

# FARS: Factor Augmented Regression Scenarios in R

Gian Pietro Bellocca 

Universidad Carlos III de Madrid

Ignacio Garrón 

Universidad Carlos III de Madrid

C. Vladimir Rodríguez-Caballero 

Department of Statistics, ITAM

Department of Economics, Duke University

Esther Ruiz 

Universidad Carlos III de Madrid

---

## Abstract

In the context of macroeconomic/financial time series, the **FARS** package provides a framework in R for the construction of conditional densities of the variable of interest based on the factor-augmented quantile regressions (FA-QRs) methodology. Within this context, the factors used to estimate the quantiles are extracted from a multi-level dynamic factor model with potential overlapping group-specific factors, while the densities are obtained by matching the estimated quantiles to a Skewed-Student density. The package also allows the construction of measures of risk as well as designing economic scenarios for the conditional densities. In particular, the package enables users to: (i) extract global and group-specific factors using a flexible multi-level factor structure and compute asymptotically valid confidence regions for the estimated factors, accounting for uncertainty in the factor loadings; (ii) obtain estimates of the parameters of the FA-QRs together with their standard deviations, and recover full predictive conditional densities from estimated quantiles; (iii) obtain risk measures based on extreme quantiles of the conditional densities; and (iv) estimate the conditional density and the corresponding extreme quantiles when the factors are stressed.

*Keywords:* Multi-level dynamic factor model, Quantile regression, Scenario analysis, R.

---

## 1. Introduction

In the context of macroeconomic/financial time series, there is a growing interest in obtaining predictions of the probability densities of specific key variables; see, for example, Granger and Pesaran (2000a) and Granger and Pesaran (2000b), who argue that point forecasts are not sufficient from the perspective of a properly informed decision-maker. In addition to being of interest in themselves, these densities can also serve to obtain measures of macroeconomic vulnerability, which are crucial for the design of resilience policies; see, for example, Delle Monache, De Polis, and Petrella (2024). Furthermore, econometricians, policy makers, and financial analysts are also interested in the construction of realistic scenarios for the distribution of key variables that can help to further understand the resilience of economic systems by providing early warning signals of what to expect should such conditions materialize in adverse outlooks; see, for example, González-Rivera, Rodríguez-Caballero, and Ruiz (2024) and Adrian, Giannone, Luciani, and West (2025).

To start with, estimation of the conditional density of interest is often based on assuming that underlying economic and/or financial latent factors drive it. As proposed by [Bai and Ng \(2008\)](#) and [Ando and Tsay \(2011\)](#), and popularized by [Adrian, Boyarchenko, and Giannone \(2019\)](#), the quantiles of the distribution of the target variable can be estimated by fitting factor-augmented quantile regressions (FA-QRs) with underlying latent factors, which summarize economic and/or financial activity, as regressors. The FA-QR model allows for different impacts of underlying factors on different quantiles of the distribution of the variable of interest, and consequently, for potential asymmetries in the downside and upside risks. After estimating the quantiles, and following [Azzalini and Capitanio \(2003\)](#), the corresponding  $h$ -step-ahead conditional density is obtained by fitting a skew-t distribution to them. The skew-t distribution has been shown to be flexible enough to provide an appropriate approximation to the conditional density of a large number of economic variables; see [Mitchell, Poon, and Zhu \(2024\)](#) for alternative estimators of densities, which are shown to outperform the popular skew-t distribution in the unlikely case of multimodal distributions. The estimated conditional density delivers any quantile of interest, and, in particular, extreme quantiles, which are often used as measures of vulnerability as, for example, growth at risk (GaR) proposed by [Adrian et al. \(2019\)](#), or inflation at risk (IaR) as in [Lopez-Salido and Loria \(2024\)](#).

The factors needed as regressors for FA-QRs can be extracted from a dynamic factor model (DFM), with the preferred estimation method being principal components (PC); see, for example, [Bai \(2003\)](#) and [Bai and Ng \(2013\)](#) for technical details.<sup>1</sup> Over the last few decades, when dealing with large systems of economic variables, it is not unusual to empirically observe that some of the latent factors, which summarize the common movements in the system, only load on particular groups of variables. This block structure may represent economic, geographical, cultural, or other characteristics. In this context, PC may face difficulties. Alternatively, factors can be extracted from multi-level dynamic factor models (MLDFMs) with the matrix of factor loadings subjected to the adequate blocks of zero restrictions. The factor structure of the MLDFM allows for pervasive (or global) factors that are common across all variables in the system, as well as group-specific factors associated with one or more blocks of variables. The MLDFM can incorporate non-overlapping or overlapping blocks of variables. The factors of MLDFMs can be extracted using the sequential least squares (SLS) estimator proposed by [Breitung and Eickmeier \(2016\)](#) for non-overlapping factors and generalized by [Rodríguez-Caballero and Caporin \(2019\)](#) to overlapping factors. It is also important to note that, when the extracted factors are used as regressors of predictive regressions, obtaining measures of their uncertainty becomes relevant; see, for example, [Amburgey and McCracken \(2022\)](#) and [Lewis, Mertens, Stock, and Trivedi \(2022\)](#). The asymptotic distribution of the factors extracted by SLS is established by [Choi, Kim, Kim, and Kwark \(2018\)](#) for DFMs without overlapping factors and by [Lu, Jin, and Su \(2025\)](#) for overlapping factors.

Finally, in order to generate stressed scenarios (or stressed factors) for the conditional densities, the methodology proposed by [González-Rivera, Maldonado, and Ruiz \(2019\)](#) can be used. Under unexpected and rare circumstances, the factors driving the distribution of the variable of interest are under stress, and thus deviate substantially from their averages. Stressed factors are probabilistically derived based on their multidimensional distribution, focusing on the observations located on its extreme autocontours.

---

<sup>1</sup>Note that on top of being used as predictors of FA-QRs, there are many other applications in which the factors can be of interest in themselves as, for example, when using them to construct economic/financial indexes or as predictors of diffusion indexes; see the survey on DFMs by [Stock and Watson \(2011\)](#).

This paper presents the **FARS** package, which provides a framework in R for modeling and forecasting conditional densities based on MLDFMs and FA-QRs.<sup>2</sup> The package enables users to:

1. Use SLS to extract pervasive, semipervasive, and block-specific factors based on a flexible specification of the MLDFM. This approach allows for overlapping blocks, enabling the extraction of factors within a hierarchical structure with more than two levels, beyond the usual global-local configuration. To the best of our knowledge, there are no other published R packages that provide this functionality. The only closely related package available in R is **GCCfactor** by Lin and Shin (2023), which allows factor estimation in MLDFMs with non-overlapping structures. The estimation method is based on the generalized canonical correlation (GCC) estimator, which is conceptually related to SLS; see Lin and Shin (2022).<sup>3</sup> Alternatively, several open-source packages provide functionalities for multilevel modeling using different methods, including **lavaan** (Rosseel 2012) and **OpenMx** (Boker, Neale, Maes, Wilde, Spiegel, Brick, Spies, Estabrook, Kenny, Bates, Mehta, and Fox 2011) in R, as well as the **mfmodel** package in Python (Parshakova, Hastie, and Boyd 2025). Commercial implementations are also available, such as **GLLAMM** in Stata and in MLwiN; see Rabe-Hesketh, Skrondal, and Pickles (2004) and Rasbash, Browne, Goldstein, Yang, Plewis, Healy, Woodhouse, Draper, Langford, and Lewis (2000), respectively.
2. Compute asymptotically valid confidence regions for the SLS factors, accounting for uncertainty in the factor loadings, and for potential cross-correlations of the idiosyncratic components. As far as we are concerned, the only software that allows inference on factors is **GCCfactor**, where the confidence regions are based on bootstrap instead of being asymptotic.
3. Estimate FA-QR models and recover full predictive conditional densities from the estimated quantiles using the quantile-matching method to determine the parameters of a skew-t distribution. Finally, obtain risk measures such as GaR and IaR. A similar implementation is presented by Lajaunie, Flament, Hurlin, and Kazemi (2025), which provides an unpublished software in R to estimate factor-augmented quantile regressions and the corresponding densities.<sup>4</sup> Furthermore, the International Monetary Fund offers a Microsoft Excel toolkit that integrates macros and a user interface to perform

---

<sup>2</sup>Version 0.8.0 of the **FARS** package is available in CRAN: <https://CRAN.R-project.org/package=FARS>.

<sup>3</sup>Other implementations of DFM (not multilevel) are available in the R, although they are not published. The **sparseDFM** package implements popular estimation methods for DFM, including the recent Sparse DFM approach by Mosley, Chan, and Gibberd (2024); see Mosley, Chan, and Gibberd (2023). The **MARSS**, **KFAS** packages provide a flexible framework for modeling DFM within state-space structures (Holmes, Ward, Scheuerell, and Wills (2023) and Helske (2017)). Furthermore, the **dfms** package offers a broad suite of DFM estimation techniques under the assumption of idiosyncratic components independently and identically distributed (*i.i.d.*) (Krantz, Bagdziunas, Tikka, and Holmes 2025). Also, commercial software can be used to extract factors from DFM in Eviews, MATLAB, Stata and Julia; see, for example, Solberger and Spanger (2020) for the estimation of the DFM in the context of state-space models.

<sup>4</sup>There are alternative R packages for estimating quantile regressions (including linear, nonlinear, censored, locally polynomial, and additive models) such as the **quantreg** package by Koenker, Portnoy, Ng, Zeileis, Grosjean, and Ripley (2025). However, these implementations do not include factor-augmented specifications. Conversely, packages like **pls** by Mevik and Wehrens (2007) allow for factor-augmented regressions but do not estimate quantiles.

GaR estimation, based on [Adrian \*et al.\* \(2019\)](#); see [Prasad, Elekdag, Jeasakul, Lafarguette, Alter, Feng, and Wang \(2019\)](#) for further details. This toolkit relies on a Python backend that performs estimation, optimization, distribution fitting, simulations, and visualization.

4. Obtain scenarios for the conditional density and associated risk measures when the factors are stressed (via extreme autocontours of their joint distribution). None of the alternative software mentioned above implements such functionality.

The main computational challenge addressed in this paper is the flexibility of the factor extraction procedure to accommodate a wide range of model specifications with potentially overlapping factor structures. Previously, the code used to extract the factors had to be adapted for each particular model specification. This made the implementation of the procedure complicated for interested users, who needed to modify their code whenever the factor structure changed. Second, although also an important advantage, the design of **FARS** is modular, allowing users to supply their own factors or estimated densities and integrate them to obtain the remaining outputs. Third, the package implements subsampling by blocks, which introduces additional computational challenges. Finally, packaging the entire workflow in an open-source R package substantially lowers the barrier for applied researchers and policy institutions to implement advanced density-based methods.

The functionalities of the **FARS** package are illustrated through two different examples. In the first example, the underlying factors of headline inflation are extracted across a large number of European countries. The factors are subsequently used to estimate the conditional density of aggregate inflation in a selected country to assess the associated risk of high inflation, both when the economy is under business-as-usual conditions and when it is under stress. A second example considers building scenarios for the density of economic growth in the United States (US), as in [González-Rivera \*et al.\* \(2024\)](#).

The rest of this paper is organized as follows. The methodology is briefly described in Section 2. Section 3 describes the functionality of the **FARS** package. Section 4 is devoted to illustrating two empirical applications regarding European inflation and US GDP growth. Finally, Section 5 concludes.

## 2. Methodology

In this section, we provide a brief description of the methodology for extracting underlying factors and obtaining conditional density forecasts of the target variable under standard economic dynamics and stressed scenarios of the underlying factors. First, we discuss the factor structures involved in the DFM and MLDFMs (with and without overlapping blocks), and describe the asymptotic distribution of the PC estimated factors, assuming that idiosyncratic components are either cross-sectionally uncorrelated or weakly correlated. Second, we describe how to obtain forecasts of the density of the target variable under both stressed and non-stressed scenarios using FA-QRs.

### 2.1. Dynamic factor model (DFM)

The DFM has been extensively studied in the literature to reduce the dimensionality of large sets of variables by assuming that they can be represented by a relatively small number of

common underlying factors; see, for example, Stock and Watson (2002a,b), Bai (2003), and Bai and Ng (2013). Consider  $\mathbf{X}_t = (x_{1t}, \dots, x_{Nt})^\top$ , the  $N \times 1$  vector of weakly stationary variables observed at time  $t = 1, \dots, T$ . The DFM is given by

$$\mathbf{X}_t = \mathbf{P}\mathbf{F}_t + \boldsymbol{\varepsilon}_t, \quad (1)$$

where  $\mathbf{P} = (\mathbf{p}_1^\top, \dots, \mathbf{p}_N^\top)^\top$  is the  $N \times r$  matrix of factor loadings,  $\mathbf{F}_t = (F_{1t}, \dots, F_{rt})^\top$  is an  $r \times 1$  vector of weakly stationary latent factors, and  $\boldsymbol{\varepsilon}_t = (\epsilon_{1t}, \dots, \epsilon_{Nt})^\top$  is the  $N \times 1$  vector of idiosyncratic components, which are assumed to be weakly stationary and cross-sectionally weakly correlated, and uncorrelated with the common factors  $\mathbf{F}_t$ . Finally, the number of factors,  $r$ , is known.

