

Article

Not peer-reviewed version

Topological Conditioning via Natural Visibility Graphs for Monte Carlo Simulation of Power Prices

[Carlo Mari](#) * and [Emiliano Mari](#)

Posted Date: 7 April 2026

doi: 10.20944/preprints202604.0292.v1

Keywords: visibility graphs; Monte Carlo simulation; power prices; volatility clustering; locally parametric methods; k-nearest neighbours; sinh-arcsinh distribution



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Topological Conditioning via Natural Visibility Graphs for Monte Carlo Simulation of Power Prices

Carlo Mari ^{1,*}  and Emiliano Mari ²

¹ DEIM – Department of Economics, Engineering, Society and Business Organization, University of Tuscia, Viterbo 01100, Italy

² Sydus, Italy

* Correspondence: carlo.mari@unitus.it

Abstract

A locally parametric framework is proposed for Monte Carlo simulation of electricity prices that jointly reproduces the key stylized facts of power markets: mean-reversion, fat tails, asymmetry, and volatility clustering. Following a two-stage pipeline in which mean-reversion is estimated separately from the innovation distribution, the paper focuses on the second stage: simulating the residual innovations via topological conditioning on Natural Visibility Graphs (NVG) built on the observed innovation sequence. At each simulation step, the local structure of the graph is used to identify historically similar market states and to draw the next innovation from a locally fitted distribution. The key methodological contribution is that this topological conditioning mechanism simultaneously determines the local scale, skewness, and tail weight of the innovation distribution — three properties that parametric models such as GARCH must address through separate equations — without any assumption on regime dynamics or transitions. The framework is locally parametric: the number of model parameters grows with the sample size rather than being fixed in advance, and the specific distributional family used as a local working model can be replaced without altering the conditioning mechanism. Applied to two power markets with contrasting distributional characteristics — the Italian Power Exchange (PUN) and PJM West Hub (US) — the framework achieves simultaneous coverage of three distributional statistics ($\hat{\sigma}$, $\hat{\gamma}$, $\hat{\kappa}$) and the first-order autocorrelation of squared innovations $\hat{\rho}_1(\varepsilon_t^2)$ for both markets, with a single neighbourhood size $k = 10$ and no market-specific re-calibration; more generally, k serves as the natural adaptation parameter for markets with substantially different distributional characteristics.

Keywords: visibility graphs; Monte Carlo simulation; power prices; volatility clustering; locally parametric methods; k -nearest neighbours; sinh-arcsinh distribution

MSC: 37M10

1. Introduction

1.1. Stylized Facts of Electricity Price Dynamics

Electricity prices exhibit a distinctive combination of statistical properties that collectively challenge standard modelling frameworks [1]. Among energy commodities, power markets are distinguished by the physical non-storability of electricity: supply and demand must be balanced instantaneously, so any transient mismatch between generation capacity and consumption produces price spikes that can reach values orders of magnitude above normal levels within minutes, only to revert once equilibrium is restored [2,3]. This mechanism generates time series with mean reversion, heavy-tailed, skewed marginal distributions and pronounced volatility clustering: periods of elevated price variability tend to persist over time, alternating with calmer regimes [4,5].

A distinctive feature of electricity prices is *mean reversion*: prices tend to revert toward a long-run equilibrium level at a speed that varies across markets and time periods. The AR(1) model

$$X_t = (1 - \alpha) X_{t-1} + \varepsilon_t, \quad (1)$$

where X_t denotes the detrended log-price and α governs the speed of mean reversion, provides a natural decomposition of the price dynamics into a persistent component and a residual innovation sequence $\{\varepsilon_t\}$. This decomposition motivates a two-stage modelling approach [6]: the mean-reversion parameter α is estimated in the first stage via a distribution-free neural estimator that requires no assumption on the innovation distribution; in the second stage, the statistical properties of the innovations — fat tails, asymmetry, and volatility clustering — are addressed. The two stages can be treated independently, avoiding the distributional misspecification that arises when both are estimated jointly under a single parametric model.

Three interrelated stylized facts define the statistical landscape of electricity innovations:

1. *Fat tails and asymmetry.* The distribution of log-price increments departs substantially from normality, exhibiting positive excess kurtosis and non-zero skewness driven by price spikes.
2. *Volatility clustering.* Large innovations tend to cluster in time. The autocorrelation function of squared innovations decays slowly from significantly positive values at short lags, confirming that the conditional variance is persistent.
3. *Non-stationarity.* Long-term trends attributable to structural changes, seasonality, and the energy transition require explicit removal before statistical analysis of short-term dynamics.

Any simulation framework aiming to generate realistic electricity price scenarios must jointly reproduce these three features. Failure to do so compromises downstream applications including risk management, derivative pricing, and investment analysis under uncertainty [7–9].

1.2. Limitations of Parametric Approaches

The standard toolkit for modelling financial volatility dynamics consists of parametric models, each of which imposes a specific functional form on the conditional distribution of innovations. While tractable and widely used, these models share structural limitations when applied to electricity markets.

GARCH models [10,11] are the workhorse of financial volatility modelling. In their standard formulation they assume symmetric, zero-skewness innovations, which is a structural limitation for electricity markets where the empirical distribution of innovations is markedly asymmetric. Extensions of the basic GARCH framework can accommodate skewness, heavy tails and other non-Gaussian features, but at the cost of substantially more complex parameterisations that are harder to estimate and interpret, and that still impose a single, globally fixed distributional shape across all market regimes.

Jump-diffusion models [3,12] augment a continuous diffusion with a Poisson jump component, explicitly representing price spikes as discrete events. Although conceptually appealing, these models require estimation of jump intensity and jump size distribution from limited historical data. In markets where spikes are rare but extreme, maximum likelihood estimation of jump parameters is notoriously unstable, and out-of-sample performance degrades rapidly when the historical spike record is short or non-representative of future dynamics.

Regime-switching models [13] assume that the price process alternates among a finite number K of latent states, each characterised by its own conditional distribution. Transition probabilities are estimated from data via the EM algorithm or Bayesian inference. These models can capture structural changes in mean and variance, but they impose several constraints: the number of regimes must be pre-specified; the within-regime distribution remains fully parametric; and the state-space representation forces a discrete partition of a volatility landscape that is inherently continuous.

The common thread across all parametric approaches is deeper than the specific limitations noted above. Every parametric model must solve two problems simultaneously: specifying the marginal distribution of innovations and specifying the mechanism that generates temporal dependence. These two problems are structurally coupled in the parametric framework — the GARCH variance recursion determines both the clustering and, through the chosen innovation distribution, the tail behaviour — so that improving one aspect of the fit (e.g., adding skewness via a skewed- t distribution, or richer dynamics via an EGARCH specification) necessarily re-parameterises the other. Distributional misspecification at any level propagates systematically through Monte Carlo simulations.

These limitations call for an approach that imposes no global distributional assumption on the innovations and conditions the temporal dependence on the data structure itself, rather than on a pre-specified parametric model. For electricity markets — where the marginal distribution is heavy-tailed, asymmetric, and market-specific — such an approach eliminates the model-selection step that any parametric framework requires.

1.3. Visibility Graphs as a Non-Parametric Tool

Visibility graphs [14] offer precisely such a mechanism: a principled way to encode the temporal structure of a time series into a network representation without any distributional assumption. Given a time series $\{\varepsilon_t\}_{t=1}^T$, two observations at times $i < j$ are connected by an edge in the Natural Visibility Graph (NVG) if and only if all intermediate observations ε_k ($i < k < j$) satisfy:

$$\varepsilon_k < \varepsilon_i + (\varepsilon_j - \varepsilon_i) \cdot \frac{k-i}{j-i}. \quad (2)$$

The right-hand side is the linear interpolant between (i, ε_i) and (j, ε_j) evaluated at time k : two points are mutually visible if no intermediate observation rises above the straight line connecting them.

The NVG encodes rich local information about the time series: a high observation connects to many neighbours and spans long temporal distances, while a low observation is shielded by surrounding values and has few, short-range connections. This topological structure is sensitive to the local volatility regime — spikes, troughs, and periods of low variability each produce characteristic connectivity patterns — without requiring any parametric model for that regime.

