Upsampled Vector Autoregressive Processes

Emilio Ruiz-Moreno[1], Luis M. López-Ramos[2], and Baltasar Beferull-Lozano[1]

[1]SIGIPRO Department, Simula Metropolitan Center for Digital Engineering, Oslo, Norway
[2]HOST Department, Simula Metropolitan Center for Digital Engineering, Oslo, Norway

***Abstract*—Vector autoregressive (VAR) processes are simple yet remarkably versatile discrete statistical models used to characterize the dynamics of a collection of variables. Although implicitly, any given VAR process assumes the highest rate, or resolution, at which those variables may vary. Hence, dynamic variations at finer rates are out of the explainability of such models by design. This paper proposes a new method to overcome this drawback. Specifically, it describes how to increase the resolution associated with any VAR process such that the process mean is unchanged and the forecast estimates at the rates given by the original resolution, and the long-run associated credible intervals, are closely preserved. Our experiments confirm the viability of the proposed method for multivariate time series analysis.**

***Index Terms*—Vector autoregressive processes, resolution refinement, multivariate time series.**

I. Introduction

A vector autoregressive (VAR) process is a discrete statistical model that describes the joint behavior of multiple interdependent variables evolving over an ordered indexed set. Each variable is expressed as a linear combination of its previous values, the previous values of other variables, and an error term. The indexed set typically corresponds to time instances, but can also consist of other ordered structures.

Albeit simple, VAR processes have proven particularly useful in multivariate time series forecasting [1]. Conversely, VAR processes suffer from two main limitations in this regard. First, they cannot model any relation beyond linearity by design, and second, they are subject to a fixed temporal resolution, e.g., when the indexed set comes from a uniform sampling of a continuous time domain. This paper focuses on loosening the latter limitation by proposing a method to increase the temporal resolution of any given VAR model.

In the context of this work, the temporal resolution specifies the uniform rate at which the variables (following a trajectory parametrized by time) are discretized when being recorded, processed, or analyzed. Therefore, it is important to note that the (selected or given) temporal resolution affects the predictive performance of VAR processes. On the one hand, overly fine temporal resolution can lead to increased computational costs. On the other hand, overly coarse temporal resolution may cause a loss of relevant information, thus leading to reduced forecasting capabilities. Therefore, adapting VAR processes to different temporal resolutions can enhance computational efficiency and improve forecasting capabilities while maintaining the integrity of temporal information.

Given a VAR process at a fixed temporal resolution, the problem of VAR forecasting at a finer temporal resolution has a potential practical utility in diverse fields, including finance, climatology, energy grids, logistics, and epidemiology. The most straightforward approach in the literature is, arguably, the use of interpolation methods; that is, estimating intermediate trajectory values, for instance, via splines [2]. However, they are prone to introduce biases and to overfit or to oversmooth, due to a priori behavioral assumptions, especially when the underlying trajectory dynamics are unknown. Moreover, many interpolation techniques do not provide uncertainty estimates, making it difficult to assess the reliability of the interpolated values. Another relatively direct approach is mixed-frequency VAR models [3]. These models allow for the integration of variables observed at distinct temporal resolutions. However, it is unclear how to vary the finest temporal resolution of a given mixed-frequency VAR model by demand. Continuous-time VAR models are conceptually different as they can extend discrete time instances to continuous time by modeling the trajectory as a stochastic differential equation (SDE) instead [4, 5]. Similarly, neural SDEs [6, 7] can be trained to approximate the dynamics of a given VAR at a fixed temporal resolution and predict at any desired temporal resolution afterward. However, both often carry an additional computational cost and lack of interpretability (especially dominant for neural SDEs) compared to ordinary VAR processes.

Contribution. Unlike existing literature, this paper presents a novel approach for adapting a given VAR process into another VAR process at a finer temporal resolution, while closely preserving the forecast estimates, at the time locations given by the original temporal resolution, and the associated credible intervals in the long run.

Organization. To present the proposed approach in a self-contained manner, this paper first provides a summary of the definition and fundamental properties of VAR processes in Sec. II. Then, Sec. III introduces the resolution grids associated with the resolution refinement, together with the concept of upsampling matrices, which represent mappings between any two grids at different resolutions. The problem of finding an adequate VAR process and upsampling matrix for a given resolution refinement is formulated in Sec. IV, and a solution is proposed in Sec. V. Experiments are presented in Sec. VI, and Sec. VII concludes the paper.

