NONPARAMETRIC MIXTURES OF MULTI-OUTPUT HETEROSCEDASTIC GAUSSIAN PROCESSES FOR VOLATILITY MODELING

Emmanouil A. Platanios
Sotirios P. Chatzis
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
PROPOSED APPROACH

- 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
MODEL FORMULATION

Input vector: \( x = [x_1^T, x_2^T, ..., x_N^T]^T \)

Output vector: \( y = [y_1^T, y_2^T, ..., y_N^T]^T \) where \( y_n = [y_1(x_n), y_2(x_n), ..., y_M(x_n)]^T \)

We assume that: \( y_m(x_n) = f_m^c(x_n) + r_m^c(x_n), \ \forall m = 1, 2, ..., M \)

and: \( r_m^c(x_n) = e^{g_m^c(x_n)} \)

We introduce the notation: \( \mathbf{r}^c = [r_1^c, r_2^c, ..., r_N^c]^T \) and \( \mathbf{r}_n^c = [r_1^c(x_n), r_2^c(x_n), ..., r_M^c(x_n)]^T \)

Then, the definition of our model comprises the assumptions:

\[
p(y_n|x_n, z_{nc} = 1) = \mathcal{N}(y_n|f_n^c, R_n^c)
\]

\[
p(f^c|x, \phi) = \mathcal{N}(f^c|0, K_{f,f})
\]

\[
p(g^c|x; \theta) = \mathcal{N}(g^c|\mu_0^c 1, K_{g,g})
\]

Where we denote: \( R_n^c = \text{diag}(r_n^c) \), \( g^c = [g_1^c, g_2^c, ..., g_N^c]^T \), \( g_n^c = [g_1^c(x_n), g_2^c(x_n), ..., g_M^c(x_n)]^T \), \( f^c = [f_1^c, f_2^c, ..., f_N^c]^T \) and \( f_n^c = [f_1^c(x_n), f_2^c(x_n), ..., f_M^c(x_n)]^T \)
MODEL FORMULATION

In the above equations (an equivalent expression exists for $K_g^c$):

$$\{K_{f,f}\}_{(m-1)\times N+n,(s-1)\times N+l} = \text{Cov}\{f_m(x_n), f_s(x_l)\}$$

Where:

$$\text{Cov}\{f_m(x_n), f_s(x_l)\} = \sum_{r=1}^{R} \int_{-\infty}^{\infty} k_{nr}^{f_m}(x_n - z) \int_{-\infty}^{\infty} k_{lr}^{f_s}(x_l - z') k_{ur,ur}(z, z') dz' dz$$

And the $k(\cdot, \cdot)$ are autoregressive kernels of order one

Finally, for the variables $z_{nc}$ we use a Pitman-Yor process prior, under a stick-breaking construction:

$$p(z_{nc} = 1|v) = \pi_c(v)$$

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

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

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 $\pi_c(v) = 0 \ \forall c > C$
EXPERIMENTS

- Forecasting in financial return time series

Asset return at time \( t, y(t) \), is defined as the one-step differential of the price \( p(t) \) of an asset, i.e. \( y(t) \equiv p(t) - p(t - 1) \), while volatility is defined as the standard deviation of a financial return series at time instant \( t \) 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

\[
y(t - 1) = [y_m(t - 1)]_m^{M},
\]

and its output comprising the return series \( y(t) \)

- To remain consistent with the existing literature, we adopt the typical assumption of zero-mean return series, setting \( f = 0 \) 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.
## EXPERIMENTS - RESULTS

Squared returns MSE performance obtained by the evaluated methods

<table>
<thead>
<tr>
<th>Method</th>
<th>1-day</th>
<th>7-day</th>
<th>30-day</th>
</tr>
</thead>
<tbody>
<tr>
<td>VHGP</td>
<td>9.87 × 10^{-7}</td>
<td>1.01 × 10^{-6}</td>
<td>1.02 × 10^{-6}</td>
</tr>
<tr>
<td>Proposed Approach: $C = 5$</td>
<td>4.79 × 10^{-7}</td>
<td>4.62 × 10^{-7}</td>
<td>4.87 × 10^{-7}</td>
</tr>
<tr>
<td>Proposed Approach: $C = 10$</td>
<td>3.78 × 10^{-7}</td>
<td>3.65 × 10^{-7}</td>
<td>3.93 × 10^{-7}</td>
</tr>
</tbody>
</table>

