

Locally Adaptive Factor Processes for Multivariate Time Series

Daniele Durante

Bruno Scarpa

Department of Statistical Sciences

University of Padua

Padua 35121, Italy

DURANTE@STAT.UNIPD.IT

SCARPA@STAT.UNIPD.IT

David B. Dunson

Department of Statistical Science

Duke University

Durham, NC 27708-0251, USA

DUNSON@STAT.DUKE.EDU

Editor: Robert E. McCulloch

Abstract

In modeling multivariate time series, it is important to allow time-varying smoothness in the mean and covariance process. In particular, there may be certain time intervals exhibiting rapid changes and others in which changes are slow. If such time-varying smoothness is not accounted for, one can obtain misleading inferences and predictions, with over-smoothing across erratic time intervals and under-smoothing across times exhibiting slow variation. This can lead to mis-calibration of predictive intervals, which can be substantially too narrow or wide depending on the time. We propose a locally adaptive factor process for characterizing multivariate mean-covariance changes in continuous time, allowing locally varying smoothness in both the mean and covariance matrix. This process is constructed utilizing latent dictionary functions evolving in time through nested Gaussian processes and linearly related to the observed data with a sparse mapping. Using a differential equation representation, we bypass usual computational bottlenecks in obtaining MCMC and online algorithms for approximate Bayesian inference. The performance is assessed in simulations and illustrated in a financial application.

Keywords: Bayesian nonparametrics, locally varying smoothness, multivariate time series, nested Gaussian process, stochastic volatility

1. Introduction

In analyzing multivariate time series data, collected in financial applications, monitoring of influenza outbreaks and other fields, it is often of key importance to accurately characterize dynamic changes over time in not only the mean of the different elements (e.g., assets, influenza levels at different locations) but also the covariance. As shown in Figure 1, it is typical in many domains to cycle irregularly between periods of rapid and slow change; most statistical models are insufficiently flexible to capture such locally varying smoothness in assuming a single bandwidth parameter. Inappropriately restricting the smoothness to be constant can have a major impact on the quality of inferences and predictions, with over-

DAX30: Squared log returns

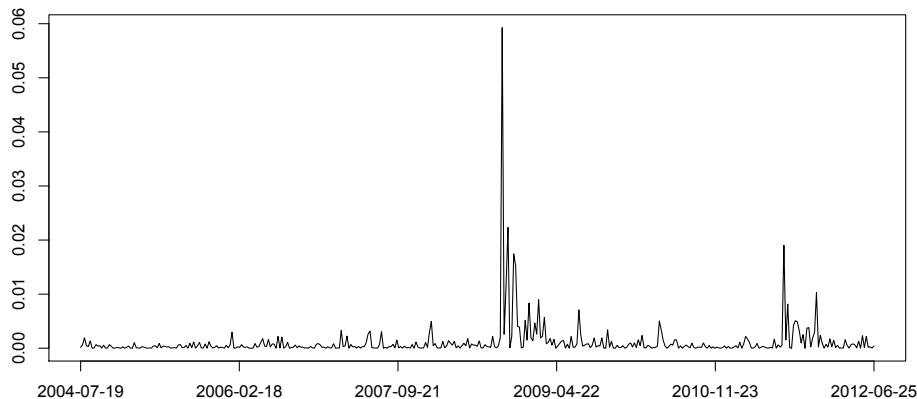


Figure 1: Squared log returns of DAX30. Weekly data from 19/07/2004 to 25/06/2012.

smoothing during times of rapid change. This leads to an under-estimation of uncertainty during such volatile times and an inability to accurately predict risk of extremal events.

Let $y_t = (y_{1t}, \dots, y_{pt})^T$ denote a random vector at time t , with $\mu(t) = E(y_t)$ and $\Sigma(t) = \text{cov}(y_t)$. Our focus is on Bayesian modeling and inference for the multivariate mean-covariance stochastic process, $\Gamma = \{\mu(t), \Sigma(t), t \in \mathcal{T}\}$ with $\mathcal{T} \subset \mathbb{R}^+$. Of particular interest is allowing locally varying smoothness, meaning that the rate of change in the $\{\mu(t), \Sigma(t)\}$ process is varying over time. To our knowledge, there is no previous proposed stochastic process for a coupled mean-covariance process, which allows locally varying smoothness. A key to our construction is the use of latent processes, which have time-varying smoothness. This results in a *locally adaptive factor* (LAF) process. We review the relevant literature below and then describe our LAF formulation.

1.1 Relevant Literature

There is a rich literature on modeling a $p \times 1$ time-varying mean vector $\mu(t)$, covering multivariate generalizations of autoregressive models (VAR) (see, e.g., Tsay, 2005), Kalman filtering (Kalman, 1960), nonparametric mean regression via Gaussian processes (GP) (Rasmussen and Williams, 2006), polynomial spline (Huang et al., 2002), smoothing spline (Hastie and Tibshirani, 1990) and kernel smoothing methods (Silverman, 1984). Such approaches perform well for slowly-changing trajectories with constant bandwidth parameters regulating implicitly or explicitly global smoothness; however, our interest is allowing smoothness to vary locally in continuous time. Possible extensions for local adaptivity include free knot splines (MARS) (Friedman, 1991), which perform well in simulations but the different strategies proposed to select the number and the locations of knots via step-wise knot selection (Friedman, 1991), Bayesian knot selection (Smith and Kohn, 1996) or MCMC methods (George and McCulloch, 1993), prove to be computationally intractable for moderately large p . Other flexible approaches include wavelet shrinkage (Donoho and

Johnstone, 1995), local polynomial fitting via variable bandwidth (Fan and Gijbels, 1995) and linear combination of kernels with variable bandwidths (Wolpert et al., 2011).

There is a separate literature on estimating a time-varying covariance matrix $\Sigma(t)$. This is particularly of interest in applications where volatilities and co-volatilities evolve through non constant paths. One popular approach estimates $\Sigma(t)$ via an exponentially weighted moving average (EWMA); see, e.g., Tsay (2005). This approach uses a single time-constant smoothing parameter $0 < \lambda < 1$, with extensions to accommodate locally varying smoothness not straightforward due to the need to maintain positive semidefinite $\Sigma(t)$ at every time. To allow for higher flexibility in the dynamics of the covariances, generalizations of EWMA have been proposed including the diagonal vector ARCH model (DVEC), (Bollerslev et al., 1988) and its variant, the BEKK model (Engle and Kroner, 1995). These models are computationally demanding and are not designed for moderate to large p . DCC-GARCH (Engle, 2002) improves the computational tractability of the previous approaches through a two-step formulation. However, the univariate GARCH assumed for the conditional variances of each time series and the higher level GARCH models with the same parameters regulating the evolution of the time-varying conditional correlations, restrict the evolution of the variance and covariance matrices. PC-GARCH (Ding, 1994; Burns, 2005) and O-GARCH (Alexander, 2001) perform dimensionality reduction through a latent factor formulation; see also van der Weide (2002). However, time-constant factor loadings and uncorrelated latent factors constrain the evolution of $\Sigma(t)$.

Such models fall far short of our goal of allowing $\Sigma(t)$ to be fully flexible with the dependence between $\Sigma(t)$ and $\Sigma(t + \Delta)$ varying with not just the time-lag Δ but also with time. In addition, these models do not handle missing data easily and tend to require long series for accurate estimation (Burns, 2005). Accommodating changes in continuous time is important in many applications, and avoids having the model be critically dependent on the time scale, with inconsistent models obtained as time units are varied.

Wilson and Ghahramani (2010) join machine learning and econometrics efforts by proposing a model for both mean and covariance regression in multivariate time series, improving previous work of Bru (1991) on Wishart processes in terms of computational tractability and scalability, allowing a more complex structure of dependence between $\Sigma(t)$ and $\Sigma(t + \Delta)$. Specifically, they propose a continuous time generalised Wishart process (GWP), which defines a collection of positive semi-definite random matrices $\Sigma(t)$ with Wishart marginals. Nonparametric mean regression for $\mu(t)$ is also considered via GP priors; however, the trajectories of means and covariances inherit the smooth behavior of the underlying Gaussian processes, limiting the flexibility of the approach across times exhibiting sharp changes.

Even for iid observations from a multivariate normal model with a single time stationary covariance matrix, there are well known problems with Wishart priors motivating a rich literature on dimensionality reduction techniques based on factor and graphical models. There has been abundant recent interest in applying such approaches to dynamic settings. Refer to Lopes and Carvalho (2007), Nakajima and West (2013) and the references cited therein for recent literature on Bayesian dynamic factor models for multivariate stochastic volatility. The Markov switching assumption for the levels of the common factor volatilities in Lopes and Carvalho (2007) improves flexibility, but may be restrictive in some applied fields and requires the additional choice of the number of possible regimes. Nakajima and West (2013) allow the factor loadings to evolve dynamically over time, while including sparsity through

a latent thresholding approach, leading to apparently improved performance in portfolio allocation. They assume a time-varying discrete-time autoregressive model, which allows the dependence in the covariance matrices $\Sigma(t)$ and $\Sigma(t + \Delta)$ to vary as a function of both t and Δ . However, the result is a richly parameterized and computationally challenging model, with selection of the number of factors proceeding by cross validation. Our emphasis is instead on developing continuous time stochastic processes for $\Sigma(t)$ and $\mu(t)$, which accommodate locally varying smoothness and provide relatively efficient MCMC computations based on a Gibbs sampler.

Fox and Dunson (2011) propose an alternative Bayesian covariance regression (BCR) model, which defines the covariance matrix as a regularized quadratic function of time-varying loadings in a latent factor model, characterizing the latter as a sparse combination of a collection of unknown Gaussian process dictionary functions. Although their approach provides a continuous time and highly flexible model that accommodates missing data and scales to moderately large p , there are two limitations motivating this article. Firstly, their proposed covariance stochastic process assumes a stationary dependence structure, and hence tends to under-smooth during periods of stability and over-smooth during periods of sharp changes. Secondly, the well known computational problems with usual GP regression are inherited, leading to difficulties in scaling to long series and issues in mixing of MCMC algorithms for posterior computation.

1.2 Contribution and Outline

Our proposed LAF process instead includes dictionary functions that are generated from nested Gaussian processes (nGP) (Zhu and Dunson, 2013), representing recently proposed priors which exploit stochastic differential equations (SDEs) to enforce GP priors for the function's m th order derivatives and favor local adaptivity by centering the latter on an higher level GP instantaneous mean. Such nGP reduces the GP computational burden involving matrix inversions from $O(T^3)$ to $O(T)$, with T denoting the length of the time series, while also allowing flexible locally varying smoothness. Marginalizing out the latent factors, we obtain a stochastic process that inherits these advantages. We also develop a different and more computationally efficient approach under this new model and propose an online implementation, which can accommodate streaming data. In Section 2, we describe the LAF structure with particular attention to prior specification. Section 3 explores the main features of the Gibbs sampler for posterior computation and outlines the steps for a fast online updating approach. In Section 4 we compare our model to BCR and to some of the most widely used models for multivariate stochastic volatility, through simulation studies. Finally in Section 5 an application to National Stock Market Indices across countries is examined.

2. Locally Adaptive Factor Processes

Our focus is on defining a novel locally adaptive factor process for $\Gamma = \{\mu(t), \Sigma(t), t \in \mathcal{T}\}$. In particular, taking a Bayesian approach, we define a prior $\Gamma \sim P$, where P is a probability measure over the space \mathcal{P} of p -variate mean-covariance processes on \mathcal{T} . In particular, each element of \mathcal{P} corresponds to a realization of the stochastic process Γ , and the measure P assigns probabilities to a σ -algebra of subsets of \mathcal{P} .

Although the proposed class of LAF processes can be used much more broadly, in conducting inferences in this article, we focus on the simple case in which data consist of vectors $y_i = (y_{1i}, \dots, y_{pi})^T$ collected at times t_i , for $i = 1, \dots, T$. These times can be unequally-spaced, or collected under an equally-spaced design with missing observations. An advantage of using a continuous-time process is that it is trivial to allow unequal spacing, missing data, and even observation times across which only a subset of the elements of y_i are observed. We additionally make the simplifying assumption that

$$y_i \sim N_p(\mu(t_i), \Sigma(t_i)).$$

It is straightforward to modify the methodology to accommodate substantially different observation models.

2.1 LAF Specification

A common strategy in modeling of large p matrices is to rely on a lower-dimensional factorization, with factor analysis providing one possible direction. Sparse Bayesian factor models have been particularly successful in challenging cases, while having advantages over frequentist competitors in incorporating a probabilistic characterization of uncertainty in the number of factors as well as the parameters in the loadings and residual covariance. For recent articles on Bayesian sparse factor analysis for a single large covariance matrix, refer to Bhattacharya and Dunson (2011), Pati et al. (2012) and the references cited therein.