This work was supported in part by the IKTPLUSS DISCO grant 338740 from the Research Council of Norway, and by the VALIDATE project grant 101057263 from the EU HORIZON-RIA.
code: https://github.com/SIGIPRO/upsampledVAR

II. Vector autoregressive processes

As outlined in Sec. I, VAR processes are statistical models normally used to describe the joint temporal evolution of a set of d variables. In its usual form, the VAR process collects these variables in a vector of length d. Then, the evolution of these variables is described as a linear function of their previous values plus a noise disturbance, typically Gaussian. The number of previous time instances used in the recursion is commonly referred to as the order of the VAR process. On the other hand, a lagged value is typically the value of a variable in a previous time instance (relative to the current time instance). So, in general, a pth-order VAR, or VAR(p), refers to a VAR model that includes the p latest lagged values. Formally,

$$\tilde{\boldsymbol{y}}_t = \boldsymbol{c} + \boldsymbol{A}_1\tilde{\boldsymbol{y}}_{t-1} + \boldsymbol{A}_2\tilde{\boldsymbol{y}}_{t-1} + \cdots + \boldsymbol{A}_p\tilde{\boldsymbol{y}}_{t-p} + \boldsymbol{u}_t, \quad (1)$$

where the matrices in $\{\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p\} \subset \mathbb{R}^{d\times d}$ contain the model parameters, $\tilde{\boldsymbol{y}}_t \in \mathbb{R}^d$ contains the variable values at time step t, $\boldsymbol{c} \in \mathbb{R}^d$ serves as the intercept on the model and the terms $\boldsymbol{u}_t \in \mathbb{R}^d$ are white Gaussian noise with covariance $\boldsymbol{\Sigma}$, i.e., $\boldsymbol{u}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ and $\mathbb{E}[\boldsymbol{u}_t\boldsymbol{u}_{t-h}^\top] = 0$ for all $h \in \mathbb{Z}\setminus\{0\}$.

A. Mean-adjusted form of a VAR(p) process

Without loss of generality, we can rewrite the VAR(p) process in (1) as

$$\boldsymbol{y}_t = \boldsymbol{A}_1\boldsymbol{y}_{t-1} + \boldsymbol{A}_2\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{A}_p\boldsymbol{y}_{t-p} + \boldsymbol{u}_t, \quad (2)$$

where $\boldsymbol{y}_t := \tilde{\boldsymbol{y}}_t - \mathbb{E}[\tilde{\boldsymbol{y}}_t]$. This is known as the mean-adjusted form. From now on, when we refer to a VAR process, we refer to its mean-adjusted form unless stated otherwise.

B. Companion form of a VAR(p) process

A VAR(p) process can always be equivalently rewritten as a VAR(1) process by appropriately rearranging its terms. The resulting representation is known as the companion form, and it is a convenient characterization for ensuing analytical derivations. Formally,

$$\boldsymbol{Y}_t = \boldsymbol{\Gamma}\boldsymbol{Y}_{t-1} + \boldsymbol{U}_t, \quad (3)$$

where $\boldsymbol{Y}_t := vec\left([\boldsymbol{y}_t, \boldsymbol{y}_{t-1}, \ldots, \boldsymbol{y}_{t-p+1}]\right) \in \mathbb{R}^{dp\times 1}$, being vec the vectorization operator, and

$$\boldsymbol{\Gamma} := \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 & \cdots & \boldsymbol{A}_{p-1} & \boldsymbol{A}_p \\ \boldsymbol{I} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I} & \boldsymbol{0} \end{bmatrix} \in \mathbb{R}^{dp\times dp}, \quad (4)$$

where $\boldsymbol{I}$ denotes an identity matrix of adequate dimensionality, in this case $d \times d$, and $\boldsymbol{U}_t := vec\left([\boldsymbol{u}_t, \boldsymbol{0}, \cdots, \boldsymbol{0}]\right) \in \mathbb{R}^{dp\times 1}$. Correspondingly, $\boldsymbol{U}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega})$, where $\boldsymbol{\Omega} := \mathbb{E}[\boldsymbol{U}_t\boldsymbol{U}_t^\top] \in \mathbb{R}^{dp\times dp}$.

For completeness, let us define $\boldsymbol{J} := \boldsymbol{J}_p = [\boldsymbol{I}, \boldsymbol{0}, \ldots, \boldsymbol{0}] \in \{0,1\}^{d\times dp}$ as an operator that selects the first d rows. Then, note that $\boldsymbol{y}_t = \boldsymbol{J}\boldsymbol{Y}_t$, $\boldsymbol{u}_t = \boldsymbol{J}\boldsymbol{U}_t$, $\boldsymbol{U}_t = \boldsymbol{J}^\top\boldsymbol{u}_t$, $\boldsymbol{\Sigma} = \boldsymbol{J}\boldsymbol{\Omega}\boldsymbol{J}^\top$, and $\boldsymbol{\Omega} = \boldsymbol{J}^\top\boldsymbol{\Sigma}\boldsymbol{J}$.

C. Stability of a VAR(p) process

A VAR process is deemed stable when the effect of the noise disturbances dissipates over time. From the companion form perspective, this occurs when all eigenvalues of $\boldsymbol{\Gamma}$ lie within the unit circle; or equivalently, when its spectral radius $\rho(\boldsymbol{\Gamma}) := \max\{|\gamma_1|, \cdots, |\gamma_{dp}|\}$, where $\gamma_1, \ldots, \gamma_{dp}$ are the eigenvalues of $\boldsymbol{\Gamma}$, is less than one.

