Gaussian Process Volatility Model

Yue Wu, José Miguel Hernández-Lobato, and Zoubin Ghahramani
Engineering Department, University of Cambridge, United Kingdom.

UNIVERSITY OF CAMBRIDGE

Neural Information Processing Systems Foundation

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.

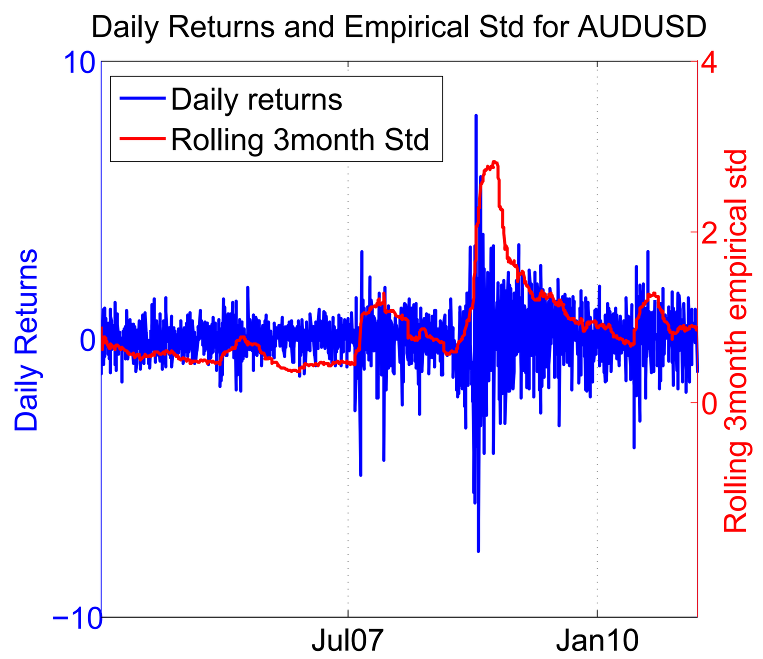
Problem: Standard econometric models, such as **GARCH**, are limited as they assume fixed, usually **linear**, functional relationships for the evolution of the variance.

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

$$x_t \sim \mathcal{N}(0, \sigma_t^2) \text{ and } \sigma_t^2 = \alpha_0 + \sum_{j=1}^{q} \alpha_j x_{t-j}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \,. \quad (1)$$

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 all still limited by assuming a **linear transition function** for the volatility.

4. Gaussian Process Volatility Model

Our contribution is the new **GP-Vol** model:

$$x_t \sim \mathcal{N}(0, \sigma_t^2) \text{ and } v_t := \log(\sigma_t^2) = f(v_{t-1}, x_{t-1}) + \epsilon_t \,, \text{where } \epsilon_t \sim \mathcal{N}(0, \sigma_n^2) \,. \quad (2)$$

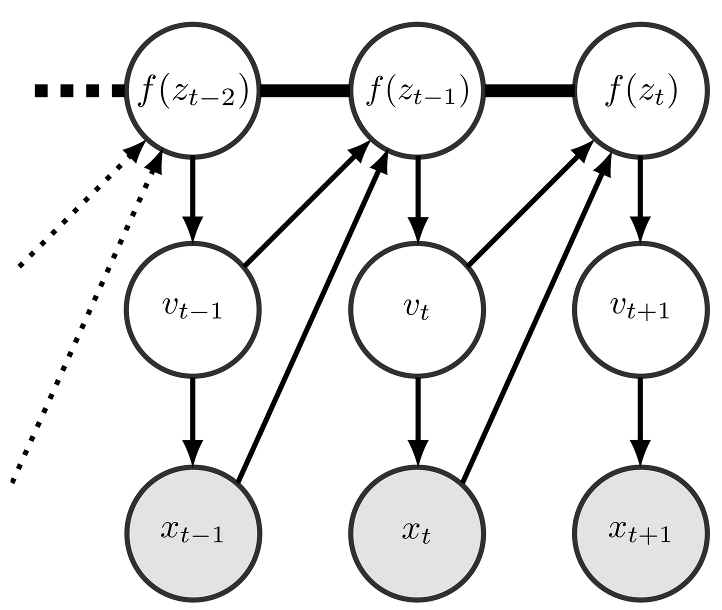
We place a **Gaussian Process** (GP) prior on the the transition function f.

