

IMPERIAL

Imperial College London
Department of Mathematics

GNAR-HARX Models for Realised Volatility: Incorporating Exogenous Predictors and Network Effects

Tom Ó Nualláin

CID: 06021498

Supervised by Professor Guy Nason

October 29, 2025

Submitted in partial fulfilment of the requirements for the MSc in
Statistics at Imperial College London

The work contained in this thesis is my own work unless otherwise stated.

Signed: Tom Ó Nualláin

Date: October 29, 2025

Acknowledgements

I would first like to express my sincere gratitude to my supervisor, Professor Guy Nason, for his generous guidance, encouragement, and advice throughout the summer. His support has been invaluable in shaping this project and in developing my understanding of the topic.

I am also grateful to Dr. Mikko Pakkanen for providing access to the main dataset used in this study. I would like to thank Peiyi Zhou for the shared discussions during our joint supervision meetings, which were both motivating and helpful.

Special thanks go to Marcos Tapia Costa, who generously shared code that informed parts of my implementation, and to Henry Palasciano, whose prior code served as a useful reference in developing my own.

Finally, I wish to thank my family and friends for their constant support and encouragement throughout the summer.

Abstract

This project introduces the GNAR-HARX model, which combines Generalised Network Autoregressive (GNAR) structure with Heterogeneous Autoregressive (HAR) dynamics and exogenous predictors such as implied volatility. The model is designed for forecasting realised volatility by capturing both temporal persistence and cross-sectional spillovers in financial markets. We apply it to daily realised variance data for ten international stock indices, generating one-step-ahead forecasts in a rolling window over an out-of-sample period of approximately 16 years (2005–2020).

Forecast accuracy is evaluated using the Quasi-Likelihood (QLIKE) loss and mean squared error (MSE), and we compare global, standard, and local variants across different network structures and exogenous specifications. The best model found by QLIKE is a local GNAR-HAR without exogenous variables, while the lowest MSE is achieved by a standard GNAR-HARX with implied volatility. Fully connected networks consistently outperform dynamically estimated graphical lasso networks.

Overall, local and standard GNAR-HAR(X) models deliver the strongest forecasts, though at the cost of more parameters than the parsimonious global variant, which nevertheless remains competitive. Across all cases, GNAR-HAR(X) models outperform univariate HAR(X) benchmarks, which often require more parameters than the GNAR-based specifications. While the top model found by QLIKE does not use exogenous variables, implied volatility and overnight returns emerge as the most useful predictors when included.

1. Introduction

Given its central role in portfolio allocation, risk management, derivatives pricing, and algorithmic trading, the accurate modelling of realised volatility has long been an active area of research. Accurate forecasts can enable market participants to hedge risk, manage uncertainty, and optimise capital allocation. This has led to an extensive literature on

statistical methods for volatility modelling, increasingly informed by the availability of high-frequency financial data.

An important empirical feature of realised volatility is its long memory, evidenced by a slowly decaying autocorrelation function. Traditional ARMA models struggle to capture this persistence unless made high-order, leading to increased estimation complexity. While ARFIMA models ([Andersen et al., 2003](#)) address this using fractional differencing, they are challenging to estimate in practice ([Baillie, 1996](#)) and may provide less direct financial interpretation compared to other models.

One widely studied approach is the Heterogeneous Autoregressive (HAR) model introduced by [Corsi \(2009\)](#), which approximates long memory using a small number of lags at daily, weekly, and monthly horizons. Motivated by the idea that market participants operate at different time scales, HAR models capture heterogeneity in volatility dynamics while remaining straightforward to estimate. As summarised by [Clements & Preve \(2021\)](#), the HAR model and its variants consistently deliver competitive, and often superior, forecast performance across a wide range of assets and market conditions.

While HAR models effectively capture persistence in univariate volatility, they are confined to individual assets or markets, and therefore cannot address cross-market spillovers. Recognising this limitation has motivated the development of multivariate approaches that explicitly model interconnected markets.

Recent developments in financial econometrics increasingly emphasise the interconnected nature of global markets. Volatility shocks often propagate across borders, particularly during turbulent periods, a phenomenon widely known as volatility spillover ([Rigobon & Sack, 2003](#)). For example, [Buncic & Gisler \(2016\)](#) and [Wang et al. \(2018\)](#) show that U.S. market volatility influences foreign asset markets, with the VIX frequently acting as a global risk barometer.

Traditional multivariate approaches such as Vector Autoregressive (VAR) models, BEKK-GARCH ([Engle & Kroner, 1995](#)), and Wishart autoregressions ([Gourieroux et al., 2009](#)) can capture cross-market dynamics but scale poorly: parameter counts grow rapidly with

the number of assets, leading to increased estimation variance and reduced predictive power (Callot et al., 2017). This curse of dimensionality has motivated the search for more parsimonious yet flexible approaches to capturing volatility spillovers, an objective for which network-based time series models are particularly well suited.

In this context, network time series models provide a promising framework by capturing both temporal and cross-sectional dependencies in multivariate time series through an underlying network representation.

An early contribution in this area is Knight et al. (2016), who introduced Network Autoregressive Moving Average (NARIMA) processes. These models extend the VARMA framework to network settings and are particularly suited to cases where the network structure itself may evolve over time.

Zhu et al. (2017) then proposed the Network Autoregressive (NAR) model, where each node evolves as a function of its own lags and those of its immediate (one-hop) neighbours. The NAR model reduces complexity relative to VAR by restricting dependence to local interactions rather than requiring interactions across all pairs of nodes. However, the NAR model assumes a binary, static network and homogeneous dynamics across nodes.

Building on Knight et al. (2016), the Generalised Network Autoregressive (GNAR) model of Knight et al. (2019, 2020) incorporates weighted edges and k -hop neighbourhoods. In GNAR models, the value at each node depends not only on its own history but also on a weighted sum of past values from neighbouring nodes up to a fixed network hop, allowing for more flexible spillover effects across nodes without substantially increasing model complexity.

Recent extensions have adopted modelling approaches similar to the GNAR framework for volatility forecasting. Zhang et al. (2025b) introduced the Graph-HAR (GHAR) model, which augments HAR-style lag structures with neighbourhood-aggregated volatility and correlation measures. In their HAR-DRD framework, the realised covariance

matrix is decomposed as $\Sigma_t = D_t R_t D_t$, where D_t is a diagonal matrix of realised standard deviations and R_t is the realised correlation matrix. Each component is then modelled separately using graph-based predictors. [Tapia Costa et al. \(2025\)](#) build on this by proposing a GNAR-HAR model for forecasting realised covariance matrices, capturing higher-order neighbourhood interactions in both volatilities and correlations. In contrast, our focus is solely on forecasting realised variance, without modelling the full covariance structure or using a DRD decomposition.

Another line of work has focused on incorporating exogenous information into network autoregressive models. While [Zhu et al. \(2017\)](#) already noted the possibility of including covariates in NAR-type specifications, [Nason & Wei \(2022\)](#) formalised this within the GNARX model. GNARX extends GNAR by including node-specific exogenous variables, such as COVID-19 stringency indices and death rates, as additional regressors. Applied to forecasting Purchasing Managers' Indices across countries using trade-determined networks, GNARX was shown to outperform both standard GNAR and traditional VAR models in terms of predictive accuracy (mean squared error).

Building on these strands of research, we introduce the GNAR-HARX model, a new framework that combines: the network-based cross-sectional modelling of GNAR, the long memory structure of HAR and the flexibility to include node-specific exogenous covariates from GNARX.

The GNAR-HARX model in this case is used to forecast daily realised variance across ten international stock indices, capturing long memory effects, network spillovers, and exogenous influences, often achieving a more parsimonious structure than similar univariate models.

Volatility spillovers are modelled via an underlying network structure, either specified a priori or estimated from historical data. The HAR lags accommodate persistence, and node-level exogenous regressors (e.g., implied volatility indices or overnight returns) allow the model to include more up to date or forward-looking predictive information.

The primary contribution of this thesis is the formal definition and empirical investigation of the GNAR-HARX model, which integrates HAR dynamics and exogenous predictors into the GNAR framework. While related components have been studied individually in previous work, this project brings them together in the context of forecasting realised variance. Building on existing literature, we also propose a set of economically motivated exogenous variables and assess their contribution to predictive performance. We assess the model’s performance using daily realised variance data for ten major international equity indices over the period from 2001 through 2020, benchmarking against established alternatives including HAR, HARX, and GNAR-HAR. To support transparency and future work, we also provide a Python implementation for estimation, forecasting, and model comparison.

The remainder of this thesis is structured as follows:

Section 2 introduces the GNAR-HARX model in detail, including its mathematical formulation, variants, and estimation procedure. Section 3 presents and summarises the data used. Section 4 outlines the model comparison framework, rolling window design, and evaluation metrics. Section 5 reports empirical findings, evaluating the forecasting performance of GNAR-HARX and the effect of network structure. Section 6 concludes with a discussion of implications, limitations, and directions for future work.

2. Model

2.1. Model Specification

2.1.1. Notation and Definitions

Let $\mathbf{Y}_t = (Y_{1,t}, \dots, Y_{N,t})^\top$ denote an N -variate time series observed over a (possibly time-varying) network $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t)$. The node set is $\mathcal{V} = \{1, \dots, N\}$, and the edge set $\mathcal{E}_t \subseteq \mathcal{V} \times \mathcal{V}$ encodes the connections between nodes at time t . We use the terms *node*

and *vertex* interchangeably. As a motivating example, $Y_{i,t}$ may represent the log realised variance of index i on day t .

Throughout this study, we consider only undirected graphs, so that $(i, j) \in \mathcal{E}_t \iff (j, i) \in \mathcal{E}_t$. The corresponding adjacency matrix A_t has entries $(A_t)_{ij} = 1$ if $(i, j) \in \mathcal{E}_t$ and 0 otherwise.

For a subset $A \subseteq \mathcal{V}$, the (stage-1) neighbour set of A is defined as

$$\mathcal{N}_t(A) = \{j \in \mathcal{V} \setminus A : \exists i \in A \text{ with } (i, j) \in \mathcal{E}_t\}.$$

For a single node i , write $\mathcal{N}_t^{(1)}(i) := \mathcal{N}_t(\{i\})$. Higher-stage neighbours are defined recursively as

$$\mathcal{N}_t^{(r)}(i) = \mathcal{N}_t(\mathcal{N}_t^{(r-1)}(i)) \setminus \left(\{i\} \cup \bigcup_{q=1}^{r-1} \mathcal{N}_t^{(q)}(i) \right), \quad r \geq 2.$$

Alternatively, nodes i and j are r -stage neighbours if the shortest path between them in $\mathcal{E}_t$ has distance r , i.e., $d_t(i, j) = r$ (Nason et al., 2024).

For $r \geq 1$ and $j \in \mathcal{N}_t^{(r)}(i)$, let $w_{i,j}^{(r)}(t) \in [0, 1]$ denote the connection weight from node i to node j at stage r . In this study we adopt uniform (equally weighted) neighbourhood averages:

$$w_{i,j}^{(r)}(t) = \begin{cases} |\mathcal{N}_t^{(r)}(i)|^{-1}, & \text{if } j \in \mathcal{N}_t^{(r)}(i) \text{ and } |\mathcal{N}_t^{(r)}(i)| > 0, \\ 0, & \text{otherwise.} \end{cases}$$

where $|\cdot|$ denotes set cardinality. Thus $\sum_{j \in \mathcal{N}_t^{(r)}(i)} w_{i,j}^{(r)}(t) = 1$ when $\mathcal{N}_t^{(r)}(i) \neq \emptyset$. The adjacency $\mathcal{E}_t$ is specified externally (see Section 4.2 for details on network construction) so that given $\mathcal{E}_t$, the weights are deterministic and not estimated parameters.

For integers $a \geq b \geq 1$, define the non-overlapping lag average

$$Y_{i,t-a:t-b} := \frac{1}{a-b+1} \sum_{k=b}^a Y_{i,t-k}.$$

which is used to form weekly and monthly HAR components.

Let $\{X_{h,i,t}\}$ be the h -th stationary exogenous regressor for node i at time t , with $h = 1, \dots, H$ and maximum lag $p'_h \in \mathbb{N}_0$.