A stable VAR process is covariance-stationary [8]. Therefore, all terms in the sequence share the same first moment, have a finite second moment, and exhibit shift-invariant covariance between any two terms in the sequence (also known as autocovariance). From now on, all VAR processes are assumed to be stable, with well-defined moments, unless stated otherwise.

D. Moments of a VAR(p) process

By definition, the first moment of a VAR process in its mean-adjusted form is zero, i.e., $\mathbb{E}[\boldsymbol{Y}_t] = \mathbb{E}[\boldsymbol{Y}_{t-h}] = 0$ for all $h \in \mathbb{Z}\setminus\{0\}$. Regarding the second moment, let us denote the autocovariance at shift h as $\boldsymbol{C}_Y(h) := \mathbb{E}[\boldsymbol{Y}_t\boldsymbol{Y}_{t-h}^\top] \in \mathbb{R}^{dp\times dp}$. Notice that by construction $\boldsymbol{C}_Y(h) = \boldsymbol{C}_Y(-h)^\top$. Then, multiplying (3) by $\boldsymbol{Y}_{t-h}^\top$ we obtain

$$\boldsymbol{Y}_t\boldsymbol{Y}_{t-h}^\top = \boldsymbol{\Gamma}\boldsymbol{Y}_{t-1}\boldsymbol{Y}_{t-h}^\top + \boldsymbol{U}_t\boldsymbol{Y}_{t-h}^\top, \quad (5)$$

and by taking the expectation we have

$$\boldsymbol{C}_Y(0) = \boldsymbol{\Gamma}\boldsymbol{C}_Y(1)^\top + \boldsymbol{\Omega}, \text{ for } h = 0, \quad (6a)$$

$$\boldsymbol{C}_Y(h) = \boldsymbol{\Gamma}\boldsymbol{C}_Y(h-1), \text{ for any } h > 0, \quad (6b)$$

because $\mathbb{E}[\boldsymbol{U}_t\boldsymbol{Y}_{t'}^\top]$ is 0 for $t \neq t'$ and equals $\boldsymbol{\Omega}$ otherwise. From here, given $\boldsymbol{\Gamma}$ and $\boldsymbol{\Omega}$, we can determine $\boldsymbol{C}_Y(0)$ as follows: from (6b), we know that $\boldsymbol{C}_Y(1) = \boldsymbol{\Gamma}\boldsymbol{C}_Y(0)$. Substituting this value in (6a) leads to

$$\boldsymbol{C}_Y(0) = \boldsymbol{\Gamma}\boldsymbol{C}_Y(0)\boldsymbol{\Gamma}^\top + \boldsymbol{\Omega}. \quad (7)$$

Theorem 1 ([9]). Let $\boldsymbol{D}$, $\boldsymbol{E}$ and $\boldsymbol{F}$ be matrices such that one can form the matrix product $\boldsymbol{D}\boldsymbol{E}\boldsymbol{F}$, then

$$vec(\boldsymbol{D}\boldsymbol{E}\boldsymbol{F}) = (\boldsymbol{F}^\top \otimes \boldsymbol{D})\, vec(\boldsymbol{E}) \quad (8)$$

holds, where $\otimes$ denotes the Kronecker product.

From **Theorem 1**, the relation in (7), and after several straightforward algebraic steps, one obtains

$$\boldsymbol{C}_Y(0) = vec^{-1}\left((\boldsymbol{I} - \boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})^{-1}\, vec(\boldsymbol{\Omega})\right), \quad (9)$$

where, in this case, the identity matrix is of dimensionality $(dp)^2 \times (dp)^2$.

Finally, the moments of the mean-adjusted form can be easily recovered from the moments computed over the companion form by recalling that $\boldsymbol{y}_t = \boldsymbol{J}\boldsymbol{Y}_t$, see Sec. II-B. That is, $\mathbb{E}[\boldsymbol{y}_t] = \boldsymbol{J}\mathbb{E}[\boldsymbol{Y}_t] = 0$ and $\boldsymbol{C}_y(0) := \mathbb{E}[\boldsymbol{y}_t\boldsymbol{y}_t^\top] = \boldsymbol{J}\boldsymbol{C}_Y(0)\boldsymbol{J}^\top$.

E. Conditional moments of a VAR(p) process

The first and second conditional moments of a VAR process can be used as a point forecast estimate and to build the associated credible interval, respectively. In the case of forecasting h future instances from instance t, the conditional mean is

$$\mathbb{E}[\boldsymbol{y}_{t+h}|\boldsymbol{Y}_t] = \boldsymbol{J}\mathbb{E}[\boldsymbol{Y}_{t+h}|\boldsymbol{Y}_t] = \boldsymbol{J}\boldsymbol{\Gamma}^h\boldsymbol{Y}_t. \tag{10}$$

