# Time Varying Factor Model: An Empirical Application to the European Area

Gloria Carloni, Luca Marchesi, Luca Orlando, Tommaso Zipoli[1]

[1]University of Bologna

November 14, 2025

**Abstract**

This paper applies the time-varying factor model proposed by Su and Wang (2017) to a large European macro-financial dataset in order to explore how the structure and transmission of common forces in the euro area have evolved over time. By allowing factor loadings to change smoothly, the model captures gradual shifts in the underlying economic and financial relationships. We investigate both the dynamics and the dimensionality of the common component, showing that the number and strength of latent factors are not constant but vary across periods. In particular, the analysis highlights how the global financial crisis of 2008 and the COVID-19 pandemic in 2020 reshaped the configuration of euro-area factors. The estimated loadings reveal clear changes in the way macroeconomic and financial variables co-move during these major shocks, indicating that the structure of the European business and financial cycles is itself time-dependent.

## 1 Introduction

Traditional static factor models, such as those developed by Stock and Watson (2002) and Bai and Ng (2002), assume that the relationships between observable variables and latent factors remain constant over time. While this assumption simplifies estimation, it may be too restrictive when data are high-dimensional or span long periods of time, potentially encompassing structural changes. In particular, the European economy has undergone major transformations over the past two decades, including the global financial crisis, the sovereign debt crisis, and the COVID-19 pandemic. These events likely altered the way in which common forces drive macroeconomic and financial variables across countries and sectors.

We adopt the *time-varying factor model* proposed by Su and Wang (2017), which extends the standard framework by allowing factor loadings to evolve smoothly over time. This approach combines kernel weighting and local principal component analysis (PCA) to estimate loadings and factors within a moving neighborhood around each point in time. By doing so, it captures gradual shifts in the structure of co-movements that a static model would fail to detect. Su and Wang also propose a statistical test to verify whether the assumption of time-invariant loadings can be rejected, providing a formal assessment of the presence of time variation.

We apply this methodology to a large European dataset consisting of 47 macro-financial variables and nearly 300 monthly observations from 2000 to 2023. The database covers a wide range of indicators—including national accounts, unemployment, credit, price, interest rates, and monetary aggregates — compiled from institutional sources such as Eurostat, the European Central Bank, and the OECD. All series are transformed and standardized to ensure stationarity and comparability across units.

As an extension of the original framework, we allow the number of factors to vary across time and select it locally using the ABC criterion of Alessi (2010). This flexible specification enables us

to assess both the evolving strength and dimensionality of the common component. The formal test provided by Su and Wang (2017) strongly rejects the null hypothesis of constant loadings ($p < 0.01$), indicating that the underlying relationships among variables change significantly over time.

Our analysis focuses on the dynamics of the first two factors. The results reveal substantial time variation in loadings, particularly around the 2008 global financial crisis and the 2020 COVID-19 shock. These findings suggest that the transmission mechanisms of macro-financial fluctuations in the Euro Area are not static over time but are very much responsive of changing economic regimes, monetary conditions and policy.

To conclude, we investigate the presence of nonlinear relationships using a frontier approach drawn from deep learning—an autoencoder factor model. Although our findings remain exploratory rather than conclusive given the constraints imposed by the dataset's limited dimensionality, this exercise opens a promising line of inquiry for nonlinear forecasting and modeling.

# 2 Methodology

We follow the framework of a *time-varying factor model* as introduced by Su and Wang (2017), which generalizes the conventional static factor model by allowing the factor loadings to evolve smoothly over time.

## 2.1 Model setup

Let $X_t = (X_{1t}, X_{2t}, \ldots, X_{Nt})'$ denote an $N \times 1$ vector of observed variables at time $t$, for $t = 1, \ldots, T$. The model assumes that each observable series can be decomposed as

$$X_{it} = \lambda_{it}' F_t + e_{it}, \qquad i = 1, \ldots, N, \tag{2.1}$$

where:

- $F_t = (F_{1t}, F_{2t}, \ldots, F_{Rt})'$ is an $R \times 1$ vector of latent *common factors*;
- $\lambda_{it} = (\lambda_{i1t}, \ldots, \lambda_{iRt})'$ is an $R \times 1$ vector of *factor loadings*, varying smoothly with time;
- $e_{it}$ is an idiosyncratic error term, weakly dependent across both $i$ and $t$.

In matrix form, for each period $t$:

$$X_t = \Lambda_t F_t + e_t, \tag{2.2}$$

where:

$$\Lambda_t = \begin{bmatrix} \lambda_{11t} & \lambda_{12t} & \cdots & \lambda_{1Rt} \\ \lambda_{21t} & \lambda_{22t} & \cdots & \lambda_{2Rt} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{N1t} & \lambda_{N2t} & \cdots & \lambda_{NRt} \end{bmatrix}_{N \times R}, \quad F_t = \begin{bmatrix} F_{1t} \\ F_{2t} \\ \vdots \\ F_{Rt} \end{bmatrix}_{R \times 1}, \quad e_t = \begin{bmatrix} e_{1t} \\ e_{2t} \\ \vdots \\ e_{Nt} \end{bmatrix}_{N \times 1}.$$

Stacking all $T$ observations, the full dataset can be written as:

$$X = F\Lambda' + E, \tag{2.3}$$

where:

$$X = \begin{bmatrix} X_1' \\ X_2' \\ \vdots \\ X_T' \end{bmatrix}_{T \times N}, \quad F = \begin{bmatrix} F_1' \\ F_2' \\ \vdots \\ F_T' \end{bmatrix}_{T \times R}, \quad \Lambda' = \begin{bmatrix} \lambda_{11} & \lambda_{21} & \cdots & \lambda_{N1} \\ \lambda_{12} & \lambda_{22} & \cdots & \lambda_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{1R} & \lambda_{2R} & \cdots & \lambda_{NR} \end{bmatrix}_{R \times N}, \quad E = \begin{bmatrix} e_1' \\ e_2' \\ \vdots \\ e_T' \end{bmatrix}_{T \times N}.$$

### 2.2 Time-varying structure

To capture gradual structural changes, the loadings are modeled as smooth functions of rescaled time:

$$\lambda_{it} = \lambda_i(\tau_t), \quad \text{where } \tau_t = \frac{t}{T}, \quad \tau_t \in (0, 1]. \tag{2.4}$$

Hence, each $\lambda_i(\cdot)$ represents a smooth curve describing how the loading of variable $i$ on each factor evolves throughout the sample.

This specification allows the model to account for long-term transitions—such as technological progress, financial integration, or policy regime shifts—without imposing discrete breaks.

### 2.3 Identification

Similarly to the conventional version of a factor model, the loadings and factors are not independently identifiable by the estimation procedure—i.e. we have $\lambda_{it}'F_t = (H_t^{-1}\lambda_{it})'(H_t'F_t)$ for any $R$x$R$ nonsingular matrix $H_t$—unless imposing $R^2$ restrictions. Following Bai and Ng (2002), we impose normalization restrictions to remove rotational indeterminacy:

$$\frac{1}{T}F'F = I_R, \qquad \frac{1}{N}\Lambda_t'\Lambda_t = \text{diag}(c_1(\tau_t), \ldots, c_R(\tau_t)), \tag{2.5}$$

where $c_r(\tau_t)$ denotes the time-varying strength of the $r$-th factor at time $\tau_t$.

These constraints ensure that the scale and rotation of factors and loadings are uniquely identified at each point in time.

### 2.4 Objective

Within this framework, the main goal is to describe how the *common component*:

$$\chi_{it} = \lambda_{it}'F_t$$

evolves over time, and to determine whether changes in the factor loadings represent genuine structural shifts in the data-generating process or gradual regime adaptations.

The subsequent section details the estimation of the local factors and loadings using a kernel-weighted version of principal component analysis (PCA), as proposed by Su and Wang (2017).

## 3 Estimation

In this section, we estimate the time-varying factor loadings and common factors by implementing a local version of principal component analysis (PCA), following the approach proposed by Su and Wang (2017). The underlying assumption is that the factor loadings $\lambda_i(\cdot)$ are smooth functions of time, so that within a small local neighborhood around $\tau = r/T$ they can be approximated by constants. Accordingly, the high-dimensional factor model

$$X_{it} = \lambda_{it}'F_t + e_{it}, \quad i = 1, \ldots, N, \ t = 1, \ldots, T, \tag{3.1}$$

can be locally approximated by

$$X_{it} \approx \lambda_{ir}'F_t + e_{it}, \qquad \text{for } t \approx r, \tag{3.2}$$

where $\lambda_{ir}$ denotes the local loading evaluated at the rescaled time $\tau = r/T$.