2.1.2. GNAR-HAR Model

The GNAR-HAR model of [Tapia Costa et al. \(2025\)](#) extends the GNAR framework to incorporate Heterogeneous Autoregressive (HAR) dynamics. The GNAR-HAR model describes the evolution of $Y_{i,t}$ for each node $i \in \{1, \dots, N\}$ at time $t \in \{1, \dots, T\}$ as:

$$\begin{aligned} Y_{i,t} = & \alpha_d Y_{i,t-1} + \alpha_w Y_{i,t-5:t-2} + \alpha_m Y_{i,t-22:t-6} \\ & + \sum_{r=1}^{r_d} \beta_{d,r} \sum_{j \in \mathcal{N}_t^{(r)}(i)} w_{i,j}^{(r)}(t) Y_{j,t-1} \\ & + \sum_{r=1}^{r_w} \beta_{w,r} \sum_{j \in \mathcal{N}_t^{(r)}(i)} w_{i,j}^{(r)}(t) Y_{j,t-5:t-2} \\ & + \sum_{r=1}^{r_m} \beta_{m,r} \sum_{j \in \mathcal{N}_t^{(r)}(i)} w_{i,j}^{(r)}(t) Y_{j,t-22:t-6} + u_{i,t} \end{aligned} \quad (1)$$

where the error terms $\{u_{i,t}\}$ are mean-zero random variables with variance σ_i^2 .

The autoregressive part follows the HAR structure of [Corsi \(2009\)](#), using daily, weekly, and monthly non-overlapping averages to capture distinct time scales. The network terms capture spillovers from neighbours at different graph distances.

2.1.3. GNAR-HARX Model

We propose the GNAR-HARX model, which extends the GNAR-HAR model in Equation 1 by incorporating exogenous regressors. Specifically, the GNAR-HARX model augments Equation 1 with the additional term

$$\sum_{h=1}^H \sum_{j'=0}^{p'_h} \lambda_{h,j'} X_{h,i,t-j'}$$

so that

$$\begin{aligned} Y_{i,t} = & \alpha_d Y_{i,t-1} + \alpha_w Y_{i,t-5:t-2} + \alpha_m Y_{i,t-22:t-6} \\ & + \sum_{r=1}^{r_d} \beta_{d,r} \sum_{j \in \mathcal{N}_t^{(r)}(i)} w_{i,j}^{(r)}(t) Y_{j,t-1} \\ & + \sum_{r=1}^{r_w} \beta_{w,r} \sum_{j \in \mathcal{N}_t^{(r)}(i)} w_{i,j}^{(r)}(t) Y_{j,t-5:t-2} \\ & + \sum_{r=1}^{r_m} \beta_{m,r} \sum_{j \in \mathcal{N}_t^{(r)}(i)} w_{i,j}^{(r)}(t) Y_{j,t-22:t-6} \\ & + \sum_{h=1}^H \sum_{j'=0}^{p'_h} \lambda_{h,j'} X_{h,i,t-j'} + u_{i,t}. \end{aligned} \quad (2)$$

Here, the additional term captures the effects of lagged exogenous covariates $X_{h,i,t}$, with $\lambda_{h,j'}$ denoting the effect of the j' -th lag of covariate h .

Parameters:

- $\alpha_d, \alpha_w, \alpha_m$: autoregressive coefficients at daily, weekly, and monthly horizons.
- $\beta_{d,r}, \beta_{w,r}, \beta_{m,r}$: network autoregressive coefficients for neighbourhood stage r at daily, weekly, and monthly horizons.
- $\lambda_{h,j'}$: coefficients for lag j' of the h -th exogenous variable.

- $\mathbf{r} = [r_d, r_w, r_m]$: maximum neighbourhood stages for each time scale.
- $\mathcal{N}_t^{(r)}(i)$: stage- r neighbours of node i .
- $w_{i,j}^{(r)}(t)$: uniform weights across neighbours at stage r .

2.1.4. Model Variations and Parameter Counts

We consider three GNAR-HARX model variants that differ in their degree of parameter sharing across nodes:

- Global- α (global): All nodes share autoregressive, network and exogenous parameters.
- Local- α (standard): Autoregressive parameters vary by node; network and exogenous parameters are shared.
- Local- $\alpha\beta$ (local): All autoregressive, network and exogenous parameters are node-specific.

A key advantage of the GNAR-HARX framework lies in its parsimony on large networks, particularly in the global- α and local- α variants. Under the global variant, the autoregressive and network components contribute only $3 + r_d + r_w + r_m$ parameters (three HAR lags plus daily/weekly/monthly network stages), regardless of N . Exogenous parameters scale with the number of variables and their lags but, in the global or standard cases, are not multiplied by N .

Table 1 summarises the number of parameters under each specification, assuming three HAR lags; daily/weekly/monthly network stages (r_d, r_w, r_m); and H exogenous variables with lag orders $p'_1, \dots, p'_H$:

Model	Autoregressive Terms	Network Terms	Exogenous Terms
Global- α	3	$r_d + r_w + r_m$	$\sum_{h=1}^H p'_h$
Local- α	$3N$	$r_d + r_w + r_m$	$\sum_{h=1}^H p'_h$
Local- $\alpha\beta$	$3N$	$N(r_d + r_w + r_m)$	$N\left(\sum_{h=1}^H p'_h\right)$

Table 1.: Parameter counts for GNAR-HARX model variants.

By comparison, fitting N independent HARX models requires $N\left(3 + \sum_{h=1}^H p'_h\right)$ parameters, which scales linearly with N . In contrast, the global- α GNAR-HARX keeps the total parameter count independent of N . The local- α specification increases only the autoregressive block to $3N$, while still sharing network and exogenous components across nodes. The more flexible local- $\alpha\beta$ variant scales the autoregressive, network and exogenous parameters with N , leading to higher complexity. Thus, the global and local- α specifications are markedly more parsimonious, which may help to reduce overfitting risk as N grows.

2.1.5. Stationarity Condition:

Assuming all exogenous regressors $X_{h,i,t}$ are stationary, the GNAR-HARX process $\mathbf{Y}_t$ is stationary if:

$$|\alpha_d| + |\alpha_w| + |\alpha_m| + \sum_{r=1}^{r_d} |\beta_{d,r}| + \sum_{r=1}^{r_w} |\beta_{w,r}| + \sum_{r=1}^{r_m} |\beta_{m,r}| < 1$$

A derivation based on the mapping of the GNAR-HARX model to a constrained GNARX process is provided in [Appendix A](#).

3. Data

This study uses a multivariate time series of daily realised variance and exogenous predictors for ten international equity indices, spanning from **2 February 2001** to **30 December 2020**. This sample period is selected because it represents the time frame during which both realised variance and implied volatility data are consistently available for the chosen indices.

3.1. Data Coverage and Indices

Table 2 lists the ten selected equity indices and their corresponding implied volatility indices. The selection covers major developed markets and reflects a broad geographic and economic representation.

Stock Index	Index Abbreviation	Volatility Index	Country/Region
AEX Index	AEX	VAEX	Netherlands
CAC 40	CAC	VCAC	France
DAX	DAX	VDAX-NEW	Germany
Dow Jones Industrial Average	DJI	VXD	United States
EURO STOXX 50	STX	VSTOXX	Eurozone
FTSE 100	FTS	VFTSE	United Kingdom
Nasdaq 100	NDX	VXN	United States
Nikkei 225	NKY	VXJ	Japan
S&P 500 Index	SPX	VIX	United States
Swiss Market Index	SMI	VSMI	Switzerland

Table 2.: Stock indices and their associated implied volatility indices

Having established the dataset and market coverage, we next describe how the realised variance is measured and constructed.

3.2. Realised Variance Measures

The realised variance (RV) data are sourced from the Oxford-Man Institute's (OMI) Realized Library (Heber et al., 2009). Specifically, we used the `rv5_ss` series, a five-minute subsampled realised variance measure. This estimator, introduced by Zhang et al. (2005), balances efficiency and robustness by averaging across multiple offset sampling grids to reduce microstructure noise. For clarity, we first outline the theoretical definition of realised variance and then describe how the `rv5_ss` series is constructed.

The estimates of RV in the OMI Realized Library implement the methodology developed by Andersen & Bollerslev (1998) and formalised in Andersen et al. (2001). These estimates are constructed from high-frequency intraday returns sampled at five-minute intervals, and reflect the ex-post variation in asset prices observed within each trading day.

Formally, let the log-price process of an asset P_t follow:

$$dP_t = \mu_t dt + \sigma_t dW_t,$$

where μ_t is a locally bounded drift, σ_t is a stochastic volatility process, and W_t is a standard Brownian motion. Over a trading day t of length H , with intraday grid $0 = \tau_{t,0} < \tau_{t,1} < \dots < \tau_{t,M_t} = H$,

define the high-frequency log returns on day t by

$$r_{t,i} := P_{\tau_{t,i}} - P_{\tau_{t,i-1}}, \quad i = 1, \dots, M_t.$$

The integrated variance on day t is

$$\text{IntVar}_t = \int_0^H \sigma_s^2 ds$$

and the realised variance estimator is

$$RV_t := \sum_{i=1}^{M_t} r_{t,i}^2.$$

Under infill asymptotics (as the mesh of the grid goes to zero, i.e. $\max_i(\tau_{t,i} - \tau_{t,i-1}) \rightarrow 0$ with $M_t \rightarrow \infty$) and in the absence of microstructure noise or jumps, $RV_t \xrightarrow{p} \text{IntVar}_t$ (Barndorff-Nielsen & Shephard, 2002). In practice, we use a finite grid (e.g., five-minute sampling, $M_t \approx 78$ over a 6.5-hour session), so RV_t is only a finite-sample estimator of IntVar_t . Compared to squared daily returns, RV_t provides a more efficient and granular measure of return variation by exploiting intraday data and typically improves volatility measurement and forecasting (Andersen et al., 2003; Barndorff-Nielsen & Shephard, 2002).

In practice, however, RV_t is affected by microstructure frictions such as bid–ask bounce and asynchronous trading, which induce noise and bias when sampling at very high frequencies. To mitigate this, Zhang et al. (2005) proposed a subsampling approach that averages realised variance estimates across multiple offset sampling grids, thereby reducing sensitivity to the exact choice of sampling times.

Accordingly, this study adopts the `rv5_ss` series from the OMI Realized Library, where the suffix “ss” denotes subsampling. This measure computes five-minute realised variance across multiple staggered grids and then averages them (Equation 3), yielding a daily volatility proxy that is less affected by microstructure noise while retaining efficiency. Formally,

$$RV_t^{(ss)} = \frac{1}{L} \sum_{\ell=1}^L \sum_{j=1}^M r_{t,j}^{(\ell)2}, \quad (3)$$

where $r_{t,j}^{(\ell)}$ is the j -th five-minute return on grid ℓ and L is the number of such grids.

Figure 1 illustrates the `rv5_ss` series for selected indices in our dataset (see Table 2 for index abbreviations), highlighting the major volatility spikes during the 2008–2009 financial crisis and the 2020 COVID-19 shock.

All subsequent analysis in this study, and references to RV , are based on the `rv5_ss` series.