In our setting, we are instead interested in letting the mean vector and the covariance matrix vary flexibly over time. Extending the usual factor analysis framework to this setting, we say that $\Gamma = \{\mu(t), \Sigma(t), t \in \mathcal{T}\} \sim \text{LAF}_{L,K}(\Theta, \Sigma_0, \Sigma_\xi, \Sigma_A, \Sigma_\psi, \Sigma_B)$ if

$$\mu(t) = \Theta \xi(t) \psi(t), \tag{1a}$$

$$\Sigma(t) = \Theta \xi(t) \xi(t)^T \Theta^T + \Sigma_0, \tag{1b}$$

where Θ is a $p \times L$ matrix of constant coefficients, $\Sigma_0 = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$, while $\xi(t)_{L \times K}$ and $\psi(t)_{K \times 1}$ are matrices comprising continuous dictionary functions evolving in time via nGP, $\xi_{lk}(t) \sim \text{nGP}([\Sigma_\xi]_{lk} = \sigma_{\xi_{lk}}^2, [\Sigma_A]_{lk} = \sigma_{A_{lk}}^2)$ and $\psi_k(t) \sim \text{nGP}([\Sigma_\psi]_k = \sigma_{\psi_k}^2, [\Sigma_B]_k = \sigma_{B_k}^2)$.

Restricting our attention on the generic element $\xi_{lk}(t) : \mathcal{T} \rightarrow \mathfrak{R}$ of the matrix $\xi(t)_{L \times K}$ (the same holds for $\psi_k(t) : \mathcal{T} \rightarrow \mathfrak{R}$), the nGP provides a highly flexible stochastic process on the dictionary functions whose smoothness, explicitly modeled by their m th order derivatives $D^m \xi_{lk}(t)$ via stochastic differential equations, is expected to be centered on a local instantaneous mean function $A_{lk}(t)$, which represents a higher-level Gaussian process, that induces adaptivity to locally varying smoothing. Specifically, we let

$$D^m \xi_{lk}(t) = A_{lk}(t) + \sigma_{\xi_{lk}} W_{\xi_{lk}}(t), \quad m \in N, \quad m \geq 2, \tag{2a}$$

$$D^n A_{lk}(t) = \sigma_{A_{lk}} W_{A_{lk}}(t), \quad n \in N, \quad n \geq 1, \tag{2b}$$

where $\sigma_{\xi_{lk}} \in \mathfrak{R}^+$, $\sigma_{A_{lk}} \in \mathfrak{R}^+$, $W_{\xi_{lk}}(t) : \mathcal{T} \rightarrow \mathfrak{R}$ and $W_{A_{lk}}(t) : \mathcal{T} \rightarrow \mathfrak{R}$ are independent Gaussian white noise processes with mean $E[W_{\xi_{lk}}(t)] = E[W_{A_{lk}}(t)] = 0$, for all $t \in \mathcal{T}$, and covariance function $E[W_{\xi_{lk}}(t)W_{\xi_{lk}}(t^*)] = E[W_{A_{lk}}(t)W_{A_{lk}}(t^*)] = 1$ if $t = t^*$, 0 otherwise. This formulation naturally induces a stochastic process for $\xi_{lk}(t)$ with varying smoothness,

where $E[D^m \xi_{lk}(t) | A_{lk}(t)] = A_{lk}(t)$, and initialization at t_1 based on the assumption

$$\begin{aligned} [\xi_{lk}(t_1), D^1 \xi_{lk}(t_1), \dots, D^{m-1} \xi_{lk}(t_1)]^T &\sim N_m(0, \sigma_{\mu_{lk}}^2 I_m), \\ [A_{lk}(t_1), D^1 A_{lk}(t_1), \dots, D^{n-1} A_{lk}(t_1)]^T &\sim N_n(0, \sigma_{\alpha_{lk}}^2 I_n). \end{aligned}$$

The Markovian property implied by SDEs in (2a) and (2b) represents a key advantage in terms of computational tractability as it allows a simple state space formulation. In particular, referring to Zhu and Dunson (2013) for $m = 2$ and $n = 1$ (this can be easily extended for higher m and n), and for $\delta_i = t_{i+1} - t_i$ sufficiently small, the process for $\xi_{lk}(t)$ along with its first order derivative $\xi'_{lk}(t)$ and the local instantaneous mean $A_{lk}(t)$ follow the approximated state equation

$$\begin{bmatrix} \xi_{lk}(t_{i+1}) \\ \xi'_{lk}(t_{i+1}) \\ A_{lk}(t_{i+1}) \end{bmatrix} = \begin{bmatrix} 1 & \delta_i & 0 \\ 0 & 1 & \delta_i \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \xi_{lk}(t_i) \\ \xi'_{lk}(t_i) \\ A_{lk}(t_i) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \omega_{i,\xi_{lk}} \\ \omega_{i,A_{lk}} \end{bmatrix}, \quad (3)$$

where $[\omega_{i,\xi_{lk}}, \omega_{i,A_{lk}}]^T \sim N_2(0, V_{i,lk})$, with $V_{i,lk} = \text{diag}(\sigma_{\xi_{lk}}^2 \delta_i, \sigma_{A_{lk}}^2 \delta_i)$.

Similarly to the nGP specification for the elements in $\xi(t)$, we can represent the nested Gaussian process for $\psi_k(t)$ with the following state equation:

$$\begin{bmatrix} \psi_k(t_{i+1}) \\ \psi'_k(t_{i+1}) \\ B_k(t_{i+1}) \end{bmatrix} = \begin{bmatrix} 1 & \delta_i & 0 \\ 0 & 1 & \delta_i \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \psi_k(t_i) \\ \psi'_k(t_i) \\ B_k(t_i) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \omega_{i,\psi_k} \\ \omega_{i,B_k} \end{bmatrix}, \quad (4)$$

for $k = 1, \dots, K$, where $[\omega_{i,\psi_k}, \omega_{i,B_k}]^T \sim N_2(0, S_{i,k})$, with $S_{i,k} = \text{diag}(\sigma_{\psi_k}^2 \delta_i, \sigma_{B_k}^2 \delta_i)$. Similarly to $\xi_{lk}(t)$, we let

$$\begin{aligned} [\psi_k(t_1), D^1 \psi_k(t_1), \dots, D^{m-1} \psi_k(t_1)]^T &\sim N_m(0, \sigma_{\mu_k}^2 I_m), \\ [B_k(t_1), D^1 B_k(t_1), \dots, D^{n-1} B_k(t_1)]^T &\sim N_n(0, \sigma_{\alpha_k}^2 I_n). \end{aligned}$$

There are two crucial aspects to highlight. Firstly, this formulation is defined at every point over a subset of the real line and allows an irregular grid of observations over t by relating the latent states at $i + 1$ to those at i through the distance between t_{i+1} and t_i where i represents a discrete order index and $t_i \in \mathcal{T}$ the time value related to the i th observation. Secondly, compared to Zhu and Dunson (2013) our approach represents an important generalization in: (i) extending the analysis to the multivariate case (i.e. y_i is a p -dimensional vector instead of a scalar) and (ii) accommodating locally adaptive smoothing not only on the mean but also on the time-varying covariance functions.

2.2 LAF Interpretation

Model (1a)–(1b) can be induced by marginalizing out the K -dimensional latent factors vector η_i , in the model

$$y_i = \Lambda(t_i) \eta_i + \epsilon_i, \quad \epsilon_i \sim N_p(0, \Sigma_0), \quad (5)$$

where $\eta_i = \psi(t_i) + \nu_i$ with $\nu_i \sim N_K(0, I_K)$ and elements $\psi_k(t) \sim \text{nGP}(\sigma_{\psi_k}^2, \sigma_{B_k}^2)$ for $k = 1, \dots, K$. In LAF formulation we assume moreover that the time-varying factor loadings

matrix $\Lambda(t)$ is a sparse linear combination, with respect to the weights of the $p \times L$ matrix Θ , of a much smaller set of continuous nested Gaussian processes $\xi_{lk}(t) \sim \text{nGP}(\sigma_{\xi_{lk}}^2, \sigma_{A_{lk}}^2)$ comprising the $L \times K$, with $L \ll p$, matrix $\xi(t)$. As a result

$$\Lambda(t_i) = \Theta \xi(t_i). \tag{6}$$

Such a decomposition plays a crucial role in further reducing the number of nested Gaussian processes to be modeled from $p \times K$ to $L \times K$ leading to a more computationally tractable formulation in which the induced $\Gamma = \{\mu(t), \Sigma(t), t \in \mathcal{T}\}$ follows a locally adaptive factor $\text{LAF}_{L,K}(\Theta, \Sigma_0, \Sigma_\xi, \Sigma_A, \Sigma_\psi, \Sigma_B)$ process where

$$\mu(t_i) = \text{E}(y_i \mid t = t_i) = \Theta \xi(t_i) \psi(t_i), \tag{7a}$$

$$\Sigma(t_i) = \text{cov}(y_i \mid t = t_i) = \Theta \xi(t_i) \xi(t_i)^T \Theta^T + \Sigma_0. \tag{7b}$$

There is a literature on using Bayesian factor analysis with time-varying loadings, but essentially all the literature assumes discrete-time dynamics on the loadings while our focus is instead on allowing the loadings, and hence the induced $\Gamma = \{\mu(t), \Sigma(t), t \in \mathcal{T}\}$ processes, to evolve flexibly in continuous time. Hence, we are most closely related to the literature on Gaussian process latent factor models for spatial and temporal data; refer, for example, to Lopes et al. (2008) and Lopes et al. (2011). In these models, the factor loadings matrix characterizes spatial dependence, with time-varying factors accounting for dynamic changes.

Fox and Dunson (2011) instead allow the loadings matrix to vary through a continuous time stochastic process built from latent $\text{GP}(0, c)$ dictionary functions independently for all $l = 1, \dots, L$ and $k = 1, \dots, K$, with c the squared exponential correlation function having $c(t, t^*) = \exp(-\kappa \|t - t^*\|_2^2)$. In our work we follow the lead of Fox and Dunson (2011) in using a nonparametric latent factor model as in (5)–(6), but induce fundamentally different behavior on $\Gamma = \{\mu(t), \Sigma(t), t \in \mathcal{T}\}$ by carefully modifying the stochastic processes for the dictionary functions.

Note that the above decomposition of $\Gamma = \{\mu(t), \Sigma(t), t \in \mathcal{T}\}$ is not unique. Potentially we could constrain the loadings matrix to enforce identifiability (Geweke and Zhou, 1996), but this approach induces an undesirable order dependence among the responses (Aguilar and West, 2000; West, 2003; Lopes and West, 2004; Carvalho et al., 2008). Given our focus on estimation of Γ we follow Ghosh and Dunson (2009) in avoiding identifiability constraints, as such constraints are not necessary to ensure identifiability of the induced mean $\mu(t)$ and covariance $\Sigma(t)$. The characterization of the class of time-varying covariance matrices $\Sigma(t)$ is proved by Lemma 2.1 of Fox and Dunson (2011) which states that for K and L sufficiently large, any covariance regression can be decomposed as in (1b). Similar results are obtained for the mean process.

2.3 Prior Specification

We adopt a hierarchical prior specification approach to induce a prior P on $\Gamma = \{\mu(t), \Sigma(t), t \in \mathcal{T}\}$ with the goal of maintaining simple computation and allowing both covariances and means to evolve flexibly over continuous time. Specifically

- $\Gamma \mid \Theta, \Sigma_0, \Sigma_\xi, \Sigma_A, \Sigma_\psi, \Sigma_B \sim \text{LAF}_{L,K}(\Theta, \Sigma_0, \Sigma_\xi, \Sigma_A, \Sigma_\psi, \Sigma_B)$.

- Recalling the assumption $\xi_{lk}(t) \sim \text{nGP}(\sigma_{\xi_{lk}}^2, \sigma_{A_{lk}}^2)$ within LAF representation, we assume for each element $[\Sigma_{\xi}]_{lk}$ and $[\Sigma_A]_{lk}$ of the $L \times K$ matrices Σ_{ξ} and Σ_A respectively, the following priors

$$\begin{aligned}\sigma_{\xi_{lk}}^2 &\sim \text{InvGa}(a_{\xi}, b_{\xi}), \\ \sigma_{A_{lk}}^2 &\sim \text{InvGa}(a_A, b_A),\end{aligned}$$

independently for each (l, k) ; where $\text{InvGa}(a, b)$ denotes the Inverse Gamma distribution with shape a and scale b .

- Similarly, the variances $[\Sigma_{\psi}]_k = \sigma_{\psi_k}^2$ and $[\Sigma_B]_k = \sigma_{B_k}^2$ in the state equation representation of the nGP for each $\psi_k(t) \sim \text{nGP}(\sigma_{\psi_k}^2, \sigma_{B_k}^2)$ are assumed

$$\begin{aligned}\sigma_{\psi_k}^2 &\sim \text{InvGa}(a_{\psi}, b_{\psi}), \\ \sigma_{B_k}^2 &\sim \text{InvGa}(a_B, b_B),\end{aligned}$$

independently for each k .

- To address the issue related to the selection of the number of dictionary elements a shrinkage prior is proposed for Θ . In particular, following Bhattacharya and Dunson (2011) we assume

$$\begin{aligned}\theta_{jl} | \phi_{jl}, \tau_l &\sim \text{N}(0, \phi_{jl}^{-1} \tau_l^{-1}), \quad \phi_{jl} \sim \text{Ga}(3/2, 3/2), \\ \vartheta_1 &\sim \text{Ga}(a_1, 1), \quad \vartheta_h \sim \text{Ga}(a_2, 1), \quad h \geq 2, \quad \tau_l = \prod_{h=1}^l \vartheta_h.\end{aligned}\tag{8}$$

Note that if $a_2 > 1$ the expected value for ϑ_h is greater than 1. As a result, as l goes to infinity, τ_l tends to infinity, shrinking θ_{jl} towards zero. This leads to a flexible prior for θ_{jl} with a local shrinkage parameter ϕ_{jl} and a global column-wise shrinkage factor τ_l which allows many elements of Θ being close to zero as L increases. Our formulation can be easily generalized to allow shrinkage over K ; see Fox and Dunson (2011). However we found reasonable to fix K to relatively small values and learn L with the shrinkage approach, to avoid higher computational complexity in sampling the K -variate vector $\psi(t_i)$, $i = 1, \dots, T$.

