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

Abstract

In this work, we present a nonparametric Bayesian method for multivariate volatility modeling. Our approach is based on postulation of a novel mixture of multi-output heteroscedastic Gaussian processes to model the covariance matrices of multiple assets. Specifically, we use the Pitman-Yor process prior as the nonparametric prior imposed over the components of our model, which are taken as multioutput heteroscedastic Gaussian processes obtained by introducing appropriate convolution kernels that combine simple heteroscedastic Gaussian processes under a multioutput scheme. We exhibit the efficacy of our approach in a volatility prediction task.

1 Introduction

In this work, we aim to address the problem of appropriately modeling the dynamic covariance matrices for high-dimensional vector-valued observations using Gaussian processes. Indeed, existing Gaussian process models typically make two rather implausible assumptions: (i) they assume a constant (noise) variance in their prior configuration, thus failing to capture the dynamic nature of the variance in the modeled data; and, (ii) they usually consider that the modeled vector-valued outputs consist of independent components, thus neglecting the covariance structure in the modeled data.

Approaches that allow for modeling the dynamic covariance matrices (volatility) for high-dimensional vector-valued observations are essential in many real-life applications. One characteristic area where volatility models are especially important is econometrics: the price of essentially every derivative security is affected by swings in volatility. Indeed, ARCH [5], and, more generally, GARCH [3] models comprise the most popular family of methods used for predicting the volatilities of returns on equity indices and currency exchanges [12, 7]. Multivariate volatility models can be used to understand the dynamic correlations (or co-movement) between equity indices, and can make better univariate predictions than univariate models.

Gaussian process (GP) models comprise one of the most popular Bayesian methods for regression, function approximation, and predictive density estimation [6]. Despite their significant flexibility and success in many application domains, GPs do also suffer from several limitations. To begin with, most GP approaches assume that the modeled vector-valued outputs consist of independent components, thus totally neglecting the actually existing covariance structure. GP models with convolved kernels have been recently proposed to resolve these shortcomings (see, e.g., [1]).

In addition, most existing GP formulations assume that the prior (noise) variance is a constant, thus failing to incorporate in their prior assumptions the dynamic nature of the variance in the modeled data. Recently, a heteroscedastic GP approach has been proposed in [8], in an effort to resolve these issues. In that work, a scalar-valued (single-output) GP is postulated, where the noise variance is considered to depend on the (multivariate) input observations of the model, with this dependence modeled by means of an additional, separately postulated GP model.

Finally, when dealing with tasks entailing non-stationary covariance functions, multi-modal output, or discontinuities, specifying appropriate priors and performing learning in GP models tends to become considerably cumbersome, if at all possible. To alleviate these issues, several researchers have considered postulating ensembles of GPs, which either divide the input space into segments, and train a different GP on each one of these segments, or obtain non-parametric mixture model-like ensembles of GPs, with suitable mixing priors, e.g. Dirichlet processes [2], or Pitman-Yor processes [11] (see, e.g., [13, 10, 9]).

Inspired by these advances, in this paper we propose a Non-Parametric Mixture of Multi-Output Heteroscedastic Gaussian Processes for Volatility Modeling. Our method comprises the postulation of a nonparametric Bayesian model based on the introduction of a Pitman-Yor process prior over the space of possible Gaussian processes with input variables in the whole input space considered each time. In addition, each one of the postulated component GPs uses a suitable convolved kernel, to allow for capturing the covariance structure in the output variables of the model, which most existing GP formulations neglect to a great extend. Finally, for each one of the postulated component mixture GPs, a dynamic noise covariance is assumed, modeled by means of an additional postulated GP; that is, the mixture component GPs in our model are considered of an heteroscedastic nature, which allows for effectively capturing the volatility in the modeled data. We experimentally exhibit the efficacy of our approach.

2 Proposed Approach

Let us consider N multidimensional input vectors $\boldsymbol{x} = \left[\boldsymbol{x_1}^T, \boldsymbol{x_2}^T, \ldots, \boldsymbol{x_N}^T\right]^T$ with corresponding multidimensional output vectors $\boldsymbol{y} = \left[\boldsymbol{y_1}^T, \boldsymbol{y_2}^T, \ldots, \boldsymbol{y_N}^T\right]^T$, where $\boldsymbol{y_n} = \left[y_1\left(\boldsymbol{x_n}\right), y_2\left(\boldsymbol{x_n}\right), \ldots, y_M\left(\boldsymbol{x_n}\right)\right]^T$. We assume that

$$y_m\left(\boldsymbol{x_n}\right) = f_m^c\left(\boldsymbol{x_n}\right) + r_m^c\left(\boldsymbol{x_n}\right) \;\forall m = 1, \ldots, M \tag{1}$$

and

$$r_m^c\left(\boldsymbol{x_n}\right) = e^{g_m^c\left(\boldsymbol{x_n}\right)} \tag{2}$$