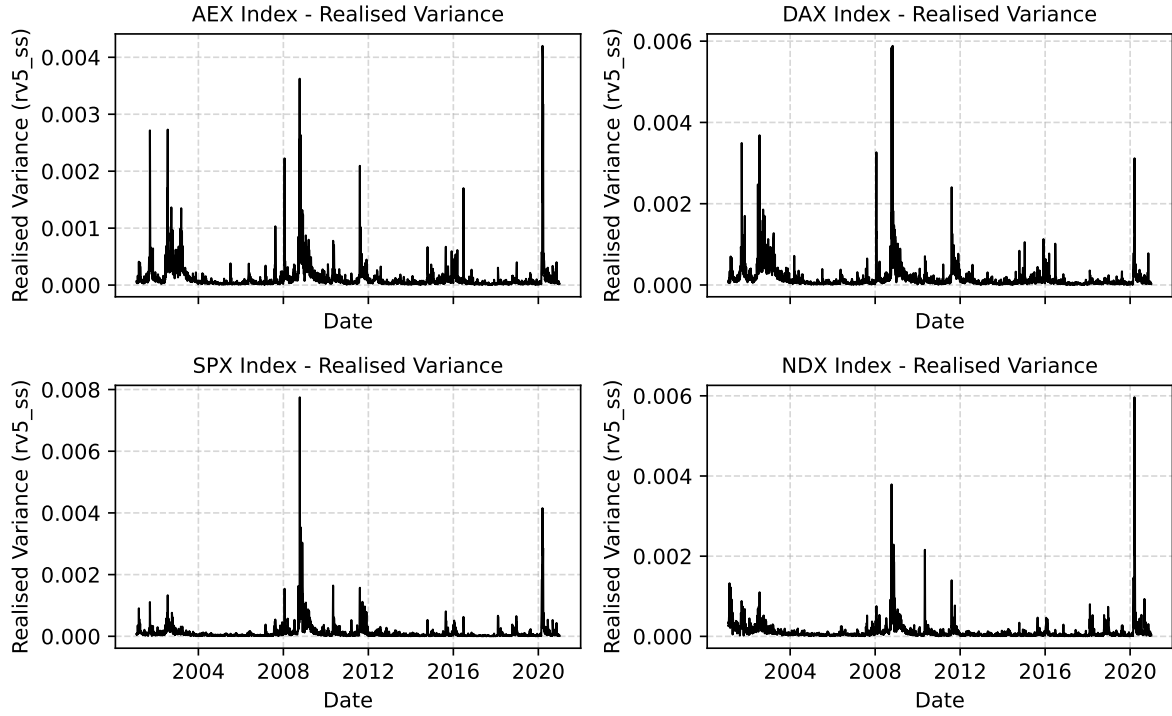


Figure 1.: Time series of realised variance (RV , measured using the `rv5_ss` estimator) for four representative equity indices.

In addition to realised variance, we also use daily log returns, sourced from the OMI Realized Library, to construct the sparse network via the graphical lasso (see Section 4.2).

3.3. Exogenous Variables

The primary methodological contribution of this study is to augment the GNAR-HAR model with several exogenous variables, motivated by prior literature on volatility forecasting, in an effort to improve forecast accuracy. This study adopts a simplified approach by including only the most recent daily (i.e., one-lag) value of each exogenous variable. This allows us to investigate whether the most up-to-date market information contains incremental predictive power beyond the autoregressive and network components. While previous work often incorporates exogenous variables using HAR-type lag structures (e.g., HAR-IV), exploring such formulations is left for future research.

- **Implied Volatility (IV):** For each stock index, we use its corresponding implied volatility index from Bloomberg (tickers listed in Table 2; e.g., S&P 500: VIX Index, field PX_LAST). This represents the market’s forward-looking expectation of future volatility over a fixed horizon (typically 30 days). These are computed using option prices and reflect risk-neutral expectations under a model-free framework. The model-free construction of the VIX is based on the theoretical results of Demeterfi et al. (1999), who show that risk-neutral expected variance can be obtained directly from a range of option prices without assuming a particular asset price model. For example, the VIX Index for the S&P 500 Index is computed using a wide range of out-of-the-money options.¹ Following Busch et al. (2011), we include lagged implied volatility to assess its incremental forecasting power.
- **Asymmetric Returns (Good/Bad Returns):** Motivated by the leverage effect (Black, 1976; Christie, 1982), and to allow for asymmetric volatility responses, we decompose daily returns for the stock indices into positive and negative parts:

$$r_t^+ = \max(r_t, 0), \quad r_t^- = \min(r_t, 0), \quad (4)$$

and include each as a separate regressor. This follows the intuition behind the

¹See the CBOE VIX White Paper for details: https://cdn.cboe.com/api/global/us_indices/governance/Cboe_Volatility_Index_Mathematics_Methodology.pdf

leverage HAR model of Corsi & Renò (2012), but without using longer-horizon averages.

- **Overnight Returns (ON):** We include lagged overnight returns,

$$r_t^{\text{overnight}} = \frac{\text{Open}_t}{\text{Close}_{t-1}} - 1, \quad (5)$$

to capture information arriving outside regular trading hours. Overnight movements can reflect global macroeconomic news, earnings announcements, or geopolitical shocks. Wang et al. (2015) show in Chinese equity markets that negative overnight returns improve volatility forecasts. Consistent evidence is found by Kambouroudis et al. (2021), who report that using overnight returns significantly improves volatility forecasting accuracy for most major international indices.

3.4. Summary Statistics

Table 3 presents summary statistics for the log of the realised variance series (as defined in Equation 3) of each index. The series exhibit strong persistence, as indicated by the high first-order autocorrelations (ranging from 0.77 to 0.86). The distributions also show moderate positive skewness, and all exhibit leptokurtosis (kurtosis ranging from 3.26 to 4.55), reflecting fat tails relative to the normal distribution, a well-documented feature of financial volatility. The SMI index stands out with the highest skewness (1.07) and kurtosis (4.55), suggesting occasional large spikes in volatility. These stylised facts support the use of long-memory or HAR-type models.

4. Methods

This section describes the forecasting framework, estimation procedure, network construction, and evaluation metrics used to assess GNAR-HAR(X) models for predicting the daily realised variance (RV) of ten major international stock indices. We benchmark

Index	Mean	Std. Dev.	Skew.	Kurt.	ACF(1)	ACF(2)	ACF(3)	PACF(1)	PACF(2)	PACF(3)
AEX	-9.644	1.015	0.639	3.376	0.856	0.819	0.797	0.857	0.320	0.193
CAC	-9.446	0.977	0.522	3.317	0.841	0.803	0.782	0.841	0.328	0.200
DAX	-9.348	1.031	0.520	3.319	0.845	0.808	0.789	0.845	0.329	0.211
DJI	-9.862	1.129	0.521	3.563	0.795	0.763	0.730	0.795	0.356	0.175
STX	-9.317	1.004	0.533	3.492	0.818	0.774	0.753	0.818	0.318	0.204
FTS	-9.647	1.010	0.693	3.712	0.785	0.751	0.728	0.785	0.353	0.207
NDX	-9.741	1.047	0.476	3.261	0.833	0.780	0.745	0.833	0.282	0.155
NKY	-9.702	0.950	0.410	3.521	0.773	0.722	0.696	0.773	0.309	0.196
SPX	-9.894	1.154	0.476	3.482	0.823	0.783	0.748	0.823	0.328	0.159
SMI	-9.890	0.912	1.068	4.554	0.856	0.821	0.803	0.856	0.331	0.211

Table 3.: Summary statistics of $\log RV$ for each index

these multivariate models against univariate HAR and HARX specifications, evaluated for both predictive accuracy and parameter efficiency. The HAR and HARX models are estimated using the HARX implementation from the `arch` Python package. This study provides the first direct comparison of GNAR-HAR(X) models with HAR/HARX models in this setting.

4.1. Forecasting Framework

Let $RV_{i,t}$ denote the realised variance of asset i on day t , constructed as described in Section 3.2. We model the log-transformed realised variance, $Y_{i,t} = \log RV_{i,t}$. Empirical results in Clements & Preve (2021) suggest that log-transformation improves forecast accuracy over using raw or square-root-transformed variance.

Each model is trained on an initial in-sample window of four years to avoid look-ahead bias (i.e., the use of future information that would not have been available at the time forecasts are made), which can lead to overly optimistic performance estimates.

For GNAR-HAR(X) models using graphical lasso (GL) networks, the initial four-year in-sample window is also used to estimate the first network structure (see Section 4.2 for details). At each subsequent refit step in the rolling window procedure, the GL network is re-estimated on the updated three-year window, allowing the network to evolve with the data.

We implement a rolling window forecasting procedure, repeating the following steps until the end of the evaluation period:

- Refit the model using the most recent three years of data (approximately 756 trading days). For models using GL networks, the network is re-estimated on this updated window.
- Generate one-step-ahead forecasts for the next 22 trading days.
- Shift the window forward by 22 trading days and repeat.

At each refit step, the response and all regressors are standardised within the three-year training window by subtracting their sample mean and dividing by their sample standard deviation. This normalisation ensures comparability across predictors and stabilises the estimation procedure.

This rolling procedure produces one-day-ahead forecasts across the out-of-sample evaluation period for each index. Estimation of the models requires only that the innovations $u_{i,t}$ are mean-zero with finite variance. However, when transforming the forecasts from $\log RV$ back to RV , an additional distributional assumption is needed to correct for Jensen’s inequality, since the exponential function is convex. We assume Gaussian residuals so that under this assumption, the unbiased estimator of RV is (see Appendix B for details):

$$\widehat{RV}_{i,t} = \exp \left(\widehat{Y}_{i,t} + \frac{1}{2} \widehat{\sigma}_i^2 \right), \quad (6)$$

where $\widehat{\sigma}_i^2$ is the in-sample residual variance from the rolling window.

In all GNAR-HAR(X) models, we restrict attention to first-order neighbourhoods, setting $r_d = r_w = r_m = 1$ for daily, weekly, and monthly spillovers. This approach, which follows Zhang et al. (2025b), is motivated by two considerations. First, in fully connected networks all nodes are directly linked, so higher-order neighbourhoods do not expand the information set. To ensure differences between fully connected and GL-based models reflect only the network structure, rather than neighbourhood orders, we also enforce

$r = 1$ in the GL case. Second, because a large number of model variants and combinations of exogenous predictors are already under comparison, we fix $r = 1$ to avoid expanding the model search space further. Preliminary experiments with $r > 1$ showed no consistent gains in forecast accuracy, similar to the findings of [Zhang et al. \(2025a\)](#), supporting this restriction. While alternative selection strategies, such as information criteria (e.g., BIC) or graphical diagnostics like the Corbit plot of [Nason et al. \(2023\)](#), could be considered, we adopt a fixed $r = 1$ specification.

4.2. Network Construction

We consider two network structures for the GNAR-based models.

(i) Fully Connected Network

Each node is connected to every other node (referred to as a complete graph in the graph theory literature). This choice reflects the high degree of integration in global financial markets, an effect observed most strongly during periods of market stress ([Korkusuz et al., 2023](#)).

(ii) Sparse Network via Graphical Lasso

To obtain a sparser, data-driven network, we estimate the network from the inverse covariance (precision) matrix of daily log returns by applying the graphical lasso ([Friedman et al., 2008](#)). The graphical lasso solves the optimisation problem:

$$\hat{\Theta} = \arg \max_{\Theta \succ 0} \{ \log \det(\Theta) - \text{tr}(S\Theta) - \rho |\Theta|_1 \}, \quad (7)$$

where S is the sample covariance matrix of the standardised log returns, Θ is the precision matrix (inverse of the covariance matrix), and ρ is a regularisation parameter controlling sparsity via the ℓ_1 norm.

In this study, the graphical lasso is implemented using the `GraphicalLasso` and `GraphicalLassoCV` classes from the `scikit-learn` Python library.

To avoid look-ahead bias, ρ is selected using ten-fold cross-validation on the initial four-year in-sample period. This chosen value is then held fixed throughout the forecasting procedure. At each refit, the graphical lasso is re-applied to the most recent three-year window with this fixed ρ , allowing the network to adapt over time while keeping the regularisation level consistent.

The adjacency matrix is then obtained by assigning an edge between nodes i and j whenever the corresponding off-diagonal entry of $\hat{\Theta}$ is non-zero, indicating conditional dependence between their daily log return series given all others. An example of a network produced by the graphical lasso is provided in Figure 10 (Appendix).

4.3. Evaluation Metrics

We assess model performance using two standard loss functions, computed on forecast errors $e_{i,t} = RV_{i,t} - \widehat{RV}_{i,t}$ and averaged over all indices $i = 1, \dots, N$ and dates $t = 1, \dots, T$:

1. Mean Squared Error (MSE):

$$\text{MSE} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T e_{i,t}^2. \quad (8)$$

2. Quasi-Likelihood (QLIKE):

$$\text{QLIKE} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left[\log \widehat{RV}_{i,t} + \frac{RV_{i,t}}{\widehat{RV}_{i,t}} \right]. \quad (9)$$

QLIKE is robust to measurement error in volatility and penalises large forecast errors less severely than MSE, particularly in the right tail. However, it is asymmetric, tending to

favour positively biased forecasts over equally sized negative ones (Patton, 2011). While QLIKE is often preferred in financial volatility contexts, reporting both metrics offers a more balanced assessment of forecasting performance.

5. Results

This section presents the empirical results from our forecasting study using the GNAR-HARX model and its benchmarks. Models are evaluated on one-step-ahead forecasting performance using mean squared error (MSE) and QLIKE loss within the rolling window framework described in Section 4.1.

