Volatility is (Mostly) Path-Dependent

Julien Guyon

Bloomberg L.P., Quantitative Research
Columbia University, Department of Mathematics
NYU, Courant Institute of Mathematical Sciences

Joint work with Jordan Lekeufack
University of California, Berkeley, Department of Statistics
Bloomberg PhD Fellow

BSDE 2022
The 9th International Colloquium on BSDEs and Mean Field Systems
June 28, 2022

jguyon2@bloomberg.net, jg3601@columbia.edu, julien.guyon@nyu.edu

Outline

1. Why Path-Dependent Volatility (PDV)?
2. Is Volatility Path-Dependent? How much? How?
3. The continuous-time empirical Markovian PDV model

This talk is dedicated to the memory of Peter Carr

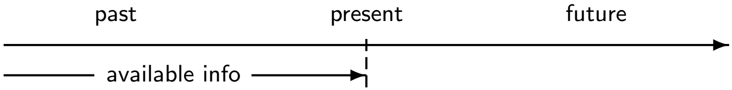
Path-Dependent Volatility

$$\frac{dS_t}{S_t} = \sigma(S_u, u \leq t)\, dW_t$$

- Zero rates, repos, dividends for simplicity
- Volatility drives the dynamics of the asset price S
- Feedback loop from prices to volatility
- Pure feedback model: volatility is an **endogenous** factor
- Main references:
  - Econometrics:
    The whole GARCH literature
  - Derivatives research (macro, pricing models, calibration):
    Hobson-Rogers '98, Guyon '14
  - Econophysics (micro, statistical models):
    Zumbach '09-10, Chicheportiche-Bouchaud '14, Blanc-Donier-Bouchaud '16
  - Recent models with a PDV component:
    Gatheral-Jusselin-Rosenbaum '20, Parent '21

Why Path-Dependent Volatility?

A philosophical argument



- The arrow of time
- Markovian assumption: the future depends on the past only through the present
- Often made just for simplicity and ease of computation, not a fundamental property
- Example: assume that the price of an option depends only on current time t and current asset price S_t: $P(t, S_t)$
- In fact, often, **the present does not capture all information from the past** $\longrightarrow P(t, (S_u, u \leq t))$

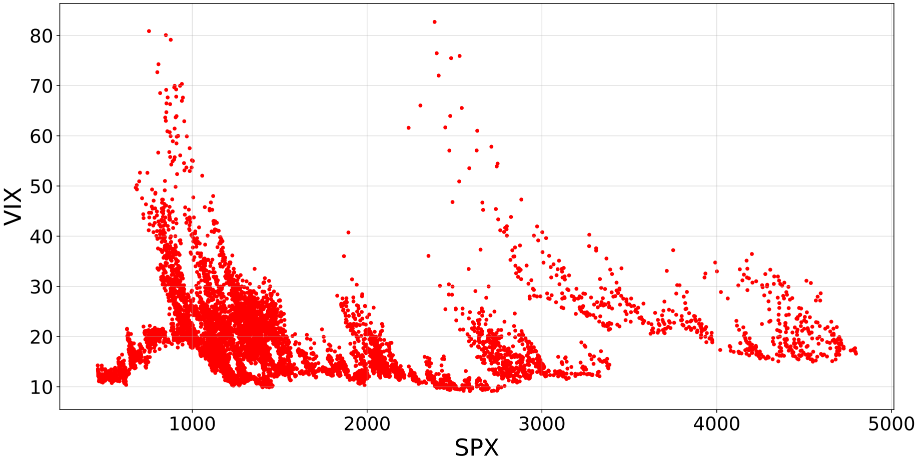
An intuitive argument: a simple quizz

	June 28, 2022		June 28, 2023
SPX	4,000		5,400
VIX			?

An intuitive argument: a simple quizz

	June 28, 2022	May 28, 2023	June 28, 2023
SPX	4,000	6,000	5,400
VIX			?

An intuitive argument: a simple quizz



A financial and scaling argument

- The two basic quantities that possess a natural scale are the **volatility levels** and the **asset returns**
- **A good model should relate these two quantities**: **Path-dependent volatility**
- **LV model** links the **volatility level** to the **asset level**. Does not make much financial sense: well chosen PDV models need not be recalibrated as often as the LV model.
- **SV models** connect the **volatility return** to the **asset price return**. Has limitations:
  - Only very high levels of vol of vol allow fast large movements of volatility
  - Typically a very large mean-reversion is postulated to keep volatility within its natural range.
- **PDV models**, by directly linking **past asset returns** to **volatility levels**, can capture fast large changes in vol more easily and naturally, while maintaining volatility in its natural range. They also provide an explanation for mean-reversion.

A financial and scaling argument

	volatility	depends on	asset
LV	level		level
SV	returns		returns
PDV	level		returns

Path-dependent volatility vs Stochastic volatility

$$\frac{dS_t}{S_t} = \sigma_t\, dW_t, \qquad \sigma_t = f(t, Y_t)$$

$$dY_t = \mu(t, Y_t)\, dt + \nu(t, Y_t) \left(\rho\, dW_t + \sqrt{1-\rho^2}\, dW_t^{\perp}\right)$$

$$Y_t = Y_0 + \int_0^t \mu(u, Y_u)\, du + \int_0^t \nu(u, Y_u) \left(\rho \frac{1}{f(u, Y_u)} \frac{dS_u}{S_u} + \sqrt{1-\rho^2}\, dW_u^{\perp}\right)$$

- $\rho = 0$: **SV is strictly path-independent**
  - The asset price is a slave process with absolutely no feedback on volatility:

$$\sigma_t = \varphi(t, (dW_u^{\perp})_{0\le u\le t}) = \psi(t, (W_u^{\perp})_{0\le u\le t})$$

- $\rho \notin \{-1, 0, 1\}$: **SV is partially path-dependent**
  - Partial feedback from asset price to volatility through spot-vol correl(s):

$$\sigma_t = \varphi\left(t, \left(\frac{dS_u}{S_u}\right)_{0\le u\le t}, \left(dW_u^{\perp}\right)_{0\le u\le t}\right) = \psi\left(t, (S_u)_{0\le u\le t}, \left(W_u^{\perp}\right)_{0\le u\le t}\right)$$

- $\rho = \pm 1$: **SV is fully path-dependent**
  - Pure feedback but path-dependence φ, ψ is complicated, implicit:

$$\sigma_t = \varphi\left(t, \left(\frac{dS_u}{S_u}\right)_{0\le u\le t}\right) = \psi(t, (S_u)_{0\le u\le t})$$

Joint calibration of SV models to SPX and VIX smiles

The joint calibration of classical parametric SV models to SPX and VIX smiles leads to

- Very large vol of vol
- Very large mean-reversions (several time scales)
- **Correlations = $\pm 1 \Longrightarrow$ Path-dependent volatility**

See:

- *Inversion of Convex Ordering in the VIX Market* (G. '20)
- *The VIX Future in Bergomi Models: Fast Approximation Formulas and Joint Calibration with S&P 500 Skew* (G. '21)