In the above equations, the index c refers to the component index of a postulated mixture model. Let us also introduce the notation $\boldsymbol{r^c} = \left[\boldsymbol{r_1^c}^T, \boldsymbol{r_2^c}^T, \ldots, \boldsymbol{r_N^c}^T\right]^T$ and $\boldsymbol{r_n^c} = \left[r_1^c\left(\boldsymbol{x_n}\right), r_2^c\left(\boldsymbol{x_n}\right), \ldots, r_M^c\left(\boldsymbol{x_n}\right)\right]^T$. Then, definition of our model comprises the assumptions:

$$p\left(\boldsymbol{y_n}|\boldsymbol{x_n}, z_{nc} = 1\right) = \mathcal{N}\left(\boldsymbol{y_n}|\boldsymbol{f_n^c}, \boldsymbol{R_n^c}\right) \tag{3}$$

$$p\left(\boldsymbol{f^c}|\boldsymbol{x}; \boldsymbol{\phi}\right) = \mathcal{N}\left(\boldsymbol{f^c}|\boldsymbol{0}, \boldsymbol{K_{f,f}^c}\right) \tag{4}$$

$$p\left(\boldsymbol{g^c}|\boldsymbol{x}; \boldsymbol{\theta}\right) = \mathcal{N}\left(\boldsymbol{g^c}|\mu_0^c\boldsymbol{1}, \boldsymbol{K_{g,g}^c}\right) \tag{5}$$

where we denote $\boldsymbol{R_n^c} = \text{diag}\left(\boldsymbol{r_n^c}\right)$, $\boldsymbol{g^c} = \left[\boldsymbol{g_1^c}^T, \boldsymbol{g_2^c}^T, \ldots, \boldsymbol{g_N^c}^T\right]^T$, $\boldsymbol{g_n^c} = \left[g_1^c\left(\boldsymbol{x_n}\right), g_2^c\left(\boldsymbol{x_n}\right), \ldots, g_M^c\left(\boldsymbol{x_n}\right)\right]^T$, $\boldsymbol{f^c} = \left[\boldsymbol{f_1^c}^T, \boldsymbol{f_2^c}^T, \ldots, \boldsymbol{f_N^c}^T\right]^T$, and $\boldsymbol{f_n^c} = \left[f_1^c\left(\boldsymbol{x_n}\right), f_2^c\left(\boldsymbol{x_n}\right), \ldots, f_M^c\left(\boldsymbol{x_n}\right)\right]^T$.

In the above equations, the $\boldsymbol{K_{f,f}^c}$ and $\boldsymbol{K_{g,g}^c}$ are Gram matrices defined over the sets of input observations $\boldsymbol{x}$. To allow for capturing the covariance structure between the modeled outputs, we use convolved kernel functions, similar to [1]. We let

$$\left\{\boldsymbol{K_{f,f}}\right\}_{(m-1)\times N+n,(s-1)\times N+l} = \text{Cov}\left\{f_m\left(\boldsymbol{x_n}\right), f_s\left(\boldsymbol{x_l}\right)\right\} \tag{6}$$

$$\{\boldsymbol{K_{g,g}}\}_{(m-1)\times N+n,(s-1)\times N+l} = \text{Cov}\left\{g_m\left(\boldsymbol{x_n}\right), g_s\left(\boldsymbol{x_l}\right)\right\} \tag{7}$$

where

$$\text{Cov}\left\{f_m\left(\boldsymbol{x_n}\right), f_s\left(\boldsymbol{x_l}\right)\right\} = \sum_{r=1}^{R}\int_{-\infty}^{\infty} k_{nr}^{f_m}\left(\boldsymbol{x_n}-\boldsymbol{z}\right)\int_{-\infty}^{\infty} k_{lr}^{f_s}\left(\boldsymbol{x_l}-\boldsymbol{z'}\right)k_{u_r,u_r}\left(\boldsymbol{z},\boldsymbol{z'}\right)d\boldsymbol{z'}d\boldsymbol{z} \tag{8}$$

$$\text{Cov}\left\{g_m\left(\boldsymbol{x_n}\right), g_s\left(\boldsymbol{x_l}\right)\right\} = \sum_{r=1}^{R}\int_{-\infty}^{\infty} k_{nr}^{g_m}\left(\boldsymbol{x_n}-\boldsymbol{z}\right)\int_{-\infty}^{\infty} k_{lr}^{g_s}\left(\boldsymbol{x_l}-\boldsymbol{z'}\right)k_{u_r,u_r}\left(\boldsymbol{z},\boldsymbol{z'}\right)d\boldsymbol{z'}d\boldsymbol{z} \tag{9}$$