To obtain the local estimates $\{\lambda_{ir}\}_{i=1}^N$ and $\{F_t\}_{t=1}^T$, we consider the kernel-weighted least squares (WLS) problem

$$\min_{\{\lambda_{ir}\},\{F_t\}} (NT)^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T} \left(X_{it} - \lambda_{ir}'F_t\right)^2 K_h\left(\frac{t-r}{T}\right), \tag{3.3}$$

where $K_h(x) = h^{-1}K(x/h)$ is a rescaled kernel function, and $h = h(T, N)$ denotes the bandwidth parameter that controls the local neighborhood width. In practice, we use the **Epanechnikov kernel**,

$$K(z) = 0.75(1 - z^2)\mathbf{1}(|z| < 1), \tag{3.4}$$

and set $h$ according to a **Silverman (1986) rule-of-thumb** adapted to the panel setting,

$$h = cT^{-1/5}N^{-1/10}, \quad \text{with } c = \frac{2.35}{\sqrt{12}}, \tag{3.5}$$

ensuring a balance between bias and variance. The effective local window is therefore defined by $\tau \in [h, 1-h]$, so that boundary observations—where the kernel support would otherwise be truncated—are excluded from the estimation grid. This trimming avoids the need for boundary kernels as in Su and Wang (2017), while preserving the uniform order of approximation over the interior region.

For each value of $\tau$, we define the normalized weights

$$w_t(\tau, h) = \frac{1}{h}K\left(\frac{t/T - \tau}{h}\right), \tag{3.6}$$

and construct the locally demeaned, kernel-weighted data matrix

$$X_r = W^{1/2}(X - \mu_\tau), \qquad \mu_\tau = \frac{\sum_t w_t(\tau, h)X_t}{\sum_t w_t(\tau, h)}, \tag{3.7}$$

where $W^{1/2} = \text{diag}(\sqrt{w_t(\tau, h)})$. The local WLS problem is equivalent to minimizing

$$\text{tr}\left[(X_r - F(\tau)\Lambda(\tau)')(X_r - F(\tau)\Lambda(\tau)')'\right], \tag{3.8}$$

subject to the standard identification restrictions

$$F(\tau)'F(\tau)/T = I_R, \quad \Lambda(\tau)'\Lambda(\tau) = \text{diagonal}. \tag{3.9}$$

Under these constraints, the solution corresponds to the principal components of $X_r$. Letting $X_rX_r'$ denote the weighted covariance matrix, the estimated factors $\widehat{F}(\tau)$ are given by $\sqrt{T}$ times the eigenvectors associated with the $R$ largest eigenvalues of $X_rX_r'$, and the local loadings are obtained as

$$\widehat{\Lambda}(\tau) = \frac{X_r'\widehat{F}(\tau)}{T}. \tag{3.10}$$

These local estimates provide $\widehat{F}(\tau)$ and $\widehat{\Lambda}(\tau)$ only up to an orthogonal rotation and sign indeterminacy. In practice, we address this by aligning the estimated loadings across the grid of $\tau$ using a **Procrustes rotation**, which ensures consistency of signs and orientation over time and prevents spurious discontinuities in the estimated loading paths.

The above procedure yields local factors and loadings that satisfy the weighted relation

$$X_t \approx \Lambda(\tau)'F_t(\tau), \qquad F_t(\tau) = \sqrt{w_t(\tau, h)}F_t, \tag{3.11}$$

implying that $\widehat{F}(\tau)$ identifies a kernel-weighted version of the true factors. To recover the unweighted factors $F_t$ on their original scale, we perform a second-stage least squares regression. Specifically, given the first-stage estimates $\widehat{\lambda}_{it}$ from local PCA, the unweighted factors at each time $t$ are obtained by solving

$$\min_{F_t \in \mathbb{R}^R} \sum_{i=1}^{N} \left(X_{it} - \widehat{\lambda}_{it}'F_t\right)^2, \tag{3.12}$$

which yields the closed-form expression

$$\widehat{F}_t = \left( \sum_{i=1}^{N} \widehat{\lambda}_{it} \widehat{\lambda}_{it}' \right)^{-1} \sum_{i=1}^{N} \widehat{\lambda}_{it} X_{it}. \tag{3.13}$$

This second step removes the local kernel weighting and provides consistent estimates of the common factors $\widehat{F}_t$ up to the same rotation used in the first stage.

In summary, the estimation proceeds as follows: the first stage performs local weighted PCA to estimate $(\widehat{F}(\tau), \widehat{\Lambda}(\tau))$ within a moving window defined by the bandwidth $h$, while the second stage, using the consistent estimator of $\widehat{\Lambda}$, recovers the unweighted factors $\widehat{F}_t$ through cross-sectional OLS regressions.

### 3.1 Determining the Number of Local Factors.

In the above estimation procedure, the number of factors $R$ was assumed to be known. In practice, it must be determined from the data. Following Su and Wang (2017), we denote the true number of factors by $R_0$, bounded above by a finite integer $R_{\max}$, and employ information criteria to identify it. While Su and Wang propose a local BIC-type criterion inspired by Bai and Ng (2002), we first evaluate the classical Bai–Ng criteria locally and subsequently refine them using the adaptive ABC approach of Alessi (2010).

Let $\widehat{F}_t(R)$ and $\widehat{\lambda}_{it}(R)$ denote the local PCA estimators of the factors and loadings under the assumption of $R$ factors, obtained by imposing the normalization

$$N^{-1}\Lambda_r'(R)\Lambda_r(R) = I_R, \qquad T^{-1}F^{(r)}(R)'F^{(r)}(R) = \mathrm{diag}(d_1, \ldots, d_R), \tag{3.14}$$

where the diagonal elements $d_j$ are arranged in descending order. Define the average residual variance function

$$V(R) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( X_{it} - \widehat{\lambda}_{ir}'(R)\, \widehat{F}_t(R) \right)^2, \tag{3.15}$$

The information criteria of Bai and Ng (2002) provide a consistent method for identifying the number of factors in large-dimensional approximate factor models, under the double asymptotic regime where both $N$ and $T$ diverge. Bai and Ng propose three alternative criteria, denoted $\mathrm{IC}_1$, $\mathrm{IC}_2$, and $\mathrm{IC}_3$, each differing in the strength of the penalization term $\rho_{NT}$. While all three are asymptotically equivalent, only the first two have been found to perform satisfactorily in finite samples; the third tends to over-penalize and may underestimate the true number of factors. We therefore focus on the first two specifications.

The generic information criterion is then given by

$$IC(R) = \log V(R) + R\, \rho_{NT}, \tag{3.16}$$

where $\rho_{NT}$ is the penalty term controlling the trade-off between model fit and parsimony. We consider the two empirically reliable versions:

$$\rho_{NT}^{(1)} = \frac{N + T_{\text{eff}}}{N\, T_{\text{eff}}} \log\!\left( \frac{N\, T_{\text{eff}}}{N + T_{\text{eff}}} \right), \tag{3.17}$$

$$\rho_{NT}^{(2)} = \frac{N + T_{\text{eff}}}{N\, T_{\text{eff}}}\, \log\!\big( \min\{\sqrt{T_{\text{eff}}},\, \sqrt{N}\} \big)^2, \tag{3.18}$$

where $T_{\text{eff}}$ denotes the effective number of time observations within each local kernel window. The first penalty (3.17) corresponds to the original $\mathrm{IC}_1$ criterion, based on the logarithm of the ratio between total and effective sample sizes; the second (3.18) corresponds to $\mathrm{IC}_2$, which adjusts the penalty according to the smaller dimension between $T_{\text{eff}}$ and $N$. Both criteria balance

the risk of overfitting against the bias induced by local smoothing, while retaining the same asymptotic consistency.

For each localization point $\tau$, we compute

$$\widehat{R}_{IC}^{(j)}(\tau) = \arg\min_{R \le R_{\max}} \Big[ \log V(R) + R\, \rho_{NT}^{(j)} \Big], \qquad j = 1, 2, \tag{3.19}$$

and store the resulting $\widehat{R}_{IC}(\tau)$ along the time grid. The comparison between $\widehat{R}_{IC}^{(1)}$ and $\widehat{R}_{IC}^{(2)}$ provides a diagnostic check of local stability in the estimated number of factors, as both criteria are consistent for $R_0$ when $N, T \to \infty$.

Although the Bai–Ng criterion performs well asymptotically, its finite-sample behavior may be unstable, often leading to over-or underestimation of $R_0$—see, for instance, Forni et al. (2009)—. To improve robustness, Alessi (2010) propose an adaptive modification—the ABC criterion—that introduces a tuning parameter $c > 0$ governing the penalization strength. While this modification leaves asymptotic consistency unaffected, it enhances the criterion's sensitivity in finite samples, especially under heteroskedastic idiosyncratic components or unbalanced cross-sections.