Following the same reasoning, the conditional covariance matrix is

$$\begin{aligned}\text{Var}[\boldsymbol{y}_{t+h}|\boldsymbol{Y}_t] &:= \boldsymbol{J}\text{Var}[\boldsymbol{Y}_{t+h}|\boldsymbol{Y}_t]J^\top && \text{(11a)}\\ &= \boldsymbol{J}\left(\sum_{i=0}^{h-1}\boldsymbol{\Gamma}^i\boldsymbol{\Omega}{\boldsymbol{\Gamma}^i}^\top\right)\boldsymbol{J}^\top && \text{(11b)}\\ &= \sum_{i=0}^{h-1}\boldsymbol{J}\boldsymbol{\Gamma}^i\,\boldsymbol{\Omega}\,\left(\boldsymbol{J}\boldsymbol{\Gamma}^i\right)^\top. && \text{(11c)}\end{aligned}$$

where $\text{Var}[\boldsymbol{y}] := \text{Cov}\left[\boldsymbol{y},\boldsymbol{y}\right] = \mathbb{E}[(\boldsymbol{y}-\mathbb{E}[\boldsymbol{y}])(\boldsymbol{y}-\mathbb{E}[\boldsymbol{y}])^\top]$, for any random vector $\boldsymbol{y}$ [10, Chapter III.5].

Theorem 2 ([9]). If $\boldsymbol{D}$, $\boldsymbol{E}$, $\boldsymbol{F}$ and $\boldsymbol{G}$ are matrices of such size that one can form the products $\boldsymbol{DF}$ and $\boldsymbol{EG}$, then the following identity holds,

$$(\boldsymbol{D}\otimes\boldsymbol{E})(\boldsymbol{F}\otimes\boldsymbol{G}) = (\boldsymbol{DF})\otimes(\boldsymbol{EG}). \tag{12}$$

Corollary 2.1. Applying **Theorem 2** recursively, one can conclude that any square matrix $\boldsymbol{T}$ satisfies

$$\boldsymbol{T}^i\otimes\boldsymbol{T}^i = \left(\boldsymbol{T}^{i-1}\otimes\boldsymbol{T}^{i-1}\right)(\boldsymbol{T}\otimes\boldsymbol{T}) = (\boldsymbol{T}\otimes\boldsymbol{T})^i. \tag{13}$$

Applying **Theorem 1**, **Theorem 2**, and **Corollary 2.1** over (11c) leads to

$$\begin{aligned}vec\left(\text{Var}[\boldsymbol{y}_{t+h}|\boldsymbol{Y}_t]\right) &= \sum_{i=0}^{h-1}(\boldsymbol{J}\boldsymbol{\Gamma}^i)\otimes(\boldsymbol{J}\boldsymbol{\Gamma}^i)\,vec(\boldsymbol{\Omega}) && \text{(14a)}\\ &= (\boldsymbol{J}\otimes\boldsymbol{J})\sum_{i=0}^{h-1}(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})^i\,vec(\boldsymbol{\Omega}). && \text{(14b)}\end{aligned}$$

Theorem 3 ([11]). For any square matrix $\boldsymbol{T}$, the finite geometric series generated by $\boldsymbol{T}$ up to the nth term satisfies

$$\boldsymbol{I}+\boldsymbol{T}+\cdots+\boldsymbol{T}^{n-1} = (\boldsymbol{I}-\boldsymbol{T})^{-1}(\boldsymbol{I}-\boldsymbol{T}^n), \tag{15}$$

if $\boldsymbol{I}-\boldsymbol{T}$ is invertible.

Corollary 3.1. The infinite geometric series generated by $\boldsymbol{T}$ in **Theorem 3** converges to $(\boldsymbol{I}-\boldsymbol{T})^{-1}$ if and only if $\rho(\boldsymbol{T}) < 1$.

From **Theorem 3** and **Corollary 2.1**, we can reach the following analytical relation

$$\begin{aligned}g(\boldsymbol{\Gamma},h) &:= \sum_{i=0}^{h-1}(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})^i && \text{(16a)}\\ &= (\boldsymbol{I}-\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})^{-1}\left(\boldsymbol{I}-\boldsymbol{\Gamma}^h\otimes\boldsymbol{\Gamma}^h\right). && \text{(16b)}\end{aligned}$$

Then, from (14b), noting that $vec(\boldsymbol{\Omega}) = (\boldsymbol{J}\otimes\boldsymbol{J})^\top vec(\boldsymbol{\Sigma})$, and making use of the auxiliary function in (16), we obtain

$$\text{Var}[\boldsymbol{y}_{t+h}|\boldsymbol{Y}_t] = vec^{-1}\left((\boldsymbol{J}\otimes\boldsymbol{J})\,g(\boldsymbol{\Gamma},h)\,(\boldsymbol{J}\otimes\boldsymbol{J})^\top vec(\boldsymbol{\Sigma})\right). \tag{17}$$

Finally, it is worth mentioning that, with the assistance of **Corollary 3.1**, it can be readily shown that the long-run conditional covariance coincides with the unshifted autocovariance of the process, i.e., $\lim_{h\to\infty}\text{Var}[\boldsymbol{y}_{t+h}|\boldsymbol{Y}_t] = \boldsymbol{C}_y(0)$.

