IMPERIAL

A new implementation of Network GARCH Model

Peiyi Zhou

September 8, 2025

Table of Contents

Table of Contents

Introduction

Definition (Volatilty)

Latent variable that measures the degree of variation of prices disregarding directions, which captures the overall trend and pattern of the financial return series and forecasting.

Definition (Volatility clustering)

The tendency of high- and low-volatility periods to occur in clusters for financial returns.

Related time series models are classified as volatility models, include:

- ARCH (Engle, 1982)-GARCH (Bollerslev, 1986)
- Multivariate GARCH (BEKK-GARCH (Engle and Kroner, 1995), DCC-GARCH (Engle, 2002), etc.)
- Network GARCH

Limitations of current volatility models

Limitations of multivariate GARCH models:

- No consideration of network interactions between assets which are ‘similar’ or ‘correlated’.
- Complexity and too many parameters.

$$X_{i,t} = Z_{i,t}\sigma_{i,t}, \quad \sigma_{it}^2 = \omega_0 + \alpha_0 X_{i,t-1}^2 + \lambda_0 \sum_{j \neq i} \frac{a_{ij}}{d_i} X_{j,t-1}^2 + \beta_0 \sigma_{i,t-1}^2 \quad (1)$$

Contemporary network GARCH model (later denoted as Zhou’s model, which is derived in 2020 (Zhou et al., 2020)) is lack of:

- Neighbouring volatility persistence.
- Covariance updates.
- Higher stage neighbouring effect.

Key notations (Nason et al., 2023)

Definition (r-stage neighbour)

We say node j is an r-stage neighbour of node i if in a network $\mathcal{G} = (\mathcal{K}, \mathcal{E})$, the shortest path connecting node i and node j is of length r. Followingly, we write $N_r(i)$ as the set of nodes being as the r-stage neighbours of node i, and $|N_r(i)|$ represents the total number of r-stage neighbours for node i.

Definition (r-stage adjacency matrix)

We define a matrix $\mathbf{S}_r$ that

$$(\mathbf{S}_r)_{ij} = \begin{cases} 1 & \text{if node } j \text{ is the } r\text{-stage neighbour of node } i \\ 0 & \text{o.w.} \end{cases} \tag{2}$$

and such matrix $\mathbf{S}_r$ is known as the r-stage adjacency matrix.

Definition (Connection weight matrix)

For a network $\mathcal{G} = (\mathcal{K}, \mathcal{E})$ with total d nodes and undirected edges, we define the connection weight matrix $\mathbf{W} \in \mathbb{R}^{d \times d}$ that

$$w_{ij} = \begin{cases} 1/|N_r(i)| & \text{if node } j \text{ is the } r\text{-stage neighbour of node } i \\ 0 & \text{if node } j \text{ is not a neighbour of } i \text{ at any stage} \end{cases} \quad (3)$$

Unlike most weight matrices, the connection weight matrices $\mathbf{W}$ are usually asymmetric, even if our network is assumed to be undirected, unweighted and no self-loops.

Table of Contents

Methods: Generalised Network GARCH (GNGARCH)

Our GNGARCH$(p, q, [s_1, \cdots, s_q], [r_1, \cdots, r_p])$ model assimilates the idea of current network GARCH model and the Generalised Network Autoregressive (GNAR) model to overcome the listed limitations.

The model is defined by three equations, included here and the next two slides:

$$\mathbf{X}_t = \Sigma_t^{1/2}\mathbf{Z}_t \iff \mathbf{X}_t \mid \mathcal{F}_{t-1} \sim D(0, \Sigma_t) \tag{4}$$

- $\mathbf{X}_t = (X_{1,t}, \cdots, X_{d,t})^T$, on a network $\mathcal{G} = (\mathcal{K}, \mathcal{E})$ with $\mathcal{K} = \{1, \cdots, d\}$ as the set of total d nodes/vertices and $\mathcal{E}$ as the set of edges.