Formally, the ABC criterion takes the form

$$IC_{ABC}(R, c) = \log V(R) + c\, R\, \rho_{NT}, \tag{3.20}$$

where $c$ varies over a grid $\mathcal{C} = [1/c_{\max}, c_{\max}]$. For each $c \in \mathcal{C}$ and subsample size $n_s$ (with $n_s \in [N - \text{nbck}, N]$), we compute the minimizing rank $\widehat{R}_{c,n_j}^{T} = \arg\min_R IC_{ABC}(R, c)$. Stability with respect to both $c$ and $n_s$ is then assessed through the empirical variance

$$S_c = \mathrm{Var}_j\Big(\widehat{R}_{c,n_j}^{T}\Big), \tag{3.21}$$

and the selected number of factors corresponds to the most stable range of $c$ where $\widehat{R}_{c,n_j}^{T}$ remains constant across subsamples.

In our implementation, we first compute $\widehat{R}_{IC}(\tau)$ from the Bai–Ng criterion for each local window. We then refine it by applying the ABC procedure to the corresponding weighted matrix $X_r$. The algorithm evaluates $IC_{ABC}(R, c)$ for a dense grid of $c$ values and multiple random subsamples of cross-sectional size $n_s$, retaining as $\widehat{R}_{ABC}(\tau)$ the value of $R$ corresponding to the most stable plateau in the function $c \mapsto \widehat{R}_{c,n_j}^{\tau_j}$. Graphically, this corresponds to the flat regions of the $\widehat{R}_{c,n_j}^{\tau_j}$ curve discussed in Alessi (2010). We notice that there are several intervals in which a plateau is reached; however following the same theoretical reasoning as in Hallin and Liška (2007), the relevant stability interval is the second—the first corresponds to $R_{max}$—. By construction, the ABC refinement produces smoother and more robust estimates $\widehat{R}_{ABC}(\tau)$ than the standard $\widehat{R}_{IC}(\tau)$, while maintaining the same large-sample consistency properties as the Bai–Ng criterion.

# 4 Testing the constancy of the factor loadings

Consider the approximate factor model

$$X_{it} = \lambda_{it}' F_t + e_{it}, \qquad i = 1, \dots, N,\ t = 1, \dots, T, \tag{4.1}$$

where $X_t$ is an $N$-dimensional vector of observations, $F_t$ is an $R$-dimensional vector of common factors, $\lambda_{it}$ are the corresponding factor loadings, and $e_{it}$ are idiosyncratic components. Under the null hypothesis the loadings are constant,

$$H_0 : \lambda_{it} = \lambda_{i0} \quad \forall\, t, \tag{4.2}$$

whereas under the alternative they vary smoothly with rescaled time $\tau = t/T$,

$$H_1 : \lambda_{it} = \lambda_i(\tau), \quad \text{smooth in } \tau \in (0,1). \tag{4.3}$$

Under $H_0$, the model reduces to

$$X_t = \Lambda_0 F_t + e_t, \tag{4.4}$$

with $\Lambda_0 = (\lambda_{10}, \ldots, \lambda_{N0})'$ and $e_t = (e_{1t}, \ldots, e_{Nt})'$. The global estimators $(\tilde{F}_t, \tilde{\Lambda}_0)$ are obtained by applying principal components to the full sample, minimizing

$$\frac{1}{NT}\sum_{t=1}^{T} \|X_t - \Lambda_0 F_t\|^2, \qquad \text{subject to } \frac{1}{T}F'F = I_R. \tag{4.5}$$

Under this normalization, the estimated factors and loadings satisfy

$$\tilde{F} = X\tilde{\Lambda}_0(\tilde{\Lambda}_0'\tilde{\Lambda}_0)^{-1}, \qquad \tilde{\Lambda}_0 = \frac{1}{T}X'\tilde{F}. \tag{4.6}$$

Hence, $\tilde{F}$ corresponds to the first $R$ eigenvectors of the covariance matrix $XX'/NT$, and $\tilde{\Lambda}_0$ is obtained by projecting $X$ onto $\tilde{F}$. The fitted common component under $H_0$ is

$$\tilde{\chi}_t = \tilde{\Lambda}_0 \tilde{F}_t. \tag{4.7}$$

Under $H_1$, local PCA with kernel weights

$$K_h(t/T - \tau) = \frac{1}{h}K\Big(\frac{t/T - \tau}{h}\Big) \tag{4.8}$$

is applied to obtain $\hat{F}_t$ and $\hat{\lambda}_{it}$ for each $\tau \in (0,1)$.

Let $\tilde{\chi}_{it} = \tilde{\lambda}_{i0}'\tilde{F}_t$ and $\hat{\chi}_{it} = \hat{\lambda}_{it}'\hat{F}_t$ denote, respectively, the estimated common components under the null and the alternative. The sample quadratic distance between the two is defined as

$$\hat{M} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\big(\hat{\chi}_{it} - \tilde{\chi}_{it}\big)^2. \tag{4.9}$$

Define the kernel and convolution weights

$$k_{h,st} = h^{-1}K\Big(\frac{s-t}{Th}\Big), \qquad \bar{k}_{st} = \int_{-1}^{1} K(v)K\Big(\frac{s-t}{Th} - v\Big)\,dv, \tag{4.10}$$

and the bias and variance normalizers

$$B_{NT} = \frac{h^{1/2}}{N^{1/2}T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\big(F_t'L_{st}F_s\big)^2 e_{is}^2, \tag{4.11}$$

$$V_{NT} = 2T^{-2}N^{-1}h^{-1}\sum_{r\neq s}\bar{k}_{sr}^2\,\mathbb{E}\Big[(F_s'H_0\bar{\Sigma}_F H_0'F_r)^2(e_{rs}')^2\Big], \tag{4.12}$$

where $L_{st}$ collects kernel-weighted rotations from local PCA and $H_0$, $\bar{\Sigma}_F$ are the population limits under $H_0$.

The standardized test statistic is

$$J_{NT} = V_{NT}^{-1/2}\Big(TN^{1/2}h^{1/2}\,\hat{M} - B_{NT}\Big), \tag{4.13}$$

which converges in distribution to the standard normal,

$$J_{NT} \xrightarrow{d} \mathcal{N}(0,1) \qquad \text{under } H_0. \tag{4.14}$$

Finite-sample inference can deviate from the asymptotic approximation. Su and Wang therefore propose a bootstrap procedure that preserves both heteroskedasticity and cross-sectional dependence in the residuals. In practice, we implement this test using the **R** package `TVMVP`, which reproduces the bootstrap algorithm developed by Su and Wang (2017). This approach provides empirical critical values that account for cross-sectional dependence and finite-sample effects, ensuring more reliable inference than asymptotic approximations.

# 5 Data

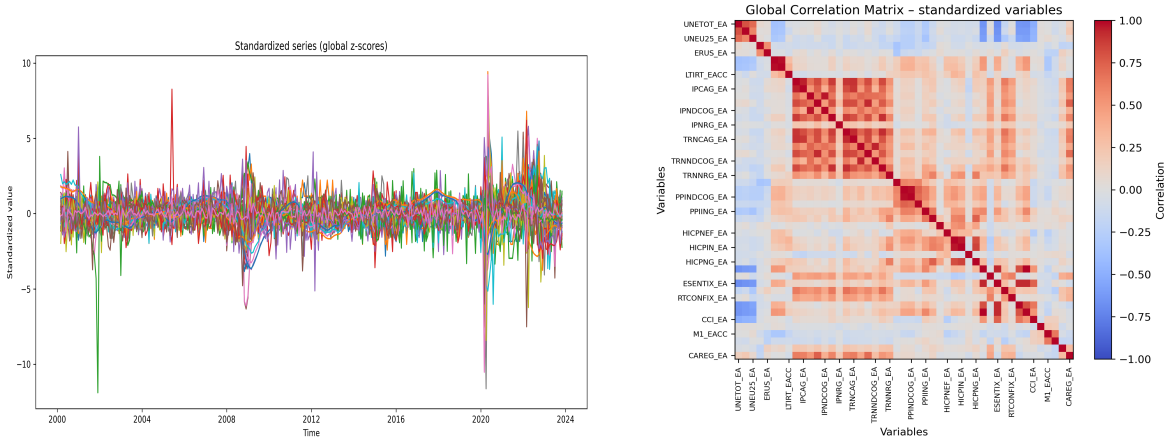
The dataset consists exclusively of Euro Area (EA) aggregate macro-financial time series. All variables are drawn from the dataset *EA-MD-QD*, compiled from official sources such as Eurostat (EUR), the European Central Bank (ECB), the OECD, and FRED.

The database includes a comprehensive set of indicators grouped into eleven broad categories: (1) national accounts, (2) labour market indicators, (3) credit aggregates, (4) labour costs, (5) exchange rates, (6) interest rates, (7) industrial production and turnover, (8) prices, (9) confidence indicators, (10) monetary aggregates, and (11) other financial variables such as share prices and car registrations. All series are expressed in consistent units (chain-linked volumes, indices, or percentages) and are seasonally and calendar adjusted (SCA or SA) when available.