- Finally for the variances of the error terms in vector ϵ_i , we assume the usual inverse gamma prior distribution. Specifically

$$\sigma_j^{-2} \sim \text{Ga}(a_{\sigma}, b_{\sigma})$$

independently for each $j = 1, \dots, p$.

3. Posterior Computation

For a fixed truncation level L^* and a latent factor dimension K^* , the algorithm for posterior computation alternates between a simple and efficient simulation smoother step (Durbin and Koopman, 2002) to update the state space formulation of the nGP in LAF prior, and standard Gibbs sampling steps for updating the parametric component parameters from their full conditional distributions. See Bhattacharya and Dunson (2011) for a method adaptively choosing the truncation levels.

3.1 Gibbs Sampling

We outline here the main features of the algorithm for posterior computation based on observations (y_i, t_i) for $i = 1, \dots, T$, while the complete algorithm is provided in Appendix A. Note that, since data are in practice observed at a finite number of times, the continuous time model is approximated in conducting inferences. This issue arises in analyzing data with any continuous time model.

- A. Given Θ and $\{\eta_i\}_{i=1}^T$, a multivariate version of the MCMC algorithm proposed by Zhu and Dunson (2013) draws posterior samples from each dictionary element's function $\{\xi_{lk}(t_i)\}_{i=1}^T$, its first order derivative $\{\xi'_{lk}(t_i)\}_{i=1}^T$, the corresponding instantaneous mean $\{A_{lk}(t_i)\}_{i=1}^T$, the variances in the state equations $\sigma_{\xi_{lk}}^2$, $\sigma_{A_{lk}}^2$ and the variances of the error terms in the observation equation σ_j^2 with $j = 1, \dots, p$.
- B. Given Θ , $\{\sigma_j^{-2}\}_{j=1}^p$, $\{y_i\}_{i=1}^T$ and $\{\xi(t_i)\}_{i=1}^T$ we implement a block sampling of $\{\psi_k(t_i)\}_{i=1}^T$, $\{\psi'_k(t_i)\}_{i=1}^T$, $\{B_k(t_i)\}_{i=1}^T$, $\sigma_{\psi_k}^2$, $\sigma_{B_k}^2$ and ν_i following a similar approach as in step A.
- C. Conditioned on $\{y_i\}_{i=1}^T$, $\{\eta_i\}_{i=1}^T$, $\{\sigma_j^{-2}\}_{j=1}^p$ and $\{\xi(t_i)\}_{i=1}^T$, and recalling the shrinkage prior for the elements of Θ in (8), we update Θ , each local shrinkage hyperparameter ϕ_{jl} and the global shrinkage hyperparameters τ_l following the standard conjugate analysis.
- D. Given the posterior samples from Θ , Σ_0 , $\{\xi(t_i)\}_{i=1}^T$ and $\{\psi(t_i)\}_{i=1}^T$ the realization of LAF process for $\{\mu(t_i), \Sigma(t_i), t_i \in \mathcal{T}\}$ conditioned on the data $\{y_i\}_{i=1}^T$ is

$$\begin{aligned} \mu(t_i) &= \Theta \xi(t_i) \psi(t_i), \\ \Sigma(t_i) &= \Theta \xi(t_i) \xi(t_i)^T \Theta^T + \Sigma_0. \end{aligned}$$

3.2 Hyperparameters Interpretation

We now focus our attention on the priors hyperparameters for $\sigma_{\xi_{lk}}^2$, $\sigma_{A_{lk}}^2$, $\sigma_{\psi_k}^2$ and $\sigma_{B_k}^2$. These quantities play an important role in facilitating local adaptivity and carefully tuning such values may improve mixing and convergence speed of our MCMC algorithm. Simulation studies have shown that the higher the variances in the latent state equations, the better our formulation accommodates locally adaptivity for sudden changes in Γ . A theoretical support for this data-driven consideration can be identified in the connection between the nGP and the nested smoothing splines. It has been shown by Zhu and Dunson (2013) that the posterior mean of the trajectory U with reference to the problem of nonparametric mean regression under the nGP prior can be related to the minimizer of the equation

$$\frac{1}{T} \sum_{i=1}^T (y_i - U(t_i))^2 + \lambda_U \int_{\mathcal{T}} (D^m U(t) - C(t))^2 dt + \lambda_C \int_{\mathcal{T}} (D^n C(t))^2 dt,$$

where C is the locally instantaneous function and $\lambda_U \in \mathfrak{R}^+$ and $\lambda_C \in \mathfrak{R}^+$ regulate the smoothness of the unknown functions U and C respectively, leading to less smoothed patterns when fixed at low values. The resulting inverse relationship between these smoothing parameters and the variances in the state equation, together with the results in the simulation studies, suggest to fix the hyperparameters in the Inverse Gamma prior for $\sigma_{\xi_{lk}}^2$, $\sigma_{A_{lk}}^2$,

$\sigma_{\psi_k}^2$ and $\sigma_{B_k}^2$ so as to allow high variances in the case in which the time series analyzed are expected to have strong changes in their covariance (or mean) dynamic. A further confirmation of the previous discussion is provided by the structure of the simulation smoother required to update the dictionary functions in our Gibbs sampling for posterior computation. More specifically, the larger the variances of $\{\omega_{i,\xi_{lk}}\}_{i=1}^T$, $\{\omega_{i,A_{lk}}\}_{i=1}^T$ and $\{\omega_{i,\psi_k}\}_{i=1}^T$, $\{\omega_{i,B_k}\}_{i=1}^T$ in the state equations, with respect to those of the vector of observations $\{y_i\}_{i=1}^T$, the higher is the weight associated to innovations in the filtering and smoothing techniques, allowing for less smoothed patterns both in the covariance and mean structures (see Durbin and Koopman, 2002).

In practical applications, it may be useful to obtain a first estimate of $\tilde{\Gamma} = \{\tilde{\mu}(t), \tilde{\Sigma}(t)\}$ to set the hyperparameters. More specifically, $\tilde{\mu}_j(t_i)$ can be the output of a standard moving average on each time series $y_j = (y_{j1}, \dots, y_{jT})^T$, while $\tilde{\Sigma}(t_i)$ can be obtained by a simple estimator, such as the EWMA procedure. With these choices, the recursive equation

$$\tilde{\Sigma}(t_i) = (1 - \lambda)\{[y_{i-1} - \tilde{\mu}(t_{i-1})][y_{i-1} - \tilde{\mu}(t_{i-1})]^T\} + \lambda\tilde{\Sigma}(t_{i-1}),$$

become easy to implement.

3.3 Online Updating

The problem of online updating represents a key point in multivariate time series with high frequency data. Referring to our formulation, we are interested in updating an approximated posterior for $\Gamma_{T+H} = \{\mu(t_{T+h}), \Sigma(t_{T+h}), h = 1, \dots, H\}$ once a new vector of observations $\{y_i\}_{i=T+1}^{T+H}$ is available, instead of rerunning posterior computation for the whole time series.

Using the posterior estimates of the Gibbs sampler based on observations available up to time T , it is easy to implement (see Appendix B) a highly computationally tractable online updating algorithm which alternates between steps A, B and D outlined in the previous section for the new set of observations, and that can be initialized at $T + 1$ using the one step ahead predictive distribution for the latent state vectors in the state space formulation. Such initialization procedure for latent state vectors in the algorithm depends on the sample moments of the posterior distribution for the latent states at T . As is known for Kalman smoothers (see, e.g., Durbin and Koopman, 2001), this could lead to computational problems in the online updating due to the larger conditional variances of the latent states at the end of the sample (i.e., at T). To overcome this problem, we replace the previous assumptions for the initial values with a data-driven initialization scheme. In particular, instead of using only the new observations for the online updating, we run the algorithm for $\{y_i\}_{i=T-k}^{T+H}$, with k small. As a result the distribution of the smoothed states at T is not anymore affected by the problem of large conditional variances leading to better online updating performance.

It is important to notice that the algorithm is not fully online in updating only the time-varying dictionary functions, while fixing the time-constant model parameters at their posterior mean. An alternative for properly propagating uncertainties while maintaining computational tractability may be to add a further step in the online updating procedure sampling the time-constant quantities conditionally on the updated dictionary functions and the quantities stored during the initial sampling. Such an approach may be reasonable if the initial time window considered is not enough large to ensure a consistent estimate

of the time-constant parameters and if the number of time series analyzed p is tractable. Since we search for a relatively fast procedure, and provided that for moderately large T the posterior for the time-stationary parameters rapidly becomes concentrated, we preferred our initially proposed algorithm in order to avoid the p draws from an L^* dimensional Gaussian in the sampling of Θ , which may slow down the online updating procedure for large p .

4. Simulation Studies

The aim of the following simulation studies is to compare the performance of our proposed LAF with respect to BCR, and to the models for multivariate stochastic volatility most widely used in practice, specifically: EWMA, PC-GARCH, GO-GARCH and DCC-GARCH. In order to assess whether and to what extent LAF can accommodate, in practice, even sharp changes in the time-varying means and covariances and to evaluate the costs of our flexible approach in settings where the mean and covariance functions do not require locally adaptive estimation techniques, we focus on two different sets of simulated data. The first is based on an underlying structure characterized by locally varying smoothness processes, while the second has means and covariances evolving in time through smooth processes. In the last subsection we also analyze the performance of the proposed online updating algorithm.

4.1 Simulated Data

- A. Locally varying smoothness processes: We generate a set of 5-dimensional observations y_i for each t_i in the discrete set $\mathcal{T}_o = \{1, 2, \dots, 100\}$, from the latent factor model in (5) with $\Lambda(t_i) = \Theta\xi(t_i)$. To allow sharp changes of means and covariances in the generating mechanism, we consider a 2×2 (i.e. $L = K = 2$) matrix $\{\xi(t_i)\}_{i=1}^{100}$ of time-varying functions adapted from Donoho and Johnstone (1994) with locally varying smoothness (more specifically we choose ‘bumps’ functions). The latent mean dictionary elements in $\{\psi(t_i)\}_{i=1}^{100}$ are simulated from a Gaussian process $\text{GP}(0, c)$ with length scale $\kappa = 10$, while the elements in matrix Θ can be obtained from the shrinkage prior in (8) with $a_1 = a_2 = 10$. Finally the elements of the diagonal matrix Σ_0^{-1} are sampled independently from $\text{Ga}(1, 0.1)$.
- B. Smooth processes: We consider the same data set of 10-dimensional observations y_i with $t_i \in \mathcal{T}_o = \{1, 2, \dots, 100\}$ investigated in Fox and Dunson (2011, Section 4.1). The settings are similar to the previous with exception of $\{\xi(t_i)\}_{i=1}^{100}$ which are 5×4 matrices of smooth GP dictionary functions with length scale $\kappa = 10$.

4.2 Estimation Performance

- A. Locally varying smoothness processes:

Posterior computation for LAF is performed by using truncation levels $L^* = K^* = 2$ (at higher level settings we found that the shrinkage prior on Θ results in posterior samples of the elements in the additional columns concentrated around 0). We place a $\text{Ga}(1, 0.1)$ prior on the precision parameters σ_j^{-2} and choose $a_1 = a_2 = 2$. As regards the nGP prior for each dictionary element $\xi_{lk}(t)$ with $l = 1, \dots, L^*$ and $k = 1, \dots, K^*$, we choose diffuse but proper priors for the initial values by setting $\sigma_{\mu_{lk}}^2 = \sigma_{\alpha_{lk}}^2 = 100$

and place an $\text{InvGa}(2, 10^8)$ prior on each $\sigma_{\xi_{lk}}^2$ and $\sigma_{A_{lk}}^2$ in order to allow less smooth behavior according to a previous graphical analysis of $\tilde{\Sigma}(t_i)$ estimated via EWMA. Similarly we set $\sigma_{\mu_k}^2 = \sigma_{\alpha_k}^2 = 100$ in the prior for the initial values of the latent state equations resulting from the nGP prior for $\psi_k(t)$, and consider $a_\psi = a_B = b_\psi = b_B = 0.005$ to balance the rough behavior induced on the nonparametric mean functions by the settings of the nGP prior on $\xi_{lk}(t)$, as suggested from previous graphical analysis. Note also that for posterior computation, we first scale the predictor space to $(0, 1]$, leading to $\delta_i = 1/100$, for $i = 1, \dots, 100$.

For inference in BCR we consider the same previous hyperparameters setting for Θ and Σ_0 priors as well as the same truncation levels K^* and L^* , while the length scale κ in GP prior for $\xi_{lk}(t)$ and $\psi_k(t)$ has been set to 10 using the data-driven heuristic outlined in Fox and Dunson (2011). In both cases we run 50,000 Gibbs iterations discarding the first 20,000 as burn-in and thinning the chain every 5 samples.