5.1. Model Comparison

Table 4 summarises forecasting performance across all models, ordered by QLIKE. Relative performance is expressed as the ratio to the best-performing model for each respective metric, with both QLIKE and MSE reported.

The best overall model as judged by QLIKE is the **local GNAR-HAR** with a fully connected (FC) network and no exogenous regressors. By MSE, the top performer is the **standard GNAR-HARX** with implied volatility (IV) as the sole exogenous variable, also with an FC network. Differences from other leading GNAR-HAR(X) specifications are minimal, with relative QLIKE and MSE values almost indistinguishable to two decimal places.

A consistent pattern is that local and standard GNAR-HAR(X) variants slightly outperform the global version, though the gap is very small. The global specification achieves nearly identical performance but with fewer parameters.

Among GNAR-HAR(X) specifications, the top models (as judged by QLIKE) for each variant exclude exogenous predictors, suggesting that once network-aggregated lags are

Model	Variant	Network	Exogenous Variables	QLIKE	Rel QLIKE	Rel MSE
GNAR-HAR	local	FC	[]	-8.5891	1.00	1.00
GNAR-HARX	local	FC	['on']	-8.5889	1.00	1.00
GNAR-HAR	standard	FC	[]	-8.5884	1.00	1.01
GNAR-HARX	standard	FC	['on']	-8.5882	1.00	1.01
GNAR-HARX	standard	FC	['iv']	-8.5881	1.00	1.00
GNAR-HARX	local	FC	['good', 'bad']	-8.5881	1.00	1.01
GNAR-HARX	standard	FC	['iv', 'on']	-8.5880	1.00	1.00
GNAR-HARX	local	FC	['good', 'bad', 'on']	-8.5877	1.00	1.01
GNAR-HARX	local	FC	['iv']	-8.5876	1.00	1.01
GNAR-HAR	global	FC	[]	-8.5876	1.00	1.01
GNAR-HARX	global	FC	['iv']	-8.5876	1.00	1.00
GNAR-HARX	standard	FC	['good', 'bad']	-8.5875	1.00	1.01
GNAR-HARX	global	FC	['on']	-8.5874	1.00	1.01
GNAR-HARX	local	FC	['iv', 'on']	-8.5874	1.00	1.01
GNAR-HARX	global	FC	['iv', 'on']	-8.5874	1.00	1.00
GNAR-HARX	standard	FC	['iv', 'good', 'bad']	-8.5873	1.00	1.00
GNAR-HARX	standard	FC	['good', 'bad', 'on']	-8.5872	1.00	1.01
GNAR-HARX	standard	FC	['iv', 'good', 'bad', 'on']	-8.5871	1.00	1.00
GNAR-HARX	local	FC	['iv', 'good', 'bad']	-8.5869	1.00	1.02
GNAR-HAR	local	GL	[]	-8.5867	1.00	1.02
GNAR-HARX	global	FC	['good', 'bad']	-8.5867	1.00	1.01
GNAR-HARX	global	FC	['iv', 'good', 'bad']	-8.5867	1.00	1.01
GNAR-HARX	local	FC	['iv', 'good', 'bad', 'on']	-8.5866	1.00	1.02
GNAR-HARX	local	GL	['on']	-8.5865	1.00	1.02
GNAR-HARX	global	FC	['iv', 'good', 'bad', 'on']	-8.5864	1.00	1.01
GNAR-HARX	global	FC	['good', 'bad', 'on']	-8.5864	1.00	1.02
GNAR-HARX	local	GL	['good', 'bad']	-8.5857	1.00	1.02
GNAR-HAR	standard	GL	[]	-8.5856	1.00	1.03
GNAR-HARX	standard	GL	['iv']	-8.5854	1.00	1.02
GNAR-HARX	standard	GL	['on']	-8.5854	1.00	1.03
GNAR-HARX	local	GL	['iv']	-8.5854	1.00	1.02
GNAR-HARX	local	GL	['good', 'bad', 'on']	-8.5852	1.00	1.02
GNAR-HARX	standard	GL	['iv', 'on']	-8.5852	1.00	1.02
GNAR-HARX	local	GL	['iv', 'on']	-8.5851	1.00	1.02
GNAR-HAR	global	GL	[]	-8.5849	1.00	1.03
GNAR-HARX	global	GL	['iv']	-8.5848	1.00	1.02
GNAR-HARX	global	GL	['on']	-8.5846	1.00	1.03
GNAR-HARX	standard	GL	['good', 'bad']	-8.5846	1.00	1.03
GNAR-HARX	global	GL	['iv', 'on']	-8.5845	1.00	1.02
GNAR-HARX	local	GL	['iv', 'good', 'bad']	-8.5845	1.00	1.03
GNAR-HARX	standard	GL	['iv', 'good', 'bad']	-8.5845	1.00	1.02
GNAR-HARX	standard	GL	['good', 'bad', 'on']	-8.5843	1.00	1.03
GNAR-HARX	standard	GL	['iv', 'good', 'bad', 'on']	-8.5842	1.00	1.02
GNAR-HARX	local	GL	['iv', 'good', 'bad', 'on']	-8.5841	1.00	1.03
GNAR-HARX	global	GL	['good', 'bad']	-8.5839	1.00	1.03
GNAR-HARX	global	GL	['iv', 'good', 'bad']	-8.5838	1.00	1.02
GNAR-HARX	global	GL	['iv', 'good', 'bad', 'on']	-8.5835	1.00	1.03
GNAR-HARX	global	GL	['good', 'bad', 'on']	-8.5835	1.00	1.03
HARX	local	None	['iv']	-8.5831	1.00	17139.38
HARX	local	None	['iv', 'good', 'bad']	-8.5824	1.00	8669.26
HARX	local	None	['iv', 'on']	-8.5823	1.00	1763807.26
HARX	local	None	['iv', 'good', 'bad', 'on']	-8.5821	1.00	616700.95
HAR	local	None	[]	-8.5785	1.00	1.10
HARX	local	None	['good', 'bad']	-8.5777	1.00	1.47
HARX	local	None	['good', 'bad', 'on']	-8.5771	1.00	10.24
HARX	local	None	['on']	-8.5759	1.00	13.31

Table 4.: Forecasting performance of GNAR-HARX and benchmark models across configurations, ordered by QLIKE from best to worst. Exogenous abbreviations: *iv* = implied volatility, *good* and *bad* = positive and negative components of the previous return, and *on* = overnight return. Networks: *FC* = fully connected, *GL* = graphical lasso.

accounted for, external variables add little incremental value. Nevertheless, when exogenous variables are included, overnight returns and implied volatility are more often associated with stronger performance, whereas the leverage effect (captured by good/bad returns) plays a more limited role.

The univariate HAR and HARX benchmarks underperform the GNAR-based models in terms of both QLIKE and MSE. This is despite their requirement of many more parameters in some cases (e.g., a total of 40 across nodes in the HARX model with IV, compared to just 7 in the global GNAR-HARX model with IV). The best HARX specification with IV achieves a QLIKE of -8.5831 , close to but still below the leading GNAR-HARX models, and shows substantial instability in MSE. In particular, the HARX models using IV as an exogenous variable record extremely large relative MSE values (e.g., exceeding 1,000 or even 1,000,000) despite otherwise reasonable QLIKE. This behaviour is concentrated around 17 March 2020, when forecasts incorporated the VIX closing value from the previous day. On Monday, 16 March 2020 the VIX closed at 82.69, its highest closing level on record (at the time of writing), exceeding the previous high of 80.86 set on 20 November 2008 and reflecting acute market stress following the U.S. national emergency declaration on Friday, 13 March (Apergis et al., 2023). This value was well above levels observed in the refit window ending 14 February 2020, as shown in Figure 12 (Appendix). Because HARX is linear in $\log RV$, such an extreme IV input drives an unusually large shift in the linear predictor, and after the exponential back-transform this leads to outsized forecasts for RV . MSE penalises these forecasts quadratically, whereas QLIKE is less sensitive to extreme over-predictions, explaining the divergence in the two metrics.

By contrast, GNAR-HARX models that include IV remain stable in terms of MSE. Their estimated IV coefficients are materially smaller than in the univariate HARX, as shown in Figure 11 (Appendix). This suggests that the network-lag (β) terms already capture much of the common volatility factor that IV proxies. As a result, IV contributes only marginally on top of the network terms, and the forecasts avoid the extreme MSE spikes seen in the univariate case.

It is notable that the top-performing models, as judged by both QLIKE and MSE, use the FC network. In contrast, models using graphical lasso (GL) networks consistently rank lower, even when augmented with the same exogenous variables. These results imply that, in this setting, the potential benefits of data-driven sparsity appear outweighed by estimation challenges, leading to inferior predictive performance.

5.2. Coefficient Dynamics

5.2.1. Alpha and Beta Coefficients

To further understand the behaviour of the best global GNAR-HARX model (FC network with IV), we examine the evolution of its estimated parameters over the full out-of-sample evaluation period (February 2005 - December 2020), using the rolling-window estimates re-fitted at each step. Figure 2 plots the estimated α parameters, which capture the relative contribution of daily, weekly, and monthly lags of $\log RV$ for the same node. All three coefficients are positive and relatively stable over time, with α_w (weekly) generally dominating. This indicates that mid-range memory effects are the most persistent driver of volatility dynamics for this particular model.

Figure 3 shows the corresponding β parameters, which reflect lagged volatility spillovers from connected nodes. Here we observe more variability and some structural shifts. For example, β_d (daily) increases notably around 2007–2010, suggesting a stronger role for contemporaneous spillovers during the global financial crisis, and rises again post-2020 during the COVID-19 pandemic and associated market turbulence. This may indicate that market participants shifted their attention from longer horizons to daily signals, reflecting heightened sensitivity to new information during crises. By contrast, the negative values of β_w and β_m may point to mean-reverting spillover effects across nodes at these horizons.

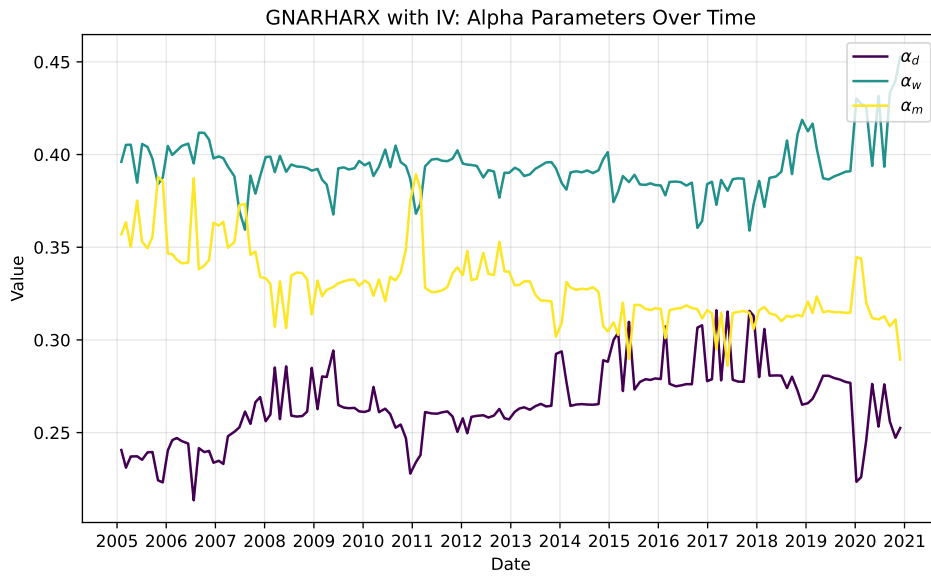


Figure 2.: Evolution of α coefficients over time for the best global GNAR-HARX model.

