Nonparametric Mixtures of Multi-Output Heteroscedastic Gaussian Processes for Volatility Modeling

Emmanouil A. Platanios
Department of Electrical and Electronic Engineering
Imperial College London
e.a.platanios@gmail.com

Sotirios P. Chatzis
Department of Electrical and Computer Engineering, and Informatics
Cyprus University of Technology
sotirios.chatzis@cut.ac.cy

Concept

- Nonparametric Bayesian method for multivariate volatility modeling (important in econometrics for example)
- Dynamic covariance matrices modeling for high-dimensional vector-valued observations using Gaussian processes
- Existing methods make implausible assumptions:
  - Consider constant (noise) variance in their prior configuration => Fail to capture the dynamic nature of the variance in the modeled data
  - Consider zero correlation between the components of the modeled vector-valued outputs => Neglect significant covariance structure in the modeled data

Proposed Approach

- We:
  - Develop a hierarchical Bayesian model for heteroscedastic Gaussian process regression
  - Noise covariance is considered a separate observation-driven Gaussian process
  - Use a convolved kernel to allow for capturing the covariance structure in the output variables of the model
- Still, Gaussian process modeling requires data that can be described sufficiently well under the Gaussian assumption
- Unrealistic in volatility modeling in econometrics: heavy-tailed and skewed data with power-law behavior
- To attack this shortcoming, we resort to Bayesian non-parametrics:
  - We postulate a Pitman-Yor process prior over the space of heteroscedastic Gaussian processes.
    - Generates a (theoretically) infinite number of component GP models, with input variables in the whole input space considered each time
  - Efficient inference by means of truncated variational Bayes

Experiments

- Forecasting in financial return time series
- The asset return at time t, $y(t)$, is defined as the one-step differential of the price $p(t)$ of an asset, i.e. $y(t) \triangleq p(t) - p(t-1)$, while volatility is defined as the standard deviation of a financial return series at time instant given the information available at time $t-1$
- We work with the Global Large-Cap Equity Indices of the period 1998-2003, available in the Econometrics Toolbox of MATLAB
- This dataset comprises daily return series for $M = 6$ assets over a 5-year period
- We postulate our model with its input driven by the one-step-back asset return values $\boldsymbol{y}(t-1) = [y_m(t-1)]_{m=1}^{M}$, and its output comprising the return series $\boldsymbol{y}(t)$
- To remain consistent with the existing literature, we adopt the typical assumption of zero-mean return series, setting $\boldsymbol{f} = \boldsymbol{0}$ in our model
- We conduct model training over windows of 120 days; this procedure is repeated every 7 days. In each case, prediction is performed one, seven, and 30 days ahead. As our performance metric, we use the MSE between the predicted volatility and (i) the **squared returns**, and (ii) the **squared standard deviation** over the employed sliding windows

Squared returns MSE performance obtained by the evaluated methods

Method	1-day	7-day	30-day
VHGP	9.87×10^{-7}	1.01×10^{-6}	1.02×10^{-6}
Proposed Approach: $C = 5$	4.79×10^{-7}	4.62×10^{-7}	4.87×10^{-7}
Proposed Approach: $C = 10$	3.78×10^{-7}	3.65×10^{-7}	3.93×10^{-7}

Sliding window variance MSE performance obtained by the evaluated methods

Method	1-day	7-day	30-day
VHGP	1.28×10^{-6}	1.27×10^{-6}	1.23×10^{-6}
Proposed Approach: $C = 5$	4.54×10^{-7}	4.20×10^{-7}	4.17×10^{-7}
Proposed Approach: $C = 10$	2.22×10^{-7}	1.99×10^{-7}	2.08×10^{-7}

Model Formulation

Input Vector: $\boldsymbol{x} = [\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, \dots, \boldsymbol{x}_N^T]^T$
Output Vector: $\boldsymbol{y} = [\boldsymbol{y}_1^T, \boldsymbol{y}_2^T, \dots, \boldsymbol{y}_N^T]^T$
where: $\boldsymbol{y}_n = [y_1(\boldsymbol{x}_n), y_2(\boldsymbol{x}_n), \dots, y_M(\boldsymbol{x}_n)]^T$
We assume that: $y_m(\boldsymbol{x}_n) = f_m^c(\boldsymbol{x}_n) + r_m^c(\boldsymbol{x}_n), \quad \forall m = 1,2,\dots,M$
and: $r_m^c(\boldsymbol{x}_n) = e^{g_m^c(\boldsymbol{x}_n)}$
We introduce the notation: $\boldsymbol{r}^c = [\boldsymbol{r}_1^{c\,T}, \boldsymbol{r}_2^{c\,T}, \dots, \boldsymbol{r}_N^{c\,T}]^T$
and: $\boldsymbol{r}_n^c = [r_1^c(\boldsymbol{x}_n), r_2^c(\boldsymbol{x}_n), \dots, r_M^c(\boldsymbol{x}_n)]^T$
Then, the definition of our model comprises the assumptions:

$$p(\boldsymbol{y}_n|\boldsymbol{x}_n, z_{nc} = 1) = \mathcal{N}(\boldsymbol{y}_n|\boldsymbol{f}_n^c, \boldsymbol{R}_n^c)$$

$$p(\boldsymbol{f}^c|\boldsymbol{x}, \phi) = \mathcal{N}(\boldsymbol{f}^c|\boldsymbol{0}, \boldsymbol{K}_{f,f}^c)$$

$$p(\boldsymbol{g}^c|\boldsymbol{x}; \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{g}^c|\mu_0^c\boldsymbol{1}, \boldsymbol{K}_{g,g}^c)$$

where we denote: $\boldsymbol{R}_n^c = diag(\boldsymbol{r}_n^c)$, $\boldsymbol{g}^c = [\boldsymbol{g}_1^{c\,T}, \boldsymbol{g}_2^{c\,T}, \dots, \boldsymbol{g}_N^{c\,T}]^T$,
$\boldsymbol{g}_n^c = [g_1^c(\boldsymbol{x}_n), g_2^c(\boldsymbol{x}_n), \dots, g_M^c(\boldsymbol{x}_n)]^T$, $\boldsymbol{f}^c = [\boldsymbol{f}_1^{c\,T}, \boldsymbol{f}_2^{c\,T}, \dots, \boldsymbol{f}_N^{c\,T}]^T$
and $\boldsymbol{f}_n^c = [f_1^c(\boldsymbol{x}_n), f_2^c(\boldsymbol{x}_n), \dots, f_M^c(\boldsymbol{x}_n)]^T$
In the above equations (an equivalent expression exists for $\boldsymbol{K}_{g,g}^c$):

$$\{\boldsymbol{K}_{f,f}\}_{(m-1)\times N+n,(s-1)\times N+l} = Cov\{f_m(\boldsymbol{x}_n), f_s(\boldsymbol{x}_l)\}$$

where:

$$Cov\{f_m(\boldsymbol{x}_n), f_s(\boldsymbol{x}_l)\} = \sum_{r=1}^{R} \int_{-\infty}^{\infty} k_{nr}^{f_m}(\boldsymbol{x}_n - \boldsymbol{z}) \int_{-\infty}^{\infty} k_{lr}^{f_s}(\boldsymbol{x}_l - \boldsymbol{z}')\, k_{u_r,u_r}(\boldsymbol{z}, \boldsymbol{z}')\, d\boldsymbol{z}'\, d\boldsymbol{z}$$

and the $\mathrm{k}(\cdot,\cdot)$ are autoregressive kernels of order one.
Finally, for the variables z_{nc} we use a Pitman-Yor process prior, under a stick-breaking construction: $p(z_{nc} = 1|\boldsymbol{v}) = \pi_c(\boldsymbol{v})$

$$\pi_c(\boldsymbol{v}) = v_c \prod_{j=1}^{c-1} (1 - v_j) \in [0,1]$$

$$p(v_c|\alpha) = Beta(1 - \delta, \alpha + \delta c)$$

We conduct inference for our model using the variational Bayesian paradigm, which results in simple and efficient predictive posterior expressions, by introducing a truncation threshold C, such that $\pi_c(\boldsymbol{v}) = 0\ \forall c > C$