1.4. Topological Conditioning for Monte Carlo Simulation

Previous work has shown that visibility graph topology carries statistically significant information about the underlying stochastic process [15,16] and has been applied to financial and energy time series to characterise distributional properties and detect structural changes [17–21]. While previous studies have used visibility graph topology as a descriptive or diagnostic tool, the present paper is the first to exploit it as a conditioning mechanism for Monte Carlo simulation: the topological structure of the innovation sequence directly governs the local distributional regime at each simulation step, without any parametric assumption on regime dynamics or transitions. Specifically, given the mean-reversion parameter $\hat{\alpha}$ estimated in the first stage [6] and the resulting innovations

$$\varepsilon_t = X_t - (1 - \hat{\alpha}) X_{t-1}, \quad (3)$$

the NVG built on the innovation sequence $\{\varepsilon_t\}$ is used to encode the local market state into a feature vector that conditions a locally parametric simulation engine.

Three backward-looking features are extracted from the NVG: the *backward degree* d_t^- (number of edges connecting the current node to past nodes), the *maximum backward edge length* ℓ_t^- (temporal span of the longest such edge), and the *squared current innovation* ε_t^2 (a dynamic proxy for the local volatility level). Together they form the conditioning vector $\mathbf{f}_t = (\varepsilon_t^2, d_t^-, \ell_t^-)$ that encodes the local volatility regime through the geometry of the time series itself, with no global assumption on the form of the innovation distribution.

The conditioning mechanism works as follows. At each point in the historical record, the feature vector $\mathbf{f}_t$ characterises the local market state: the squared innovation captures the current volatility level, while the two topological quantities describe the geometric structure of the recent price trajectory. States with similar feature vectors are interpreted as instances of the same volatility regime, regardless of when they occurred. The k historical states most similar to the current one are identified by nearest-neighbour search in the feature space, and the innovations that followed those states are used to fit a local parametric distribution. As the local working model, a Jones–Pewsey Sinh-ArcSinh (SAS) distribution [23] is adopted: a flexible family that simultaneously accommodates heavy tails, positive or negative asymmetry, and a wide range of kurtosis values. Its scale, skewness, and tail weight are estimated locally for each historical state, so that a single conditioning operation jointly determines all three distributional properties.

This mechanism extends naturally to Monte Carlo simulation: starting from the last observed state, the feature vector is updated at each step using the most recent simulated innovations, a new set of nearest neighbours is identified in the historical record, and the next innovation is drawn from the corresponding local SAS distribution. The result is a collection of synthetic price paths that inherit the distributional and temporal structure of the observed series — fat tails, asymmetry, and volatility clustering — without any global assumption on the innovation distribution. The simulated innovations are then fed back into the AR(1) structure to generate synthetic price paths directly applicable to risk quantification and derivative pricing.

The framework is validated on two electricity markets with contrasting distributional characteristics: the Italian Power Exchange (negative skewness, moderate kurtosis) and PJM West Hub (strong positive skewness, heavy tails). With a single neighbourhood size $k = 10$ and no market-specific re-calibration, the framework achieves simultaneous coverage of $\hat{\sigma}$, $\hat{\gamma}$, $\hat{\kappa}$, and $\hat{\rho}_1(\varepsilon_t^2)$ for both markets, assessed via a coverage test on $N = 500$ Monte Carlo simulated paths. The neighbourhood size k is the sole tunable parameter of the framework and serves as a natural adaptation lever for markets with substantially different distributional characteristics.

The remainder of the paper is organised as follows. Section 2 describes the data and preprocessing pipeline. Section 3 details the methodology: NVG construction, feature extraction, and local SAS fit. Section 4 presents the empirical results: coverage analysis ($\hat{\sigma}$, $\hat{\gamma}$, $\hat{\kappa}$, $\hat{\rho}_1(\varepsilon_t^2)$) and ACF reproduction for both markets. Section 5 discusses the single-mechanism nature of the topological conditioning, the locally parametric and distribution-free character of the framework, and its market-specific interpretation. Section 6 concludes and outlines limitations and directions for future research. The overall pipeline is summarised in Figure 1.

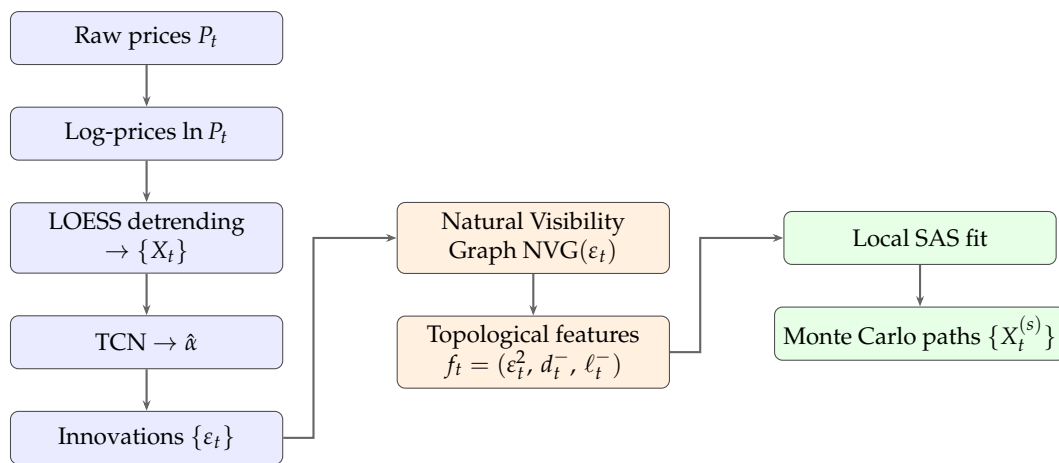


Figure 1. Overall pipeline. Raw prices are preprocessed into AR(1) innovations ε_t ; a Natural Visibility Graph is constructed on the innovation sequence and backward topological features $f_t = (\varepsilon_t^2, d_t^-, \ell_t^-)$ are extracted (squared innovation ε_t^2 , backward degree d_t^- , and maximum backward edge length ℓ_t^-). The feature vector conditions the local SAS fit, which implements the topological conditioning mechanism and generates the simulated paths.

2. Data and Preprocessing

2.1. Markets and Data

The empirical analysis is conducted on two electricity spot markets with contrasting geographic and structural characteristics.

PUN (Prezzo Unico Nazionale) is the reference price of the Italian day-ahead power exchange, operated by GME (Gestore dei Mercati Energetici SpA). Daily price observations are used, yielding $T = 1825$ AR(1) innovations after the first lag is consumed. PUN prices reflect Italy's dependence on imported gas and hydropower, and exhibit recurrent negative spikes associated with periods of excess renewable generation.

PJM West Hub is the day-ahead locational marginal price at the PJM Western Hub, the largest organised electricity market in the United States. Business day price observations are used (weekends and US market holidays excluded), yielding $T = 1224$ AR(1) innovations after the first lag is consumed. PJM prices are driven by temperature-sensitive demand, thermal generation costs, and transmission constraints, and exhibit recurrent upward spikes associated with cold or heat waves.

Both datasets cover the same five-year window, 1 January 2019 – 31 December 2023, providing a consistent temporal basis for comparison. This period encompasses major common shocks: the COVID-19 demand disruption (2020), the post-pandemic recovery, and the extreme price volatility triggered by the 2022 energy crisis following the Russia–Ukraine conflict. PUN is characterised by moderate kurtosis and negative skewness, while PJM exhibits higher kurtosis and pronounced positive asymmetry (Table 1).

2.2. Preprocessing Pipeline

Raw price series P_t are transformed into innovations ε_t through three sequential steps, illustrated in Figure 1.

Step 1 — Logarithmic transformation. Log-prices $\tilde{P}_t = \ln P_t$ are used, which stabilise variance and convert multiplicative price dynamics into additive form.

Step 2 — LOESS detrending. A nonparametric LOESS smoother [24,25] with bandwidth $h = 0.10$, selected by leave-one-out cross-validation over the grid $\{0.05, 0.10, 0.20\}$, is fitted to $\tilde{P}_t$, yielding the smooth trend $\hat{m}_t$. The detrended series is $X_t = \tilde{P}_t - \hat{m}_t$. LOESS is preferred over parametric detrending because it imposes no functional form on the long-term trend and is robust to outliers.