Joint calibration of SV models to SPX and VIX smiles

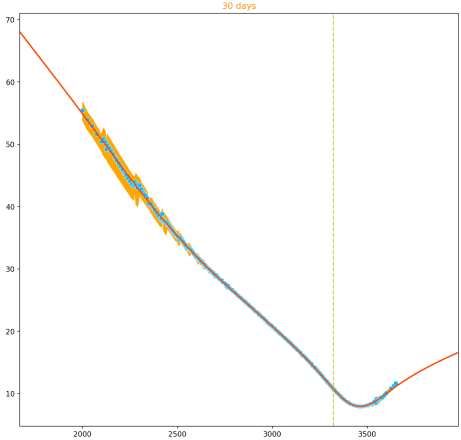


Figure: SPX smile as of January 22, 2020, $T = 30$ days

Joint calibration of SV models to SPX and VIX smiles

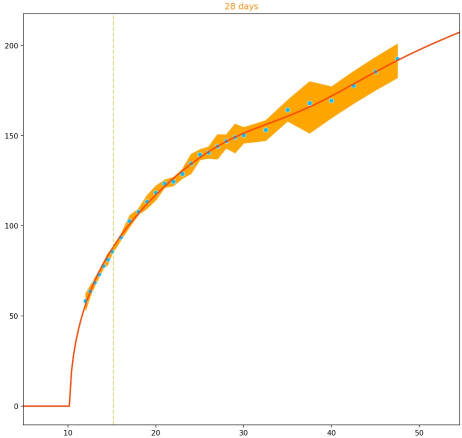


Figure: VIX smile as of January 22, 2020, $T = 28$ days

Joint calibration of SV models to SPX and VIX smiles

- ATM skew:

$$\text{Definition:} \quad \mathcal{S}_T = \left.\frac{d\sigma_{\text{BS}}(K,T)}{\frac{dK}{K}}\right|_{K=F_T}$$

$$\text{SPX, small } T\text{:} \quad \mathcal{S}_T \approx -1.5$$

$$\text{Classical one-factor SV model:} \quad \mathcal{S}_T \underset{T\to 0}{\longrightarrow} \frac{1}{2} \times \text{spot-vol correl} \times \text{vol of vol}$$

- Calibration to short-term ATM SPX skew $\Longrightarrow$

  vol of vol $\geq 3 = 300\% \gg$ short-term ATM VIX implied vol

- $\Longrightarrow$ Use
  - **very large vol of vol**
  - **very large mean-reversion(s)** (so that VIX implied vol $\ll$ vol of vol)
  - -1 **spot-vol correlation(s)**

An information-theoretical/financial economics argument

- Contrary to SV models, PDV models do not require adding extra sources of randomness to generate rich spot-vol dynamics: they explain volatility in a purely **endogenous** way.
- $\Longrightarrow$ Unlike SV models, PDV models are **complete models**: derivatives have a unique, unambiguous price, independent of any preferences or utility functions.
- **All the information exchanged by market participants is recorded in the underlying asset prices**, not just in current prices, but in the history of all past prices.
- Reality is a bit more complex, but we will show that it is actually quite close to this, so it makes sense to **start building a model by extracting all the information that past asset prices contain about volatility**.

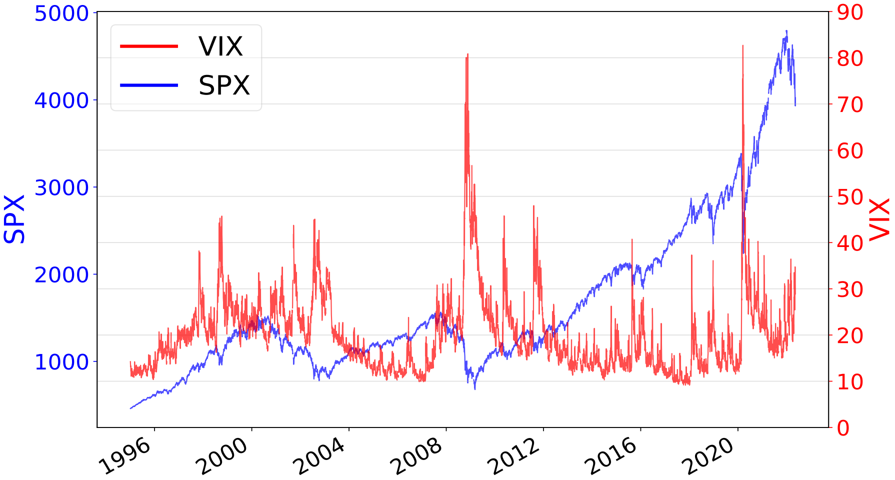
Path-dependent volatility is generic for option pricing

- **All SV models have an equivalent PDV model** in the sense that **all path-dependent options (not only vanilla options)** written on the underlying asset **have the same prices in both models**.
- Brunick and Shreve '13: Given a general Itô process $dS_t = \sigma_t S_t \, dW_t$, there exists a PDV model $d\hat{S}_t = \sigma(t, (\hat{S}_u)_{u \leq t})\hat{S}_t \, d\hat{W}_t$ such that the distributions of the **processes** $(S_t)_{t \geq 0}$ and $(\hat{S}_t)_{t \geq 0}$ are equal; the equivalent PDV is given by

$$\sigma(t, (S_u)_{u \leq t})^2 = \mathbb{E}[\sigma_t^2 | (S_u)_{u \leq t}].$$

- $\Longrightarrow$ The price process $(S_t)_{t \geq 0}$ produced by any SV or stochastic local volatility (SLV) model **can be exactly reproduced by a PDV model**.

Empirical evidence



Empirical evidence

- Much of the GARCH literature
- **Time reversal asymmetry in finance**: Zumbach-Lynch '01, Zumbach '09, Chicheportiche-Bouchaud '14...: "Financial time series are not statistically symmetrical when past and future are interchanged" (BDB '16)
- Leverage effect:
  - Past returns affect (negatively) future realized volatilities, but not the other way round" (BDB '16)
  - $t \to -t$ and $r \to -r$ asymmetry
- ZL '01: time reversal asymmetry even in absence of leverage effect:
  - Weak Zumbach effect: "Past large-scale realized volatilities are more correlated with future small-scale realized volatilities than vice versa" (BDB '16). Most easily captured by PDV models.
  - $t \to -t$ asymmetry, but $r \to -r$ symmetry
- Strong Zumbach effect: "Conditional dynamics of volatility with respect to the past depend not only on past volatility trajectory but also on the historical price path" (GJR '20) $\iff$ **There is some price-path-dependency in the volatility dynamics**

Empirical evidence

Our Machine Learning approach confirms those findings and moreover answer two crucial questions:

1. **How exactly does volatility depend on past price returns (price trends and past squared returns)?**
2. **How much of volatility is path-dependent, i.e., purely endogenous?**

That is, explain volatility as an endogenous factor as best as we can, empirically.

Objectives

(1) Learn path-dependent volatility empirically

- Learn how much of volatility is path-dependent, and how it depends on past asset returns.
- Empirical study: learn implied volatility (VIX) and future Realized Volatility (RV) from SPX path [+ other equity indexes].
- Historical PDV or Empirical PDV or $\mathbb{P}$-PDV.

(2) Build continuous-time Markovian version of empirical PDV model

- Extremely realistic sample paths + SPX and VIX smiles.

(3) Jointly calibrate Model (2) to SPX and VIX smiles

- Modify parameters of historical PDV model to fit market smiles: $\mathbb{P} \neq \mathbb{Q}$.
- Implied PDV or Risk-neutral PDV or $\mathbb{Q}$-PDV.

