

# **Introduction to Multi-step DFM-based Forecasts of Multiple Volatility Indexes**

Hui “Fox” Ling, [h.fox.ling@gmail.com](mailto:h.fox.ling@gmail.com), i4cast LLC.

Initial Version: February 2025, This Version: December 2025.

## **Abstract**

Values of volatility indexes represent expected volatility levels of underlying time-series, expected at a specified future time. Therefore, volatility indexes are closely related to forecasted volatilities of underlying time-series. This paper introduces a modeling method to make (a) multi-step forecasts of multiple volatility indexes and (b) dynamic attributions of forecasted volatilities by applying dynamic factor model (DFM) volatility analysis on DFM-input vector time-series including both (1) price time-series underlying the volatility indexes and (2) other information-enhancing time-series. Two forecasting approaches are detailed: (I) forecasting changes in volatility indexes with forecasted changes in volatilities of underlying time-series and (II) making quadratic autoregressive (QAR) forecasts of volatility indexes with QAR equation of volatilities of underlying time-series. With DFM modeling, these approaches are able to attribute volatility index forecasts to various dynamic sources of volatility, e.g. variance of common factors and idiosyncratic UDCs (unobserved dynamic components), autocovariance of factors and UDCs, and prediction/residual errors.

**Keywords:** Volatility index, multi-step forecasts of multiple volatility indexes, volatility forecasts by DFM-based volatility model, dynamic volatility attributions, forecasting volatility

components, volatility prediction errors.

## **List of Sections**

1. Introduction
2. Data: Four Sets of Multiple Time-Series
3. Volatility Model on DFM-Input Time-Series
4. Forecasts by DFM Volatility Model
5. Expressions of Underlying Volatilities
6. Dynamic Forecasts of Volatility Indexes
7. Evaluating Forecasts of Volatility Indexes
8. Dynamic Volatility Attributions
9. Examples
10. Discussion
  - A.1. DFVCM Algorithm
  - A.2. LMDFM Algorithm
  - A.3. YWpcAR Algorithm
- Reference

## **1. Introduction**

Widely watched, reported and utilized volatility indexes include VIX and others published by CBOE (<https://www.cboe.com/>), such as VIX (on S&P 500), VXD (on DJIA), VXN (on Nasdaq 100), RVX (on Russel 2000), VXEEM (on emerging market), VXEWZ (on Brazil ETF), GVZ (on gold), OVX (on crude oil), EVZ (on Euro), VXAZN (on Amazon),

VXAPL (on Apple), VXGS (on Goldman Sachs), VXGOG (on Google), VXIBM (on IBM), etc.

Volatility indexes published by global financial exchanges are regarded as among crucial indicators by many economy and market participants all over the world. Well-established holistic data-driven models capable to analyze and forecast volatility indexes could serve as important tools.

Many volatility indexes, all price time-series underlying these volatility indexes, and many other relevant financial time-series are “dynamically correlated”: correlated both synchronously cross-section and asynchronously over time. A set of “dominant dynamic correlation characteristics” of these large number of time-series can be extracted by widely utilized dynamic factor models (DFMs). Those “other relevant time-series” are information-enhancing inputs of the models. Dynamic relationships represented by dominant dynamic factors are less likely contaminated by random noises and, as a result, more likely to produce more robust forecasts.

Our modeling approach is to carry out DFM volatility analysis on DFM-input time-series, which are multiple time-series including both price time-series underlying the volatility indexes and other information-enhancing time-series. Then, the volatility forecasts of the underlying time-series are transformed into multi-step forecasts of the multiple volatility indexes.

Our Dynamic Factor Based Volatility Index Forecast (DFbVIF) model will be discussed in details step by step as follows: four separate but related sets of multiple (or vector) time-series; DFM-based volatility modeling on DFM-input time-series; forecasted volatilities of DFM-input time-series; formulas and expressions of variances of underlying time-series; DFM-based dynamic forecasts of volatility indexes; dynamic volatility attributions; evaluation of forecasts of volatility indexes; numeric examples; and concluding discussions.

## **2. Data: Four Sets of Multiple Time-Series**

Our DFM-based multi-step forecasts of multiple volatility indexes involve following four related sets of multiple time-series:

1. volatility index time-series,
2. underlying time-series,
3. info-enhancing time-series and
4. DFM-input time-series.

Underlying (CBOE's and other) volatility index time-series, VIX, VXD, etc. are market price/index time-series, S&P 500, DJIA, etc. An underlying time-series, underlying a given volatility index, is a time-series of percentage or logarithmic changes in price/value/index, with the value of volatility index representing expected volatility level of the underlying time-series, expected at a specified future time. This fact about volatility expectation indicates that forecasting volatility indexes can be made by volatility analysis on, and volatility forecast of, their underlying time-series.

Dynamic factor models (DFMs) are the most researched and useful frameworks able to model multiple dynamically correlated time-series. There is an enormous volume of widely accepted academic publications and technical whitepapers on DFMs. Doz and Fuleky (2020) is among the best overviews on DFM researches and developments. In addition to analyzing and forecasting values of multiple time-series, DFMs can also effectively analyze and forecast variance-covariance of the same set of time-series (i4cast LLC 2024 on DFVCM algorithm; Alessi, Barigozzi and Capasso 2007). This fact about multiple time-series volatility analysis indicates that forecasting volatility indexes can be made by DFM volatility modeling on the

underlying time-series.

DFMs built on dynamic PCA (principal component analysis) are able to accommodate large number of time-series with greater order of vector autoregression (VAR). Some publicly available machine learning algorithms are among such DFMs of big data (e.g. LMDFM, YWpcAR and DFVCM algorithms available on AWS and offered by i4cast LLC). This crucial advantage of big data analysis indicates that we can combine the underlying time-series with many other information-enhancing time-series dynamically correlated with the underlying ones to serve as an expanded input data set of the DFM volatility analysis.

Combining underlying and info-enhancing time-series into a single set of DFM-input time-series will provide the DFM volatility analysis/forecast with more relevant information than the underlying time-series alone. The higher degree of information completeness of the DFM-input time-series indicates that they are able to increase predictive power of the DFM-based forecasts of the volatility indexes.

Forecasted variance-covariance matrix of underlying time-series is a sub-matrix of that of DFM-input time-series. Forecasts of variances of underlying time-series can be transformed into forecasts of implied variances which are squares of volatility indexes. Two different transformations are developed: one is “underlying volatility forecasts (UVF) as predictors”, and the other “quadratic autoregressive (QAR) forecasts” with QAR coefficients of underlying volatilities. Their formulars will be presented in the forms of equations of matrixes.

We denote multiple volatility index time-series as time-series  $vi_t$  of  $h \times 1$  vector  $vi$ , representing  $h$  individual volatility indexes. When multiple (forecasted) volatility indexes need to be expressed in equations of matrixes, we also denote the same set of volatility indexes as time-series  $VI_t$  of  $h \times h$  diagonal matrix  $VI$ . A set of  $h$  “implied variances” (squares of

volatility indexes) is denoted as time-series  $VI_t^2$  of  $h \times h$  diagonal matrix  $VI^2$ ,

$$VI_t^2 = VI_t VI_t^T = VI_t^T VI_t = VI_t VI_t . \quad (2.1)$$

A set of  $h$  underlying time-series is denoted as time-series  $y_t^{(und)}$  of  $h \times 1$  vector  $y^{(und)}$ .

Similarly, a set of  $n - h$  additional info-enhancing time-series is denoted as time-series  $y_t^{(info)}$  of  $(n - h) \times 1$  vector  $y^{(info)}$ . Therefore, a set of  $n$  DFM-input time-series is denoted as time-series  $y_t$  of  $n \times 1$  vector  $y$ ,

$$y_t = [ (y_t^{(und)})^T, (y_t^{(info)})^T ]^T . \quad (2.2)$$

Now, the DFM-input vector time-series  $y_t$  is ready for DFM volatility modeling.