Step 3 — Innovation extraction. The price process X_t is modelled as an AR(1) mean-reverting process:

$$X_t = (1 - \alpha) X_{t-1} + \varepsilon_t, \quad \text{Cov}(\varepsilon_t, \varepsilon_s) = 0 \quad (t \neq s), \quad (4)$$

where α is the mean-reversion speed; the condition $|1 - \alpha| < 1$ ensures stationarity of $\{X_t\}$. No further assumption is imposed on ε_t : in particular, innovations may be heteroscedastic (as in GARCH-type processes), and the Yule–Walker identification of α remains valid under zero serial correlation alone. Given an estimate $\hat{\alpha}$, innovations are recovered as $\varepsilon_t = X_t - (1 - \hat{\alpha}) X_{t-1}$.

The mean-reversion parameter α is estimated by a Temporal Convolutional Network (TCN) [22] trained on simulated AR(1) paths with diverse innovation distributions [6]. Unlike classical estimators (OLS, Yule–Walker, GARCH-based), which condition on a specific innovation model and therefore yield different estimates of α on the same sample depending on the assumed distribution, the TCN learns to read the autocorrelation structure of the series directly, without any distributional conditioning. The values $\hat{\alpha} = 0.2857$ (PUN) and $\hat{\alpha} = 0.3540$ (PJM) are taken directly from [6].

2.3. Descriptive Statistics

Table 1 reports the main descriptive statistics of the innovation sequences. Both markets exhibit substantial departures from normality: Pearson kurtosis of 5.48 (PUN) and 9.90 (PJM), well above the Gaussian benchmark of 3, and significant asymmetry. The autocorrelation of squared innovations at lag 1, $\rho_1(\varepsilon_t^2)$, is positive and statistically significant in both cases, confirming the presence of volatility

clustering. Figure 2 shows the innovation time series and empirical densities alongside a Gaussian reference.

Table 1. Descriptive statistics of AR(1) innovations ε_t . T = number of innovations (one less than the number of price observations). Kurtosis is Pearson (Gaussian = 3). $\rho_1(\varepsilon_t^2)$ is the first-order autocorrelation of squared innovations.

Market	T	$\hat{\alpha}$	Std	Skewness	Kurtosis	$\rho_1(\varepsilon_t^2)$
PUN	1825	0.2857	0.1365	-0.351	5.48	0.168
PJM	1224	0.3540	0.1799	+1.178	9.90	0.210

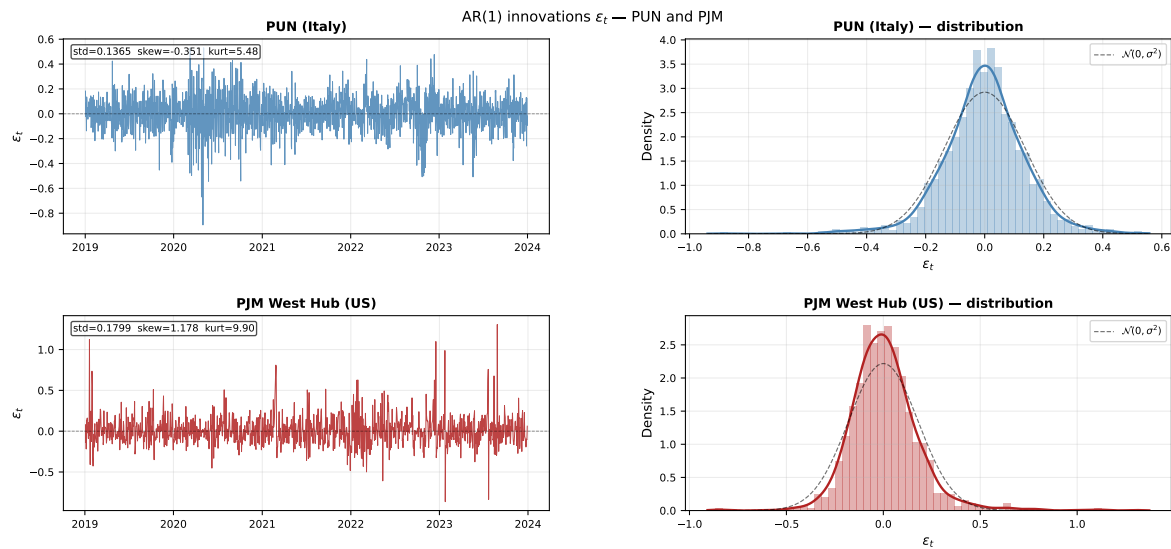


Figure 2. AR(1) innovations ε_t for PUN (top) and PJM (bottom). Left panels: time series with annotated moments. Right panels: empirical density (histogram and KDE) versus Gaussian reference (dashed). Both series exhibit heavy tails and departure from symmetry.

3. Methodology

3.1. Natural Visibility Graph

Given the innovation sequence $\{\varepsilon_t\}_{t=1}^T$, the Natural Visibility Graph (NVG) is the undirected graph $G = (V, E)$ where $V = \{1, \dots, T\}$ is the set of nodes, each representing one observation in the time series, and $E \subseteq V \times V$ is the set of edges defined by the visibility condition: $(i, j) \in E$ if and only if $i \neq j$ and condition (2) holds for the pair (i, j) with $i < j$. Each edge $(i, j) \in E$ therefore represents a direct line-of-sight between observations ε_i and ε_j , with no intermediate observation rising above the straight line connecting them. The NVG is constructed once on the full historical sequence using an $O(T^2)$ algorithm [14,26].

The topological properties of the NVG encode local dynamical information about the time series. A node i with a large value ε_i dominates its neighbourhood: it is visible from many distant nodes, yielding high degree and long edges. Conversely, a node in a low-volatility trough is partially occluded by surrounding observations and has short, local connections. This sensitivity to the local volatility regime is the key property exploited in the simulation framework.

Figure 3 illustrates the geometric structure of the NVG on a representative window of 80 innovations for each market. The left panels display the arc representation: each arc connects a pair of observations satisfying condition (2) (arcs are drawn as curves for visual clarity; the underlying criterion is based on straight lines, as defined in (2)), with the arc colour encoding the temporal distance between the two nodes. High-valued observations (spikes) spawn long-range arcs spanning the entire window, while low-valued observations generate only short, local connections. The right panels show the same graph in a Kamada–Kawai network layout: node size is proportional to degree and node

colour encodes the innovation value, revealing how spike nodes (large, red) occupy structurally central positions in the network while trough nodes (small, blue) cluster at the periphery.

