[1] Dynamics of Implied Volatility Surfaces [Rama Cont and Jose da Fonseca, 2002]

- Link: https://www.researchgate.net/publication/227624113_Dynamics_of_Implied_Volatility_Surfaces
- Short summary
    - The paper looks at how the implied volatility surface (IVS) of option prices on the SP500 and FTSE indices moves over time, not just its static smile structure shape.
    - The authors argue that option markets have their own sources of randomness beyond the underlying index, and that IV surfaces behave like a random surface driven by a few factors, which creates "vega risk" for option portfolios.
    - They propose an empirical factor model of the IVS based on a Karhunen–Loeve (PCA) decomposition of daily IV changes.
- Key theory features
    - Implied volatility surface as state variable
        - They argue it's better to model implied vol directly than local volatility, because IV is observable, directly linked to traded options, and familiar to practitioners.
    - Smile rules vs stochastic surface
        - They discuss "sticky moneyness" (surface constant in moneyness coordinates)

$$\forall(m,\tau), \qquad I_{t+\Delta t}(m,\tau) = I_t(m,\tau)$$

and "sticky strike" (IV fixed for each (K,T))

$$\forall(K,T), \qquad \sigma^{\mathrm{BS}}_{t+\Delta t}(K,T) = \sigma^{\mathrm{BS}}_{t}(K,T)$$

, which traders use as deterministic update rules for IVS.

        - By showing large day-to-day variation in IV (e.g. ATM IV moving 15-40% on S&P in a few months

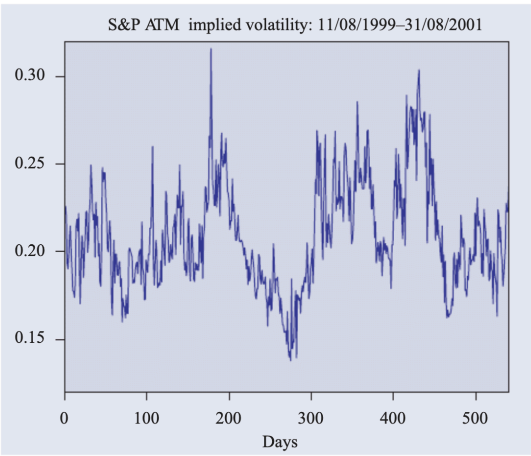


Figure 2. Evolution of at the money implied volatility for SP500 options, Aug. 1999–Aug. 2001.

), they argue these rules are too simplistic and can't capture real volatility risk.

- Implied volatility as a random surface
    - They model the (log) implied volatility surface as a stationary random field on the 2D domain (moneyness, time-to-maturity).
- Karhunen–Loeve decomposition - functional PCA of surfaces
    - They apply a Karhunen–Loève (KL) decomposition, i.e. PCA of the random surface:

$$U(\omega, \cdot) = \sum U_n(\omega) f_n(\cdot) = \sum U_n f_n.$$

where f_n (x) are deterministic eigen-surfaces (principal components in m–τ space) and U_n are uncorrelated factor loadings.

    - Each eigen-surface corresponds to a 'mode of deformation' of the IV surface: level, slope (skew), curvature (smile thickness),
- Mean-reverting factor model
    - After finding the eigenmodes (f_k), they model the factor loadings (x_k(t)) (projections of the IV surface onto these modes) as mean-reverting Ornstein-Uhlenbeck processes, driven by independent noise sources, which can be Wiener or jump processes
    - This gives a low-dimensional stochastic model of the entire IV surface
- Key practical features
    - Data, smoothing, and surface construction
        - They use end-of-day SP500 and FTSE index options over ~1-2 years, focusing on liquid out-of-the-money options with moneyness in [0.5, 1.5] and maturities from around 1 month to 1 year
        - For each day, they construct a smooth IV surface on a fixed grid using a Nadaraya–Watson kernel estimator (Gaussian kernel in m and τ with data-driven bandwidths). This is crucial: the KL decomposition is done on smoothed surfaces, not raw noisy quotes.
    - Dynamics, correlations, and risk implications
        - They explicitly connect this to Vega risk: since multiple factors move (not just level, and not perfectly linked to the underlying), delta hedging is not enough; a factor-based Vega hedge is needed. They also show how their model extends "sticky delta" by adding stochastic deformations around the smile.
- Relevance
    - The paper provides canonical empirical stylized facts about IV surface dynamics:
        - low-dimensional factor structure (level / skew / curvature),
        - mean-reversion times around 1–2 months,
        - strong negative correlation of IV level with underlying (leverage),
        - relatively weak link between underlying and shape factors.
    - Justification for modelling implied vol surfaces directly
        - Our approach is training a diffusion model directly on forward curves + IV surfaces, so fits exactly this philosophy. We cite them as the classical

justification for a "market-based" modelling approach rather than modelling the underlying S or instantaneous volatility only.