### **3. Volatility Model on DFM-Input Time-Series**

To make multi-step forecasts of multivariate volatilities of observed DFM-input vector time-series  $y_t$ , we first estimate all coefficients and attributes of a dynamic factor model representation of  $y_t$  expressed as follows:

$$y_t = \mu_t + X_{t,0} f_t + u_t , \quad (3.1)$$

$$f_t = A_{t,1} f_{t-1} + A_{t,2} f_{t-2} + \cdots + A_{t,p} f_{t-p} + v_t , \quad (3.2)$$

$$u_t = g_t + r_t , \quad (3.3)$$

$$g_t = D_{t,1} g_{t-1} + D_{t,2} g_{t-2} + \cdots + D_{t,q} g_{t-q} + e_t , \quad (3.4)$$

where (random and non-random) variables

- $y_t$  is time-series of  $n \times 1$  vector of observed data,
- $\mu_t$  is time-series of  $n \times 1$  vector of mean values of  $y_t$ ,
- $f_t$  is time-series of  $m \times 1$  vector of common dynamic factor scores,
- $X_{t,0}$  is  $n \times m$  matrix of factor loadings of dynamic factors  $f_t$ ,

- $A_{t,j}$  are  $m \times m$  matrixes of VAR (vector autoregressive) coefficients of factors  $f_t$ , here  $n > or \gg m$ ,
- $v_t$  is time-series of  $m \times 1$  vector of VAR prediction errors of factors  $f_t$ ,
- $u_t$  is time-series of  $n \times 1$  vector of idiosyncratic components of individual observed time-series,
- $g_t$  is time-series of  $n \times 1$  vector of unobserved dynamic components (UDCs) of idiosyncratic components  $u_t$ ,
- $D_{t,k}$  are  $n \times n$  diagonal matrixes of AR (autoregressive) coefficients of UDCs  $g_t$ ,
- $e_t$  is time-series of  $n \times 1$  vector of AR prediction errors of UDCs  $g_t$ , and
- $r_t$  is time-series of  $n \times 1$  vector of residual random errors.

For  $j, k = 0, 1, \dots$ , variances and autocovariances of common dynamic factors and idiosyncratic UDCs are estimated as:

$$V_{t-j} = \text{Var}(f_{t-j}) = E(f_{t-j} f_{t-j}^T), \quad (3.5)$$

$$V_{(t-j),(t-k)} = \text{Cov}(f_{t-j}, f_{t-k}) = E(f_{t-j} f_{t-k}^T), \quad (3.6)$$

$$W_{t-j} = \text{Var}(g_{t-j}) = E(g_{t-j} g_{t-j}^T), \quad (3.7)$$

$$W_{(t-j),(t-k)} = \text{Cov}(g_{t-j}, g_{t-k}) = E(g_{t-j} g_{t-k}^T), \quad (3.8)$$

where estimated matrixes

- $V_{t-j}$  are  $m \times m$  diagonal matrixes of variance of dynamic factors  $f_t$ ,
- $V_{(t-j),(t-k)}$  are  $m \times m$  matrixes of autocovariance of factors  $f_t$ ,
- $W_{t-j}$  are  $n \times n$  diagonal matrixes of variance of UDCs  $g_t$ , and
- $W_{(t-j),(t-k)}$  are  $n \times n$  diagonal matrixes of autocovariance of UDCs  $g_t$ .

Variances of random errors are estimated as:

$$R_t^{(v)} = \text{Var}(v_t), \quad (3.9)$$

$$R_t^{(e)} = \text{Var}(e_t), \quad (3.10)$$

$$R_t^{(r)} = \text{Var}(r_t); \quad (3.11)$$

where estimated matrixes

- $R_t^{(v)}$  is time-series of  $m \times m$  diagonal matrix of variances of VAR prediction errors  $v_t$ ,
- $R_t^{(e)}$  is time-series of  $n \times n$  diagonal matrix of variances of AR prediction errors  $e_t$ , and
- $R_t^{(r)}$  is time-series of  $n \times n$  diagonal matrix of variances of residual errors  $r_t$ .

The DFM estimates summarized with Eqs. (3.1) through (3.11) above can be performed using DFM-based volatility model formulated and detailed in Ling (2025 B) which offers advantages over other volatility models, including capability to make multi-step forecasts of multivariate volatilities of large number of time-series. Volatility components forecasted by DFM-based model include those separately attributable to (1) common or (2) idiosyncratic (a) variances or (b) autocovariances or (c) prediction errors. Common volatility components are forecasted using DFM by spectral PCA and conjugate 2D DFT. Idiosyncratic volatility components are forecasted using AR model by Yule-Walker-PCA estimates. “Dynamic Factor Variance-Covariance Model” algorithm (DFVCM,

<https://aws.amazon.com/marketplace/pp/prodview-yvaulquatt3v2?sr=0>) available on AWS is an implementation of this DFM-based volatility forecasts.

The estimation of the coefficients and attributes in Eqs (3.1) through (3.11) applied following assumptions widely proposed, accepted and practiced in dynamic factor model

research literature:

- mean value (of observed vector time-series)  $\mu_t = 0$ ,
- covariance matrix (of dynamic factor scores)  $V_{t-j}$  is diagonal matrix,
- covariance matrix (of idiosyncratic components)  $W_{t-j}$  is approximated by diagonal matrix, ignoring allowed but “mild” cross-correlation,
- assuming all cross- and serial-correlations between common factors  $f_t$  and idiosyncratic UDCs  $g_t$  are 0's.

As a result, for time-lag  $j = 0, 1, \dots$ , all  $n \times n$  variance-covariance matrixes of DFM-input time-series  $y_t$  are estimated as

$$\begin{aligned}
C_{t-j} &= \text{Var}(y_{t-j}) \\
&= \text{Var}(X_{t,0} f_{t-j} + g_{t-j} + r_t) \\
&= X_{t,0} \text{Var}(f_{t-j}) X_{t,0}^T + \text{Var}(g_{t-j}) + \text{Var}(r_t) \\
&= X_{t,0} V_{t-j} X_{t,0}^T + W_{t-j} + R_t^{(r)}. \tag{3.12}
\end{aligned}$$

#### **4. Forecasts by DFM Volatility Model**

According to the factor model representation discussed in the Section 3 above,  $s$ -step forecasts of time-series,  $f_t$ ,  $g_t$  and  $y_t$ , based on data observed until time  $t$  can be made by dynamic equations as

$$f_{(t+s)|t} = A_{t,1} f_{t+s-1} + A_{t,2} f_{t+s-2} + \dots + A_{t,p} f_{t+s-p} + v_{t+s}, \tag{4.1}$$

$$g_{(t+s)|t} = D_{t,1} g_{t+s-1} + D_{t,2} g_{t+s-2} + \dots + D_{t,q} g_{t+s-q} + e_{t+s}, \tag{4.2}$$

$$y_{(t+s)|t} = X_{t,0} f_{(t+s)|t} + g_{(t+s)|t} + r_{t+s}, \tag{4.3}$$

where  $s = 1, 2, \dots$ . The random errors  $v_{t+s}$ ,  $e_{t+s}$  and  $r_{t+s}$  cannot be forecasted, but can be characterized by assumed diagonal variance matrixes as

$$R_{t+s}^{(v)} = R_t^{(v)}, \quad (4.4)$$

$$R_{t+s}^{(e)} = R_t^{(e)}, \quad (4.5)$$

$$R_{t+s}^{(r)} = R_t^{(r)}. \quad (4.6)$$

Therefore,  $s$ -step forecast of variance of common dynamic factors  $f_t$  is

$$\begin{aligned} V_{(t+s)|t} &= \text{Var}(f_{(t+s)|t}) = E(f_{(t+s)|t} f_{(t+s)|t}^T) \\ &= \text{Diag}(E((\sum_{j=1}^p A_{t,j} f_{t+s-j})(\sum_{k=1}^p A_{t,k} f_{t+s-k})^T)) + R_t^{(v)} \\ &= \sum_{j=1}^p \sum_{k=1}^p \text{Diag}(A_{t,j} E(f_{t+s-j} f_{t+s-k}^T) A_{t,k}^T) + R_t^{(v)} \\ &= \sum_{j=1}^p \sum_{k=1}^p \text{Diag}(A_{t,j} V_{(t+s-j),(t+s-k)} A_{t,k}^T) + R_t^{(v)}, \end{aligned} \quad (4.7)$$

where  $s = 1, 2, \dots$ , and variance and autocovariance of factors,  $E(f_{t+s-j} f_{t+s-k}^T)$ , are evaluated by estimates Eqs. (3.5) and (3.6), or by “earlier-step” forecasts Eqs. (4.7) and (4.8).