(4) Add SV to account for the (small) exogenous part: PDSV

- SV component built from the analysis of residuals $\frac{\text{true vol}}{\text{predicted PDV vol}} \approx 1$.

Is Volatility Path-Dependent?

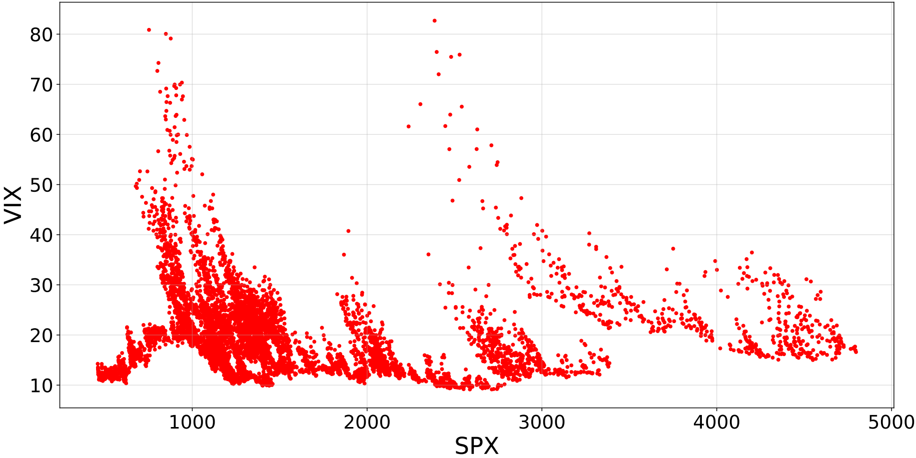
Is volatility path-dependent? A Machine Learning approach

- Objective: **learn from data how much the volatility level depends on past asset returns**.
- Learn Volatility (VIX or RV) from SPX path:

$$\text{Volatility}_t = f(S_u, u \leq t) + \varepsilon$$

- $\longrightarrow$ Historical PDV / Empirical PDV / $\mathbb{P}$-PDV
- Feature engineering: find relevant SPX path features.
- Try various models: various sets of features and parametric forms for f_θ.
- Select the one(s) with the best validation score.
- Check how the models perform on the test set.
- Training set: 2004–18; test set: 2019–22.
- A very challenging test set! Due to the Covid-19 pandemic, the test set includes very different volatility regimes
- **As a result of this analysis, we propose a new, simple PDV model that performs better than existing models**.

Feature engineering



Price path features should be **scale-invariant**

Feature engineering

We focus on **two main types of features**:

[1] Features that capture a **recent trend** in the asset price:

- in order to learn the **leverage effect**: volatility tends to be higher when asset prices fall.
- in order to capture the **strong Zumbach effect**.

[2] Features that capture **recent activity (volatility)** in the asset price (regardless of trend):

- in order to learn **volatility clustering**:
  - periods of large volatility tend to be followed by periods of large volatility.
  - implied volatility tends to be larger when historical volatility is larger.
- in order to capture both the **weak and strong Zumbach effects**.

Trend features

- The most important example of a trend feature is a weighted sum of past daily returns

$$R_{1,t} := \sum_{t_i \leq t} K_1(t - t_i)\, r_{t_i}$$

where

$$r_{t_i} := \frac{S_{t_i} - S_{t_{i-1}}}{S_{t_{i-1}}} \qquad \text{(scale invariance)}$$

- K_1: convolution kernel that typically decreases towards zero: the impact of a given daily return fades away over time.
- Another example:

$$N_t := \sum_{t_i \leq t} K_N(t - t_i) r_{t_i}^- \quad \text{or more generally } N_t^\varphi := \sum_{t_i \leq t} K_N(t - t_i)\varphi(r_{t_i})$$

with, e.g., $\varphi(r) = r^+$ or r^3 or $(r^-)^2$.
- Another example: spot-to-moving-average ratio

$$U_t := \frac{S_t}{A_t}, \qquad A_t := \sum_{t_i \leq t} K_A(t - t_i)\, S_{t_i}.$$

Activity features (or volatility features)

- The most important example of a volatility feature is a weighted sum of past squared daily returns

$$R_{2,t} := \sum_{t_i \le t} K_2(t - t_i)\, r_{t_i}^2.$$

- K_2-weighted historical volatility:

$$\Sigma_t := \sqrt{R_{2,t}}$$

- Higher even moments of past daily returns may also be considered.

Our model

$$\text{Volatility}_t = \beta_0 + \beta_1 R_{1,t} + \beta_2 \Sigma_t, \qquad \beta_0 > 0, \;\; \beta_1 < 0, \;\; \beta_2 \in (0,1)$$

- Volatility_t denotes either some implied volatility (e.g., the VIX) observed at t, or the future realized volatility RV_t (realized over day "$t+1$").
- Leverage effect: $\beta_1 < 0$.
- Volatility clustering, like in GARCH models: $\beta_2 \in (0,1)$.
- Importantly, **both factors $R_{1,t}$ and Σ_t are needed to satisfactorily explain the volatility**.
- We find that **a simple linear model does the job, explaining a very large part of the variability observed in the volatility.**

Kernels

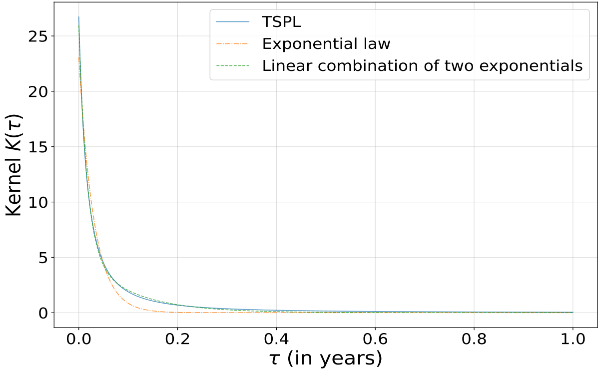


Figure: Typical kernel shapes

Kernels

- The two kernels K_1 and K_2 are distinct.
- Both mix short and long memory
- We consider kernels K_1, K_2 with power-law decay because the data shows that
  1. Very recent daily returns are given much more weight than older daily returns: the weights $K_n(\tau)$ decrease fast for small lags τ.
  2. Nevertheless, volatility has long memory: the weights $K_n(\tau)$ decrease slowly for large τ: persistence of volatility.

The power law aggregates the various time horizons of investors.

Kernels

- This was checked by running a multivariate lasso regression with variables $R_{1,t}^{(\lambda_j)}$ and $\sqrt{R_{2,t}^{(\mu_k)}}$, where

$$R_{n,t}^{(\lambda)} := \sum_{t_i \le t} K^{(\lambda)}(t - t_i)\, r_{t_i}^n, \qquad K^{(\lambda)}(\tau) := \lambda e^{-\lambda\tau}, \qquad \lambda > 0.$$

- For both $n = 1$ and $n = 2$, lasso selects a multitude of λ's which, combined, form a kernel that looks like a power law, except that for vanishing lags τ the kernels do not seem to blow up (the largest λ's are not selected).
- $\Longrightarrow$ We choose both kernels to be **time-shifted power laws** (TSPL):

$$K(\tau) = K_{\alpha,\delta}(\tau) := Z_{\alpha,\delta}^{-1}(\tau + \delta)^{-\alpha}, \qquad \tau \ge 0, \qquad \alpha > 1, \ \delta > 0,$$

  with only two parameters α, δ.
- The time shift δ (one to a few weeks) guarantees that $K_{\alpha,\delta}(\tau)$ **does not blow up when the lag τ vanishes**.
- If we force δ to be 0, we recover the power-law kernel of rough volatility models. However, our empirical tests always select **positive δ**.

