3. MODELLING NON-STATIONARY FINANCIAL TIME SERIES WITH INPUT WARPED STUDENT T-PROCESSES[1]

Gheorghe RUXANDA*
Sorin OPINCARIU*
Stefan IONESCU**

Abstract

The evolution of financial assets is known to be non-stationary and to present long tails and non-Gaussian. Gaussian processes are a flexible and general Bayesian nonparametric generative model that provide flexible priors on function spaces and interpretable uncertainty quantification. While GP are extremely flexible function approximators, their Gaussian marginal distribution makes them inappropriate to model financial assets returns distributions. We present the Student t-processes that are known to fit heavier tail. We also augment the model with input warping to account with the financial time series non stationarity. We present a case study of fitting the evolution of SP500 index stressing the importance of good uncertainty estimates, especially when the series manifests structural breaks.

Keywords: Bayesian nonparametric, Student processes, Gaussian processes, stylized facts.

JEL Classification: C45 C53

1. Introduction

The evolution of financial assets is a research subject of vast dimensions. The literature dedicated to modeling the evolution of financial returns focused overwhelmingly on the task of finding a quantitative relationship able to accurately describe and predict the financial returns. While we acknowledge the importance of quantitative predictions of the model, we also think that having a measure of the model's confidence in its predictions is of equal importance. The bayesian methodology offers a principled way of quantifying uncertainty.

An important branch of bayesian statistics is the bayesian nonparametrics where the concept of probability distribution is extended on more abstract spaces. The gaussian processes, proposed by Rasmussen (2003) is an example of the above mentioned extension. A gaussian process (whose technical definition will be elaborated the sections below) is a stochastic process for which each realization is a continuous function. Hence a gaussian

*Bucharest University of Economic Studies.
**Romanian American University.

process can be considered as a probability distribution on the space of continuous functions. The gaussian processes are used in functional regression settings. A gaussian process regression does not need to posit a functional form of the regression model. The functional form and the parameters of the regression model are estimated from data under the form pf a posterior distribution (bayesian inference).

Student t-processes, studied in literature (see Jylanki (2011), Vanhatalo (2009), Rios (2018)) are more robust to outliers alternative to Gaussian Process regression. At the same time Student t-processes retain the universality property of the gaussian processes. Both gaussian and Student t-processes are probability distributions on functional spaces, they only differ in the probability attached to each function. We use the Student t-processes to model the evolution of financial returns, as they are more capable to model deviations from gaussianity that we observe in the financial assets' evolution,

We also make use of the warping function Bayesian framework described in Lazaro (2012) and extend it to Student t-processes. The warping functions are a way to impose constraints on the input data (positivity, boundedness). We use the warping function framework to incorporate domain knowledge (stylized facts about financial returns). In this paper we propose a principled Bayesian nonparametric model based on warping functions and Student t-processes of the evolution of financial returns. The model proposed is non-linear autoregressive with no exogenous variables (only the history of the given asset is considered). We discuss the fitted Bayesian model for the evolution of the equity index S&P 500 and discuss how it recovers the stylized facts about financial assets returns.

2. Student T-processes

In this section, we describe the properties of gaussian and Student-t processes and the connection between them. We also present the most typical covariance functions describing how the functions sampled from a gaussian/student processed are influenced by the form of covariance functions.

2.1 Gaussian Processes

Gaussian processes are stochastic processes that define a probability distribution on a function space. Following Rassmussen (2003):

Definition 1 A Gaussian process is a stochastic process which for any finite realization x1, . . .,xn the collection of random variables $x_1, \dots, x_n$ has a joint Gaussian distribution:

$$(f(x_1), \dots, f(x_n)) \sim \aleph(m(x_1, \dots, x_n), \Sigma).$$

The Gaussian processes are completely specified by the mean and the covariance function of the stochastic process *f*(*x*):