In the model reported in Equation 1, the loadings and factors cannot be separately identified. They can only be estimated consistently up to a rotation of the factor space. Consequently, the standard identification restrictions often assumed in the literature are that  $\frac{1}{T}\mathbf{F}^\top\mathbf{F} = \mathbf{I}_r$ , and that  $\frac{1}{N}\mathbf{P}^\top\mathbf{P}$  is a diagonal matrix with distinct elements on the main diagonal, ordered from largest to smallest. Under these restrictions, estimated factors are identified up to a sign transformation; see Bai and Ng (2013) for more details about the identification of DFMs in the context of PC estimation.

In practice, factors are often estimated using PC. Let  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_T)^\top$  denote the  $T \times N$  matrix of observed data. The PC-estimated factors,  $\hat{\mathbf{F}}_t$ , are obtained as  $\sqrt{T}$  times the eigenvectors associated with the  $r$  highest eigenvalues of the matrix  $\mathbf{X}\mathbf{X}^\top$ , ordered in decreasing magnitude. The corresponding loading matrix is then estimated by  $\hat{\mathbf{P}}^\top = \frac{1}{T}\hat{\mathbf{F}}^\top\mathbf{X}$ .

## 2.2. Multi-level dynamic factor model (MLDFM)

In many economic/financial applications, the variables in  $\mathbf{X}_t$  are naturally grouped into blocks, such as countries, geographical regions, or economic sectors. In some cases, not all variables in  $\mathbf{X}_t$  load onto all factors in the DFM, which implies the presence of zeros in the matrix of loadings,  $\mathbf{P}$ . The standard PC approach is suboptimal in this context, as it neglects the block structure. Consequently, when the block structure is known, a more appropriate approach is to extract the factors from a MLDFM, where the relevant zero restrictions are imposed directly on  $\mathbf{P}$ . In what follows, we present two alternative specifications of the MLDFM, depending on whether the blocks of variables overlap.<sup>5</sup>

### *MLDFM without overlapping blocks*

Breitung and Eickmeier (2016) propose the following MLDFM with non-overlapping blocks

$$\begin{bmatrix} \mathbf{X}_{1t} \\ \vdots \\ \mathbf{X}_{Kt} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 & \boldsymbol{\lambda}_1 & 0 & \dots & 0 \\ \boldsymbol{\mu}_2 & 0 & \boldsymbol{\lambda}_2 & \dots & 0 \\ \vdots & & 0 & \ddots & 0 \\ \boldsymbol{\mu}_K & 0 & 0 & \dots & \boldsymbol{\lambda}_K \end{bmatrix} \begin{bmatrix} \mathbf{G}_t \\ \mathbf{F}_{1t} \\ \mathbf{F}_{2t} \\ \vdots \\ \mathbf{F}_{Kt} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{1,t} \\ \vdots \\ \boldsymbol{\varepsilon}_{K,t} \end{bmatrix}, \quad (2)$$

---

<sup>5</sup>It is important to note that the **FARS** package assumes that the MLDFM structure is known in the sense that the block structure and the number of global factors and factors within each block are previously determined by the user.

where, for  $k = 1, \dots, K$ ,  $\mathbf{X}_{kt}$  is the  $N_k \times 1$  vector of variables within block  $k$ , such that the cross-sectional dimension of  $\mathbf{X}_t = (\mathbf{X}_{1t}, \dots, \mathbf{X}_{Kt})'$  is  $N = \sum_{k=1}^K N_k$ , and  $\boldsymbol{\mu}_i$  and  $\lambda_i$  are  $N_i \times r_g$  and  $N_i \times r_k$  loading matrices of the global and block-specific factors on the variables within block  $i$ , respectively. Furthermore,  $\mathbf{G}_t = (\mathbf{G}_{1t}, \dots, \mathbf{G}_{r_G t})^\top$  is the  $r_G \times 1$  vector of pervasive factors that load on all variables in the system, while  $\mathbf{F}_{kt} = (\mathbf{F}_{1t}, \dots, \mathbf{F}_{r_k t})^\top$  is the  $r_k \times 1$  vector of block-specific factors that load only within the block  $\mathbf{X}_{kt}$ . The idiosyncratic noise are defined conformably; see [Wang \(2010\)](#), [Breitung and Eickmeier \(2016\)](#), [Choi et al. \(2018\)](#) and [Bellocca, Garrón, Rodríguez-Caballero, and Ruiz \(2026\)](#), for further technical details and identification conditions.

### MLDFM with overlapping blocks

For clarity of exposition of the MLDFM with overlapping blocks, consider the case with  $K = 3$ ; see [Rodríguez-Caballero and Caporin \(2019\)](#) for a detailed description.<sup>6</sup> Assume the presence of pervasive factors,  $\mathbf{G}_t$ , and block-specific factors,  $\mathbf{F}_{k,t} = (\mathbf{F}_{1t}^\top, \mathbf{F}_{2t}^\top, \mathbf{F}_{3t}^\top)^\top$ , as described above. In addition, a general factor structure may also include pairwise (or semipervasive) factors,  $\mathbf{F}_{kj,t} = (\mathbf{F}_{12,t}^\top, \mathbf{F}_{13,t}^\top, \mathbf{F}_{23,t}^\top)^\top$ . For instance, the factor  $\mathbf{F}_{12,t}$  loads on the variables in blocks  $\mathbf{X}_{1t}$  and  $\mathbf{X}_{2t}$ ; that is, the semipervasive factor captures the commonality only between blocks 1 and 2 without any dependence on block 3. This type of factor structure is illustrated in Figure 1, which represents the relationships between pervasive, semipervasive, and block-specific factors, when  $K = 3$ .

The MLDFM with overlapping blocks is given by

$$x_{kit} = \boldsymbol{\mu}_{ki}^\top \mathbf{G}_t + \boldsymbol{\kappa}_{kji}^\top \mathbf{F}_{kj,t} + \boldsymbol{\lambda}_{ki}^\top \mathbf{F}_{kt} + \varepsilon_{kit},$$

where  $k = 1, 2, 3$  indicates the block, index  $i = 1, \dots, N_k$  denotes the  $i$ 'th cross-section unit of block  $k$ ,  $t = 1, \dots, T$  is the period of time, and  $kj$  means interaction between blocks  $k$  and  $j \in (1, 2, 3)$  with  $k \neq j$ .  $\boldsymbol{\mu}_{ki}$ ,  $\boldsymbol{\kappa}_{kji}$ , and  $\boldsymbol{\lambda}_{ki}$  are the  $r_G$ ,  $r_{F_{kj}}$ , and  $r_{F_k}$ -dimensional vectors of factor loadings. The number of pervasive, pairwise, and block-specific factors can naturally vary in each block  $k$ . The idiosyncratic term denoted by  $\varepsilon_{k,it}$  satisfies the standard assumptions of the DFM defined in Equation 1.

The vector representation of the three-block MLDFM with overlapping blocks is given by

$$\begin{bmatrix} \mathbf{X}_{1t} \\ \mathbf{X}_{2t} \\ \mathbf{X}_{3t} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 & \boldsymbol{\kappa}_{12_1} & \boldsymbol{\kappa}_{13_1} & 0 & \boldsymbol{\lambda}_1 & 0 & 0 \\ \boldsymbol{\mu}_2 & \boldsymbol{\kappa}_{12_2} & 0 & \boldsymbol{\kappa}_{23_2} & 0 & \boldsymbol{\lambda}_2 & 0 \\ \boldsymbol{\mu}_3 & 0 & \boldsymbol{\kappa}_{13_3} & \boldsymbol{\kappa}_{23_3} & 0 & 0 & \boldsymbol{\lambda}_3 \end{bmatrix} \begin{bmatrix} \mathbf{G}_t \\ \mathbf{F}_{12,t} \\ \mathbf{F}_{13,t} \\ \mathbf{F}_{23,t} \\ \mathbf{F}_{1t} \\ \mathbf{F}_{2t} \\ \mathbf{F}_{3t} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix}. \quad (3)$$

Note that the total number of unobserved common factors involved in Equation 3 is  $r_G + r_{F_{12}} + r_{F_{13}} + r_{F_{23}} + r_{F_1} + r_{F_2} + r_{F_3}$ . [Hallin and Liška \(2011\)](#) and [Ergemen and Rodríguez-Caballero \(2023\)](#) propose a simple methodology based on the inclusion-exclusion principle of

<sup>6</sup>The **FARS** package supports  $K > 3$  blocks, including triple-wise (and higher-order) interactions. However, the computational burden naturally increases when the number of blocks and/or the order of interactions increases.

set theory to determine the number of pervasive, semipervasive and block-specific factors. However, the **FARS** package assumes that this number is known.

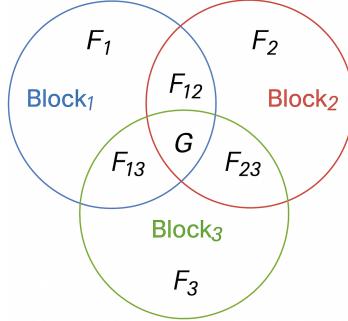


Figure 1: Factor structure of the MLDFM with three different overlapping blocks of data.

#### *Sequential least squares estimation*

The estimation of the MLDFM is based on the SLS approach proposed by Breitung and Eickmeier (2016) in which the main goal is to minimize the following residual sums of squares (RSS) function:

$$S(\hat{\mathbf{F}}_t, \hat{\mathbf{P}}) = \sum_{t=1}^T (\mathbf{X}_t - \hat{\mathbf{P}}\hat{\mathbf{F}}_t)^\top (\mathbf{X}_t - \hat{\mathbf{P}}\hat{\mathbf{F}}_t), \quad (4)$$

by a sequence of LS regressions. The algorithm can be executed for the general case of  $K$  blocks with overlapping factors as follows:

1. Obtain initial values of the factors as follows:
  - (a) Employ canonical correlation analysis (CCA) on  $\mathbf{X}_{k,t}$  to obtain initial estimates of the global factor,  $\hat{\mathbf{G}}^{(0)} = (\hat{\mathbf{G}}_1^{(0)}, \hat{\mathbf{G}}_2^{(0)}, \dots, \hat{\mathbf{G}}_T^{(0)})^\top$ .
  - (b) Filter out the global component by regressing  $\mathbf{X}_{k,t}$  on  $\hat{\mathbf{G}}^{(0)}$ , and get the corresponding residuals,  $\mathbf{X}_{kt}^{*(0)}$ , from each of the  $K$  separate regressions.
  - (c) Employ CCA on  $\mathbf{X}_{kt}^{*(0)}$  to obtain the following lower-level factors, selecting the corresponding blocks.
  - (d) Regress  $\mathbf{X}_{kt}^{*(0)}$  on the respective lower-level factors involved and get the residuals.
  - (e) Steps c) and d) are executed sequentially until the initial estimates of the pairwise block factors are obtained. Denote by  $\mathbf{X}_{kit}^{**(0)}$  the residuals after filtering the pairwise factors of each block  $k$ .
  - (f) Run PC on  $\mathbf{X}_{k,t}^{**(0)}$  to obtain the block-specific factors  $\hat{\mathbf{F}}^{(0)} = (\hat{\mathbf{F}}_{1t}^{(0)}, \hat{\mathbf{F}}_{2t}^{(0)}, \dots, \hat{\mathbf{F}}_{kt}^{(0)})^\top$ .
  - (g) The initial loading matrix,  $\hat{\mathbf{P}}^{(0)}$ , is estimated through time-series regressions of  $\mathbf{X}_{k,t}$  on global factors,  $\mathbf{X}_{kt}^*$  on semi-pervasive factors, and  $X_{kt}^{**}$  on non-pervasive factors.

2. Updated estimates for the unobserved factors,  $\hat{\mathbf{F}}^{(1)}$ , are obtained by LS regression of  $\mathbf{X}_{kt}$  on  $\hat{\mathbf{P}}^{(0)}$  as follows  $\hat{\mathbf{F}}^{(1)} = (\hat{\mathbf{P}}^{(0)\top} \hat{\mathbf{P}}^{(0)})^{-1} \hat{\mathbf{P}}^{(0)\top} \mathbf{X}_{k,t}$ .
3. The updated factors  $\hat{\mathbf{F}}^{(1)}$  are used to obtain the associated loadings matrix,  $\hat{\mathbf{P}}^{(1)}$ , as in Step 1.
4. Steps 2 and 3 are repeated until the RSS converges to a minimum, from which  $\hat{\mathbf{F}}^*$  and  $\hat{\mathbf{P}}^*$  are obtained.