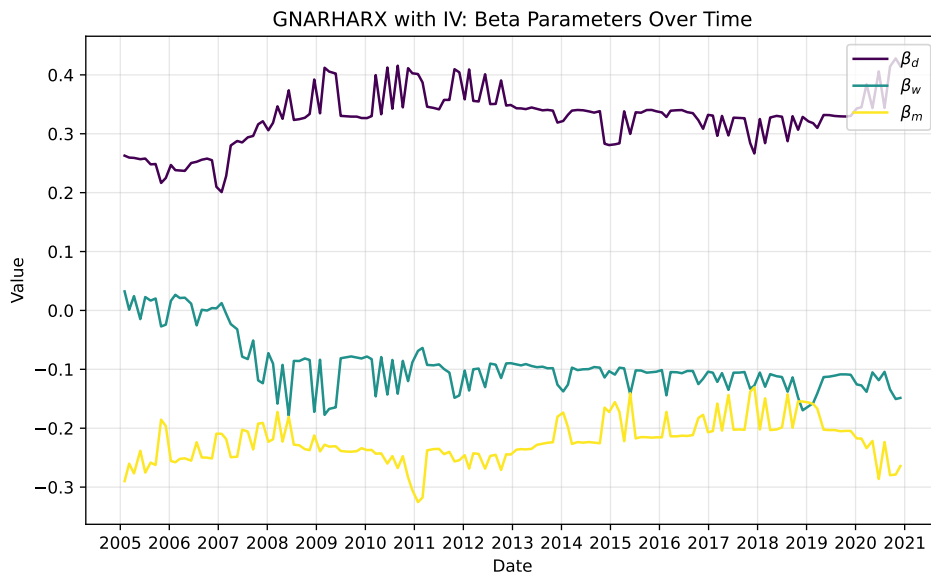


Figure 3.: Evolution of β coefficients over time for the best global GNAR-HARX model.

5.2.2. Exogenous Coefficients

Finally, Figure 4 presents the exogenous coefficient λ_{IV} for implied volatility. Although relatively small in magnitude throughout, the coefficient becomes more pronounced on

several occasions, most notably when the model is refitted on 23 March 2020, coinciding with the impact of the COVID-19 pandemic on financial markets. This aligns with the notion that IV can carry strong predictive power during periods of market uncertainty.

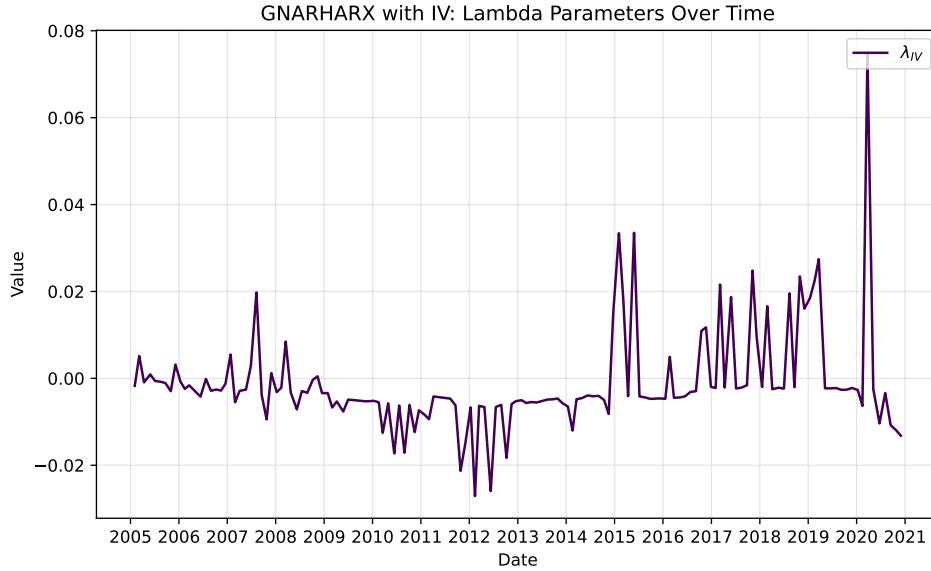


Figure 4.: Time-varying exogenous coefficients in the best global GNAR-HARX model using implied volatility.

5.3. Residual Analysis

To further evaluate the performance of the best global GNAR-HARX model (FC network with IV), we conduct a residual analysis for a representative node, the SPX Index (S&P 500).

Figure 5 presents predicted versus actual values for realised volatility, i.e. the square-root of realised variance ($\sqrt{RV}$). This transformation is commonly used for visualisation because it compresses large spikes while preserving interpretability in volatility units. Visually, the model tracks periods of market turbulence (e.g., the 2008 crisis and 2020 COVID shock) reasonably well, though some underestimation is evident at extreme peaks.

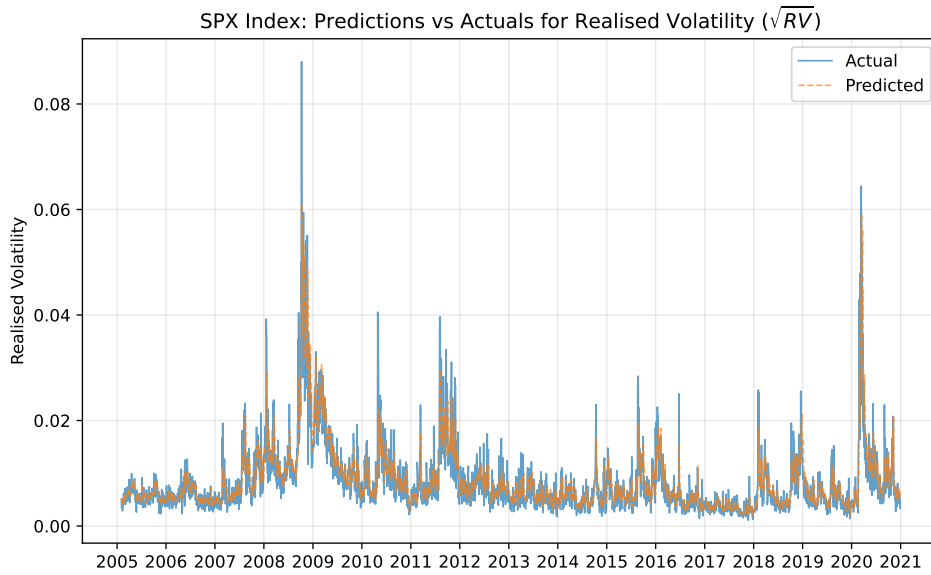


Figure 5.: Predicted vs actual realised volatility ($\sqrt{RV}$) for the SPX Index using the best global GNAR-HARX model.

Figures 6 and 7 show residuals in the log-realised variance space, the scale on which the model was estimated. Analysing residuals on this scale is more appropriate, as it aligns with the assumed error distribution and ensures that diagnostic properties (e.g., normality and mean-zero errors) reflect the model's true fitting behaviour.

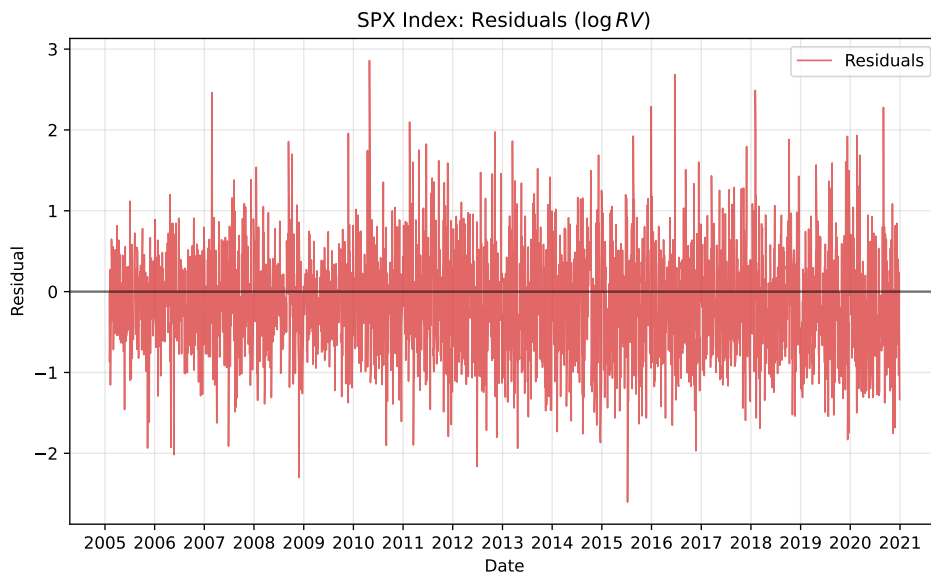


Figure 6.: Residuals (log-realised variance) over time for the SPX Index.

The residual time series (Figure 6) is centred around zero, with occasional large residuals during volatile periods. This suggests that while the model may adapt well to regular market conditions, some extreme shocks are not fully captured.

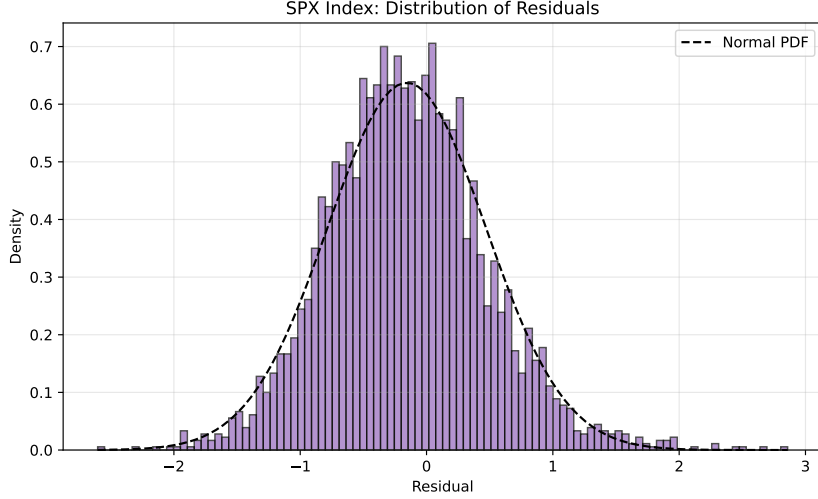


Figure 7.: Distribution of residuals (log-realised variance) for the SPX Index, with overlaid normal density.

The histogram in Figure 7 shows a roughly symmetric, bell-shaped distribution with heavier tails than a normal distribution, particularly on the right. The slight positive skew and excess kurtosis suggest occasional under-prediction of realised variance, motivating the consideration of heavier-tailed error distributions.

Similar analyses for the remaining indices are provided in Appendix D. These confirm that residuals are generally centred around zero (Figure 13) and approximately bell-shaped with heavier tails than normality would suggest (Figure 14). This indicates that the diagnostic features observed for the SPX Index also hold across the full panel.

5.4. Network Structure Analysis

We further explore the performance gap between models using FC and GL networks. While the GL network is dynamically refitted at each training window to capture time-varying relationships between indices, this flexibility does not translate into improved

forecasting performance. This may be due to instability or over-fitting in sparse graphs, especially in high-volatility periods.

5.5. Graphical Lasso Network Dynamics

To better understand the evolving structure of the GL network, we investigate the evolution of the adjacency matrices used. Although these networks are refitted monthly to capture time-varying relationships between nodes, their use did not consistently improve out-of-sample performance relative to the simpler FC network.

Figure 8 shows the number of non-zero edges in the graphical lasso network over time. From 2005 through 2015, the number of active connections remains relatively stable, ranging between 26 and 32, suggesting a relatively consistent network structure during this period. However, from 2016 onward, we observe more variance in edge count, with sharp drops in connectivity, especially from 2018 to 2020. These changes may be driven either by over-shrinkage of the penalised estimator in noisy conditions or by genuine structural breaks.

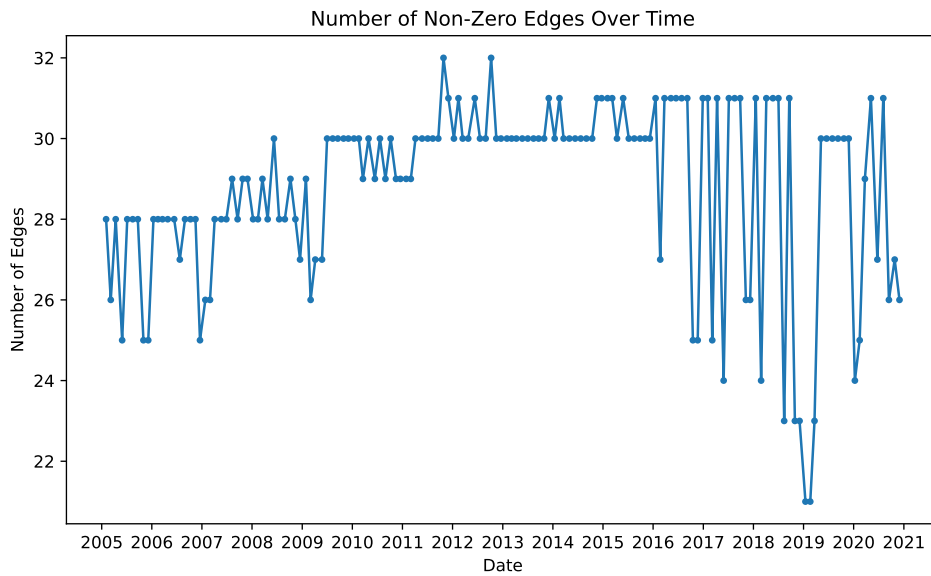


Figure 8.: Number of non-zero edges in the GL network over time, reflecting changing estimated market connectivity.