Then,  $s$ -Step forecast of  $k$ -lag autocovariance matrix of dynamic factor scores  $f_t$  is

$$\begin{aligned} V_{(t+s),(t+s-k)} &= \text{Cov}(f_{(t+s)|t}, f_{t+s-k}) \\ &= E(f_{(t+s)|t} f_{t+s-k}^T) = E(\sum_{j=1}^p A_{t,j} f_{t+s-j} f_{t+s-k}^T) \\ &= \sum_{j=1}^p A_{t,j} E(f_{t+s-j} f_{t+s-k}^T) = \sum_{j=1}^p A_{t,j} V_{(t+s-j),(t+s-k)}, \end{aligned} \quad (4.8)$$

where  $s = 1, 2, \dots$ ,  $k = 1, 2, \dots$ , and, again, variance and autocovariance,  $E(f_{t+s-j} f_{t+s-k}^T)$ , are evaluated by estimates Eqs. (3.5) and (3.6), or by earlier forecasts Eqs. (4.7) and (4.8).

Similarly,  $s$ -Step forecast of variance of idiosyncratic unobserved dynamic components (UDCs)  $g_t$  is

$$\begin{aligned} W_{(t+s)|t} &= \text{Var}(g_{(t+s)|t}) = E(g_{(t+s)|t} g_{(t+s)|t}^T) \\ &= \text{Diag}(E((\sum_{j=1}^q D_{t,j} g_{t+s-j})(\sum_{k=1}^q D_{t,k} g_{t+s-k})^T)) + R_t^{(e)} \\ &= \sum_{j=1}^q \sum_{k=1}^q \text{Diag}(D_{t,j} E(g_{t+s-j} g_{t+s-k}^T) D_{t,k}) + R_t^{(e)} \\ &= \sum_{j=1}^q \sum_{k=1}^q D_{t,j} W_{(t+s-j),(t+s-k)} D_{t,k} + R_t^{(e)}, \end{aligned} \quad (4.9)$$

where  $s = 1, 2, \dots$ , and variance and autocovariance of UDCs,  $E(g_{t+s-j} g_{t+s-k}^T)$ , are

evaluated by estimates Eqs. (3.7) and (3.8), or by “earlier-step” forecasts Eqs. (4.9) and (4.10).

Then,  $s$ -step forecast of  $k$ -lag autocovariance of UDCs  $g_t$  is

$$\begin{aligned}
 W_{(t+s),(t+s-k)} &= \text{Cov}(g_{(t+s)|t}, g_{t+s-k}) \\
 &= E(g_{(t+s)|t} g_{t+s-k}^T) = \text{Diag}(E(\sum_{j=1}^q D_{t,j} g_{t+s-j} g_{t+s-k}^T)) \\
 &= \sum_{j=1}^q \text{Diag}(D_{t,j} E(g_{t+s-j} g_{t+s-k}^T)) \\
 &= \sum_{j=1}^q D_{t,j} W_{(t+s-j),(t+s-k)}, \tag{4.10}
 \end{aligned}$$

where  $s = 1, 2, \dots$ ,  $k = 1, 2, \dots$ , and, again, variance and autocovariance,  $E(g_{t+s-j} g_{t+s-k}^T)$ ,

are evaluated by estimates Eqs. (3.7) and (3.8), or by earlier forecasts Eqs. (4.9) and (4.10).

Having

- estimated factor loadings matrix  $X_{t,0}$  in Eq. (3.1),
- forecasted diagonal variance matrix  $V_{(t+s)|t}$  of dynamic factors  $f_t$  by Eq. (4.7),
- forecasted diagonal variance matrix  $W_{(t+s)|t}$  of UDCs  $g_t$  by Eq. (4.9), and
- “forecasted” diagonal variance matrix of residual errors  $r_t$  by Eq. (4.6),

$s$ -step forecast of variance-covariance matrix of DFM-input vector time-series  $y_t$  is

$$\begin{aligned}
 C_{(t+s)|t} &= \text{Var}(y_{(t+s)|t}) = E(y_{(t+s)|t} y_{(t+s)|t}^T) \\
 &= X_{t,0} V_{(t+s)|t} X_{t,0}^T + W_{(t+s)|t} + R_t^{(r)} \\
 &= X_{t,0} \sum_{j=1}^p \sum_{k=1}^q \text{Diag}(A_{t,j} V_{(t+s-j),(t+s-k)} A_{t,k}^T) X_{t,0}^T \\
 &\quad + \sum_{j=1}^q \sum_{k=1}^q D_{t,j} W_{(t+s-j),(t+s-k)} D_{t,k} \\
 &\quad + X_{t,0} R_t^{(v)} X_{t,0}^T + R_t^{(e)} + R_t^{(r)}, \tag{4.11}
 \end{aligned}$$

where  $s = 1, 2, \dots$ .

## 5. Expressions of Underlying Volatilities

Our objective is to make forecasts of volatility indexes based on volatility expressions of underlying time-series which, in turn, can be obtained by volatility expressions of DFM-input

time-series, presented in Sections 3 and 4 above.

For time-lag  $j = 0, 1, \dots$ , and forecast step  $s = 1, 2, \dots$ , basic volatility attributes

specifically for underlying time-series  $y_t^{(und)}$  include:

- common factor loadings, an  $h \times m$  matrix  $X_{t,0}^{(und)}$ , is upper  $h \times m$  block of  $n \times m$  matrix  $X_{t,0}$ ,
- autoregressive (AR) coefficients of UDCs,  $h \times h$  diagonal matrixes  $D_{t,j}^{(und)}$ , are upper-left  $h \times h$  diagonal blocks of  $n \times n$  diagonal matrixes  $D_{t,j}$ ,
- estimated/forecasted variance of UDCs,  $h \times h$  diagonal matrixes  $W_{t+s-j}^{(und)}$ , are upper-left  $h \times h$  diagonal blocks of  $n \times n$  diagonal matrixes  $W_{t+s-j}$ ,
- estimated variances of UDC prediction errors, an  $h \times h$  diagonal matrix  $R_t^{(e|und)}$ , is upper-left  $h \times h$  diagonal block of  $n \times n$  diagonal matrix  $R_t^{(e)}$ , and
- estimated variances of residual errors, an  $h \times h$  diagonal matrix  $R_t^{(r|und)}$ , is upper-left  $h \times h$  diagonal block of  $n \times n$  diagonal matrix  $R_t^{(r)}$ .