- Baseline generative model to compare against
  - Their factor-OU model is essentially a linear generative model for IV surfaces: simulate OU factors -> reconstruct surface via eigenmodes -> price options via Black–Scholes.

[2] Deep Learning from Implied Volatility Surfaces [Kelly, Bryan T. and Kuznetsov, Boris and Malamud, Semyon and Xu, Teng Andrea, 2023]

- Link
  - https://dx.doi.org/10.2139/ssrn.4531181
- Context
  - IV surface = image (moneyness × maturity) containing rich info about state-contingent risk premia and return distribution.
  - Economic theory:
    - local derivatives of IV (Breeden–Litzenberger, Dupire) link surface geometry to Arrow–Debreu prices and volatility.
  - Empirical problem:
    - IV grid is discrete, noisy, illiquid –> hard to compute theory-driven local features directly.
- Theory part – key concepts
  - IV surface as structured image
    - 2D grid of IVs on (delta / moneyness, maturity); "universal local features" = non-linear functions of neighbouring pixels.
  - Economic link
    - Cross-moneyness derivatives → Arrow–Debreu state prices (Breeden–Litzenberger).
    - Term-structure slope → local variance via Dupire.
  - CNN inductive bias
    - locality, translation / rotation invariance
    - shared filters across the surface
    - better suited than fully-connected DNNs for structured IV data
  - Ensemble complexity
    - many local minima in CNN loss landscape
    - averaging many randomly initialized CNNs boosts performance (virtue of complexity)
  - Gradient outer product & principal linear features
    - new ML object; eigenvectors define "principal linear features" (PC-analogue) → no linear feature sparsity; >100 linear features needed to explain predictive content of IV.

$$w_{*,t} \;=\; \arg\min_{w} \ell(w),\;\; \ell(w) \;=\; \sum_{\theta=t-T}^{t} \sum_{i=1}^{N_\theta} (R_{i,\theta+1} - f(IV_{i,\theta}; w))^2 \,,$$

- Practical part – what they actually do

- Data / preprocessing
    - Critical because we'll have to gather data for each day options based on some logic to keep sizing constant in our paper as well
    - OptionMetrics IvyDB IV surfaces → normalized grid (~10 maturities × 34 deltas) per stock-day; remove very short (10-day) expiry; handle missing / incomplete images.
- Models
    - CNN1, CNN4, CNN5 = convolutional nets with 1 / 4 / 5 conv layers; depth = complexity; complexity metric ~ #parameters / sample size.

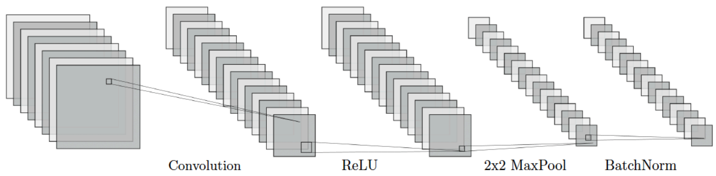


Figure 3: The figure above shows a building block of the CNN model consisting of a convolutional layer with a 3×3 filter, a ReLU layer, 2×2 max-pooling, and batch normalization layers. Note the max-pooling layer shrinks the height and width of the input by half and keeps the same depth.

    - Ensembles of up to 100 randomly initialized CNNs
    - also baselines: linear "kitchen-sink" ridge model and fully-connected NN1
- Main findings:
    - Deep CNN ensembles significantly predict 1-month stock returns using only month-end IV surface; out-of-sample Sharpe for H–L strategy rises from ~0.9 (single model) to ~2.7 (ensemble of 100 for deepest CNNs).
    - Alphas remain significant relative to many equity and option-based factors; robust to transaction costs and short-sale constraints.
    - Principal linear features: performance keeps improving as number of features P grows; need >100 to capture predictive content → very high feature complexity