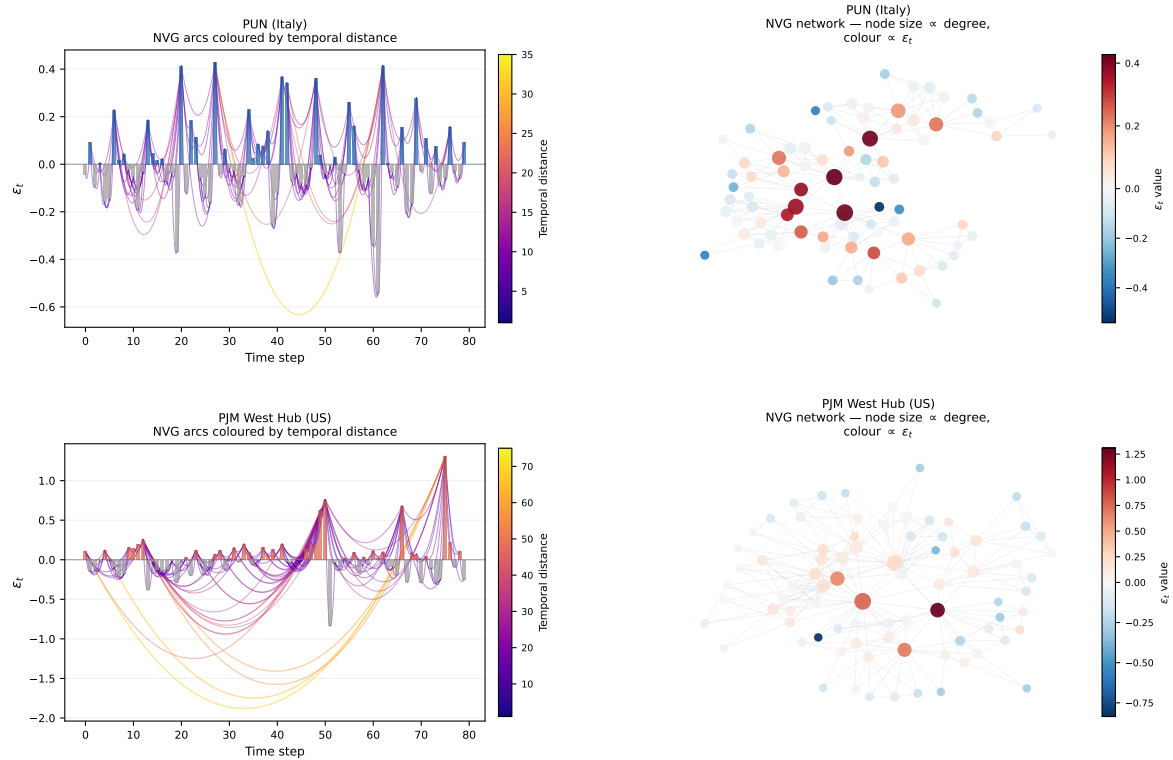


Figure 3. NVG geometric structure on an 80-innovation window for PUN (top) and PJM (bottom). *Left:* arc representation of the NVG edges; arc colour encodes temporal distance between the two connected nodes (dark = short, bright = long). *Right:* Kamada-Kawai network layout; node size is proportional to NVG degree, node colour encodes the innovation value (red = large positive spike, blue = large negative value). Both representations highlight that high-valued observations act as topological hubs with long-range connections, while low-valued observations have short, local links.

3.2. Backward Topological Features

At any simulation step t , the future is not yet observed: only $\varepsilon_1, \dots, \varepsilon_t$ are available. Computing the NVG on the full history would require $O(T^2)$ operations and would yield features whose meaning changes as T grows. The graph is therefore restricted to a *rolling window* of fixed length $W = 100$ observations ending at t :

$$\mathcal{W}_t = \{\varepsilon_{t-W+1}, \dots, \varepsilon_t\}, \quad (5)$$

and build a local NVG $G_t = (V_t, E_t)$ on $\mathcal{W}_t$ at each step (for $t < W$, the window is zero-padded on the left to maintain fixed length). This choice serves two purposes: (i) it enforces *causality*, since $\mathcal{W}_t$ uses only observations available at time t ; (ii) it ensures *temporal locality*, so that the topological features reflect the current market regime rather than a long-run average over the entire series.

Attention is restricted to *backward* edges of G_t — edges connecting the last node of the window to all earlier nodes it can see — and the following causal feature vector is defined:

$$\mathbf{f}_t = (\varepsilon_t^2, d_t^-, l_t^-) \in \mathbb{R}^3, \quad (6)$$

where

$$d_t^- = \sum_{\substack{j \in \mathcal{W}_t \\ j < t}} \mathbf{1}[(j, t) \in E_t] \quad (7)$$

is the *backward degree* — the number of past observations in $\mathcal{W}_t$ that are visible from t — and

$$\ell_t^- = \max\{t - j : j \in \mathcal{W}_t, j < t, (j, t) \in E_t\} \quad (8)$$

is the *maximum backward edge length* — the temporal distance to the farthest visible predecessor within the window.

The three components carry complementary information. It is worth noting that ε_t^2 is not a topological property of the NVG: it is the observed value of the current innovation, included in the feature vector to anchor the conditioning to the absolute volatility level. The two genuinely topological components d_t^- and ℓ_t^- capture the *relative* structure of the local window, which may not fully identify the volatility level on its own. ε_t^2 is preferred over the signed innovation ε_t : conditioning on the signed value would make the drawn distribution depend on the direction of the innovation, creating a systematic link between successive innovations and inducing $\text{Cov}(\varepsilon_{t+1}, \varepsilon_t) \neq 0$, in violation of the AR(1) assumption. The squared value captures the volatility level symmetrically, avoiding this directional bias. d_t^- measures how many past observations are visible from t : a high value indicates that ε_t lies above its recent neighbours (low-volatility regime, unobstructed line of sight to many predecessors), while a low value signals that ε_t is buried in a volatile cluster where nearby spikes block the view. ℓ_t^- captures the temporal horizon of visibility: it is large for isolated spikes (visible far back in time) and small for values embedded in a volatile cluster (visibility blocked by recent high-amplitude neighbours).

For computational efficiency, the backward features are computed via a vectorised algorithm that precomputes the NVG index arrays for a window of length W once, requiring $O(W^2)$ memory and reducing the per-step cost to $O(W)$ floating-point operations. Figure 4 illustrates the NVG structure and backward features on a representative window of PJM innovations: the current node (red bar) is connected to all past nodes visible from it, with backward edges highlighted in red. The annotation shows the resulting values of d_t^- and ℓ_t^- for that node.

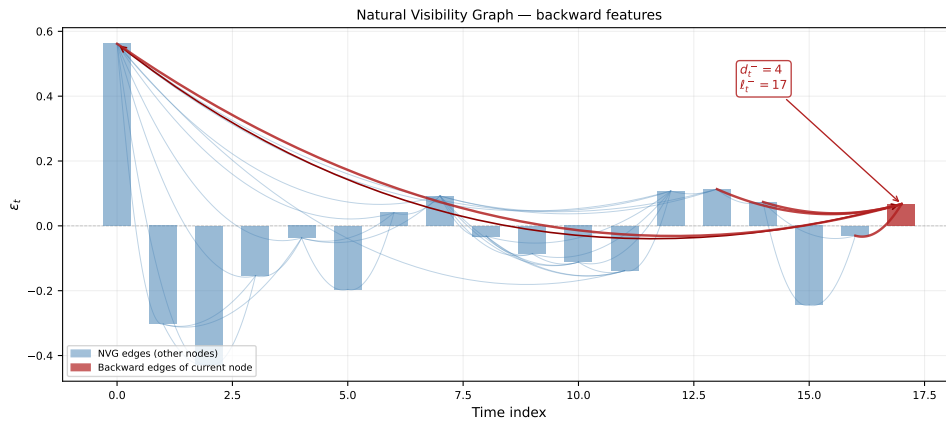


Figure 4. Natural Visibility Graph on a rolling window of $W = 18$ PJM innovations. The current node (red bar) is connected to all past nodes it can “see”; backward edges are highlighted in red, all other NVG edges in blue. The backward degree d_t^- and maximum backward edge length ℓ_t^- are annotated. Edges are drawn as curves for visual clarity; the actual visibility criterion is based on straight lines between observations, as defined in condition (2).

Before nearest-neighbour search, features are standardised:

$$\tilde{\mathbf{f}}_t = \frac{\mathbf{f}_t - \hat{\boldsymbol{\mu}}_f}{\hat{\sigma}_f}, \quad (9)$$

where $\hat{\mu}_f = (\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3)$ and $\hat{\sigma}_f = (\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3)$ are computed component-wise over all $T - 1$ historical nodes:

$$\hat{\mu}_j = \frac{1}{T-1} \sum_{t=1}^{T-1} f_{t,j}, \quad \hat{\sigma}_j = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T-1} (f_{t,j} - \hat{\mu}_j)^2}. \quad (10)$$

The division in (9) is performed component by component, so that each feature is mapped to approximate zero mean and unit variance before the Euclidean distance is computed. This step is necessary because the three components operate on very different numerical scales: ε_t^2 is typically small (of the order of the innovation variance), while d_t^- counts visible predecessors and ℓ_t^- measures temporal distances, both of which can be orders of magnitude larger. Without standardisation, the Euclidean distance in feature space would be dominated by ℓ_t^- , making the nearest-neighbour search effectively blind to the volatility signal carried by ε_t^2 .