III. Resolution grid

Suppose we are given a sequence of terms and we arrange them in a uniform grid. Then, any other sequence of terms that can be arranged with aligned terms and with $r-1$ intermediate (additional) terms is considered a sequence with an integer resolution r with respect to the given sequence. For instance,

```
- x - - - - - - - x - - - - - - - - x -
- ○ - - - ○ - - - ○ - - - ○ - - - ○ -
- * - * - * - * - * - * - * - * - * - * -
```

the sequence of $*$'s has a resolution $r = 4$ with respect to the sequence of x's and a resolution $r = 2$ with respect to the sequence of ○'s.

A. Upsampling matrices

Suppose we have two sequences of d-dimensional values, namely sequences A and B, where sequence A is obtained from uniformly subsampling every rth term of sequence B. In this setting, sequence B has a resolution r with respect to the resolution grid given by sequence A. Now, suppose we want to observe the last q values of sequence B, but we only have access to the last p values of sequence A. That is, at any time instance we have partial information about the values of sequence B.

For example, if we consider the ordered sets $\{\boldsymbol{y}_n^B\}_{n\in\mathbb{Z}}$ and $\{\boldsymbol{y}_n^A : \boldsymbol{y}_n^A = \boldsymbol{y}_{rn}^B\}_{n\in\mathbb{Z}}$ as the sequences B and A, respectively, and we set $p = 2$, $q = 4$ and $r = 2$, then, at any arbitrary instance t,

–	–	–	$\boldsymbol{y}_{t-1}^A$	–	–	–	$\boldsymbol{y}_t^A$	–
–	$\boldsymbol{y}_{2t-3}^B$	–	$\boldsymbol{y}_{2t-2}^B$	–	$\boldsymbol{y}_{2t-1}^B$	–	$\boldsymbol{y}_{2t}^B$	–

the aligned terms $\boldsymbol{y}_{2t}^B = \boldsymbol{y}_t^A$ and $\boldsymbol{y}_{2t-2}^B = \boldsymbol{y}_{t-1}^A$ are equal. On the other hand, the values of $\boldsymbol{y}_{2t-1}^B$ and $\boldsymbol{y}_{2t-3}^B$ remain unknown.

To solve this problem, we can rely on any imputation method that best suits the characteristics of sequences A and B. Arguably, the most straightforward approach is to use a linear transformation that preserves the value of the terms that are aligned. That is,

$$\boldsymbol{Y}_{rt}^B = \boldsymbol{M}_{A\to B}\boldsymbol{Y}_t^A, \tag{18}$$

where $\boldsymbol{Y}_t^A \in \mathbb{R}^{dp\times 1}$ and $\boldsymbol{Y}_{rt}^B \in \mathbb{R}^{dq\times 1}$ are the vectorization of the values at hand, and $\boldsymbol{M}_{A\to B} \in \mathbb{R}^{q\times p}\otimes\mathbb{R}^{d\times d}$ is a block matrix such that aligned values in the resolution grid are equal, henceforth an upsampling matrix.

Continuing with the previous example, (18) can expand to

$$\begin{bmatrix} \boldsymbol{y}^B_{2t} \\ \boldsymbol{y}^B_{2t-1} \\ \boldsymbol{y}^B_{2t-2} \\ \boldsymbol{y}^B_{2t-3} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{M}_{2,1} & \boldsymbol{M}_{2,2} \\ \boldsymbol{0} & \boldsymbol{I} \\ \boldsymbol{M}_{4,1} & \boldsymbol{M}_{4,2} \end{bmatrix} \begin{bmatrix} \boldsymbol{y}^A_t \\ \boldsymbol{y}^A_{t-1} \end{bmatrix}, \tag{19}$$

where every $\boldsymbol{M}_{i,j} \in \mathbb{R}^{d\times d}$ denotes each of the unconstrained blocks that comprise the upsampling matrix $\boldsymbol{M}_{A\to B}$.

In general, any valid upsampling matrix is constructed as

$$[\boldsymbol{M}_{A\to B}]_{i,j} = \begin{cases} \boldsymbol{I}, & \text{if } \underline{i} \bmod r = 0,\ k < p \text{ and } \underline{j} = k, \\ \boldsymbol{0}, & \text{if } \underline{i} \bmod r = 0,\ k < p \text{ and } \underline{j} \neq k, \\ \boldsymbol{M}_{i,j}, & \text{otherwise,} \end{cases} \tag{20}$$

where $i = 1,\dots,q$ and $j = 1,\dots,p$ are the block row and block column indices, $\underline{i} := i-1$, $\underline{j} := j-1$, and $k := \lfloor \underline{i}/r \rfloor$.

B. Upsampled forecasting

Upsampling matrices can assist VAR forecasts from an observation at a coarser resolution. To see this, let us consider again the case discussed in Sec. III-A. At any instance t, one can use the last observation of process A, i.e., $\boldsymbol{Y}^A_t$, to forecast h instances into the future of process B. That is, $\mathbb{E}\left[\boldsymbol{y}^B_{rt+h} | \boldsymbol{Y}^B_{rt}\right] = \mathbb{E}\left[\boldsymbol{y}^B_{rt+h} | \boldsymbol{M}_{A\to B}\boldsymbol{Y}^A_t\right]$. Furthermore, notice that the conditional covariance (17) is independent of the observation it is conditioned to, and so are the credible intervals of the forecast.