Similar models

- QARCH (Sentana '95):

$$\text{Volatility}_t^2 = \beta_0 + \beta_1 R_{1,t} + \beta_2 R_{2,t}^{\mathrm{Q}}, \qquad R_{2,t}^{\mathrm{Q}} := \sum_{t_i,t_j \le t} K_2^{\mathrm{Q}}(t-t_i, t-t_j)\, r_{t_i} r_{t_j}$$

- Diagonal QARCH model (CB '14, $K_2(\tau) := K_2^{\mathrm{Q}}(\tau,\tau)$):

$$\text{Volatility}_t^2 = \beta_0 + \beta_1 R_{1,t} + \beta_2 R_{2,t} \qquad \text{(M1)}$$

- ZHawkes process (BDB '16):

$$\text{Volatility}_t^2 = \beta_0 + \beta_1 R_{1,t}^2 + \beta_2 R_{2,t} \qquad \text{(M2)}$$

- Discrete-time version of the quadratic rough Heston model (GJR '20, $\theta_0 = 0$):

$$\text{Volatility}_t^2 = \beta_0 + \beta_1 (R_{1,t} - \beta_2)^2 \qquad \text{(M3)}$$

with Mittag-Leffler kernel K_1.

- Discrete-time version of the threshold EWMA Heston model (Parent '21):

$$\text{Volatility}_t^2 = \beta_0 + \beta_1 (R_{1,t} - \beta_2)^2 \mathbf{1}_{\{R_{1,t} \le \beta_2\}} \qquad \text{(M4)}$$

with K_1 an exponential kernel, $K_1(\tau) = \lambda e^{-\lambda\tau}$.

Our model differs in several ways

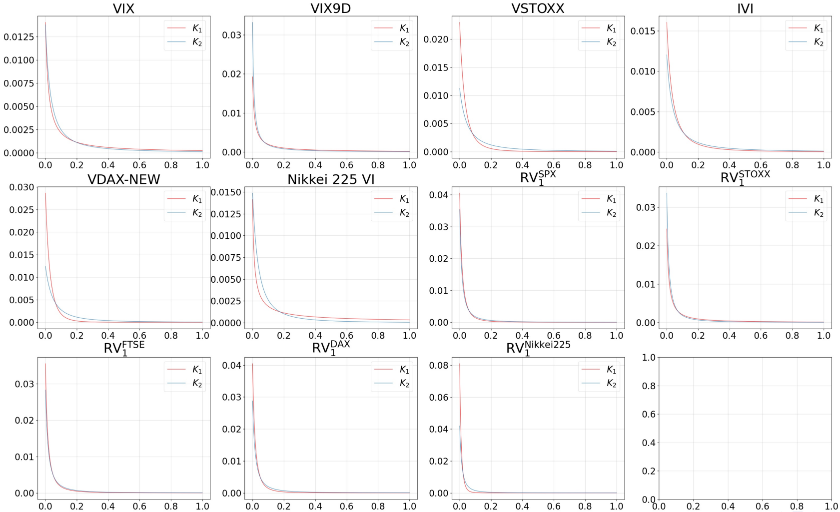
1. All the above models, like almost all ARCH models, model the **square** of the volatility, the variance. Instead, **we directly model the volatility itself**.
2. We use the square root Σ_t of $R_{2,t}$ rather than $R_{2,t}$ itself as one of the linear factors.
3. We use new, explicit parametric forms for the kernels K_1 and K_2, capturing **non-blowing-up power-law-like decays**.
4. Compared with (M3) and (M4), we empirically prove the **importance of including the historical volatility factor Σ_t**.
5. Compared with (M2), we argue that it is **not necessary to include a quadratic factor $R_{1,t}^2$**, as the quadratic-like dependence of the volatility (resp. variance) on $R_{1,t}$ is already captured by the factor Σ_t (resp. $R_{2,t}$).

Results: Implied volatility

	β_0	α_1	δ_1	β_1	α_2	δ_2	β_2
VIX	0.057	1.057	0.020	-23.829	1.597	0.052	0.819
VIX9D	0.045	0.993	0.011	-30.655	1.252	0.011	0.884
VSTOXX	0.032	3.959	0.127	-9.192	1.895	0.089	0.966
IVI	0.022	2.262	0.081	-14.640	1.630	0.063	0.991
VDAX-NEW	0.036	5.540	0.156	-6.149	2.207	0.103	0.922
Nikkei 225 VI	0.055	0.778	0.008	-17.337	2.090	0.077	0.855

Table: Table of optimal parameters for different implied volatility indexes.

Results: Implied volatility



Results: Implied volatility

	Train RMSE	Train r^2	Test RMSE	Test r^2
VIX	0.020	0.946	0.035	0.855
VIX9D	0.023	0.876	0.034	0.914
VSTOXX	0.026	0.929	0.029	0.913
IVI	0.024	0.925	0.030	0.871
VDAX-NEW	0.025	0.934	0.028	0.918
Nikkei 225 VI	0.030	0.890	0.031	0.799

Table: Table of r^2 scores and RMSE for various implied volatility indexes.

Results: Implied volatility

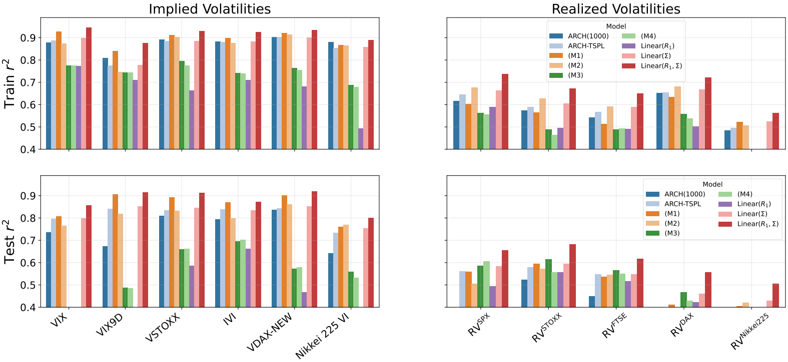


Figure: Comparison of r^2 scores for the different models (M1)-(M4) and our linear models. Top: r^2 score on train set. Bottom: r^2 score on test set. Left: Implied volatilities. Right: Realized volatilities.

Results: Implied volatility

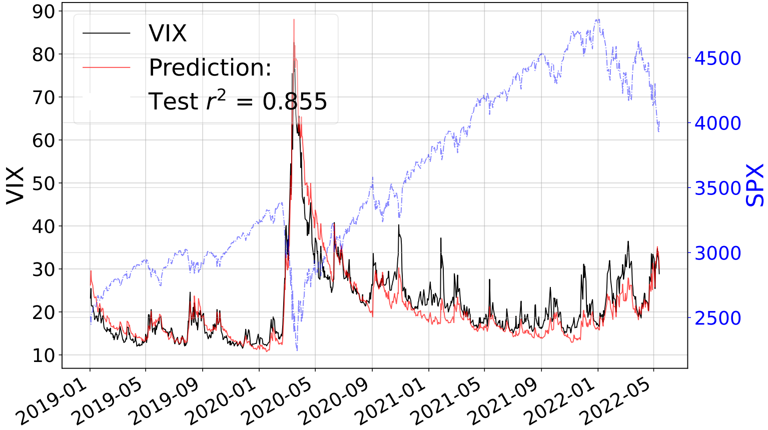


Figure: Time series of VIX and predicted values on test set. The dashed blue line represents the SPX time series.

