High-Dimensional Forecasting Dynamics and Comovements in European Financial Markets through the FARM Model

Filippo Nardoni (0001172086) **Luca Danelli** (0001173490)
Federica Carrieri (0001176000) **Anita Scambia** (0001179942)

University of Bologna LM(EC)^2, Academic Year 2025/2026

Abstract

This paper investigates forecasting performance in high-dimensional European equity markets using daily data for 119 large-cap stocks across nine countries from 2003 to 2025. We compare a range of univariate, multivariate, and factor-based forecasting models—including AR, VAR, VAR–LASSO, PCR, AR–PCR, and FARM Predict—estimated under rolling-window schemes of different lengths. We follow (7), but adopt a block design scheme, using all available information to generate forecasts for all stocks under a given model configuration defined by the diffusion indexes, as proposed by (9). Forecast accuracy is evaluated using the Mean Squared Forecast Error (MSFE) and standard information criteria, allowing us to identify the optimal model configuration. Results show that the *FARM Predict* framework, which integrates factor extraction and sparse regression, consistently outperforms traditional VAR models and performs comparably to penalized specifications such as VAR–LASSO, particularly in highly integrated markets like Core Europe. Factor analysis reveals the dominance of a single pervasive component driving most of the cross-sectional variation, while residual dependence tests indicate that idiosyncratic correlations remain but become largely sparse once partial correlations are considered.

Introduction

Forecasting in high-dimensional financial markets requires models capable of extracting relevant information from large panels while accounting for strong cross-sectional dependencies. These challenges are particularly salient in Europe, where financial integration remains limited and national markets often respond heterogeneously to common shocks. As noted by Draghi (6), approximately 70% of local shocks in the United States are absorbed through integrated financial markets, compared to only 25% in the euro area, reflecting persistent fragmentation. Understanding how common and idiosyncratic components behave across European markets is therefore essential both for improving forecasting accuracy and for assessing the degree of financial integration.

Using daily data for 119 large-cap stocks from nine European countries over more than two decades, we construct a balanced panel of standardized returns and evaluate the forecasting performance of a wide set of models. Our analysis includes univariate (AR), multivariate (VAR, VAR–LASSO), and factor-based approaches (PCR, AR–PCR, and FARM Predict), all estimated under rolling-window schemes of varying lengths. Following Fan et al. (7), we implement the *FARM Predict* model, which integrates factor extraction and sparse regression to jointly capture systematic and idiosyncratic components. Unlike traditional applications, we adopt a **block design** framework—inspired by Stock and Watson (9)—where all available information is used to generate forecasts for the entire cross-section simultaneously. Under this design, hyperparameters such as lag length and the number of diffusion indexes are determined collectively, ensuring consistency and comparability across assets.

Beyond forecasting, we investigate the residual dependence structure that remains after extracting common factors. Using the novel covariance and partial covariance tests of Fan et al. (7), we evaluate whether the estimated idiosyncratic components are approximately independent or whether residual interdependencies persist. By applying these tests to both the raw residuals ($\hat{U}_t$) and their LASSO-adjusted counterparts ($\hat{V}_t$), we formally assess how much of the cross-sectional dependence is explained by common shocks versus local spillovers among European equities.

The remainder of the paper proceeds as follows. Section 1 presents the dataset and preliminary analysis, describing the selection of European stocks, the preprocessing steps, and the examination of cross-country

correlations that motivate the identification of regional market blocks. Section 2 outlines the theoretical foundations of the forecasting framework, introducing the use of information criteria for model selection and defining the Mean Squared Forecast Error (MSFE) as our main metric of predictive accuracy. Section 3 develops the empirical implementation, moving from univariate benchmarks to multivariate and factor-augmented models—AR, VAR, PCR, AR–PCR, VAR–LASSO, and FARM Predict—all estimated under rolling windows to assess out-of-sample performance. Section 4 analyzes the residual dependence structure after factor extraction using high-dimensional covariance and partial covariance tests, distinguishing between common and idiosyncratic shocks across European markets. Finally, Section 5 concludes by summarizing the main empirical results and discussing their implications for forecasting and financial integration in high-dimensional environments.

Taken together, these steps provide a unified empirical framework that links **forecasting performance** with the analysis of the **behavior of idiosyncratic components** across European financial markets.

1 Data description and Analysis

Our analysis relies on daily financial market data consisting of stock prices for a wide set of European firms from nine countries: Italy (11), Spain (21), France (22), Germany (22), Belgium (10), the Netherlands (10), Sweden (10), Denmark (10), and Norway (13). The data were obtained from *Yahoo Finance* using adjusted closing prices,[1] covering the period from 1 January 2003 to 1 January 2025. For each country, we selected the most capitalized and liquid firms listed on the main national exchanges, ensuring that our sample reflects the core components of each domestic market and broadly aligns with the composition of the EURO STOXX family indices. This guarantees that the panel captures the behavior of major market participants and the dynamics driving aggregate price movements.

During the preprocessing stage, where a larger initial pool of stocks was considered, all series with non-positive adjusted prices were removed, since these can arise from corporate revaluation or back-adjustment inconsistencies in the original source. To maintain consistency across assets and avoid unreliable return calculations, such series were excluded from the final dataset. The resulting balanced panel contains 5,667 trading days and 119 individual stock series, spanning multiple phases of market expansion and contraction.

Missing values were imputed to ensure continuity across time using a symmetric interpolation scheme based on adjacent prices:

$$\begin{cases} p_{i,1} = p_{i,2}, & \text{for } t = 1, \\ p_{i,t} = \dfrac{p_{i,t+1} + p_{i,t-1}}{2}, & \text{for } t \in (1, T), \\ p_{i,T} = p_{i,T-1}, & \text{for } t = T. \end{cases}$$

This procedure preserves the smoothness of each price trajectory while avoiding artificial jumps that could distort volatility estimates. After completing the imputation, all series were standardized to ensure comparability across assets and countries.

Returns were computed as simple percentage changes in prices:

$$r_{i,t} = \frac{p_{i,t+1} - p_{i,t}}{p_{i,t}}.$$

We use this formulation instead of logarithmic returns to prevent numerical distortions that may arise during episodes of sharp price swings—such as those observed in 2007–2008—when the log transformation could exaggerate extreme movements.

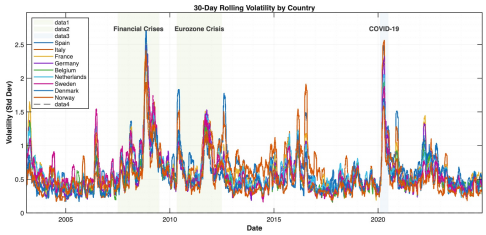
Figure 1a displays the evolution of daily standardized returns across all assets. The series appear visually stationary, a result confirmed by Augmented Dickey–Fuller (ADF) tests, which reject the null hypothesis of a unit root for all individual stocks ($z_t = 1$).[2]. Episodes of volatility clustering are clearly visible during major stress periods such as the *2007–2008 Global Financial Crisis*, the *Eurozone Sovereign Debt Crisis*, and the *COVID-19 pandemic*.

To better capture how market uncertainty evolves over time, we compute the *rolling volatility*, shown in Figure 1b. Rolling volatility measures the standard deviation of returns over a moving window—in

[1] Adjusted prices account for corporate actions such as stock splits, dividends, or other structural adjustments that affect valuation.

[2] Formally, $z_t := \sum_{i=1}^{119} h_i$, where $h_i = \mathbb{I}(p_{\text{ADF},i} < 0.05)$ is an indicator function equal to one if the ADF test rejects the null of a unit root for stock i at the 5% significance level.

(a) Daily standardized returns for all European stocks. Shaded areas correspond to major crisis periods.



(b) Thirty-day rolling volatility by country. Shaded areas indicate major financial crises.

Figure 1: Dynamics of returns and volatility across European financial markets.

our case, 30 consecutive trading days—providing a local estimate of how variable the market has been in the recent past. Higher values indicate phases of elevated uncertainty and stronger short-term price fluctuations, while lower values correspond to calmer, more stable market conditions.

Figure 1b highlights how volatility surged dramatically during crisis periods and gradually subsided in subsequent recoveries, confirming the presence of time-varying risk and common regional patterns in market behavior. Between 2003 and 2005, we observe relatively erratic yet low-magnitude movements, likely reflecting early integration dynamics among European markets before the mid-2000s credit expansion. To better visualize cross-country heterogeneity, returns are further disaggregated by country in Figure 2. The plot highlights differences in both the amplitude and persistence of volatility across European markets, with crisis periods (shaded areas) clearly corresponding to sharp increases in return variability.