  $X_{i,t}$ denotes the time series variable observed at node i and time t, assumed zero conditional mean at all time. We usually treat $X_{i,t}$ as the return of asset i at time t once each node represents a financial asset, either in the form of simple return or log-return.
- $\mathbf{Z}_t = (Z_{1,t}, \cdots, Z_{d,t})^T$ is a vectorised strict white noise (SWN) variable with zero mean $\mathbb{E}[\mathbf{Z}_t] = 0$ and unit variance $\mathrm{var}(\mathbf{Z}_t) = \mathbf{I}_d$, where every vectorised variable in the SWN process $\{\mathbf{Z}_t\}$ is i.i.d. across both time t and assets i.
- Filtration: $\mathcal{F}_{i,t-1} = \sigma(X_{i,t-1}, \ldots, X_{i,0})$, and $\mathcal{F}_{t-1} = \bigcup_{i=1}^{d} \mathcal{F}_{i,t-1}$.
- Σ_t is the overall conditional covariance matrix with $(\Sigma_t)_{ii} = \sigma_{i,t}^2$ and $(\Sigma_t)_{ij} = \sigma_{ij,t}$ (these will be discussed in the next two slides).

Conditional variance $\sigma_{i,t}^2$ of $X_{i,t}$ given all past information $\mathcal{F}_{i,t-1} = \sigma(X_{i,t-1}, \ldots, X_{i,0})$.

$$\sigma_{i,t}^2 = \underbrace{\alpha_0 + \sum_{k=1}^{q} \alpha_k X_{i,t-k}^2 + \sum_{\ell=1}^{p} \gamma_\ell \sigma_{i,t-\ell}^2}_{\text{own GARCH}(p,\,q)\text{ component}}$$
$$+ \left[\underbrace{\sum_{k=1}^{q} \sum_{r=1}^{s_k} \beta_{kr} \sum_{j \in N_r(i)} w_{ij} X_{j,t-k}^2}_{\text{neighbouring clustering effect}} + \underbrace{\sum_{\ell=1}^{p} \sum_{r'=1}^{r_\ell} \delta_{\ell r'} \sum_{j \in N_{r'}(i)} w_{ij} \sigma_{j,t-\ell}^2}_{\text{neighbouring persistence effect}} \right] \tag{5}$$

Conditional covariance $\sigma_{ij,t}$ between $X_{i,t}$ and $X_{j,t}$ conditional on $\mathcal{F}_{i,t-1} \cup \mathcal{F}_{j,t-1}$.

$$
\begin{aligned}
\sigma_{ij,t} = &\underbrace{\alpha_0 + \sum_{k=1}^{q} \alpha_k X_{i,t-k} X_{j,t-k} + \sum_{\ell=1}^{p} \gamma_\ell \sigma_{ij,t-\ell}}_{\text{GARCH}(p,\,q)\text{ related, node }(i,\,j)\text{ covariance}} \\
&+ \frac{1}{2} \sum_{k=1}^{q} \sum_{r=1}^{s_k} \beta_{kr} \left[\underbrace{\sum_{\substack{u \in N_r(i) \\ u \neq j}} w_{iu} X_{u,t-k} X_{j,t-k}}_{\substack{\text{node } i\text{'s neighbouring effect} \\ \text{on clustering w.r.t node } j}} + \underbrace{\sum_{\substack{v \in N_r(j) \\ v \neq i}} w_{jv} X_{i,t-k} X_{v,t-k}}_{\substack{\text{node } j\text{'s neighbouring effect} \\ \text{on clustering w.r.t node } i}} \right] \\
&+ \frac{1}{2} \sum_{\ell=1}^{p} \sum_{r'=1}^{r_\ell} \delta_{\ell r'} \left[\underbrace{\sum_{\substack{u \in N_{r'}(i) \\ u \neq j}} w_{iu} \sigma_{uj,t-\ell}}_{\substack{\text{node } i\text{'s neighbouring effect} \\ \text{on persistence w.r.t node } j}} + \underbrace{\sum_{\substack{v \in N_{r'}(i) \\ v \neq i}} w_{jv} \sigma_{vi,t-\ell}}_{\substack{\text{node } j\text{'s neighbouring effect} \\ \text{on persistence w.r.t node } i}} \right]
\end{aligned}
\tag{6}
$$