3.3. Local SAS Fit

Let $k \geq 1$ be a fixed integer (the neighbourhood size) and, for each historical node $i \in \{1, \dots, T - 1\}$, let $\mathcal{N}_k(i)$ denote the set of the k indices $j \in \{1, \dots, T - 1\} \setminus \{i\}$ with smallest Euclidean distance $\|\tilde{\mathbf{f}}_i - \tilde{\mathbf{f}}_j\|$ to node i in the standardised feature space (9). Note that the feature vectors $\tilde{\mathbf{f}}_i$ are each computed from a local rolling window of length W ending at i (Section 3.2), so they capture the local dynamical state at each point in time. The nearest-neighbour search, however, ranges over the *entire* historical record $\{1, \dots, T - 1\}$: it identifies which past episodes — wherever they occurred in the series — most closely resembled the current local state. The rolling window defines the geometry of each node; the k -NN search compares these geometries globally across all available history.

The *successor* of any node j is the immediately following innovation ε_{j+1} , i.e. the observation recorded one step after j in the historical sequence. Rather than drawing directly from the observed successors of the k nearest neighbours — which would confine the simulation to the historical range — a Sinh-ArcSinh (SAS) distribution [23] is fitted by maximum likelihood on those successors, allowing the simulation to generate innovations beyond the observed sample.

The SAS Family

Given the successors $\{\varepsilon_{j+1} : j \in \mathcal{N}_k(i)\}$ of the k nearest neighbours of node i , a SAS distribution is fitted by maximum likelihood. The family is parametrised as:

$$X \stackrel{d}{=} \lambda \sinh\left(\frac{\operatorname{arcsinh}(Z) + \eta}{\delta}\right), \quad Z \sim \mathcal{N}(0, 1), \quad (11)$$

where $\lambda > 0$ is the scale parameter, $\eta \in \mathbb{R}$ controls skewness (positive η induces right-skewed distributions), and $\delta > 0$ controls tail weight (smaller δ yields heavier tails). The SAS family is flexible enough to capture the full range of skewness and kurtosis observed in power markets and encompasses symmetric unimodal distributions as a special case ($\eta = 0$).

Before any simulation path is generated (*offline phase*), parameters $(\hat{\lambda}_i, \hat{\eta}_i, \hat{\delta}_i)$ are estimated by maximum likelihood (Nelder-Mead) for every historical node $i \in \{1, \dots, T - 1\}$: the SAS is fitted on the set of successors $\{\varepsilon_{j+1} : j \in \mathcal{N}_k(i)\}$. Since all nodes and their successors belong to the historical record, no future observations are required at this stage.

The Unifying Principle

The critical property of this construction is that all three parameters — scale λ_i , skewness η_i , and tail weight δ_i — are functions of the topological neighbourhood of node i and therefore vary continuously with the local dynamical state. In high-volatility regimes (large ε_t^2 , high backward degree d_t^- , long maximum edge ℓ_t^-), the k -NN identifies historically similar spike-time nodes: their successors have large variance, pronounced skewness, and heavy tails, yielding a fitted SAS with large $\hat{\lambda}_i$, non-zero $\hat{\eta}_i$, and small $\hat{\delta}_i$. In low-volatility regimes, the SAS contracts toward a near-Gaussian shape. *Scale*,

skewness, and tail weight are thus jointly determined by a single conditioning operation — the topological k -NN search — without any additional parameters or regime-switching equations.

Locally Parametric and Distribution-Free

The framework is locally parametric in the sense of [27]: within each topological neighbourhood a parametric family (SAS) is fitted as a working approximation to the local distribution, while globally no parametric form is imposed on the innovation process and the total parameter count grows with the sample size. This is the defining property that distinguishes locally parametric methods from both fully parametric models (fixed parameter count, global distributional assumption) and classical non-parametric methods (no local approximating family). The SAS family serves as a flexible working model; any universal approximator (Gaussian mixture, normal-inverse-Gaussian, generalised hyperbolic) could be substituted without altering the conditioning mechanism.

3.4. Monte Carlo Simulation

In the *online phase*, at each simulation step t , the current feature vector $\tilde{\mathbf{f}}_t$ acts as a real-time regime indicator: it jointly encodes the current volatility level (through ε_t^2) and the local topological structure (through d_t^- and ℓ_t^-), and is matched against all precomputed historical feature vectors to identify j^* , the historical node whose feature fingerprint most closely resembles the current one. Proximity in this hybrid feature space is interpreted as similarity of market regime. The next innovation is then drawn from the SAS already fitted for j^* , which encodes how the process behaved *after* historically similar regimes. The simulation never requires a future successor of t : the entire conditional distribution is inherited from the closest historical analogue. The per-step cost reduces to a single nearest-neighbour lookup and one scalar draw.

Formally, the procedure at each step t is as follows:

1. Compute $\tilde{\mathbf{f}}_t = (\tilde{\varepsilon}_t^2, \tilde{d}_t^-, \tilde{\ell}_t^-)$ from the current rolling window.
2. Find the closest historical analogue

$$j^* = \arg \min_{i \in \{1, \dots, T-1\}} \|\tilde{\mathbf{f}}_t - \tilde{\mathbf{f}}_i\|.$$

3. Draw $\varepsilon_{t+1} \sim \text{SAS}(\hat{\lambda}_{j^*}, \hat{\eta}_{j^*}, \hat{\delta}_{j^*})$ using transformation (11).
4. Update the detrended log-price: $X_{t+1} = (1 - \hat{\alpha}) X_t + \varepsilon_{t+1}$.

Iterating steps 1–4 over $t = T, T+1, \dots, T+H$ produces a Monte Carlo trajectory $\{X_t^{(s)}\}$ of detrended log-prices of length H , where each X_{t+1} is generated from ε_{t+1} via the AR(1) recursion in step 4. Repeating the procedure for $s = 1, \dots, N$ independent seeds yields an ensemble of N synthetic price paths, each initialised at the last observed state X_T . The loop implements a continuous regime-matching mechanism: at every step, the current topological fingerprint selects the historical market state that most resembles the present one, and the simulation inherits the distributional properties that followed that state in the past. Regime transitions are not governed by a parametric equation but emerge naturally from the geometry of the feature space.

Internal consistency. The model assumption $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0$ for $t \neq s$, required for Yule–Walker identification of α , is preserved by the simulation. The SAS parameters $(\hat{\lambda}_{j^*}, \hat{\eta}_{j^*}, \hat{\delta}_{j^*})$ are estimated in the offline phase on historical innovations that satisfy this assumption. In the online phase, each draw is made from a *fixed* distribution indexed by j^* : the topological conditioning determines *which* distribution to use, but does not introduce a linear dependence between ε_{t+1} and ε_t . Serial dependence in the simulated sequence is therefore confined to the second moment (volatility clustering through ε_t^2), which is precisely the structure targeted by the conditioning mechanism. The use of ε_t^2 rather than the signed innovation ε_t in the feature vector is essential to this argument: conditioning on the signed value would make the drawn distribution depend on the direction of the innovation, creating a systematic link between successive innovations and inducing $\text{Cov}(\varepsilon_{t+1}, \varepsilon_t) \neq 0$, in violation of $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0$.

4. Results

4.1. Coverage Analysis

Throughout this section, the following statistics are used to characterise the innovation distribution: $\hat{\sigma}$ (standard deviation), $\hat{\gamma}$ (skewness), $\hat{\kappa}$ (Pearson kurtosis, Gaussian = 3), and $\hat{\rho}_1(\varepsilon_t^2)$ (first-order autocorrelation of squared innovations). The last quantity captures volatility clustering and is not a distributional moment, but an autocorrelation measure of the squared process.

Table 2 reports the main coverage results for both markets. For each statistic, the empirically observed value is compared to the median and [5th,95th] percentile band across $N = 500$ simulated paths of length T . All three statistics and $\hat{\rho}_1(\varepsilon_t^2)$ are covered simultaneously for both markets with a single choice of $k = 10$ and no market-specific re-calibration.

Table 2. Coverage analysis ($N = 500$ paths, $k = 10$, $\mathbf{f}_t = (\varepsilon_t^2, d_t^-, \ell_t^-)$): observed values versus simulated $[p_5, p_{95}]$ band. $\hat{\sigma}$ = standard deviation, $\hat{\gamma}$ = skewness, $\hat{\kappa}$ = Pearson kurtosis (Gaussian = 3), $\hat{\rho}_1$ = first-order ACF of ε_t^2 . All observed values fall within the band.