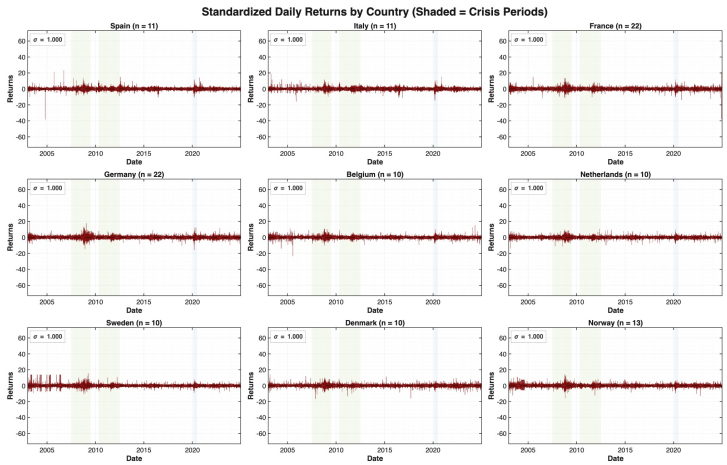


Figure 2: Standardized daily returns by country (shaded areas indicate crisis periods).

These heterogeneous volatility patterns suggest that cross-country interactions may also vary substantially over time and space. To investigate this dimension, we examine the correlation structure across all assets, providing a complementary perspective on how national markets co-move and cluster.

We first plot the correlation surface across all assets to detect potential structural patterns in the data. This preliminary step reveals several clusters of high correlations, suggesting that the cross-sectional dependence is far from uniform across European markets. Based on these patterns, we group countries into macro-areas that will later be used to evaluate and compare model performance:

- **Southern Europe:** Spain and Italy display strong domestic correlations and moderate bilateral dependence, forming a relatively cohesive sub-block characterized by similar market reactions and co-movements.
- **Core Europe:** France and Germany exhibit a more homogeneous and symmetric correlation structure, consistent with higher financial integration and stronger synchronization across assets.
- **Northern Europe:** The Netherlands and Sweden show intense within-country correlations but weaker inter-country links, indicating more localized dependence structures.

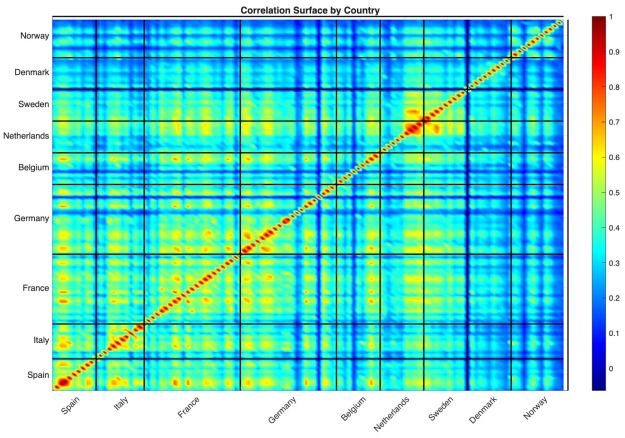


Figure 3: Cross-country correlation surface among standardized returns.

Although Southern and Core Europe appear more interconnected, forming a partially integrated block, the presence of strong within-country correlations in the South suggests that idiosyncratic components still play a relevant role. This aspect becomes crucial in the context of the FARM model, where such idiosyncratic dynamics are explicitly modeled alongside common factors.

Overall, the correlation surface partially confirms our initial hypothesis regarding the heterogeneity of European capital markets: they appear less integrated than the U.S. market and more influenced by country-specific dynamics. Nonetheless, beyond countries such as Denmark and Norway—which remain relatively weakly connected—there is still evidence of a set of latent common factors driving returns across European markets.

These findings are consistent with the existing literature on European financial integration, which highlights the persistence of regional clustering and heterogeneous degrees of market co-movement across the continent (see, among others, 2). In line with these studies, our results confirm that while common factors drive part of the cross-country dynamics, substantial idiosyncratic and region-specific components still characterize the European market structure.

2 Forecasting Theoretical Foundations

In our empirical analysis, we first fine-tune each model by minimizing the Mean Squared Forecast Error (MSFE) across alternative hyperparameter values, selecting the configuration that achieves the lowest MSFE. We then complement this analysis using information criteria—AIC and BIC—computed for each individual series. Following the approach of Stock and Watson (9), we identify the optimal lag length or number of diffusion indexes as the value most frequently selected across all equations, rather than by averaging AIC or BIC values. This procedure provides a robust, panel-consistent measure of model fit and parsimony, ensuring that the selected specification reflects the prevailing dynamics in the data. Since our framework estimates and forecasts all series simultaneously, we treat diffusion indexes as global hyperparameters. As shown by Stock and Watson (9), such diffusion-based specifications yield reliable forecasts under mild conditions, particularly at longer horizons, enhancing both model stability and interpretability.

2.1 MSFE

In order to evaluate the forecasting performance of the models presented in this work, we rely on the MSFE (Mean Square Forecasting Error), which allows us to assess the predictive accuracy of our models' estimates for the actual data. In our empirical setting, we call "MSFE" the one-day-ahead forecast estimation of the theoretical MSE.
This theoretical value is given by:

$$\mathrm{MSE} = \mathbb{E}_T\Big[\mathbb{E}_{(Y,X)}\big[(Y - \hat{f}_T(X))^2 \mid T\big]\Big] = \mathbb{E}_T[\mathrm{MSE}_T].$$

where T refers to the training sample, $\hat{f}_T$ is the function (model) estimated on T, X refers to the

test sample data and Y refers to the original actual data. The expected Test MSE embeds a double expectation: it averages across different out-of-sample draws of Y and X and also across different training samples T, which implies different estimates $\hat{f}_T$ on the test data X.

In our empirical setting the MSFE, namely the forecasted prediction of the theoretical MSE, is estimated through a rolling window procedure: at each step, the model is estimated on a window of length w (e.g., 150, 252, or 500 days), which constitutes the training sample, and then used to forecast the observation for the following day, which constitutes the test data. Thus, our empirical MSFE is computed as:

$$\widehat{\text{MSFE}} = \frac{1}{NH} \sum_{i=1}^{N} \sum_{h=1}^{H} \left(Y_{i,t+h} - \hat{f}_T(X_{i,t+h}) \right)^2 .$$

where N is the number of firms, H the number of forecasts (out-of-sample test periods), $Y_{i,t+h}$ the realized return, and $\hat{f}_T(X_{i,t+h})$ the corresponding forecast obtained from the model estimated on the training sample T. Hence, lower MSFE values indicate higher predictive accuracy. In our analysis, it is used to select the optimal lag length, window size and hyperparameters (depending on the model) to compare the forecasting performance of the AR, VAR, PCR, AR–PCR, and FARM models.

2.2 AIC and BIC

To determine the *admissible range of lag lengths* for our models, we rely on two widely used information criteria: the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). These criteria provide complementary benchmarks for balancing model fit and complexity:

$$\text{AIC} = -2 \log L + 2k,$$
$$\text{BIC} = -2 \log L + k \log T,$$

where L denotes the model likelihood, k is the number of estimated parameters, and T is the sample size. The first term rewards better fit, while the second penalizes over-parameterization. Because the penalty term in BIC increases more rapidly with T, it is generally more conservative than AIC and tends to favor more parsimonious models.

To illustrate the practical implications of these criteria, we estimate several models over the full sample and compute AIC and BIC values across a range of lag orders for each model. For each individual time series within a model, we identify the lag that minimizes AIC and the lag that minimizes BIC separately. We then summarize the results by reporting the frequency with which each lag is selected across all series, showing which lags are most commonly chosen by each criterion.

This highlights the general tendencies of the two measures: AIC, with its relatively mild penalty for model complexity, tends to select higher lag orders, while BIC, which imposes a stronger penalty that increases with sample size, favors simpler, more parsimonious models. These patterns align with the theoretical distinction between the two measures—AIC emphasizes fit, while BIC emphasizes simplicity.

Importantly, this procedure also informs our subsequent lag selection based on out-of-sample performance: the lags most frequently chosen by AIC and BIC provide a plausible range, which we then refine using the MSFE to select the lag lengths that optimize predictive accuracy.

2.3 Models

The models are implemented in increasing order of complexity, not as part of a horse-race comparison, but to sequentially incorporate additional structural features of the data. This progressive design aims to prevent overfitting and to ensure that each model extension captures new sources of predictability rather than noise. Our approach differs from (7) in that we estimate and fine-tune all models across the entire cross-section of assets, rather than focusing on a single representative stock i. Formally, each specification is estimated and evaluated for every stock $i \in N$, where $N = [1, 119]$.