Remarks of GNGARCH

1. Similar to the univariate GARCH model, in order to ensure the positivity of conditional variance in (5), we need to put constraints on these global parameters that:

$$\alpha_0 > 0, \alpha_k \geq 0, \gamma_\ell \geq 0, \beta_{kr} \geq 0, \delta_{\ell r'} \geq 0 \tag{7}$$

   for $k = 1, \cdots, q$, $\ell = 1, \cdots, p$, $r = s_1, \cdots, s_q$ and $r' = r_1, \cdots, r_p$.
2. We include coefficient $1/2$ in (6) for the neighbouring clustering and persistence terms based on the assumption that we have equal neighbouring clustering and persistence effects on volatility from node i's r-stage neighbours interacting on node j, and node j's r-stage neighbours interacting on node i respectively. The sum of each node's neighbours would result the overall normalised effects and avoid double-counting.
3. To make the model meaningful, we also employ suitable numerical approximations to ensure the positive-definiteness of Σ_t in every t. Practically it can be performed by:
   - Adding a small 'jitter' on the diagonal to force all eigenvalues to be positive.
   - (Or) projecting onto the nearest positive-definite matrix using, for example, Python's `cov_nearest` function.
4. Conversion: GNGARCH $\leftrightarrow$ VARMA.

Table of Contents

GNGARCH specification

Consider GNGARCH$(1,1,[1],[1])$, in which each node's conditional variance depends on its own one-period lag and the one-period lag of its first-stage neighbours.

$$\mathbf{X}_t = \Sigma_t^{1/2}\mathbf{Z}_t \iff \mathbf{X}_t \mid \mathcal{F}_{t-1} \sim D(0, \Sigma_t) \tag{8a}$$

$$\sigma_{i,t}^2 = \alpha_0 + \alpha_1 X_{i,t-1}^2 + \gamma_1 \sigma_{i,t-1}^2 + \left[\beta_{11} \sum_{j \in N_1(i)} w_{ij} X_{j,t-1}^2 + \delta_{11} \sum_{j \in N_1(i)} w_{ij} \sigma_{j,t-1}^2\right] \tag{8b}$$

$$\begin{aligned}\sigma_{ij,t} &= \alpha_0 + \alpha_1 X_{i,t-1} X_{j,t-1} + \gamma_1 \sigma_{ij,t-1} \\ &\quad + \frac{1}{2}\beta_{11}\left[\sum_{u \in N_1(i), u \neq j} w_{iu} X_{u,t-1} X_{j,t-1} + \sum_{v \in N_1(j), v \neq i} w_{jv} X_{i,t-1} X_{v,t-1}\right] \\ &\quad + \frac{1}{2}\delta_{11}\left[\sum_{u \in N_r(i), u \neq j} w_{iu}\sigma_{uj,t-1} + \sum_{v \in N_r(j), v \neq i} w_{jv}\sigma_{vi,t-1}\right]\end{aligned} \tag{8c}$$

Our parameters for this model are $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \gamma_1, \beta_{11}, \delta_{11})$.

Simulation network and scheme

Under the assumption that the SWN process $\{\mathbf{Z}_t\}$ is normally distributed, with a preset initial value of $\mathbf{X}_0$ and Σ_0, the simulation process goes by sequentially updates the conditional covariance Σ_t via (8b) and (8c), and then simulate $\mathbf{X}_t$ from Gaussian distribution $\mathcal{N}(0, \Sigma_t)$.