Results: Implied volatility

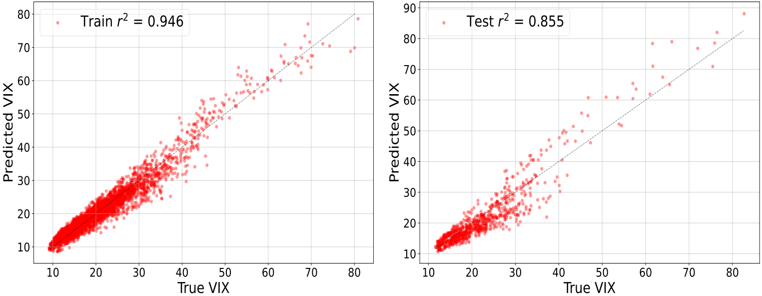


Figure: Predicted VIX vs true VIX on train/test set.

Results: Implied volatility

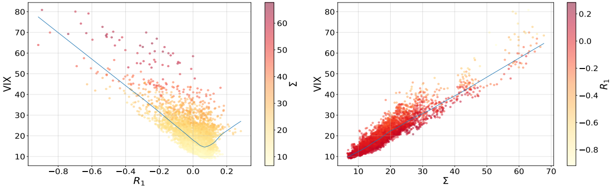


Figure: VIX vs features on the train data set.

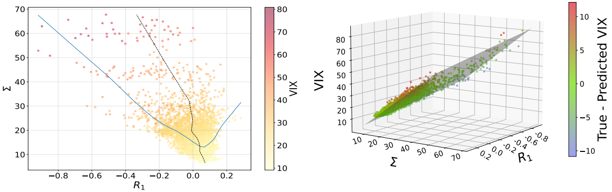


Figure: Σ vs R_1 on the train data set and 3D scatter plot of VIX vs R_1 and Σ

Results: Implied volatility

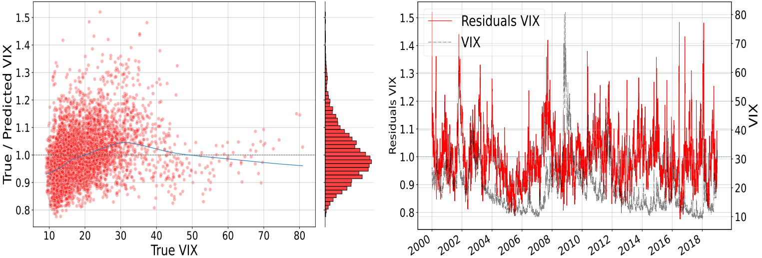


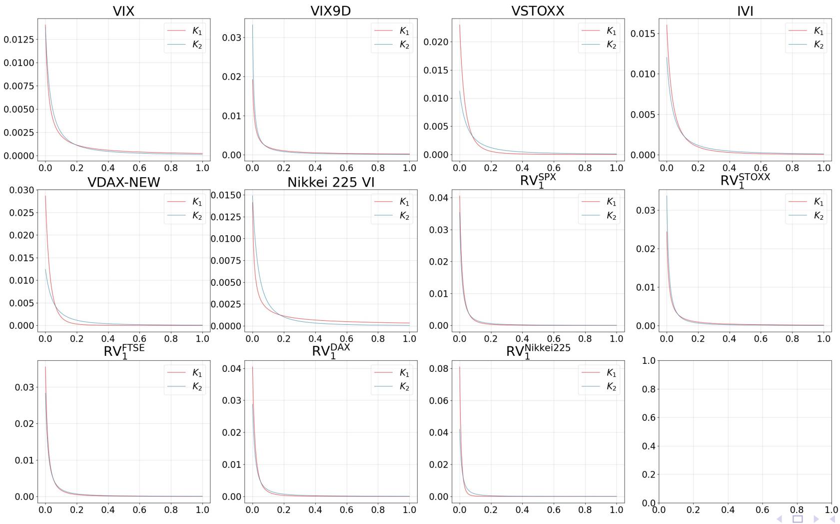
Figure: Residuals plots for VIX predictions.

Results: Realized volatility

	β_0	α_1	δ_1	β_1	α_2	δ_2	β_2
SPX	0.018	2.821	0.044	-10.490	1.860	0.025	0.708
STOXX	0.023	1.306	0.017	-15.567	1.787	0.024	0.697
FTSE	0.017	2.216	0.034	-10.753	1.837	0.031	0.762
DAX	0.001	2.868	0.045	-7.570	1.800	0.029	0.812
NIKKEI	0.032	6.296	0.063	-2.802	2.292	0.030	0.511

Table: Table of optimal parameters for the realized volatility for different indexes.

Results: Realized volatility



Results: Realized volatility

	Train RMSE	Train r^2	Test RMSE	Test r^2
SPX	0.049	0.738	0.063	0.654
STOXX	0.060	0.672	0.064	0.682
FTSE 100	0.055	0.650	0.066	0.617
DAX	0.057	0.722	0.059	0.557
NIKKEI	0.051	0.563	0.051	0.504

Table: Table of r^2 scores and RMSE for the realized volatility of several indexes

Results: Realized volatility

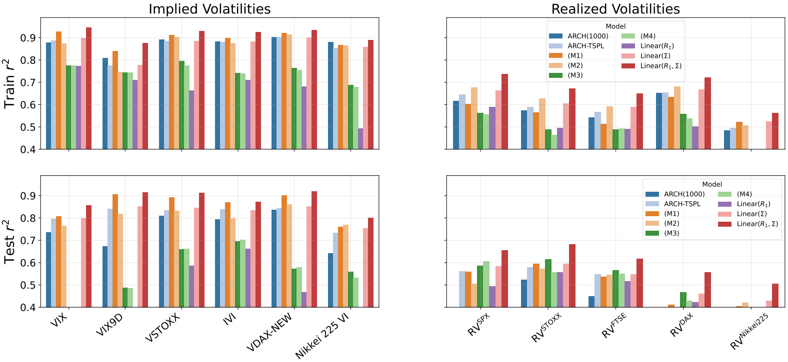


Figure: Comparison of r^2 scores for the different models (M1)-(M4) and our linear models. Top: r^2 score on train set. Bottom: r^2 score on test set. Left: Implied volatilities. Right: Realized volatilities.