2.3.1 AR

An autoregressive model AR(p) of order p expresses the current value of a time series as a linear combination of its own past p realizations. In our setting, we apply the AR(p) model to each firm i to predict its returns through the following relation:

$$\hat{y}_{i,t+1} = \hat{\phi}_{i,0} + \sum_{j=1}^{p} \hat{\phi}_{i,j}\, y_{i,t+1-j}.$$

where $y_{i,\hat{t+1}}$ denotes the one-step-ahead predicted return of firm i at time $t+1$ and $\phi_{i,0\ldots p}$ are OLS estimates. We estimate each firm's AR(p) model under a rolling-window scheme, then we select the optimal lag length p^* and window size w^* by comparing the corresponding MSFE values.

2.3.2 VAR and VAR-LASSO

A Vector Autoregressive model VAR(p) of order p generalizes the univariate AR model to a multivariate framework, allowing for interdependencies among multiple time series. The model is defined as:

$$\hat{\mathbf{y}}_{t+1} = \hat{\boldsymbol{\phi}}_0 + \sum_{j=1}^{p} \hat{\boldsymbol{\Phi}}_j\, \mathbf{y}_{t+1-j},$$

where $\mathbf{y}_t$ is the vector of the predicted returns of our 119 firms at time $t+1$, $\hat{\boldsymbol{\phi}}_0$, is a vector of OLS estimated intercepts, and $\hat{\boldsymbol{\Phi}}_j$ is the OLS matrix of OLS coefficients estimated equation by equation. As for the **AR** model, we use the resulting MSFE to make comparisons, even though we work in a more restrictive rolling-window and lag length framework due to computational limits.

Then we extend our model to a VAR-LASSO one by introducing an ℓ_1-norm penalty on the coefficients to perform shrinkage and allow for dimensionality reduction. The LASSO regularization ideally yields sparse estimated coefficient matrices, assessing the overfitting issue we often encounter in high-dimensional settings without any penalization. The parameters are estimated by minimizing the residual sum of squares and accounting for a dimensionality penalization, thus solving:

$$\hat{\boldsymbol{\phi}}_0, \hat{\boldsymbol{\Phi}}_j = \arg\min_{\boldsymbol{\phi}_0, \boldsymbol{\Phi}_j} \mathrm{RSS}(\boldsymbol{\Phi}_0, \ldots, \boldsymbol{\Phi}_p) + \lambda \sum_{j=1}^{p} \|\boldsymbol{\Phi}_j\|_1$$

where λ controls the degree of penalization and $\|\cdot\|_1$ denotes the element-wise ℓ_1 norm.

In order to make MSFE comparisons, we estimate the VAR-LASSO under the same rolling window and lag scheme of the VAR.

2.3.3 PCR and AR-PCR

Principal Component Regression (PCR) is a dimension-reduction technique that combines Principal Component Analysis (PCA) with linear regression. First, PCR identifies the orthogonal principal components through PCA, then it selects the most relevant ones — namely, those components (factors) that capture the largest variance in the predictors — and finally performs a OLS regression of the target variables (in our case, the 119 firms' returns) on these selected components. Even in this case, this method is helpful to avoid overfitting. We get our returns' PCR estimates via the following relation:

$$\hat{\mathbf{Y}}_{t+1} = \hat{\boldsymbol{\Pi}}_0 + \sum_{j=1}^{p} \hat{\boldsymbol{\Pi}}_j\, \hat{\mathbf{F}}_{t+1-j}$$

where $\hat{\mathbf{F}}_t$ is a k-dimensional vector of factors at time t obtained by the k selected principal components from PCA and the model parameters $\hat{\boldsymbol{\Pi}}_0$ and $\hat{\boldsymbol{\Pi}}_j$, are OLS estimates on p lags of $\hat{\mathbf{F}}_t$. We compute our forecasting predictions for the firms' returns through a rolling-window process. As before, we select the optimal number of lags p and window size w^* by comparing the relative MSFE values.

We then extend the Principal Component Regression (PCR) model with an AC-PCR model, which combines common factors selected through PCA with firm-specific autoregressive dynamics. Just like before, we first apply PCA to select the most relevant factors. Then, for each firm i, the model augments these factors with its own lagged returns, in order for the prediction to incorporate both common and

idiosyncratic sources of variation. Finally, an OLS regression is performed to estimate the coefficients associated with both the autoregressive and the factor components according to the following relation:

$$\hat{y}_{i,t+1} = \hat{\phi}_{i,0} + \sum_{j=1}^{p} \hat{\phi}_{i,j}\, y_{i,t+1-j} + \sum_{j=1}^{p} \hat{\boldsymbol{\pi}}'_{i,j}\, \hat{\mathbf{F}}_{t+1-j},$$

where $y_{i,\hat{t+1}}$ denotes the one-step-ahead predicted return of firm i at time $t+1$, $y_{i,t+1-j}$ are the firm-specific lagged returns, $\hat{\mathbf{F}}_t$ is the vector of k relevant factors obtained by PCA and $\hat{\phi}_{i,j}$ and $\hat{\pi}_{i,j}$ are OLS estimates for the respective coefficients. Even in this case we adopt a rolling-window framework and compare MSFE values to select the optimal lag lenght p^* and window size w^*.

2.3.4 FARM

The *FARM* framework combines the dense dimensional-reduced structure of principal–component factor models with the sparse dependence captured through LASSO. The construction begins by regressing each return series $Y_{i,t}$ on the corresponding set of observable predictors $X_{i,t}$, and by collecting the resulting partial residuals $R_{i,t} = Y_{i,t} - \gamma_i' X_{i,t}$. After removing the contribution of $X_{i,t}$, the covariance of the residual vector $R_t = (R_{1,t}, \ldots, R_{n,t})'$ is examined to determine whether substantial cross-sectional dependence remains. Whenever such dependence is present, the residuals are decomposed via principal component analysis,

$$\widehat{R}_t = \widehat{\Lambda}\, \widehat{F}_t + \widehat{U}_t,$$

where $\widehat{F}_t$ denotes the estimated common factors, $\widehat{\Lambda}$ the associated loadings, and $\widehat{U}_t$ the idiosyncratic component.

To capture the temporal dynamics of the latent structure, the estimated factors are projected onto a suitable vector of their own lags. Let $\widehat{G}_t = (\widehat{F}'_t, \widehat{F}'_{t-1}, \ldots, \widehat{F}'_{t-\ell_F})'$ be the collection of lagged factors. For each factor index r, the autoregressive representation

$$\widehat{F}_{r,t} = \widehat{\rho}'_r\, \widehat{G}_t$$

is estimated, and the collection of coefficients $\widehat{P} = (\widehat{\rho}_1, \ldots, \widehat{\rho}_r)'$ is then used to construct the one-step-ahead factor forecast $\widehat{F}_{t+1} = \widehat{P}\, \widehat{G}_t$.

The idiosyncratic component $\widehat{U}_t$ is subsequently inspected for additional cross-sectional dependence. When such dependence is statistically significant, a sparse representation is imposed by regressing each $\widehat{U}_{i,t}$ on a vector of contemporaneous and lagged residuals,

$$\widehat{U}_{i,t} = \theta_i'\, \widehat{W}_{i,t} + V_{i,t}, \qquad \widehat{W}_{i,t} = \big(\widehat{U}'_{-i,t}, \widehat{U}'_{t-1}, \ldots, \widehat{U}'_{t-\ell_U}\big)',$$

where LASSO regularization is used to enforce sparsity in the coefficients θ_i and to isolate the most relevant idiosyncratic interactions.

Finally, combining the projected factor dynamics with the sparse idiosyncratic component yields the one-step-ahead prediction for the return of asset i,

$$\widehat{Y}_{i,t} = \widehat{\gamma}_i' X_{i,t} + \widehat{\lambda}_i' \widehat{P}\, \widehat{G}_t + \widehat{\theta}_i' \widehat{W}_{i,t}.$$

The performance of the FARM–Predict model is evaluated through its mean squared forecast error (MSFE), with the forecasting window and lag orders aligned with those selected in the benchmark models to ensure comparability.

3 Forecasting Empirical Analysis

We start our empirical application by first extracting the common factors from the dataset, which serve as the foundation for the subsequent forecasting stage. After estimating the factors, we proceed by estimating several models with different levels of specificity. For each model, we iterate the estimation procedure over three different rolling window lengths: 150, 252, and 500 observations, whenever feasible (the *VAR* model due to high dimensionality, and the *FARM* model is restricted for computational reasons). In each iteration, we compute one-step-ahead squared prediction errors. Within each rolling window, we estimate the specific model using alternative settings, depending on the model's structure and dimensionality.

At the end of each iteration (for each rolling window size and model specification), we compute the corresponding Mean Squared Forecast Error (MSFE). This iterative procedure allows us to approximate the optimal combination of rolling window length and model structure by minimizing the MSFE. The optimal parameters are selected following (9) using Diffusion Indexes.