Note that the SLS final estimates of the pervasive, semipervasive and non-pervasive factors estimate the space spanned by the factors and not the factors themselves. Therefore, particular rotations of interest of the factors may also be obtained after estimation. The **FARS** package provide the rotation for which the usual identification conditions are satisfied.

In step 1 of the algorithm, the initialization of  $\mathbf{P}^*$  and  $\mathbf{F}_t^*$  is carried out using CCA. Alternatively, the **FARS** package provides the alternative of using PC. Although both approaches produce approximately the same estimated common components  $\hat{\mathbf{P}}^* \hat{\mathbf{F}}_t^*$ , the convergence of CCA is typically faster, requiring fewer iterations to minimize RSS. However, when the factor structure is highly complex, initializing with PC tends to be computationally more efficient; see also [Breitung and Eickmeier \(2016\)](#) for the comparison of the small sample properties of the SLS estimator initialized with CCA and PC for the two-level DFM.

### 2.3. Asymptotic distribution of factors

The construction of probabilistic scenarios for the unobserved factors requires knowledge of their joint distribution. The asymptotic distribution of PC factors obtained from the DFM in Equation 1 is derived by [Bai \(2003\)](#). If  $\frac{\mathbf{F}'\mathbf{F}}{T} = \mathbf{I}_r$  and  $\frac{\sqrt{N}}{T} \rightarrow 0$  when  $N, T \rightarrow \infty$ , the asymptotic distribution of  $\hat{\mathbf{F}}_t$ , at each moment,  $t$ , is given by<sup>7</sup>

$$\sqrt{N} (\hat{\mathbf{F}}_t - \mathbf{F}_t) \xrightarrow{d} N(0, \Sigma_P^{-1} \Gamma_t \Sigma_P^{-1}), \quad (5)$$

where  $\Sigma_P = \lim_{N \rightarrow \infty} \frac{\mathbf{P}'\mathbf{P}}{N}$  and  $\Gamma_t = \lim_{N \rightarrow \infty} \sum_{i=1}^N \sum_{j=1}^N \mathbf{p}_i \mathbf{p}_j' E(\varepsilon_{it} \varepsilon_{jt})$  with  $\mathbf{p}_i$  and  $\varepsilon_{it}$  defined as in the DFM in Equation 1. The finite sample approximation of the asymptotic covariance matrix of  $\hat{\mathbf{F}}_t$  can be estimated as follows:

$$MSE_t = \left( \frac{\hat{\mathbf{P}}'\hat{\mathbf{P}}}{N} \right)^{-1} \frac{\hat{\Gamma}_t}{N} \left( \frac{\hat{\mathbf{P}}'\hat{\mathbf{P}}}{N} \right)^{-1}, \quad (6)$$

where  $\hat{\Gamma}_t$  is a consistent estimator for  $\Gamma_t$ . Under the assumption of cross-sectionally uncorrelated idiosyncratic components, [Bai and Ng \(2006\)](#) propose the following estimator:

$$\hat{\Gamma}_t^{BN} = \frac{1}{N} \sum_{i=1}^N \hat{\mathbf{p}}_i \hat{\mathbf{p}}_i' \hat{\varepsilon}_{it}^2, \quad (7)$$

where  $\hat{\varepsilon}_{it} = x_{it} - \hat{\mathbf{p}}_i' \hat{\mathbf{F}}_t$  are the residuals from the DFM model.

---

<sup>7</sup>Note that PC factors satisfy the identification restrictions. As such, they estimate the space spanned by the true factors and not the true factors themselves. Then, it is always possible rotating the factors to obtain an interpretation of interest. In this case, the asymptotic distribution should be accordingly modified.

In many empirical settings, assuming that the idiosyncratic covariance matrix  $\Sigma_\epsilon$  is diagonal imposes a stringent restriction that may not hold in practice. Therefore, alternatively, we relax this assumption allowing the idiosyncratic components to be weakly cross-sectionally correlated. Under these circumstances,  $\Gamma_t$  can be consistently estimated as proposed by [Fresoli, Poncela, and Ruiz \(2024\)](#) using adaptive thresholding of the sample covariances of the idiosyncratic residuals,  $\hat{\sigma}_{ij}$ , as follows:

$$\tilde{\Gamma}^{FPR} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \hat{\mathbf{p}}_i \hat{\mathbf{p}}_j' \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_{it} \hat{\varepsilon}_{jt} I(|\hat{\sigma}_{ij}| \geq c_{ij}), \quad (8)$$

where  $I(\cdot)$  is the indicator function that takes value one when the argument is true and zero otherwise, and  $c_{ij} = \delta \omega_{NT} [\widehat{Var}[\hat{\varepsilon}_{it} \hat{\varepsilon}_{jt}]]^{1/2}$ , with  $\widehat{Var}[\hat{\varepsilon}_{it} \hat{\varepsilon}_{jt}] = \frac{1}{T} \sum_{t=1}^T [\hat{\varepsilon}_{it} \hat{\varepsilon}_{jt} - \hat{\sigma}_{ij}]^2$ ,  $\omega_{NT} = \frac{1}{\sqrt{N}} + \sqrt{\frac{\log(N)}{T}}$ , and  $\delta$  chosen as proposed by [Qiu and Liyanage \(2019\)](#). It is important to note that the estimator of  $\Gamma_t$  in (8) requires stationarity and, consequently, is constant over time.

Regardless of whether  $\Gamma_t$  is obtained from (7) or (8), the estimated asymptotic covariance matrix in (6) does not account for the uncertainty arising from the estimation of the loading matrix. In this light, [Maldonado and Ruiz \(2021\)](#) propose a correction of the asymptotic MSE based on subsampling subsets of series of size  $N^* < N$  in the cross-sectional space, with each series in the subsample containing all temporal observations. For each subsample, the loadings and factors are estimated by PC, obtaining  $\hat{\mathbf{F}}_t^{*(s)}$  and  $\hat{\mathbf{P}}^{*(s)}$ , for  $s = 1, \dots, S$ . The corrected finite sample approximation of the asymptotic MSE of  $\hat{\mathbf{F}}_t$  can be estimated as follows:

$$MSE_t^* = \frac{1}{N} \left( \frac{\hat{\mathbf{P}}' \hat{\mathbf{P}}}{N} \right)^{-1} \hat{\Gamma}_t \left( \frac{\hat{\mathbf{P}}' \hat{\mathbf{P}}}{N} \right)^{-1} + \frac{N^*}{NS} \sum_{s=1}^S \left( (\hat{\mathbf{F}}_t^{*(s)} - \hat{\mathbf{F}}_t) (\hat{\mathbf{F}}_t^{*(s)} - \hat{\mathbf{F}}_t)' \right). \quad (9)$$

Based on the asymptotic normality in (5), [Maldonado and Ruiz \(2021\)](#) construct confidence ellipsoids for the estimated factors with coverage probability  $100 \times \alpha\%$  as follows:

$$g(\mathbf{F}_t, \alpha) = \{ \mathbf{F}_t \in \mathbb{R}^r | (\mathbf{F}_t - \hat{\mathbf{F}}_t) M S E_t^{*-1} (\mathbf{F}_t - \hat{\mathbf{F}}_t) \leq \chi_{r(\alpha)}^2 \}, \quad (10)$$

where  $\chi_{r(\alpha)}^2$  is the  $\alpha$ -quantile of the  $\chi^2$  distribution with  $r$  degrees of freedom, with  $r$  being the number of factors. Each point on the surface of the ellipsoid represents a possible joint realization of all factors in the DFM. These boundary points correspond to extreme, yet plausible, stress conditions.

## 2.4. Density forecasts under stressed and non-stressed conditions

The estimated factors, which summarize the information contained in a large set of predictors  $X_t$ , are used to estimate the temporal evolution of the conditional density of a target variable. In this section, we describe how these densities can be obtained under stressed and non-stressed conditions for the underlying factors.

Let  $y_t$  be the observation at time  $t$  of the target variable. We start by obtaining  $h$ -step-ahead forecasts of the  $\tau^*$ -quantile of the conditional distribution of  $y_t$  by estimating the following

FA-QR:

$$q_{\tau^*}(y_{t+h} | y_t, \mathbf{F}_t) = \mu(\tau^*, h) + \phi(\tau^*, h)y_t + \sum_{k=1}^r \beta_k(\tau^*, h)F_{kt}, \quad (11)$$

where  $\mu(\tau^*, h)$ ,  $\phi(\tau^*, h)$ , and  $\beta_k(\tau^*, h)$ , for  $k = 1, \dots, r$ , are parameters, and  $\mathbf{F}_t$  is the  $r \times 1$  vector of the underlying unobserved factors at time  $t$ . In practice, the underlying factors in (11) are replaced by their estimates,  $\hat{\mathbf{F}}_t$ , obtained as described above.

The parameters of the FA-QR model in (11) are estimated using the algorithm by Koenker and D’Orey (1987), which implements the quantile regression method originally developed by Koenker and Bassett (1978). When the error terms are assumed to be independently distributed according to a Laplace distribution, the estimator coincides with the Maximum Likelihood (ML) estimator; see Ando and Tsay (2011). Bai and Ng (2008) establishes its asymptotic normality.

The FA-QR provides estimates of the quantile function of the target variable,  $\hat{q}_{\tau^*}(y_{t+h}|y_t, \mathbf{F}_t)$ , for several values of  $\tau^*$ . However, in practice, it is challenging to map these estimates into a probability distribution function because of approximation errors and estimation noise. Consequently, as in Adrian *et al.* (2019), we use the skew-t distribution proposed by Azzalini and Capitanio (2003) to smooth the quantile function and estimate the conditional density of  $y_t$ . The skew-t density depends on four parameters as follows:

$$f(y; \mu, \sigma, \alpha, v) = \frac{2}{\sigma} st\left(\frac{y - \mu}{\sigma}; v\right) sT\left(\alpha \frac{y - \mu}{\sigma} \sqrt{\frac{v+1}{v + \left(\frac{y-\mu}{\sigma}\right)^2}}; v+1\right), \quad (12)$$

where  $st(\cdot)$  and  $sT(\cdot)$  denote the probability density function and the cumulative distribution function of the Student’s t distribution, respectively. The skew-t distribution is specified by its location  $\mu$ , scale  $\sigma$ , shape  $\alpha$ , and tail heaviness  $v$ . At each time  $t$ , a skew-t distribution is fitted by choosing the parameters that minimize the squared differences between the quantile estimates and the skew-t implied quantiles,  $q_{\tau^*}(y; \mu, \sigma, \alpha, v)$ , as follows:

$$(\hat{\mu}_{t+h}, \hat{\sigma}_{t+h}, \hat{\alpha}_{t+h}, \hat{v}_{t+h}) = \underset{\mu, \sigma, \alpha, v}{\operatorname{argmin}} \sum_{t=1}^{T-h} (\hat{q}_{\tau^*}(y_{t+h} | y_t, \mathbf{F}_t) - q_{\tau^*}(y_t; \mu, \sigma, \alpha, v))^2. \quad (13)$$

The methodology described above estimates the conditional density of  $y_t$  under non-stressed conditions. To construct conditional densities based on stressed scenarios, González-Rivera *et al.* (2019) and González-Rivera *et al.* (2024) use the confidence ellipsoids defined in Equation 10, and determine the value of the factors in the  $\alpha\%$ -contour (stress level of the underlying factors) that minimize (or maximize) a given quantile ( $\tau$ ) of the conditional distribution of the target variable. For instance, consider that we are interested in deriving a stress scenario for  $\tau = 0.05$ , with the factors stressed at their  $\alpha\%$  level, **FARS** solves the following optimization problem at each  $t$

$$\begin{aligned} & \min_{\mathbf{F}_t^{(S)}} \hat{q}_{0.05}(y_{t+h} | y_t, \mathbf{F}_t^{(S)}) \\ & \text{s.t. } g(\mathbf{F}_t^{(S)}, \alpha) = 0, \end{aligned} \quad (14)$$

where  $g(\mathbf{F}_t^{(S)}, \alpha) = 0$  is a predetermined  $\alpha$ -contour of the factors, that is, an ellipsoid that contains  $\mathbf{F}_t$  with probability  $\alpha$ .