The dataset spans from 2000Q1 to 2023Q4 and primarily features **monthly** observations. Before estimation, all series were transformed to achieve stationarity. Following the procedures outlined in the file `routine_data.m`, a **light transformation** was applied to all variables.

A few remaining non-stationary series were rendered stationary through an additional first-difference transformation.

Finally, all series were standardized to have zero mean and unit variance, ensuring comparability across variables and preventing scale differences from affecting the extraction of common factors.



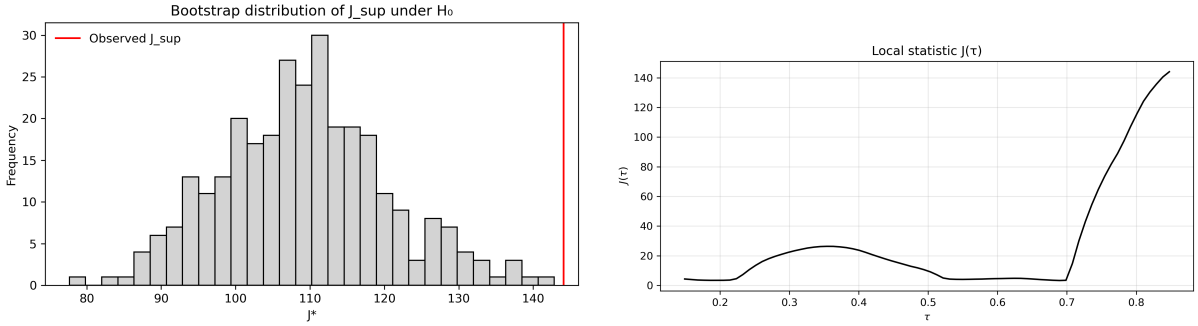

(a) Standardized series (z-scores)

(b) Global correlation matrix

Figure 1: Overview of standardized data and correlation structure.

# 6 Empirical Results

## 6.1 Stability test




(a) Bootstrap distribution of $J_{\text{sup}}$ under $H_0$. The red line marks the observed statistic.

(b) Local statistic $J(\tau)$ across the rescaled time domain.

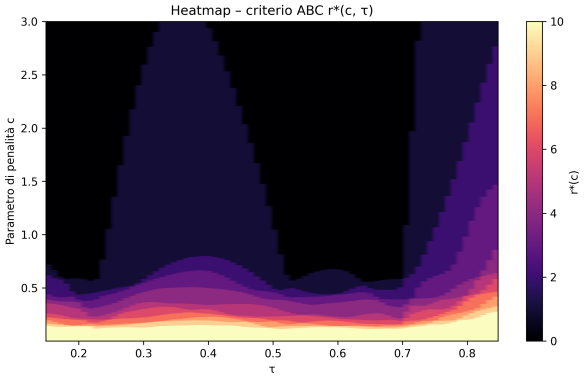
Figure 2: Bootstrap-based inference for the constancy of factor loadings. The left panel reports the simulated null distribution, while the right panel shows the local test statistic $J(\tau)$ indicating time variation.

As shown in Figure 2, the local statistic $J(\tau)$ exhibits pronounced instability, particularly in the post-COVID period, where the values rise sharply and signal strong time variation in the factor loadings. Consistently, the bootstrap distribution of $J_{\text{sup}}$ (left panel) places the observed statistic in the far upper tail, leading to the rejection of the null hypothesis of constant loadings at conventional significance levels (the corresponding $p$-value is effectively zero). These findings clearly justify the adoption of a time-varying factor model, as they highlight the presence of structural shifts in the underlying relationships. Allowing the loadings to evolve over time enables us to capture how the relative contribution of each variable to the identification of the common factors changes across different economic phases, providing a more accurate and dynamic representation of the euro-area comovements.

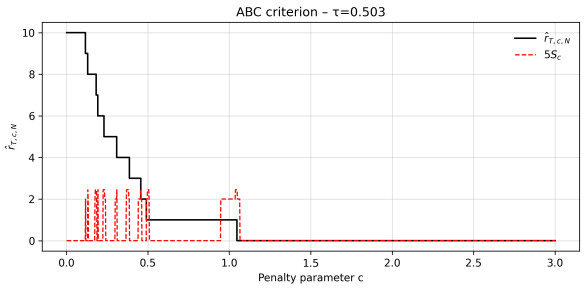
## 6.2 Bai-Ng and ABC

The determination of the number of factors across local windows is conducted using both standard Information Criteria by Bai and Ng (2002) and the refined ABC criterion by Alessi (2010). Our goal is a local cross-validation of factor dimensionality within the time-varying setting, so that selected factors reflect structural changes rather than sampling noise. In each window we identify stability regions – plateaux – where the empirical variance explained by the extracted components is locally constant; following Alessi (2010), persistent plateaux are taken to indicate economically meaningful latent dimensions.

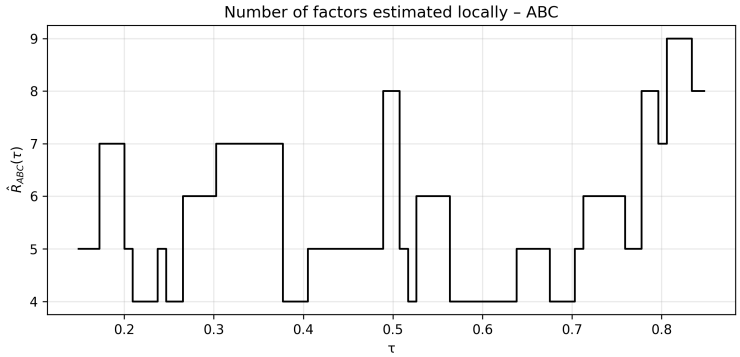
To make this selection transparent, we report a heatmap of the selected factor count over the $\tau$-grid and, for representative $\tau$'s, window-level plots showing the stepwise changes in the estimated number of factors together with the corresponding empirical variance. As shown in Alessi (2010), due to the monotonicity of $\widehat{R}^T_{c,n_j}$ we expect the step function to display well-defined stability ranges and for moderate values of $c$ to be a stable function of the sample size $(n_j, \tau_j)$. The stability is furthermore illustrated by the empirical variance.




(a) Heatmap of the selected number of factors across taus, conditioned on the grid of penalties




(b) $\widehat{R}^T_{c,n_j}$ and $S_c$ as functions of $c$ for a specific tau




(c) Selected number of factors $\widehat{R}(\tau)$ over the $\tau$–grid

Figure 3: Selection of the number of factors across local windows. The left panel shows the heatmap of the selected number of factors obtained through the refined ABC criterion over the $\tau$–grid, depending on the penalizing parameter $c$. The right panel focuses on a representative $\tau$–specific window, displaying $\widehat{R}^T_{c,n_j}$ and the corresponding scaled empirical variance $S_c$. The bottom panel reports $\widehat{R}(\tau)$ across time, summarizing the local dimensionality.

## 6.3 Factors

Before discussing the estimated factors, it is important to recall that the model is based on a two–sided kernel. This specification implies that each local estimate incorporates information not only from past observations but also from future ones within the bandwidth of the kernel. As a consequence, the factors may appear to anticipate macroeconomic shocks by roughly one or two years. In reality, this apparent anticipation simply reflects the smoothing window: the kernel starts to integrate the information content of the upcoming observations before the shock fully materializes. Therefore, the dynamics shown in the figures below should be interpreted as locally smoothed representations of the underlying time–varying relationships rather than as true forecasts. Since the interpretation of time–varying loadings requires a fixed factor dimension, we set the number of factors equal to the (rounded) average of the values selected locally by the ABC criterion.

### 6.3.1 Interpreting the Time-Varying Loadings

Factor 1 represents an inverse real business cycle. It increases during periods of economic slowdown or monetary tightening, when the euro appreciates, unemployment rises, and monetary aggregates such as M2 expand, while industrial production and stock prices weaken. Conversely,

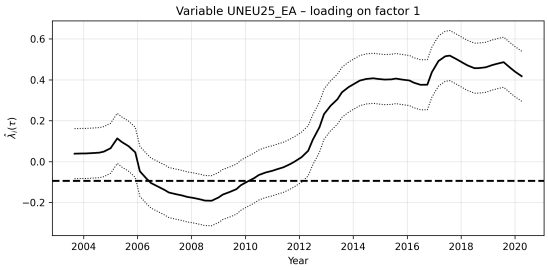
it declines during phases of economic expansion, characterized by exchange rate depreciation, higher output, and rising asset prices. The factor therefore reflects macroeconomic stress when positive and real expansion when negative.
During the global financial crisis (2008–2013), the loadings of **REER42** and **M2** sharply increased, while those of **IPNRG** and **SHIX** turned negative. This configuration suggests that during the crisis, the latent factor was primarily driven by the dynamics of monetary liquidity and currency value, with **M2** and **REER42** reflecting these movements, while **IPNRG** and **SHIX** captured the contraction in real activity and financial markets.