IV. Problem formulation

Given a VAR(p) process A, the ultimate goal is to find a VAR(q) process B with a resolution r with respect to process A, and an upsampling matrix between them, such that the forecasts of processes A and B, and their long-run conditional covariances, are approximately equal according to some closeness criterion. Mathematically,

$$\mathbb{E}\left[\boldsymbol{y}^B_{rt+rh} | \boldsymbol{M}_{A\to B}\boldsymbol{Y}^A_t\right] \simeq \mathbb{E}\left[\boldsymbol{y}^A_{t+h} | \boldsymbol{Y}^A_t\right], \tag{21a}$$

$$\boldsymbol{C}^B_y(0) \simeq \boldsymbol{C}^A_y(0), \tag{21b}$$

for any instance t, horizon h and observation $\boldsymbol{Y}^A_t$.

Intuitively, any process B and upsampling matrix that satisfy (21), can produce the same point forecast, in the coinciding grid locations, and has the same long-run conditional covariance as the given coarser process A. Moreover, it can also produce r intermediate point forecasts. In this sense, process B can be seen as a finer temporal resolution counterpart of process A.

On the other hand, the solution to the system of equations (21) may not lie on the set of stable VAR processes or valid upsampling matrices. In that case, one needs to seek an approximate solution that satisfies such constraints.

V. Proposed solution

To tackle problem (21a), we adopt the strategy of minimizing a cost function based on the expected Euclidean norm of the difference between the forecasts $\mathbb{E}\left[\boldsymbol{y}^A_{t+h} | \boldsymbol{Y}^A_t\right]$ and $\mathbb{E}\left[\boldsymbol{y}^B_{rt+rh} | \boldsymbol{M}_{A\to B}\boldsymbol{Y}^A_t\right]$ over all possible starting values $\boldsymbol{Y}^A_t$ and finite forecast horizons $h = 1,\dots,H$. The optimization variables are thus the VAR parameters of process B and the unconstrained entries of the upsampling matrix $\boldsymbol{M}_{A\to B}$.

Using the analytical form of the VAR forecasts (10), one can define an instantaneous loss between the forecasts for a single horizon h and input observation $\boldsymbol{Y}^A_t$ as

$$\ell(h, \boldsymbol{Y}^A_t) := \left\| \boldsymbol{J}_p \boldsymbol{\Gamma}^h_A \boldsymbol{Y}^A_t - \boldsymbol{J}_q \boldsymbol{\Gamma}^{hr}_B \boldsymbol{M}_{A\to B} \boldsymbol{Y}^A_t \right\|^2_2, \tag{22}$$

where only relevant dependencies for ensuing derivations are made explicit. Such an instantaneous loss can be readily used to compute a cost function by sample average approximation. However, this numerical approach is limited to the number of input observations used.

Theorem 4. The expected instantaneous loss over all input observations, defined as $\mathcal{L}(h) := \mathbb{E}_{\boldsymbol{Y}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{C}^A_Y(0))}\left[\ell(h,\boldsymbol{Y})\right]$, admits the following closed-form representation

$$\mathcal{L}(h) = \left\| \left(\boldsymbol{J}_p \boldsymbol{\Gamma}^h_A - \boldsymbol{J}_q \boldsymbol{\Gamma}^{hr}_B \boldsymbol{M}_{A\to B}\right) \boldsymbol{L}_A \right\|^2_F, \tag{23}$$

where $\boldsymbol{L}_A$ is the lower triangular matrix in the Cholesky decomposition of $\boldsymbol{C}_Y(0)$.

Proof. Let us define $\boldsymbol{S} := \boldsymbol{J}_p \boldsymbol{\Gamma}^h_A - \boldsymbol{J}_q \boldsymbol{\Gamma}^{hr}_B \boldsymbol{M}_{A\to B} \in \mathbb{R}^{d\times dp}$ as an auxiliary matrix. Then,

$$\begin{aligned} \mathcal{L}(h) &= \mathbb{E}\left[\ell(h,\boldsymbol{Y})\right] = \mathbb{E}\left[\|\boldsymbol{S}\boldsymbol{Y}\|^2_2\right] && \text{(24a)} \\ &= \mathbb{E}\left[\boldsymbol{Y}^\top \boldsymbol{S}^\top \boldsymbol{S}\boldsymbol{Y}\right] = \mathbb{E}\left[\operatorname{tr}\left(\boldsymbol{Y}^\top \boldsymbol{S}^\top \boldsymbol{S}\boldsymbol{Y}\right)\right] && \text{(24b)} \\ &= \operatorname{tr}\left(\boldsymbol{S}^\top \boldsymbol{S}\, \mathbb{E}\left[\boldsymbol{Y}\boldsymbol{Y}^\top\right]\right) = \operatorname{tr}\left(\boldsymbol{S}\, \boldsymbol{C}^A_Y(0)\, \boldsymbol{S}^\top\right) && \text{(24c)} \\ &= \operatorname{tr}\left(\boldsymbol{S}\boldsymbol{L}_A\boldsymbol{L}_A^\top\boldsymbol{S}^\top\right) = \|\boldsymbol{S}\boldsymbol{L}_A\|^2_F. && \text{(24d)} \end{aligned}$$