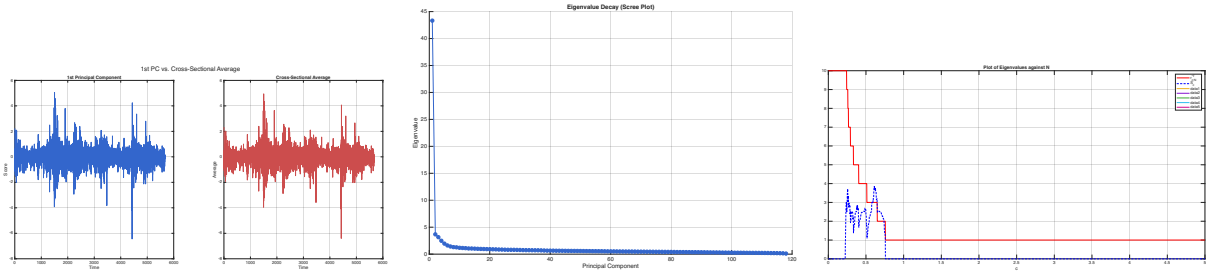
Once the optimal configuration has been identified, we re-estimate the corresponding model over the full available sample. The out-of-sample evaluation then begins right after the first rolling window, allowing us to compute one-step-ahead predictive errors for each subsequent period. This forecasting exercise is carried out for every variable in the dataset, and model performance is evaluated both at the level of individual variables and for groups of related assets.

3.1 Factor Extraction and Dimensionality Reduction

To extract the main sources of common variation across assets, we apply a **Singular Value Decomposition (SVD)** to the standardized data matrix $\mathbf{Y}$:

$$\mathbf{Y} = \mathbf{U}\mathbf{D}\mathbf{V}',$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices and $\mathbf{D}$ contains the singular values $d_1 \geq d_2 \geq \cdots \geq d_p \geq 0$. The first r components of $\mathbf{UD}$ capture the dominant directions of co-movement, defining the estimated factors $\widehat{\mathbf{F}} = \mathbf{U}_r\mathbf{D}_r$ and corresponding loadings $\widehat{\mathbf{L}} = \mathbf{V}_r$. These factors summarize the high-dimensional return space into a few latent variables that drive most of the common dynamics. They are then used as explanatory regressors in subsequent models—such as PCR, AR–PCR, and FARM—to capture both individual and cross-sectional dependencies efficiently.



(a) First PC vs. average. (b) Eigenvalue decay. (c) ABC criterion.

Figure 4: Principal component extraction and factor selection.

The first panel of Figure 4a compares the first principal component (PC1) with the cross-sectional average of standardized returns. The two series move almost identically, confirming that a single factor dominates the common variation across European equities. This interpretation is further supported by the scree plot (middle panel), where the first eigenvalue accounts for the vast majority of total variance, and the remaining eigenvalues decay sharply—evidence of a strong "market" component followed by relatively weak idiosyncratic structure.

To determine the number of latent factors, we apply the **Alessi–Barigozzi–Capasso (ABC) criterion** introduced by Alessi et al. (1), which refines the information criteria of Bai and Ng (3) through an adaptive penalization scheme. While the (3) criteria rely on a fixed penalty and can yield unstable estimates of r in finite samples, the ABC method introduces an adaptive constant c that scales the penalty according to the signal-to-noise ratio of the dataset. Formally, the modified information criterion is defined as

$$IC_a^*(k) = \log[V(k)] + c\,k\left(\frac{N+T}{NT}\right)\log(\psi_a) \tag{1}$$

where $V(k)$ is the residual variance from a k-factor model and ψ_a is a dimensional adjustment term. The constant c is varied over a grid of values, and the estimated number of factors $\widehat{r}_{c,N}^{T}$ is recorded for each configuration. The optimal number of factors is chosen as the value of r that remains stable across different penalty levels, ensuring robustness against over- or under-penalization.

In our dataset, both versions of the ABC estimator ($\widehat{r}_1$ and $\widehat{r}_2$) exhibit convergence to $r = 1$ across all values of c (see the right panel of Figure 4c). This consistency indicates the presence of a single pervasive **market-wide factor**, whose dominance persists regardless of the penalization level. Combined with

the eigenvalue decay and first principal component plots, these results confirm that the cross-sectional dynamics of European stock returns are primarily driven by one common component, reflecting strong market integration across countries.

3.2 Models

We now present the empirical results for all the models employed in our analysis. For each specification, we report the relevant tuning statistics and briefly discuss the corresponding outcomes. The section concludes with a comparative summary of all models, where we rank them according to their estimated predictive performance.

3.2.1 AR

We start by estimating a univariate Autoregressive model AR(p) for each of the $N = 119$ firms' returns in the dataset.

We evaluate the model under three alternative rolling window lengths w — 150, 252, and 500 observations — and, due to computational reasons, for lag orders p ranging from 4 to 10. At each iteration, the model is re-estimated over the training window and used to produce one-step-ahead forecasts for the subsequent observation. For every w and p combination, we compute the Mean Squared Forecast Error (MSFE) averaged across all firms.

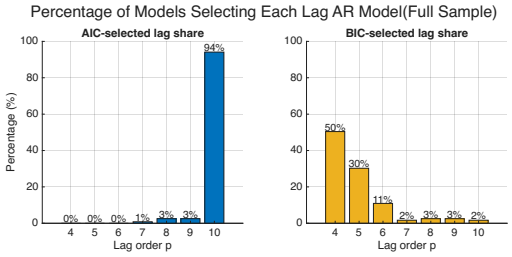


Figure 5: Share of models selecting each lag under AIC and BIC for the AR model.

Table 1: MSFE values for AR model

RW	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8	Lag 9	Lag 10
150	1.0169	1.0388	1.0358	1.0501	1.0632	1.0840	1.0824
252	1.0007	1.0125	1.0112	1.0209	1.0284	1.0413	1.0375
500	**0.9977**	1.0056	1.0026	1.0085	1.0140	1.0213	1.0184

Our results indicate that, as expected, BIC tends to select relatively low lag orders ($p = 4$) than the AIC ($p = 10$) since it penalizes more the number of parameters to avoid overfitting.

According to the MSFE comparison, we find that the optimal configuration corresponds to a rolling window of $w^* = 500$ and a a lag order of $p^* = 4$, which is consistent with those suggested by the AIC and BIC, since it is embedded in their range.

3.2.2 VAR and VAR LASSO

We estimate a classical Vector Autoregressive (VAR) model. In this case, the number of parameters to be estimated is considerably larger than in the AR model. Indeed, in a VAR(p) with k variables, the number of coefficients grows approximately as k^2p. This quickly inflates model dimensionality and may lead to overfitting or instability if the sample size is limited.

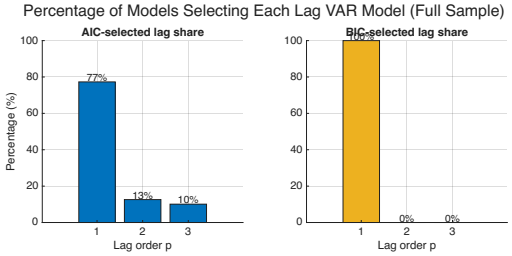


Figure 6: Share of models selecting each lag under AIC and BIC for the VAR model.

Table 2: MSFE values for VAR model

RW	Lag 1	Lag 2	Lag 3
500	**1.3825**	2.0637	3.8408

For this reason, we restrict our search for the near-optimal specification to a rolling window of 500 observations. This window size provides enough data for a stable estimation of the multivariate dynamics while keeping computational feasibility.

Within this window, we test three alternative model formulations including one, two, and three lags. Although, in principle, VAR models with fewer lags could also be estimated under smaller rolling windows (e.g., 150 or 252 observations), our previous results from the AR model already indicated that 500 observations minimize the MSFE on average. Therefore, we maintain this window size to preserve comparability and to allow a richer temporal structure to emerge.

Comparing the AIC, BIC, and MSFE values, we find that the specification with one lag performs best across all criteria. However, the MSFE values obtained for the VAR model are higher than those reported for the AR model. This difference can be attributed to three main factors:

- Due to data constraints, the VAR model was only evaluated under a single rolling window length (500), which may provide a less precise approximation of the best-performing model within a given class.
- Since no penalization or regularization was applied, the model may capture additional noise arising from cross-asset dependencies rather than systematic co-movements.
- Given the bigger amounts of parameters to estimate the VAR model can presents overfitting issues leading to greater variance across rolling estimates and a loss in predictive efficiency.

Overall, the VAR(1) specification emerges as the preferred one.