To measure structural persistence, we compute the Jaccard similarity index between consecutive networks. The Jaccard index is well suited to this task because it directly quantifies the proportion of common edges across two networks, relative to the total number of distinct edges. Unlike raw edge counts, it provides a scale-free measure of overlap that lies in $[0, 1]$, enabling us to track the degree of structural stability or turnover in the estimated adjacency matrices over time. Specifically, for two binary adjacency matrices $A^{(t)}$ and $A^{(t-1)}$, we first extract the upper-triangular entries (excluding the diagonal) to avoid double-counting. Let $\mathbf{a}^{(t)}$ and $\mathbf{a}^{(t-1)}$ denote these binary edge indicator vectors. The Jaccard similarity is then defined as

$$J(\mathbf{a}^{(t)}, \mathbf{a}^{(t-1)}) = \frac{|\mathbf{a}^{(t)} \cap \mathbf{a}^{(t-1)}|}{|\mathbf{a}^{(t)} \cup \mathbf{a}^{(t-1)}|},$$

where the numerator counts the number of edges present in both networks, and the denominator counts the number of edges present in at least one of the two networks. This yields a value between zero and one, with one indicating identical edge sets and zero indicating no common edges.

As shown in Figure 9, the Jaccard index remains close to one for much of the sample, indicating that many edges persist across consecutive refit windows. From around 2016 onwards, however, the index becomes more erratic, with frequent dips below 0.8. This coincides with the period of greater edge-count variability and suggests that the estimated network structure is less stable in later years.

This instability can be interpreted in two ways. On one hand, it may reflect heightened sensitivity of the graphical lasso to noise, reducing the robustness of the estimated networks for forecasting. On the other hand, it may be capturing genuine time variation in cross-market linkages, in which case the shifting networks could be seen as appropriately adapting to changing market regimes. A possible explanation for the weaker forecasting performance of GL-based models is that, while they can adapt, the monthly refitting on a three-year rolling window may be too slow to fully capture more abrupt structural shifts, leaving parts of the estimated network stale between refits.

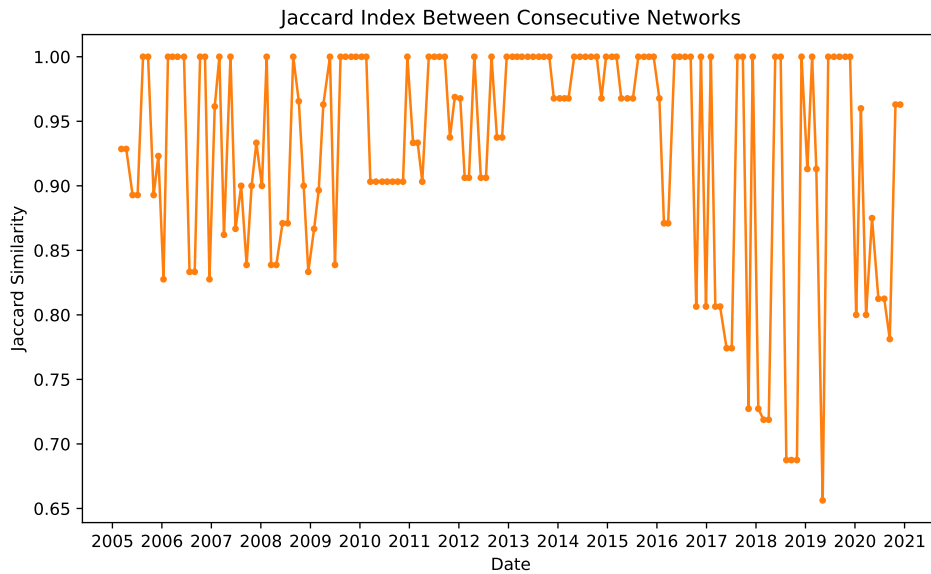


Figure 9.: Jaccard similarity index between consecutive GL networks. Values closer to one indicate higher structural persistence.

Summary of results:

- The best-performing model found by QLIKE is a **local GNAR-HAR** with an FC network and no exogenous variables.
- The lowest MSE is achieved by a **standard GNAR-HARX** model with implied volatility as the sole exogenous input, also using an FC network.
- **FC networks** generally outperformed GL networks.
- **Overnight returns** and **implied volatility** appear frequently in the top-performing GNAR-HARX configurations.
- Standard univariate models (HAR, HARX) were consistently outperformed by GNAR-HAR(X) models, while often having more estimated parameters.
- Estimated model coefficients, the autoregressive (α), network spillover (β), and exogenous (λ) terms, showed meaningful variation over time.
- Analysis of GL network structure revealed stability in earlier years, but increasing variance and reduced structural persistence from 2016 onward.

- Residual analysis indicated generally good model fit, but some underestimation during extreme volatility events, with evidence of skewness and heavy tails.

6. Discussion and Future Work

6.1. Discussion

Our results demonstrate that GNAR-HAR(X) models outperform univariate HAR/HARX benchmarks, even when there are considerably fewer parameters to estimate. The top-performing specification as judged by QLIKE was a local GNAR-HAR model with an FC network and no exogenous variables, while the best model by MSE was a standard GNAR-HARX model using IV. Notably, the univariate HARX models, while competitive when judged by QLIKE, show greater sensitivity to very large spikes in implied volatility, leading to occasional extreme forecast errors and higher MSE.

The consistent underperformance of GL networks, relative to FC networks, highlights an important trade-off. Although the GL networks are refitted dynamically to adapt to changing market conditions, this adaptivity came at the cost of stability. Early in the sample the estimated networks were relatively persistent, but from 2016 onward they became more volatile, with sharper drops in edge count and more frequent dips in Jaccard similarity of consecutive networks. This instability likely reflects both estimation noise and genuine structural change, but in either case it reduced forecasting accuracy. In this setting, the variance costs of adaptivity appear to outweigh its potential benefits, making the simpler FC network more effective.

Analysing model coefficient estimates over time for the best global GNAR-HARX model (FC network with IV) offers further insight into model behaviour. Autoregressive coefficients (α) remained stable, with mid-horizon (weekly) lags dominant. Network spillovers (β), however, were less stable, with more variation evident following crisis periods such as

2008 and 2020. The exogenous IV coefficient (λ_{IV}) was generally small but exhibited distinct spikes during market stress, consistent with its interpretation as a forward-looking risk indicator.

Residual analysis of the best global GNAR-HARX model confirmed a good overall fit but revealed slight underestimation during extreme volatility events. The residuals showed mild positive skew and excess kurtosis in the log RV space, suggesting that alternative error models with heavier tails (e.g., Student- t) may be more appropriate.

6.2. Limitations and Future Work

While the GNAR-HARX framework demonstrates promising empirical performance, several limitations remain that suggest directions for future research.

First, we do not compare directly against multivariate volatility models without a network structure, such as VAR-HAR (Bubák et al., 2011; Souček & Todorova, 2013). These could serve as a natural benchmark to isolate the added value of incorporating a network structure. Given the large number of models already evaluated, such comparisons were left for future work, but would clarify whether the GNAR structure offers unique benefits over more standard multivariate alternatives. Future work might focus on a narrower comparison between the best GNAR-type models identified here and alternative multivariate approaches.

Second, the assumption of a known or correctly estimated network structure is strong. While we implemented graphical lasso to construct networks from the estimated time-varying sparse precision matrices, the resulting networks appeared less stable in later years of the sample, potentially undermining the forecasting performance of GL-based models. Alternative strategies may perform better. For example, Nason & Palasciano (2025) report success using randomly generated networks in the RaGNAR model for UK inflation forecasting. Such findings suggest that stochastic or ensemble-based network designs could offer a promising direction for future research.

Third, our analysis was limited to ten international stock indices with available high-frequency and implied volatility data. This restriction was driven by data availability and the need for reliable implied volatility series. Expanding to a broader cross-section of markets, or applying the framework to individual stocks, where IV could be constructed from options data, would allow for a broader test of model scalability and generalisability.

Fourth, model estimation relied on ordinary least squares (OLS), which may become unreliable as the number of predictors grows with the number of nodes or exogenous variables. This is especially true in local GNAR-HAR(X) models. Regularised regression methods, such as ridge or lasso, could improve estimation stability and predictive accuracy. Preliminary results with ridge regression on simulated data from local models support this idea, and future work could explore penalised or Bayesian estimation.

Fifth, the residual analysis suggests departures from normality, with heavier tails and occasional large outliers, a common feature in financial volatility data. It would be natural to consider heavier-tailed error distributions, such as the Student- t , to account for occasional large outliers more effectively. Some work in this direction has already been undertaken for GNAR-type models ([Olawale Olanrewaju et al., 2023](#)). Another promising extension is to model continuous and jump components of volatility separately, for example by combining GNAR-type dependence structures with jump-robust realised variance estimators ([Andersen et al., 2007](#)).

Finally, while all models in this analysis were evaluated using one-step-ahead forecasts, many practical applications require multi-step horizons. Extending the framework to multi-horizon forecasting, while also enriching the lag structure of exogenous variables beyond the single daily lag considered here, would provide a more comprehensive assessment of model performance and may enhance forecast accuracy.

To summarise, while the GNAR-HARX framework offers a flexible and interpretable structure for multivariate volatility forecasting, there remains considerable scope for methodological refinement and broader empirical validation.

7. Endmatter

- **Data and reproducibility:** The code and data required to reproduce the results of this study are available in a private GitHub repository at <https://github.com/tomonuallain/msc-project>. Access can be granted on request via email (tso24@ic.ac.uk or tomonuallain@gmail.com).
- **Use of AI tools:** Conversational AI tools (ChatGPT) were occasionally used to assist with rephrasing or shortening sentences. All substantive content, analysis, and interpretation are my own.

References

- Andersen, T. G. & Bollerslev, T. (1998). Towards a unified framework for high and low frequency return volatility modeling. *Statistica Neerlandica*, 52(3), 273–302. DOI: [10.1111/1467-9574.00085](https://doi.org/10.1111/1467-9574.00085).
- Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility. *Review of Economics and Statistics*, 89(4), 701–720. DOI: [10.1162/rest.89.4.701](https://doi.org/10.1162/rest.89.4.701).
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61(1), 43–76. DOI: [10.1016/S0304-405X\(01\)00055-1](https://doi.org/10.1016/S0304-405X(01)00055-1).
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and Forecasting Realized Volatility. *Econometrica*, 71(2), 579–625. DOI: [10.1111/1468-0262.00418](https://doi.org/10.1111/1468-0262.00418).
- Apergis, N., Mustafa, G., & Malik, S. (2023). The role of the COVID-19 pandemic in US market volatility: Evidence from the VIX index. *The Quarterly Review of Economics and Finance*, 89, 27–35. DOI: [10.1016/j.qref.2023.03.004](https://doi.org/10.1016/j.qref.2023.03.004).
- Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73(1), 5–59. DOI: [10.1016/0304-4076\(95\)01732-1](https://doi.org/10.1016/0304-4076(95)01732-1).
- Barndorff-Nielsen, O. E. & Shephard, N. (2002). Econometric Analysis of Realized Volatility and its Use in Estimating Stochastic Volatility Models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 64(2), 253–280. DOI: [10.1111/1467-9868.00336](https://doi.org/10.1111/1467-9868.00336).
- Black, F. (1976). Studies of Stock Price Volatility Changes. In *Proceedings of the American Statistical Association, Business and Economics Statistics Section* (pp. 177–181). Washington, DC: American Statistical Association.
- Bubák, V., Kočenda, E., & Žikeš, F. (2011). Volatility transmission in emerging European foreign exchange markets. *Journal of Banking & Finance*, 35(11), 2829–2841. DOI: [10.1016/j.jbankfin.2011.03.012](https://doi.org/10.1016/j.jbankfin.2011.03.012).
- Buncic, D. & Gisler, K. I. M. (2016). Global equity market volatility spillovers: A broader role for the United States. *International Journal of Forecasting*, 32(4), 1317–