Therefore, estimated variance of underlying time-series  $y_t^{(und)}$ , an  $h \times h$  diagonal matrix  $\sigma_{t-j}^2$ , is a “sub-matrix” by Eq. (3.12) as

$$\begin{aligned}\sigma_{t-j}^2 &= \text{Diag}(C_{t-j}^{(und)}) \\ &= \text{Diag}(X_{t,0}^{(und)} V_{t-j} (X_{t,0}^{(und)})^T) + W_{t-j}^{(und)} + R_t^{(r|und)},\end{aligned}\quad (5.1)$$

where time-lag  $j = 0, 1, \dots$ . Estimated volatility of  $y_t^{(und)}$  is  $\sigma_{t-j} = (\sigma_{t-j}^2)^{1/2}$ . Forecasted variance of underlying time-series  $y_t^{(und)}$ , an  $h \times h$  diagonal matrix  $\sigma_{(t+s)|t}^2$ , is a “sub-matrix” by Eq. (4.11) as

$$\begin{aligned}
\sigma_{(t+s)|t}^2 &= \text{Diag}(C_{(t+s)|t}^{(und)}) \\
&= \text{Diag}(X_{t,0}^{(und)} \sum_{j=1}^p \sum_{k=1}^p \text{Diag}(A_{t,j} V_{(t+s-j),(t+s-k)} A_{t,k}^T) (X_{t,0}^{(und)})^T) \\
&\quad + \sum_{j=1}^q \sum_{k=1}^q D_{t,j}^{(und)} W_{(t+s-j),(t+s-k)}^{(und)} D_{t,k}^{(und)} \\
&\quad + \text{Diag}(X_{t,0}^{(und)} R_t^{(v)} (X_{t,0}^{(und)})^T) + R_t^{(e|und)} + R_t^{(r|und)}, \tag{5.2}
\end{aligned}$$

where forecast step  $s = 1, 2, \dots$ . Forecasted volatility of  $y_t^{(und)}$  is  $\sigma_{(t+s)|t} = (\sigma_{(t+s)|t}^2)^{1/2}$ .

All predictors of first-step forecasts are carriers of information from current, i.e. the most recent, observations. Substituting  $s = 1$  into Eq. (5.2) gives the first-step forecast of variance of underlying time-series  $y_t^{(und)}$  as

$$\begin{aligned}
\sigma_{(t+1)|t}^2 &= \text{Diag}(C_{(t+1)|t}^{(und)}) \\
&= \sum_{j=1}^p \sum_{k=1}^p \text{Diag}(X_{t,0}^{(und)} \text{Diag}(A_{t,j} V_{(t+1-j),(t+1-k)} A_{t,k}^T) (X_{t,0}^{(und)})^T) \\
&\quad + \sum_{j=1}^q \sum_{k=1}^q D_{t,j}^{(und)} W_{(t+1-j),(t+1-k)}^{(und)} D_{t,k}^{(und)} \\
&\quad + \text{Diag}(X_{t,0}^{(und)} R_t^{(v)} (X_{t,0}^{(und)})^T) + R_t^{(e|und)} + R_t^{(r|und)}. \tag{5.3}
\end{aligned}$$

Attributes about underlying time-series to be discussed based on Eq. (5.3) include

- basic parts of variance,
- quadratic autoregression (QAR), and
- dynamic attribution of variance.

Individual additive terms in Eq. (5.3) can be regarded as basic parts of variances of  $y_t^{(und)}$ . Common parts of variance of underlying time-series are based on variances and autocovariances of common dynamic factors,  $f_t$ , as

$$\begin{aligned}
&(\sigma_{t,j,k}^{(V)})^2 \\
&= \text{Diag}(X_{t,0}^{(und)} \text{Diag}(A_{t,j} V_{(t+1-j),(t+1-k)} A_{t,k}^T) (X_{t,0}^{(und)})^T), \tag{5.4}
\end{aligned}$$

where  $j, k = 1, 2, \dots, p$ . Idiosyncratic parts are based on variances and autocovariances of

idiosyncratic UDCs,  $g_t$ , as

$$\begin{aligned} & (\sigma_{t,j,k}^{(W)})^2 \\ &= D_{t,j}^{(und)} W_{(t+1-j),(t+1-k)}^{(und)} D_{t,k}^{(und)}, \end{aligned} \quad (5.5)$$

where  $j, k = 1, 2, \dots, q$ . And random error part incorporating variances of all prediction/residual errors is

$$\begin{aligned} & (\sigma_t^{(R)})^2 \\ &= \text{Diag}(X_{t,0}^{(und)} R_t^{(v)} (X_{t,0}^{(und)})^T) + R_t^{(e|und)} + R_t^{(r|und)}. \end{aligned} \quad (5.6)$$

Then the first-step forecasts of variances of underlying time-series are sum totals of all parts of variances as

$$\begin{aligned} & \sigma_{(t+1)|t}^2 \\ &= \sum_{j=1}^p \sum_{k=1}^p (\sigma_{t,j,k}^{(V)})^2 + \sum_{j=1}^q \sum_{k=1}^q (\sigma_{t,j,k}^{(W)})^2 + (\sigma_t^{(R)})^2. \end{aligned} \quad (5.7)$$

A common part,  $(\sigma_{t,j,k}^{(V)})^2$ , by Eq. (5.4) is actually factor-based portion of covariance of pair of time-lags  $(j, k)$ , while an idiosyncratic part,  $(\sigma_{t,j,k}^{(W)})^2$ , by Eq. (5.5) is UDC-based portion of covariance of  $(j, k)$ . Similar to definition of correlation coefficients, define “factor-based QAR (quadratic autoregressive) coefficients” as

$$a_{t,j,k}^{(V)} = \sigma_{t+1-j}^{-1} (\sigma_{t,j,k}^{(V)})^2 \sigma_{t+1-k}^{-1}, \quad (5.8)$$

where  $j, k = 1, 2, \dots, p$ , and  $\sigma_{t+1-j}^{-1}$  or  $\sigma_{t+1-k}^{-1}$  is estimated by Eq. (5.1); and define “UDC-based QAR coefficients” as

$$a_{t,j,k}^{(W)} = \sigma_{t+1-j}^{-1} (\sigma_{t,j,k}^{(W)})^2 \sigma_{t+1-k}^{-1}, \quad (5.9)$$

where  $j, k = 1, 2, \dots, q$ . With coefficients by Eqs. (5.8) and (5.9), the first-step forecasts of variances of underlying time-series  $y_t^{(und)}$  can be expressed by a quadratic autoregressive (QAR) equation as

$$\begin{aligned}\sigma_{(t+1)|t}^2 &= \sum_{j=1}^p \sum_{k=1}^p \sigma_{t+1-j} a_{t,j,k}^{(V)} \sigma_{t+1-k} \\ &\quad + \sum_{j=1}^q \sum_{k=1}^q \sigma_{t+1-j} a_{t,j,k}^{(W)} \sigma_{t+1-k} + (\sigma_t^{(R)})^2.\end{aligned}\quad (5.10)$$

This is why  $a_{t,j,k}^{(V)}$  and  $a_{t,j,k}^{(W)}$  by Eqs. (5.8) and (5.9) are referred to as QAR coefficients. The QAR equation in Eq. (5.10) will be adapted to make QAR forecasts of implied variances (squares of volatility indexes).