Alternatively, from **Theorem 4**, one can construct an analytical cost function $\mathcal{C}\left(\boldsymbol{B}_1,\dots,\boldsymbol{B}_q,\boldsymbol{M}\right) := \sum_{h=1}^H \mathcal{L}(h)$. Hence, the aforementioned optimization problem can be cast as

$$\begin{aligned} \boldsymbol{M}^*, \{\boldsymbol{B}^*_i\}^q_{i=1} \in \ \arg\min_{\boldsymbol{M},\{\boldsymbol{B}_i\}^q_{i=1}} \ & \mathcal{C}\left(\boldsymbol{B}_1,\dots,\boldsymbol{B}_q,\boldsymbol{M}\right) && \text{(25a)} \\ \text{subject to: } & \{\boldsymbol{B}_i\}^q_{i=1} \in \mathcal{G}_{(d,q)} && \text{(25b)} \\ & \boldsymbol{M} \in \mathcal{M}_{(d,q,p,r)}, && \text{(25c)} \end{aligned}$$

where $\mathcal{G}_{(d,q)} := \left\{\{\boldsymbol{G}_i\}^q_{i=1} \subset \mathbb{R}^{d\times d} : \rho(\boldsymbol{\Gamma}_G) < 1\right\}$ is the set of all parameter configurations that conform a stable d-dimensional VAR(q) model, $\mathcal{M}_{(d,q,p,r)}$ is the set of all valid upsampling matrices, i.e., as in (20), for this setting, and $\boldsymbol{M}^*$, $\boldsymbol{B}^*_1,\dots,\boldsymbol{B}^*_q$ denote the optimal upsampling matrix and process B parameter matrices, respectively.

Note that problem (25) is nonconvex. In this work, we tackle it by projected gradient descent, using a default setting Adam optimizer [12] in combination with random initializations to find well-performing local minima. Alternative optimization methods are outside the scope of the present paper and are left as potential future work.

Next, we aim to find the noise covariance matrix to fully model process B. To do this, note that the relation in (21b) can be expanded to

$$\boldsymbol{J}_p\, \boldsymbol{C}^A_Y(0)\, \boldsymbol{J}_p^\top \simeq \boldsymbol{J}_q\, \boldsymbol{C}^B_Y(0)\, \boldsymbol{J}_q^\top, \tag{26a}$$

$$\left(\boldsymbol{J}_p \otimes \boldsymbol{J}_p\right) vec\left(\boldsymbol{C}^A_Y(0)\right) \simeq \left(\boldsymbol{J}_q \otimes \boldsymbol{J}_q\right) vec\left(\boldsymbol{C}^B_Y(0)\right), \tag{26b}$$

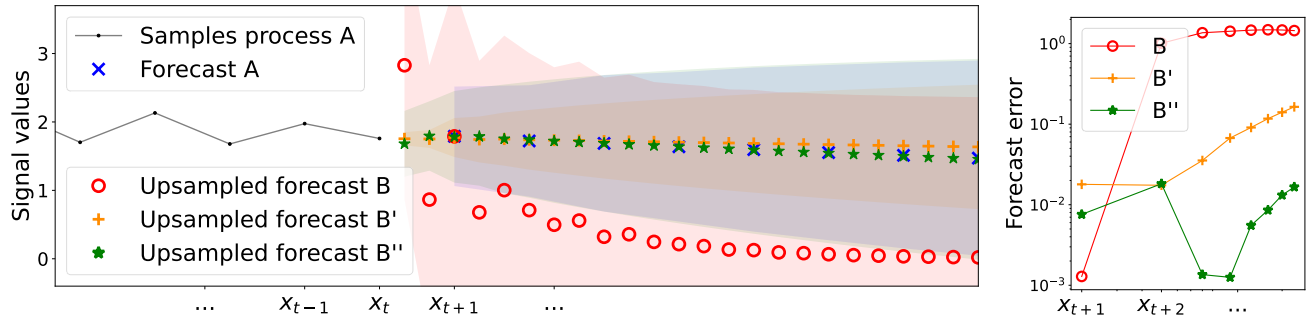

Fig. 1. A single variable visualization of the experiment described in Sec. VI. The shaded areas represent one conditional standard deviation, i.e., the square root of the diagonal elements of (17), above and below the corresponding forecast. The forecast error is defined as the absolute difference between the upsampled forecast samples from processes B, B', or B", and the forecast samples from process A. The forecast error is plotted on a double logarithmic scale.