The values of  $\mathbf{F}_t^{(S)}$  on the boundary of the ellipsoid  $g(\mathbf{F}_t^{(S)}, \alpha) = 0$  represent extreme events of the factors. After solving the optimization problem in Equation 14, these optimized values are plugged into the estimated FA-QRs. The conditional density of  $y_t$  under stress is then obtained by smoothing the corresponding quantiles as described in Equation 13.<sup>8</sup>

### 3. The FARS package

In this section, we provide a detailed overview of the **FARS** package functionalities and explain how users can implement the methodology described in Section 2 using the available functions.

#### 3.1. MLDFM in FARS

We begin by introducing the `mldfm()` function, which provides users with a flexible tool for extracting factors using DFM or MLDFM, with non-overlapping or overlapping blocks. In the case of a simple DFM, the function requires two arguments. The first is `data`, which contains the  $N$  variables from which the factors are extracted, structured as a  $T \times N$  matrix. The second argument is `global`, which specifies the number of factors  $r$  to be extracted from the data. By default, the input data are centered and scaled prior to factor extraction; this behavior can be controlled via the optional logical arguments `center` and `scale`, both set to `TRUE` by default.

In the case of the MLDFM without overlapping blocks, additional arguments must be provided to the `mldfm` function: i) the argument `blocks` defines the number of blocks  $K$  that make up the data sample (the default is 1, corresponding to the DFM case); ii) `block_ind` requires a vector that indicates the indices of the end column for each block  $k$ . For example, if  $K = 3$  and  $N = N_1 + N_2 + N_3$ , the argument `block_ind` should contain  $[N_1, N_1 + N_2, N_1 + N_2 + N_3]$ ; iii) the argument `local` is a vector of integers, indicating the number of block-specific factors  $r_{F_k}$  to be extracted from each block  $k$ ; iv) `global` specifies the number of pervasive factors  $r_G$ ; v) `method` defines the factor initialization strategy for the sequential LS estimation: 0 for the CCA (default) and 1 for PC<sup>9</sup>; vi) the arguments `tol` and `max_iter` define the tolerance level and the maximum number of iterations allowed for the RSS minimization process, with default values set to  $10^{-6}$  and 1000, respectively.

In the case of the MLDFM with overlapping blocks, an additional `middle_layer` argument must be provided. `middle_layer` is a named list, where each name is a string specifying a group of overlapping blocks (e.g.,  $kj$  in the case of pairwise groups), and each value is the number of factors  $r_{kj}$  to extract from that group. For example, if we want to extract one pairwise factor from blocks 1 and 3 ( $r_{13} = 1$ ), the list should be defined as `list("1-3" = 1)`.

Regardless of the particular specification of the model, the `mldfm()` function returns an S3 object of class `mldfm` as output. The object is a list containing several attributes described in Table 1.

The `mldfm` object has typical S3 methods: `print()`, `summary()` and `plot()`. The first two

---

<sup>8</sup>Note that the stressed scenarios are slightly different from that in González-Rivera *et al.* (2019) and González-Rivera *et al.* (2024), who obtain stressed factors for each quantile of the distribution.

<sup>9</sup>PC is implemented using the `prcomp()` function from the package `stats`.

Attribute	Description
<code>factors</code>	$T \times r$ matrix containing all the extracted factors.
<code>loadings</code>	$N \times r$ matrix of factor loadings with necessary zero restrictions.
<code>residuals</code>	$T \times N$ residual matrix from the model fit.
<code>fitted</code>	$T \times N$ matrix containing the fitted values $\hat{\mathbf{X}}$ .
<code>method</code>	The initialization strategy used (CCA or PCA).
<code>iterations</code>	Number of iterations performed until convergence (0 in DFM).
<code>factors_list</code>	A summary list indicating the number of factors extracted at each level.

Table 1: A description of `mldfm` object attributes. The data stored in the factors and loadings matrices follow the hierarchical order (from global to local) described in `factors_list`

functions offer a brief overview of the model estimation outcome, while `plot()` offers pre-configured visualization tools. The call of the `plot` function on a `mldfm` object generates distinct line charts for all estimated factors, each enriched with confidence interval bands that assume cross-sectionally independent and homoskedastic idiosyncratic components. Furthermore, an optional input argument `dates` can be provided. `dates` is a vector of dates to be displayed on the x-axis, replacing the default integer time index ranging from 1 to  $T$ . An additional optional argument, `flip`, can be supplied to improve the interpretability of the plots. `flip` is a binary vector (with values 0 or 1) indicating whether each estimated factor or the corresponding loadings should be sign-flipped before plotting. A value of 1 for a given element reverses its sign, while 0 leaves it unchanged. This is useful when the arbitrary sign indeterminacy of factor models leads to less interpretable visualizations. An optional argument, `fpr`, can be set to `TRUE` to estimate the asymptotic MSE of the factors using  $\hat{\Gamma}_{FPR}^t$  as defined in Equation 8. Differently, the default setup (`FALSE`) uses  $\hat{\Gamma}_t^{BN}$  as described in Equation 7. Moreover, using the `plot()` function, it is possible to visualize the estimated loadings or residuals, specifying a `which` argument with values "`loadings`" or "`residuals`". With "`loadings`", a set of bar charts is generated, one for each factor. Each bar chart displays the estimated loadings along with their corresponding confidence intervals. Differently, with "`residuals`", a figure depicting the correlation heatmap of the residuals is produced. In both cases, the user can provide a list of variable names using the optional `var_names` argument. This enables the replacement of the default indexes from VAR 1 to VAR  $N$  with the appropriate variable names. Specific attributes of the `mldfm` object can be accessed using appropriate functions, `factors()`, `loadings()`, `residuals()` and `fitted()`.

### 3.2. Probability distribution of factors in FARS

A two-step procedure is implemented in **FARS** to obtain the asymptotic joint probability density of the factors with the subsampling correction.

The first step involves running a subsampling method to extract factors from subsets of  $N^*$  variables, selected from the entire data sample. This is implemented using the `mldfm_subsampling()` function. The function iteratively generates `n_samples` subsamples of size `sample_size` and estimates factors using the MLDFM approach through the `mldfm()` function<sup>10</sup>. This approach offers two main advantages. First, the arguments of `mldfm_subsampling()`

<sup>10</sup>The argument `n_samples` is the number of samples, while `sample_size` is the proportion of the cross-

are the same as those of `mldfm()`, plus two additional arguments to define the number and size of the subsamples. Second, the function returns an object `mldfm_subsample` containing a list of `mldfm` objects, enabling the user to apply standard methods to each of the subsample results. In addition, an optional `seed` argument can be provided to ensure the reproducibility of the results. A `mldfm_subsample` object contains the attributes listed in Table 2 and provides `print()`, `summary()` and `plot()` methods, as well as `get_mldfm_list()` and `get_mldfm_model()` functions to access the entire list or a specific `mldfm` object, respectively.

Attribute	Description
<code>models</code>	A list containing the <code>n_samples</code> <code>mldfm</code> objects.
<code>n_samples</code>	The number of subsamples generated.
<code>sample_size</code>	The proportion of the sample used for each subsample $\frac{N^*}{N}$ .
<code>seed</code>	The seed used for random sampling.

Table 2: A description of `mldfm_subsample` object attributes.

The second step involves constructing confidence regions for the factors, as described in Equation 10. This operation is performed by the `create_scenario()` function, which requires three main arguments. The first is `model`, which contains the result of the `mldfm()` function applied to the full dataset and serves as the center of the ellipsoid. The second is `subsamples`, which uses the output of `mldfm_subsampling()` to compute the MSE correction as defined in Equation 9. The third is `alpha`, which defines the coverage probability (i.e., the level of stress) of the ellipsoids. An optional argument, `fpr`, can be set to `TRUE` to estimate the asymptotic MSE of the factors using  $\tilde{\Gamma}^{FPR}$  as defined in Equation 8. Differently, the default setup (`FALSE`) uses  $\hat{\Gamma}_t^{BN}$  as described in Equation 7. The output of `create_scenario()` is a `fars_scenario` object whose attributes are presented in Table 3. A `fars_scenario` object is provided with the standard S3 methos (`print()`, `summary()` and `plot()`) and with `get_ellipsoids()` and `get_sigma_list()` functions to access specific arguments. In particular, `get_ellipsoids()` returns a list of  $T$  matrices of size  $z \times r$  representing the ellipsoid points in  $r$  dimensions at each time  $t$ . The number of points  $z$  depends on the number of dimensions  $r$ . In the case of only one factor ( $r = 1$ ), only a confidence interval is built based on the specified `alpha` level; for this reason,  $z = 2$  (i.e., the upper and the lower bounds). In the case of two dimensions ( $r = 2$ ), the 2-D ellipsoid is composed of  $z = 300$  points and is built using the `ellipse` package; see Murdoch and Chow (2023). Lastly, in the case of more than two dimensions ( $r > 2$ ), the  $r$ -D ellipsoid is generated through the `hyperellipsoid()` and `hypercube_mesh()` functions from the `SyScSelection` package (Kopfmann 2023). In this case, the number of points composing the ellipsoid depends on the `phi` parameter of the `hypercube_mesh()` function, which defines the scalar fineness of the mesh. In **FARS**, `phi` is set to 8.

### 3.3. Conditional density under stressed and non-stressed conditions in FARS

In this section, we present the tools provided by **FARS** for obtaining conditional density forecasts in both the non-stressed and stressed scenarios.

---

sectional dimension,  $N$ , which composes the subsamples (e.g., 0.9 to select 90% of the original variables). In the case of multiple blocks, the proportion is maintained in all the blocks.

Attribute	Description
<code>ellipsoids</code>	A list containing $T$ matrices of dimensions $r \times z$ .
<code>center</code>	$T \times r$ matrix containing all the factors used as center for the ellipsoids.
<code>sigma</code>	A list of $T$ covariance matrices of dimensions $r \times r$ .
<code>periods</code>	Number of time periods $T$ .
<code>n_points</code>	Number of points $z$ used to define each ellipsoid.
<code>alpha</code>	Confidence level for the ellipsoids.

Table 3: A description of `fars_scenario` object attributes.

The first step is to estimate the FA-QRs<sup>11</sup>. This operation is performed through the `compute_fars()` function, which estimates the parameter of the FA-QR in Equation 11. The function requires only three arguments to work. First, `dep_variable`, which contains the dependent variable  $y_t$ . Second, `factors`, which includes the factors the user wants to add to the quantile regression model.<sup>12</sup> Third, `h`, which defines the forecast horizon (the default is  $h = 1$ ). The function estimates the FA-QRs for a fixed set of quantiles: 0.05, 0.25, 0.50, 0.75, and 0.95, as these are later used for the skew-t density fit. Alternatively, the user can modify the extreme quantiles by setting an optional `edge` argument. For example, setting `edge = 0.01` forces the edge quantiles to 0.01 and 0.99. The default value is 0.05. The output of `compute_fars()` is an S3 object of type `fars`, which contains a set of attributes listed in Table 4.

Attribute	Description
<code>models</code>	A list of 5 <code>rq</code> objects, each for a specific quantile regression.
<code>periods</code>	Number of time periods $T$ .
<code>n_factors</code>	Number of factors used in the regressions.
<code>h</code>	Lag order used in the regressions.
<code>levels</code>	The list of estimated quantiles.

Table 4: A description of `fars` object attributes.

Like the `mldfm` object, the `fars` object has standard S3 methods. The `print()` function provides a brief overview of the FA-QRs. The `summary()` function returns a detailed summary of each `rq` object by iteratively calling `summary()` on each of them. Lastly, the `plot()` function generates a line charts of the fitted values. The function can display customized dates on the x-axis by setting the corresponding optional argument `dates`. Additional methods are provided. A specific `rq` object can be accessed through `get_rq_model()`. The functions `fitted()` and `residuals()` return, respectively, a  $T \times Q$  matrix of fitted values and residuals estimated for the five quantiles ( $Q = 5$ ). The function `coef()` displays the estimated coefficients for all quantiles, while `logLik()` returns a `logLik` object corresponding to a user-specified quantile. Moreover, `predict()` allows out-of-sample estimation for the five quantiles, taking the required input data as a matrix through the `newdata` argument. An important feature of the `predict()` function is the ability to estimate stressed quantiles

<sup>11</sup>FARS estimate FA-QRs using the `quantreg` package (Koenker *et al.* 2025). The standard deviations of the estimated parameters are calculated using the sandwich formula proposed by Powell (1989) under the option `ker`, which is commonly used in practice.

<sup>12</sup>These can be easily accessed through the `factors` attribute of the `mldfm` object obtained after estimating the MLDFM by `mldfm()`.

by providing stressed factors in the `newdata` matrix instead of the baseline factors. These stressed factors can be obtained using the `compute_stressed_factors()` function, which implements the optimization procedure described in Section 2 (see Equation 13). In brief, the function estimates a FA-QR for the selected quantile level `qtau` and lag order `h`, and identifies the factor values on the contour ellipsoid that minimize or maximize the conditional quantile of the dependent variable. The resulting stressed factors can be then used in `predict()` to compute quantiles under stressed scenarios.