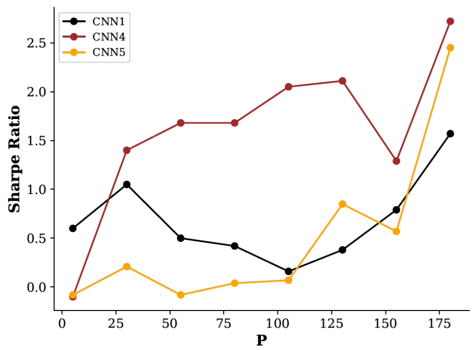


Figure 5: The figures above show the Sharpe Ratio of our H-L strategy (5) as a function of P, the number of principal features, based on the function $f_P(x)$ constructed using Algorithm 1. The experiment is run separately for each of the CNN1, CNN4, and CNN5 models.

- Why this paper matters for our diffusion-IV project
  - Justification for modelling the full surface
    - shows that local geometry of the entire IV surface carries rich, non-linear predictive info that cannot be captured by a few summary statistics (level, slope, skew, convexity)
    - our conditional diffusion model over full surfaces is aligned with this "feature-rich" view
  - Structured-data perspective
    - treats IV as an image with locality and spatial structure; supports the idea of using structured generative models (diffusion with moneyness / maturity structure) rather than flat feature vectors.
  - Complexity vs parsimony
    - their evidence of ensemble and feature complexity provides a motivation to use flexible high-dimensional generative models
    - our work is going to base economic structure (no-arbitrage, P–Q split, forward curve consistency) on top of ML flexibility
  - Positioning
    - their CNNs are discriminative (return prediction);
    - our diffusion is generative (joint future paths of forward curve + IV surface with risk-management P&L metrics) – so we can explicitly present the model as complementary: instead of extracting predictive features, we will try to simulate economically consistent scenarios that could be fed into similar CNN-style predictors or used directly for hedging / risk.

[3] VolGAN: A Generative Model for Arbitrage-Free Implied Volatility Surfaces [Milena Vuletić and Rama Cont, 2024]

- Link
  - https://www.tandfonline.com/doi/full/10.1080/1350486X.2025.2471317
- Context
  - Conditional GAN for joint dynamics of underlying return + IV surface.
  - Trained on SPX options; aims at realistic scenarios + static no-arbitrage.
  - Used for forecasting, VIX simulation, and hedging option portfolios.
- Theory part
  - Static arbitrage penalty

Following Cont and Vuletic (2023), we define the *arbitrage penalty* associated with the (discretely sampled) volatility surface $\boldsymbol{\sigma}(\boldsymbol{m}, \boldsymbol{\tau})$ as:

$$\Phi\left(\boldsymbol{\sigma}(\boldsymbol{m}, \boldsymbol{\tau})\right) = p_1(\boldsymbol{\sigma}(\boldsymbol{m}, \boldsymbol{\tau})) + p_2(\boldsymbol{\sigma}(\boldsymbol{m}, \boldsymbol{\tau})) + p_3(\boldsymbol{\sigma}(\boldsymbol{m}, \boldsymbol{\tau})). \tag{2}$$

where the functions p_1, p_2, p_3 measure violations of calendar, call and butterfly arbitrage constraints, respectively:

$$p_1(\boldsymbol{\sigma}(\boldsymbol{m}, \boldsymbol{\tau})) = \sum_{i=1}^{N_m} \sum_{j=1}^{N_\tau} \left(\tau_j \frac{c(m_i, \tau_j) - c(m_i, \tau_{j+1})}{\tau_{j+1} - \tau_j} \right)^+, \tag{3}$$

$$p_2(\boldsymbol{\sigma}(\boldsymbol{m}, \boldsymbol{\tau})) = \sum_{i=1}^{N_m} \sum_{j=1}^{N_\tau} \left(\frac{c(m_{i+1}, \tau_j) - c(m_i, \tau_j)}{m_{i+1} - m_i} \right)^+, \tag{4}$$

$$p_3(\boldsymbol{\sigma}(\boldsymbol{m}, \boldsymbol{\tau})) = \sum_{i=1}^{N_m} \sum_{j=1}^{N_\tau} \left(\frac{c(m_i, \tau_j) - c(m_{i-1}, \tau_j)}{m_i - m_{i-1}} - \frac{c(m_{i+1}, \tau_j) - c(m_i, \tau_j)}{m_{i+1} - m_i} \right)^+. \tag{5}$$

Static arbitrage constraints (Davis and Hobson 2007) are then equivalent to

$$\Phi(\boldsymbol{\sigma}(\boldsymbol{m}, \boldsymbol{\tau})) = 0$$

and the magnitude of $\Phi(\boldsymbol{\sigma}(\boldsymbol{m}, \boldsymbol{\tau}))$ can be considered as a 'distance' from the set of arbitrage-free implied volatility surfaces.