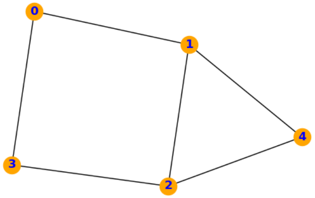


Figure: The simple network used for our simulation task and model evaluation. Here we have $\mathbf{X}_t = (X_{0,t}, X_{1,t}, X_{2,t}, X_{3,t}, X_{4,t})^T$.

We generate a total of 2000 samples of $\mathbf{X}_t$ through (8a), and discard the first 20% samples (so 400 burnin samples). The 'true' parameters we fit for our GNGARCH$(1, 1, [1], [1])$ model for simulation of $\{\mathbf{X}_t\}$ are

$$\alpha_0 = 0.05, \alpha_1 = 0.20, \gamma_1 = 0.60, \beta_{11} = 0.05, \delta_{11} = 0.05 \tag{9}$$

Simulation results: Stylised Facts (Cont, 2001)

We check the model validity via stylised facts (SF) of asset return series.

SF1: The unconditional distribution of returns has a heavy tail and mild asymmetry.

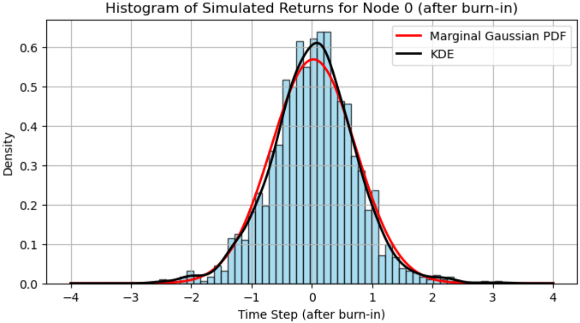


Figure: Histogram of node 0 returns with KDE and normal distribution curve.

Pearson's kurtosis and simulated series skewness are computed that

$$\kappa = 3.903 > 3, \quad \beta = 0.008 \neq 0 \tag{10}$$

A Pearson kurtosis greater than 3 and a non-zero skewness together indicate the heavy-tail and asymmetric nature of simulated return's distribution, aligning with SF1.

SF2: Autocorrelations of asset returns are often negligible that decay to zero rapidly.

We compute the sample/empirical autocorrelation function (ACF) of the raw return series

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{T}(X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=1}^{T}(X_t - \bar{X})^2} \tag{11}$$

for lags $k = 1, \cdots, 20$, with the 95% confidence bound of $\pm 1.96/\sqrt{T}$ (and $T = 1600$).

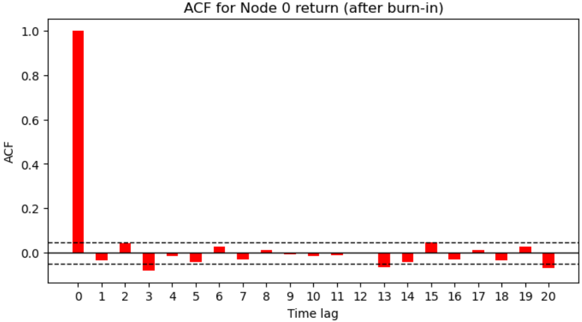


Figure: Empirical autocorrelation of the first 20 lags of node 0 return.

SF3: Volatility clustering and persistence effect: different measures of volatility shows a positive autocorrelation and slow decay over time lags.

We can assess this either by computing the sample autocorrelation of the absolute return series $|X_{i,t}|$ or by using the sequence of model-implied conditional standard deviations $\{\sigma_{i,t}\}$ directly: both measures can be seen as our volatility proxies.