The second step to obtaining a density forecast is to estimate the density of the target variable  $y_t$  by fitting a skewed-t distribution. This operation is performed via the `compute_density()` function, which requires a `quantiles` argument, containing the quantiles estimated before<sup>13</sup>. Depending on the quantiles provided, the density function returns the non-stressed or the stressed density. Additional arguments can be provided to `compute_density()`, including `est_points`, which set the number of estimation points (default is 512), `random_samples`, which define the number of random samples to be drawn from the estimated distribution (default is 5000) and `support`, which select the lower and upper bounds of the random variable support (default is `c(-10, 10)`). For each period  $t$ , `compute_density()` initializes the skewed-t distribution by setting three parameters (location, scale, and shape) using the quantile values provided as input. The function implements two optimization procedures to fit the skew-t distribution. The default is a linear optimization using `optim()` from `stats`, which implements the L-BFGS-B method. The second is a non-linear optimization method that can be selected by setting the argument `n1 = TRUE`. The non-linear method is from the `nloptr` package and is based on NLOPT\_LN\_SBPLX (Johnson (2008)). In both cases, the theoretical quantiles and the probability distribution function (pdf) of the fitted skewed-t distribution are computed using `qst()` and `dst()` from `sn` (Azzalini (2026)), respectively. Finally, a `seed` argument can be provided to ensure the reproducibility of the results. The `compute_density()` function returns a `fars_density` object that provides the attributes listed in Table 5.

Attribute	Description
<code>density</code>	The estimated densities at time $t$ .
<code>distribution</code>	The random draws from the fitted skew-t distribution at each $t$ .
<code>optimization</code>	The optimization method implemented: linear or non-linear.
<code>eval_points</code>	The sequence of evaluation points used to compute the density.

Table 5: A description of `fars_density` object attributes. Both `density` and `distribution` are provided in matrix form with one row for each time  $t$ .

The `fars_density` object is equipped with standard S3 methods. The `print()` function provides a brief overview of the estimated density. The `summary()` function returns the mean, median, and standard deviation of the distribution at each time  $t$ . Finally, the `plot()` function generates a 3D plot of the density, with evaluation points (`eval_points`) on the x-axis, time indices on the y-axis, and density values on the z-axis. The function can also display custom dates on the y-axis by setting the optional argument `time_index`. The distribution is accessible through the `get_distribution()` function.

<sup>13</sup>If the quantiles computed with `compute_fars()` have been modified via the `edge` argument, the density function must be informed of the correct quantiles levels. This can be done by setting the `levels` argument using levels obtained by calling `get_quantile_levels()` on the `fars` object.

The final step in obtaining a conditional density forecast is to extract the conditional quantile from the estimated skew-t distribution. This can be performed using the function `quantile_risk()`. This function requires two parameters: an object of class `fars_density` and the quantile that must be extracted `qtau`. The quantile extraction is implemented via `quantile()` from `stats`. Depending on the `fars_density` object provided, either a non-stressed or a stressed density, the `quantile_risk()` extracts a non-stressed quantile or a stressed quantile of the target variable (e.g., in the case of GDP growth with `qtau = 0.05`, it extracts growth at risk or growth in stress).

Figure 2 shows a recap of the **FARS** package workflow for both the non-stressed and the stressed scenarios.

## 4. Illustration of FARS package functionalities

In this section, we illustrate the functionalities of the **FARS** package by extracting factors, estimating conditional densities and obtaining stress scenarios in the context of: i) aggregate inflation in Europe; and ii) growth in US. Regardless of the particular application, the first step is to install and load the package **FARS**, which is publicly available on CRAN under the GPL-3 license, as follows:

```
R> install.packages("FARS")
```

The development version is available on GitHub at <https://github.com/GPEBellocca/FARS>. This can be downloaded using the `devtools` package with the following command:

```
R> devtools::install_github("GPEBellocca/FARS")
```

After installing the package from CRAN or GitHub, it should be loaded as follows:

```
R> library("FARS")
```

### 4.1. Measuring inflation risk in Europe using FARS

In the first illustration, we analyze the risk of an inflation increase in Europe. To do this, we collect monthly headline CPI data (Ha, Kose, and Ohnsorge 2023) from January 2005 to December 2024 ( $T = 240$ ) for a set of  $N = 38$  European countries. The countries considered are divided into three different blocks, depending on geographical location:

- West ( $N_1 = 11$ ): Austria, Belgium, France, Germany, Ireland, Italy, Luxembourg, Portugal, Spain, Switzerland, United Kingdom;
- East ( $N_2 = 21$ ): Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czech Republic, Estonia, Greece, Hungary, Kosovo, Latvia, Lithuania, Malta, Moldova, Rep., North Macedonia, Poland, Romania, Slovakia, Slovenia, Turkey, Ukraine;
- North ( $N_3 = 6$ ): Denmark, Finland, Iceland, The Netherlands, Norway, Sweden;

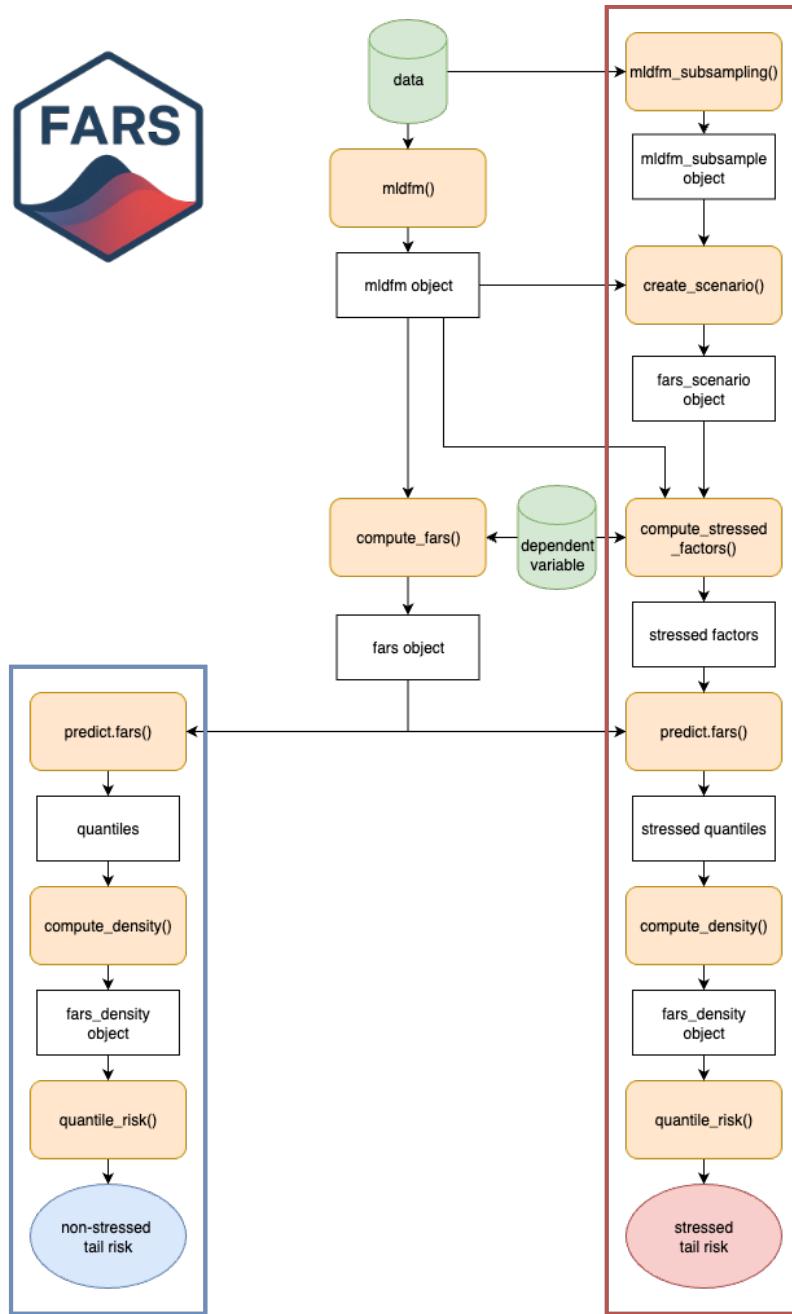


Figure 2: **FARS** package workflow for both non-stressed and stressed scenarios.

For each country, CPI prices are transformed into annualized month-on-month (mom) inflation, with each inflation series sequentially cleaned of seasonal effects and outliers. The processed data can be imported using:

```
R> data("inflation_data", package = "FARS")
```

To estimate a MLDFM through `mldfm()`, we first need to decide how many factors to extract from each block. We extract one global factor common to all  $N$  countries, and one block-

specific factor common to countries in each of the three blocks. This operation is performed as follows:

```
R> mldfm_result <- mldfm(inflation_data,
+   blocks = 3,
+   block_ind = c(11, 32, 38),
+   global = 1,
+   local = c(1, 1, 1))
```

Since we do not provide any `method`, `tol`, and `max_iter`, the default values are enforced. Also, by default, the dataset is standardized. The `mldfm` object returned is stored in the `mldfm_result` variable. After completion, the function `summary()` can be used to display an overview of the estimated MLDFM, including the number of factors extracted at each level of the hierarchical structure used in the SLS estimation.

```
R> summary(mldfm_result)

Summary of Multilevel Dynamic Factor Model (MLDFM)
=====
Number of periods      : 239
Number of factors       : 4
Number of nodes         : 4
Initialization method    : CCA
Number of iterations to converge: 33

Factor structure:
- 1-2-3 : 1 factor(s)
- 1 : 1 factor(s)
- 2 : 1 factor(s)
- 3 : 1 factor(s)

Residual diagnostics:
- Total residual sum of squares (RSS): 4471.70
- Average RSS per time period       : 18.71
```

Additionally, using `plot()`, it is possible to obtain a graphical representation of the estimated factors, loadings, and residuals. This is performed by calling the `plot` function three times in sequence. For a more precise result, we provide the `plot` function with appropriate arrays composed of dates and country names using the optional arguments. Also, we specify that the global factor must be flipped in sign. The results are plotted in Figures 3, 4 and 5, respectively.

```
R> plot(mldfm_result, dates = dates, flip = c(1, 0, 0, 0))
R> plot(mldfm_result, which = "loadings", var_names = countries,
+   flip = c(1, 0, 0, 0))
R> plot(mldfm_result, which = "residuals", var_names = countries)
```

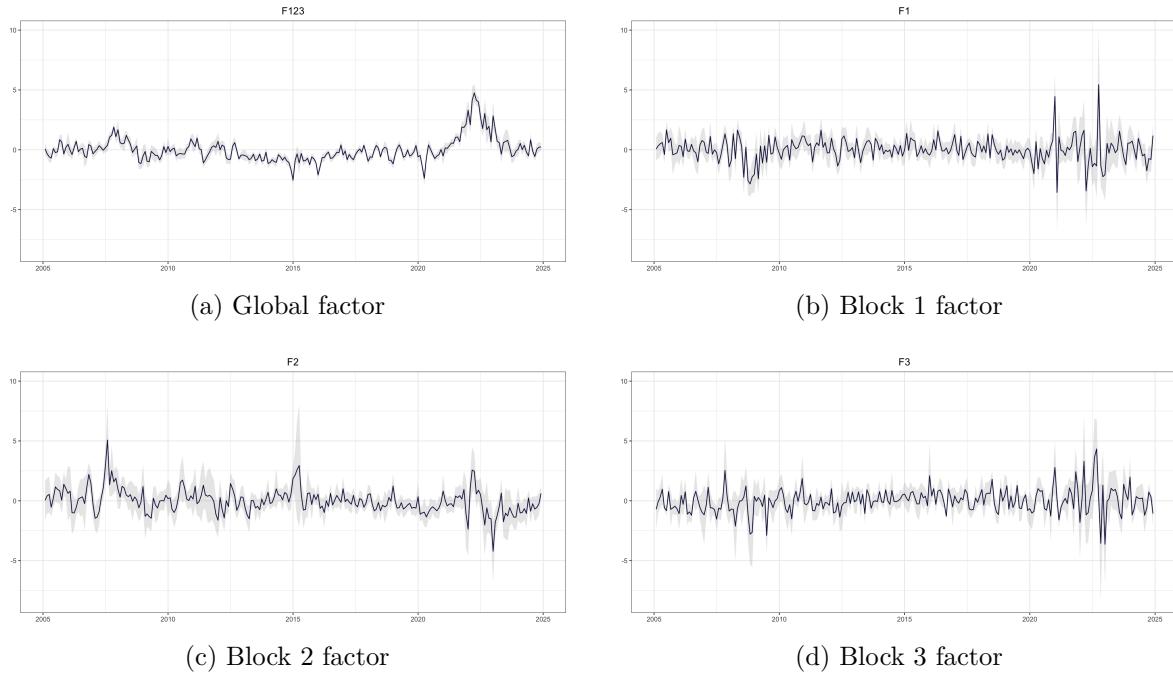


Figure 3: Estimated factors of headline inflation in Europe together with 95% confidence bounds.