To mitigate these issues, we next estimate a VAR-LASSO model, which introduces an l_1-penalization term on the coefficient matrix. This regularization shrinks less relevant coefficients toward zero, effectively reducing the model's complexity while preserving key cross-variable dependencies. For computational consistency and comparability, we maintain the same rolling window length (500) and the same lag order (one) identified in the baseline VAR. Finally we fine tune the λ hyperparameter of the LASSO component by applying the 5 fold Time series Cross Validation, preserving the order of the data ensuring consistency of our analysis.

The resulting model shows a substantial improvement in predictive performance: the MSFE decreases to 0.985, indicating a remarkable gain relative to the unpenalized VAR (MSFE = 1.5091) and even outperforming the univariate AR benchmark. Hence, the introduction of l_1 regularization proves effective in addressing overfitting while also capturing additional, relevant interdependencies across the series that were missed by the AR model.

3.2.3 PCR

We next estimate the Principal Component Regression (PCR) model, The model is estimated under the same rolling-window framework ($w = 150, 252, 500$) and lag orders ($p = 4, ..., 10$).

By comparing information criteria across lag orders, we obtain the same optimal values founded for the AR model, namely $p = 4$ for the BIC and $p = 10$ for the AIC. When it comes to the MSFE, its lowest values is obtained for $p = 4$ (which belongs to the range found from the AIC and BIC) and $w = 500$, just like in the AR. Overall, the MSFE value of PCR (0.9976) is very similar to the one computed for the AR (0.9977): this very modest improvement suggests that the inclusion of common latent factors extracted through PCA provides a very small gain in predictive accuracy with respect to AR, which does a little poorly when it comes to overfitting. However, the best forecasting model so far is still the VAR-Lasso.

3.2.4 AR–PCR

We then estimate the Autoregressive Principal Component Regression (AR–PCR) model, which extends the PCR framework by including firm-specific autoregressive dynamics alongside the common factors extracted through PCA.

The model is evaluated under the same rolling-window framework ($w = 150, 252, 500$) and lag orders ($p = 4, ..., 10$) used for the AR and PCR benchmarks.

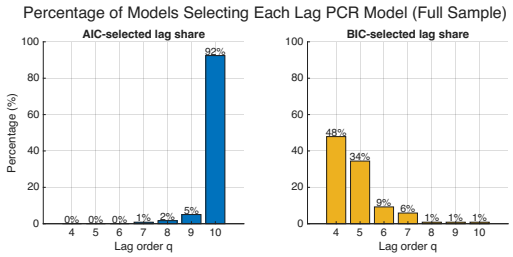


Figure 7: Share of models selecting each lag under AIC and BIC for the PCR model.

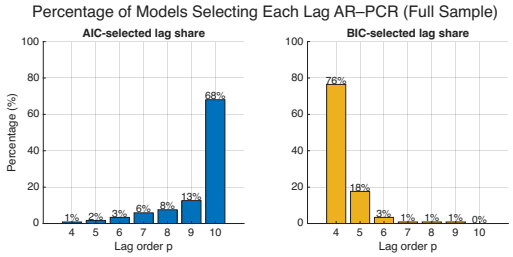


Figure 8: Share of models selecting each lag under AIC and BIC for the AR–PCR model.

Table 3: MSFE values for PCR model

RW	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8	Lag 9	Lag 10
150	1.0224	1.0331	1.0471	1.0594	1.0692	1.0855	1.0989
252	1.0030	1.0092	1.0173	1.0233	1.0288	1.0383	1.0467
500	**0.9976**	1.0005	1.0049	1.0079	1.0110	1.0166	1.0215

Table 4: MSFE values for AR–PCR model

RW	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8	Lag 9	Lag 10
150	1.0720	1.0999	1.1279	1.1568	1.1808	1.2148	1.2474
252	1.0294	1.0449	1.0595	1.0735	1.0849	1.1017	1.1178
500	**1.0100**	1.0170	1.0244	1.0309	1.0368	1.0454	1.0533

Comparing the AIC and BIC values across lag orders, once again we obtain the same optimal results as before, respectively $p = 4$ for BIC and $p = 10$ for the AIC.

When turning to the MSFE criterion, even in this case the best performance is achieved for $w = 500$ and $p = 4$, which is included in the AIC and BIC range.

The MSFE value (1.100) however, is a little higher than the one obtained for the AR and PCR separately. This could be due to overfitting, as the AR–PCR models implies a larger number of parameters with respect to the AR and PCR solely, but also to multicollinearity among the AR lags and PCA factors, further increasing estimation variance and slightly deteriorating out-of-sample forecast accuracy.

3.2.5 FARM PREDICT

The *FARM Predict* model allows us to jointly account for the presence of common factors and for the sparse dependence structure in the idiosyncratic components. After extracting the factors, each idiosyncratic component $U_{i,t}$ is regressed on the remaining components $U_{-i,t}$ using LASSO regularization, capturing the high-dimensional residual dependence across series. The LASSO hyperparameter λ is tuned using a time-series 3-fold cross-validation procedure (as in the VAR-LASSO case) to select the optimal value.

In our empirical implementation, differently from other model tuning specifications, forecasting is carried out using a rolling window of 252; the reason for such choice rely on the high computational costs needed to run the entire model. At each iteration we re-estimate the common factors and construct the predictor sets entering the LASSO step. The predictive information is split into two blocks:

- **Factor predictors:** we include 4 lags of the common factor (the first principal component), collected in
$$\widehat{G}_t = \left(\widehat{F}'_{1,t-1}, \widehat{F}'_{1,t-2}, \widehat{F}'_{1,t-3}, \widehat{F}'_{1,t-4}\right)'.$$
- **Idiosyncratic predictors:** we include 1 lag of all idiosyncratic components, gathered in
$$\widehat{W}_t = \left(\widehat{U}_{1,t-1}, \widehat{U}_{2,t-1}, \ldots, \widehat{U}_{N,t-1}\right)'.$$

For each residual equation, the LASSO penalty parameter λ is selected through a data-driven grid search over 30 logarithmically spaced values in the interval $[10^{-4}, 10^{1}]$. We adopt a three-block cross-validation scheme that minimizes the out-of-sample prediction error, ensuring that the tuning procedure adapts to the temporal structure encoded in both $\hat{G}_t$ and $\hat{W}_t$.

Throughout the entire procedure, the number of factors is fixed at $K = 1$, although this number could in principle vary across windows. This choice is consistent with the findings of Fan et al. (7) and with the empirical evidence presented in Section 4.1, where a single dominant factor is identified as the main driver of co-movement across European equity returns. After running the full procedure, we obtain an overall MSFE of 0.9935, which is relatively low compared to most competing models (only the VAR Lasso specification achieves a lower MSFE). While the results are encouraging, the gains from FARM Predict are somewhat limited by the high computational cost of the tuning process: a finer optimization of λ would likely require substantially greater computational resources to reach a near-optimal configuration.

3.3 Final Overview

Table 5: Comparison of Forecasting Models and Optimal Specifications

Model	Optimal Window	Optimal Lags	MSFE	MSFE Ratio	% Worse
AR	500	4	0.9977	1.0127	1.27
VAR	500	1	1.3825	1.4033	40.33
VAR–LASSO	500	1	**0.9852**	**1.0000**	**0.00**
PCR	500	4	0.9976	1.0126	1.26
AR–PCR	500	4	1.0100	1.0251	2.51
FARM	252	4	0.9935	1.0084	0.84

The results reported in Table 5 highlight a clear hierarchy in forecasting performance across the different model classes. Among all specifications, the *VAR–LASSO* achieves the lowest MSFE, setting the benchmark with an MSFE ratio of 1.000. This confirms the advantage of combining multivariate information with sparsity-inducing regularization, which effectively captures cross-sectional dependencies while mitigating overfitting. Traditional univariate and factor-based models such as AR and PCR perform comparably well, yielding MSFE ratios around 1.01, suggesting that much of the short-term dynamics can be explained by their parsimonious structures. The *FARM Predict* model also attains competitive results (MSFE = 0.9935), improving slightly over AR and PCR, though its gains remain modest relative to its computational cost. Finally, the unrestricted VAR performs substantially worse, with an MSFE roughly 40% higher than the benchmark, underscoring the inefficiency of dense parameterization in high-dimensional settings. Overall, these findings suggest that penalized and factor-augmented approaches strike the best balance between flexibility and parsimony, delivering robust forecasting performance in large cross-sectional panels.

4 Covariance Analysis

In this section, we examine the residual dependence structure that remains after extracting the common market factors. The goal is to determine whether the idiosyncratic components of returns are approximately independent—as implied by a correctly specified factor model—or whether additional sources of co-movement persist across European markets. To do so, we analyze both the covariance and partial covariance matrices of the estimated residuals, applying recent high-dimensional inference methods to formally test for remaining cross-sectional dependence. This analysis provides a diagnostic evaluation of the factor structure introduced in the previous section and helps identify whether local or regional shocks still drive the dynamics of European stock returns.