As regards the other approaches, EWMA has been implemented by choosing the smoothing parameter λ that minimizes the mean squared error (MSE) between the estimated covariances and the true values. PC-GARCH algorithm follows the steps provided by Burns (2005) with GARCH(1,1) assumed for the conditional volatilities of each single time series and the principal components. GO-GARCH and DCC-GARCH recall the formulations provided by van der Weide (2002) and Engle (2002) respectively, assuming a GARCH(1,1) for the conditional variances of the processes analyzed, which proves to be a correct choice in many financial applications and also in our setting. Note that, differently from LAF and BCR, the previous approaches do not model explicitly the mean process $\{\mu(t_i)\}_{i=1}^{100}$ but work directly on the innovations $\{y_i - \mu(t_i)\}_{i=1}^{100}$. Therefore in these cases we first model the conditional mean via smoothing spline and in a second step we estimate the models working on the innovations. The smoothing parameter for spline estimation has been set to 0.7, which was found to be appropriate to best reproduce the true dynamic of $\{\mu(t_i)\}_{i=1}^{100}$.

B. Smooth processes:

We mainly keep the same setting of the previous simulation study with few differences. Specifically, L^* and K^* has been fixed to 5 and 4 respectively (also in this case the choice of the truncation levels proves to be appropriate, reproducing the same results provided in the simulation study of Fox and Dunson (2011) where $L^* = 10$ and $K^* = 10$). Moreover the scale parameters in the Inverse Gamma prior on each $\sigma_{\xi_{lk}}^2$ and $\sigma_{A_{lk}}^2$ has been set to 10^4 in order to allow a smoother behavior according to a previous graphical analysis of $\tilde{\Sigma}(t_i)$ estimated via EWMA, but without forcing the nGP prior to be the same as a GP prior. Following Fox and Dunson (2011) we run 10,000 Gibbs iterations which proved to be enough to reach convergence, and discarded the first 5,000 as burn-in.

In the first set of simulated data, we analyzed mixing by the Gelman-Rubin procedure (see, e.g., Gelman and Rubin, 1992), based on potential scale reduction factors computed for each chain by splitting the sampled quantities in 6 pieces of same length. The analysis shows slower mixing for BCR compared with LAF. Specifically, in LAF 95% of the chains have a potential reduction factor lower than 1.35, with a median equal to 1.11, while in

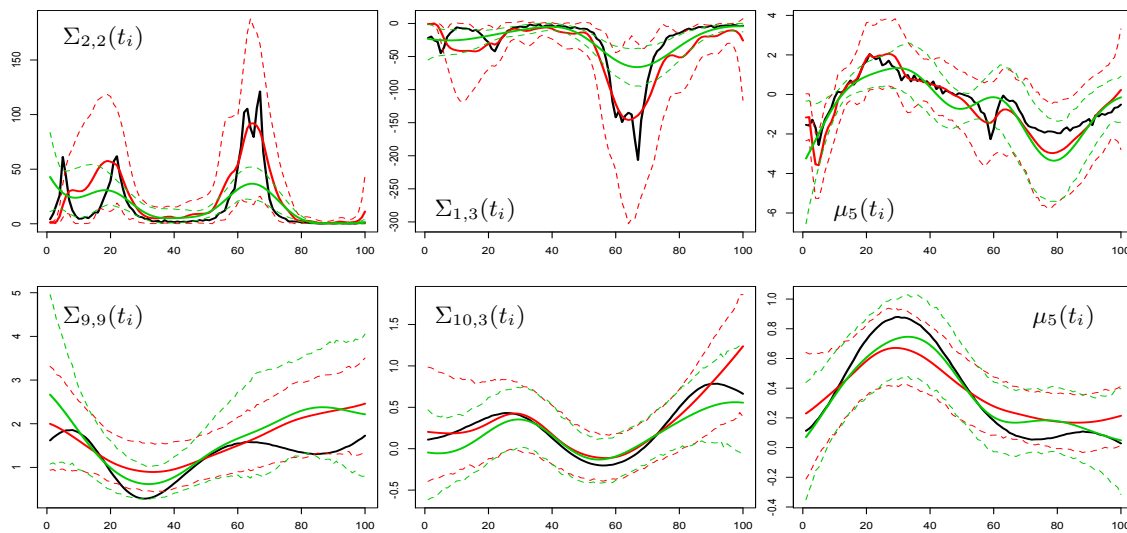


Figure 2: For locally varying smoothness simulation (top) and smooth simulation (bottom), plots of truth (black) and posterior mean respectively of LAF (solid red line) and BCR (solid green line) for selected components of the variance (left), covariance (middle), mean (right). For both approaches the dotted lines represent the 95% highest posterior density intervals.

BCR the 95% quantile is 1.44 and the median equals 1.18. Less problematic is the mixing for the second set of simulated data, with potential scale reduction factors having median equal to 1.05 for both approaches and 95% quantiles equal to 1.15 and 1.31 for LAF and BCR, respectively.

Figure 2 compares, in both simulated samples, true and posterior mean of the process $\Gamma = \{\mu(t_i), \Sigma(t_i), i = 1, \dots, 100\}$ over the predictor space \mathcal{T}_o together with the point-wise 95% highest posterior density (hpd) intervals for LAF and BCR. From the upper plots we can clearly note that our approach is able to capture conditional heteroscedasticity as well as mean patterns, also in correspondence of sharp changes in the time-varying true functions. The major differences compared to the true values can be found at the beginning and at the end of the series and are likely to be related to the structure of the simulation smoother which also causes a widening of the credibility bands at the very end of the series; for references regarding this issue see Durbin and Koopman (2001). However, even in the most problematic cases, the true values are within the bands of the 95% hpd intervals. Much more problematic is the behavior of the posterior distributions for BCR which over-smooth both covariance and mean functions leading also to many 95% hpd intervals not containing the true values. Bottom plots in Figure 2 show that the performance of our approach is very close to that of BCR, when data are simulated from a model where the covariances and means evolve smoothly across time and local adaptivity is not required. This happens even if the hyperparameters in LAF are set in order to maintain separation between nGP and GP prior, suggesting large support property for the proposed approach.

	Mean	90% Quantile	95% Quantile	Max
Covariance $\{\Sigma(t_i)\}$				
EWMA	1.37	2.28	5.49	85.86
PC-GARCH	1.75	2.49	6.48	229.50
GO-GARCH	2.40	3.66	10.32	173.41
DCC-GARCH	1.75	2.21	6.95	226.47
BCR	1.80	2.25	7.32	142.26
LAF	0.90	1.99	4.52	36.95
Mean $\{\mu(t_i)\}$				
SPLINE	0.064	0.128	0.186	2.595
BCR	0.087	0.185	0.379	2.845
LAF	0.062	0.123	0.224	2.529

Table 1: LOCALLY VARYING SMOOTHNESS PROCESSES: Summaries of the standardized squared errors between true values $\{\mu(t_i)\}_{i=1}^{100}$ and $\{\Sigma(t_i)\}_{i=1}^{100}$ and estimated quantities $\{\hat{\Sigma}(t_i)\}_{i=1}^{100}$ and $\{\hat{\mu}(t_i)\}_{i=1}^{100}$ computed with different approaches.

	Mean	90% Quantile	95% Quantile	Max
Covariance $\{\Sigma(t_i)\}$				
EWMA	0.030	0.081	0.133	1.119
PC-GARCH	0.018	0.048	0.076	0.652
GO-GARCH	0.043	0.104	0.202	1.192
DCC-GARCH	0.022	0.057	0.110	0.466
BCR	0.009	0.019	0.039	0.311
LAF	0.009	0.022	0.044	0.474
Mean $\{\mu(t_i)\}$				
SPLINE	0.007	0.019	0.027	0.077
BCR	0.005	0.015	0.024	0.038
LAF	0.005	0.017	0.026	0.050

Table 2: SMOOTH PROCESSES: Summaries of the standardized squared errors between true values $\{\mu(t_i)\}_{i=1}^{100}$ and $\{\Sigma(t_i)\}_{i=1}^{100}$ and estimated quantities $\{\hat{\Sigma}(t_i)\}_{i=1}^{100}$ and $\{\hat{\mu}(t_i)\}_{i=1}^{100}$ computed with different approaches.

The comparison of the summaries of the squared errors between true process $\Gamma = \{\mu(t_i), \Sigma(t_i), i = 1, \dots, 100\}$ and the estimated quantities $\hat{\Gamma} = \{\hat{\mu}(t_i), \hat{\Sigma}(t_i), i = 1, \dots, 100\}$ standardized with the range of the true processes $r_\mu = \max_{i,j} \{\mu_j(t_i)\} - \min_{i,j} \{\mu_j(t_i)\}$ and $r_\Sigma = \max_{i,j,k} \{\Sigma_{j,k}(t_i)\} - \min_{i,j,k} \{\Sigma_{j,k}(t_i)\}$ respectively, once again confirms the overall better performance of our approach relative to all the considered competitors. Table 1 shows that, when local adaptivity is required, LAF provides a superior performance having standardized residuals lower than those of the other approaches. EWMA seems to provide quite accurate estimates, but it is important to underline that we choose the optimal smoothing parameter λ in order to minimize the MSE between estimated and true parameters, which are clearly not known in practical applications. Different values of λ reduces significantly the performance of EWMA, which shows also lack of robustness. The closeness of the summaries of LAF and BCR in Table 2 confirms the flexibility of LAF even in settings where local adaptivity is not required and highlights the better performance of the two approaches with respect to the other competitors also when smooth processes are investigated.

To better understand the improvement of our approach in allowing locally varying smoothness and to evaluate the consequences of the over-smoothing induced by BCR on the

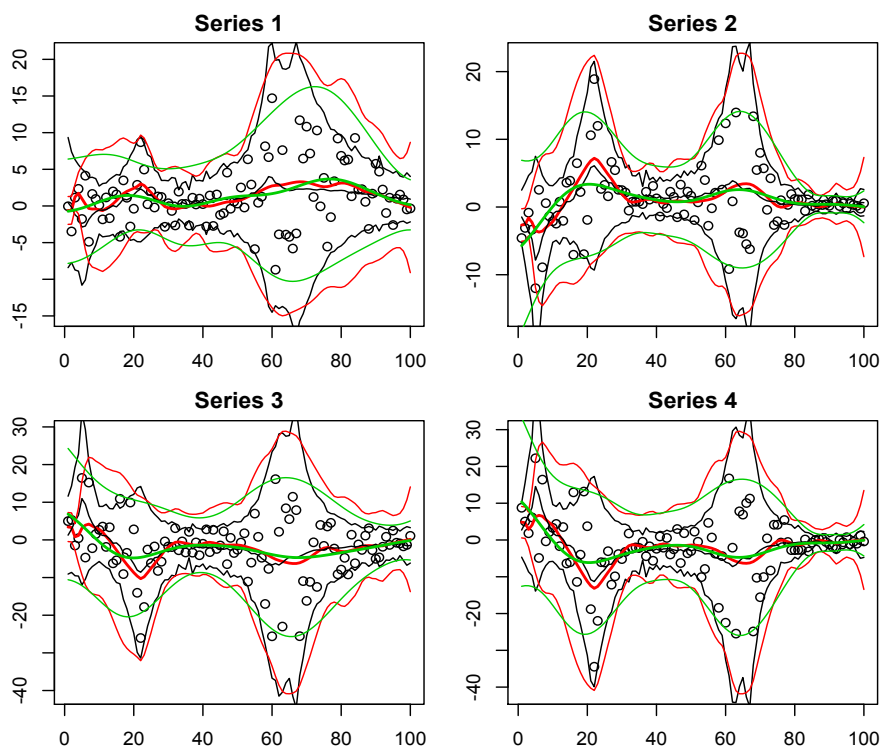


Figure 3: For 4 selected simulated series: time-varying mean $\mu_j(t_i)$ and 2.5% and 97.5% quantiles of the marginal distribution of y_{ji} with true mean and variance (black), mean and variance from posterior mean of LAF (red), mean and variance from posterior mean of BCR (green). Black points represent the simulated data.

distribution of y_i with $i = 1, \dots, 100$ consider Figure 3 which shows, for some selected series $\{y_{ji}\}_{i=1}^{100}$ in the first simulated data set, the time-varying mean together with the point-wise 2.5% and 97.5% quantiles of the marginal distribution of y_{ji} induced respectively by the true mean and true variance, the posterior mean of $\mu_j(t_i)$ and $\Sigma_{jj}(t_i)$ from our proposed approach and the posterior mean of the same quantities from BCR. We can clearly see that the marginal distribution of y_{ji} induced by BCR is over-concentrated near the mean, leading to incorrect inferences. Note that our proposal is also able to accommodate heavy tails, a typical characteristic in financial series.