Between 2013 and 2019, the loadings of **M2** remained positive but showed a gradual decline, signaling a weakening impact of the earlier expansionary monetary policies. Conversely, **REER42** experienced a clear stability, indicating that the relationship between the euro's value and the factor became constant as the economy stabilized. During this period, **IPNRG** and **SHIX** displayed more stability, with **SHIX** moving toward a neutral position, suggesting that financial market tensions decreased as the economy entered recovery.

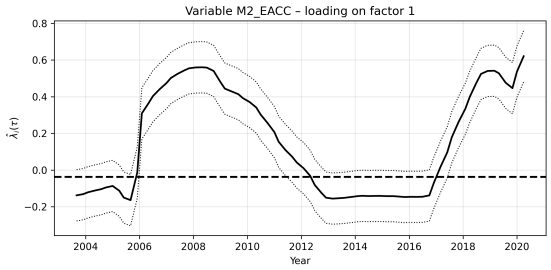
With the outbreak of the COVID-19 pandemic (2020–2021), the loadings of **M2** surged once again, reflecting the renewed monetary expansion. However, **REER42** started its decline, suggesting that the euro's role in the factor became even less pronounced during this period of crisis. **IPNRG** and **SHIX**, instead of becoming more negative, maintained a relatively stable or even slightly positive relationship with the factor, highlighting that real economic activity and financial markets behaved differently compared to previous crisis periods.

The **UNEU25** loading, which represents youth unemployment, became notably positive during the pandemic period, indicating that youth unemployment dynamics played an increasingly significant role in driving the common factor during the crisis. This shift underscores how labor market tensions, particularly among younger workers, became more important in shaping macro-financial dynamics during the pandemic.

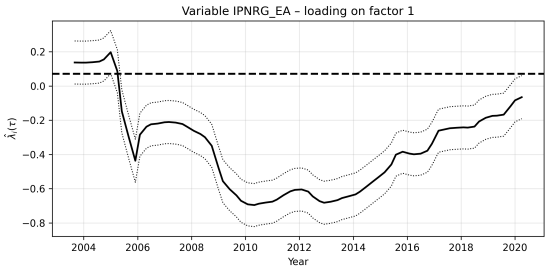
Overall, Factor 1 can be interpreted as a measure of macro-financial tension, capturing the evolving influence of monetary liquidity, real economic activity, and labor market conditions. The factor tends to be positive during periods of economic distress, with high loadings for **M2** and **UNEU25**, and negative during more stable or expanding periods, where **IPNRG** and **SHIX** show stronger relationships. The time-varying nature of this factor allows us to identify shifts in the structure of macro-financial interactions, with the most significant changes occurring during the global financial crisis and the COVID-19 pandemic.



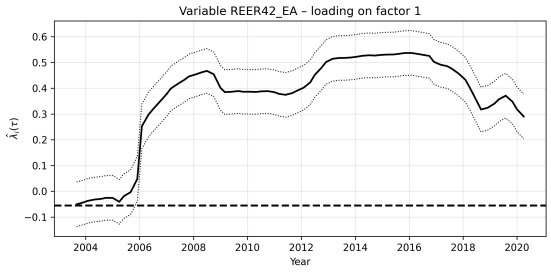

UNEU25 loading



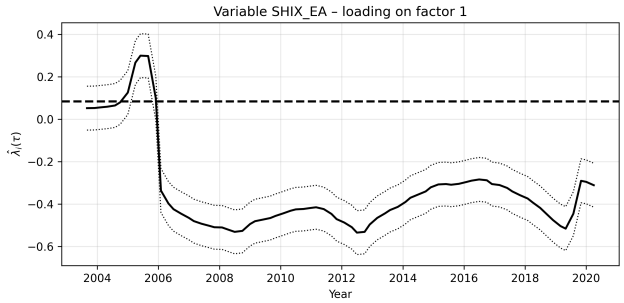

M2 loading




IPNRG loading




REER loading




SHIX loading

Figure 4: Loadings on the first factor

### 6.3.2 Second Factor

While $F_1$ reflects real economic activity, $F_2$ captures broad economic expectations, acting as a forward-looking factor since investment and monetary intentions usually precede changes in output or unemployment. "Economic sentiment" refers to the overall confidence or pessimism of consumers and firms regarding current and future economic conditions: high sentiment fosters spending and investment, while low sentiment leads to saving and sluggish growth.

The main loadings of latent variables such as **ICONFIX_EA** (−0.612), **ESENTIX_EA** (−0.610), **BCI_EA** (−0.606), and **SCONFIX_EA** (−0.550) indicate a strong negative correlation between Euro Area sentiment and $F_2$. Conversely, unemployment indicators like **UNETOT_EA** and **UNEO25_EA** have positive loadings (around 0.50), showing that higher confidence coincides with lower unemployment.

After a period of relative stability, confidence dropped sharply, reaching its lowest point in 2009 and recovering to baseline only after 2013. This aligns with the 2008 Lehman Brothers collapse, when global financial turmoil and uncertainty fueled risk aversion. Between 2013 and 2019, trade and industrial sentiment improved but remained fragile due to slow growth, gradually declining unemployment, and moderate inflation.

In 2020 this trend reversed with the outbreak of COVID-19, never returning to pre-crisis levels. The same pattern is visible in **SCONFIX_EA**, though its 2008 dip is milder, while the post-2017 fall is unprecedentedly sharp. Between 2007 and 2009, real estate collapsed but

government stimulus and low rates temporarily sustained investment. By contrast, the 2020 lockdown disrupted production and supply chains, pushing sentiment to record lows. Subsequent easing and stimulus led to a partial rebound, uneven across sectors.

Overall, the curve for **SCONFIX_EA** is smoother than that of **RTCONFIX_EA**, reflecting the industrial sector's alignment with broader cycles, whereas retail confidence reacts more quickly to short-term shocks such as the lockdown.

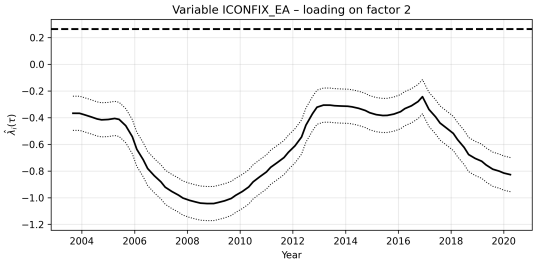



Figure 5: ICONFIX_EA loading on F2

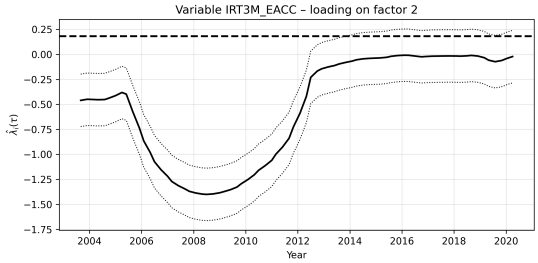



Figure 6: IRT3M loading on F2

Equally relevant indexes contributing to our analysis of the second factor are **IRT3M_EACC** ($-0.529$) and **IRT6M_EACC** ($-0.515$), interest-rate indicators that partly confirm the trend previously discussed above and provide additional insightful information for our analysis of time-varying loadings (Figure 2).
Their negative loadings suggest that as general sentiment improves, both three months (**IRT3M**) and six months (**IRT6M**) interest rates decrease – and vice versa. The strongest negative co-movement is again experienced around 2007–2009, when the ECB moved from tightening to emergency cuts and rates fell from around 4.25% to 1%. The 2006–2010 financial crisis, triggered by the U.S. subprime mortgage collapse, led to a severe credit crunch where the Fed aggressively cut short-term rates to near zero to re-stabilize the banking system.

After the global financial crisis, the loading rose steadily from around $-1.5\%$ in 2009 to nearly zero, marking a shift from a cyclical to a policy-driven dynamic. Before 2008, short-term interest rates tended to move counter-cyclically – falling when confidence weakened and rising as sentiment improved. In the post-crisis period, however, this link progressively vanished as the ECB adopted an ultra-accommodative stance. From 2014 onwards, with policy rates close to or even below zero, short-term rates ceased to reflect the state of the economy or the prevailing sentiment, remaining artificially low as a deliberate policy tool to sustain demand and anchor credibly expectations.