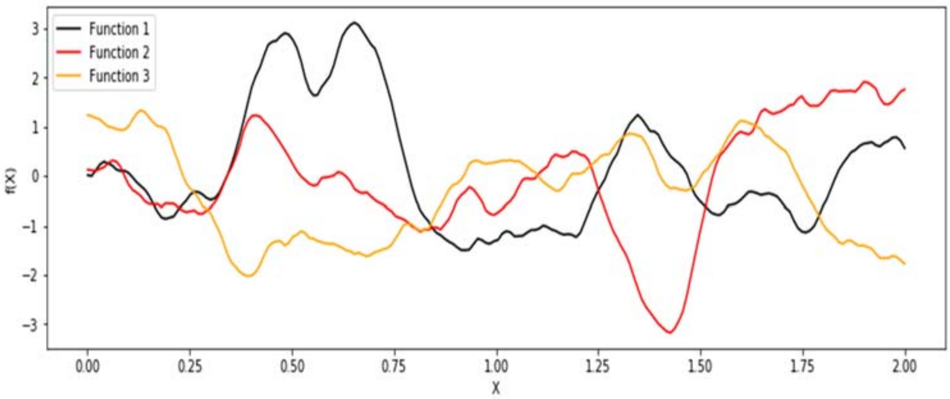
$$m(x) = \mathbb{E}[f(x)]$$
$$k(x, x') = \mathbb{E}\big[\big(f(x) - m(x)\big)\big(f(x') - m(x')\big)\big]$$
$$f(x) \sim \mathcal{GP}\big(m(x), k(x, x')\big)$$
$$f(x) \sim \mathcal{GP}\big(0, k(x, x')\big)$$

The definition imposes on the Gaussian process a consistency/marginalization. The distribution specified by the finite realization *x*1,. . .,*xn* must be consistent with any subset *x*11,. . .,*xn*1 (*i.e.* if we marginalize out the distribution of the full set with respect to the random variables not present in the subset, we recover the distribution of the subset). The consistency property of the random processes allows us to consider the Gaussian processes as probability distributions on function spaces, as it allows one to iteratively sample new points to the others already sampled.

The forms of the functions sampled from the Gaussian processes depend on the form of the mean and covariance functions. Usually the mean function $m(x)$ is set to 0 hence one samples functions from:

$$f(x) \sim \mathcal{GP}\big(0, k(x, x')\big)$$

Figure 1
Sample Functions Drawn from a Gaussian Process with Matern 3/2 Covariance



The properties of the functions sampled from the Gaussian process are controlled by the analytical form of the covariance function allowing one to express prior beliefs such as smoothness, periodicity, stationarity.

Stationary covariance functions. If the covariance function does not separately depend on x and x' but only on the distance $\tau = \|x - x'\|$ between the points we call the covariance function stationary. Gaussian processes with stationary covariance kernels sample functions which have the same properties across the whole domain. The most used stationary kernels are:

1. Squared exponential covariance:

   $$k_{SE(\tau)} = e^{-\tau^2/2l^2} \ ,$$

   where: l is the characteristic length controlling the interval needed for the sample function to significantly change;

2. Matern covariance:

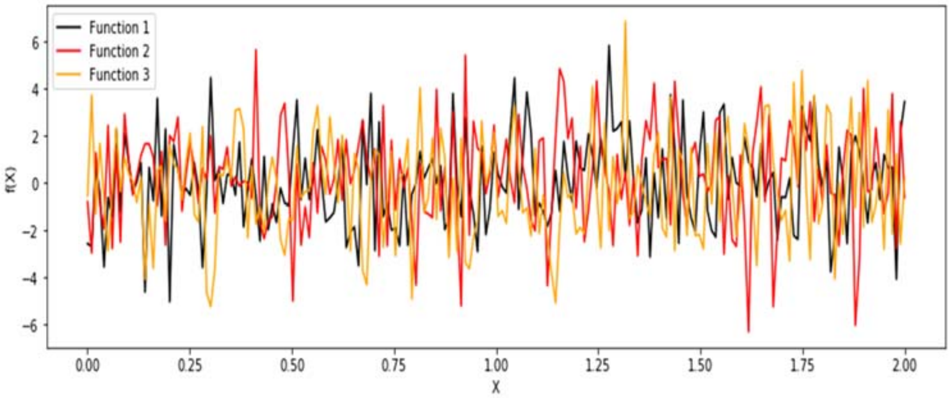
   $$k_{Matern}(x, x') = \frac{e2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu\tau}}{l}\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu\tau}}{l}\right),$$

   where: $K\nu$ is the modified Bessel function. The case ν = 3⁄2 corresponds to continuous version of the autoregressive model AR(1);

3. White noise covariance:

   $$k(x, x') = \sigma^2 I_{xx}.$$

Figure 2
Sample Functions Drawn from Gaussian Process with White Noise Covariance



Sample functions drawn from different covariance kernel are presented in Figures 1 and 2.