4.3 Online Updating Performance

To analyze the performance of the online updating algorithm in LAF model, we simulate 50 new observations $\{y_i\}_{i=101}^{150}$ with $t_i \in \mathcal{T}_o^* = \{101, \dots, 150\}$, considering the same Θ and Σ_0 used in the generating mechanism for the first simulated data set and taking the 50 subsequent observations of the bumps functions for the dictionary elements $\{\xi(t_i)\}_{i=101}^{150}$; finally the additional latent mean dictionary elements $\{\psi(t_i)\}_{i=101}^{150}$ are simulated as before maintaining the continuity with the previously simulated functions $\{\psi(t_i)\}_{i=1}^{100}$. According to

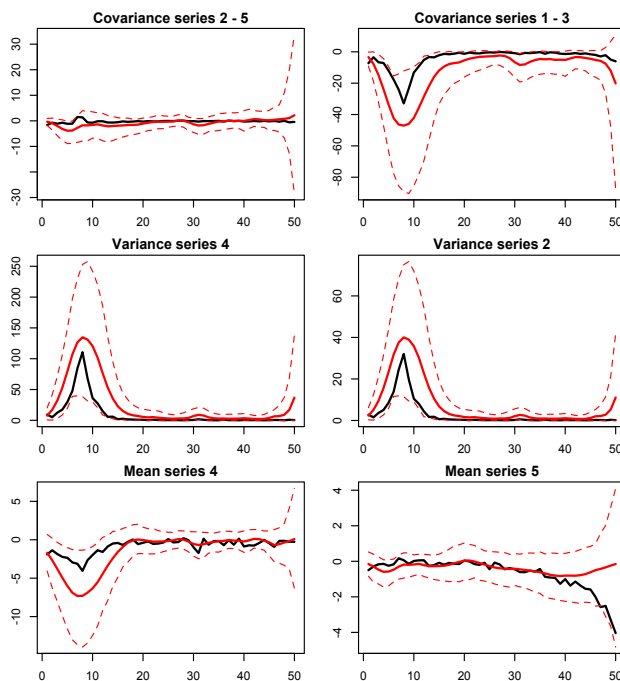


Figure 4: Plots of truth (black) and posterior mean of the online updating procedure (solid red line) for selected components of the covariance (top), variance (middle), mean (bottom). The dotted lines represent the 95% highest posterior density intervals.

the algorithm described in Subsection 3.3, we fix Θ , Σ_0 , Σ_ξ , Σ_A, Σ_ψ and Σ_B at their posterior mean from the previous Gibbs sampler and consider the last three observations y_{98} , y_{99} and y_{100} (i.e. $k = 3$) to initialize the simulation smoother in $i = 101$ through the proposed data-driven initialization approach. Posterior computation shows good performance in terms of mixing, and convergence is assessed after 5,000 Gibbs iterations with a small burn-in of 500.

Figure 4 compares true mean and covariance to posterior mean of a selected set of components of $\Gamma_* = \{\mu(t_i), \Sigma(t_i), i = 101, \dots, 150\}$ including also the 95% hpd intervals. The results clearly show that the online updating is characterized by a good performance which allows to capture the behavior of new observations conditioning on the previous estimates. Note that the posterior distribution of the approximated mean and covariance functions tends to slightly over-estimate the patterns of the functions at sharp changes, however also in these cases the true values are within the bands of the credibility intervals. Finally note that the data-driven initialization ensures a good behavior at the beginning of the series, while the results at the end have wider uncertainty bands as expected.

5. Application Study

Spurred by the recent growth of interest in the dynamic dependence structure between financial markets in different countries, and in its features during the crises that have

followed in recent years, we applied our LAF to the multivariate time series of the main National Stock Market Indices.

5.1 National Stock Market Indices, Introduction and Motivation

National Stock Market Indices represent technical tools that allow, through the synthesis of numerous data on the evolution of the various stocks, to detect underlying trends in the financial market, with reference to a specific basis of currency and time. More specifically, each Market Index can be defined as a weighted sum of the values of a set of national stocks, whose weighting factors is equal to the ratio of its market capitalization in a specific date and overall of the whole set on the same date.

In this application we focus our attention on the multivariate weekly time series of the main 33 (i.e. $p = 33$) National Stock Market Indices from 12/07/2004 to 25/06/2012. Figure 5 shows the main features in terms of stationarity, mean patterns and volatility of two selected National Stock Market Indices downloaded from <http://finance.yahoo.com/>. The non-stationary behavior, together with the different bases of currency and time, motivate the use of logarithmic returns $y_{ji} = \log(I_{ji}/I_{ji-1})$, where I_{ji} is the value of the Stock Market Index j at time t_i . Beside this, the marginal distribution of log returns shows heavy tails and irregular cyclical trends in the nonparametric estimation of the mean, while EWMA estimates highlight rapid changes of volatility during the financial crises observed in the recent years. All these results, together with large p settings and high frequency data typical in financial fields, motivate the use of our approach to obtain a better characterization of the time-varying dependence structure among financial markets.

5.2 LAF for National Stock Market Indices

We consider the heteroscedastic model $y_i \sim N_{33}(\mu(t_i), \Sigma(t_i))$ for $i = 1, \dots, 415$ and t_i in the discrete set $\mathcal{T}_o = \{1, 2, \dots, 415\}$, where the elements of $\Gamma = \{\mu(t_i), \Sigma(t_i), i = 1, \dots, 415\}$, defined by (7a)-(7b), are induced by the dynamic latent factor model outlined in (5)-(6).

Posterior computation is performed by first rescaling the predictor space \mathcal{T}_o to $(0, 1]$ and using the same setting of the first simulation study, with the exception of the truncation levels fixed at $K^* = 4$ and $L^* = 5$ (which we found to be sufficiently large from the fact that the last few columns of the posterior samples for Θ assumed values close to 0) and the hyperparameters of the nGP prior for each $\xi_{lk}(t)$ and $\psi_k(t)$ with $l = 1, \dots, L^*$ and $k = 1, \dots, K^*$, set to $a_\xi = a_A = a_\psi = a_B = 2$ and $b_\xi = b_A = b_\psi = b_B = 5 \times 10^7$ to capture also rapid changes in the mean functions according to Figure 5. Missing values in our data set do not represent a limitation since the Bayesian approach allows us to update our posterior considering solely the observed data. We run 10,000 Gibbs iterations with a burn-in of 2,500. Examination of trace plots of the posterior samples for $\Gamma = \{\mu(t_i), \Sigma(t_i), i = 1, \dots, 415\}$ showed no evidence against convergence.

Posterior distributions for the variances in Figure 6 demonstrate that we are clearly able to capture the rapid changes in the dynamics of volatility that occur during the world financial crisis of 2008, in early 2010 with the Greek debt crisis and in the summer of 2011 with the financial speculation in government bonds of European countries together with the rejection of the U.S. budget and the downgrading of the United States rating. Moreover,

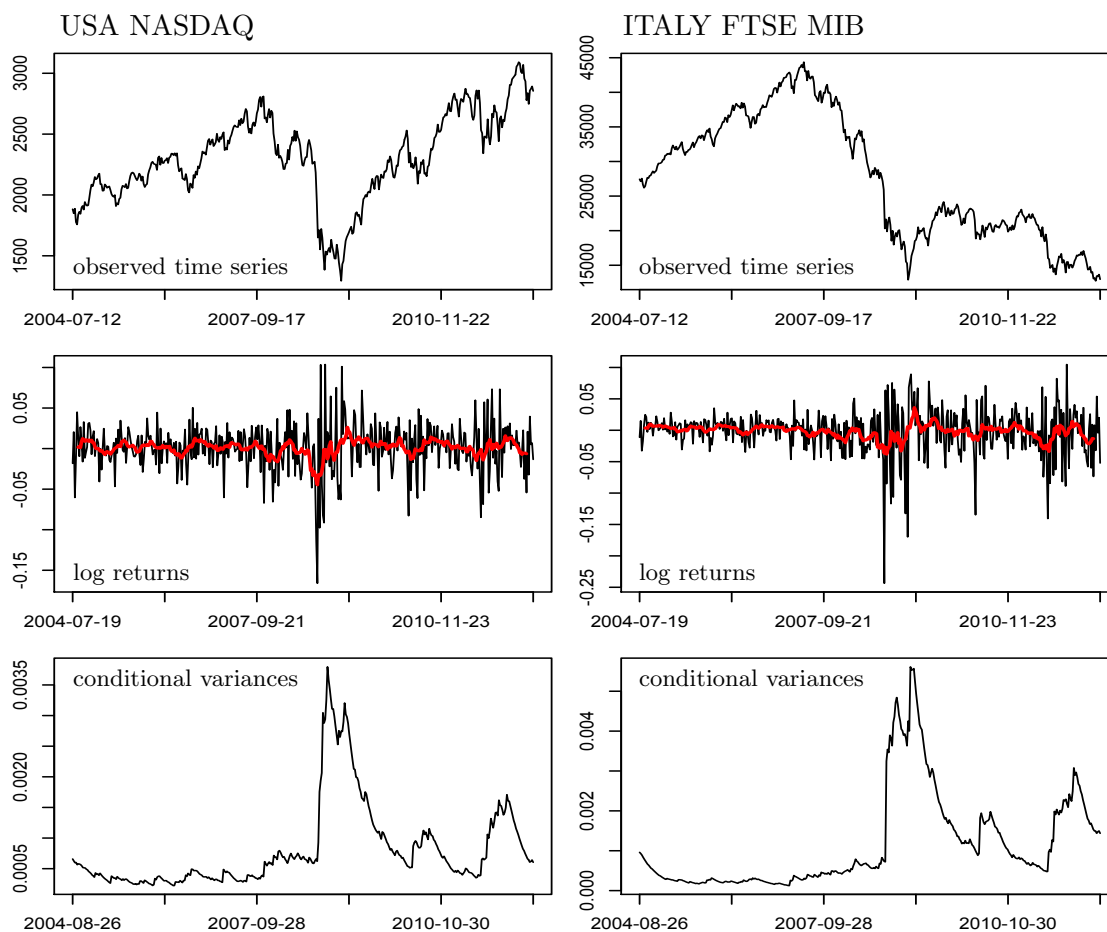


Figure 5: Plots of the main features of USA NASDAQ (left) and ITALY FTSE MIB (right). Specifically: observed time series (top), log returns series with nonparametric mean estimation via 12 week Equally Weighted Moving Average (red) in the middle, EWMA volatility estimates (bottom).

the resulting marginal distribution of the log returns induced by the posterior mean of $\mu_j(t)$ and $\Sigma_{jj}(t)$, shows that we are also able to accommodate heavy tails as well as mean patterns cycling irregularly between slow and more rapid changes.

Important information about the ability of our model to capture the evolution of world geo-economic structure during different finance scenarios is provided in Figures 7 and 8. From the correlations between NASDAQ and the other National Stock Market Indices (based on the posterior mean $\{\hat{\Sigma}(t_i)\}_{i=1}^{415}$ of the covariances function) in Figure 7, we can immediately notice the presence of a clear geo-economic structure in world financial markets (more evident in LAF than in BCR), where the dependence between the U.S. and European countries is systematically higher than that of South East Asian Nations (Economic Tigers), showing also different reactions to crises. Plots at the top of the Figure 8 confirms the above considerations showing how Western countries exhibit more connection with countries closer

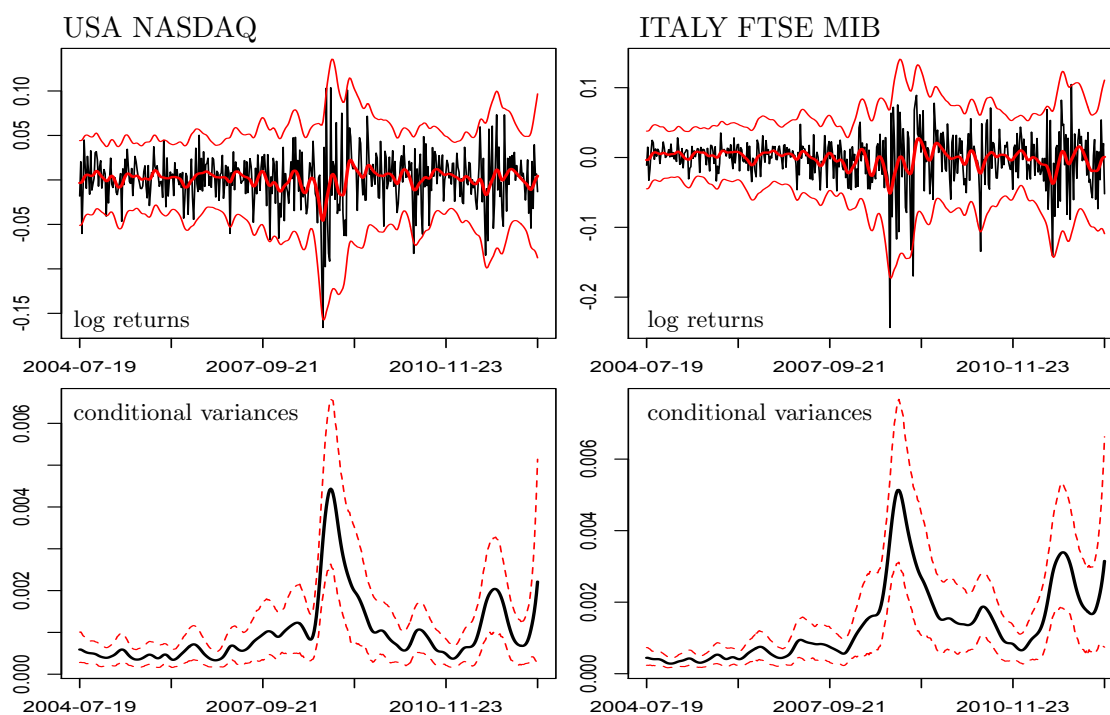


Figure 6: Top: Plot for 2 National Stock Market Indices, respectively USA NASDAQ (left) and ITALY FTSE MIB (right), of the log returns (black) and the time-varying estimated mean $\{\hat{\mu}_j(t_i)\}_{i=1}^{415}$ together with the time-varying 2.5% and 97.5% quantiles (red) of the marginal distribution of y_{ji} from LAF. Bottom: posterior mean (black) and 95% hpd (dotted red) for the variances $\{\Sigma_{jj}(t_i)\}_{i=1}^{415}$.

in terms of geographical, political and economic structure; the same holds for Eastern countries where we observe a reversal of the colored curves. As expected, Russia is placed in a middle path between the two blocks. A further element that our model captures about the structure of the markets is shown in the plots at the bottom of Figure 8. The time-varying regression coefficients obtained from the standard formulas of the conditional normal distribution based on the posterior mean of $\Gamma = \{\mu(t_i), \Sigma(t_i), i = 1, \dots, 415\}$ highlight clearly the increasing dependence of European countries with higher crisis in sovereign debt and Germany, which plays a central role in Eurozone as expected.