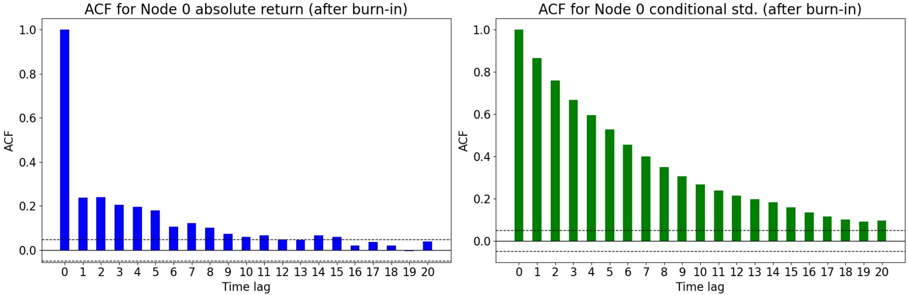


Figure: Empirical autocorrelation of the first 20 lags of node 0 absolute return (left) and the first 20 lags of simulated conditional standard deviation (right).

SF4: Aggregational Gaussianity: the distribution of aggregated returns converges toward a normal distribution as the aggregation time scale increases.

We evaluate by considering 1-day (daily), 7-day (weekly), and 30-day (monthly) aggregation windows, and for each we standardise aggregated returns and plot QQ-plots against the standard normal distribution $\mathcal{N}(0,1)$.

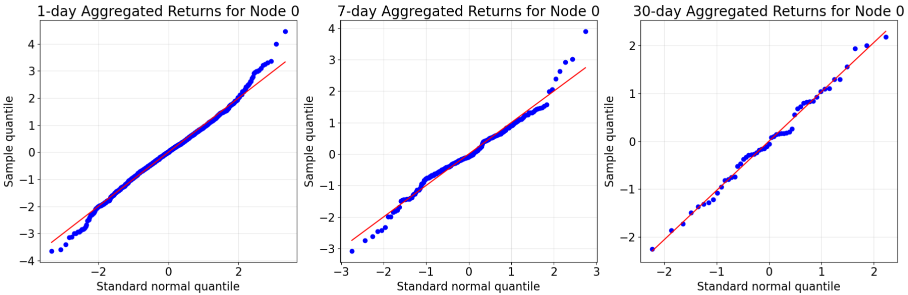


Figure: QQ plot for node 0 aggregated returns over days, weeks and months.

Parameter estimation

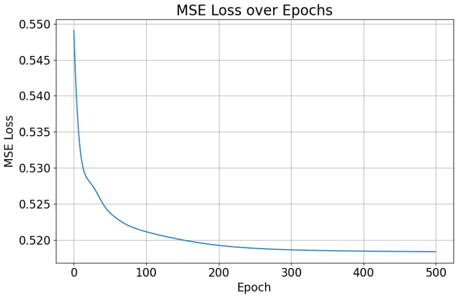
Our parameter estimation scheme is based on a numerical optimisation process with **a proxy of latent variable volatility** and **a certain loss function**, then employ a gradient-based optimiser to minimise it.

- Variance proxy: **squared return** ($\mathbf{X}_t\mathbf{X}_t^T$) as a proxy for the true, unobservable conditional variance, which is conditionally unbiased with respect to the model conditional variance $\widehat{\Sigma}_t$.
- Loss function: with simulated $\{\mathbf{X}_t\}_{t=0}^{T-1}$ (with $\mathbf{X}_0$ as the initial return), and $\{\widehat{\Sigma}_t\}_{t=1}^{T-1}$ as the corresponding conditional variance sequence.
  1. Mean squared error (**MSE**):
  $$L_{\mathrm{MSE}}(\mathbf{X}_t\mathbf{X}_t^T, \widehat{\Sigma}_t) = \frac{1}{T-1}\sum_{t=1}^{T-1}\left[d^{-2}\sum_{i=1}^{d}\sum_{j=1}^{d}(\mathbf{X}_t\mathbf{X}_t^T - \widehat{\Sigma}_t)_{ij}^2\right] \tag{12}$$
  2. Quasi-likelihood (**QLIKE**): invariant up to additive constants and an overall scale factor to negative log-likelihood (**NLL**) function.
  $$L_{\mathrm{QLIKE}}(\mathbf{X}_t\mathbf{X}_t^T, \widehat{\Sigma}_t) = \frac{1}{T-1}\sum_{t=1}^{T-1}\left(\log|\widehat{\Sigma}_t| + \mathbf{X}_t^T\widehat{\Sigma}_t^{-1}\mathbf{X}_t\right) \tag{13}$$