By re-grouping variance parts and autocovariance parts separately, underlying variance formula Eq. (5.7) can be re-written to represent different sources of variance:

$$\begin{aligned}\sigma_{(t+1)|t}^2 &= \sum_{j=1}^p \sum_{k=1}^p \delta_{ij} (\sigma_{t,j,k}^{(V)})^2 + \sum_{j=1}^p \sum_{k=1}^p (1 - \delta_{ij}) (\sigma_{t,j,k}^{(V)})^2 \\ &\quad + \sum_{j=1}^q \sum_{k=1}^q \delta_{ij} (\sigma_{t,j,k}^{(W)})^2 + \sum_{j=1}^q \sum_{k=1}^q (1 - \delta_{ij}) (\sigma_{t,j,k}^{(W)})^2 \\ &\quad + (\sigma_t^{(R)})^2,\end{aligned}\quad (5.11)$$

where  $\delta_{ij}$  is Kronecker delta ( $\delta_{ij} = 1$  when  $j = i$  and  $\delta_{ij} = 0$  when  $j \neq i$ ). Major dynamic sources of variances of underlying time-series  $y_t^{(und)}$  shown in Eq. (5.11) are:

- “common volatility”,  $\sum_{j=1}^p \sum_{k=1}^p \delta_{ij} (\sigma_{t,j,k}^{(V)})^2$ , contributions from variances of common dynamic factors,
- “common serial-correlation”,  $\sum_{j=1}^p \sum_{k=1}^p (1 - \delta_{ij}) (\sigma_{t,j,k}^{(V)})^2$ , contributions from vector autocovariances of common factors,
- “idiosyncratic volatility”,  $\sum_{j=1}^q \sum_{k=1}^q \delta_{ij} (\sigma_{t,j,k}^{(W)})^2$ , contributions from variances of idiosyncratic UDCs,
- “idiosyncratic serial-correlation”,  $\sum_{j=1}^q \sum_{k=1}^q (1 - \delta_{ij}) (\sigma_{t,j,k}^{(W)})^2$ , contributions from autocovariances of idiosyncratic UDCs, and
- “prediction error”,  $(\sigma_t^{(R)})^2$ , contributions from dynamic factor prediction errors,

UDC prediction errors, and residual errors.

A complete set of (percentage) portions of these dynamic sources (i.e. all portions adding to 1.0 or 100%) of underlying variance represent dynamic attributions of variances of underlying time-series  $y_t^{(und)}$ .

## **6. Dynamic Forecasts of Volatility Indexes**

With all preparations discussed in Sections 2 through 5, we are now ready to make multi-step forecasts of multiple volatility indexes,  $VI_t$ . As outlined in Section 2, forecasts of variances of underlying time-series can be transformed into forecasts of implied variances  $VI_t^2$ . Two different transformations are developed:

- UVF forecast, and
- QAR forecast.

The UVF forecast is “volatility index forecasts with underlying volatility forecasts (UVF) as predictors”, or “forecasting changes in volatility indexes with forecasted changes in underlying volatilities”. It is achieved by forecasting implied variances,  $VI_t^2$ , with forecasted underlying variances,  $\sigma_{(t+s)|t}^2$ , as

$$VI_{(t+s)|t}^2 = \beta_t \sigma_{(t+s)|t}^2, \quad (6.1)$$

where forecast step  $s = 1, 2, \dots$ , and  $\beta_t$  is multipliers of value-adjustment for any forecast step. Since nowcast by Eq. (6.1) should yield the known value of implied variance, i.e.  $\beta_t \sigma_{(t+0)|t}^2 = VI_t^2$ , the multipliers is

$$\beta_t = \beta_t^{(uvf)} = VI_{(t+0)|t}^2 \sigma_{(t+0)|t}^{-2} = VI_t^2 \sigma_t^{-2}, \quad (6.2)$$

where  $\beta_t^{(uvf)}$  is an  $h \times h$  diagonal matrix of value-adjustment multiplier for UVF forecasts

and  $\sigma_t^2$  is of estimated underlying variances. Then, the UVF forecasts of implied variances are

$$(VI_{(t+s)|t}^{(uvf)})^2 = \beta_t^{(uvf)} \sigma_{(t+s)|t}^2, \quad (6.3)$$

and the UVF forecasts of volatility indexes are

$$VI_{(t+s)|t}^2 = (VI_{(t+s)|t}^{(uvf)})^2, \quad (6.4)$$

$$VI_{(t+s)|t} = (VI_{(t+s)|t}^2)^{1/2}, \quad (6.5)$$

where forecast step  $s = 1, 2, \dots$ .

The QAR forecast is “volatility index forecasts by quadratic autoregressive (QAR) equation with QAR coefficients of underlying volatilities”, or “making quadratic autoregressive (QAR) forecasts of volatility indexes with QAR equation of underlying volatilities”. It is achieved based on underlying volatility QAR equation in Eq. (5.10) as

$$\begin{aligned} VI_{(t+s)|t}^2 &= \alpha_t + \beta_t (\sum_{j=1}^p \sum_{k=1}^p VI_{t+s-j} a_{t,j,k}^{(V)} VI_{t+s-k} \\ &\quad + \sum_{j=1}^q \sum_{k=1}^q VI_{t+s-j} a_{t,j,k}^{(W)} VI_{t+s-k}), \end{aligned} \quad (6.6)$$

where forecast step  $s = 1, 2, \dots$ , QAR coefficients  $a_{t,j,k}^{(V)}$  and  $a_{t,j,k}^{(W)}$  are estimated by Eqs. (5.8) and (5.9), respectively, and  $\alpha_t$  and  $\beta_t$  are additive constants and multipliers, respectively, of value-adjustment for any forecast step, Since “1-step forecast from 1-step back” by Eq. (6.6) should yield the known value of implied variance, i.e.

$$\begin{aligned} VI_{t|(t-1)}^2 &= \beta_t (\sum_{j=1}^p \sum_{k=1}^p VI_{t-j} a_{t,j,k}^{(V)} VI_{t-k} \\ &\quad + \sum_{j=1}^q \sum_{k=1}^q VI_{t-j} a_{t,j,k}^{(W)} VI_{t-k}) + \alpha_t \\ &= VI_t^2, \end{aligned} \quad (6.7)$$

and the additive constant  $\alpha_t$  is functionally the same as, then should be proportional to, the prediction error term in Eq. (5.10), i.e.

$$\alpha_t VI_t^{-2} = (\sigma_t^{(R)})^2 \sigma_{(t+1)|t}^{-2}, \quad (6.8)$$

the additive constants of value-adjustment are

$$\alpha_t = \alpha_t^{(qar)} = VI_t^2 (\sigma_t^{(R)})^2 \sigma_{(t+1)|t}^{-2}, \quad (6.9)$$

and the multipliers of value-adjustment are

$$\begin{aligned} \beta_t &= \beta_t^{(qar)} \\ &= (VI_t^2 - \alpha_t^{(qar)}) (\sum_{j=1}^p \sum_{k=1}^p VI_{t-j} a_{t,j,k}^{(V)} VI_{t-k} \\ &\quad + \sum_{j=1}^q \sum_{k=1}^q VI_{t-j} a_{t,j,k}^{(W)} VI_{t-k})^{-1}, \end{aligned} \quad (6.10)$$

where  $\alpha_t^{(qar)}$  is an  $h \times h$  diagonal matrix of value-adjustment additive constants for QAR forecasts,  $\beta_t^{(qar)}$  is an  $h \times h$  diagonal matrix of value-adjustment multipliers for QAR forecasts, and  $\sigma_{(t+1)|t}^2$  is the first-step forecasts of underlying variances. Then the QAR forecasts of implied variances are

$$\begin{aligned} (VI_{(t+s)|t}^{(qar)})^2 \\ = \alpha_t^{(qar)} + \beta_t^{(qar)} (\sum_{j=1}^p \sum_{k=1}^p VI_{t+s-j} a_{t,j,k}^{(V)} VI_{t+s-k} \\ + \sum_{j=1}^q \sum_{k=1}^q VI_{t+s-j} a_{t,j,k}^{(W)} VI_{t+s-k}), \end{aligned} \quad (6.11)$$

and the QAR forecasts of volatility indexes are

$$VI_{(t+s)|t}^2 = (VI_{(t+s)|t}^{(qar)})^2, \quad (6.12)$$

$$VI_{(t+s)|t} = (VI_{(t+s)|t}^2)^{1/2}, \quad (6.13)$$

where forecast step  $s = 1, 2, \dots$ . Forecasting by Eqs. (6.11) to (6.13) is an iterative process: forecasted volatility indexes at step  $s$  are inputs of forecasting implied variances at step  $s + 1$ .