4.1 Theoretical Foundation

In high-dimensional settings, analyzing the correlation structure is fundamental for performing and assessing dimensionality reduction, which relies on linear transformations of the variables that capture most of the overall variability, and thus aim to summarize the idiosyncratic component of the data. In our project, we adopt the (partial) covariance extension of the high-dimensional hypothesis test proposed by (7), which extends the bootstrap-based inference framework of (8). These tests allow us to formally assess whether the covariance matrix exhibits a specific structure—most notably, whether its off-diagonal elements are zero—thereby providing inference on residual correlations among variables that remain after

factor extraction.

The main innovation of this framework is that it delivers reliable and consistent inference even when testing *partial* covariances, i.e., residual dependencies that remain after controlling for the effect of all other variables. When the factor structure is correctly specified and informative, the idiosyncratic covariance matrix is expected to be approximately sparse. The general estimator of the covariance structure is given by

$$\hat{\Sigma} := \frac{1}{T}\sum_{t=1}^{T} \hat{\mathbf{U}}_t \hat{\mathbf{U}}_t',$$

which provides an overview of inter-variable relationships. However, to measure the *direct* connection between two components $(U_{i,t}, U_{j,t})$, removing the contamination effect of the remaining components, we consider the *partial covariance*, defined as $\hat{V}_{i,j,t} := \hat{U}_{i,t} - \hat{\theta}_{i,j}' \hat{U}_{-ij,t}$ which corresponds to the residual of the LASSO regression of $\hat{U}_{i,t}$ onto $\hat{U}_{-ij,t}$.

$$\pi_{i,j} := \mathbb{E}[V_{i,j,t} V_{j,i,t}],$$

with empirical counterpart

$$\hat{\pi}_{i,j} := \frac{1}{T}\sum_{t=1}^{T} \hat{V}_{i,j,t} \hat{V}_{j,i,t}.$$

Based on these quantities, the null hypothesis for the covariance structure can be written as

$$\mathcal{H}_{0,D}^{\Sigma} : \Sigma_D = \Sigma_D^0, \qquad \text{or equivalently for partial covariances,} \qquad \mathcal{H}_{0,D}^{\Pi} : \Pi_D = \Pi_D^0,$$

where $D \subset [n]^2$ is a set of index pairs and $d = |D|$ may grow with the sample size T. The corresponding test statistic, belonging to the Kolmogorov–Smirnov family, is defined as

$$S_D^{\Sigma} := \left\| \sqrt{T}\,(\hat{\Sigma}_D - \Sigma_D^0) \right\|_{\max}, \qquad S_D^{\Pi} := \left\| \sqrt{T}\,(\hat{\Pi}_D - \Pi_D^0) \right\|_{\max}.$$

Since both $\hat{\Sigma}$ and $\hat{\Pi}$ are obtained from estimated idiosyncratic components rather than directly observed variables, the bootstrap procedure accounts for this estimation uncertainty. Critical values are derived through a Gaussian bootstrap approximation based on

$$\mathbf{Z}_{\Sigma}^{*} \mid \mathbf{X}, \mathbf{Y} \sim \mathcal{N}\big(\mathbf{0}, \hat{\mathcal{Y}}_{\Sigma}\big),$$

where $\hat{\mathcal{Y}}_{\Sigma}$ denotes a long-run covariance matrix estimated using a Newey–West–type approach:

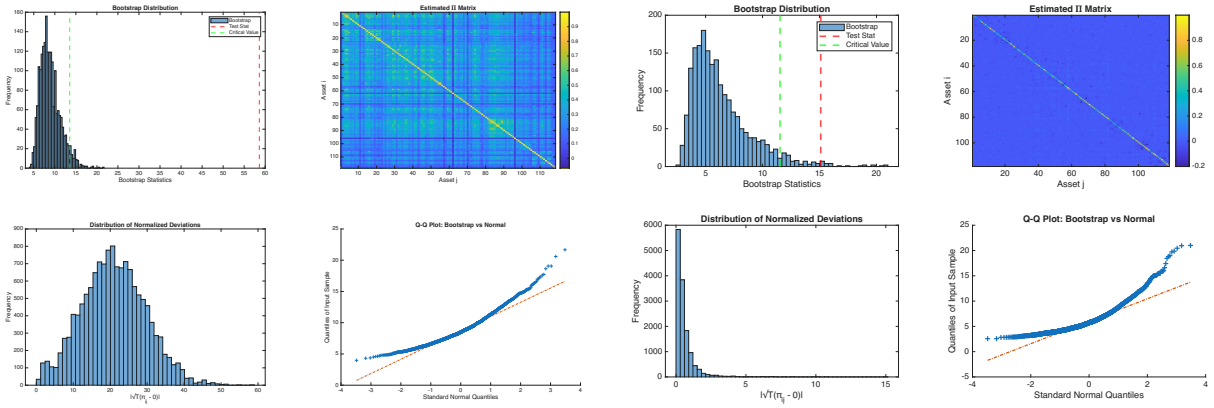
$$\hat{\mathcal{Y}}_{\Sigma} = \sum_{|b| \le B_T} K\!\left(\frac{b}{h}\right) \hat{M}_{\Sigma,b}, \qquad \hat{M}_{\Sigma,b} = \frac{1}{T}\sum_{t=b+1}^{T} \hat{D}_{\Sigma,t} \hat{D}_{\Sigma,t-b}'.$$

Here $\hat{D}_{\Sigma,t}$ collects the centered products $\hat{U}_{i,t}\hat{U}_{j,t} - \hat{\sigma}_{i,j}$ for $(i,j) \in D$. This construction ensures correct size control and consistent inference even when both the dimension n and the number of tested covariances d increase with T.

4.2 Empirical Foundings

We apply the high-dimensional covariance and partial-covariance tests introduced in the theoretical section to evaluate the residual dependence structure of the idiosyncratic components extracted from the FARM Predict model. Figure 9a reports the results for the covariance matrix of the raw residuals $\widehat{U}$, while Figure 9b presents the same diagnostics after the LASSO adjustment, based on the partial covariance matrix $\widehat{V}$.
In both tests, the statistic itself is computed from the sample covariance (or partial covariance) matrix, while the bootstrap is used solely to approximate its null distribution and to obtain critical values under the hypothesis of weak cross-sectional dependence ($B = 2000$ bootstrap repetitions). This ensures valid inference even in high-dimensional settings where the theoretical limit distribution is intractable.

In panel (a), the top-left histogram shows the bootstrap distribution of the test statistic under the null hypothesis, with the vertical dashed lines marking the observed value and critical thresholds. The



(a) Correlation matrix of residuals $\widehat{U}$

(b) Correlation matrix of LASSO residuals $\widehat{V}$

Figure 9: Comparison of residual correlation structures before and after LASSO adjustment. Each panel shows, clockwise from top left: (i) the bootstrap null distribution of the test statistic with the observed value and critical thresholds, (ii) the estimated correlation matrix, (iii) the Q–Q plot comparing bootstrap and normal quantiles, and (iv) the distribution of normalized deviations $|\sqrt{T}(\hat{\Sigma}_{ij} - \Sigma^0_{ij})|$.

top-right heatmap displays the estimated covariance matrix $\widehat{\Sigma}$, where distinct blocks of positive correlations reveal residual interdependencies among assets. The lower-left plot illustrates the distribution of normalized deviations, which remains wide and skewed, while the Q–Q plot (lower-right) exhibits heavy tails relative to the normal benchmark. Together, with results from 7 we confirm that, even after removing the common factor, substantial cross-sectional dependence persists among the idiosyncratic components; in fact while the test statistic reports a value of $S^{\Sigma}_{D} = 58.6842$, the critical value lies while below $c_{0.05} = 13.5845$ therefore prompting us to reject H_0.

Table 6: Covariance Test Results on $\hat{V}$

Test statistic	15.1394
Critical value (5%)	11.3324
P-value	0.0095
Decision	**REJECT** H_0

Table 7: Covariance Test Results on $\hat{U}$

Test statistic	58.6842
Critical value (5%)	13.5845
P-value	0.0000
Decision	**REJECT** H_0

Panel (b) shows the same diagnostics for the LASSO-adjusted residuals $\widehat{V}$. Here, the estimated matrix $\widehat{\Pi}$ becomes nearly diagonal, indicating that most residual correlations vanish once conditional dependencies are partialled out. The bootstrap distribution of the test statistic aligns less well with the theoretical normal distribution than in the previous case. Its heavier tails lead to lower estimated critical values, which in turn increases the likelihood that the bootstrap critical value lies closer to the observed statistic, potentially distorting inference. We can confirm our analysis by looking at the results from 6 which show a value for the test statistic of $S^{\Sigma}_{D} = 15.13494$, which is very close to the critical value $c_{0.05} = 11.3324$. This confirms that the sparse regression step effectively removes the remaining cross-sectional structure, yielding an approximately independent set of idiosyncratic components.