- Shape constraints for IV surface
  - Monotonicity in maturity ($\partial\tau C \geq 0$), decreasing in moneyness ($\partial m C \leq 0$), convex in moneyness ($\partial^2 m C \geq 0$).
  - Translated into constraints on $\sigma(m,\tau)$ and its derivatives
- Conditional GAN architecture (VolGAN)
  - Condition vector
    - previous IV surface g□(m,τ)
    - last two log-returns
    - previous realized vol
  - Generator G(a□, z) → (next log-return, log-IV increment Δg□(m,τ))

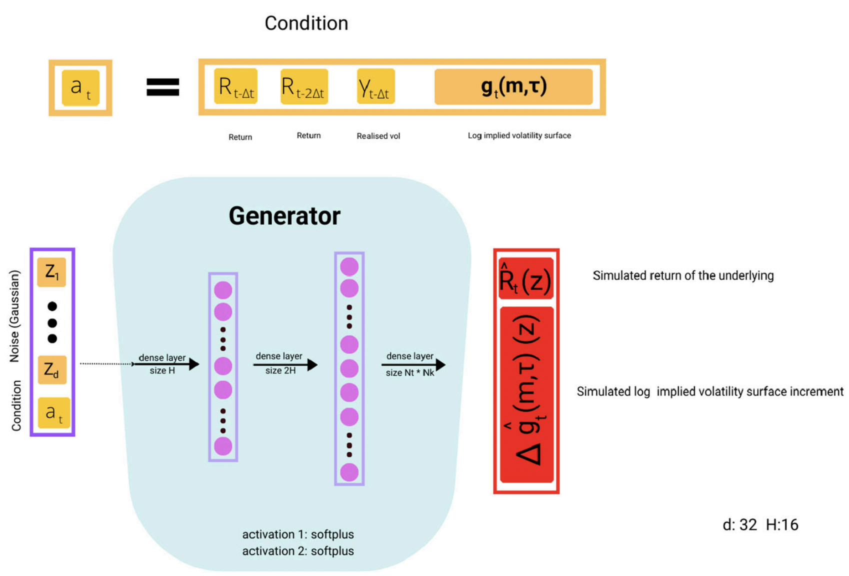


Figure 1. VoLGAN generator architecture.

