1. Introduction

Motivation: The prediction of time-changing variances is important when modeling financial time series. Relevant applications range from the estimation of financial risk to portfolio construction and derivative pricing.

Problem: Standard econometric models, such as GARCH, are limited as they assume fixed, usually linear, functional relationships for the evolution of the variance.

Solution: We model the unknown functional relationship in our new GP-Vol model using Gaussian Processes. We perform fast online inference by adapting particle filters.

2. Characteristics of Financial Volatility

Financial returns exhibit:
- Time-dependent standard deviation or volatility.
- Volatility clustering.
- Asymmetric effect on volatility from positive and negative returns.

3. Standard Models

GARCH is the best known and most popular volatility model [1]:

\[ \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2. \]  

(1)

GARCH can model a) time-dependence and b) volatility clustering, but does not account for the asymmetric effect of positive and negative returns on volatility. GARCH variants, such as EGARCH [2] and GJR-GARCH [3], do account for the asymmetric effect, but they are still limited by assuming a linear transition function for the volatility.

4. Gaussian Process Volatility Model

Our contribution is the new GP-Vol model:

\[ x_t \sim N(0, \sigma_t^2) \quad \text{and} \quad v_t := \log(\sigma_t^2) = f(v_{t-1}, x_t) + \epsilon_t, \]  

(2)

where \( \epsilon_t \sim N(0, \sigma^2) \).

We place a Gaussian Process (GP) prior on the transition function \( f \).

Advantages of GP-Vol:
- GP-Vol Models the unknown transition function in a non-parametric manner.
- GP-Vol reduces the risk of overfitting by following a full Bayesian approach.

5. GP-Vol Graphical Model

GP-Vol is a Gaussian Process state space model (GP-SSM), a generalization of HMM in which the transition function is unknown and represented by a GP.

6. Learning with Particle Filters

We develop a new algorithm called Regularized Auxiliary Chain Particle filter (RAPCF) based on [4]. RAPCF has similar performance to the state-of-the-art method Particle Gibbs with Ancestor Sampling (PGAS) [5], but RAPCF is much faster.

Algorithm 1 RAPCF

1. Input: data \( z_{1:T} \), number of particles \( N \), shrinkage parameter \( \lambda > 0 \), initial states \( \theta_0 \), prior \( \pi(\theta) \).
2. Sample \( N \) parameter particles from the prior: \( \theta_1^{(1)}, \ldots, \theta_1^{(N)} \sim \pi(\theta) \).
3. Set initial importance weights, \( W_0 = 1/N \).
4. for \( t = 1 \) to \( T \) do
   5. Shrink parameter towards their empirical mean \( \theta_{t-1} = \frac{1}{N} \sum_{i=1}^{N} \theta_{t-1}^{(i)} \).
   6. Compute the new expected states: \( v_t^i = \mathbb{E}(v_t | Q_t, \theta_{t-1}) \).
   7. Compute importance weights proportional to the likelihood of the new expected states:
   \[ W_t^i = \frac{1}{W_{t-1}^i} \rho(v_t^i), \]  
   \[ \rho(v_t^i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(v_t^i - \mu(v_t^i))^2}{2\sigma^2}\right). \]  
   (3)
   8. Resample \( N \) auxiliary indices \( j_t \) according to weights \( \pi(v_t^j) \).
   9. Propagate the corresponding chains of hidden states forward, that is, \( v_{t+1}^{j_t} \).
   10. Add jitter: \( \theta_t^j = (1 - \lambda)\theta_{t-1} + \lambda v_{t+1}^{j_t} \), where \( \lambda \) is the empirical covariance of \( \theta_{t-1} \).
   11. Propose new states \( v_t^j \sim \pi(v_t^j | \theta_t^j) \).
   12. Compute importance weights adjusting for the modified proposal:
   \[ W_t^j = \frac{1}{W_{t-1}^j} \rho(v_t^j), \]  
   \[ \rho(v_t^j) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(v_t^j - \mu(v_t^j))^2}{2\sigma^2}\right). \]  
   (4)
   13. end for
   14. Output: particles for chains of states \( v_{t+1}^{j_t} \), particles for parameters \( \theta_t \) and particle weights \( W_t^j \).

7. Experiment Setup

Data:
- 50 time series, consisting of 20 daily FX and 30 daily Equity returns.
- Each time series contains 780 observations, from Jan 2008 - Jan 2011.
- Each time series was normalized to have zero mean and unit standard deviation.

Model comparison:
- The models compared were GP-Vol, GARCH(1,1), EGARCH(1,1) and GJR-GARCH(1,1).
- Each model receives an initial time series of length 100.
- The models are trained on the training data and a one-step forward prediction is made.
- Then the training data is augmented with a new observation and the process is repeated.
- For GP-Vol, an RAPCF with \( N = 200 \) and \( \lambda = .95 \) was used.

8. Comparison of Model Predictive Performance

We perform a multiple comparison test, where all the methods are ranked according to their performance on the 50 time series or tasks. The average ranks are then tested for statistical significant differences. GP-Vol is the best model with significant evidence.

9. RAPCF vs. PGAS

RAPCF has similar predictive performance as the state-of-the-art method PGAS (see paper), but RAPCF is much faster.

10. Predicted Volatility Surface

Surface generated by plotting the mean predicted outputs \( v_t \) against a grid of inputs for \( v_{t-1} \) and \( v_{t-2} \).

11. References