Overall, the transition from $\widehat{U}$ to $\widehat{V}$ highlights the central contribution of the FARM Predict framework: by combining factor extraction with LASSO-based adjustment, it isolates the dominant common components while enforcing sparsity in the residual dependence network. The near-diagonal structure of $\widehat{V}$ thus provides strong evidence that the remaining idiosyncratic shocks are largely uncorrelated once common effects are controlled for.

5 Evaluation of Predictions

To evaluate the predictive performance of the FARM model relative to the other estimated models, we adopt two complementary perspectives. First, we examine forecasting accuracy for a single representative asset—specifically, the most capitalized Italian stock, ENI. Second, we extend the same analysis

to the most capitalized firm in each of the countries included in our dataset. Finally, we assess predictive performance at an aggregated level, grouping assets according to the three macro-areas previously identified: Southern, Core, and Northern Europe.

In particular, when evaluating performance at the aggregated level, we compute the ratio of cumulative Mean Squared Forecast Errors (MSFEs). By placing the MSFE of the FARM model in the numerator, this ratio provides a direct measure of relative efficiency: values greater than one indicate that the FARM model produces higher forecast errors than the benchmark, whereas values below one suggest superior predictive accuracy.

This comparison allows us to assess:

- whether, on average, the *FARM* model outperforms more traditional specifications;
- under which conditions and time periods the *FARM* model delivers more accurate predictions.

Furthermore, analyzing these ratios at the aggregated (regional) level makes it possible to identify whether the *FARM* model captures information more effectively in specific geographical or economic contexts—for instance, in highly integrated markets such as the Core area, or in more segmented environments like Northern Europe.

5.0.1 Forecasting ENI stock returns

We start by analyzing the forecasting performance for the **Eni** company. Figure 10 reports the out-of-sample predictions over the period 1 January 2005 to 1 January 2025, comparing the realized and predicted returns across all models considered.

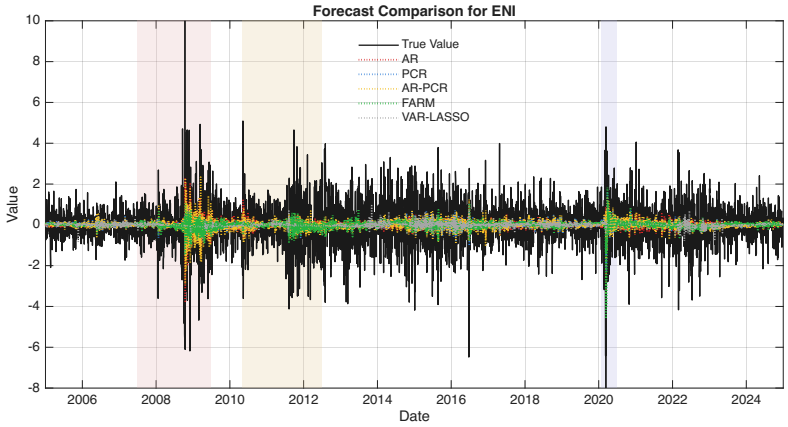


Figure 10: Forecast comparison for ENI stock returns. Shaded areas correspond to major crisis periods.

As shown in Figure 10, the actual return series (in black) exhibits pronounced fluctuations during the 2008–2009 financial crisis and the 2020 pandemic episode, consistent with sharp increases in market volatility. Simpler models such as AR and PCR display smoother trajectories that tend to underreact to these extreme events, generating forecasts that lag behind the true dynamics. The AR–PCR specification improves slightly by incorporating both autoregressive and principal component features, yet it still fails to fully capture large shocks.

In contrast, the *Farm Predict* model (in green) shows a markedly closer alignment with the realized series, especially during turbulent periods. This improved fit highlights the model's ability to integrate common factors extracted from the cross-section of stocks, which enhance its responsiveness to systemic market movements. The VAR-LASSO, although flexible, produces slightly noisier predictions and occasionally overshoots actual values.

Overall, the visual evidence supports the superior adaptability of the *Farm Predict* approach in capturing both tranquil and volatile phases of the market, offering forecasts that are not only more accurate but also more robust to structural changes in financial conditions.

5.0.2 Forecasting High-Cap Companies by Country

Extending our exercise on the high-cap companies' stock returns, for each country we consider the largest and most representative stock: *Telefónica S.A.* (Spain), *Intesa Sanpaolo* (Italy), *BNP Paribas*

(France), *Volkswagen Group* (Germany), *Lotus Bakeries* (Belgium), *SBM Offshore* (Netherlands), *Assa Abloy* (Sweden), *Vestas Wind Systems* (Denmark), and *Mowi ASA* (Norway). The resulting forecasts are displayed in Figure 11.

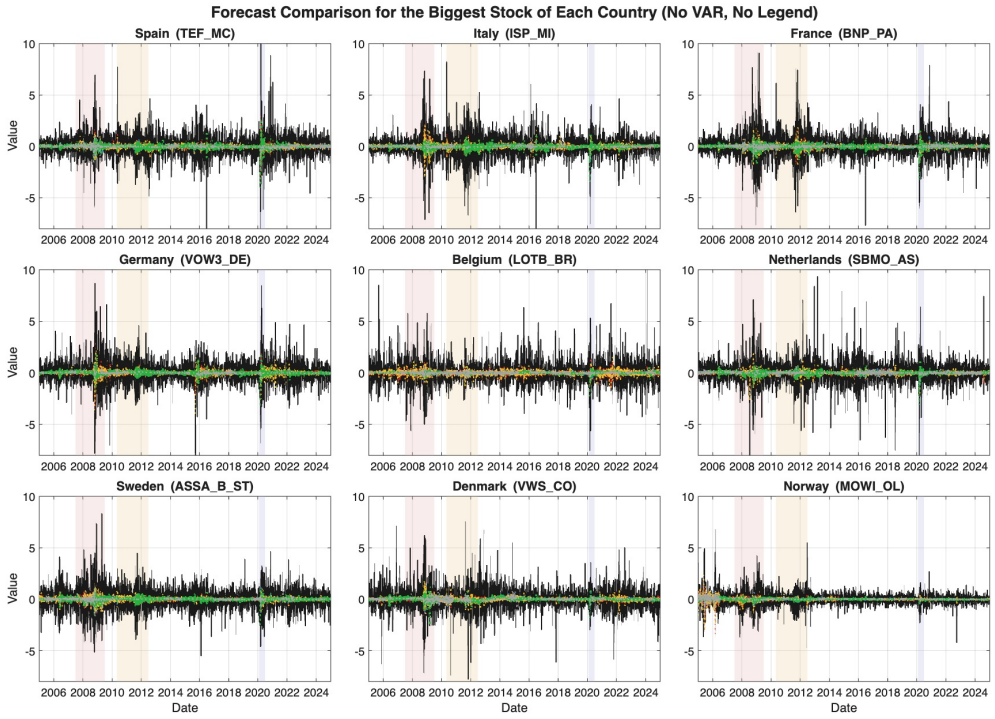


Figure 11: Forecast comparison for the largest stock in each country. The figure shows actual returns (black) and model forecasts (colored lines) across European regions, with shaded areas indicating major financial crises.

This analysis provides a useful perspective on the regional heterogeneity of model performance across Europe. We observe that the *FARM Predict* model systematically outperforms alternative forecasting approaches for stocks from countries belonging to Southern Europe (Spain, Italy), Core Europe (France, Germany), and Northern Europe (Sweden, the Netherlands). This pattern suggests the presence of a dominant common driver—interpretable as a broad market factor—that captures most of the co-movements in returns within each regional block. In other words, the factor structure appears particularly effective where financial markets are more integrated, consistent with the hypothesis that these regional groups share similar exposure to aggregate European shocks and global risk components.

For stocks of out-of-block countries such as Belgium, Denmark, and Norway, we observe that simpler autoregressive (AR) or sparse models tend to perform comparably or even slightly better than the factor-based approach. This may reflect more idiosyncratic return dynamics or weaker linkage to the common European market factor, suggesting that integration intensity varies across the continent.

5.1 Evaluation of Local Financial Markets

When we evaluate for geografical area individuated before, we can observe that the *FARM* model consistently outperforms the traditional *VAR* throughout the entire period considered. This result is not surprising, given the previously discussed evidence on the limited adaptability of *VAR* models to high-dimensional data structures.

When comparing *FARM* to the other benchmarks, the interpretation becomes less straightforward, and some heterogeneity emerges across geographical areas.