- Discriminator D(a□, (R,Δg)) distinguishes real vs generated

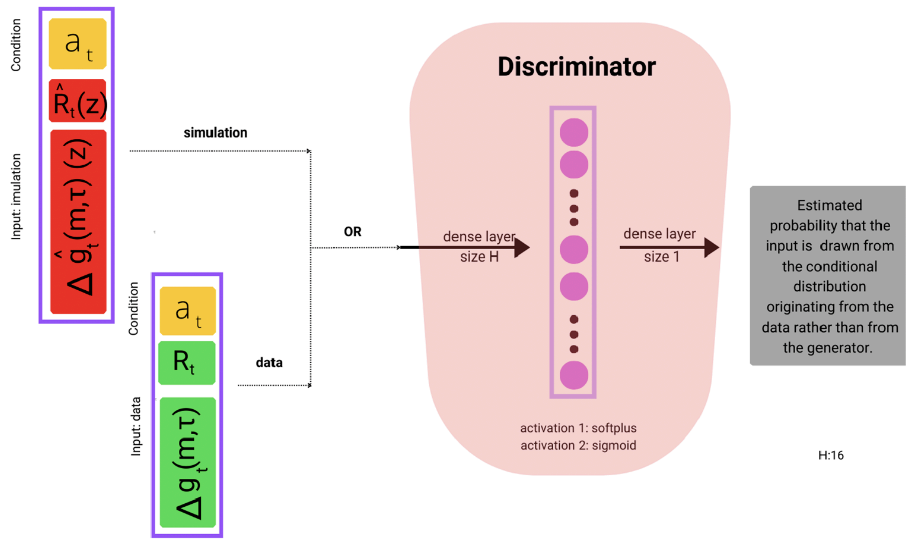
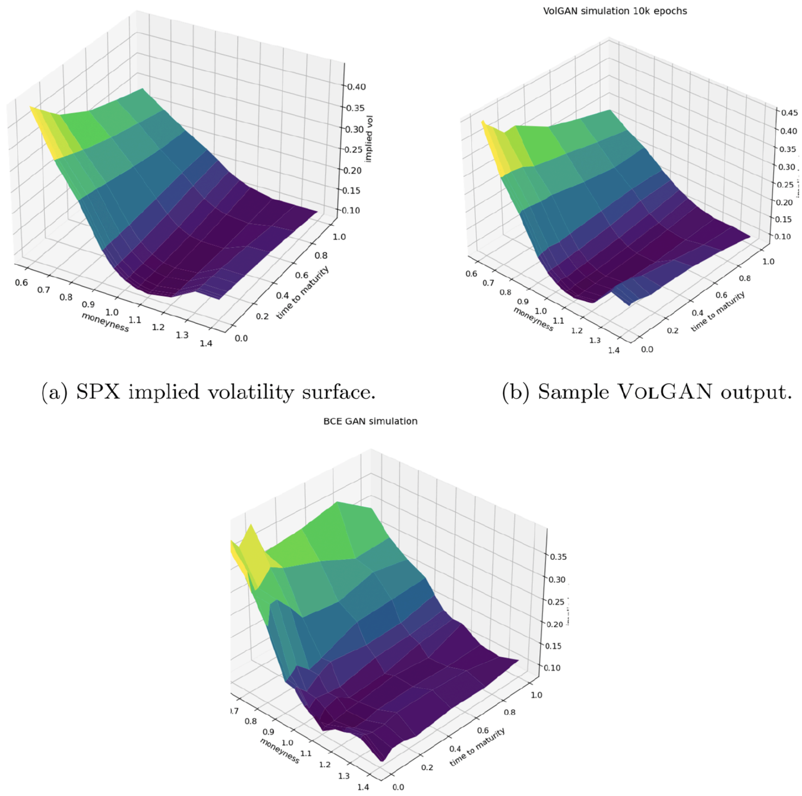


Figure 2. VoLGAN discriminator architecture.

- Smoothness penalty (Sobolev seminorm)
    - L□ and L_τ penalties on discrete derivatives in m and τ for log-IV surface; encourage smooth, “PDE-like” surfaces
    - The result of their work is presented below - we can see it is super smooth, so we have to keep that discovery in our work as well



(a) SPX implied volatility surface. (b) Sample VolGAN output. (c) Sample BCE GAN output.

Figure 6. Implied volatility surfaces generated using (b) VolGAN (c) classical GAN, compared with (a) SPX implied volatility surface.

- Empirical part
  - Hedging application
    - Scenario-based regression hedging of a 1-month straddle using VolGAN scenarios
    - Compare BS delta, BS delta-vega, and VolGAN + regression (LASSO / ATM only)
- Relevance to our conditional diffusion paper
  - Same object, different generator
    - VolGAN provides a GAN-based baseline for our diffusion-based generator.
  - Arbitrage handling paradigm:
    - VolGAN uses smoothness penalties + scenario re-weighting based on a discrete arbitrage functional $\phi(\sigma)$.
    - We can borrow their arbitrage penalty functional and shape constraints
  - Conditional structure & inputs:
    - Their conditioning set (past IV surface + returns + realized vol) is a good template for our conditioning variables in a diffusion model (we just extend it with forward/futures curve, extra history, etc.).
  - Our contribution
    - replace GAN with conditional diffusion
    - add joint forward curve + IV surface, more explicit P/Q separation
    - emphasise P&L-based evaluation using a similar hedging setup

[4] Controllable Generation of Implied Volatility Surfaces with Variational Autoencoders [Wang, Liu, Vuik, 2025]

- Link
  - https://arxiv.org/abs/2509.01743
- Context
  - Proposes a controllable VAE that generates IV surfaces with specified shape features (level, slope, curvature, term-structure slope).
  - Adds post-generation arbitrage repair in latent space (calendar + butterfly).
- Experiment features
  - Dataset
    - 60k synthetic IVSs on fixed 28×28 grid in (m,τ), from Heston + SABR parameter ranges
    - VAE: latent dim z=5, ResNet encoder/decoder, β tuned per experiment

Table 1: Hyperparameters used in the controllable VAE model.

