioenergía	0.0101	0.0127	
AFC_Carbón	0.0264	0.0073	***
AFC_Eólica	0.0018	0.0026	
AFC_Gas	-0.0014	0.0017	
AFC_Hidráulica	-0.0063	0.0024	**
AFC_Nuclear	0.0131	0.0260	
AFC_Petróleo	0.0055	0.0048	
AFC_Renovables_total	0.0014	0.0021	
AFC_Solar	0.0066	0.0028	*
GEP_Bioenergía	0.0067	0.0052	
GEP_Carbón	0.0089	0.0059	
GEP_Eólica	0.0132	0.0119	
GEP_Gas	-0.0001	0.0021	
GEP_Hidráulica	0.0038	0.0023	†
GEP_Nuclear	0.-train0088	0.0128	
GEP_Petróleo	0.0288	0.0165	†
GEP_Renovables_total	-0.0896	0052	
GEP_Solar	0.0047	0.0123	
AFC_Bioenergía$_{\text{lag5}}$	-0.0088	0.0163	
AFC_Carbón$_{\text{lag5}}$	-0.0069	0.0043	
AFC_Eólica$_{\text{lag5}}$	-0.0013	0.0043	
AFC_Gas$_{\text{lag5}}$	0.0022	0.0021	
AFC_Hidráulica$_{\text{lag5}}$	-0.0096	0.0035	**
AFC_Nuclear$_{\text{lag5}}$	0.0259	0.0113	*
AFC_Petróleo$_{\text{lag5}}$	0.0102	0.0051	*
AFC_Renovables_total$_{\text{lag5}}$	0.0020	0.0024	
AFC_Solar$_{\text{lag5}}$	0.0134	0.0061	*
GEP_Bioenergía$_{\text{lag5}}$	-0.0061	0.0056	
GEP_Carbón$_{\text{lag5}}$	-0.0018	0.0071	
GEP_Eólica$_{\text{lag5}}$	-0.0110	0.0138	
GEP_Gas$_{\text{lag5}}$	0.0014	0.0029	
GEP_Hidráulica$_{\text{lag5}}$	0.0085	0.0039	*
GEP_Nuclear$_{\text{lag5}}$	0.0000	0.0094	
GEP_Petróleo$_{\text{lag5}}$	-0.0006	0.0139	
GEP_Renovables_total$_{\text{lag5}}$	-0.2810	0.1864	
GEP_Solar$_{\text{lag5}}$	-0.0001	0.0116	
Within R^2	0.2550		
Between R^2	0.1697		
Overall R^2	0.1702		
Observations	655		
F-statistic	15.637 ($p < 0.001$)		

Notes: $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Empty cells = not significant.

Appendix 5: SIEC Cluster Membership by Country (2000–2020)

Table 6. Cluster assignments for each country under the SIEC-based clustering for 2000, 2010 and 2020.

AL	2	1	0
AT	1	0	2
BE	1	0	2
BG	0	3	1
CY	2	1	0
CZ	0	0	2
DE	1	0	2
DK	1	0	2
EE	3	2	3
EL	0	3	1
ES	1	0	2
FI	1	0	2
FR	1	0	2
HR	0	3	1
HU	0	0	2
IE	2	3	2
IS	2	1	0
IT	1	0	2
LT	2	1	0
LU	2	1	0
LV	2	1	0
MK	0	3	0
MT	2	1	0
NL	1	0	2
NO	1	0	2
PL	0	0	2
PT	1	0	2
RO	0	3	1
RS	0	3	1
SE	1	0	2
SI	0	3	1
SK	0	0	2
TR	0	3	1