Gaussian process for regression. Following Rasmussen (2003) let the set of the known observations $(x\,,y\,) = (x\,,f\,(x)\,)$ and let *f** be the predictions for new points *x**. Suppose *f* and *f** are sampled from the same Gaussian process then we have:

$$f_*|x_*, x, f \sim \mathcal{N}\big(k(x_*\,,x\,)k(x\,,x\,)^{-1}f\,, k(x_*\,,x_*\,) - k(x_*,x\,)k(x\,,x\,)^{-1}k(x,x_*\,)\big) \tag{1}$$

The covariance of the predictive distribution described in equation 1 shows no dependence on the observed values of *f*. This means that the observed values of *f* only control the means of the predictions having no influence on the uncertainty estimation of the predictions.

The Gaussian process regression are a very flexible nonparametric method (the only constraints on the sampled functions are the weak ones imposed by the chosen form of the covariance functions). However, the prediction is sampled from a Gaussian distribution which is not very robust to outliers and the covariance of the sampled distribution is not influenced by the observed *f* which makes the uncertainty estimates not reliable.

2.2 Student T-processes

The Student t-processes are a generalization of Gaussian Processes being part of the larger class of elliptical process. While for a Gaussian process the marginal distribution is a multivariate normal distribution, the marginal distribution of a t-process is a multivariate Student distribution. In the limit of the infinite degrees of freedom the Student distribution converges to the multivariate gaussian distribution and in this limit the Student t-process are generalization of the Gaussian processes. We will describe below a fully Bayesian hierarchical derivation of the Student t-processes due to Shah (2014). Having a fully specified hierarchical Bayesian model will allow us to include it in the more complex hierarchies needed to model financial assets' returns.

Inverse Wishart Process. The Gaussian processes $f \sim \mathcal{G}P(0, K)$ suppose the specification of a covariance matrix whose propertiy is symmetry and positive definiteness.

The covariance matrices presented in the previous section are parametric forms specified by the covariance function. A full Bayesian treatment will suppose putting a prior on the possible covariance matrices, a probability distribution over the space Π(*n*) of all *nxn* symmetric, real valued and positive definite matrices.

An obvious choice of probability distribution over the space of covariance matrices would be the Wishart distribution. However as proved in Shah (2004), the Wishart distribution is not consistent under marginalization. Shah proposed the usage of the inverse Wishart distribution which is consistent under marginalization and is defined as:

Definition 2 A random matrix $\Sigma \in \Pi(n)$ is inverse Wishart distributed with parameters $\nu \in \mathbb{R}_+$ and $\mathrm{K} \in \Pi(n) p(\Sigma) \sim \mathrm{IW}_n(\nu, K)$ if its density function is:

$$p(\Sigma) = \mathrm{c}_n(\nu, K)|\Sigma|^{-(\nu+2n)/2} exp\left[-\frac{1}{2} Tr(K\Sigma^{-1})\right],$$

where:

$$\mathrm{c}_n(\nu, K) = \frac{|\mathrm{K}|^{(\nu+n-1)/2}}{2^{(\nu+n-1)/2}\Gamma_n[(\nu+n-1)/2]}.$$

The inverse Wishart prior is equivalent to the Wishart matrix prior because $\Sigma \sim W(\mu, K)$ if and only if $\Sigma^{-1} \sim IW(\mu, K)$. The inverse Wishart distribution allows one to construct the inverse Wishart process defined as:

Definition 3 σ is an inverse Wishart process on $\mathcal{X}$ with parameters $\mu \in \mathbb{R}_+$ and a base kernel $k: \mathcal{X} x \mathcal{X} \to \mathbb{R}$ if for any finite collection x1,. . .,xn $\in \mathcal{X}$ we have $\sigma(x_1, \dots, x_n) \sim IW(\nu, K)$ where $k_{ij} = k(x_i, x_j)$. We say that $\sim IWP(\nu, K)$.