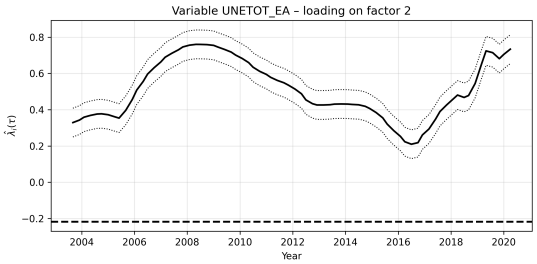



Figure 7: UNETOT_EA loading on F2

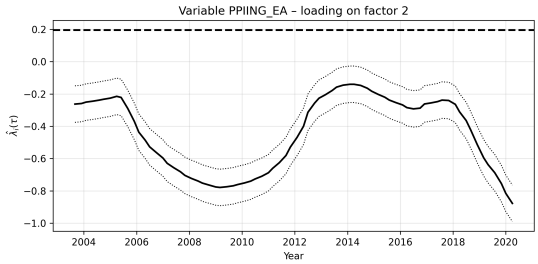



Figure 8: PPIING_EA loading on F2

From 2003 to 2020, the loading of unemployment on economic sentiment fluctuated markedly, reflecting how the link between joblessness and collective outlook evolved through major crises. The impact of unemployment on $F_2$ is captured by **UNETOT** (0.510) and **UNEU25** (0.498), both loading positively on the factor. During the two major crises, unemployment loadings rose sharply, showing an almost symmetric pattern to previous indicators. Between 2006 and 2012, the loading increased from about 0.35 to 0.8, indicating that pessimism and labor market

weakness became closely intertwined.

After 2008, loadings declined as sentiment recovered faster than employment, suggesting a temporary decoupling likely driven by policy support and slow labor adjustments. Confidence improved with Euro Area stabilization and quantitative easing despite persistent joblessness. Between 2016 and 2021, loadings rose again to around 0.7, reflecting the late recovery and the renewed sensitivity of sentiment to unemployment ahead of COVID-19. Following the pandemic, unemployment once more became a key driver of $F_2$, signaling that confidence remained highly responsive to labor market fragility.

Lastly, the role of **PPIING**, a Producer Price Index measure, shows an average negative loading of $-0.442$, implying that inflation generally moves opposite to sentiment. This inverse relationship strengthened from $-0.3$ to $-0.8$ between 2003 and 2008, during the global financial crisis, when rising input costs coincided with declining confidence. It weakened in the following years, as both sentiment and PPI stagnated, before turning sharply negative again between 2018 and 2020 ($-0.9$) amid the COVID-19 shock, when both collapsed simultaneously.

**Interpretation of Factors 3–6**

**Factor 3 – Monetary and labour conditions:** This factor is negatively correlated with both nominal (ERUS) and real (REER42) exchange rates, meaning that higher factor values correspond to a weaker euro. It also loads positively on unemployment and monetary aggregates (M1–M2), while confidence indicators move in the opposite direction. Overall, it captures phases of economic slack and monetary easing, when euro depreciation, rising unemployment, and liquidity expansion coexist with low confidence and weak activity.

**Factor 4 – Goods and energy inflation:** This factor shows strong positive loadings on consumer and energy price indices (HICP, PPI) and negative ones on industrial production and turnover in consumer goods. It captures episodes of supply-side inflation, where higher prices coincide with falling output and rising long-term interest rates. The combination of inflationary pressures and real contraction points to cost-push dynamics driven by energy and input shocks.

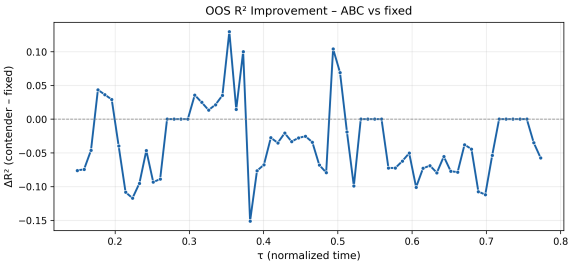
**Factors 5–6 – Residual real and financial dynamics:** The fifth and sixth factors are more difficult to interpret clearly. Factor 5 mixes energy-related price pressures with labour market slack and competitiveness, suggesting a residual cost-push component overlapping with the main inflation factor. Factor 6, instead, loads negatively and uniformly on industrial production, turnover and equity prices, hinting at a broad but less distinct real–financial cycle. Both likely capture secondary co-movements—sectoral or cyclical—rather than autonomous macroeconomic forces.

### 6.4 Out-of-sample performance (one-sided)

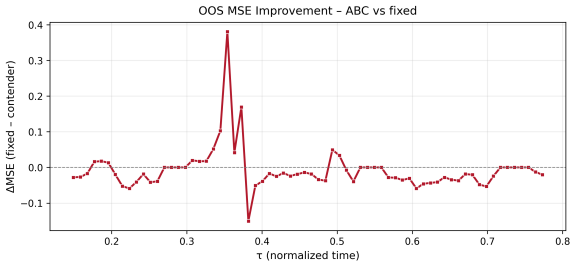
We emphasize that forecasting is not the primary objective of this study. The Time-Varying Factor Model is estimated in a two-sided framework, where both past and future information contribute to the extraction of latent factors. The one-sided out-of-sample (OOS) evaluation presented here is thus intended only as a robustness exercise, assessing whether the model retains reasonable predictive ability when estimated recursively using information available up to each $\tau$.

In this strictly one-sided setting, the adaptive ABC selection does not outperform the fixed-$r$ benchmark: mean and median $\Delta R^2$ are negative ($-0.031$ and $-0.035$), and a paired $t$-test rejects equality ($t = -4.66$, $p = 0.0000$). For squared loss the evidence is weak: the average $\Delta$MSE is slightly negative but not statistically different from zero ($t = 1.01$, $p = 0.3181$). In absolute terms, the ABC specification attains a mean $R^2_{\text{OOS}}$ of 0.419 (min 0.232, max 0.648) and a mean $\text{MSE}_{\text{OOS}}$ of 0.363 (min 0.193, max 1.498). The spike observed in the out-of-sample MSE is largely driven by strong shifts in local means within the one–sided windows, which inflate the absolute error whenever the forecast deviates even slightly from the true level. Since $R^2$ is a relative measure – normalized by the local variability – it is much less sensitive to level shifts and therefore provides a more reliable basis for comparing model performance across specifications.

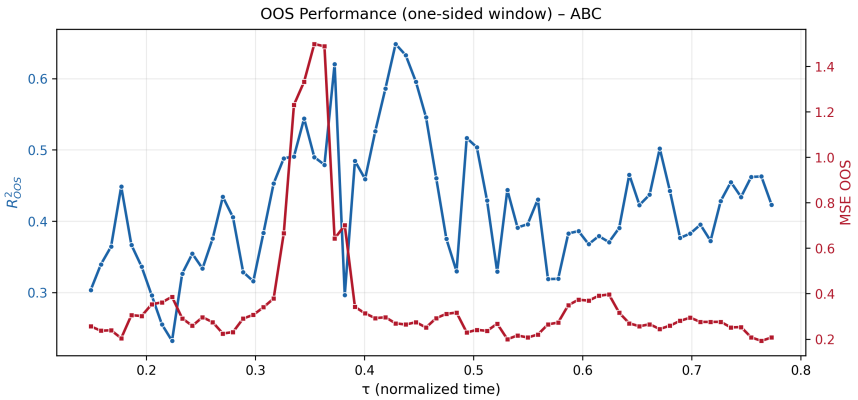
These results indicate that allowing for time variation in factor dimension does not materially deteriorate the model's predictive accuracy. The modest differences observed are likely influenced by the fact that roughly half of the available observations are used for model estimation, while forecasts are generated for the subsequent 30% of the sample – a relatively short evaluation window that limits the potential benefit of adaptive selection.




(a) $\Delta R^2$ (ABC – Fixed)




(b) $\Delta$MSE (Fixed - ABC)




(c) Absolute OOS performance – ABC

Figure 9: One-sided OOS evaluation. Panels (a)–(b) show differences vs a fixed-$r$ benchmark; panel (c) reports absolute performance of the ABC specification.

Table 1: OOS comparison – ABC vs fixed selection and absolute OOS performance (one-sided)

| **Statistic** | **Value** | | |
|---|---|---|---|
| **ABC vs Fixed (68 common $\tau$)** | | | |
| Mean $\Delta R^2$ | $-0.031$ | Median $\Delta R^2$ | $-0.035$ |
| Share($\Delta R^2 > 0$) | 19.12% | | |
| Mean $\Delta$MSE | $-0.007609$ | Median $\Delta$MSE | $-0.018637$ |
| Share($\Delta$MSE $> 0$) | 19.12% | | |
| Paired t-test $\Delta R^2$ | $t = -4.66$, $p = 0.0000$ | | |
| Paired t-test $\Delta$MSE | $t = 1.01$, $p = 0.3181$ | | |
| **Absolute OOS performance – ABC** | | | |
| Mean $R^2_{OOS}$ | 0.419 | Min | 0.232 |
| Max $R^2_{OOS}$ | 0.648 | | |
| Mean MSE$_{OOS}$ | 0.362886 | Min | 0.192510 |
| Max MSE$_{OOS}$ | 1.497662 | | |

# 7 Do Nonlinearities Matter?

In this section, we assess curvature in the data manifold using an *autoencoder*. A few caveats apply. First, this procedure is not a formal statistical test for nonlinearity, as it lacks variance estimates beyond practical robustness checks; it is intended for forecasting rather than inference. Second, this analysis is *complementary* to the time-varying model. Whereas the latter identifies structural breaks in parameters, the autoencoder explores alternative non-affine mappings that are assumed to be time-invariant (though this could be relaxed with appropriate dummies). The primary point of comparison between the two approaches is their out-of-sample performance (Section 6.4). Finally, although our analysis is constrained by the limited dimensionality of the dataset, it still serves as an exploratory exercise aimed at illustrating the potential of this machinery for addressing these questions.