Market		Observed	Median	p_5	p_{95}
PUN	$\hat{\sigma}$	0.1365	0.1342	0.127	0.142
	$\hat{\gamma}$	−0.351	−0.322	−1.037	0.071
	$\hat{\kappa}$	5.48	6.29	4.70	14.18
	$\hat{\rho}_1(\varepsilon_t^2)$	0.168	0.138	0.052	0.237
PJM	$\hat{\sigma}$	0.1799	0.1734	0.159	0.195
	$\hat{\gamma}$	+1.178	+0.984	0.076	2.945
	$\hat{\kappa}$	9.90	11.33	5.89	40.47
	$\hat{\rho}_1(\varepsilon_t^2)$	0.210	0.139	0.029	0.313

The coverage results warrant several observations. First, the observed values generally lie near the lower portion of the $[p_5, p_{95}]$ band for skewness and the median for kurtosis, indicating that the simulated distribution of these moments is wide. The width reflects genuine sampling variability: for a series of length $T \leq 1825$, a single observed path is one realisation of a stochastic process whose moment estimates carry substantial uncertainty. The topological SAS framework correctly reproduces this uncertainty rather than artificially concentrating simulated moments around a point estimate.

Second, the coverage of $\hat{\rho}_1(\varepsilon_t^2)$ — the first-order autocorrelation of squared innovations, the direct measure of volatility clustering — is achieved without any explicit model for the conditional variance. The clustering emerges endogenously from the topological conditioning: in high-volatility states (large ε_t^2 , high d_t^-), the k -NN selects historically similar spike-time nodes whose successors remain in the elevated-volatility tail, perpetuating the cluster.

4.2. Sensitivity to Neighbourhood Size k

Table 3 reports, for each $k \in \{10, 20, 30, 50\}$, the median and 90% prediction interval $[p_5, p_{95}]$ of $\hat{\sigma}$, $\hat{\gamma}$, $\hat{\kappa}$, and $\hat{\rho}_1(\varepsilon_t^2)$ across $N = 500$ simulated paths. All other settings are unchanged.

For PUN, all three statistics and $\hat{\rho}_1(\varepsilon_t^2)$ are covered across the entire range $k \in \{10, 20, 30, 50\}$: the simulation band consistently contains the observed values regardless of neighbourhood size. The median estimates remain close to the observed values, while the interval width narrows modestly as k increases, reflecting the smoothing effect of larger neighbourhoods.

For PJM, coverage holds for $k \leq 30$ but deteriorates at $k = 50$: the observed standard deviation (0.180), skewness (1.178), and kurtosis (9.90) fall above the respective 95th percentiles (0.175, 0.910, 9.26), while $\hat{\rho}_1(\varepsilon_t^2)$ remains covered. The failure is systematic: at $k = 50$ the median skewness drops to 0.419 and the median kurtosis to 5.75, far below the observed values, indicating that the large neighbourhood averages over regimes with substantially different skewness and scale. This compresses the local SAS parameters toward a global mean and erases the heavy-tailed, right-skewed character of the PJM

distribution. The first-order ACF of squared innovations is more robust to this averaging because volatility clustering is a coarser property than the precise shape of the marginal distribution.

Table 3. Sensitivity to neighbourhood size $k \in \{10, 20, 30, 50\}$. Each cell reports the median (top line) and 90% prediction interval $[p_5, p_{95}]$ (bottom line, smaller font) of $N = 500$ simulated paths. The *Observed* row gives the historical values. [†] marks cells where the observed value falls outside $[p_5, p_{95}]$. Notation: $\hat{\sigma}$ = standard deviation, $\hat{\gamma}$ = skewness, $\hat{\kappa}$ = Pearson kurtosis, $\hat{\rho}_1 =$ first-order ACF of ε_t^2 .

	$\hat{\sigma}$	$\hat{\gamma}$	$\hat{\kappa}$	$\hat{\rho}_1(\varepsilon^2)$
<i>Panel A — PUN (Italy)</i>				
Observed	0.137	−0.351	5.48	0.168
$k = 10$	0.134 [0.127, 0.142]	−0.322 [−1.037, 0.071]	6.29 [4.70, 14.18]	0.138 [0.052, 0.237]
$k = 20$	0.136 [0.130, 0.143]	−0.191 [−0.576, 0.080]	5.20 [4.36, 7.79]	0.157 [0.077, 0.226]
$k = 30$	0.137 [0.130, 0.144]	−0.173 [−0.416, 0.072]	4.95 [4.17, 6.85]	0.157 [0.084, 0.229]
$k = 50$	0.135 [0.129, 0.142]	−0.168 [−0.410, 0.025]	4.78 [4.10, 6.29]	0.154 [0.091, 0.219]
<i>Panel B — PJM (West Hub, US)</i>				
Observed	0.180	1.178	9.90	0.210
$k = 10$	0.173 [0.159, 0.195]	0.984 [0.076, 2.945]	11.33 [5.89, 40.47]	0.139 [0.029, 0.313]
$k = 20$	0.169 [0.157, 0.185]	0.672 [0.042, 1.619]	7.99 [5.10, 17.24]	0.124 [0.034, 0.296]
$k = 30$	0.168 [0.158, 0.182]	0.547 [0.005, 1.366]	6.87 [4.80, 14.63]	0.118 [0.037, 0.262]
$k = 50$	0.164 [†] [0.154, 0.175]	0.419 [†] [0.044, 0.910]	5.75 [†] [4.42, 9.26]	0.115 [0.045, 0.222]

These results confirm that $k = 10$ is a conservative and well-motivated choice, and that the framework remains valid across $k \leq 30$ for both markets. More generally, k is the natural adaptation parameter of the framework: markets with more heterogeneous dynamics may benefit from a smaller k to preserve local specificity, while larger k provides stability at the cost of smoothing over distinct regimes.

4.3. Volatility Clustering: Full ACF Profile

Figure 5 displays the median $\text{ACF}(\varepsilon_t^2)$ profile over lags 1–30, together with the [5th, 95th] percentile simulation band and the empirical profile.

For *PUN* ($\text{MAE-ACF} = 0.080$), the empirical ACF oscillates between positive and negative values, consistent with a weak clustering signal ($\hat{\rho}_1 = 0.168$). The observed profile exits the $[p_5, p_{95}]$ band at a few lags; with 30 lags tested simultaneously at the 90% level, approximately three exceedances are expected even under a correctly specified model. The higher MAE-ACF relative to *PJM* reflects the oscillatory character of the observed profile: the median of 500 simulated paths is inherently smooth and cannot track the individual oscillations of a single realisation with a weak dependence signal. The coverage of $\hat{\rho}_1(\varepsilon_t^2)$ in Table 2 confirms that the overall clustering level is correctly reproduced.

For *PJM* ($\text{MAE-ACF} = 0.031$), the empirical ACF exhibits a clear positive structure at short lags ($\hat{\rho}_1 = 0.210$), consistent with pronounced spike clustering. The simulation band covers the empirical profile across all 30 lags, and the median simulated ACF tracks the observed decay profile closely.

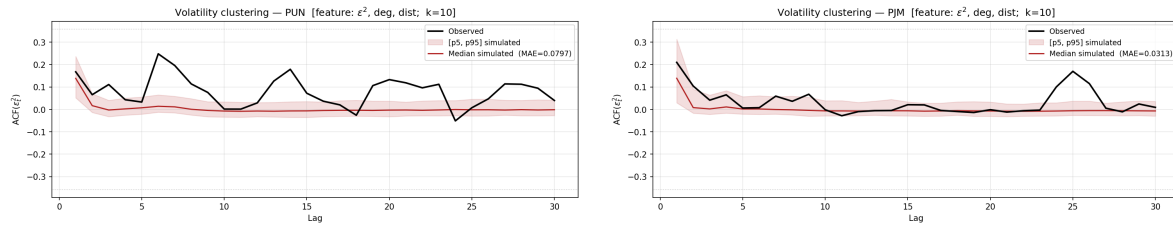


Figure 5. $ACF(\varepsilon_t^2)$ for PUN (left) and PJM (right). Black solid line: empirical profile. Dark red line: median over $N = 500$ simulated paths, $k = 10$. Shaded band: [5th, 95th] percentile simulation interval.