1339. DOI: [10.1016/j.ijforecast.2016.05.001](https://doi.org/10.1016/j.ijforecast.2016.05.001).
- Busch, T., Christensen, B. J., & Nielsen, M. (2011). The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics*, 160(1), 48–57. DOI: [10.1016/j.jeconom.2010.03.014](https://doi.org/10.1016/j.jeconom.2010.03.014).
- Callot, L. A. F., Kock, A. B., & Medeiros, M. C. (2017). Modeling and Forecasting Large Realized Covariance Matrices and Portfolio Choice. *Journal of Applied Econometrics*, 32(1), 140–158. DOI: [10.1002/jae.2512](https://doi.org/10.1002/jae.2512).
- Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics*, 10(4), 407–432. DOI: [10.1016/0304-405X\(82\)90018-6](https://doi.org/10.1016/0304-405X(82)90018-6).
- Clements, A. & Preve, D. P. A. (2021). A Practical Guide to harnessing the HAR volatility model. *Journal of Banking & Finance*, 133, 106285. DOI: [10.1016/j.jbankfin.2021.106285](https://doi.org/10.1016/j.jbankfin.2021.106285).
- Corsi, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics*, 7(2), 174–196. DOI: [10.1093/jjfinec/nbp001](https://doi.org/10.1093/jjfinec/nbp001).
- Corsi, F. & Renò, R. (2012). Discrete-Time Volatility Forecasting With Persistent Leverage Effect and the Link With Continuous-Time Volatility Modeling. *Journal of Business & Economic Statistics*, 30(3), 368–380. DOI: [10.1080/07350015.2012.663261](https://doi.org/10.1080/07350015.2012.663261).
- Demeterfi, K., Derman, E., Kamal, M., & Zou, J. (1999). *More Than You Ever Wanted to Know About Volatility Swaps*. Technical report, Goldman Sachs. URL: https://www.researchgate.net/publication/246869706_More_Than_You_Ever_Wanted_to_Know_About_Volatility_Swaps.
- Engle, R. F. & Kroner, K. F. (1995). Multivariate Simultaneous Generalized ARCH. *Econometric Theory*, 11(1), 122–150. DOI: [10.1017/S0266466600009063](https://doi.org/10.1017/S0266466600009063).
- Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3), 432–441. DOI: [10.1093/biostatistics/kxm045](https://doi.org/10.1093/biostatistics/kxm045).
- Gourieroux, C., Jasiak, J., & Sufana, R. (2009). The Wishart Autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*, 150(2), 167–181. DOI: [10.1016/j.jeconom.2008.12.016](https://doi.org/10.1016/j.jeconom.2008.12.016).
- Heber, G., Lunde, A., Shephard, N., & Sheppard, K. (2009). OMI's realised library.

Version 0.1, Oxford & Man Institute, University of Oxford.

- Kambouroudis, D. S., McMillan, D. G., & Tsakou, K. (2021). Forecasting realized volatility: The role of implied volatility, leverage effect, overnight returns, and volatility of realized volatility. *Journal of Futures Markets*, 41(10), 1618–1639. DOI: [10.1002/fut.22241](https://doi.org/10.1002/fut.22241).
- Knight, M., Leeming, K., Nason, G., & Nunes, M. (2019). Generalised Network Autoregressive Processes and the GNAR package. DOI: [10.48550/arXiv.1912.04758](https://doi.org/10.48550/arXiv.1912.04758).
- Knight, M., Leeming, K., Nason, G., & Nunes, M. (2020). Generalized Network Autoregressive Processes and the GNAR Package. *Journal of Statistical Software*, 96(5). DOI: [10.18637/jss.v096.i05](https://doi.org/10.18637/jss.v096.i05).
- Knight, M. I., Nunes, M. A., & Nason, G. P. (2016). Modelling, Detrending and Decorrelation of Network Time Series. DOI: [10.48550/arXiv.1603.03221](https://doi.org/10.48550/arXiv.1603.03221).
- Korkusuz, B., McMillan, D. G., & Kambouroudis, D. (2023). Complex network analysis of volatility spillovers between global financial indicators and G20 stock markets. *Empirical Economics*, 64(4), 1517–1537. DOI: [10.1007/s00181-022-02290-w](https://doi.org/10.1007/s00181-022-02290-w).
- Nason, G., Salnikov, D., & Cortina-Borja, M. (2023). New tools for network time series with an application to COVID-19 hospitalisations. DOI: [10.48550/arXiv.2312.00530](https://doi.org/10.48550/arXiv.2312.00530).
- Nason, G., Salnikov, D., & Cortina-Borja, M. (2024). Modelling clusters in network time series with an application to presidential elections in the USA. DOI: [10.48550/arXiv.2401.09381](https://doi.org/10.48550/arXiv.2401.09381).
- Nason, G. P. & Palasciano, H. A. (2025). Forecasting UK consumer price inflation with RaGNAR: Random generalised network autoregressive processes. *International Journal of Forecasting*. DOI: [10.1016/j.ijforecast.2025.04.005](https://doi.org/10.1016/j.ijforecast.2025.04.005) (in press).
- Nason, G. P. & Wei, J. L. (2022). Quantifying the Economic Response to COVID-19 Mitigations and Death Rates Via Forecasting Purchasing Managers' Indices Using Generalised Network Autoregressive Models with Exogenous Variables. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185(4), 1778–1792. DOI: [10.1111/rssa.12875](https://doi.org/10.1111/rssa.12875).
- Olawale Olanrewaju, R., Ranjan, R. P., C. Chukwudum, Q., & Olanrewaju, S. A. (2023). Vulgarized neighboring network of multivariate autoregressive processes with Gaussian and Student-t distributed random noise. *Malaysian Journal of Computing (MJoC)*,

- 8(2), 1574–1588. DOI: [10.24191/mjoc.v8i2.23103](https://doi.org/10.24191/mjoc.v8i2.23103).
- Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1), 246–256. DOI: [10.1016/j.jeconom.2010.03.034](https://doi.org/10.1016/j.jeconom.2010.03.034).
- Rigobon, R. & Sack, B. (2003). *Spillovers Across U.S. Financial Markets*. Working Paper 9640, National Bureau of Economic Research. DOI: [10.3386/w9640](https://doi.org/10.3386/w9640).
- Souček, M. & Todorova, N. (2013). Realized volatility transmission between crude oil and equity futures markets: A multivariate HAR approach. *Energy Economics*, 40, 586–597. DOI: [10.1016/j.eneco.2013.08.011](https://doi.org/10.1016/j.eneco.2013.08.011).
- Tapia Costa, M., Cucuringu, M., & Nason, G. P. (2025). Higher Order Dynamic Network Linear Models for Covariance Forecasting. DOI: [10.2139/ssrn.5113698](https://doi.org/10.2139/ssrn.5113698) (submitted for publication).
- Wang, X., Wu, C., & Xu, W. (2015). Volatility forecasting: The role of lunch-break returns, overnight returns, trading volume and leverage effects. *International Journal of Forecasting*, 31(3), 609–619. DOI: [10.1016/j.ijforecast.2014.10.007](https://doi.org/10.1016/j.ijforecast.2014.10.007).
- Wang, Y., Pan, Z., & Wu, C. (2018). Volatility spillover from the US to international stock markets: A heterogeneous volatility spillover GARCH model. *Journal of Forecasting*, 37(3), 385–400. DOI: [10.1002/for.2509](https://doi.org/10.1002/for.2509).
- Zhang, C., Pu, X., Cucuringu, M., & Dong, X. (2025a). Forecasting realized volatility with spillover effects: Perspectives from graph neural networks. *International Journal of Forecasting*, 41(1), 377–397. DOI: [10.1016/j.ijforecast.2024.09.002](https://doi.org/10.1016/j.ijforecast.2024.09.002).
- Zhang, C., Pu, X., Cucuringu, M., & Dong, X. (2025b). Graph-Based Methods for Forecasting Realized Covariances. *Journal of Financial Econometrics*, 23(2). DOI: [10.1093/jjfinc/nbae026](https://doi.org/10.1093/jjfinc/nbae026).
- Zhang, L., Mykland, P. A., & Aït-Sahalia, Y. (2005). A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data. *Journal of the American Statistical Association*, 100(472), 1394–1411. DOI: [10.1198/016214505000000169](https://doi.org/10.1198/016214505000000169).
- Zhu, X., Pan, R., Li, G., Liu, Y., & Wang, H. (2017). Network Vector Autoregression. *The Annals of Statistics*, 45(3). DOI: [10.1214/16-AOS1476](https://doi.org/10.1214/16-AOS1476).

A. Derivation of Stationarity Constraint

Assuming stationarity of the exogenous processes $\{X_{h,i,t}\}$, we follow Appendix A.1 of [Tapia Costa et al. \(2025\)](#), where GNAR-HAR is formulated as a GNAR(22, $\mathbf{r}$) model. The HAR coefficients correspond to:

$$\begin{aligned}\alpha_{i,1} &= \alpha_d, & \alpha_{i,j} &= \alpha_w/4 \text{ for } j = 2, \dots, 5, & \alpha_{i,j} &= \alpha_m/17 \text{ for } j = 6, \dots, 22, \\ \beta_{1,r} &= \beta_{d,r}, & \beta_{j,r} &= \beta_{w,r}/4 \text{ for } j = 2, \dots, 5, & \beta_{j,r} &= \beta_{m,r}/17 \text{ for } j = 6, \dots, 22.\end{aligned}$$

Here the per-lag maximum neighbourhood stages are $s_1 = r_d$, $s_j = r_w$ for $j = 2, \dots, 5$, and $s_j = r_m$ for $j = 6, \dots, 22$.

For a GNARX($p, \mathbf{s}, \mathbf{p}'$) process with stationary exogenous inputs, the sufficient condition for stationarity from [Nason & Wei \(2022\)](#) is:

$$\sum_{j=1}^p \left(|\alpha_{i,j}| + \sum_{s=1}^{s_j} |\beta_{j,s}| \right) < 1 \quad \forall i.$$

Substituting in the HAR-to-GNARX parameter mappings, we obtain:

$$\begin{aligned}\sum_{j=1}^{22} \left(|\alpha_{i,j}| + \sum_{r=1}^{s_j} |\beta_{j,r}| \right) &= |\alpha_d| + \sum_{r=1}^{r_d} |\beta_{d,r}| \\ &\quad + \sum_{j=2}^5 \left(\frac{|\alpha_w|}{4} + \sum_{r=1}^{r_w} \frac{|\beta_{w,r}|}{4} \right) \\ &\quad + \sum_{j=6}^{22} \left(\frac{|\alpha_m|}{17} + \sum_{r=1}^{r_m} \frac{|\beta_{m,r}|}{17} \right) \\ &= |\alpha_d| + |\alpha_w| + |\alpha_m| + \sum_{r=1}^{r_d} |\beta_{d,r}| + \sum_{r=1}^{r_w} |\beta_{w,r}| + \sum_{r=1}^{r_m} |\beta_{m,r}|.\end{aligned}$$

This yields the constraint stated in the main text.

B. Jensen's Inequality Adjustment

In this project, models are estimated on the log of realised variance, so that forecasts are obtained for $Y_{i,t} = \log RV_{i,t}$.

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\mathbb{E}[e^X] = \int_{-\infty}^{\infty} e^x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx.$$

Completing the square inside the exponential gives

$$\mathbb{E}[e^X] = \exp\left(\mu + \frac{1}{2}\sigma^2\right).$$

Applying this conditionally with $X = \widehat{Y}_{i,t} + u_{i,t}$, where $u_{i,t} \mid \mathcal{F}_{t-1} \sim \mathcal{N}(0, \sigma_{i,t}^2)$, yields

$$\mathbb{E}[RV_{i,t} \mid \mathcal{F}_{t-1}] = \exp\left(\widehat{Y}_{i,t} + \frac{1}{2}\sigma_{i,t}^2\right).$$

which is the corrected back-transformation in (6).