We find no evidence of curvature: non-affine models do not outperform affine benchmarks out of sample.

**Autoencoder Architecture.** Autoencoders are a class of neural networks characterized by a distinctive "two-funnel" architecture: an *encoder* that compresses high-dimensional inputs into a compact latent representation, and a reversed *decoder* that reconstructs the original signal while filtering out noise. The network parameters are optimized to minimize the reconstruction loss—the dissimilarity between input and output, typically measured by the $L_2$ norm—with optional regularization to mitigate overfitting.

Autoencoders have been employed recently to estimate signal in factor models (Andreini et al., 2023; Xiu and Shen, 2024). The interest in this class of models arises from their potential to capture nonlinearities in data manifolds. Despite this, several challenges remain, including conditions for universality in the ambient space, non-identifiability of latent space representations, and invariance to overspecification of the true number of factors.[1]

Following Andreini et al. (2023), we use an asymmetric autoencoder with a deep, one-hidden-layer encoder and a purely affine decoder. While this restricts the hypothesis space, limiting the network's expressive power, it substantially reduces the number of parameters to learn—a necessary advantage given our limited dataset.[2] The loss function incorporates two

[1] See the Technical Appendix for a discussion of theoretical foundations, literature, and open questions on these issues.

[2] In some configurations, the ratio of data points to parameters can be as low as 10%.

**For each activation function:**

**Step 1: Model Tuning**

a) Tune linear model hyperparameters via 5-fold CV.

b) Tune nonlinear model hyperparameters via 5-fold CV.

c) Generate out-of-sample predictions and compute residual correlations between models.

**Step 2: Ensemble Optimization** — Tune ensemble weight $\alpha$ to maximize averaged $R^2$ over validation folds; record global and feature-wise metrics.

**Step 3: Selection Across Activations** — Select the activation function (and ensemble) with highest out-of-sample $R^2$.

Table 2: Pseudo-code overview of the two-step cross-validation and ensemble procedure.

| **Fixed parameters** | |
|---|---|
| Number of layers | 1 |
| Optimizer | `adam` |
| Learning rate | $10^{-2}$ |
| Length of validation set | 27 |
| **Cross-validated parameters (fine grid)** | |
| Activation functions | Tanh, SiLU, Sine |
| Neurons in hidden layer | 10, 20 |
| Number of latent factors | 4, 5, 6 |
| $\ell_1$ penalty coefficient | 2e-2, 5e-2 |
| Orthogonalization penalty coefficient | 1e-3, 1e-2 |

Table 3: Configuration of key hyperparameters.

regularization components: (i) an $\ell_1$ penalty to enforce sparsity, and (ii) an orthogonalization penalty that penalizes factor correlation.[3] The corresponding sensitivity parameters are selected via cross-validation.

**Model Selection.** Model selection was performed using a two-step procedure, repeated for each activation function, and summarized in Table 2.

First, linear and nonlinear models were tuned separately. Cross-validation was implemented with `TimeSeriesSplit` from `scikit-learn`[4], which adapts standard $k$-fold CV to time series by ensuring that training sets always precede validation sets. Hyperparameter tuning focused on a subset of key parameters to balance coverage and computational cost, using a coarse-to-fine strategy: random search followed by a full grid search (see Table 3 for the parameter configuration). Model performance was measured using the average out-of-sample $R^2$ across the five folds, providing a variance-insensitive assessment of generalization error, preferred over reconstruction MSE.

Second, the ensemble weight $\alpha$ was optimized on the validation set. Global and feature-wise $\Delta R^2$ were computed to quantify the contribution of the nonlinear model relative to the linear baseline. Third, the activation function and corresponding ensemble achieving the highest out-of-sample $R^2$ were selected as optimal.

[3]Formally, the orthogonalization penalty minimizes $\|\mathbf{F}^\top\mathbf{F} - \text{diag}(\mathbf{F}^\top\mathbf{F})\|_F^2$, with $\mathbf{F} \in \mathbb{R}^{T\times r}$ denoting the matrix of normalized factors. It is worth emphasizing that this penalty acts as a *soft* constraint: correlation between factors is penalized, but allowed. This constitutes another deviation from the classical linear framework.

[4]See `TimeSeriesSplit` for documentation.

| Model | Parameter | Value | OOS $R^2$ |
|---|---|---|---|
| Linear model | Number of latent factors | 6 | |
| | Neurons in hidden layer | 20 | |
| | $\ell_1$ penalty coefficient | 0.02 | |
| | Orthogonalization penalty | 0.01 | 0.3150 |
| SiLU model | Number of latent factors | 4 | |
| | Neurons in hidden layer | 10 | |
| | $\ell_1$ penalty coefficient | 0.05 | |
| | Orthogonalization penalty | 0.001 | 0.3083 |
| | Residual correlation | 0.8957 | |
| Ensemble model | Optimal weight $\alpha$ | 0.50 | 0.3440 |

Table 4: Best configurations, out-of-sample (OOS) performance, and ensemble parameters with `seed = 42`.

**Quantitative results and robustness.** SiLU emerges as the optimal nonlinear activation, and Table 4 reports the corresponding optimal parameters for individual models along with key ensemble metrics. The data manifold exhibits only moderate curvature, and out-of-sample performance of the linear and SiLU models is within seed variance—and consistent with the linear model after accounting for time-varying effects—with residual correlation around 90%. Improvements in out-of-sample $R^2$ with the ensemble are minimal yet consistent across seeds, both globally and feature-wise, indicating that the ensemble primarily stabilizes variance rather than capturing additional nonlinear structure (see Figure 10).

These findings are robust across different autoencoder architectures (symmetrical with deep decoder, asymmetric with one extra hidden layer) and hyperparameter search methods (including Bayesian optimization via the `Optuna` package). Performance metrics—both global and feature-wise—remain stable across multiple seeds, despite occasional variations in optimal hyperparameters.

**Qualitative discussion.** From the feature-wise $R^2$ plot (Figure 10), we see that the model captures real-economy indicators——industrial production, unemployment, and confidence measures——quite well. In contrast, its performance is weaker for nominal variables such as prices and monetary aggregates. This pattern is broadly consistent with the linear analysis, where the first two factors corresponded to real-activity and sentiment components, respectively.

**Limitations.** A key limitation is the trade-off between expressiveness and generalization: while more complex network structures increase representational capacity, they can overfit, especially with limited data as in our case. The alignment between linear and nonlinear models suggests overfitting is largely mitigated, though the ability to capture more complex patterns remains constrained. Additionally, the non-i.i.d. nature of time series data makes cross-validation metrics sensitive to outliers, reflected in sensitivity to the choice of validation set parameters.[5]

# 8 Conclusions

This empirical analysis highlights the benefits of allowing for time variation in the factor structure vis-à-vis a model in which the factor loadings remain constant over time. Compared to a static

[5]We partially addressed this by fixeing `len_valset = 28`, to ensure the validation set is large enough to mitigate this sensitivity, while keeping it below 20% of the training set to preserve sufficient training data for reliable parameter estimation.