The flexibility of the proposed approach and the possibility of accommodating varying smoothness in the trajectories over time, allow us to obtain a good characterization of the dynamic dependence structure according with the major theories on financial crisis. The top plot in Figure 7 shows how the change of regime in correlations occurs exactly in correspondence to the burst of the U.S. housing bubble (A), in the second half of 2006. Moreover we can immediately notice that the correlations among financial markets increase significantly during the crises, showing a clear international financial contagion effect in agreement with other theories on financial crisis (see, e.g., Baig and Goldfajn, 1999; Claessens and Forbes, 2001). As expected the persistence of high levels of correlation is evident during the global

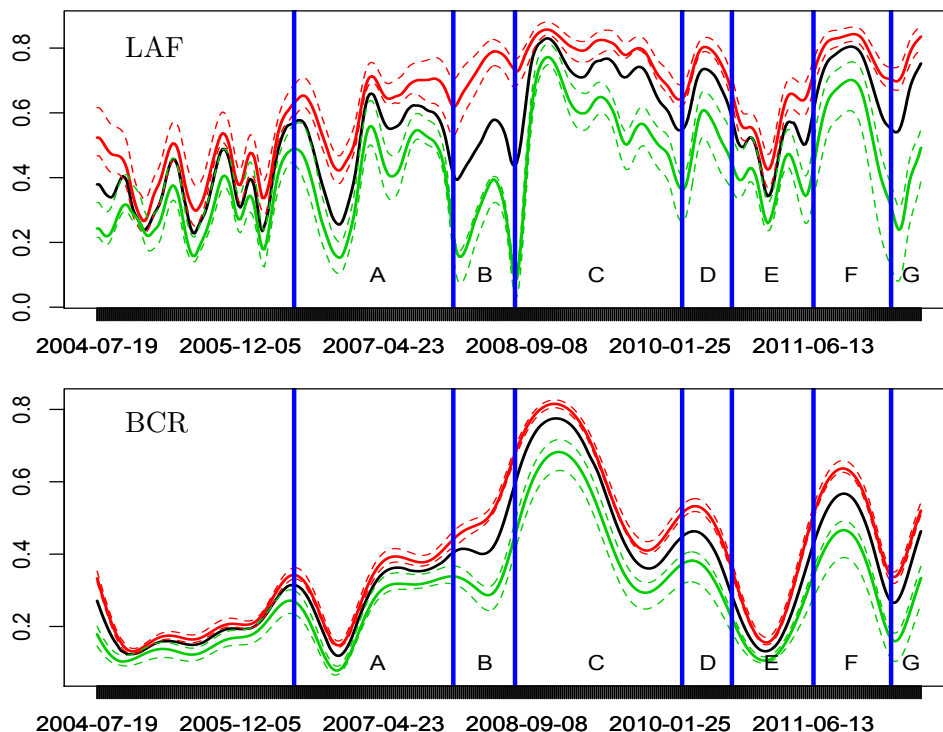


Figure 7: Black line: For USA NASDAQ median of correlations with the other 32 National Stock Market Indices based on posterior mean of $\{\Sigma(t_i)\}_{i=1}^{415}$. Red lines: 25%, 75% (dotted lines) and 50% (solid line) quantiles of correlations between USA NASDAQ and European countries (without considering Greece and Russia which present a specific pattern). Green lines: 25%, 75% (dotted lines) and 50% (solid line) quantiles of correlations between USA NASDAQ and the countries of Southeast Asia (Asian Tigers and India). Timeline: (A) burst of U.S. housing bubble; (B) risk of failure of the first U.S. credit agencies (Bear Stearns, Fannie Mae and Freddie Mac); (C) world financial crisis after the Lehman Brothers' bankruptcy; (D) Greek debt crisis; (E) financial reform launched by Barack Obama and E.U. efforts to save Greece (the two peaks represent respectively Irish debt crisis and Portugal debt crisis); (F) worsening of European sovereign-debt crisis and the rejection of the U.S. budget; (G) crisis of credit institutions in Spain and the growing financial instability of the Eurozone.

financial crisis between late-2008 and end-2009 (C), at the beginning of which our approach also captures a sharp variation in the correlations between the U.S. and Economic Tigers, which lead to levels close to those of Europe. Further rapid changes are identified in correspondence of Greek crisis (D), the worsening of European sovereign-debt crisis and the rejection of the U.S. budget (F) and the recent crisis of credit institutions in Spain together with the growing financial instability Eurozone (G). Finally, even in the period of U.S. financial reform launched by Barack Obama and E.U. efforts to save Greece (E), we

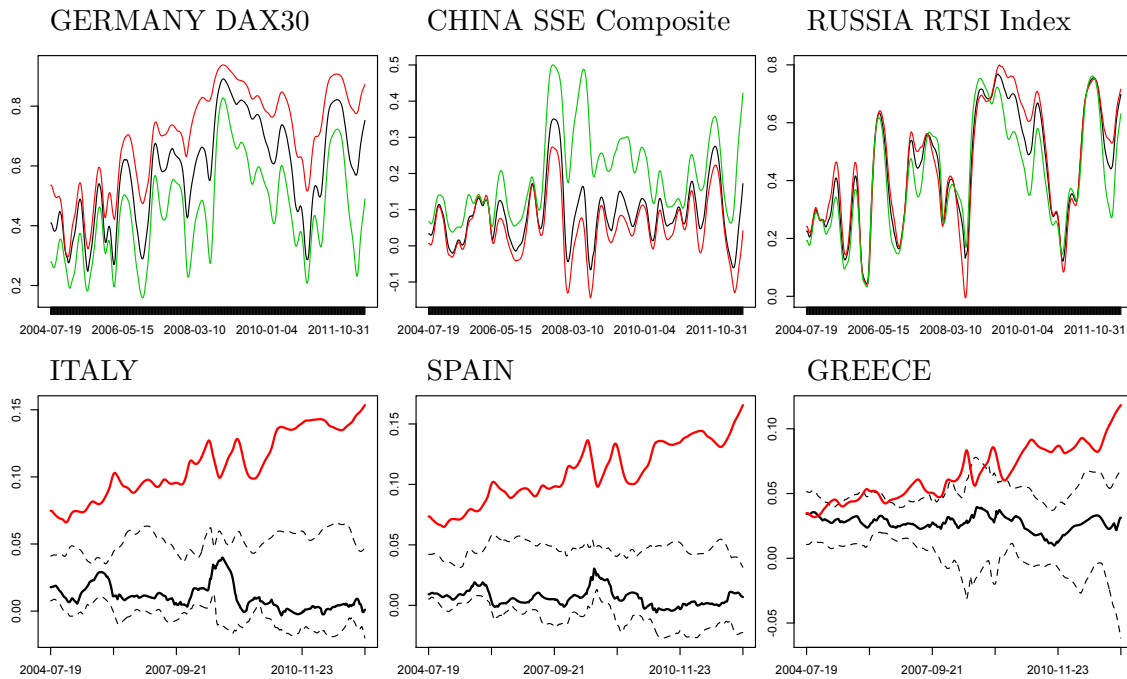


Figure 8: Top: For 3 selected National Stock Market Indices, plot of the median of the correlation based on posterior mean of $\{\Sigma(t_i)\}_{i=1}^{415}$ with the other 32 world stock indices (black), the European countries without considering Greece and Russia (red) and the Asian Tigers including India (green). Bottom: For 3 of the European countries more subject to sovereign debt crisis, plot of 25%, 50% and 75% quantiles of the time-varying regression parameters based on posterior mean $\{\hat{\Sigma}(t_i)\}_{i=1}^{415}$ with the other countries (black) and Germany (red).

can notice two peaks representing respectively Irish debt crisis and Portugal debt crisis. Note also that BCR, as expected, tends to over-smooth the dynamic dependence structure during the financial crisis, proving to be not able to model the sharp change in the correlations between USA NASDAQ and Economic Tigers during late-2008, and the two peaks representing respectively Irish and Portugal debt crisis at the beginning of 2011.

5.3 National Stock Market Indices, Updating and Predicting

The possibility to quickly update the estimates and the predictions as soon as new data arrive, represents a crucial aspect to obtain quantitative informations about the future scenarios of the crisis in financial markets. To answer this goal, we apply the online updating algorithm presented in Subsection 3.3, to the new set of weekly observations $\{y_i\}_{i=416}^{422}$ from 02/07/2012 to 13/08/2012 conditioning on posterior estimates of the Gibbs sampler based on observations $\{y_i\}_{i=1}^{415}$ available up to 25/06/2012. We initialized the simulation smoother algorithm with the last 8 observations of the previous sample.

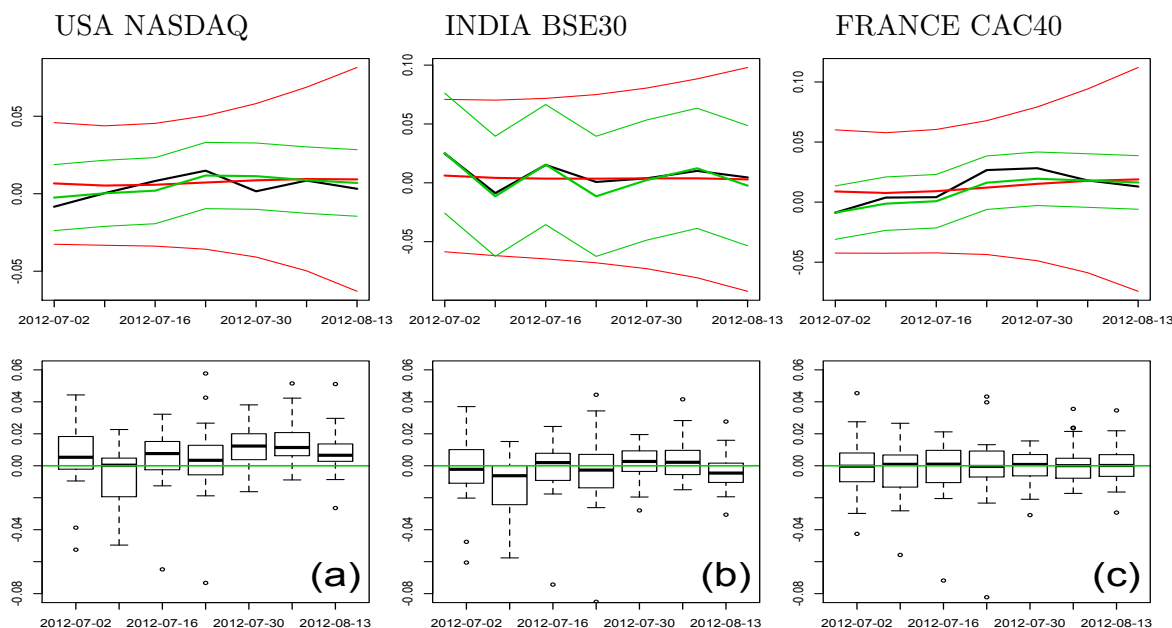


Figure 9: Top: For 3 selected National Stock Market Indices, respectively USA NASDAQ (left), INDIA BSE30 (middle) and FRANCE CAC40 (right), plot of the observed log returns (black) together with the mean and the 2.5% and 97.5% quantiles of the marginal distribution (red) and conditional distribution given the other 32 National Stock Market Indices (green) based on the posterior mean of $\Gamma_* = \{\mu(t_i), \Sigma(t_i), i = 416, \dots, 422\}$ from the online updating procedure for the new observations from 02/07/2012 to 13/08/2012. Bottom: boxplots of the one step ahead prediction errors for the 33 National Stock Market Indices, where the predicted values are respectively: (a) unconditional mean $\{\tilde{y}_{i+1}\}_{i=415}^{421} = 0$, (b) marginal mean of the one step ahead predictive distribution, (c) conditional mean given the log returns of the other 32 NSI at $i + 1$ of the one step ahead predictive distribution. Predictions for (b) and (c) are induced by the posterior mean of $\{\mu(t_{i+1|i}), \Sigma(t_{i+1|i}), i = 415, \dots, 421\}$ of LAF.

Plots at the top of Figure 9 show, for 3 selected National Stock Market Indices, the new observed log returns $\{y_{ji}\}_{i=416}^{422}$ (black) together with the mean and the 2.5% and 97.5% quantiles of the marginal distribution (red) and conditional distribution (green) of $y_{ji}|y_i^{-j}$ with $y_i^{-j} = \{y_{qi}, q \neq j\}$. We use standard formulas of the multivariate normal distribution based on the posterior mean of the updated $\Gamma_* = \{\mu(t_i), \Sigma(t_i), i = 416, \dots, 422\}$ after 5,000 Gibbs iterations with a burn in of 500. This is sufficient for convergence based on examining trace plots of the time-varying mean and covariance matrices. From these results, we can clearly notice the good performance of our proposed online updating algorithm in obtaining a characterization for the distribution of new observations. Also note that the multivariate approach together with a flexible model for the mean and covariance, allow for significant improvements when the conditional distribution of an index given the others is analyzed.