4.4. Simulated Price Paths

Figure 6 provides a qualitative assessment of the simulation framework by displaying $N = 30$ simulated detrended log-price paths together with the observed series. The single path whose maximum and minimum most closely match the observed extremes is highlighted in blue; all other paths are shown as thin grey lines.

The fan of grey paths spans a plausible range of scenarios anchored to the empirical volatility structure, without any path collapsing to zero or diverging to unrealistic levels. The highlighted path (blue) reproduces the overall amplitude of the observed series, confirming that the local SAS extrapolation generates tail realisations consistent with the observed extremes. The clustering of large deviations visible in the observed series (particularly the 2022 energy crisis spike in PUN and the cold-wave spikes in PJM) reappears across the simulated fan.

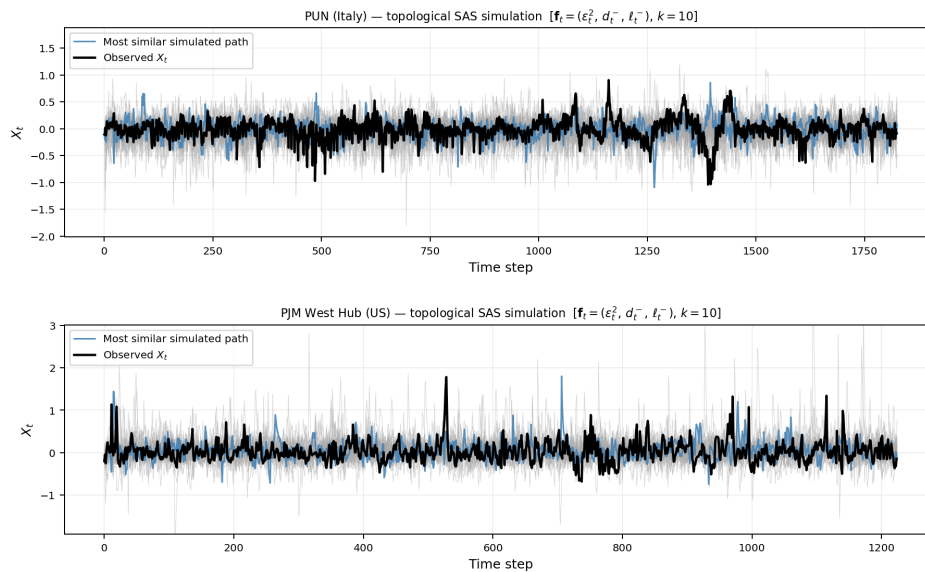


Figure 6. Simulated detrended log-price paths from the topological SAS framework ($N = 30$ paths, $k = 10$) for PUN (top) and PJM (bottom). Black: observed series X_t . Blue: the simulated path whose maximum and minimum most closely match the observed extremes. Grey: all other simulated paths (very thin).

5. Discussion

5.1. A Single Mechanism for All Distributional Properties

The central conceptual contribution of this paper is the identification of a single conditioning mechanism — topological k -nearest neighbours in the NVG backward feature space — that jointly determines the scale, skewness, and tail weight of the local innovation distribution. This is not merely an empirical observation but a structural consequence of the framework design.

All three SAS parameters $(\hat{\lambda}_i, \hat{\eta}_i, \hat{\delta}_i)$ are estimated jointly on the k neighbours of node i : they share the same conditioning variable, the same neighbourhood, and the same estimation procedure. No separate equations govern their dynamics. The mechanism is *parsimonious at the global level* — one

conditioning variable, one neighbourhood size k — while being *flexible at the local level*: each node carries its own distributional triplet, adapted to the local volatility regime.

5.2. Locally Parametric and Distribution-Free

A natural question is whether the introduction of the SAS family compromises the locally parametric character of the framework. The answer depends on the level at which parametric assumptions are evaluated.

At the *global level*, no parametric assumptions are imposed: the total number of SAS parameters is $3n$ (one triplet per node), so it grows with the sample size — a defining property of non-parametric methods. No fixed distributional form is assumed for the innovation distribution across all observations.

At the *local level*, within each topological neighbourhood, a three-parameter SAS distribution is fitted to k successor values. This is a form of local maximum-likelihood estimation [27]: the parametric model is a working approximation to the local distribution, not a global structural assumption.

The SAS was selected for its parsimony and flexibility: three parameters suffice to represent a wide range of asymmetric, heavy-tailed distributions encountered in power markets. However, the conditioning mechanism is independent of this choice: any sufficiently rich family (Gaussian mixture, normal-inverse-Gaussian, generalised hyperbolic) would serve the same role, producing a different working approximation of the same locally varying distribution.

5.3. Structural Comparison with GARCH-Type Models

In GARCH-type models, the conditional variance is driven by a parametric recursion on past squared innovations and past variances, so scale adapts over time. Skewness and tail weight, however, remain global constants: a symmetric innovation distribution (e.g. Student- t) forces zero skewness regardless of the market state, and the tail parameter is fixed across all regimes. Any attempt to improve one component — adding skewness via a skewed- t innovation distribution [28], or richer scale dynamics via an EGARCH specification — necessarily re-parameterises the others. Further extensions with time-varying higher moments [29] add dedicated parametric equations for skewness and kurtosis dynamics, substantially increasing model complexity without providing a unified interpretable mechanism.

In the topological framework, scale $\hat{\lambda}_i$, skewness $\hat{\eta}_i$, and tail weight $\hat{\delta}_i$ are all functions of the local topological state, estimated jointly on the k neighbours of node i . No global parameters govern the distributional shape; all three properties vary endogenously with the market regime through the geometry of the NVG, without any additional equations or parameters in the global model.

5.4. Market-Specific Interpretation

The two markets present complementary profiles that illuminate the behaviour of the proposed approach.

PUN (Italian power exchange) has moderate kurtosis ($\hat{\kappa} = 5.48$) and negative skewness ($\hat{\gamma} = -0.351$), driven by the 2022 energy crisis which produced large but relatively symmetric price movements. The volatility clustering signal is moderate ($\hat{\rho}_1(\varepsilon_t^2) = 0.168$). The topological conditioning captures the alternation between high- and low-volatility regimes; coverage of all three statistics and $\hat{\rho}_1(\varepsilon_t^2)$ is achieved with a wide $[p_5, p_{95}]$ band for skewness and kurtosis, reflecting the genuine sampling variability of these statistics at $T = 1825$.

PJM West Hub exhibits stronger clustering ($\hat{\rho}_1(\varepsilon_t^2) = 0.210$), higher kurtosis ($\hat{\kappa} = 9.90$), and pronounced positive skewness ($\hat{\gamma} = +1.178$) from cold- and heat-wave spikes. The larger empirical skewness requires the local SAS fit to assign systematically positive $\hat{\eta}_i$ in high-volatility regimes. The fact that coverage is achieved for both skewness and kurtosis simultaneously confirms that the topological neighbourhood correctly identifies the spike-clustering regimes in which both properties are most extreme.

6. Conclusions

This paper has introduced a framework for Monte Carlo simulation of power price innovations whose core principle is topological conditioning: at each simulation step, the local distribution of the next innovation is determined by the k -nearest neighbours of the current node in the Natural Visibility Graph feature space, rather than by a parametric recursion for the conditional variance or a global distributional assumption.

The main finding is that a single conditioning mechanism — k -nearest neighbours in the space $(\varepsilon_t^2, d_t^-, \ell_t^-)$ — simultaneously determines the scale, skewness, and tail weight of the local innovation distribution, through the Jones–Pewsey Sinh-ArcSinh family fitted on the topological neighbourhood. These three distributional properties are not modelled separately: they emerge jointly from one geometric operation, with no additional parameters or regime-switching equations. The framework is locally parametric — the total parameter count grows with the sample size — and the SAS family serves as a working model, not a structural assumption; any universal approximator could substitute it without altering the conditioning mechanism.