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

$$p\left(z_{nc}=1|\boldsymbol{v}\right) = \pi_c\left(\boldsymbol{v}\right) \tag{10}$$

$$\pi_c\left(\boldsymbol{v}\right) = v_c\prod_{j=1}^{c-1}\left(1-v_j\right) \in [0,1] \tag{11}$$

$$p\left(v_c|\alpha\right) = \text{Beta}\left(1-\delta,\alpha+\delta c\right) \tag{12}$$

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 (maximum allowed number of mixture components), such that $\pi_c\left(\boldsymbol{v}\right) = 0\ \forall c > C$.

3 Experiments

To show the efficacy of our approach, we apply it to perform volatility modeling and 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. $y(t) \triangleq p(t) - p(t-1)$, while volatility is defined as the standard deviation of a financial return series at time instant t given the information available at time $t-1$.

In our experiments, 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 $\boldsymbol{f} = 0$ in our model. Note though that we could as well relax this assumption, to allow for our model to capture drifts in the mean of the return series.

We conduct model training over windows of 120 days; this procedure is repeated every 7 days, i.e. we retrain our model using sliding windows 120 values-wide with overlap of 7 values. 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. These are two of the few consistent ways to measure volatility, as discussed in [4].

To obtain some comparative results, we also evaluate the VHGP model of [8] in the same task. Our considered competitor consists a popular approach that yields state-of-the-art results in financial return series volatility modeling. Our results are provided in Tables 1 and 2 for the maximum allowed number of mixture components (variational truncation theshold) equal to $C = 5$ and 10. As we observe, our approach offers a very significant improvement over its competitor, approximately of one order of magnitude.

We would like to mention that we could obtain a much larger improvement by setting the truncation level higher, but this would be quite cumbersome, due to the computational time required. Our

Table 1: Squared returns MSE performance obtained by the evaluated methods.

Method	One-day prediction	Seven-day prediction	30-day prediction
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}

Table 2: Sliding window variance MSE performance obtained by the evaluated methods.

Method	One-day prediction	Seven-day prediction	30-day prediction
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}

ongoing research efforts are targeted at coming up with suitable sparse approximation techniques that would allow for increasing the scalability of our method.

References

[1] Mauricio A. Álvarez and Neil D. Lawrence. Computationally efficient convolved multiple output gaussian processes. *Journal of Machine Learning Research*, 12(5):1459–1500, 2011.

[2] C. Antoniak. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, 2(6):1152–1174, 1974.

[3] T. Bollerslev. Modeling the coherence in short-term nominal exchange rates: A multivariate generalized $arch$ approach. *Review of Economics and Statistics*, 72:498–505, 1990.

[4] C.T. Brownlees, R.F. Engle, and B.T. Kelly. A practical guide to volatility forecasting through calm and storm, 2009. URL http://ssrn.com/abstract=1502915, 2009.

[5] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987– 1007, 1982.

[6] D. Gu and H. Hu. Spatial gaussian process regression with mobile sensor networks. *IEEE Transactions on Neural Networks and Learning Systems*, 23(8):1279 – 1290, 2012.

[7] P. R. Hansen and A. Lunde. A forecast comparison of volatility models: Does anything beat a $garch(1,1)$. *Journal of Applied Econometrics*, 20(7):873–889, 2005.

[8] Miguel Lázaro-Gredilla and Michalis Titsias. Variational heteroscedastic Gaussian process regression. In *Proc. ICML*, 2011.

[9] L.Xu, M.I.Jordan, and G.E.Hinton. Analternative model for mixtures of experts. In *Advances in Neural Information Processing Systems*, volume 7, pages 633–640. MIT Press, 1995.

[10] Edward Meeds and Simon Osindero. An alternative infinite mixture of Gaussian process experts. In *Advances in Neural Information Processing Systems*, volume 18, pages 883–890. MIT Press, Cambridge, 2006.

[11] J. Pitman and M Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. In *Annals of Probability*, volume 25, pages 855–900, 1997.

[12] S.-H. Poon and C. W. Granger. Practical issues in forecasting volatility. *Financial Analysts Journal*, 61(1):45–56, 2005.

[13] C.E. Rasmussen and Z. Ghahramani. Infinite mixtures of Gaussian process experts. In *In Advances in Neural Information Processing Systems*, volume 14, pages 881–888. MIT Press, 2002.