## 7. Evaluating Forecasts of Volatility Indexes

Generally, potential future performances of time-series forecasts can be evaluated by comparing most recent out-of-sample forecasts against later actually observed true values being

forecasted.

Specifically, the most recent out-of-sample volatility index forecasts can be denoted as

$$VI_{(\tau(i)+s)|\tau(i)}^{(Forecast)} = VI_{(\tau(i)+s)|\tau(i)}^{(uvf)} \text{ or } VI_{(\tau(i)+s)|\tau(i)}^{(qar)}, \quad (7.1)$$

where  $s$ -step forecasts of volatility indexes made at time  $\tau(i)$ ,  $VI_{(\tau(i)+s)|\tau(i)}^{(Forecast)}$  are either UVF forecasts  $VI_{(\tau(i)+s)|\tau(i)}^{(uvf)}$  or QAR forecasts  $VI_{(\tau(i)+s)|\tau(i)}^{(qar)}$ . A sample of (volatility index and DFM-input time-series) data points utilized to calculate out-of-sample forecasts can be labeled by either the sample ID number  $i \geq 0$  or the last time-stamp  $\tau(i)$  of the sample. Forecast step  $s \geq 1$  and  $s \leq s_{max}$ . As long as the last time-stamp  $\tau(i) = t - s_{max} - i$ , here  $i = 0, 1, \dots$ , the time-stamp of the true values being forecasted  $\tau(i) + s \leq t$ , i.e. these true values have already been observed at time  $t$ :

$$VI_{(\tau(i)+s)|\tau(i)}^{(True)} = VI_{\tau(i)+s}^{(True)} = VI_{\tau(i)+s}. \quad (7.2)$$

The forecasts  $VI_{(\tau(i)+s)|\tau(i)}^{(Forecast)}$  are still literally “out-of-sample” when all data points observed later than  $\tau(i)$ , i.e. all volatility index values  $VI_\tau$  and DFM-input time-series values  $y_\tau$  with  $\tau > \tau(i)$ , are completely kept out of any calculations leading to the forecasts  $VI_{(\tau(i)+s)|\tau(i)}^{(Forecast)}$ .

Then the goodness of fit of the most recent out-of-sample forecasts,  $VI_{(\tau(i)+s)|\tau(i)}^{(Forecast)}$ , can be estimated by some evaluation metrics measuring certain similarities or dissimilarities between the forecasted-true value pairs,  $VI_{(\tau(i)+s)|\tau(i)}^{(Forecast)}$  and  $VI_{\tau(i)+s}^{(True)}$ , with sample ID  $i = 0, 1, \dots$ . The most popular evaluation metrics can be estimated by open-source Python library “scikit-learn” (<https://scikit-learn.org/stable/>, and [https://scikit-learn.org/stable/modules/model\\_evaluation.html#regression-metrics](https://scikit-learn.org/stable/modules/model_evaluation.html#regression-metrics)), which include:

- “r2\_score”,

- “d2\\_absolute\\_error\\_score”,
- “mean\_squared\_error”,
- “mean\_absolute\_error”, and
- “mean\_absolute\_percentage\_error”.

Furthermore, in general, to make evaluation of time-series forecasts genuinely meaningful, we must compare the goodness of fit of the model-based forecasts with an objective benchmark, e.g. “goodness of fit of the most relevant random forecasts”.

Specifically, a set of most relevant random forecasts of volatility index is of a normal distribution:

$$VI_{(\tau(i)+s)|\tau(i)}^{(Random)} \sim N(VI_{\tau(i)}, \sigma_{\tau(i)}^2(s)), \quad (7.3)$$

where  $\sigma_{\tau(i)}^2(s)$ ,  $i \geq 0$ , is variance of  $s$ -step difference of the volatility indexes. When using an evaluation metric based on mean values, as those listed above, to calculate the benchmark of goodness, we can replace a Monte Carlo simulation with Eq. (7.3) by simple “flat forecasts” as

$$VI_{(\tau(i)+s)|\tau(i)}^{(Flat)} = VI_{\tau(i)}. \quad (7.4)$$

Therefore, an objective evaluation score,  $Score(s)$ , of  $s$ -step volatility index forecasts can be estimated as follows:

$$\begin{aligned} & Goodness(s) \\ &= Metric(VI_{(\tau(i)+s)|\tau(i)}^{(Forecast)}, VI_{(\tau(i)+s)|\tau(i)}^{(True)}; i = 0, 1, \dots), \end{aligned} \quad (7.5)$$

$$\begin{aligned} & Benchmark(s) \\ &= Metric(VI_{(\tau(i)+s)|\tau(i)}^{(Flat)}, VI_{(\tau(i)+s)|\tau(i)}^{(True)}; i = 0, 1, \dots), \end{aligned} \quad (7.6)$$

$$\begin{aligned} & Score(s) \\ &= Goodness(s) - Benchmark(s). \end{aligned} \quad (7.7)$$

When denoting:

$$y_{pred} = \{ VI_{(\tau(i)+s)|\tau(i)}^{(Forecast)} ; i = 0,1,\dots \},$$

or  $\{ VI_{(\tau(i)+s)|\tau(i)}^{(Flat)} ; i = 0,1,\dots \},$  (7.8)

$$y_{true} = \{ VI_{(\tau(i)+s)|\tau(i)}^{(True)} ; i = 0,1,\dots \}, \quad (7.9)$$

and using Python scikit-learn library to calculate evaluation scores, the evaluation formulas Eqs. (7.5) and (7.6) can be calculated as:

$$\begin{aligned} & Metric(VI_{(\tau(i)+s)|\tau(i)}^{(Forecast \ or \ Flat)}, VI_{(\tau(i)+s)|\tau(i)}^{(True)}; i = 0,1,\dots) \\ &= Metric(y_{pred}, y_{true}) \\ &= r2\_score(y_{true}, y_{pred}, \dots) \\ &\quad \text{or } d2\_absolute\_error\_score(y_{true}, y_{pred}, \dots) \\ &\quad \text{or } mean\_squared\_error(y_{true}, y_{pred}, \dots) \\ &\quad \text{or } mean\_absolute\_error(y_{true}, y_{pred}, \dots) \\ &\quad \text{or } mean\_absolute\_percentage\_error(y_{true}, y_{pred}, \dots). \end{aligned} \quad (7.10)$$

For a “metric of similarity”, e.g. `r2_score` or `d2_absolute_error_score`, a positive evaluation score,  $Score(s) > 0$ , indicates effective forecasts, and the higher score the better. On the other hand, for a “metric of dissimilarity”, e.g. `mean_squared_error`, `mean_absolute_error`, or `mean_absolute_percentage_error`, a negative evaluation score,  $Score(s) < 0$ , indicates effective forecasts, and the lower score the better.

## 8. Dynamic Volatility Attributions

One of the primary objectives of volatility analysis is to make “volatility attribution”: evaluating contributions from various sources to estimated, nowcasted or forecasted volatility measures. The simplest source of volatility is a pure random noise – assuming all multivariate/individual serial-correlations are zeros. In real data sets, however, serial-correlations are quite often not zeros. In general, multivariate volatility values are sums of contributions from

two sources: (1) variance-covariances and (2) vector and/or individual autocovariances. Positive serial-correlations increase volatility levels, while negative ones decrease them. Non-zero autocovariances also make term-structures of volatilities different from that of pure random noise: “volatility level is proportional to square root of time horizon”.