Parameter fitting results

Parameter	True	Estimate (MSE)	Estimate (NLL)
α_0	0.05	0.06	0.05
α_1	0.20	0.16	0.16
γ_1	0.60	0.59	0.62
β_{11}	0.05	0.03	0.05
δ_{11}	0.05	0.09	0.04

Table: Parameter estimates (2 decimal places) for simulated returns with seed 0 and 500 epochs, with the adaptive optimiser `Adam` with learning rate 0.01.



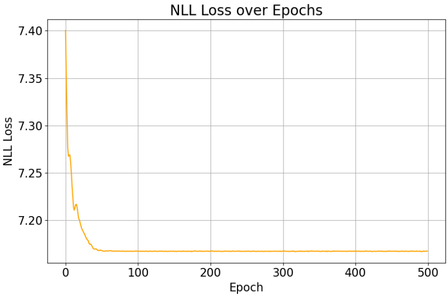


Table of Contents

Data used

The return series dataset used for training the model in this section is taken as log-returns derived from the most active 75 US stocks' daily closing price in the market (updated on 19th June, 2025), including well-known stocks like Nvidia (NVDA), Tesla (TSLA) and Intel (INTC), etc., in a duration from 29th April, 2022 to 31st December, 2024.

Virtual Network Construction

We apply correlation-based network construction, and the result graph adjacency matrix is named the correlation-of-correlation (CoC) adjacency matrix (Tapia Costa et al., 2025).

Formally, if we denote the empirical sample correlation matrix as $\mathbf{R}$ and the CoC adjacency matrix as $\mathbf{A}$, with threshold λ, then

$$\mathbf{A}_{ij} = \mathbf{1}(\mathbf{R}_{ij} > \lambda) \tag{14}$$

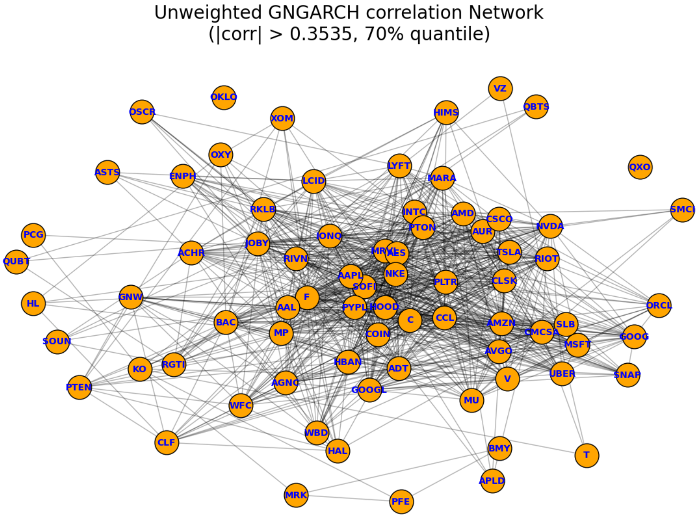


Figure: Constructed virtual stock network based on correlation method, with the linking threshold setting as the 70% quantile of the absolute value of empirical correlation. Node names represent the stock name abbreviations.

Parameter fitting results

Parameter	Estimate (MSE)	Estimate (NLL)
α_0	0.0232	0.0005
α_1	0.0702	0.1648
γ_1	0.1693	0.7072
β_{11}	0.0236	0.0008
δ_{11}	0.0432	0.0039

Table: Parameter estimates on the dataset, results up to 4 decimal places.