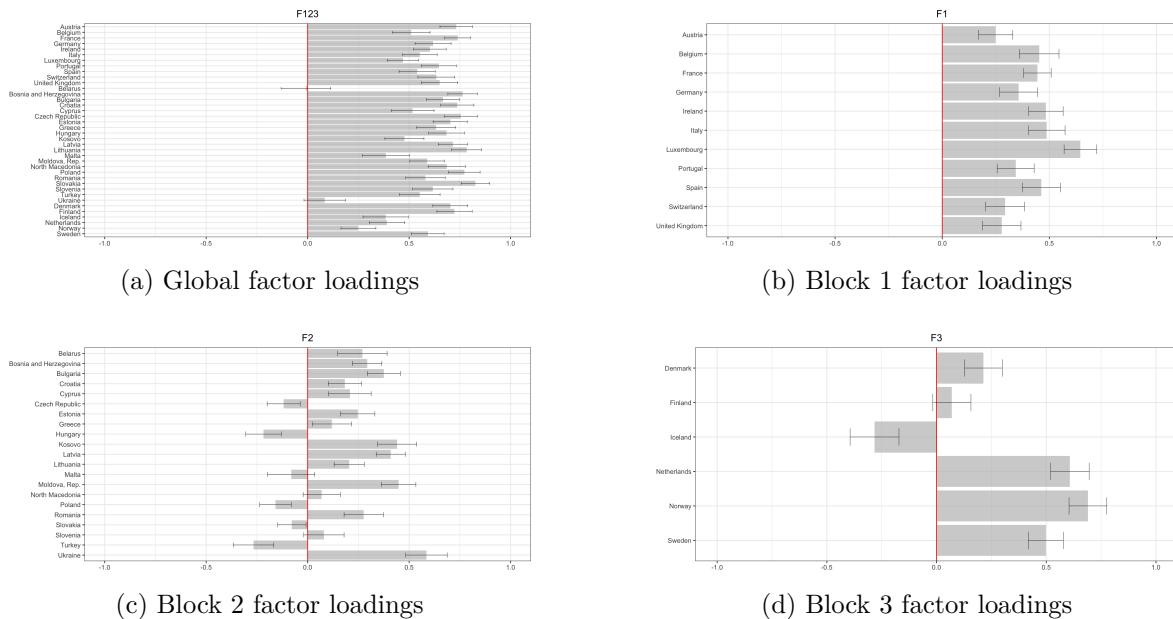


Figure 4: Estimated factor loadings of headline inflation in Europe together with 95% confidence bounds.

In order to analyze potential inflation risk in Europe we utilize Germany as an example. To do this, we extract the corresponding inflation series from the data set and the factors from `mldfm_result`.

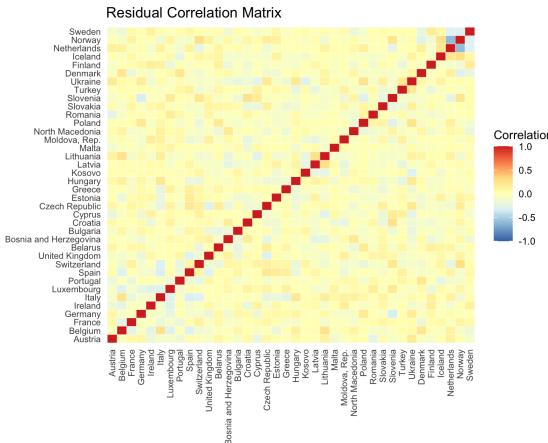


Figure 5: Correlation heatmap of estimated idiosyncratic components of headline inflation in Europe.

```
R> dep_variable <- as.numeric(inflation_data[[4]])
R> factors <- factors(mldfm_result)
```

The next step is to estimate the FA-QRs which will be used as baseline for our scenario analysis:<sup>14</sup>

```
R> fars_result <- compute_fars(dep_variable,
+   factors,
+   h = 1)
```

```
Running Factor-Augmented Quantile Regressions (FA-QRs)...
Completed
```

After this, we can print a recap of the FA-QRs and inspect the estimated coefficients across all quantile regressions. Additional information can be observed calling `summary()` function.

```
R> print(fars_result)

Factor-Augmented Quantile Regressions (FARS)
=====
Number of periods: 238
Number of factors: 4
Lag: 1
Quantile levels: 0.05 0.25 0.50 0.75 0.95
```

```
R> coef(fars_result)

      0.05    0.25    0.50    0.75    0.95
(Intercept) -2.554   0.030   2.066   3.475   6.964
```

<sup>14</sup>For this task, we consider the simplest case with `h = 1`.

LagY	0.096	0.046	0.024	-0.013	-0.215
F1	-0.909	-0.889	-1.216	-1.420	-2.531
F2	0.962	0.220	-0.216	-0.139	-0.211
F3	1.029	0.435	0.217	0.102	0.164
F4	-0.093	0.340	0.370	0.256	1.086

If the user is interested in performing additional diagnostic analyses on a specific quantile, the corresponding `rq` object can be retrieved as follows:

```
R> get_rq_model(fars_result, tau = 0.25)

Call:
quantreg::rq(formula = y ~ ., tau = tau_i, data = df)

Coefficients:
(Intercept)      LagY          F1          F2          F3          F4 
 0.03043957  0.04618208 -0.88931471  0.22043590  0.43533163  0.34005343 

Degrees of freedom: 238 total; 232 residual
```

### *Non-stressed scenario*

To estimate the density under the non-stressed scenario we first need the quantiles' series. For this example, we use the in-sample estimates obtained with the `fitted()` method of `fars` object.<sup>15</sup> The fitted values can be visualized using the `plot` function (see Figure 6, panel a).

```
R> quantiles <- fitted(fars_result)
R> plot(fars_result, dates = dates[-1])
```

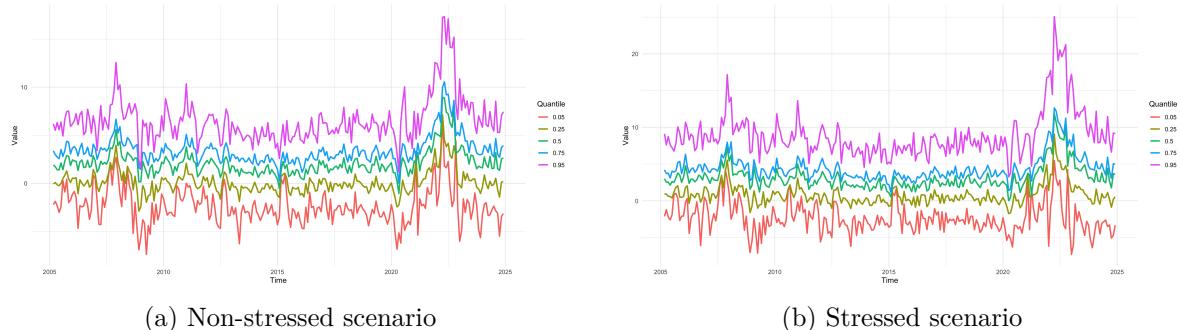


Figure 6: Non-stressed (left panel) and stressed scenario (right panel) quantiles for Germany headline inflation.

The estimates, stored in the `quantiles` variable, are then used to fit a skew-t distribution, generating the density for the non-stressed scenario. This is done by applying the non-linear optimization method and providing an appropriate support for the inflation case.

---

<sup>15</sup>The same result could be obtained using `predict()` and passing as `newdata` the same data used for fitting the FA-QRs.

```
R> ns_density <- compute_density(quantiles,
+ support = c(-30, 30),
+ seed = 42,
+ nl = TRUE)
```

```
Estimating skew-t densities from forecasted quantiles...
Completed
```

The generated `fars_density` object can be used to plot the non-stressed density (see Figure 7, panel a) and visualize an overview of the density estimation.

```
R> plot(ns_density, time_index = dates[-1])
R> print(ns_density)
```

#### FARS Density

---

```
Time observations : 238
Estimation points : 512
Random samples    : 5000
Support range     : [ -30 , 30 ]
Optimization       : Non-linear
```

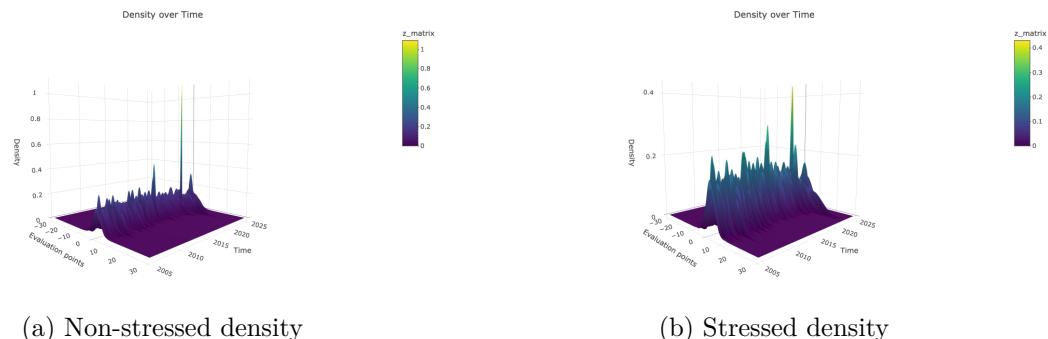


Figure 7: Non-stressed (left panel) and stressed (right panel) densities for Germany headline inflation.

Finally, we derive the inflation at risk (IaR) at  $q_{\tauau} = 0.99$  applying the `quantile_risk()` function to the non-stressed density.

```
R> IaR <- quantile_risk(ns_density, qtau = 0.99)
```

#### Stressed scenarios

As explained in Section 3, the computation of stressed scenarios can be performed in two steps. First, we need to obtain the asymptotic distribution of the factors. For this goal, we implement the subsampling procedure using the appropriate function. In our case, we generate 100 samples by extracting 95% of the countries in each block.

```
R> mldfm_ss_result <- mldfm_subsampling(inflation_data,
+   blocks = 3,
+   block_ind = c(11, 32, 38),
+   global = 1,
+   local = c(1, 1, 1),
+   n_samples = 100,
+   sample_size = 0.95,
+   seed = 42)
```

Generating 100 subsamples...  
Subsampling completed.

Each of the 100 models stored in `mldfm_ss_result` can be manipulated as a distinct `mldfm` object. For example, we can visualize the summary of the MLDFM estimated for sample number 10.

```
R> summary(get_mldfm_model(mldfm_ss_result, index = 10))
```

```
Summary of Multilevel Dynamic Factor Model (MLDFM)
=====
Number of periods      : 239
Number of factors       : 4
Number of nodes         : 4
Initialization method    : CCA
Number of iterations to converge: 23

Factor structure:
- 1-2-3 : 1 factor(s)
- 1 : 1 factor(s)
- 2 : 1 factor(s)
- 3 : 1 factor(s)

Residual diagnostics:
- Total residual sum of squares (RSS): 4197.49
- Average RSS per time period       : 17.56
```

The second step is to generate the stressed scenario by calling the `create_scenario()` function. For this exercise, we consider the highest stress level of `alpha = 0.99` and default  $\hat{\Gamma}_t^{\text{BN}}$ .

```
R> scenario <- create_scenario(model = mldfm_result,
+   subsample = mldfm_ss_result,
+   alpha = 0.99)
```

Constructing scenario using 100 subsamples, alpha = 0.99  
and standard time-varying Gamma...  
Scenario construction completed.