Student processes as hierarchical gaussian. Shah *et al.* ([6]) propose the following 2 stage hierarchical model for a base kernel *k*0 and a continuous mean function $\Phi: \mathcal{C}\mathcal{X} \to \mathbb{R}$:

$$\sigma \sim IWP(\nu, k_0);$$

$$f|\sigma \sim \mathcal{G}P(\phi, (\nu-2)\sigma). \quad (2)$$

The conjugacy of inverse Wishart prior to a gaussian likelihood allow Shah et all to derive from the posterior distribution:

$$p(f|\mu, K) \sim MVT_n(\mu, \phi, \mathrm{K}),$$

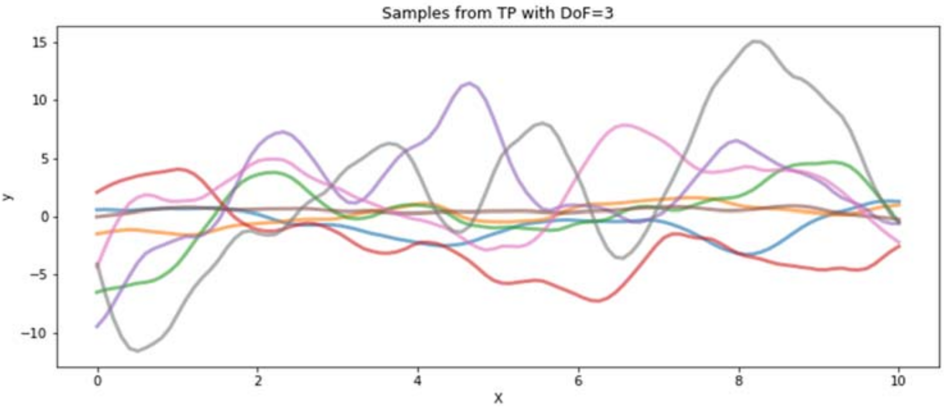
where: *MV Tn* denotes the multivariate Student t-distribution. From the above one can define the Student t-process as:

Definition 4 f is a Student t-process on $\mathcal{X}$ with parameters μ > 2 and $\Phi: \mathcal{X} \to \mathbb{R}$ and kernel function $k: \mathcal{X} x \mathcal{X} \to \mathbb{R}$ if any finite collection of function values is a multivariate Student t-process. We write:

$$f \sim TP(\mu, \Phi, k).$$

Figure 3

Sample functions drawn from a Student process 3 degrees of freedom with Matern 3/2 covariance



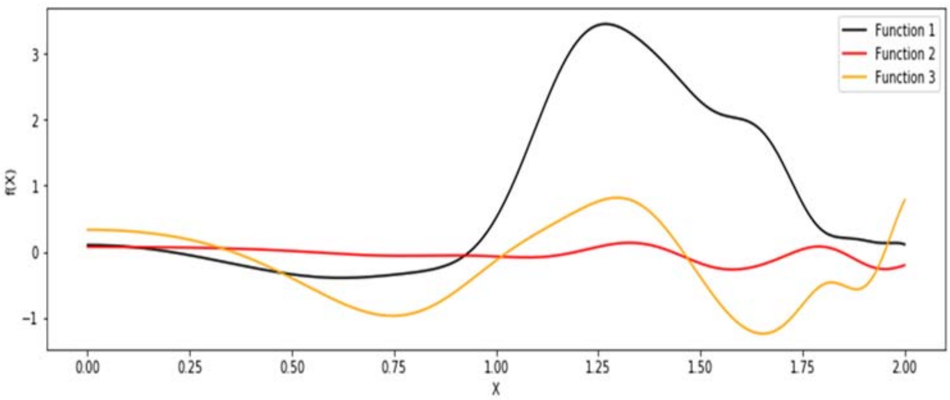
In Figure 3 we illustrate a few sample functions drawn from Student t-process. The fat tails of t multivariate distribution results in a greater variability of sample functions compared to the ones drawn from the equivalent Gaussian process.

2.3 Power Transforms

Power transforms are introduced to address the non-gaussianity of positive data. A standard approach in financial stochastic theory is to apply the logarithm to the positive non gaussian data.

Figure 4

Sample functions drawn from a warped Student process 3 degrees of freedom with Matern 3/2 covariance



The Box-Cox transformation are a generalization of the logarithmic transform and are widely used in the statistics literature (see for example Bickel (1981)). The Box-Cox transform of parameter λ are defined as:

$$\phi_\lambda(x) = \frac{sgn(x)|x|^\lambda - 1}{\lambda}.$$

In [7] is shown that even if the transformed data is not gaussian at least the first moments are symmetrical.