and then, by defining $\boldsymbol{K}_p := \boldsymbol{J}_p \otimes \boldsymbol{J}_p$, using that $vec(\boldsymbol{\Omega}_A) = \boldsymbol{K}_p^\top vec(\boldsymbol{\Sigma}_A)$, and defining $\boldsymbol{\Psi}_A := \boldsymbol{K}_p(\boldsymbol{I} - \boldsymbol{\Gamma}_A \otimes \boldsymbol{\Gamma}_A)^{-1}\boldsymbol{K}_p^\top$ (same for process B), it follows that

$$\boldsymbol{\Psi}_A \, vec(\boldsymbol{\Sigma}_A) \simeq \boldsymbol{\Psi}_B \, vec(\boldsymbol{\Sigma}_B). \tag{26c}$$

From (26c), and using the model parameters in $\boldsymbol{B}_1^*, \ldots, \boldsymbol{B}_q^*$, the noise covariance of process B can be computed as

$$\boldsymbol{\Sigma}_B^* = \text{proj}_{\succeq 0}\left(vec^{-1}\left(\boldsymbol{\Psi}_{B^*}^{-\dagger}\boldsymbol{\Psi}_A vec(\boldsymbol{\Sigma}_A)\right)\right), \tag{27}$$

where $\text{proj}_{\succeq 0}$ refers to the projection onto the space of positive semidefinite matrices, and $^{-\dagger}$ denotes the Moore-Penrose pseudo-inverse from the left.

Closed-form solution. The problem in which $p = q = 1$ and $\boldsymbol{A}_1$ is positive semidefinite, admits an analytical solution. Notice that in that case, $\boldsymbol{Y}_t^A = \boldsymbol{y}_t^A$, $\boldsymbol{Y}_{rt}^B = \boldsymbol{y}_{rt}^B$, and thus, $\boldsymbol{M}_{A\to B} = \boldsymbol{I}$. Then, by solving (25) one obtains $\boldsymbol{B}_1^* = \boldsymbol{A}_1^{\frac{1}{r}}$, which satisfies (21a) for all h.

VI. Experiments

To validate our upsampled VAR process, we perform the following experiment. A 2-dimensional VAR(2) process A is used as a starting point. The goal is to estimate another 2-dimensional VAR(3) process, operating at a finer resolution $r = 3$, together with an upsampling matrix of size 6×4.

To do so, we present three formulations, based on the problem discussed in Sec. IV and V, which differ only in the farthest forecasting horizon H. We examine H values of 1, 2, and 3, resulting in processes B, B', and B", respectively. To show the original process and subsequent forecasts, samples of the coarser process A are shown up until a given time x_t, and from there onwards, the forecast samples are estimated from this process A. These are shown in comparison with the upsampled forecast samples from processes B, B', and B". We observe that process B", with $H = 3$, closely matches the forecast samples from process A. Conversely, for process B, with $H = 1$, the upsampled forecast samples closely align with the first forecast but diverge from the subsequent forecasts from process A. Lastly, process B', with $H = 2$, exhibits an intermediate behavior. These results are visualized in Fig. 1.

As a concluding remark, choosing a short farthest horizon H reduces the complexity of the loss function (24) by limiting the number of cross-products among the learnable parameters, thereby facilitating the convergence to well-performing local minima. However, a low value of H may lead to overfitting and, consequently, undesirable performance. Therefore, H should be considered as a hyperparameter of our method.

VII. Conclusion

This work presents the upsampled VAR process, a discrete statistical model that increases the resolution at which any given VAR process operates. The proposed method is validated for multivariate time series forecasting. Future work includes studying nonlinear imputation methods, thus extending beyond upsampling matrices, and nonconvex optimization approaches efficiently exploiting the analytical structure of the posed optimization problem.

References

[1] Eric Zivot and Jiahui Wang. Vector autoregressive models for multivariate time series. *Modeling financial time series with S-PLUS®*, pages 385–429, 2006.

[2] Grace Wahba. Spline models for observational data. *Society for Industrial and Applied Mathematics*, 1990.

[3] Frank Schorfheide and Dongho Song. Real-time forecasting with a mixed-frequency VAR. *Journal of Business & Economic Statistics*, 33 (3):366–380, 2015.

[4] Andrew C Harvey and James H Stock. The estimation of higher-order continuous time autoregressive models. *Econometric Theory*, 1(1):97–117, 1985.

[5] Oisín Ryan and Ellen L Hamaker. Time to intervene: A continuous-time approach to network analysis and centrality. *Psychometrika*, 87 (1):214–252, 2022.

[6] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

[7] Junteng Jia and Austin R Benson. Neural jump stochastic differential equations. *Advances in Neural Information Processing Systems*, 32, 2019.

[8] Helmut Lütkepohl. *New introduction to multiple time series analysis*. Springer Science & Business Media, 2005.

[9] Jan R Magnus and Heinz Neudecker. *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons, 2019.

[10] William Feller. *An introduction to probability theory and its applications, Volume 2*, volume 2. John Wiley & Sons, 1991.

[11] Carl D Meyer. *Matrix analysis and applied linear algebra*. SIAM, 2023.

[12] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.