Sliding window variance MSE performance obtained by the evaluated methods

<table>
<thead>
<tr>
<th>Method</th>
<th>1-day</th>
<th>7-day</th>
<th>30-day</th>
</tr>
</thead>
<tbody>
<tr>
<td>VHGP</td>
<td>1.28 × 10^{-6}</td>
<td>1.27 × 10^{-6}</td>
<td>1.23 × 10^{-6}</td>
</tr>
<tr>
<td>Proposed Approach: $C = 5$</td>
<td>4.54 × 10^{-7}</td>
<td>4.20 × 10^{-7}</td>
<td>4.17 × 10^{-7}</td>
</tr>
<tr>
<td>Proposed Approach: $C = 10$</td>
<td>2.22 × 10^{-7}</td>
<td>1.99 × 10^{-7}</td>
<td>2.08 × 10^{-7}</td>
</tr>
</tbody>
</table>
THANK YOU

Emmanouil A. Platanios
Sotirios P. Chatzis
The variational free energy of the model reads (ignoring constant terms):

\[
\mathcal{L}(q) = \sum_{c=1}^{C} \sum_{m=1}^{M} \int df_m q(f_m) \log \frac{p(f_m|0, K^c_{f,f})}{q(f_m)} + \sum_{c=1}^{C} \sum_{m=1}^{M} \int dg_m q(g_m) \log \frac{p(g_m|\mu_0^c 1, K^c_{g,g})}{q(g_m)}
\]

\[
+ \sum_{c=1}^{C-1} \int d\alpha q(\alpha) \int dv_c q(v_c) \log \frac{p(v_c|\alpha)}{q(v_c)} + \sum_{c=1}^{C} \sum_{n=1}^{N} q(z_{nc} = 1)
\]

\[
\left\{ \int dv q(v) \log p(z_{nc} = 1|v) - \log q(z_{nc} = 1) + \sum_{m=1}^{M} \int df_m dg_m q(f_m) q(g_m) \log p(y_{nd}|f_m(x_n), R^c_n) \right\}
\]
1. Regarding the PYP stick-breaking variables $v_c$, we have:

$$q(v_c) = \text{Beta}(v_c | \beta_{c,1}, \beta_{c,2})$$

where:

$$\beta_{c,1} = 1 - \delta + \sum_{n=1}^{N} q(z_{nc} = 1)$$

$$\beta_{c,2} = \langle \alpha \rangle + c\delta + \sum_{c'=c+1}^{C} \sum_{n=1}^{N} q(z_{nc'} = 1)$$

2. Regarding the posteriors over the latent functions $f^c_n$, we have:

$$q(f^c_m) = \mathcal{N}(f^c_m | \mu^c_m, \Sigma^c_m)$$

where:

$$\Sigma^c_m = \left( K^c_{f,f}^{-1} + B^c_m \right)^{-1}$$

$$\mu^c_m = \Sigma^c_m B^c_m y_m$$

$$B^c_m \triangleq \text{diag}\left( \frac{1}{\langle R^c_n \rangle} q(z_{nc} = 1) \right)_{n=1}^{N}$$
3. Similarly, regarding the posteriors over the latent noise variance processes $g_m^c$, we have:

$$q(g_m^c) = \mathcal{N}(g_m^c|m_m^c, s_m^c)$$

where:

$$s_m^c = (K_{g,g}^{-1} + \Lambda_m^c)^{-1}$$

$$m_m^c = K_{g,g}^{c} \left( \Lambda_m^c + \frac{1}{2} \text{diag}([q(z_{nc} = 1)]_{n=1}^{N}) \right) 1 + \mu_0^c 1$$

and $\Lambda_m^c$ 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:

$$q(z_{nc} = 1) \propto \exp(\langle \log \pi_c(v) \rangle) \exp(r_{nc})$$

where:

$$\langle \log \pi_c(v) \rangle = \sum_{c'=1}^{c-1} \langle \log(1 - v_{c'}) \rangle + \langle \log v_c \rangle$$

and:

$$r_{nc} \triangleq -\frac{1}{2} \sum_{d=1}^{D} \left\{ \frac{1}{(R_n^c)^2} \left[ (y_{nm} - [\mu_m^c]_n)^2 + [\Sigma_m^c]_{nn} \right] + [m_m^c]_{n} \right\}$$