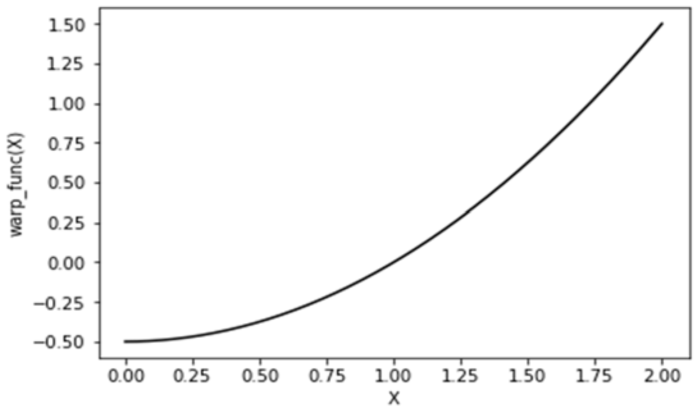
In this paper we will use the Box-Cox transform as input warping function for the Gaussian and Student t-processes, An input warping function is of the form:

$$\phi(x) \sim \mathcal{GP}(m, K)) \text{ or } \phi\ (x) \sim \mathcal{TP}(m, \Phi, K).$$

By using a warping function, the covariance function becomes non-stationary as MacKay shown in MacKay (2003). This property makes desirable the use of warping function in the analysis of non-stationary time series as it allows the use of stationary covariance functions to model non=stationary series. In Figure 5, the Box Cox transform is presented and samples from a warped Student process are presented in Figure 4. As one may see in Figure 4, the sampled functions are changing shape dependent on variable *x*.

Figure 5

Box Cox power transform



3. Modeling Financial Time Series Student T-processes

We set out to use the Gaussian and Student t-processes to model the evolution of the SP500. The objective of the modeling exercise is obtaining predictions of the future unknown evolution of the index at the same time with quantifying the uncertainty in making those predictions. To get uncertainty estimates necessary for establishing a a grounded and robust approach to forecasting we will specify a full Bayesian hierarchical model in which Gaussian or Student processes are a vital components The choice of covariance functions is critical, as we have seen above, to the shape of the function sampled and it will be guided by a set of stylized facts about the evolution of financial assets returns collected from the literature

3.1 Stylized Facts

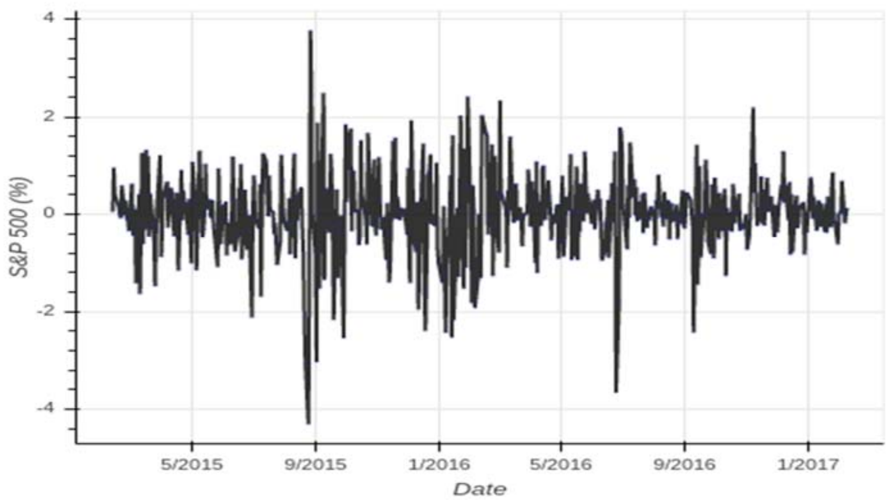
Stylized facts are obtained by taking the common denominator among the properties observed in the studies of different markets and instruments. This means that stylized facts are presented in terms of qualitative properties.

In his survey of the financial econometrics' literature, Rama Cont (2011) has identified the following set of stylized facts

1. **Non-stationarity.** The stationarity hypothesis of financial asset returns affirms that for any set of times $t1, \ldots, tk$ the distribution of returns along with its moments is largely unchanged. This hypothesis is largely believed to be untrue as we can see from Figure 5 where we can see that there are periods of high and low volatility (the standard deviation is not constant).