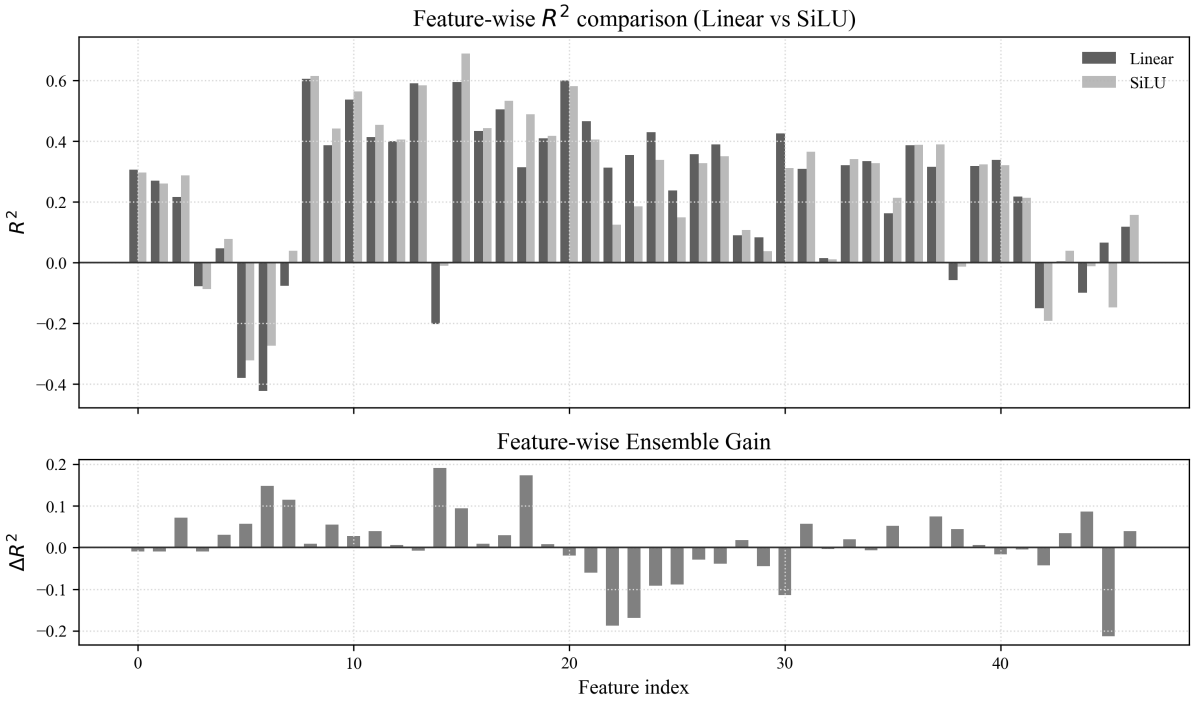



Figure 10: Upper panel: Feature-wise out-of-sample $R^2$ for affine and SiLU models. Lower panel: $\Delta R^2$ between ensemble and affine model.

specification, the time-varying model captures more accurately the evolving relationships among macro-financial variables, which appear to change significantly in response to major economic and policy shocks. Evidence suggests that the common factors shaping aggregate co-movements are themselves affected by these shocks, both in their relevance and composition, reflecting the shifting nature of underlying economic drivers.

In addition, it is worth noting that the optimal number of factors tends to rise during episodes of heightened turbulence—such as financial or sovereign crises—when additional dimensions are worth being included to better account for the complexity and interdependence of economic dynamics. Overall, these findings underline the importance of incorporating structural changes and regime-dependent behavior when modeling macro-financial linkages in an evolving economic environment.

Complementing the time-varying linear analysis, our exploratory nonlinear exercise suggests that affine and non-affine (autoencoder-based) models deliver broadly similar forecasting performance, confirming our baseline findings. This indicates that, within the limits of the available data, departures from linearity do not materially improve predictive accuracy.

# A Technical Appendix

This appendix serves a dual objective: it examines the theoretical foundations of autoencoder factor models, focusing on their representational power and limitations, complementing the applied-oriented discussion in Section 7; and it provides an overview of the state-of-the-art in the literature on these issues, highlighting open questions for further research.

**Data-Generating Process.** Let us introduce a new set of assumptions. Formally, we claim that there exists a (possibly nonlinear) mapping $\varphi : \mathbb{R}^n \to \mathbb{R}^r$ such that

$$\boldsymbol{x}_t = \varphi(\boldsymbol{x}_t) + \boldsymbol{\varepsilon}_t, \tag{A.1}$$

where the signal $\varphi(\boldsymbol{x}_t)$ and noise $\boldsymbol{\varepsilon}_t$ are uncorrelated in expectation,

$$\mathbb{E}[\varphi(\boldsymbol{x}_t)'\boldsymbol{\varepsilon}_t] = 0, \tag{A.2}$$

and the signal variance is approximately low-rank, with most variance concentrated along $r \ll n$ principal directions. That is,

$$\frac{\sum_{i=1}^{r} \lambda_i}{\sum_{j=1}^{n} \lambda_j} \approx 1, \tag{A.3}$$

where $\lambda_1 \geq \cdots \geq \lambda_n$ are the eigenvalues of $\mathrm{Cov}[\varphi(\boldsymbol{x}_t)]$.

This framework and the linear model have in common the assumption that the dominant variability in the data is captured by a few latent dimensions. Geometrically, data lie near an $r$-dimensional differentiable manifold. Unlike the linear case, however, this manifold may be non-affine—curved or twisted—and we require $\varphi$ to be fixed rather than time-dependent, a choice primarily motivated by computational tractability.

**On Universality in Latent (Ambient) Space.** Given this manifold structure, a natural question arises: Can a neural network recover such a mapping $\varphi$? The universal approximation theorem provides a formal foundation for this possibility, stating that a feedforward neural network with a single hidden layer and a non-polynomial activation function can approximate any continuous function on a compact domain to arbitrary precision, given sufficient hidden units (Cybenko, 1989; Hornik et al., 1989; Hornik, 1991). Equivalently, such networks are dense in $C(K)$—the space of continuous mappings on a compact subset $K \subset \mathbb{R}^r$—under the uniform norm.

Formally, for any continuous target mapping $f^\star : \mathcal{X} \to \mathcal{Z}$ and any $\varepsilon > 0$, there exists a neural network $f_\theta \in \mathcal{F}$ such that

$$\|f_\theta - f^\star\|_\infty < \varepsilon. \tag{A.4}$$

We say that the model is *universal in latent space* if its encoder family $\mathcal{F}$ satisfies the above property for mappings $\mathbb{R}^n \to \mathbb{R}^r$.

In our framework, the encoder is a multilayer perceptron with a non-polynomial activation (SiLU or sine). In principle, this allows recovery of any smooth factor structure, regardless of curvature. In practice, however, universality is bounded by the hidden layer's limited capacity, imposed by data scarcity. Conversely, the purely affine decoder violates the conditions for *universality in ambient space*: it can reconstruct only a best-fitting affine (flattened) approximation of the curved signal manifold.[6]

[6] Even with a nonlinear decoder, the conditions ensuring universality in ambient space—particularly regarding the dimensionality of the bottleneck—remain an open question.

**On Latent Space Representation.** While universal approximation ensures that a sufficiently flexible encoder can, in principle, capture the mapping $\varphi$, the corresponding latent-space representation is generally non-unique, even at the population level.

Consider the decomposition

$$\varphi = \gamma \circ \delta, \tag{A.5}$$

where $\delta : \mathbb{R}^n \to \mathbb{R}^r$ maps observations to a latent space and $\gamma : \mathbb{R}^r \to \mathbb{R}^n$ maps them back. Defining $\boldsymbol{f}_t = \delta(\boldsymbol{x}_t)$ yields a latent representation. For any diffeomorphism $\mu$ on $\mathbb{R}^r$, one can equivalently write

$$\varphi = \underbrace{(\gamma \circ \mu^{-1})}_{\text{new decoder}} \circ \underbrace{(\mu \circ \delta)}_{\text{new encoder}}, \tag{A.6}$$

yielding an alternative but equally valid representation.[7] Thus, the intrinsic latent structure is an equivalence class of parameterizations on an $r$-dimensional manifold. Structural constraints—such as sparsity or orthogonality—can guide learning toward more interpretable solutions.

**On the Number of Factors.** Having established the general structure of the latent space, a natural question concerns the appropriate number of latent dimensions. In classical linear settings, consistent estimators allow recovery of the true number of factors. In nonlinear models, however, this is more challenging because the reconstruction loss is insensitive to over-specification: adding an extra latent factor with zero loadings on all variables leaves the loss unchanged, yielding the same global minimum. Moreover, information criteria cannot be applied *ex post*—after cross-validating hyperparameters—since all hyperparameters, including the latent dimensionality, should be selected jointly.

In practice, one cross-validates the number of neurons in the bottleneck layer, which serves as the model's latent dimensionality, along with other hyperparameters (Xiu and Shen, 2024). This approach does not guarantee recovery of the true number of latent factors, but it selects a dimension that balances reconstruction accuracy and model complexity. A promising direction for future work is to penalize over-specification directly in the loss function via an information-criterion-style regularizer.

[7] Analogous to the linear case, where latent factors are identifiable only up to orthogonal rotations.

## References


Alessi, Barigozzi, C. (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics  Probability Letters*, 80(23-24):1806–1813.

Andreini, P., Izzo, C., and Ricco, G. (2023). Deep dynamic factor models. Working Papers 2023-08, Center for Research in Economics and Statistics.

Bai and Ng (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS)*, 2(4):303–314.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2009). Opening the black box: structural factor models versus structural vars. *Econometric Theory*, 25(5):1319–1347.

Hallin and Liška (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102(478):603–617.

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2):251–257.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366.

Stock and Watson (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

Su and Wang (2017). On time-varying factor models: Estimation and testing. *Journal of Econometrics*, 198(1):84–101.

Xiu, D. and Shen, Z. (2024). Deep autoencoders for nonlinear factor models: Theory and applications. Research Paper 25-14, Chicago Booth School of Business.