Dynamic volatility forecasts, with Eqs. (3.7) to (3.11) in whitepaper by i4cast LLC (2025) and with Eqs. (6.6) to (6.11) here in Section 6 of this paper, demonstrate that equations for immediate (as opposed to “forward”) one-step volatility forecasts incorporate all volatility sources expressed in the latest available data values and model estimates. Therefore, any specific formula of dynamic volatility forecasts has a corresponding set of volatility attributions associated with it.

Specifically, associated with the UVF forecasts Eq. (6.3) of volatility indexes are dynamic volatility attributions the same as volatility attributions of the underlying time-series discussed in i4cast LLC (2025). It can be summarized with contributions from 5 sources of volatility:

- “common volatility”, from variances of dynamic common factors;
- “common serial-correlation”, from vector autocovariances of dynamic common factors;
- “idiosyncratic volatility”, from variance of idiosyncratic UDCs;
- “idiosyncratic serial-correlation”, from autocovariance of idiosyncratic UDCs;
- “prediction error”, from (a) dynamic factor prediction errors, (b) UDC prediction errors, and (c) residual errors.

Similarly, associated with the QAR forecasts Eq. (6.11) of volatility indexes are dynamic volatility attributions with the same 5 sources of volatility, but with a different set of

calculations: those based on QAR equation Eq. (6.11), of course. With QAR forecasts Eqs. (6.11) to (6.13), the first 1-step forecast of implied variance is

$$\begin{aligned} & VI_{(t+1)|t}^2 \\ &= \alpha_t^{(qar)} + \beta_t^{(qar)} \left( \sum_{j=1}^p \sum_{k=1}^p VI_{t+1-j} a_{t,j,k}^{(V)} VI_{t+1-k} \right. \\ &\quad \left. + \sum_{j=1}^q \sum_{k=1}^q VI_{t+1-j} a_{t,j,k}^{(W)} VI_{t+1-k} \right). \end{aligned} \quad (8.1)$$

We can define factor-explained parts of implied variance by QAR forecast as

$$(VI_{t,j,k}^{(V)})^2 = \beta_t^{(qar)} VI_{t+1-j} a_{t,j,k}^{(V)} VI_{t+1-k}, \quad (8.2)$$

where  $j, k = 1, 2, \dots, p$ , define UDC-explained parts of implied variance as

$$(VI_{t,j,k}^{(W)})^2 = \beta_t^{(qar)} VI_{t+1-j} a_{t,j,k}^{(W)} VI_{t+1-k}, \quad (8.3)$$

where  $j, k = 1, 2, \dots, q$ . Therefore, the first 1-step QAR forecasts can be expressed in five volatility sources as

$$\begin{aligned} & VI_{(t+1)|t}^2 \\ &= \sum_{j=1}^p \sum_{k=1}^p \delta_{ij} (VI_{t,j,k}^{(V)})^2 + \sum_{j=1}^p \sum_{k=1}^p (1 - \delta_{ij}) (VI_{t,j,k}^{(V)})^2 \\ &\quad + \sum_{j=1}^q \sum_{k=1}^q \delta_{ij} (VI_{t,j,k}^{(W)})^2 + \sum_{j=1}^q \sum_{k=1}^q (1 - \delta_{ij}) (VI_{t,j,k}^{(W)})^2 \\ &\quad + \alpha_t^{(qar)}. \end{aligned} \quad (8.4)$$

Here, each of the five terms represents an aggregate dynamic source of volatility, measured by contribution to forecasted implied variance from the source:

- the first term, can be labeled as “common volatility”, is of variances of dynamic common factors;
- the second term, as “common serial-correlation”, is of vector autocovariances of dynamic common factors;
- the third term, as “idiosyncratic volatility”, is of variance of idiosyncratic UDCs;
- the fourth term, as “idiosyncratic serial-correlation”, is of autocovariance of

- idiosyncratic UDCs;
- the fifth term, as “prediction error”, is of (a) dynamic factor prediction errors, (b) UDC prediction errors, and (c) residual errors.

Volatility analysis based on dynamic factor models jointly estimate both variance-covariances and vector/individual autocovariances of large number of time-series. Traditional static factor model volatility analyses, however, do not estimate any autocovariances at all. Therefore, dynamic volatility attributions provide more and deeper insights than static volatility attributions.

## **9. Examples**

“Dynamic Factor Based Volatility Index Forecast” model (DFbVIF, <https://aws.amazon.com/marketplace/pp/prodview-p6f77e3yfnxbu?sr=0>) publicly available on AWS is an implementation of DFM-based volatility index forecasting method presented in detail in this paper.

To test Dynamic Factor Based Volatility Index Forecast (DFbVIF) model, an example of input data table is tabulated with publicly available real data sets ([https://github.com/i4cast/aws/blob/main/dfm-based\\_volatility\\_index\\_forecast\\_model/input/Weekly\\_MTS\\_6Yr.csv](https://github.com/i4cast/aws/blob/main/dfm-based_volatility_index_forecast_model/input/Weekly_MTS_6Yr.csv)). The table contains multiple weekly time-series for several years ending as of the last Friday of the most recent calendar year. Each row is of an individual time-series. The first 14 rows are of 14 volatility indexes published by CBOE (<https://www.cboe.com/>) listed here, in a format of,

- [symbol of volatility index], on [name of underlying index/price/value] ( [symbol of underlying time-series] ),

as follows:

- VIX, on S&P 500 (SPY);
- VXD, on DJIA (DIA);
- VXN, on Nasdaq 100 (QQQM);
- RVX, on Russel 2000 (VTWO);
- VXEEM, on emerging market (EEM);
- VXEWZ, on Brazil ETF (EWZ);
- GVZ, on gold (GLD);
- OVX, on crude oil (USO);
- EVZ, on Euro FX (FXE);
- VXAZN, on Amazon (AMZN);
- VXAPL, on Apple (AAPL);
- VXGS, on Goldman Sachs (GS);
- VXGOG, on Google (GOOGL);
- VXIBM, on IBM: (IBM).

Remaining 55 rows of the table are of “DFM-input time-series”, including logarithmic changes in 14 underlying time-series (indexes/prices/values underlying the 14 volatility indexes) and in 41 information-enhancing time-series dynamically correlated with the underlying time-series. The info-enhancing ones include logarithmic changes in indexes/prices of equities, fixed incomes, commodities, and FXs.

A dynamic factor model, DFVCM algorithm (Appendix A.1), embedded in DFbVIF model estimated and forecasted the variance-covariances and vector autocovariances of the DFM-input time-series. Then, DFbVIF transformed the forecasted variances of the underlying

time-series into forecasted implied variances of the volatility indexes, by either UVF or QAR transformation. Forecasted volatility index levels are square roots of the forecasted implied variances.

## **10. Discussion**

The numeric example described above demonstrates that the Dynamic Factor Based Volatility Index Forecast (DFbVIF) model can generate statistically good and consistent multi-step forecasts of multiple volatility indexes. The performance of DFbVIF forecasts is consistently better than an objective benchmark: performance of simple flat forecasts, even at multiple forecasting steps, such as 10 steps.

Dynamic attributions of forecasts by DFbVIF can reveal useful information in terms of common vs. idiosyncratic factors, and cross-sectional vs. serial correlations.

Numbers of common factors and idiosyncratic UDCs for better QAR (quadratic autoregressive) forecasts may be different from those for better UVF (underlying volatility forecasts as predictors) forecasts. Or, in another words, QAR and UVF forecasts may need to be tuned separately.