Figure 6

Evolution of S&P 500 returns between February 2015 and February 2017



2 **Fat tails.** The fat tails phenomenon is concerned with the probability mass that is to be found in the tails of the probability distribution (extreme events far away from the mean). The presence or the absence of the fat tails is judged relative to the tails of the normal distribution (there is no expectation to find significant probability mass 2 standard deviations away from the mean in the case of a normal distribution) and is measured kurtosis. A value significantly higher than 3 (the value for a normal distribution) indicates the fat tail phenomenon.

3 **No linear autocorrelations.** Linear autocorrelations are often insignificant except for small time scales (high frequency). The absence of autocorrelations means that the classical time series models like ARMA cannot distinguish between white noise and asset returns. The dependence of asset returns is nonlinear.

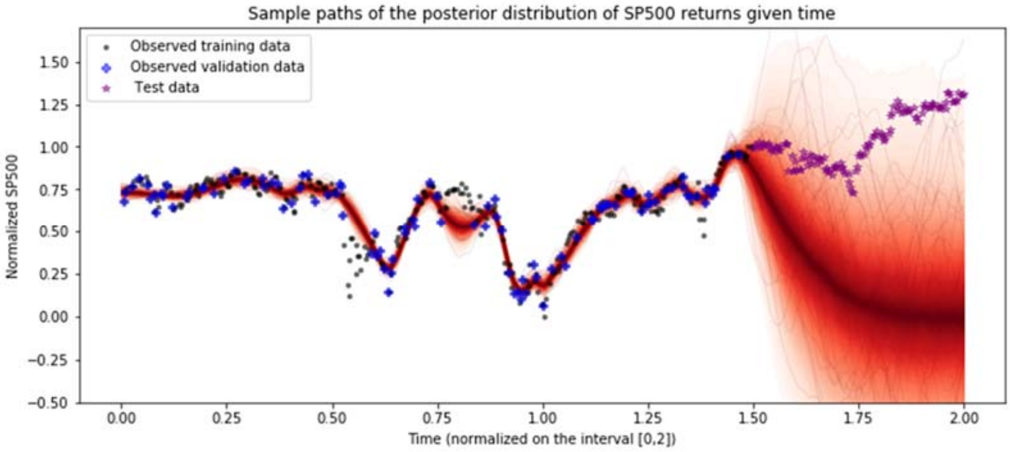
3.2 Modeling SP500 with Input Warped T-process

Given the stylize facts presented below we propose a hierarchical Bayesian model in which we use:

- A warped Student t-process to account for heavy tails and non-normality of financial asset evolution. The functions drawn from the process are the filtered (de-noised) evolution of financial returns.
- The covariance function of the Student t-process is a squared exponential. The length scale of the covariance function has a gamma prior distribution.
- The number of degrees of freedom *v* has a positive Cauchy prior distribution.

Figure 7

Bayesian model of S&P 500 returns with normal likelihood and a warped Student t-process with square exponential covariance function.



- A Box-Cox warping function to account for the non-stationarity in the data. The Box-Cox transformation account for the non-constant variance stylized fact discussed above. The parameter λ which controls the shape of the warping function has a positive Cauchy prior distribution.
- An observational model with normally distributed noise. The standard deviation of the observational noise a positive Cauchy prior distribution. This accounts for the bursty behavior.

We can observe that the majority of prior distributions are positive Cauchy distributions. We choose this distribution to express our preference for small values of the parameters (the mode of the distribution is 0) while also allowing for larger values of the parameters (non-zero probability on the $[0,\infty)$) should the data require.

The full hierarchical Bayesian model is presented below:

$$\lambda \sim \mathcal{C}^{+}(\beta = 3);$$

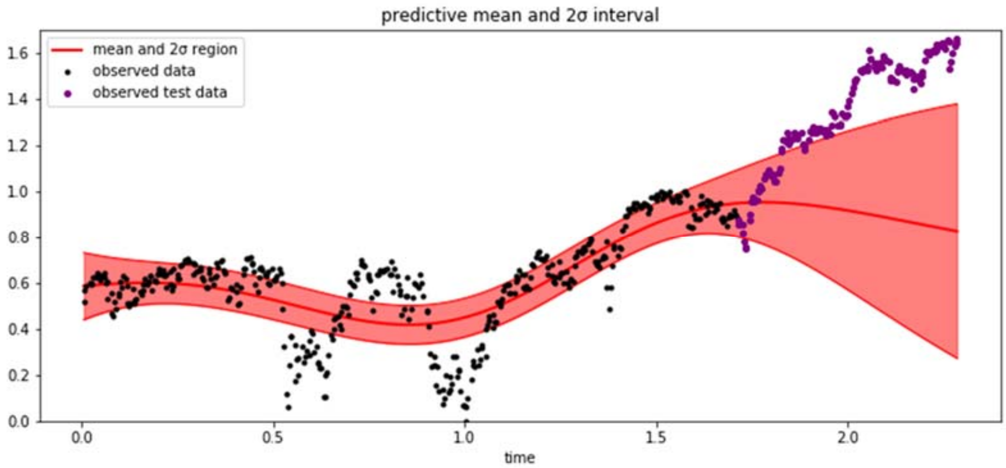
$$\phi_{\lambda}(x) = \frac{sgn(x)|x|^{\lambda} - 1}{\lambda};$$

$$\nu \sim \mathcal{C}^{+}(\beta = 3);$$

$$l \sim \mathcal{G}(\alpha = 2, \beta = 2);$$
$$k_{SE(\tau)} = e^{-\tau^2/2l^2};$$
$$\phi_\lambda(x) \sim TP\big(0, \nu, k_{SE}(\tau)\big);$$
$$eps \sim \mathcal{C}^+(\beta = 3);$$
$$y \sim \mathcal{N}(\phi\,(x), eps).$$

Figure 8

Bayesian model of S&P 500 returns with normal likelihood and a Gaussian process with square exponential covariance function.



By using a Hamiltonian Monte Carlo, we extract 1000 possible evolutions (1000 sample functions from the model). These sample functions will show us to what extent the fitted model recovers the SP500 dynamics. The sampled functions will localize around the observed values of SP500 and spread when the uncertainty is high.

As one may see in Figure 7, when the evolution is predictable the sampled function will localize giving precise estimates with low standard deviations and will spread around the space when its predictions are of low confidence (as is the case with the upward trend in the latter part of the evolution). The sampled functions are non-linear and follow the quasi-periodicity in the data.

Even with the non-stationary evolution the model can still with some probability forecast the correct evolution.

For reference we also fitted a Gaussian process but as we can see in Figure 8 the model is not robust to fat tails, a majority of points observed or not being situated outside the model’s confidence bounds.

5. Conclusions

In this paper we presented a full Bayesian hierarchical model based on Student t-process to account for the stylized facts in the evolution of financial assets returns. Although the fitted model was not perfect, we managed to recover the fundamental stylized facts of the evolution: nonlinear dependencies, fat tails and non-stationarity. As we future work we propose including exogenous inputs to improve the predictive quality of the model.

References

Rasmussen, C.E., 2003, February. Gaussian processes in machine learning. In *Summer School on Machine Learning* (pp. 63-71). Springer, Berlin, Heidelberg.

Jylänki, P., Vanhatalo, J. and Vehtari, A., 2011. Robust Gaussian process regression with a Student-t likelihood. *Journal of Machine Learning Research*, *12*(Nov), pp.3227-3257. Dolado, J.J. and Lütkepohl, H., 1996. Making Wald Test Work for Cointegrated VAR Systems. *Econometric Reviews*, 15, pp.369-386.

Vanhatalo, J., Jylänki, P. and Vehtari, A., 2009. Gaussian process regression with Student-t likelihood. In *Advances in neural information processing systems* (pp. 1910-1918).

Rios, G. and Tobar, F., 2018, July. Learning non-Gaussian time series using the Box-Cox Gaussian process. In *2018 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.

Lázaro-Gredilla, M., 2012. Bayesian warped Gaussian processes. In: *Advances in Neural Information Processing Systems*, pp. 1619-1627.

Shah, A., Wilson, A. and Ghahramani, Z., 2014, April. Student-t processes as alternatives to Gaussian processes. In *Artificial intelligence and statistics* (pp. 877-885).

Bickel, P.J. and Doksum, K.A., 1981. An analysis of transformations revisited. *Journal of the american statistical association*, *76*(374), pp.296-311.

MacKay, D.J. and Mac Kay, D.J., 2003. *Information theory, inference and learning algorithms*. Cambridge University Press.

Cont, R., 2001. *Empirical properties of asset returns: stylized facts and statistical issues*.