Recurrent Latent Variable Conditional Heteroscedasticity

Sotirios P. Chatzis

Department of Electrical Engineering, Computer Engineering, and Informatics
Cyprus University of Technology sotirios.chatzis@cut.ac.cy

Introduction

- Generalized autoregressive conditional heteroscedasticity (GARCH) models are one of the most successful families of approaches for volatility modeling in financial return signals.
- However, they employ quite rigid assumptions regarding the evolution of the variance.
- We address these issues by introducing a recurrent latent variable model, capable of capturing highly flexible functional relationships for the variances.
- We derive a fast, scalable, and robust to overfitting Bayesian inference algorithm.
- Our approach avoids the need to compute per-data point variational parameters, but can instead compute a set of global variational parameters valid for inference at both training and test time.

Motivation

- The changes in the log price of financial market indices (*returns*) may be non-linear, non-stationary and/or heavy-tailed, while their marginal distributions may be asymmetric, leptokurtic and/or show conditional heteroscedasticity.
- GARCH-type models make a specific assumption of what the functional dynamics of the volatility look like; in reality, this functional form is completely unknown.
- Hence, we need a new modeling paradigm that *infers this functional form* from the data.
- To this end, we leverage recent advances in the field of deep learning, namely *deep generative models* treated under the *amortized variational inference* (AVI) paradigm.

Deep Generative Models with Amortized Variational Inference

- Let us consider a dataset $X = \{\boldsymbol{x}_n\}_{n=1}^N$ consisting of N samples of some observed random variable $\boldsymbol{x}$; here, the modeled data constitute time-series signals of asset returns.
- We assume that the observed random variable is generated by some random process, involving an unobserved continuous random variable $\boldsymbol{z}$.
- We introduce a conditional independence assumption for the observed variables $\boldsymbol{x}$ given the corresponding latent variables $\boldsymbol{z}$; we adopt the conditional likelihood function $p(\boldsymbol{x}|\boldsymbol{z};\boldsymbol{\theta})$.
- To perform Bayesian inference for the postulated model, we impose some prior distribution $p(\boldsymbol{z};\boldsymbol{\varphi})$.
- We yield the following evidence lower bound (ELBO) expression:

$$\log p(X) \geq \mathcal{L}(\boldsymbol{\theta},\boldsymbol{\varphi},\boldsymbol{\phi}|X) = \sum_{i=1}^{N}\Big\{ -\mathrm{KL}\big[q(\boldsymbol{z}_i;\boldsymbol{\phi})||p(\boldsymbol{z}_i;\boldsymbol{\varphi})\big] + \mathbb{E}_{q(\boldsymbol{z}_i;\boldsymbol{\phi})}[\log p(\boldsymbol{x}_i|\boldsymbol{z}_i;\boldsymbol{\theta})]\Big\} \quad (1)$$

  where $\mathrm{KL}\big[q||p\big]$ is the KL divergence, $q(\boldsymbol{z};\boldsymbol{\phi})$ is the sought approximate (variational) posterior over the latent variable $\boldsymbol{z}$, while $\mathbb{E}_{q(\boldsymbol{z};\boldsymbol{\phi})}[\cdot]$ is the (posterior) expectation of a function w.r.t. the random variable $\boldsymbol{z}$, the distribution of which is taken to be the posterior $q(\boldsymbol{z};\boldsymbol{\phi})$.
- AVI assumes that the likelihood function and the resulting latent variable posterior, $q(\boldsymbol{z};\boldsymbol{\phi})$, are parameterized via deep neural networks (DNs).
- This is a non-conjugate construction; hence, $\mathbb{E}_{q(\boldsymbol{z}_i;\boldsymbol{\phi})}[\log p(\boldsymbol{x}_i|\boldsymbol{z}_i;\boldsymbol{\theta})]$ and its gradient are intractable.
- AVI resolves these issues by drawing random samples of $\boldsymbol{z} \sim q(\boldsymbol{z};\boldsymbol{\phi})$, which are reparameterized via an appropriate differentiable transformation of an (auxiliary) random noise variable $\boldsymbol{\epsilon}$:

$$\mathcal{L}(\boldsymbol{\theta},\boldsymbol{\varphi},\boldsymbol{\phi}|X) = \sum_{i=1}^{N}\Big\{ -\mathrm{KL}\big[q(\boldsymbol{z}_i;\boldsymbol{\phi})||p(\boldsymbol{z}_i;\boldsymbol{\varphi})\big] + \frac{1}{L}\sum_{l=1}^{L}\log p(\boldsymbol{x}_i|\boldsymbol{z}_i^{(l)};\boldsymbol{\theta})\Big\} \quad (2)$$