To obtain further informations about the predictive performance of our LAF, we can easily use our online updating algorithm to obtain h step-ahead predictions for $\Gamma_{T+h|T} = \{\mu(t_{T+h|T}), \Sigma(t_{T+h|T}), h = 1, \dots, H\}$. In particular, referring to Durbin and Koopman (2001), we can generate posterior samples of $\Gamma_{T+h|T}$ merely by treating $\{y_i\}_{i=T+1}^{T+H}$ as missing values in the proposed online updating algorithm. Here, we consider the one step ahead prediction (i.e. $H = 1$) problem for the new observations. More specifically, for each i from 415 to 421, we update the mean and covariance functions conditioning on informations up to t_i through the online algorithm and then obtain the predicted posterior distribution for $\Sigma(t_{i+1|i})$ and $\mu(t_{i+1|i})$ by adding to the sample considered for the online updating a last column y_{i+1} of missing values.

Plots at the bottom of Figure 9, show the boxplots of the one step ahead prediction errors for the 33 National Stock Market Indices obtained as the difference between the predicted value $\tilde{y}_{j,i+1}$ and, once available, the observed log return $y_{j,i+1}$ with $i + 1 = 416, \dots, 422$ corresponding to weeks from 02/07/2012 to 13/08/2012. In (a) we forecast the future log returns with the unconditional mean $\{\tilde{y}_{i+1}\}_{i=415}^{421} = 0$, which is what is often done in practice under the general assumption of zero mean, stationary log returns. In (b) we consider $\tilde{y}_{i+1} = \hat{\mu}(t_{i+1|i})$, the posterior mean of the one step ahead predictive distribution of $\mu(t_{i+1|i})$, obtained from the previous proposed approach after 5,000 Gibbs iteration with a burn in of 500. Finally in (c) we suppose that the log returns of all National Stock Market Indices except that of country j (i.e., $y_{j,i+1}$) become available at t_{i+1} and, considering $y_{i+1|i} \sim N_p(\hat{\mu}(t_{i+1|i}), \hat{\Sigma}(t_{i+1|i}))$ with $\hat{\mu}(t_{i+1|i})$ and $\hat{\Sigma}(t_{i+1|i})$ posterior mean of the one step ahead predictive distribution respectively for $\mu(t_{i+1|i})$ and $\Sigma(t_{i+1|i})$, we forecast $y_{j,i+1}$ with the conditional mean of $y_{j,i+1|i}$ given the other log returns at time t_{i+1} . Comparing boxplots in (a) with those in (b) we can see that our model allows to obtain improvements also in terms of prediction. Furthermore, by analyzing the boxplots in (c) we can notice how our ability to obtain a good characterization of the time-varying covariance structure can play a crucial role also in improving forecasting, since it enters into the standard formula for calculating the conditional mean in the normal distribution.

6. Discussion

In this paper, we have presented a continuous time multivariate stochastic process for time series to obtain a better characterization for mean and covariance temporal dynamics. Maintaining simple conjugate posterior updates and tractable computations in moderately large p settings, our model increases significantly the flexibility of previous approaches as it captures sharp changes both in mean and covariance dynamics while accommodating heavy tails. Beside these key advantages, the state space formulation enables development of a fast online updating algorithm particularly useful for high frequency data.

The simulation studies highlight the flexibility and the overall better performance of LAF with respect to the models for multivariate stochastic volatility most widely used in practice, both when adaptive estimation techniques are required, and also when the underlying mean and covariance structures do not show sharp changes in their dynamic.

The application to the problem of capturing temporal and geo-economic structure between the main financial markets demonstrates the utility of our approach and the improvements that can be obtained in the analysis of multivariate financial time series with

reference to (i) heavy tails, (ii) locally adaptive mean regression, (iii) sharp changes in covariance functions, (iii) high dimensional data set, (iv) online updating with high frequency data (v) missing values and (vi) predictions. Potentially further improvements are possible using a stochastic differential equation model that explicitly incorporates prior information on dynamics.

Acknowledgments

This research was partially supported by grant R01ES17240 from the National Institute of Environmental Health Sciences (NIEHS) of the National Institutes of Health (NIH) and by grant CPDA121180/12 from the University of Padua, Italy.

Appendix A. Posterior Computation

For a fixed truncation level L^* and a latent factor dimension K^* the detailed steps of the Gibbs sampler for posterior computations are:

1. Define the vector of the latent states and the error terms in the state space equation resulting from nGP prior for dictionary elements as

$$\begin{aligned}\Xi_i &= [\xi_{11}(t_i), \xi_{21}(t_i), \dots, \xi_{L^*K^*}(t_i), \xi'_{11}(t_i), \dots, \xi'_{L^*K^*}(t_i), A_{11}(t_i), \dots, A_{L^*K^*}(t_i)]^T, \\ \Omega_{i,\xi} &= [\omega_{i,\xi_{11}}, \omega_{i,\xi_{21}}, \dots, \omega_{i,\xi_{L^*K^*}}, \omega_{i,A_{11}}, \omega_{i,A_{21}}, \dots, \omega_{i,A_{L^*K^*}}]^T.\end{aligned}$$

Given Θ , $\{\eta_i\}_{i=1}^T$, $\{y_i\}_{i=1}^T$, Σ_0 and the variances in latent state equations $\{\sigma_{\xi_{lk}}^2\}$, $\{\sigma_{A_{lk}}^2\}$, with $l = 1, \dots, L^*$ and $k = 1, \dots, K^*$; update $\{\Xi_i\}_{i=1}^T$ by using the simulation smoother in the following state space model

$$y_i = [\eta_i^T \otimes \Theta, 0_{p \times (2 \times K^* \times L^*)}] \Xi_i + \epsilon_i, \quad (9)$$

$$\Xi_{i+1} = T_i \Xi_i + R_i \Omega_{i,\xi}, \quad (10)$$

where the observation equation in (9) results by applying the *vec* operator in the latent factor model $y_i = \Theta \xi(t_i) \eta_i + \epsilon_i$. More specifically recalling the property $\text{vec}(ABC) = (C^T \otimes A) \text{vec}(B)$ we obtain

$$\begin{aligned}y_i = \text{vec}(y_i) &= \text{vec}\{\Theta \xi(t_i) \eta_i + \epsilon_i\} \\ &= \text{vec}\{\Theta \xi(t_i) \eta_i\} + \text{vec}(\epsilon_i) \\ &= (\eta_i^T \otimes \Theta) \text{vec}\{\xi(t_i)\} + \epsilon_i.\end{aligned}$$

The state equation in (10) is a joint representation of the equations resulting from the nGP prior on each $\xi_{lk}(t)$ defined in (3). As a result, the $(3 \times L^* \times K^*) \times (3 \times L^* \times K^*)$ matrix T_i together with the $(3 \times L^* \times K^*) \times (2 \times L^* \times K^*)$ matrix R_i reproduce, for each dictionary element the state equation in (3) by fixing to 0 the coefficients relating latent states with different (l, k) (from the independence between the dictionary elements). Finally, recalling the assumptions on $\omega_{i,\xi_{lk}}$ and $\omega_{i,A_{lk}}$, $\Omega_{i,\xi}$ is normally distributed with $E[\Omega_{i,\xi}] = 0$ and $E[\Omega_{i,\xi} \Omega_{i,\xi}^T] = \text{diag}(\sigma_{\xi_{11}}^2 \delta_i, \dots, \sigma_{\xi_{L^*K^*}}^2 \delta_i, \sigma_{A_{11}}^2 \delta_i, \dots, \sigma_{A_{L^*K^*}}^2 \delta_i)$.

2. Given $\{\Xi_i\}_{i=1}^T$ sample each $\sigma_{\xi_{lk}}^2$ and $\sigma_{A_{lk}}^2$ respectively from

$$\begin{aligned}\sigma_{\xi_{lk}}^2 | \{\Xi_i\} &\sim \text{InvGa} \left(a_\xi + \frac{T}{2}, b_\xi + \frac{1}{2} \sum_{i=1}^{T-1} \frac{(\xi'_{lk}(t_{i+1}) - \xi'_{lk}(t_i) - A_{lk}(t_i)\delta_i)^2}{\delta_i} \right), \\ \sigma_{A_{lk}}^2 | \{\Xi_i\} &\sim \text{InvGa} \left(a_A + \frac{T}{2}, b_A + \frac{1}{2} \sum_{i=1}^{T-1} \frac{(A_{lk}(t_{i+1}) - A_{lk}(t_i))^2}{\delta_i} \right).\end{aligned}$$

3. Similarly to Ξ_i and $\Omega_{i,\xi}$ let

$$\begin{aligned}\Psi_i &= [\psi_1(t_i), \psi_2(t_i), \dots, \psi_{K^*}(t_i), \psi'_1(t_i), \dots, \psi'_{K^*}(t_i), B_1(t_i), \dots, B_{K^*}(t_i)]^T, \\ \Omega_{i,\psi} &= [\omega_{i,\psi_1}, \omega_{i,\psi_2}, \dots, \omega_{i,\psi_{K^*}}, \omega_{i,B_1}, \omega_{i,B_2}, \dots, \omega_{i,B_{K^*}}]^T,\end{aligned}$$

be the vectors of the latent states and error terms in the state space equation resulting from nGP prior for $\psi(t)$. Conditional on Θ , $\{\xi(t_i)\}_{i=1}^T$, $\{y_i\}_{i=1}^T$, Σ_0 , and the variances in latent state equations $\{\sigma_{\psi_k}^2\}$, $\{\sigma_{B_k}^2\}$, with $k = 1, \dots, K^*$; sample $\{\Psi_i\}_{i=1}^T$ from the simulation smoother in the following state space model

$$y_i = [\Theta\xi(t_i), 0_{p \times (2 \times K^*)}] \Psi_i + \varpi_i, \quad (11)$$

$$\Psi_{i+1} = G_i \Psi_i + F_i \Omega_{i,\psi}, \quad (12)$$

$\varpi_i \sim N(0, \Theta\xi(t_i)\xi(t_i)^T\Theta^T + \Sigma_0)$. The observation equation in (11) results by marginalizing out ν_i in the latent factor model with nonparametric mean regression $y_i = \Theta\xi(t_i)\psi(t_i) + \Theta\xi(t_i)\nu_i + \epsilon_i$. Analogously to Ξ_i , the state equation in (12) is a joint representation of the state equation induced by the nGP prior on each $\psi_k(t)$ defined in (4); where the $(3 \times K^*) \times (3 \times K^*)$ matrix G_i and the $(3 \times K^*) \times (2 \times K^*)$ matrix F_i are constructed with the same goal of the matrices T_i and R_i in the state space model for Ξ_i . Finally, $\Omega_{i,\psi} \sim N_{2 \times K^*}(0, \text{diag}(\sigma_{\psi_1}^2 \delta_i, \sigma_{\psi_2}^2 \delta_i, \dots, \sigma_{\psi_{K^*}}^2 \delta_i, \sigma_{B_1}^2 \delta_i, \sigma_{B_2}^2 \delta_i, \dots, \sigma_{B_{K^*}}^2 \delta_i))$.

4. Given $\{\Psi_i\}_{i=1}^T$ update each $\sigma_{\psi_k}^2$ and $\sigma_{B_k}^2$ respectively from

$$\begin{aligned}\sigma_{\psi_k}^2 | \{\Psi_i\} &\sim \text{InvGa} \left(a_\psi + \frac{T}{2}, b_\psi + \frac{1}{2} \sum_{i=1}^{T-1} \frac{(\psi'_k(t_{i+1}) - \psi'_k(t_i) - B_k(t_i)\delta_i)^2}{\delta_i} \right), \\ \sigma_{B_k}^2 | \{\Psi_i\} &\sim \text{InvGa} \left(a_B + \frac{T}{2}, b_B + \frac{1}{2} \sum_{i=1}^{T-1} \frac{(B_k(t_{i+1}) - B_k(t_i))^2}{\delta_i} \right).\end{aligned}$$

5. Conditioned on Θ , Σ_0 , y_i , $\xi(t_i)$ and $\psi(t_i)$, and recalling $\nu_i \sim N_{K^*}(0, I_{K^*})$; the standard conjugate posterior distribution $\nu_i | \Theta, \Sigma_0, \tilde{y}_i, \xi(t_i), \psi(t_i)$ is

$$N_{K^*} \left((I + \xi(t_i)^T \Theta^T \Sigma_0^{-1} \Theta \xi(t_i))^{-1} \xi(t_i)^T \Theta^T \Sigma_0^{-1} \tilde{y}_i, (I + \xi(t_i)^T \Theta^T \Sigma_0^{-1} \Theta \xi(t_i))^{-1} \right),$$

with $\tilde{y}_i = y_i - \Theta\xi(t_i)\psi(t_i)$.