C. Graphical Lasso Network Example

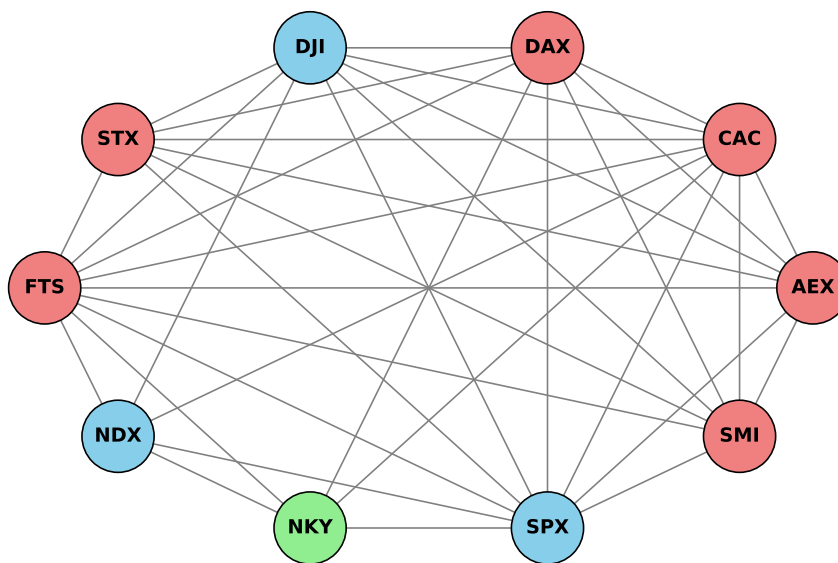


Figure 10.: Example of a graphical lasso network, with nodes coloured by region.

D. Further Residual Analysis

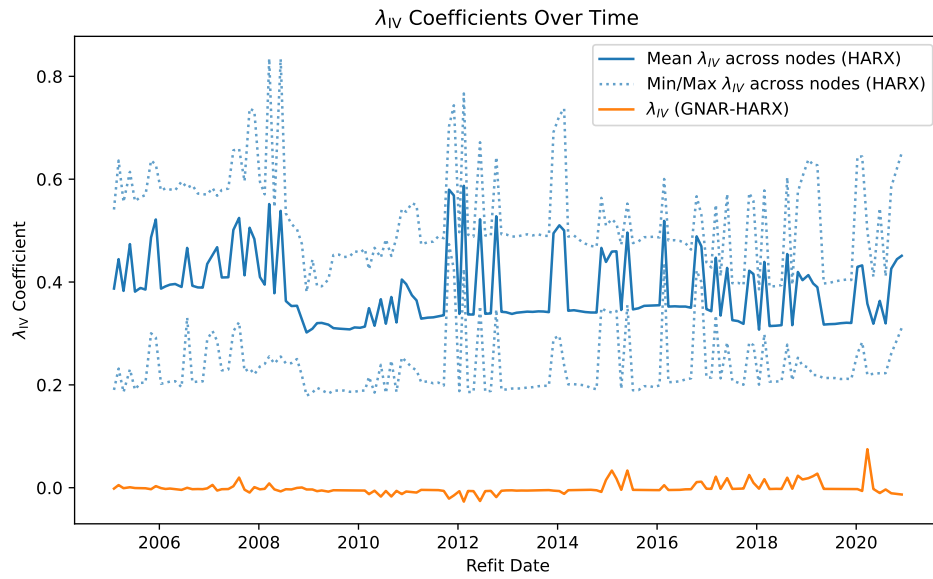


Figure 11.: Estimated IV coefficients (λ_{IV}) over refits for HARX (node-wise mean and range) and global GNAR-HARX. HARX shows larger and more variable loadings on IV, whereas GNAR-HARX loadings are near zero and stable.

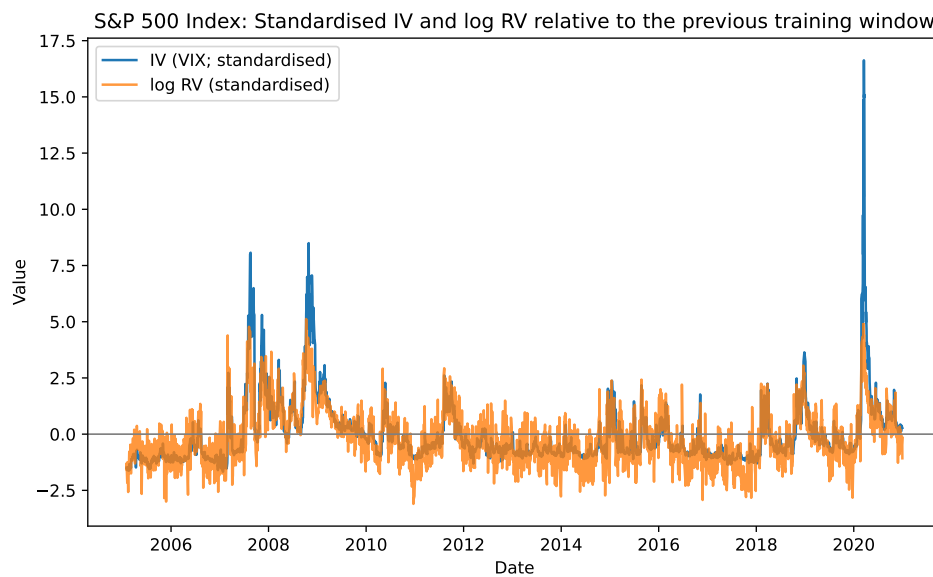


Figure 12.: S&P 500 (SPX) Index: IV (VIX) vs log RV , both standardised relative to the previous refit window. The IV spike in March 2020 helps explain HARX(IV) MSE instability.

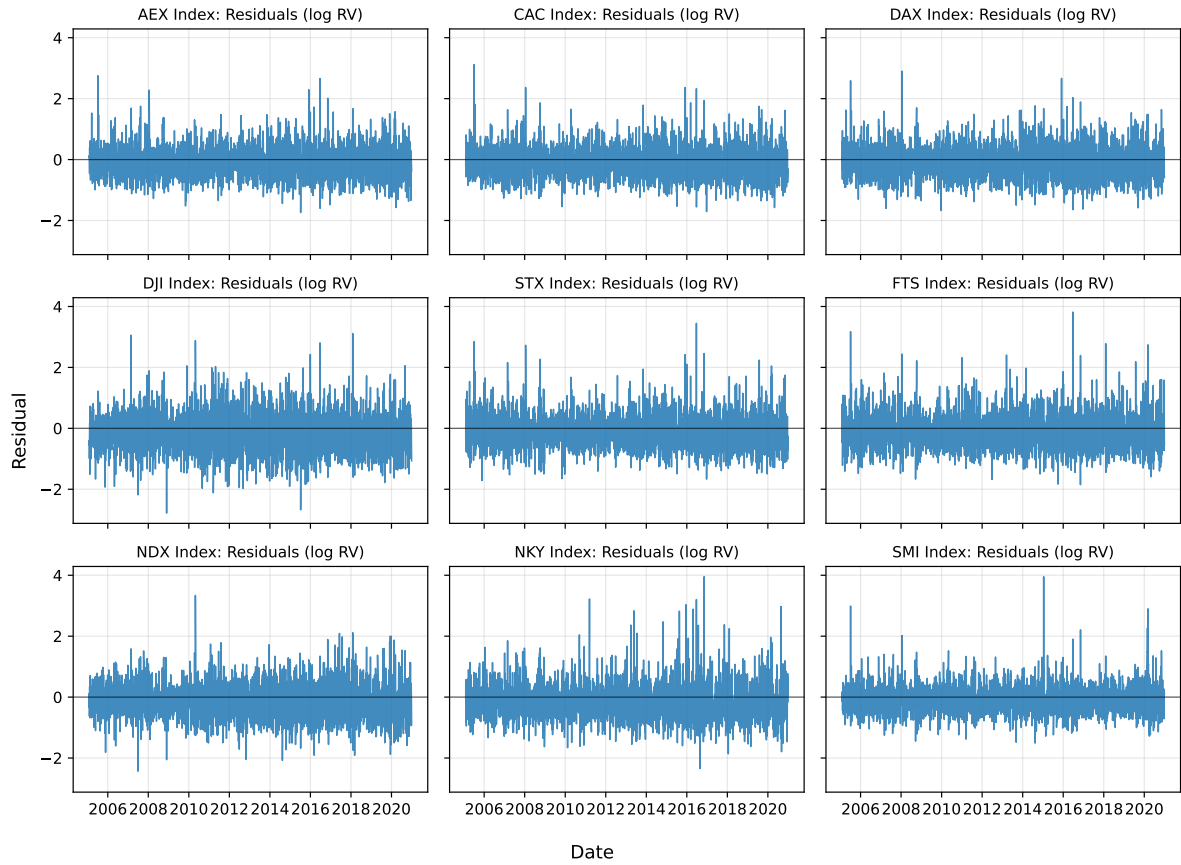


Figure 13.: Residuals (log-realised variance) over time for the other nine indices under the best global GNAR-HARX model (FC network with IV). Residuals are generally centred around zero, with occasional large deviations.

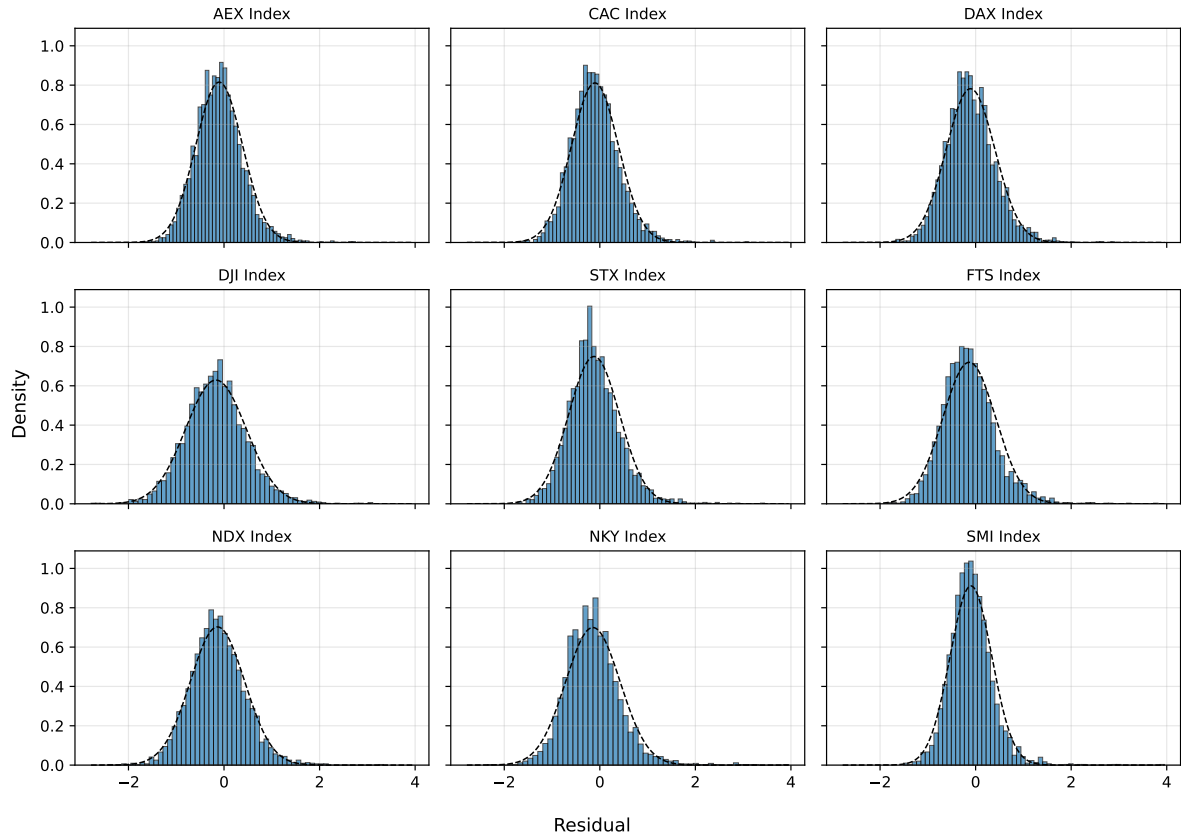


Figure 14.: Distribution of residuals (log-realised variance) for the other nine indices under the best global GNAR-HARX model (FC network with IV). Histograms are overlaid with a normal density fit. Distributions are broadly symmetric but exhibit heavier tails than normality, consistent with the SPX case.