- Specifically, considering a Gaussian posterior of the form

$$q(\boldsymbol{z}_i;\boldsymbol{\phi}) = \mathcal{N}(\boldsymbol{z}_i|\boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{x}_i), \mathrm{diag}\,\boldsymbol{\sigma}^2_{\boldsymbol{\phi}}(\boldsymbol{x}_i)) \quad (3)$$

  we have:

$$\boldsymbol{z}_i^{(l)} = \boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{x}_i) + \boldsymbol{\sigma}_{\boldsymbol{\phi}}(\boldsymbol{x}_i)\cdot\boldsymbol{\epsilon}_i^{(l)} \quad (4)$$

  In Eq. (4), $\boldsymbol{\epsilon}_i^{(l)}$ is white random noise, $\boldsymbol{\epsilon}_i^{(l)} \sim \mathcal{N}(\boldsymbol{0},\boldsymbol{I})$, the $\boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{x}_i)$ and $\boldsymbol{\sigma}^2_{\boldsymbol{\phi}}(\boldsymbol{x}_i)$ are parameterized via deep neural networks, and $\mathrm{diag}\,\boldsymbol{\chi}$ is a diagonal matrix with $\boldsymbol{\chi}$ on its main diagonal.
- Nevertheless, the employed *diagonal Gaussian assumption is quite limiting*.
- To allow for capturing the true model posterior, we adopt the principle of *normalizing flows*.
- We postulate the *auxiliary* latent variables $\boldsymbol{z}_i'$, for which we consider that the Gaussian assumption regarding their posterior is accurate.
- We perform a series of *invertible* transforms, $\{f_k(\cdot)\}_{k=1}^K$, that converts the *auxiliary* latent variables $\boldsymbol{z}_i'$ to the original ones, $\boldsymbol{z}_i$.
- This yields a tractable, non-Gaussian posterior over them, $q(\boldsymbol{z}_i)$, which reads

$$\log q(\boldsymbol{z}_i) = \log q(\boldsymbol{z}_i') - \sum_k \log\det|\nabla f_k| \quad (5)$$

Proposed Approach: The ReLaVaCH model

- On this basis, ReLaVaCH postulates a conditional independence assumption, where the conditioning variables $\boldsymbol{z}_n$ are some latent variables defined in a D-dimensional space with support in $\mathbb{R}$:

$$x_n|\boldsymbol{z}_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\sigma_n^2) \quad (6)$$

  where

$$\sigma_n^2 = g_{\boldsymbol{\theta}}(\boldsymbol{z}_n) \quad (7)$$

  and $g_{\boldsymbol{\theta}}(\cdot)$ is a deep neural network (DN) comprising *rectified* linear units, with parameters set $\boldsymbol{\theta}$.
- We consider a latent variables prior that captures the temporal dynamics of volatility:

$$\boldsymbol{z}_n \sim \mathcal{N}(\tilde{\boldsymbol{m}}_n, \mathrm{diag}(\tilde{\boldsymbol{s}}_n^2)) \quad (8)$$

  where

$$[\tilde{\boldsymbol{m}}_n; \tilde{\boldsymbol{s}}_n^2] = g_{\boldsymbol{\varphi}}(\boldsymbol{\rho}_{n-1}) \quad (9)$$

  while $g_{\boldsymbol{\varphi}}(\cdot)$ is a DN comprising *rectified* linear units, with parameters set $\boldsymbol{\varphi}$.
- Here, $\boldsymbol{\rho}_{n-1}$ is a *state vector* that encodes the history of observed return values, $\{x_\tau\}_{\tau=1}^{n-1}$, and inferred latent vectors, $\{\boldsymbol{z}_\tau\}_{\tau=1}^{n-1}$, in the form of a high-dimensional representation:

$$\boldsymbol{\rho}_\tau = r([r_x(x_\tau);\ r_z(\boldsymbol{z}_\tau);\ \boldsymbol{\rho}_{\tau-1}]) \quad (10)$$

  where $r(\cdot)$, $r_x(\cdot)$, and $r_z(\cdot)$ are DNs composed of *rectified* linear units.
- Hence, ReLaVaCH variational posterior over $\boldsymbol{z}_n$ will be a function of both the current observation, x_n, as well as the recurrently-generated high-dimensional history representation, $\boldsymbol{\rho}_{n-1}$, $\forall n$.
- To allow for inferring the true underlying posterior over the $\boldsymbol{z}_n$, we postulate the *auxiliary latent variables* $\boldsymbol{z}_n' \in \mathbb{R}^D$, which we assume that yield an (accurate) Gaussian posterior of the form:

$$p(\boldsymbol{z}_n'|\boldsymbol{x}_n,\boldsymbol{h}_{n-1};\boldsymbol{\phi}) = \mathcal{N}(\boldsymbol{z}_n'|\hat{\boldsymbol{m}}_n, \mathrm{diag}(\hat{\boldsymbol{s}}_n^2)) \quad (11)$$

$$[\hat{\boldsymbol{m}}_n; \hat{\boldsymbol{s}}_n^2] = g_{\boldsymbol{\phi}}([\boldsymbol{x}_n; \boldsymbol{\rho}_{n-1}]) \quad (12)$$

- Then, we assume that the original postulated latent variables, $\boldsymbol{z}_n \in \mathbb{R}^D$, can be obtained by transforming the auxiliary ones, $\boldsymbol{z}_n'$, via a series of *planar normalizing flows* of the form:

$$f_k(\boldsymbol{z}) = \boldsymbol{z} + \boldsymbol{u}_k h(\boldsymbol{w}_k^T\boldsymbol{z} + b_k) \quad (13)$$

  This way, application of (5) yields the following posterior over the $\boldsymbol{z}_n \in \mathbb{R}^D$:

$$\log p(\boldsymbol{z}_n|\boldsymbol{x}_n,\boldsymbol{h}_{n-1};\boldsymbol{\phi}) = \log p(\boldsymbol{z}_n'|\boldsymbol{x}_n,\boldsymbol{h}_{n-1};\boldsymbol{\phi}) - \sum_k \log|1 + \boldsymbol{u}_k^T\psi_k(\boldsymbol{z}_n^k)| \quad (14)$$

  where $\boldsymbol{z}_n^k \triangleq f_k \circ f_{k-1}\cdots\circ f_1(\boldsymbol{z}_n')$.
- We use *Adagrad* to train our model (i.e. maximize the ELBO (1)).

Experimental Evaluation

- We consider the daily closing prices of 25 NYSE equity indices, January 2008 to January 2011.
- Initially, we train on the first 100 data points, $x_{1:100}$. We perform one-step-ahead prediction and evaluate our model on the test-data log-likelihood pertaining to x_{100}.
- Then, we add x_{100} to the training set, and rerun training/evaluation. We repeat, one step at a time.
- The employed inference DNs comprised 2 layers of 100 hidden units each. The dimensionality of the latent variables $\boldsymbol{z}_n$ was set to $D = 50$. The used normalizing flows comprised $K = 5$ transforms.

Table 1: Average predictive log-likelihood of the evaluated methods (the higher the better).

Equity Index	GARCH	GJR	GPMCH	ReLaVaCH
A	-1.328	-1.298	-1.280	**-1.264**
AA	-1.215	-1.223	-1.213	**-1.201**
AAPL	-1.222	-1.211	-1.211	**-1.198**
ABC	-1.352	-1.340	-1.322	**-1.311**
ABT	**-1.283**	**-1.283**	**-1.283**	**-1.283**
ACE	-1.070	-1.074	-1.067	**-1.060**
ADBE	-1.352	-1.393	-1.293	**-1.282**
ADI	-1.357	-1.334	-1.331	**-1.317**
ADM	-1.210	-1.210	-1.210	**-1.206**
ADP	-1.235	-1.219	-1.215	**-1.198**
ADSK	-1.028	-1.042	**-1.020**	-1.022
AEE	-1.283	-1.269	-1.159	**-1.140**
AEP	-1.138	-1.131	-1.130	**-1.121**
AES	-1.215	-1.215	-1.199	**-1.182**
AET	-1.268	-1.260	-1.243	**-1.228**
AFL	-1.044	-1.046	-1.109	**-1.024**
AGN	-1.257	-1.253	-1.256	**-1.249**
AIG	-1.142	-1.173	-1.055	**-1.005**
AIV	-1.021	-1.032	-1.003	**-1.002**
AIZ	-1.304	-1.336	-1.264	**-1.227**
AKAM	-1.343	-1.329	-1.342	**-1.302**
AKS	-1.211	-1.240	-1.182	**-1.158**
ALL	-1.250	-1.183	**-1.182**	-1.186
ALTR	-1.070	-1.067	-1.056	**-1.044**
AMAT	-1.223	-1.218	-1.235	**-1.211**