The estimates obtained here differ markedly, and we will prefer using the QLIKE/NLL-based estimates as when the variance proxy is set as the squared return, the MSE-based estimation is heavily affected from the noise effect accumulated in the complex network with many nodes, causing the parameter estimates to be dominated by noise.

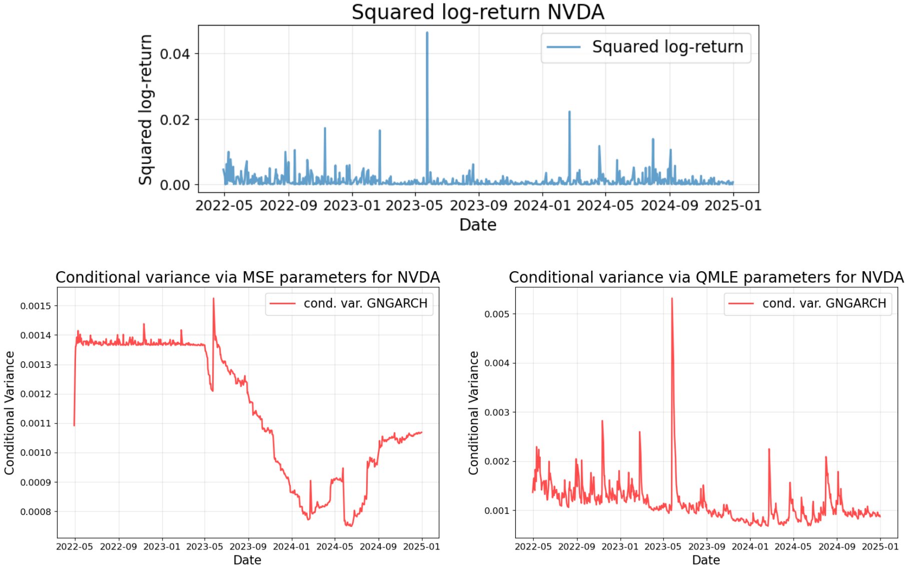


Figure: The fitted GNGARCH$(1, 1, [1], [1])$ model with rescaled conditional variance by using MSE-based parameters (left) and QLIKE/NLL-based QMLE parameters (right).

Model comparison - conditional variance for a single stock

We evaluate the conditional variance for the example stock, NVDA, for the volatility study, against four benchmarks:

1. The best-fitting univariate GARCH model.
2. The RiskMetrics procedure defined as $\widehat{\sigma}^2_{i,t} = \lambda \widehat{\sigma}_{i,t-1} + (1-\lambda) r^2_{i,t-1}$, with $\lambda = 0.94$.
3. Our GNGARCH$(1, 1, [1], [1])$ estimated by QMLE.
4. Zhou's network GARCH model, also estimated by QMLE.

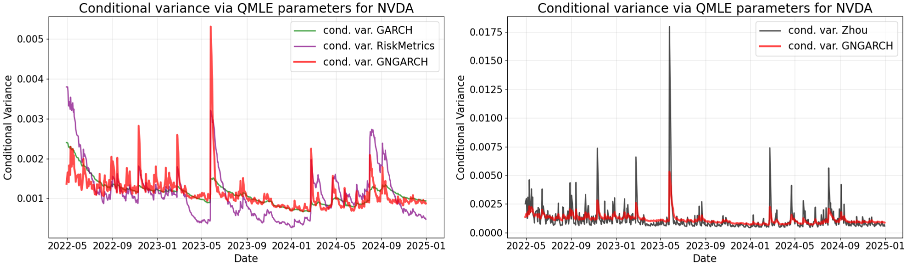
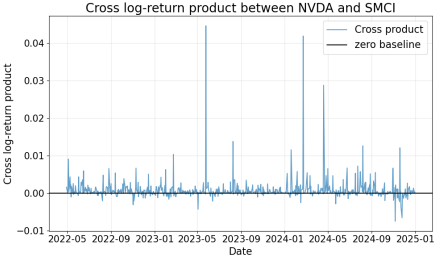


Figure: Conditional variance comparison for NVDA across models. Left: comparison between the best-fit univariate GARCH (green), RiskMetrics ($\lambda = 0.94$, purple) and the fitted GNGARCH$(1, 1, [1], [1])$ (thick red). Right: comparison between Zhou's network GARCH with QMLE-fitted parameters (black) and the fitted GNGARCH$(1, 1, [1], [1])$ (thick red).