Results: Realized volatility

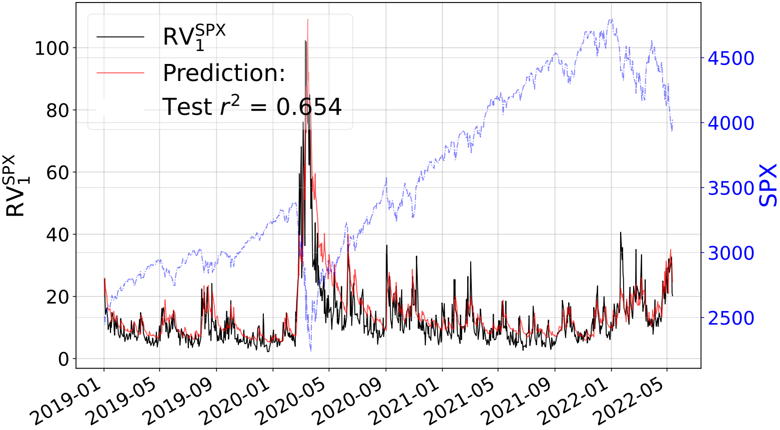


Figure: Time series of VIX and predicted values on test set. The dashed blue line represents the SPX time series.

Results: Realized volatility

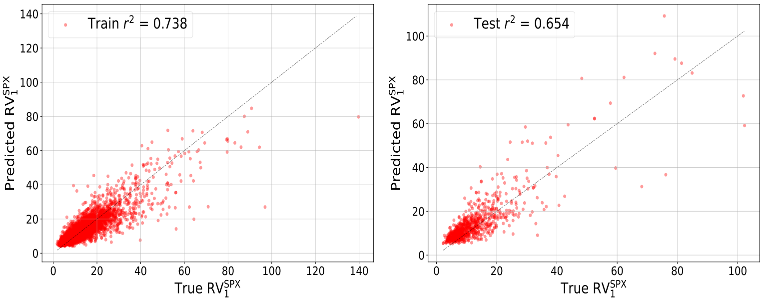


Figure: Predicted VIX vs true VIX on train/test set.

Results: Realized volatility

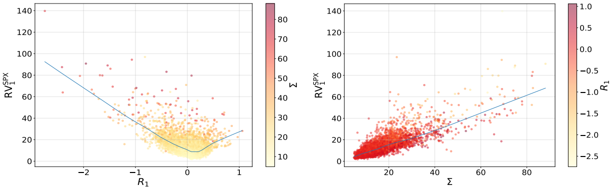


Figure: $\mathrm{RV}^{\mathrm{SPX}}$ vs features on the train data set.

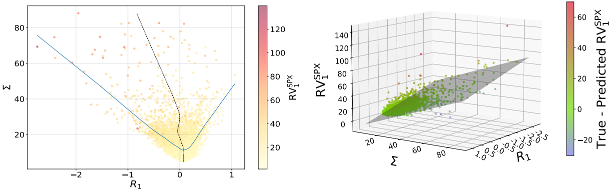


Figure: Σ vs R_1 on the train data set and 3D scatter plot of $\mathrm{RV}^{\mathrm{SPX}}$ vs R_1 and Σ.

Results: Realized volatility

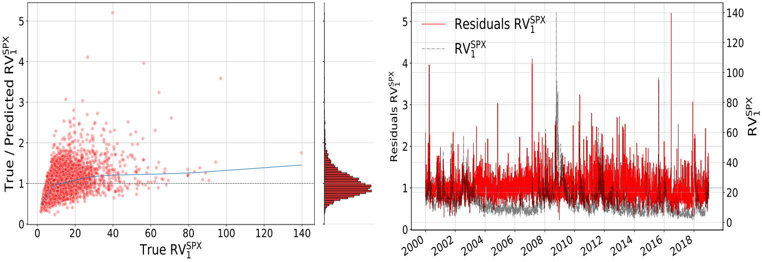


Figure: Residuals plots for RV^{SPX} predictions.

The Continuous-Time Empirical Path-Dependent Volatility Model

The Continuous-Time Empirical Path-Dependent Volatility Model

We now consider the **continuous-time limit** of our empirical PDV model, where we identify Volatility_t as the instantaneous volatility σ_t:

$$\begin{aligned}
\frac{dS_t}{S_t} &= \sigma_t \, dW_t, \\
\sigma_t &= \sigma(R_{1,t}, R_{2,t}) \\
\sigma(R_1, R_2) &= \beta_0 + \beta_1 R_1 + \beta_2 \sqrt{R_2} \\
R_{1,t} &= \int_{-\infty}^{t} K_1(t-u) \frac{dS_u}{S_u} = \int_{-\infty}^{t} K_1(t-u) \sigma_u \, dW_u, \\
R_{2,t} &= \int_{-\infty}^{t} K_2(t-u) \left(\frac{dS_u}{S_u} \right)^2 = \int_{-\infty}^{t} K_2(t-u) \sigma_u^2 \, du.
\end{aligned} \tag{1}$$

The Continuous-Time Empirical Path-Dependent Volatility Model

The dynamics of $R_{1,t}$ and $R_{2,t}$

$$\begin{aligned}
dR_{1,t} &= \left(\int_{-\infty}^{t} K_1'(t-u)\frac{dS_u}{S_u}\right) dt + K_1(0)\frac{dS_t}{S_t} \\
&= \left(\int_{-\infty}^{t} K_1'(t-u)\sigma_u dW_u\right) dt + K_1(0)\sigma_t\, dW_t \\
dR_{2,t} &= \left(\int_{-\infty}^{t} K_2'(t-u)\left(\frac{dS_u}{S_u}\right)^2\right) dt + K_2(0)\left(\frac{dS_t}{S_t}\right)^2 \\
&= \left(K_2(0)\sigma_t^2 + \int_{-\infty}^{t} K_2'(t-u)\sigma_u^2 du\right) dt
\end{aligned}$$

are in general non-Markovian, since for general kernels K_1 and K_2 the integrals in the above drifts are not functions of $(R_{1,t}, R_{2,t})$.

A (too) simple Markovian approximation: the 2-Factor PDV model

- The simplest kernels yielding a Markovian model are the (normalized) exponential kernels $K_1(\tau) := \lambda_1 e^{-\lambda_1 \tau}$ and $K_2(\tau) := \lambda_2 e^{-\lambda_2 \tau}$, $\lambda_1, \lambda_2 > 0$.
- $K_1' = -\lambda_1 K_1$ and $K_2' = -\lambda_2 K_2$ so both $(R_{1,t}, R_{2,t})$ and $(S_t, R_{1,t}, R_{2,t})$ have Markovian dynamics:

$$\frac{dS_t}{S_t} = \sigma(R_{1,t}, R_{2,t})\, dW_t, \qquad \sigma(R_1, R_2) = \beta_0 + \beta_1 R_1 + \beta_2 \sqrt{R_2},$$

$$dR_{1,t} = \lambda_1 \left(\frac{dS_t}{S_t} - R_{1,t}\, dt \right) = \lambda_1 \big(\sigma(R_{1,t}, R_{2,t})\, dW_t - R_{1,t}\, dt \big),$$

$$dR_{2,t} = \lambda_2 \left(\left(\frac{dS_t}{S_t} \right)^2 - R_{2,t}\, dt \right) = \lambda_2 \left(\sigma(R_{1,t}, R_{2,t})^2 - R_{2,t} \right) dt.$$

- We call this model the **2-Factor PDV model** (2FPDV model).