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

- GARCH, EGARCH and GJR-GARCH are all Hidden Markov models (**HMM**). The dynamic variances are the hidden states.
- 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 $x_{1:T}$, number of particles N, shrinkage parameter $0 < \lambda < 1$, prior $p(\boldsymbol{\theta})$.
2: Sample N parameter particles from the prior: $\{\boldsymbol{\theta}_0^i\}_{i=1,\ldots,N} \sim p(\boldsymbol{\theta})$.
3: Set initial importance weights, $W_0^i = 1/N$.
4: **for** $t = 1$ **to** T **do**
5: Shrink parameter particles towards their empirical mean $\bar{\boldsymbol{\theta}}_{t-1} = \sum_{i=1}^N W_{t-1}^i \boldsymbol{\theta}_{t-1}^i$ by setting

$$\widetilde{\boldsymbol{\theta}}_t^i = \lambda \boldsymbol{\theta}_{t-1}^i + (1-\lambda)\bar{\boldsymbol{\theta}}_{t-1} \,. \quad (6)$$

6: Compute the new expected states:

$$\mu_t^i = \mathbb{E}(v_t | \widetilde{\boldsymbol{\theta}}_t^i, v_{1:t-1}^i, x_{1:t-1}) \,. \quad (7)$$

7: Compute importance weights proportional to the likelihood of the new expected states:

$$g_t^i \propto W_{t-1}^i p(x_t | \mu_t^i, \widetilde{\boldsymbol{\theta}}_t^i) \,. \quad (8)$$

8: Resample N auxiliary indices $\{j\}$ according to weights $\{g_t^i\}$.
9: Propagate the corresponding chains of hidden states forward, that is, $\{v_{1:t-1}^j\}_{j \in J}$.
10: Add jitter: $\boldsymbol{\theta}_t^j \sim \mathcal{N}(\widetilde{\boldsymbol{\theta}}_t^j, (1-\lambda^2)\mathbf{V}_{t-1})$, where $\mathbf{V}_{t-1}$ is the empirical covariance of $\boldsymbol{\theta}_{t-1}$.
11: Propose new states $v_t^j \sim p(v_t | \boldsymbol{\theta}_t^j, v_{1:t-1}^j, x_{1:t-1})$.
12: Compute importance weights adjusting for the modified proposal:

$$W_t^j \propto p(x_t | v_t^j, \boldsymbol{\theta}_t^j) / p(x_t | \mu_t^j, \widetilde{\boldsymbol{\theta}}_t^j) \,, \quad (9)$$

13: **end for**
14: **Output:** particles for chains of states $v_{1:T}^j$, particles for parameters $\boldsymbol{\theta}_t^j$ 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.

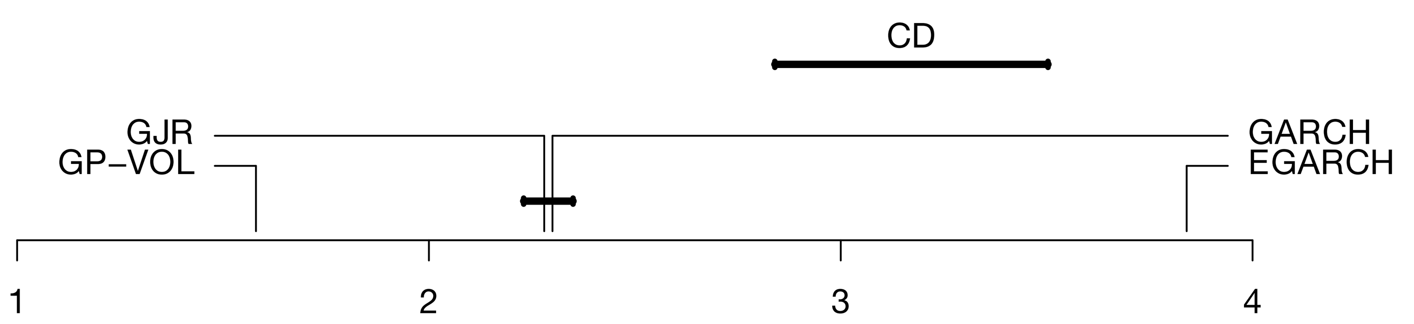
Model comparison:

- 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.