Hyperparameter	**Setting**
Input Dimension ($\mathbf{x}$)	784
Control Dimension ($\mathbf{y}$)	problem-specific
Latent Dimension ($\mathbf{z}$)	5
Encoder hidden layers	[256, 128]
Decoder hidden layers	[128, 256]
Batch size	64
Activation function	ReLU
Optimizer	Adam
Learning rate	3×10^{-4}
β	problem-specific
Max epochs	5000

- Relevance
  - Their VAE is a static generator with feature knobs; the diffusion we are gonna use is a dynamic conditional generator (next-day surface + forward curve).
  - Here we can position our model as: same economic features + richer temporal dynamics and joint forward/IV structure.
  - Arbitrage treatment template:
    - They use post-hoc latent optimization with explicit calendar/butterfly penalties.
    - Here we can adopt their penalty functions and diagnostics as evaluation tools;
    - In the contrary, our diffusion aims to build no-arbitrage into the generative dynamics, not only fix outputs afterwards.

[5] Diffusion-Based Generative Modeling of Financial Time Series [2025]

- https://hdl.handle.net/10012/22497
- Problem: generative modeling of multi-asset financial time series beyond GBM/Heston-type parametrics.
- Proposed method: Elucidated Diffusion Model + NCSN++ backbone adapted to returns, with “Ambient Diffusion” variance correction and analytic noise schedule.

- Data/metrics: synthetic GBM/Heston/Merton + real SPY, NVDA, BTC etc; evaluated on distribution fit, SDE parameter recovery, option pricing and risk metrics (VaR/CVaR).
- Result: ambient-corrected diffusion markedly reduces volatility bias and pricing error versus vanilla EDM across assets.

[6] Volatility Surface Completion using Score-Based Generative Models [2023]

- Problem: "volatility completion" – filling missing IV quotes on a strike–maturity grid while enforcing static no-arbitrage.
- Proposed method: treat IV surface as an image on a fixed grid and use a noise-conditional score network with Langevin inpainting, modified to impose butterfly + calendar constraints during sampling.
- Data and metrics: Heston-generated IVS on 8×8 and 16×16 grids; interpolation and randomized masks up to 80% missing.
- Result: interpolation errors $\approx 10^{-4}$ and max error <0.5% even with 80% missing, while keeping surfaces essentially arbitrage-free.

[7] Forecasting Implied Volatility Surface with Generative Diffusion Models [2025]

- https://arxiv.org/abs/2511.07571
- Problem: 1-day-ahead forecasting of arbitrage-free IV surfaces on SPX using a data-driven model that handles path dependence and residual arbitrage in training data.
- Proposed method: conditional DDPM on a 9×9 (moneyness, maturity) grid, conditioned on EWMAs of past IV surfaces, returns/squared returns, and VIX; loss combines reconstruction with SNR-weighted arbitrage penalty.
- Key theory: shows the arbitrage penalty introduces a small, controllable bias while steering the model toward the arbitrage-free manifold.
- Experimental result: on SPX OTM options, diffusion model beats VolGAN variants on MAPE and delivers better-calibrated 90% CIs for vols across ATM/OTM/ITM buckets and horizons.

[8] Generating the Term Structure of Interest Rates with Diffusion Models [2025]

- https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5493026
- Problem: generative modeling of zero-coupon yield curves (term structures) conditional on macro/market variables across currencies.
- Proposed method: conditional DDPM with v-parameterization and cross-attention U-Net, generating dense OIS-based yield curves for JPY/USD/GBP given macro and rate inputs.
- Variants: direct curve generation vs generating day-to-day differences; plus a faster Nelson-Siegel-Svensson (NSS) factor-based diffusion.
- Results: realistic yield-curve shapes and dynamics, good 6-month out-of-sample behavior; NSS factor model cuts training/inference time by ≈40–44% with similar quality.

[9] A Neural Network Approach to Understanding Implied Volatility Movements [2019]

- https://www.researchgate.net/publication/341383844_A_neural_network_approach_to_understanding_implied_volatility_movements
- Problem: empirical modeling of how the IV surface moves with index returns, moneyness and maturity for S&P 500 options.
- Proposed method: “three-feature” NN (return, delta-moneyness, time to maturity) and “four-feature” NN (adds VIX) trained on ~2M daily SPX call observations (2010–2017).
- Use cases: compare empirical surface dynamics to stochastic-vol models; compute minimum-variance delta that accounts for expected IV changes.
- Key result: NNs significantly outperform a simple analytic regression model across regimes; VIX as fourth feature further improves fit, especially in high-vol markets.