The empirical results confirm this principle. Applied to two power markets with contrasting distributional characteristics (PUN: negative skewness, moderate kurtosis; PJM: strong positive skewness, heavy tails), the framework achieves simultaneous coverage of $\hat{\sigma}$, $\hat{\gamma}$, $\hat{\kappa}$, and $\hat{\rho}_1(\varepsilon_t^2)$ — for both markets with a single neighbourhood size $k = 10$ and no market-specific re-calibration. The observed values fall within the [5th, 95th] percentile band of $N = 500$ simulated paths of length T for all eight moment-market combinations.

The framework contrasts structurally with GARCH-type models, not merely in accuracy. In GARCH-skewed- t , scale varies through a parametric recursion while skewness and tail weight remain global constants. Extending GARCH to time-varying higher moments requires dedicated equations for each property, increasing parametric complexity without providing a unified mechanism. The topological approach subsumes all three into a single operation, with complexity emerging from the geometry of the NVG rather than from parametric specification.

The framework is portable by design: the same algorithm, with the same hyperparameters, was applied without modification to markets with opposite skewness signatures and different tail weights. The neighbourhood size k is the sole tunable parameter and serves as a natural adaptation lever for markets with substantially different distributional characteristics. The synthetic paths produced by the framework are directly applicable to operational tasks such as derivative pricing, Value-at-Risk estimation, and scenario generation for investment analysis under uncertainty.

Several limitations of the present study point to directions for future research. The analysis is restricted to two electricity markets; gas and oil markets, where kurtosis is dominated by rare extreme events [6], represent a natural next step. For such markets, a smaller k would better preserve the local specificity of spike regimes, but the SAS fit on very few points becomes unreliable: hybrid approaches combining topological conditioning with explicit extreme-value modelling for the far tails are a promising extension. The feature vector $\mathbf{f}_t = (\varepsilon_t^2, d_t^-, \ell_t^-)$ was chosen for parsimony; enriching it with additional backward topological features — mean backward edge length, lagged squared innovations, or rolling variance estimates — may sharpen the conditioning signal for markets with complex multi-step clustering patterns, though the trade-off between feature richness and overfitting in the k -NN search warrants careful evaluation. Data-driven selection of k via cross-validation or information-theoretic criteria would eliminate the sole remaining hyperparameter. Finally, the $O(T^2)$ NVG construction cost, feasible for the daily series considered here ($T \leq 1825$), would require approximate nearest-neighbour structures for application to high-frequency intraday data.

Author Contributions: Conceptualization, C.M. and E.M.; methodology, C.M. and E.M.; software, E.M.; validation, C.M. and E.M.; formal analysis, C.M. and E.M.; investigation, C.M. and E.M.; data curation, E.M.; writing - original draft preparation, C.M. and E.M.; writing - review and editing, C.M. and E.M.; visualization, E.M.; supervision, C.M.; project administration, C.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The code and synthetic data generation scripts are publicly available at <https://github.com/DeepCE/VG>. Original Italian energy market data (PUN) are available from GME (Gestore dei Mercati Energetici, <https://www.mercatoelettrico.org>). US energy market data (PJM West Hub) are available from the U.S. Energy Information Administration (EIA, <https://www.eia.gov>).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Weron, R. *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*; Wiley: Hoboken, NJ, USA, 2006.
2. Weron, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *Int. J. Forecast.* **2014**, *30*, 1030–1081. <https://doi.org/10.1016/j.ijforecast.2014.08.008>
3. De Sanctis, A.; Mari, C. Modelling spikes in electricity markets using excitable dynamics. *Physica A* **2007**, *384*, 457–467.
4. Cont, R. Empirical properties of asset returns: Stylized facts and statistical issues. *Quant. Finance* **2001**, *1*, 223–236.
5. Huisman, R.; Huurman, C.; Mahieu, R. Hourly electricity prices in day-ahead markets. *Energy Econ.* **2007**, *29*, 240–248.
6. Mari, C.; Mari, E. A Distribution-Free Neural Estimator for Mean Reversion, with Application to Energy Commodity Markets. *Preprints* **2026**, Preprint, posted 13 March 2026. doi:10.20944/preprints202603.0996.v1
7. Geman, H. *Commodities and Commodity Derivatives: Modelling and Pricing for Agriculturals, Metals and Energy*; Wiley: Chichester, UK, 2005.
8. Mari, C.; Lucheroni, C. Internal hedging of intermittent renewable power generation and optimal portfolio selection. *Ann. Oper. Res.* **2019**. <https://doi.org/10.1007/s10479-019-03221-2>
9. Mari, C. Stochastic NPV based vs stochastic LCOE based power portfolio selection under uncertainty. *Energies* **2020**, *13*, 3677.
10. Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *J. Econom.* **1986**, *31*, 307–327.
11. Tsay, R.S. *Analysis of Financial Time Series*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2005.
12. Mari, C.; Mari, E. Gaussian clustering and jump-diffusion models of electricity prices: A deep-learning analysis. *Decis. Econ. Finance* **2021**. <https://doi.org/10.1007/s10203-021-00332-z>
13. Mari, C.; Baldassari, C. Ensemble methods for jump-diffusion models of power prices. *Energies* **2021**, *14*, 2084.
14. Lacasa, L.; Luque, B.; Ballesteros, F.; Luque, J.; Nuño, J.C. From time series to complex networks: The visibility graph. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 4972–4975.
15. Luque, B.; Lacasa, L.; Ballesteros, F.; Luque, J. Horizontal visibility graphs: Exact results for random time series. *Phys. Rev. E* **2009**, *80*, 046103.
16. Lacasa, L.; Luque, B.; Luque, J.; Nuño, J.C. The visibility graph: A new method for estimating the Hurst exponent of fractional Brownian motion. *Europhys. Lett.* **2009**, *86*, 30001.
17. Wang, N.; Li, D.; Wang, Q. Visibility graph analysis on quarterly macroeconomic series of China based on complex network theory. *Physica A* **2012**, *391*, 6543–6555.
18. Yang, Y.H.; Liu, Y.L.; Shao, Y.H. Visibility graph analysis of crude oil futures markets: Insights from the COVID-19 pandemic and Russia-Ukraine conflict. *Fluct. Noise Lett.* **2025**, *24*, 2550021. <https://doi.org/10.1142/S021947752550021X>
19. Mari, C.; Baldassari, C. Optimization of mixture models on time series networks encoded by visibility graphs: An analysis of the US electricity market. *Comput. Manag. Sci.* **2023**. <https://doi.org/10.1007/s10287-023-00460-4>
20. Ravetti, M.G.; Carpi, L.C.; Gonçalves, B.A.; Frery, A.C.; Rosso, O.A. Distinguishing noise from chaos: Objective versus subjective criteria using horizontal visibility graph. *PLoS ONE* **2014**, *9*, e108004.
21. Zou, Y.; Donner, R.V.; Marwan, N.; Donges, J.F.; Kurths, J. Complex network approaches to nonlinear time series analysis. *Phys. Rep.* **2019**, *787*, 1–97.
22. Bai, S.; Kolter, J.Z.; Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv* **2018**, arXiv:1803.01271. <https://arxiv.org/abs/1803.01271>
23. Jones, M.C.; Pewsey, A. Sinh-arcsinh distributions. *Biometrika* **2009**, *96*, 761–780.
24. Cleveland, W.S. Robust locally weighted regression and smoothing scatterplots. *J. Am. Stat. Assoc.* **1979**, *74*, 829–836.

25. Cleveland, W.S.; Devlin, S.J. Locally weighted regression: An approach to regression analysis by local fitting. *J. Am. Stat. Assoc.* **1988**, *83*, 596–610.
26. Lan, X.; Mo, H.; Chen, S.; Liu, Q.; Deng, Y. Fast transformation from time series to visibility graphs. *Chaos* **2015**, *25*, 083105.
27. Fan, J.; Gijbels, I. *Local Polynomial Modelling and Its Applications*; Chapman & Hall: London, UK, 1996.
28. Hansen, B.E. Autoregressive conditional density estimation. *Int. Econ. Rev.* **1994**, *35*, 705–730.
29. Harvey, C.R.; Siddique, A. Autoregressive conditional skewness. *J. Financ. Quant. Anal.* **1999**, *34*, 465–487.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.