Focusing on **Southern Europe** (Figure 12), *FARM* initially accumulates a higher forecasting error compared to the classical models, but it soon stabilizes around a ratio close to one. Over time, the *VAR-LASSO* appears to deliver the lowest cumulative MSFE, confirming its superior flexibility in capturing both temporal and cross-sectional dependencies while controlling for dimensionality.

The *FARM* model performs slightly better than the *AR* model, suggesting that it captures integration components that a purely autoregressive structure fails to represent. From the end of the financial crisis

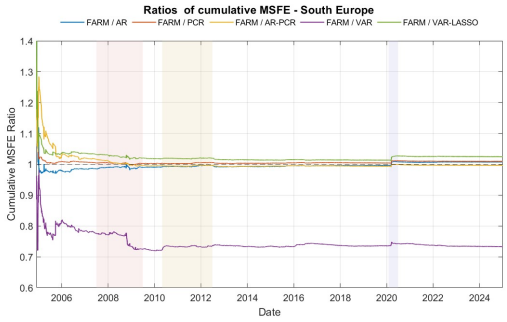


Figure 12: Ratios of cumulative MSFEs (Southern Europe)

onward, the cumulative ratio between *FARM* and *AR-PCR* slightly decreases, indicating that *FARM* accumulates marginally fewer forecasting errors than the *AR-PCR* model. However, its performance remains very close to that of the *PCR* model. This similarity indicates that, in this setup, *FARM* does not seem to fully exploit the idiosyncratic component, behaving instead like a standard principal component model.

This outcome may stem from the rigid specification we adopted for cross-area comparability: the *FARM* structure was estimated jointly on the entire dataset, without re-optimizing its parameters for each variable or area (9). Consequently, the model sacrifices some flexibility compared with more adaptive approaches such as *VAR-LASSO*, which are better suited to local dynamics and cross-sectional heterogeneity.

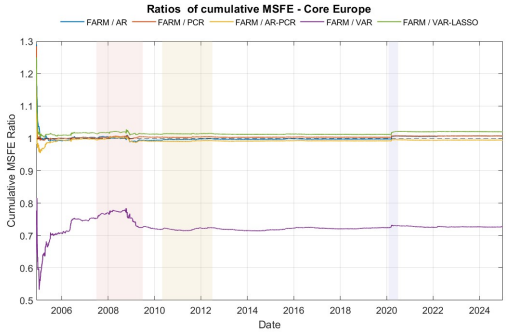


Figure 13: Ratios of cumulative MSFEs (Core Europe)

In the **Core Europe** area, the results are overall consistent with those previously observed. The *FARM* model systematically outperforms the traditional *VAR*, while the *VAR-LASSO* remains the best-performing approach across all alternatives. The *FARM* behaves similarly to the *AR* (slightly better), the *PCR* (slightly worse), and the *AR-PCR* (slightly better, from 2008 onward), suggesting that its predictive ability lies between these benchmark models.

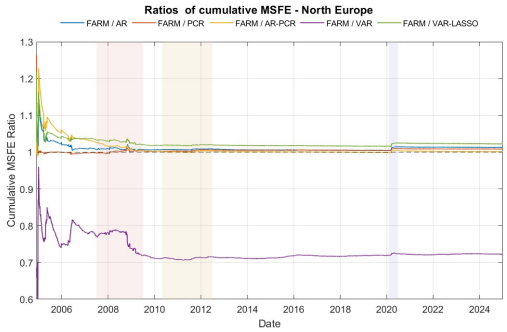


Figure 14: Ratios of cumulative MSFEs (Northern Europe)

In the **Northern Europe** area, the same general patterns hold. However, in this case, the *AR* model tends to outperform *FARM* for most of the sample period. This finding is coherent with the results discussed in the correlation analysis: in this region, the within-country correlations are stronger, meaning that local autoregressive dynamics explain a larger portion of the variance. As a result, the common-factor structure of *FARM* may add limited information beyond what is already captured by the own-lag terms of each variable.

It is important to stress that this comparison method - based on cumulative ratios - is not fully informative about the relative explanatory capacity of *FARM* with respect to the other models. Since the ratio accumulates errors over time, the fact that *FARM* starts with higher MSFE mechanically affects the subsequent path. Thus, even when *FARM* performs better during specific periods, the cumulative ratio may not fall below one simply because past accumulated errors continue to dominate.

For instance, during the financial crisis in Southern and Core Europe, the *FARM* ratio decreases noticeably, indicating that in this period *FARM* accumulates fewer forecasting errors than the competing models. The fact that the ratio does not drop below one does not imply that *FARM* performs worse; rather, it reflects the weight of earlier periods in which *FARM* exhibited higher errors. Indeed, as shown in the single–asset forecasting exercises (e.g., the ENI stock and the representative assets for each region), *FARM* often outperforms *VAR-LASSO* precisely during turbulent periods, suggesting that the model captures crisis-specific common dynamics more effectively.

Therefore, when interpreting these results, it is crucial to keep this feature in mind. Since we are evaluating the *mean* behaviour of variables aggregated by area, the cumulative-ratio metric - although limited - remains the most coherent method for visual comparison across regions.

Additionally, in Northern Europe, unlike the other areas, the *FARM* ratio increases during the 2008 crisis. This suggests that the common-factor structure contributes less to predictive accuracy in this region under high volatility, reinforcing the evidence that Northern markets are more strongly driven by idiosyncratic, country-specific autoregressive dynamics rather than by shared common components.

6 Conclusion

This paper provides a comprehensive evaluation of forecasting models for high-dimensional European equity markets, emphasizing the role of common factors and sparse dependence structures in improving predictive performance. Using daily data for 119 large-cap stocks from nine European countries over the period 2003–2025, we compared traditional time-series approaches (AR, VAR), penalized regressions (VAR–LASSO), and factor-augmented frameworks (PCR, AR–PCR, and FARM Predict) within a unified rolling-window setup.

Our empirical findings reveal that the *FARM Predict* model, which integrates factor extraction and sparse regression, achieves robust predictive accuracy across all regional groups. It systematically outperforms traditional *VAR* specifications and performs comparably to the *VAR–LASSO*, particularly in the more financially integrated Core European markets. Factor analysis confirms the dominance of a single pervasive market-wide component, while covariance and partial-covariance diagnostics indicate that residual dependencies become largely sparse once common factors are removed, although they remain statistically significant. By applying the novel covariance test proposed by Fan et al. (7), we show that the residuals terms still exhibit non-negligible co-movements. After performing the partial-covariance test using LASSO-adjusted residuals, we find stronger evidence of idiosyncratic components in the residual structure.

However, the analysis is subject to certain limitations. First, the dataset includes only a subset of the most capitalized firms from a limited number of European countries, which may bias results toward larger and more liquid markets. Second, in the regional grouping, some macro-areas (such as Southern Europe) contain more firms than others, potentially affecting cross-area comparability. Future research could address these limitations by expanding the sample to include a broader set of firms and countries, and by adopting a dynamic high-dimensional framework such as the *FNETS* model of Barigozzi et al. (4), which explicitly accounts for time-varying network interactions among assets and evolving factor structures.

Even without a deep analysis, we can observe an interesting pattern in European financial markets: central European countries appear to form a highly integrated capital market, whereas northern

European countries seem less integrated, possibly because they are more exposed to country-specific factors.However, it is worth noting that additional methods should be considered when analysing financial market connectedness. In particular, the framework proposed by (Barigozzi et al.) offers a valuable extension, as it decomposes financial shocks into local, regional, and market-wide components.

Overall, the study demonstrates that high-dimensional forecasting frameworks integrating common factors and sparse dependencies provide valuable tools for understanding and predicting the complex dynamics of European financial markets, while pointing toward promising avenues for further methodological development.

References

[1] Alessi, L., Barigozzi, M., and Capasso, M. (2008). A robust criterion for determining the number of static factors in approximate factor models. Working Paper Series 903, European Central Bank, Frankfurt am Main.

[2] Baele, L., Ferrando, A., Hördahl, P., Krylova, E., and Monnet, C. (2004). Measuring financial integration in the euro area. *Oxford Review of Economic Policy*.

[3] Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*.

[4] Barigozzi, M., Cho, H., and Owens, D. (2024). Fnets: Factor-adjusted network estimation and forecasting for high-dimensional time series. *Journal of Business & Economic Statistics*.

[Barigozzi et al.] Barigozzi, M., Hallin, M., and Soccorsi, S. Identification of global and local shocks in international financial markets via general dynamic factor models. *Journal of Financial Econometrics*.

[6] Draghi, M. (2018). Speech at hertie school of government. Speech.

[7] Fan, J., Masini, R. P., and Medeiros, M. C. (2023). Bridging factor and sparse models. *Journal of Business & Economic Statistics*.

[8] Giessing, A. and Fan, J. (2023). A bootstrap hypothesis test for high-dimensional mean vectors.

[9] Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*.