The 2-Factor PDV model

- Choosing K_1 and K_2 to be single exponential kernels fails to capture the mix of short and long memory in both R_1 and R_2 observed in the data.
- We will capture this mix of short and long memory in a Markovian way by choosing K_1 and K_2 to be **linear combinations** of exponential kernels.
- Dynamics of the volatility $\sigma_t = \beta_0 + \beta_1 R_{1,t} + \beta_2 \sqrt{R_{2,t}}$ reads

$$d\sigma_t = \left(-\beta_1 \lambda_1 R_{1,t} + \frac{\beta_2 \lambda_2}{2} \frac{\sigma_t^2 - R_{2,t}}{\sqrt{R_{2,t}}} \right) dt + \beta_1 \lambda_1 \sigma_t \, dW_t. \qquad (2)$$

- **Constant instantaneous vol of instantaneous vol** but rich drift
- **Volatility clustering via mean-reversion** + **explanation for mean-reversion**
- Price-path-dependence of volatility dynamics: **strong Zumbach effect**
- Nonnegativity of volatility guaranteed if $\lambda_2 < 2\lambda_1$

A better Markovian approximation: the 4-Factor PDV model

- Approximate TSPL kernel $\tau \mapsto Z_{\alpha,\delta}^{-1}(\tau+\delta)^{-\alpha}$ by a linear combination of two exponential kernels, $\tau \mapsto (1-\theta)\lambda_0 e^{-\lambda_0 \tau} + \theta\lambda_1 e^{-\lambda_1 \tau}$ with $\theta \in [0,1]$ and $\lambda_0 > \lambda_1 > 0$.
- **Short memory: large** λ_0.
- **Long memory: small** λ_1.
- θ is a mixing factor.

TSPL vs linear combination of two exponentials

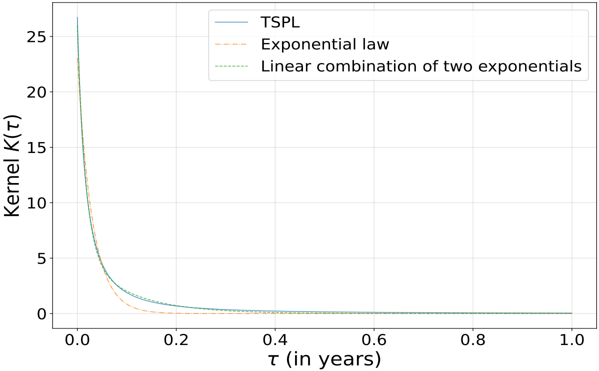


Figure: TSPL kernel K_1 and its approximations by an exponential and by a linear combination of two exponentials.

The 4-Factor PDV model

Introduce parameters $\theta_1, \lambda_{1,0}, \lambda_{1,1}$ and $\theta_2, \lambda_{2,0}, \lambda_{2,1}$ for the approximation of the TSPL kernels K_1 and K_2. For $n \in \{1,2\}$ and $j \in \{0,1\}$, denote

$$R_{n,j,t} := \int_{-\infty}^{t} \lambda_{n,j} e^{-\lambda_{n,j}(t-u)} \left(\frac{dS_u}{S_u}\right)^n.$$

$$\begin{aligned}
\frac{dS_t}{S_t} &= \sigma_t \, dW_t \\
\sigma_t &= \sigma(R_{1,t}, R_{2,t}) \\
\sigma(R_1, R_2) &= \beta_0 + \beta_1 R_1 + \beta_2 \sqrt{R_2} \\
R_{1,t} &= (1-\theta_1) R_{1,0,t} + \theta_1 R_{1,1,t} \\
R_{2,t} &= (1-\theta_2) R_{2,0,t} + \theta_2 R_{2,1,t} \\
dR_{1,j,t} &= \lambda_{1,j} \left(\frac{dS_t}{S_t} - R_{1,j,t}\, dt\right) = \lambda_{1,j} \left(\sigma(R_{1,t}, R_{2,t})\, dW_t - R_{1,j,t}\, dt\right), \\
dR_{2,j,t} &= \lambda_{2,j} \left(\left(\frac{dS_t}{S_t}\right)^2 - R_{2,j,t}\, dt\right) = \lambda_{2,j} \left(\sigma(R_{1,t}, R_{2,t})^2 - R_{2,j,t}\right) dt.
\end{aligned}$$

The 4-Factor PDV model

The dynamics of the instantaneous volatility reads

$$d\sigma_t = \beta_1 \left((1-\theta_1)\lambda_{1,0} + \theta_1\lambda_{1,1}\right)\sigma_t\, dW_t + \Big\{ -\beta_1 \left((1-\theta_1)\lambda_{1,0} R_{1,0,t} + \theta_1\lambda_{1,1} R_{1,1,t}\right) + \frac{\beta_2}{2} \frac{\left((1-\theta_2)\lambda_{2,0} + \theta_2\lambda_{2,1}\right)\sigma_t^2 - \left((1-\theta_2)\lambda_{2,0} R_{2,0,t} + \theta_2\lambda_{2,1} R_{2,1,t}\right)}{\sqrt{R_{2,t}}} \Big\}\, dt \qquad (3)$$

and satisfies similar qualitative properties as dynamics (2):

- The drift of σ_t produces **volatility clustering via a clear trend of mean reversion of volatility**.
- The **lognormal volatility of σ_t is constant**.
- The **dynamics of (σ_t) are price-path-dependent**: the drift of σ_t cannot be written as a function of just the past values $(\sigma_u)_{u\leq t}$ of the volatility; it depends on the past asset returns through $R_{1,0,t}$ and $R_{1,1,t}$.

The 4-Factor PDV model: drift of the volatility

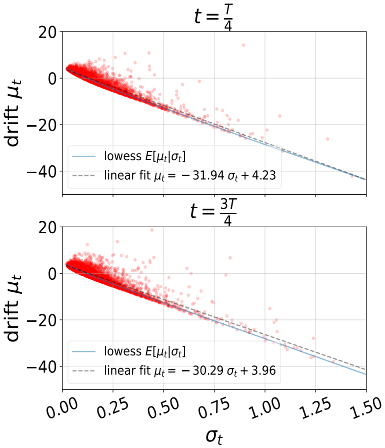


Figure: Drift of σ_t vs σ_t for different maturities and for $N = 10$k paths, $T = 1$ year.

The 4-Factor PDV model: sample paths

β_0	β_1	$\lambda_{1,0}$	$\lambda_{1,1}$	θ_1	β_2	$\lambda_{2,0}$	$\lambda_{2,1}$	θ_2
0.04	-0.105	62	10	0.21	0.6	40	3	0.42

Table: Parameters for the simulation of the 4-Factor PDV Model

The 4-Factor PDV model: sample paths

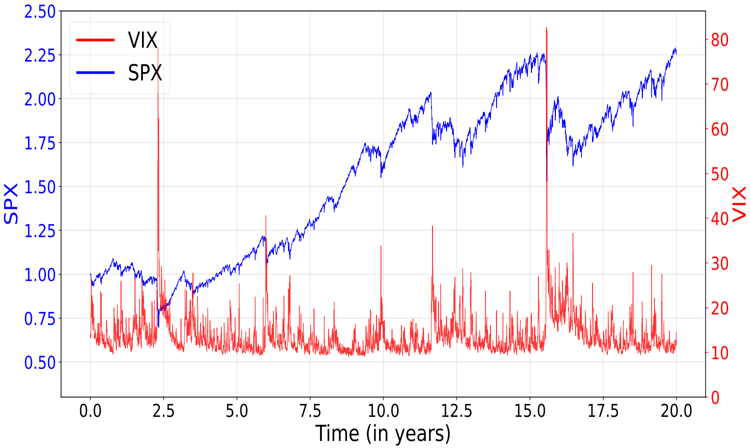


Figure: SPX and VIX timeseries on a typical path of 20 years.