A summary of the scenario can be displayed as follows:

```
R> summary(scenario)
```

```
FARS Scenario Summary
=====
Number of periods      : 239
Ellipsoid dimensions : 4
Points per ellipsoid : 1072
Confidence level       : 99 %
FPR Gamma              : FALSE

Center (factor estimates):
  Mean      : 0
  Std. Dev  : 1.0005
  Min       : -4.7547
  Max       : 5.4379

Ellipsoid variability (diagonal of Sigma):
  Mean      : 0.3285
  Std. Dev  : 0.5681
  Min       : 0.0079
  Max       : 7.2701
```

Now that we have our scenario we can stress accordingly our factors. Since we are interested in inflation risk, our objective is to maximize the dependent variable for the chosen quantile ( $q_{\tauau} = 0.99$ ).

```
R> ellipsoids <- get_ellipsoids(scenario)
R> stressed_factors <- compute_stressed_factors(dep_variable,
+   factors,
+   h = 1,
+   qtau = 0.99,
+   direction = "max",
+   ellipsoids = ellipsoids)
```

Using the new factors, the stressed quantiles can be estimated with the `predict()` method of our `fars` object.<sup>16</sup> Moreover, by providing `newdata` to the `plot()` function, it is also possible to visualize the stressed quantiles (see Figure 6 panel b).

```
R> newdata <- cbind(dep_variable, stressed_factors)
R> newdata <- newdata[-nrow(newdata), ]
R> stressed_quantiles <- predict(fars_result, newdata)
R> plot(fars_result, newdata, dates[-1])
```

---

<sup>16</sup>The last row of the `newdata` matrix is dropped to keep only the in-sample period.

As in the non-stressed case, we fit a skew-t distribution to obtain the stressed density, this time using the matrix of stressed quantiles as input. Again, we can visualize the density with the `plot` function (see Figure 7, panel b).

```
R> s_density <- compute_density(stressed_quantiles,
+ support = c(-30, 30),
+ seed = 42,
+ nl = TRUE)
```

```
Estimating skew-t densities from forecasted quantiles...
Completed
```

The last step is deriving the inflation in stress (IiS) for  $\text{qtau} = 0.99$  by feeding `quantile_risk()` with the stressed densities.

```
R> IiS <- quantile_risk(s_density, qtau = 0.99)
```

In Figure 8, we plot the final IaR and IiS estimates along with the dependent variables for the period spanning from March 2005 to December 2024. We observe that IiS is higher than IaR. This worse outcome would be neglected if we only estimated IaR, which assumes that factors evolve according to an average scenario.

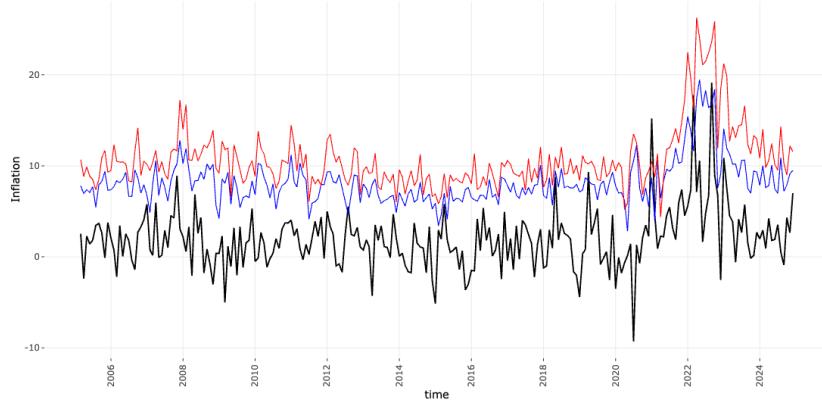


Figure 8: Germany monthly mom headline inflation (black), together with 99% IaR (blue) and 99% IiS stressed with  $\alpha = 99\%$ .

#### 4.2. Measuring growth risk in the United States using FARS

In our second illustration, we follow González-Rivera *et al.* (2024) and construct densities for annualized quarterly GDP growth in the US with the underlying factors extracted in the context of a MLDFM using a data sample composed of three blocks. The first block contains  $N_1 = 63$  international macroeconomic variables (GDP growth for 63 countries), the second block contains  $N_2 = 248$  domestic macroeconomic variables, and the third block contains  $N_3 = 208$  international financial variables. All variables are observed quarterly from 2005Q3

to 2020Q1.<sup>17</sup> The dataset, composed of  $N = N_1 + N_2 + N_3 = 519$  variables and the US GDP growth can be imported using:

```
R> data("mf_data", package = "FARS")
R> data("dep_variable", package = "FARS")
```

We extract one global factor common to all  $N$  variables, a pairwise factor common to all international variables (international macroeconomic and international financial blocks), and one block-specific factor common to the variables in each of the three blocks. Then, we check the summary of the model.

```
R> mldfm_result <- mldfm(mf_data,
+   blocks = 3,
+   block_ind = c(63, 311, 519),
+   global = 1,
+   local = c(1, 1, 1),
+   middle_layer = list("1-3" = 1))
R> summary(mldfm_result)

Summary of Multilevel Dynamic Factor Model (MLDFM)
=====
Number of periods : 59
Number of factors : 5
Number of nodes : 5
Initialization method : CCA
Number of iterations to converge: 47

Factor structure:
 - 1-2-3 : 1 factor(s)
 - 1-3 : 1 factor(s)
 - 1 : 1 factor(s)
 - 2 : 1 factor(s)
 - 3 : 1 factor(s)

Residual diagnostics:
 - Total residual sum of squares (RSS): 14770.03
 - Average RSS per time period : 250.34
```

Using the extracted factors, we fit our FA-QR models, this time changing the extreme quantiles to 1% and 99%.

```
R> factors <- factors(mldfm_result)
R> fars_result <- compute_fars(dep_variable,
+   factors,
+   h = 1,
+   edge = 0.01)
```

---

<sup>17</sup>Data are retrieved from the replication files of González-Rivera *et al.* (2024).

Running Factor-Augmented Quantile Regressions (FA-QRs) . . .  
Completed

```
R> print(fars_result)

Factor-Augmented Quantile Regressions (FARS)
=====
Number of periods: 58
Number of factors: 5
Lag: 1
Quantile levels: 0.01 0.25 0.50 0.75 0.99
```

### *Non-stressed scenario*

As in the inflation exercise, we use the estimated quantiles to fit the non-stressed density and compute the growth at risk for the lower tail at the 1% quantile. The quantiles and the resulting density distribution are shown in Figure 9, panel a, and Figure 10, panel a, respectively.

```
R> quantiles <- fitted(fars_result)
R> plot(fars_result, dates = dates[-1])
```

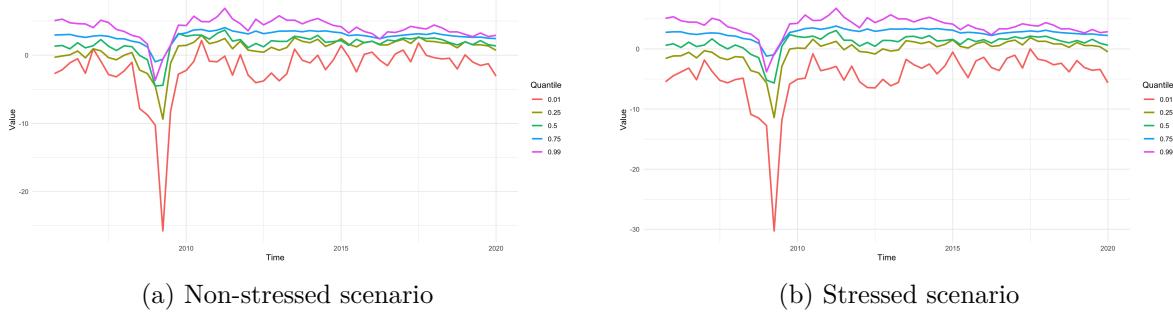


Figure 9: Non-stressed (left panel) and stressed scenario (right panel) quantiles for US GDP growth.

```
R> levels <- get_quantile_levels(fars_result)
R> ns_density <- compute_density(quantiles,
+ support = c(-30, 10),
+ levels = levels,
+ seed = 42)
```

Estimating skew-t densities from forecasted quantiles...  
Completed

```
R> plot(ns_density, time_index = dates[-1])
R> GaR <- quantile_risk(ns_density, qtau = 0.01)
```

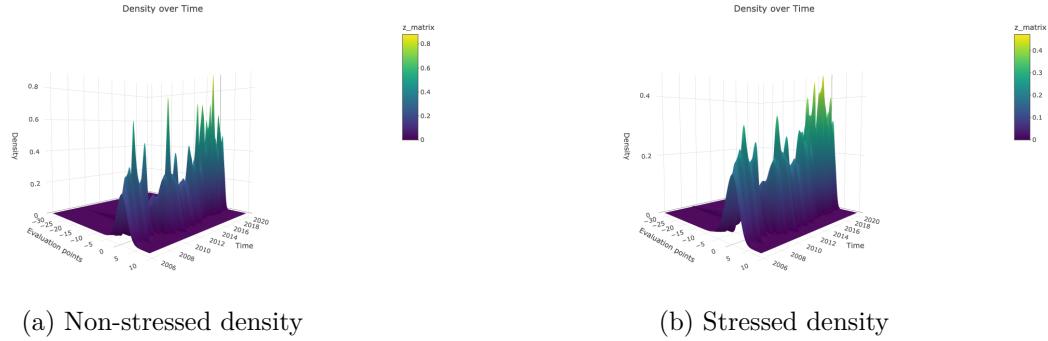


Figure 10: Non-stressed (left panel) and stressed (right panel) densities for US GDP growth.

### *Stressed scenario*

To build the stressed scenario we implement the same two-step procedure using the `mldfm_subsampling()` and `create_scenario()` functions. As for the inflation case, we generate 100 samples by extracting 95% of the variables from each block and consider the highest stress level of  $\alpha = 0.99$  with default  $\hat{\Gamma}_t^{\text{BN}}$ .

```
R> mldfm_ss_result <- mldfm_subsampling(mf_data,
+   blocks = 3,
+   block_ind = c(63, 311, 519),
+   global = 1,
+   local = c(1, 1, 1),
+   middle_layer = list("1-3" = 1),
+   n_samples = 100,
+   sample_size = 0.95,
+   seed = 42)

Generating 100 subsamples...
Subsampling completed.
```

```
R> scenario <- create_scenario(model = mldfm_result,
+   subsample = mldfm_ss_result,
+   alpha = 0.99)

Constructing scenario using 100 subsamples, alpha = 0.99
and standard time-varying Gamma...
Scenario construction completed.
```

Now we can stress our factors. Since we are interested in GDP growth risk, our objective is to minimize the dependent variable for the chosen low quantile (`qtau = 0.01`).

```
R> ellipsoids <- get_ellipsoids(scenario)
R> stressed_factors <- compute_stressed_factors(dep_variable,
+   factors,
```

```
+ h = 1,
+ qtau = 0.01,
+ direction = "min",
+ ellipsoids = ellipsoids)
```

The stressed factors are input into the FA-QRs to estimate GDP growth quantiles under stress, from which we derive the density distribution and the corresponding growth in stress. The resulting quantiles and density distribution are shown in Figure 9, panel b), and Figure 10, panel b), respectively.

```
R> newdata <- cbind(dep_variable, stressed_factors)
R> newdata <- newdata[-nrow(newdata), ]
R> stressed_quantiles <- predict(fars_result, newdata)
R> plot(fars_result, newdata, dates[-1])
R> s_density <- compute_density(stressed_quantiles,
+ support = c(-30, 10),
+ levels = levels,
+ seed = 42)
```

```
Estimating skew-t densities from forecasted quantiles...
Completed
```

```
R> plot(s_density, time_index = dates[-1])
R> GiS <- quantile_risk(s_density, qtau = 0.01)
```

In Figure 11, we plot the in-sample GaR and GiS estimates along with the dependent variables. As in González-Rivera *et al.* (2024), we observe that GiS is more negative than GaR. This negative outcome would be neglected if we only estimated GaR, which assumes that factors evolve according to an average scenario.

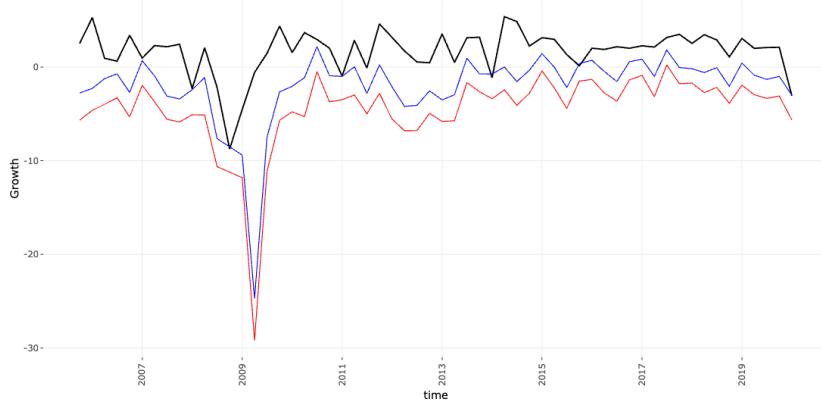


Figure 11: US quarterly growth: observed annualized rates (black), 1% GaR (blue), and 1% GiS stressed with  $\alpha = 99\%$  (red).