Nemenyi Test

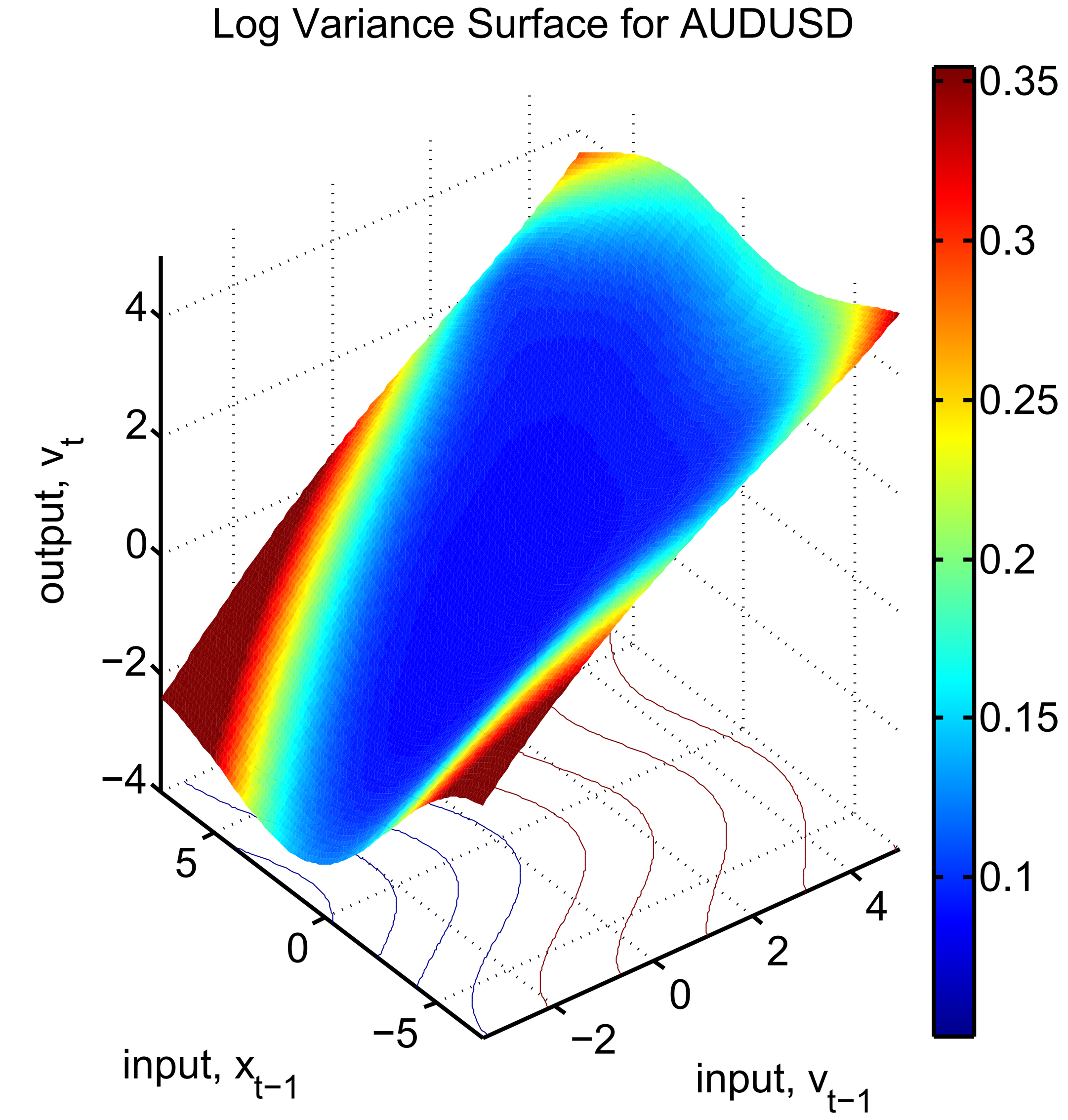


9. RAPCF vs. PGAS

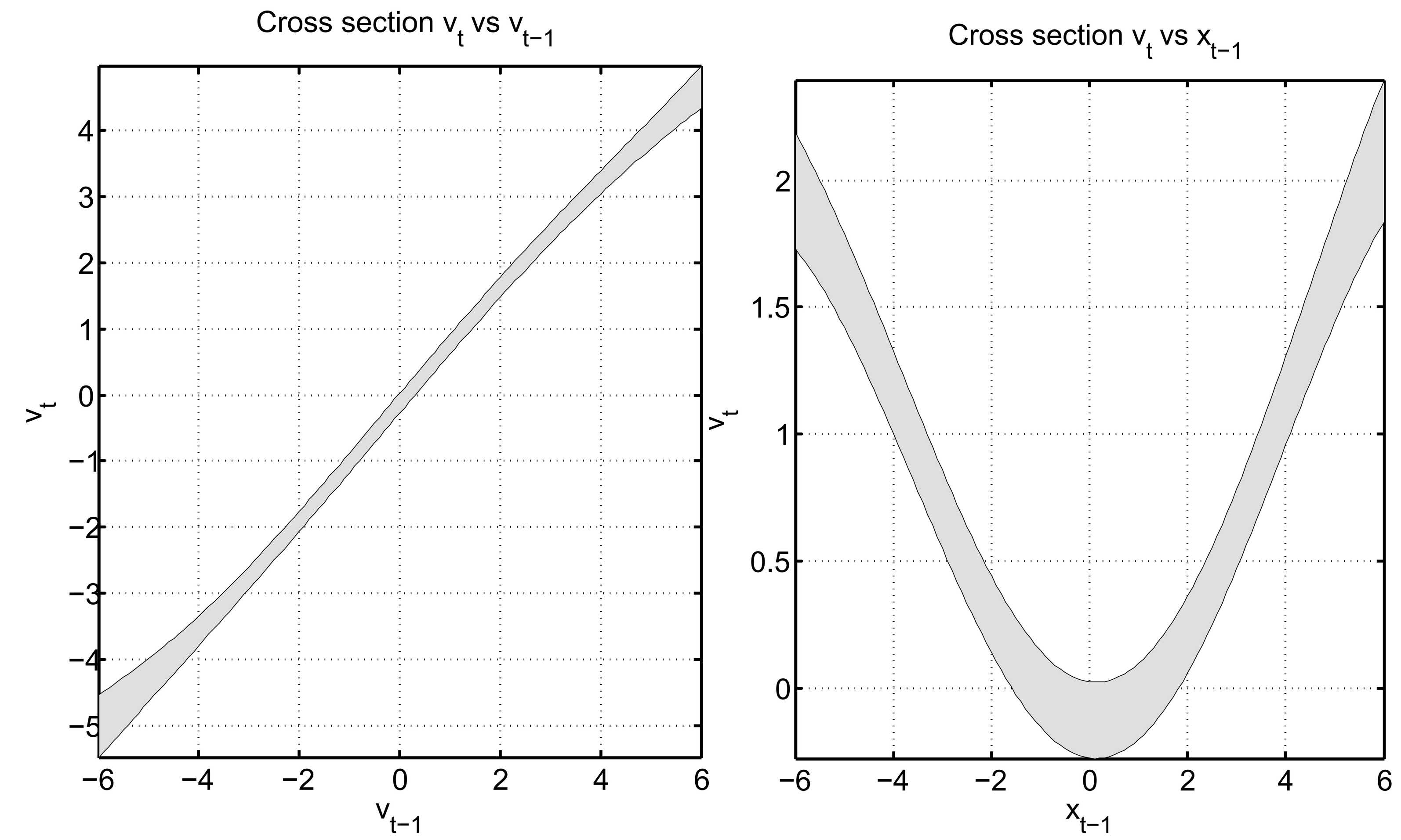
RAPCF has similar predictive performance as the state-of-the-art method PGAS (see paper), but RAPCF but is much faster.

Method	Configuration	Avg. Time
RAPCF	$N = 200$	6 min.
PGAS	$N = 10, M = 100$	732 min.

10. Predicted Volatility Surface



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



Left, predicted $v_t \pm 2$ s.d. for inputs $(0, x_{t-1})$. Right, predicted $v_t \pm 2$ s.d. for inputs $(0, x_{t-1})$.

11. References

[1] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

[2] D.B. Nelson.Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370, 1991.

[3] L.R. Glosten, R. Jagannathan, and D.E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801, 1993.

[4] J. Liu and M. West. Combined parameter and state estimation in simulation-based filtering. Institute of Statistics and Decision Sciences, Duke University, 1999.

[5] F. Lindsten, M. Jordan, and T. Schön. Ancestor Sampling for Particle Gibbs. In *Advances in Neural Information Processing Systems 25*, pages 2600–2608, 2012.