The 4-Factor PDV model: scatter plots

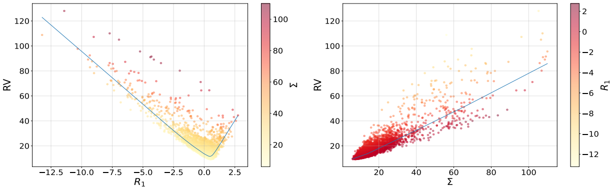


Figure: VIX vs features on 5 simulated paths of 20 years.

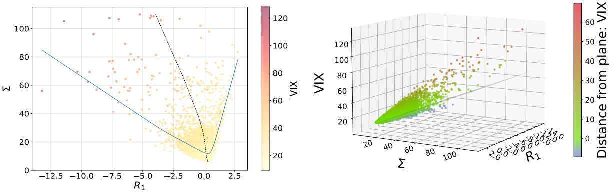


Figure: Σ vs R_1 and 3D scatter plot of VIX vs R_1 and Σ.

The 4-Factor PDV model: scatter plots

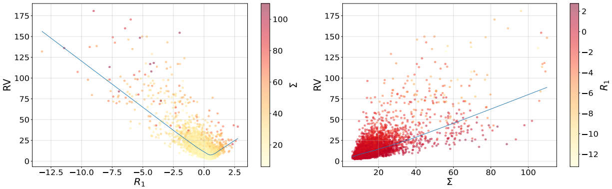


Figure: RV vs features on 5 simulated paths of 20 years.

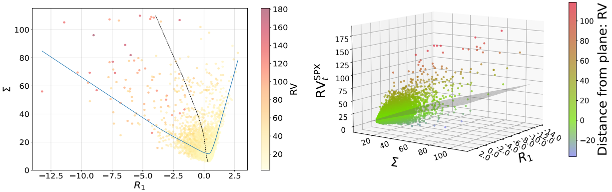


Figure: Σ vs R_1 and 3D scatter plot of RV vs R_1 and Σ.

The 4-Factor PDV model: very realistic smiles

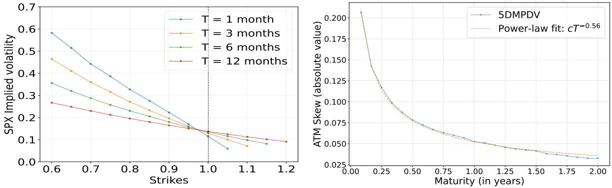


Figure: Model SPX smiles and term-structure of ATM skew.

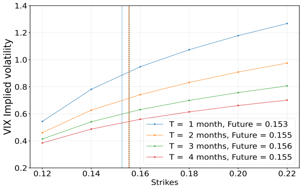


Figure: Model VIX smiles.

Conclusion

- **Volatility is (mostly) path-dependent, endogenous.**
- Volatility is very well explained by recent past asset returns only: train $r^2 \approx \mathbf{0.9}$, test $r^2 \approx \mathbf{0.9}$ on implied volatility data; train $r^2 \approx \mathbf{0.7}$, test $r^2 \approx \mathbf{0.7}$ on (noisy) realized volatility data.
- We have found a simple path-dependent volatility model that accurately explains the current VIX or RV value by recent SPX returns.
- We directly model the volatility level (not the vol changes).
- By design, dependence on trend features (MA of past returns) $\Longrightarrow$ leverage effect + strong Zumbach effect...
- ...but it is not enough: volatility features (MA of past squared returns = historical volatility) are needed too; they capture volatility clustering + weak Zumbach effect.
- Multi-scale trading memory: different time scales of path-dependence are needed $\Longleftrightarrow$ various time horizons of investors/traders
- Using EWMA yields easy-to-simulate Markovian models...
- ...but still generates spurious roughness caused by noisy estimation of RV.

Conclusion

- Volatility is not purely path-dependent: some of it depends on news, new information.
- The (small) exogenous part can then be incorporated using another source of randomness, e.g.,

$$\frac{dS_t}{S_t} = a_t\,\sigma(S_u, u \leq t)\,dW_t$$

  where a_t is some stochastic volatility, for instance: **PDSV**
- The ratio residuals $\frac{\mathrm{VIX}_t}{f(S_u, u \leq t)}$ help define relevant stochastic dynamics for (a_t).

We believe this is the right way of modeling volatility:
(1) Model the purely endogenous part of volatility as best as we can.
(2) Then add the exogenous part, if needed.

Some references

Blanc, P., Donier, B., Bouchaud, J.-P.: *Quadratic Hawkes processes for financial prices*, Quantitative Finance, 17(2):171–188, 2017.

Bergomi, L.: *Smile dynamics II*, Risk, October, 2005.

Bollerslev, T.: *Generalized Autoregressive Conditional Heteroskedasticity*, Journal of Econometrics, 31:307–327, 1986.

Bouchaud, J.-P.: *Radical complexity*, Entropy 23:1676, 2021.

Brunick, G. and Shreve, S.: *Mimicking an Itô process by a solution of a stochastic differential equation*, Ann. Appl. Prob., 23(4):1584–1628, 2013.

Chicheportiche, R., Bouchaud, J.-P.: *The fine-structure of volatility feedback I: Multi-scale self-reflexivity*, Physica A 410:174–195, 2014.

Engle, R.: *Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation*, Econometrica 50:987–1008, 1982.

Figà-Talamanca, G. and Guerra, M.L.: *Fitting prices with a complete model*, J. Bank. Finance 30(1), 247–258, 2006.

Foschi, P, and Pascucci, A.: *Path Dependent Volatility*, Decisions Econ. Finan., 2007.

Guyon, J.: *Path-dependent volatility*, Risk, October 2014.

Some references

- Guyon, J.: *Path-dependent volatility: practical examples*, presentation at Global Derivatives, May 2017.
- Guyon, J.: *Inversion of Convex Ordering in the VIX Market*, Quantitative Finance 20(10):1597–1623, 2020.
- Guyon, J.: *The VIX Future in Bergomi Models: Fast Approximation Formulas and Joint Calibration with S&P 500 Skew*, in revision, 2021.
- Guyon, J., Lekeufack, J.: *Volatility is (mostly) path-dependent*, in preparation, 2022.
- Hobson, D. G. and Rogers, L. C. G.: *Complete models with stochastic volatility*, Mathematical Finance 8(1):27–48, 1998.
- Hubalek, F., Teichmann, J. and Tompkins, R.: *Flexible complete models with stochastic volatility generalising Hobson-Rogers*, working paper, 2004.
- Jacquier, A., Lacombe, C.: *Path-dependent volatility models*, preprint, 2020.
- Gatheral, J., Jusselin, P., Rosenbaum, M.: *The quadratic rough Heston model and the joint calibration problem*, Risk, May 2020.
- Parent, L.: *The EWMA Heston model*, preprint, 2021.
- Sentana, E.: *Quadratic ARCH models*, Rev. Econ. Stud., 62(4):639–661, 1995.
- Zumbach, G.: *Time reversal invariance in finance*, Quantitative Finance 9(5):505–515, 2009.
- Zumbach, G.: *Volatility conditional on price trends*, Quantitative Finance 10(4):431–442, 2010.