## 5. Summary and discussion

The **FARS** package offers a suite of tools in R for modeling and designing economic scenarios based on conditional densities derived from MLDFMs and FA-QRs. It provides an integrated framework to extract multilevel factors with a flexible hierarchical structure, estimate their relationship with macroeconomic variables, and build stressed scenarios. The empirical applications presented in this paper illustrate how **FARS** can be used to analyze the distribution of European inflation and US GDP growth under adverse conditions, linking factor dynamics to macroeconomic outcomes. The **FARS** package is available on the Comprehensive R Archive Network (CRAN).

## Acknowledgments

Financial support from the Spanish Government grant PID2022-139614NB-C22/AIE/10.13039/501100011033 (MINECO/FEDER) is gratefully acknowledged by all authors. We are also very grateful to the editors and referees for their very constructive and useful comments, which helped to improve both the paper and the associated **FARS** package. Thanks are also due to Enrico Carraro, who, by using the **FARS** package, has been very helpful in finding some bugs, and to participants at the Computational and Financial Econometrics Conference (London, December 2025) and the Econometrics Seminar of the Statistics Department of the Universidad Carlos III de Madrid, for their comments. Any remaining errors are the sole responsibility of the authors.

## References

- Adrian T, Boyarchenko N, Giannone D (2019). “Vulnerable Growth.” *American Economic Review*, **109**, 1263–1289.
- Adrian T, Giannone D, Luciani M, West M (2025). “Scenario Synthesis and Macroeconomic Risk.” *IMF Working Papers*, **2025**(105), 1. ISSN 1018-5941. doi:10.5089/9798229009225.001. URL <http://dx.doi.org/10.5089/9798229009225.001>.
- Amburgey A, McCracken M (2022). “On the Real-Time Predictive Content of Financial Condition Indices for Growth.” *Journal of Applied Econometrics*, **38**, 137–163.
- Ando T, Tsay RS (2011). “Quantile Regression Models with Factor-Augmented Predictors and Information Criterion.” *Econometrics Journal*, **14**(1), 1–24.
- Azzalini A (2026). *The R Package sn: The Skew-Normal and Related Distributions Such as the Skew-t and the SUN (Version 2.1.2)*. Università degli Studi di Padova, Italia. URL <https://cran.r-project.org/package=sn>.
- Azzalini A, Capitanio A (2003). “Distributions Generated by Perturbation of Symmetry With Emphasis on a Multivariate Skew t-Distribution.” *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, **65**, 367–389.

- Bai J (2003). “Inferential Theory for Factor Models of Large Dimensions.” *Econometrica*, **71**(1), 135–171.
- Bai J, Ng S (2006). “Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions.” *Econometrica*, **74**, 1133–1150.
- Bai J, Ng S (2008). “Forecasting Economic Time Series Using Targeted Predictors.” *Journal of Econometrics*, **146**, 304–317.
- Bai J, Ng S (2013). “Principal Components Estimation and Identification of Static Factors.” *Journal of Econometrics*, **176**, 18–29.
- Bellocca GP, Garrón I, Rodríguez-Caballero V, Ruiz E (2026). “The Empirical Distribution of Sequential LS Factors in Multi-Level Dynamic Factor Models.” ArXiv preprint, URL <https://arxiv.org/abs/2602.14813>.
- Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, Spies J, Estabrook R, Kenny S, Bates T, Mehta P, Fox J (2011). “**OpenMx**: An Open-Source Extended Structural Equation Modeling Framework.” *Psychometrika*, **76**(2), 306–317. doi:10.1007/s11336-010-9200-6. URL <https://doi.org/10.1007/s11336-010-9200-6>.
- Breitung J, Eickmeier S (2016). “Analyzing International Business and Financial Cycles Using Multi-Level Factor Models: A Comparison of Alternative Approaches.” In *Advances in Econometrics*, volume 35.
- Choi I, Kim D, Kim YJ, Kwark NS (2018). “A Multilevel Factor Model: Identification, Asymptotic Theory and Applications.” *Journal of Applied Econometrics*, **33**, 355–377.
- Delle Monache D, De Polis A, Petrella I (2024). “Modeling and Forecasting Macroeconomic Downside Risk.” *Journal of Business & Economic Statistics*, **42**(3), 1010–1025.
- Ergemen YE, Rodríguez-Caballero CV (2023). “Estimation of a Dynamic Multi-Level Factor Model with Possible Long-Range Dependence.” *International Journal of Forecasting*, **39**(1), 405–430.
- Fresoli D, Poncela P, Ruiz E (2024). “Dealing with Idiosyncratic Cross-Correlation when Constructing Confidence Regions for PC Factors.” *arXiv preprint*. URL <https://arxiv.org/pdf/2407.06883v1.pdf>.
- González-Rivera G, Maldonado J, Ruiz E (2019). “Growth in Stress.” *International Journal of Forecasting*, **35**, 948–966.
- González-Rivera G, Rodríguez-Caballero CV, Ruiz E (2024). “Expecting the Unexpected: Stressed Scenarios for Economic Growth.” *Journal of Applied Econometrics*, **39**, 926–942.
- Granger C, Pesaran M (2000a). “Decision Theoretic Approach to Forecast Evaluation.” In W Chan, W Li, H Tong (eds.), *Statistics and Finance: An Interface-Proceesing of the Hong Kong International Workshop on Statistics and Finance*. World Scientific.
- Granger C, Pesaran M (2000b). “Economic and Statistical Measures of Forecast Accuracy.” *Journal of Forecasting*, **19**, 537–560.

- Ha J, Kose MA, Ohnsorge F (2023). “One-stop Source: A Global Database of Inflation.” *Journal of International Money and Finance*, **137**, 102896. ISSN 0261-5606. doi:<https://doi.org/10.1016/j.jimonfin.2023.102896>. URL <https://www.sciencedirect.com/science/article/pii/S0261560623000979>.
- Hallin M, Liška R (2011). “Dynamic Factors in the Presence of Blocks.” *Journal of Econometrics*, **163**(1), 29–41.
- Helske J (2017). “**KFAS**: Exponential Family State Space Models in R.” *Journal of Statistical Software*, **78**, 1–39.
- Holmes EE, Ward EJ, Scheuerell MD, Wills K (2023). **MARSS**: Multivariate Autoregressive State-Space Modeling. R package version 3.11.9, URL <https://CRAN.R-project.org/package=MARSS>.
- Johnson SG (2008). *The NLOpt nonlinear-optimization package*. URL <https://github.com/stevengj/nlopt>.
- Koenker R, Bassett G (1978). “Regression Quantiles.” *Econometrica*, **46**, 33–50.
- Koenker R, Portnoy S, Ng PT, Zeileis A, Grosjean P, Ripley BD (2025). **quantreg**: Quantile Regression. R package version 6.1, URL <https://CRAN.R-project.org/package=quantreg>.
- Koenker RW, D’Orey V (1987). “Computing Regression Quantiles.” *Journal of the Royal Statistical Society Series C: Applied Statistics*, **36**, 383–393.
- Kopfmann M (2023). **SyScSelection**: Systematic Scenario Selection for Stress Testing . R package version 1.0.2, URL <https://CRAN.R-project.org/package=SyScSelection>.
- Krantz S, Bagdziunas R, Tikka S, Holmes E (2025). **dfms**: Dynamic Factor Models. R package version 0.3.0, URL <https://CRAN.R-project.org/package=dfms>.
- Lajaunie Q, Flament G, Hurlin C, Kazemi S (2025). *At Risk*. R package version 0.2.0, URL <https://CRAN.R-project.org/package=atRisk>.
- Lewis D, Mertens K, Stock J, Trivedi M (2022). “Measuring Real Activity Using a Weekly Economic Index.” *Journal of Applied Econometrics*, **37**, 667–687.
- Lin R, Shin Y (2022). “Generalised Canonical Correlation Estimation of the Multilevel Factor Model.” *SSRN Electronic Journal*. doi:[10.2139/ssrn.4295429](https://doi.org/10.2139/ssrn.4295429).
- Lin R, Shin Y (2023). **GCCfactor**: GCC Estimation of the Multilevel Factor Model. R package version 1.0.1, URL <https://CRAN.R-project.org/package=GCCfactor>.
- Lopez-Salido D, Loria F (2024). “Inflation at Risk.” *Journal of Monetary Economics*, **145, Supplement**, 103570.
- Lu X, Jin S, Su L (2025). “Three-Dimensional Factor Models with Global and Local Factors.” *Econometric Theory*.
- Maldonado J, Ruiz E (2021). “Accurate Confidence Regions for Principal Components Factors.” *Oxford Bulletin of Economics and Statistics*, **83**, 1432–1453.

- Mevik BH, Wehrens R (2007). “The **pls** Package: Principal Components and Partial Least Squares Regression in R.” *Journal of Statistical Software*, **18**, 1–23.
- Mitchell J, Poon A, Zhu D (2024). “Constructing Density Forecasts from Quantile Regressions: Multimodality in Macrofinancial Dynamics.” *Journal of Applied Econometrics*, **39**, 790–812.
- Mosley L, Chan TS, Gibberd A (2023). **sparseDFM**: Estimate Dynamic Factor Models with Sparse Loadings. R package version 1.0, URL <https://CRAN.R-project.org/package=sparseDFM>.
- Mosley L, Chan TST, Gibberd A (2024). “The Sparse Dynamic Factor Model: a Regularised Quasi-Maximum Likelihood Approach.” *Statistics and Computing*, **34**, 1–19.
- Murdoch D, Chow ED (2023). **ellipse**: Functions for Drawing Ellipses and Ellipse-Like Confidence Regions . R package version 0.5.0, URL <https://CRAN.R-project.org/package=ellipse>.
- Parshakova T, Hastie T, Boyd S (2025). “Fitting Multilevel Factor Models.” *SIAM Journal on Matrix Analysis and Applications*, **46**(3), 1930–1963.
- Powell J (1989). “Estimation of Monotonic Regression Models under Quantile Restrictions.” In *Barnett, W.A., J.L. Powell and G.E. Tauchen (eds.), Nonparametric and Semiparametric Methods in Econometric and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics*.
- Prasad A, Elekdag S, Jeasakul P, Lafarguette R, Alter A, Feng AX, Wang C (2019). “Growth at Risk: Concept and Application in IMF Country Surveillance.” *IMF Working Papers*, **19**, 1. ISSN 1018-5941. doi:[10.5089/9781484397015.001](https://doi.org/10.5089/9781484397015.001).
- Qiu Y, Liyanage J (2019). “Threshold Selection for Covariance Estimation.” *Biometrics*, **75**, 895–905.
- Rabe-Hesketh S, Skrondal A, Pickles A (2004). “GLLAMM Manual.” *Technical report*, Berkeley Electronic Press.
- Rasbash J, Browne W, Goldstein H, Yang M, Plewis I, Healy M, Woodhouse G, Draper D, Langford I, Lewis T (2000). “A User’s Guide to **MLwiN**.” *London: Institute of Education*, **286**.
- Rodríguez-Caballero CV, Caporin M (2019). “A Multilevel Factor Approach for the Analysis of CDS Commonality and Risk Contribution.” *Journal of International Financial Markets, Institutions and Money*, **63**, 101144.
- Rosseel Y (2012). “**lavaan**: An R Package for Structural Equation Modeling.” *Journal of statistical software*, **48**(1), 1–36.
- Solberger M, Spanger E (2020). “Estimating a Dynamic Factor Model in EViews Using the Kalman Filter and Smoother.” *Computational Economics*, **55**, 875–900.
- Stock JH, Watson MW (2002a). “Forecasting Using Principal Components from a Large Number of Predictors.” *Journal of the American Statistical Association*, **97**, 1167–1179.

- Stock JH, Watson MW (2002b). “Macroeconomic Forecasting Using Diffusion Indexes.” *Journal of Business and Economic Statistics*, **20**(2), 147–162.
- Stock JH, Watson MW (2011). “Dynamic Factor Models.” *The Oxford Handbook of Economic Forecasting*.
- Wang P (2010). “Large dimensional factor models with a multi-level factor structure: Identification, estimation, and inference.” *Manuscript, New York University*.

**Affiliation:**

Gian Pietro Bellocca, Ignacio Garrón, Esther Ruiz

Department of Statistics

Universidad Carlos III de Madrid

E-mail: [gbellocc@est-econ.uc3m.es](mailto:gbellocc@est-econ.uc3m.es), [igarron@est-econ.uc3m.es](mailto:igarron@est-econ.uc3m.es), [ortega@est-econ.uc3m.es](mailto:ortega@est-econ.uc3m.es)

C. Vladimir Rodríguez-Caballero

Department of Statistics

Instituto Tecnológico Autónomo de México

Department of Economics

Duke University

E-mail: [vladimir.rodriguez@itam.mx](mailto:vladimir.rodriguez@itam.mx)