6. Conditioned on Θ , $\{\eta_i\}_{i=1}^T$, $\{y_i\}_{i=1}^T$, and $\{\xi(t_i)\}_{i=1}^T$ (obtained from Ξ_i), the standard conjugate posterior from which to update σ_j^{-2} is

$$\sigma_j^{-2} | \Theta, \{\eta_i\}, \{y_i\}, \{\xi(t_i)\} \sim \text{Ga} \left(a_\sigma + \frac{T}{2}, b_\sigma + \frac{1}{2} \sum_{i=1}^T (y_{ji} - \theta_j \xi(t_i) \eta_i)^2 \right).$$

Where $\theta_j = [\theta_{j1}, \dots, \theta_{jL^*}]$

7. Given $\{\eta_i\}_{i=1}^T$, $\{y_i\}_{i=1}^T$, $\{\xi(t_i)\}_{i=1}^T$ and the hyperparameters ϕ and τ the shrinkage prior on Θ combined with the likelihood for the latent factor model lead to the Gaussian posterior

$$\theta_j | \{\eta_i\}, \{y_i\}, \{\xi(t_i)\}, \phi, \tau \sim \text{N}_{L^*} \left(\tilde{\Sigma}_\theta \tilde{\eta}^T \sigma_j^{-2} \begin{bmatrix} y_{j1} \\ \vdots \\ y_{jT} \end{bmatrix}, \tilde{\Sigma}_\theta \right),$$

where $\tilde{\eta}^T = [\xi(t_1)\eta_1, \xi(t_2)\eta_2, \dots, \xi(t_T)\eta_T]$ and

$$\tilde{\Sigma}_\theta^{-1} = \sigma_j^{-2} \tilde{\eta}^T \tilde{\eta} + \text{diag}(\phi_{j1}\tau_1, \dots, \phi_{jL^*}\tau_{L^*}).$$

8. The Gamma prior on the local shrinkage hyperparameter ϕ_{jl} implies the standard conjugate posterior given θ_{jl} and τ_l

$$\phi_{jl} | \theta_{jl}, \tau_l \sim \text{Ga} \left(2, \frac{3 + \tau_l \theta_{jl}^2}{2} \right)$$

9. Conditioned on Θ and τ , sample the global shrinkage hyperparameters from

$$\vartheta_1 | \Theta, \tau^{(-1)} \sim \text{Ga} \left(a_1 + \frac{pL^*}{2}, 1 + \frac{1}{2} \sum_{l=1}^{L^*} \tau_l^{(-1)} \sum_{j=1}^p \phi_{jl} \theta_{jl}^2 \right),$$

$$\vartheta_h | \Theta, \tau^{(-h)} \sim \text{Ga} \left(a_2 + \frac{p(L^* - h + 1)}{2}, 1 + \frac{1}{2} \sum_{l=h}^{L^*} \tau_l^{(-h)} \sum_{j=1}^p \phi_{jl} \theta_{jl}^2 \right),$$

where $\tau_l^{(-h)} = \prod_{t=1, t \neq h}^l \tau_t$ for $h = 1, \dots, L^*$.

10. Given the posterior samples from Θ , Σ_0 , $\{\xi(t_i)\}_{i=1}^T$ and $\{\psi(t_i)\}_{i=1}^T$ the realization of the LAF process for $\{\mu(t_i), \Sigma(t_i), t_i \in \mathcal{T}\}$ conditioned on the data $\{y_i\}_{i=1}^T$ is

$$\begin{aligned} \mu(t_i) &= \Theta \xi(t_i) \psi(t_i), \\ \Sigma(t_i) &= \Theta \xi(t_i) \xi(t_i)^T \Theta^T + \Sigma_0. \end{aligned}$$

Appendix B. Online Updating Algorithm

Consider Θ , Σ_0 , $\{\sigma_{\xi_{lk}}^2\}$, $\{\sigma_{A_{lk}}^2\}$, $\{\sigma_{\psi_k}^2\}$ and $\{\sigma_{B_k}^2\}$ fixed at their posterior mean $\hat{\Theta}$, $\hat{\Sigma}_0$, $\{\hat{\sigma}_{\xi_{lk}}^2\}$, $\{\hat{\sigma}_{A_{lk}}^2\}$, $\{\hat{\sigma}_{\psi_k}^2\}$, $\{\hat{\sigma}_{B_k}^2\}$ respectively, and let $\hat{\Xi}_T$, $\hat{\Sigma}_{\Xi_T}$ and $\hat{\Psi}_T$, $\hat{\Sigma}_{\Psi_T}$ be the sample mean and covariance matrix of the posterior distribution respectively for Ξ_T and Ψ_T obtained from the posterior estimates of the Gibbs sampler conditioned on $\{y_i\}_{i=1}^T$.

1. Given $\hat{\Theta}$, $\hat{\Sigma}_0$, $\{\hat{\sigma}_{\xi_{lk}}^2\}$, $\{\hat{\sigma}_{A_{lk}}^2\}$, $\{\eta_i\}_{i=T+1}^{T+H}$ and $\{y_i\}_{i=T+1}^{T+H}$ update $\{\Xi_i\}_{i=T+1}^{T+H}$ by using the simulation smoother in the following state space model

$$\begin{aligned} y_i &= [\eta_i^T \otimes \hat{\Theta}, 0_{p \times (2 \times K^* \times L^*)}] \Xi_i + \epsilon_i, \\ \Xi_{i+1} &= T_i \Xi_i + R_i \Omega_{i,\xi}, \end{aligned}$$

where Ξ_{T+1} can be initialized from the standard one step ahead predictive distribution for the state space model $\Xi_{T+1} \sim N(T_T \hat{\Xi}_T, T_T \hat{\Sigma}_{\Xi_T} T_T^T + R_T E[\Omega_{T,\xi} \Omega_{T,\xi}^T] R_T^T)$.

2. Conditioned on $\hat{\Theta}$, $\hat{\Sigma}_0$, $\{\hat{\sigma}_{\psi_k}^2\}$, $\{\hat{\sigma}_{B_k}^2\}$, $\{\xi(t_i)\}_{i=T+1}^{T+H}$ and $\{y_i\}_{i=T+1}^{T+H}$ sample $\{\Psi_i\}_{i=T+1}^{T+H}$ through the simulation smoother in the state space model

$$\begin{aligned} y_i &= [\hat{\Theta} \xi(t_i), 0_{p \times (2 \times K^*)}] \Psi_i + \varpi_i, \\ \Psi_{i+1} &= G_i \Psi_i + F_i \Omega_{i,\psi}. \end{aligned}$$

Similarly to Ξ_{T+1} , $\Psi_{T+1} \sim N(G_T \hat{\Psi}_T, G_T \hat{\Sigma}_{\Psi_T} G_T^T + F_T E[\Omega_{T,\psi} \Omega_{T,\psi}^T] F_T^T)$.

3. Given $\hat{\Theta}$, $\hat{\Sigma}_0$, $\{y_i\}$, $\xi(t_i)$ and $\psi(t_i)$, for $i = T + 1, \dots, T + H$, sample ν_i from the standard conjugate posterior distribution for $\nu_i | \hat{\Theta}, \hat{\Sigma}_0, \tilde{y}_i, \xi(t_i), \psi(t_i)$:

$$N_{K^*} \left((I + \xi(t_i)^T \hat{\Theta}^T \hat{\Sigma}_0^{-1} \hat{\Theta} \xi(t_i))^{-1} \xi(t_i)^T \hat{\Theta}^T \hat{\Sigma}_0^{-1} \tilde{y}_i, (I + \xi(t_i)^T \hat{\Theta}^T \hat{\Sigma}_0^{-1} \hat{\Theta} \xi(t_i))^{-1} \right),$$

with $\tilde{y}_i = y_i - \hat{\Theta} \xi(t_i) \psi(t_i)$.

4. Compute the updated covariance $\{\Sigma(t_i)\}_{i=T+1}^{T+H}$ and mean $\{\mu(t_i)\}_{i=T+1}^{T+H}$ from the usual equations

$$\begin{aligned} \Sigma(t_i) &= \hat{\Theta} \xi(t_i) \xi(t_i)^T \hat{\Theta}^T + \hat{\Sigma}_0, \\ \mu(t_i) &= \hat{\Theta} \xi(t_i) \psi(t_i). \end{aligned}$$

References

- O. Aguilar and M. West. Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics*, 18(3):338–357, 2000.
- C.O. Alexander. Orthogonal GARCH. *Mastering Risk*, 2:21–38, 2001.
- T. Baig and I. Goldfajn. Financial Market Contagion in the Asian Crisis. *IMF Staff Papers*, 46(2):167–195, 1999.

- A. Bhattacharya and D.B. Dunson. Sparse Bayesian infinite factor models. *Biometrika*, 98(2):291–306, 2011.
- T. Bollerslev, R.F. Engle and J.M. Wooldridge. A capital-asset pricing model with time-varying covariances. *Journal of Political Economy*, 96(1):116–131, 1988.
- M. Bru. Wishart Processes. *Journal of Theoretical Probability*, 4(4):725–751, 1991.
- P. Burns. Multivariate GARCH with Only Univariate Estimation. Available electronically at <http://www.burns-stat.com/pages/Working/multgarchuni.pdf>, 2005.
- C.M. Carvalho, J.E. Lucas, Q. Wang, J. Chang, J.R. Nevins and M. West. High-dimensional sparse factor modeling - Applications in gene expression genomics. *Journal of the American Statistical Association*, 103(484):1438–1456, 2008.
- S. Claessens and K. Forbes. *International Financial Contagion*. Springer, New York, 2001.
- Z. Ding. *Time series analysis of speculative returns*. PhD thesis, University of California, San Diego, 1994.
- D.L. Donoho and J.M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.
- D.L. Donoho and J.M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224, 1995.
- J. Durbin and S. Koopman. *Time Series Analysis by State Space Methods*. Oxford University Press Inc., New York, 2001.
- J. Durbin and S. Koopman. A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–615, 2002.
- R.F. Engle and K.F. Kroner. Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11(1):122–150, 1995.
- R.F. Engle. Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350, 2002.
- J. Fan and I. Gijbels. Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society. Series B*, 57(2):371–394, 1995.
- E. Fox and D.B. Dunson. Bayesian Nonparametric Covariance Regression. Available electronically at <http://arxiv.org/pdf/1101.2017.pdf>, 2011.
- J.H. Friedman. Multivariate Adaptive Regression Splines. *Annals of Statistics*, 19(1):1–67, 1991.
- A. Gelman and D.B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 1992.

- E.I. George and R.E. McCulloch. Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889, 1993.
- J. Geweke and G. Zhou. Measuring the pricing error of the arbitrage pricing theory. *Review of Financial Studies*, 9(2):557–587, 1996.
- J. Ghosh and D.B. Dunson. Default Prior Distributions and Efficient Posterior Computation in Bayesian Factor Analysis. *Journal of Computational and Graphical Statistics*, 18(2):306–320, 2009.
- T.J. Hastie and R.J. Tibshirani. *Generalized Additive Models*. Chapman & Hall, London, 1990.
- J.Z. Huang, C.O. Wu and L. Zhou. Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika*, 89(1):111–128, 2002.
- R.E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.
- H.F. Lopes and C.M. Carvalho. Factor stochastic volatility with time-varying loadings and Markov switching regimes. *Journal of Statistical Planning and Inference*, 137(10):3082–3091, 2007.
- H.F. Lopes, D. Gamerman and E. Salazar. Generalized spatial dynamic factor models. *Computational Statistics & Data Analysis*, 55(3):1319–1330, 2011.
- H.F. Lopes, E. Salazar and D. Gamerman. Spatial Dynamic Factor Analysis. *Bayesian Analysis*, 3(4):759–792, 2008.
- H.F. Lopes and M. West. Bayesian model assessment in factor analysis. *Statistica Sinica*, 14(1):41–67, 2004.
- J. Nakajima and M. West. Bayesian analysis of latent threshold dynamic models. *Journal of Business and Economic Statistics*, 31(2):151–164, 2013.
- D. Pati, A. Bhattacharya, N.S. Pillai and D.B. Dunson. Bayesian high-dimensional covariance matrix estimation. Available electronically at <http://ftp.stat.duke.edu/WorkingPapers/12-05.html>, 2012.
- C.E. Rasmussen and C.K.I Williams. *Gaussian processes for machine learning*. MIT Press, Boston, 2006.
- B.W. Silverman. Spline Smoothing: The Equivalent Variable Kernel Method. *Annals of Statistics*, 12(3):898–916, 1984.
- M. Smith and R. Kohn. Nonparametric regression using Bayesian variable selection. *Journal of Econometrics*, 75(2):317–343, 1996.
- R.S. Tsay. *Analysis of Financial Time Series*. II ed., Wiley, New York, 2005.

- R. van der Weide. GO-GARCH: a multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics*, 17(5):549–564, 2002.
- M. West. Bayesian factor regression models in the large p , small n paradigm. In *Proceedings of the Seventh Valencia International Meeting*, pages 723–732, Tenerife, Spain, 2003.
- A.G. Wilson and Z. Ghahramani. Generalised Wishart Processes. Available electronically at <http://arxiv.org/pdf/1101.0240.pdf>, 2010.
- R.L. Wolpert, M.A. Clyde and C. Tu. Stochastic expansions using continuous dictionaries: Levy adaptive regression kernels. *Annals of Statistics*, 39(4):1916–1962, 2011.
- C.O. Wu, C.T. Chiang and D.R. Hoover. Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *Journal of the American Statistical Association*, 93(444):1388–1402, 1998.
- B. Zhu and D.B. Dunson. Locally Adaptive Bayes Nonparametric Regression via Nested Gaussian Processes. *Journal of the American Statistical Association*, 108(504):1445–1456, 2013.