### **A.1. DFVCM Algorithm**

The DFVCM algorithm (<https://aws.amazon.com/marketplace/pp/prodview-yvaulquatt3v2?sr=0>) is an implementation of DFM-based volatility model estimation, formulated and detailed in Ling (2025 B). This algorithm can be summarized as making multi-step forecasts of multivariate volatilities in three separate steps:

1. Step One, estimate and forecast all values related to “common dynamic factors”

(common to many or all DFM-input time-series), e.g. Eqs. (3.1), (3.2), (3.5), (3.6) and (3.9), by using LMDFM (long-memory dynamic factor model) algorithm (see Appendix A.2).

2. Step Two, estimate and forecast all values related to “idiosyncratic UDCs” (specific to individual DFM-input time-series), e.g. Eqs. (3.3), (3.4), (3.7), (3.8), (3.10) and (3.11), by using YWpcAR (Yule-Walker-PCA autoregressive model) algorithm (see Appendix A.3).
3. Step Three, combine estimates and forecasts of common factors and idiosyncratic UDCs together to make multi-step forecasts of multivariate volatilities.

## A.2. LMDFM Algorithm

The LMDFM algorithm (<https://aws.amazon.com/marketplace/pp/prodview-da6ffrp4mlopg?sr=0>) is an implementation of DFM estimation by spectral principal components analysis (spectral PCA) with conjugate two-dimensional discrete Fourier transform (C2D-DFT), formulated and detailed in Ling (2025 A). This algorithm can be summarized as follows:

- Estimating variance-covariance matrixes (VCMs) and autocovariance matrixes (ACMs) of observed vector (i.e. multiple) time-series  $y_t$ , as  $C(j, k) = Cov(y_{t-j}, y_{t-k}) = E(y_{t-j} y_{t-k}^T)$ ,  $j, k = 0, 1, \dots, p$ , assuming  $E(y_t) = 0$ .
- Combining VCMs,  $C(j, j)$  when  $k = j$ , and ACMs,  $C(j, k)$  when  $k \neq j$ , by applying conjugate two-dimensional discrete Fourier transform (C2D-DFT) on  $C(j, k)$ ,  $j, k = 0, 1, \dots, p$ .
- Referring resulted transform by C2D-DFT as spectral density matrixes (SDMs),  $S(m, n) = DFT_{C2D}(\{C(j, k)\})$ ,  $m, n = 0, 1, \dots, p$ , of vector time-series,  $y_t$ .

- Estimating spectral principal components of “on-diagonal SDMs”,  $S(m, m)$ ,  $m = n$ , and SPCA representations of “off-diagonal SDMs”,  $S(m, n)$ ,  $m \neq n$ .
- Estimating full-spectrum principal components (FSPCs) of original VCMs by applying inverse C2D-DFT on SPCA representations of SDMs.
- This way, FSPCs of original VCMs contain dynamic information in all VCMs and ACMs.
- Dynamic factor loadings and vector autoregressive (VAR) coefficients of factor scores can be estimated by full-spectrum principal components of VCMs.

The LMDFM estimates two different simplest forms of DFM. DFM of Form I estimates:

$$y_t = \mu_t + X_{t,0} f_t + X_{t,1} f_{t-1} + \cdots + X_{t,p} f_{t-p} + u_t,$$

$$f_t = g(e_t),$$

where  $e_t$  is a vector of white noises of unit variance,  $g(\cdot)$  is a linear transformation, and elements of vector  $f_t$  are independent of each other and over time. Other application facts about Form I include:

- classic form of DFM for classic dynamic analysis,
- suitable for simple Monte Carlo simulation,
- estimating a whole set of matrixes of factor loadings  $X_{t,k}$ ,
- not applicable to “big data set” due to large number of elements of loadings,
- question on time-series forecast: sizable impact by assuming unknown factor scores  $f_{t+s} = 0$ ?
- for variance-covariance forecast: assuming stable variance of  $f_t$ .

LMDFM’s DFM of Form II estimates Eqs. (3.1) and (3.2) as:

$$y_t = \mu_t + X_{t,0} f_t + u_t,$$

$$f_t = A_{t,1} f_{t-1} + A_{t,2} f_{t-2} + \cdots + A_{t,p} f_{t-p} + v_t,$$

where the only loadings matrix  $X_{t,0}$  in Form II is NOT the same as that in Form I (just share the same symbol). Other application facts about Form II include:

- only one matrix of factor loadings,  $X_{t,0}$ ,
  - estimating a whole set of VAR matrixes,  $A_{t,k}$ , of common factors,
  - non-zero vector autocovariance of comm factors,  $Cov(f_{t-j}, f_{t-j-k}) \neq 0$ ,
- therefore, not for simple Monte Carlo simulation,
- applicable to “big data set” due to limited number of elements of VAR matrixes,
  - question on time-series forecast: smaller impact by assuming unknown errors  $v_{t+s} = 0$  ?
  - for variance-covariance forecast: assuming stable variance of errors  $v_t$ .

DFM of Form II, i.e. Eqs. (3.1) and (3.2), is utilized to estimate common components, including  $X_{t,0}$ ,  $f_t$ ,  $A_{t,j}$ ,  $R_t^{(v)}$ ,  $V_{t-j}$ ,  $V_{(t-j),(t-k)}$  and  $u_t$ , shown in Section “3. DFM Volatility Model on DFM-Input TS”.

### A.3. YWpcAR Algorithm

The YWpcAR algorithm (<https://aws.amazon.com/marketplace/pp/prodview-prndys7tr7go6?sr=0>) is an implementation of estimating non-stationary autoregressive (AR) model by Yule-Walker dimension reduction method, formulated and detailed in Ling (2026, to be published in 2026). This algorithm can be summarized as follows:

- A set of AR coefficients of AR model on observed univariate time-series is a solution vector of linear Yule-Walker (YW) equation of the AR time-series.
- Design matrix and response vector of the YW equation are expressed in terms of

variances and autocovariances of the time-series at a range of time-lags.

- The YW solution vector can be expressed with all eigenvalues and all eigenvectors of symmetric (sample) variance-autocovariance matrixes (VACMs) associated with the YW equation.
- Replacing the set of all eigenpairs with selected principal eigenvalues and principal eigenvectors of the VACMs, the YW solution vector becomes dimension-reduced set of AR coefficients.
- Combe principal eigen score time-series into an unobserved dynamic component (UDC) time-series of the observed time-series.
- The UDC time-series is of an AR process governed by the dimension-reduced AR coefficients.
- Forecast values of observed time-series by dimension-reduced AR forecasts of the UDC time-series.
- Forecast variances of observed time-series by dimension-reduced AR forecasts of variances of the UDC as well as estimated residual variance.

The YWpcAR can be utilized to estimate idiosyncratic components, including  $g_t$ ,  $D_{t,k}$ ,  $R_t^{(e)}$ ,  $W_{t-j}$ ,  $W_{(t-j),(t-k)}$  and  $R_t^{(r)}$ , shown in Section “3. DFM Volatility Model on DFM-Input TS”.

## References

- L. Alessi, M. Barigozzi and M. Capasso (2007). “Dynamic factor GARCH: Multivariate volatility forecast for a large number of series”. LEM Working Paper Series, No. 2006/25, Laboratory of Economics and Management (LEM), Pisa.
- C. Doz and P. Fuleky (2020). “Chapter 2, Dynamic Factor Models”, in

Macroeconomic Forecasting in the Era of Big Data: Theory and Practice, Ed. P. Fuleky.

Advanced Studies in Theoretical and Applied Econometrics, Volume 52. Springer.

H.F. Ling (2025 A). “Introduction to Non-Stationary Dynamic Factor Models Estimated by Spectral PCA with Conjugate 2D-DFT”. To be available on Social Science Research Network, (December 2025), <https://ssrn.com/>.

H.F. Ling (2025 B). “Introduction to Multi-step Forecast of Multivariate Volatility with Dynamic Factor Model”. To be available on Social Science Research Network, (December 2025), <https://ssrn.com/>.

[ EOF ]