Model conditional covariances between two stocks

On the other hand, the model conditional covariance is often used to estimate the latent (squared) co-volatility, and the cross-product can be analogously seen as a noisy proxy.



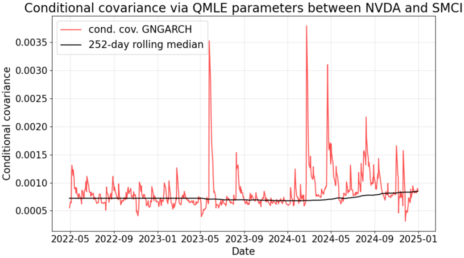


Figure: Left: Cross-product of log-returns over time between stock NVDA and SMCI. Right: conditional covariance between NVDA and SMCI, plotted against its (252-day) rolling median. Conditional covariances above the base rolling median indicate the same direction co-movement, vice versa.

Our fitted GNGARCH$(1, 1, [1], [1])$ is again able to track the spikes, but also the some troughs that are below the baseline, showing a general consistent trend along with the cross-product of log-returns.

Table of Contents

Discussion

We develop a novel framework of network GARCH model, the Generalised Network GARCH (GNGARCH). Analytical work in this presentation includes:

- Model derivation and remarks.
- Simulation experiments on a specified GNGARCH$(1, 1, [1], [1])$ model.
- Apply our GNGARCH$(1, 1, [1], [1])$ on a real-world example on 75 active US stocks.

Other things I have done (in thesis but not in the presentation):

- A brief sketch of stationarity of GNGARCH.
- Complete analysis of model vectorisation and conversion.
- Spatial illustration of the model conditional variances via network volatility autocorrelation function (NVACF).
- Discussion and evaluation on noise accumulation effect in MSE rather than QLIKE/NLL.
- Thorough explanation of CoC network construction.
- Fitted model performance on a validation dataset.

Future perspectives

Limitations:

- We do not provide a general proof of stationarity for GNGARCH.
- We only consider the GNGARCH$(1, 1, [1], [1])$ specification and do not explore extensions that incorporate higher stage neighbouring nodes.
- Our model comparisons are qualitative by just plotting conditional variances against squared (log-)returns and covariances against cross-product of corresponding (log-)returns, a quantitative evaluation of predictive accuracy is therefore necessary.

Future work will focus on these gaps, along with using the Diebold-Mariano-West tests to quantify predictive accuracy among different volatility models, and apply these network GARCH models to asset's value-at-risk (VaR) forecasting.

References

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *J. Econometrics.*, 31(3):307–327.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quant. Financ.*, 1(2):223–236.

Engle, R. (2002). Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *J. Bus. Econ. Stat.*, 20(3):339–350.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4):987–1007.

Engle, R. F. and Kroner, K. F. (1995). Multivariate Simultaneous Generalized ARCH. *Economet. Theor.*, 11(1):122–150.

Nason, G. P., Salnikov, D., and Cortina Borja, M. (2023). New tools for network time series with an application to COVID-19 hospitalisations. *J. R. Stat. Soc. A (to appear)*.

Tapia Costa, M., Cucuringu, M., and Nason, G. P. (2025). Higher Order Dynamic Network Linear Models for Covariance Forecasting. *Available at SSRN 5113698*.

Zhou, J., Li, D., Pan, R., and Wang, H. (2020). Network GARCH Model. *Stat. Sinica.*, 30(4):1723–1740.

Thank you for your listening!

Questions?