Variational Bayes Update Equations

The variational free energy of the model reads (ignoring constant terms):

$$\mathcal{L}(q) = \sum_{c=1}^{C}\sum_{m=1}^{M} \int d\boldsymbol{f}_m^c q(\boldsymbol{f}_m^c) \log \frac{p(\boldsymbol{f}_m^c|\boldsymbol{0}, \boldsymbol{K}_{f,f}^c)}{q(\boldsymbol{f}_m^c)} + \sum_{c=1}^{C}\sum_{m=1}^{M} \int d\boldsymbol{g}_m^c q(\boldsymbol{g}_m^c) \log \frac{p(\boldsymbol{g}_m^c|\mu_0^c\boldsymbol{1}, \boldsymbol{K}_{g,g}^c)}{q(\boldsymbol{g}_m^c)}$$
$$+ \sum_{c=1}^{C-1} \int d\alpha q(\alpha) \int dv_c q(v_c) \log \frac{p(v_c|\alpha)}{q(v_c)} + \sum_{c=1}^{C}\sum_{n=1}^{N} q(z_{nc} = 1)$$
$$\left\{ \int d\boldsymbol{v} q(\boldsymbol{v}) \log p(z_{nc} = 1|\boldsymbol{v}) - \log q(z_{nc} = 1) \right.$$
$$\left. + \sum_{m=1}^{M} \int d\boldsymbol{f}_m^c d\boldsymbol{g}_m^c q(\boldsymbol{f}_m^c) q(\boldsymbol{g}_m^c) \log p(y_{nd}|\boldsymbol{f}_m^c(\boldsymbol{x}_n), \boldsymbol{R}_n^c) \right\}$$

1. Regarding the PYP stick-breaking variables v_c, we have: $q(v_c) = Beta(v_c|\beta_{c,1}, \beta_{c,2})$
   where:
   $$\beta_{c,1} = 1 - \delta + \sum_{n=1}^{N} q(z_{nc} = 1) \qquad \beta_{c,2} = \langle\alpha\rangle + c\delta + \sum_{c'=c+1}^{C}\sum_{n=1}^{N} q(z_{nc'} = 1)$$
2. Regarding the posteriors over the latent functions $\boldsymbol{f}_n^c$, we have: $q(\boldsymbol{f}_m^c) = \mathcal{N}(\boldsymbol{f}_m^c|\boldsymbol{\mu}_m^c, \boldsymbol{\Sigma}_m^c)$
   where:
   $$\boldsymbol{\mu}_m^c = \boldsymbol{\Sigma}_m^c \boldsymbol{B}_m^c \boldsymbol{y}_m \qquad \boldsymbol{\Sigma}_m^c = \left(\boldsymbol{K}_{f,f}^{c\;-1} + \boldsymbol{B}_m^c\right)^{-1}$$
   $$\boldsymbol{B}_m^c \triangleq diag\left(\left[\frac{\boldsymbol{1}}{\langle \boldsymbol{R}_n^c \rangle} \boldsymbol{q}(\boldsymbol{z}_{nc} = 1)\right]_{\boldsymbol{n}=\boldsymbol{1}}^{\boldsymbol{N}}\right)$$
3. Similarly, regarding the posteriors over the latent noise variance processes $\boldsymbol{g}_m^c$, we have:
   where: $q(\boldsymbol{g}_m^c) = \mathcal{N}(\boldsymbol{g}_m^c|\boldsymbol{m}_m^c, \boldsymbol{S}_m^c)$
   $$\boldsymbol{m}_m^c = \boldsymbol{K}_{g,g}^c \left(\boldsymbol{\Lambda}_m^c + \frac{1}{2} diag([q(z_{nc} = 1)]_{n=1}^{N})\right)\boldsymbol{1} + \mu_0^c\boldsymbol{1} \qquad \boldsymbol{S}_m^c = \left(\boldsymbol{K}_{g,g}^{c\;-1} + \boldsymbol{\Lambda}_m^c\right)^{-1}$$
   and $\boldsymbol{\Lambda}_m^c$ is a positive semi-definite diagonal matrix, whose components comprise variational parameters that can be freely set.
4. Finally, the posteriors over the latent variables Z yield:
   where: $q(z_{nc} = 1) \propto \exp(\langle \log \pi_c(\boldsymbol{v})\rangle) \exp(r_{nc})$
   $$\langle \log \pi_c(\boldsymbol{v}) \rangle = \sum_{c'=1}^{c-1} \langle \log(1 - v_{c'})\rangle + \langle \log v_c \rangle$$
   and:
   $$r_{nc} \triangleq -\frac{1}{2}\sum_{d=1}^{D} \left\{ \frac{\boldsymbol{1}}{\langle \boldsymbol{R}_n^{c\,2} \rangle} [(y_{nm} - [\boldsymbol{\mu}_m^c]_n)^2 + [\boldsymbol{\Sigma}_m^c]_{nn}] + [\boldsymbol{m}_m^c]_n \right\}$$