Volatility Is (Mostly) Path-Dependent

Jordan Lekeufack

University of California, Berkeley, Department of Statistics

Joint work with Julien Guyon
Ecole des Ponts ParisTech

CDAR
March 07, 2023

jordan.lekeufack@berkeley.edu

Outline

1. Why Path-Dependent Volatility (PDV)?
2. Is Volatility Path-Dependent? How much? How?
3. Continuous-time Markovian extension: the 4-Factor PDV model

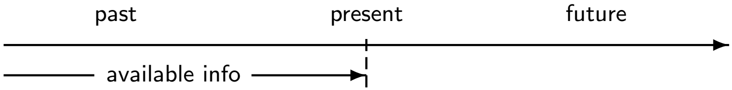
Path-Dependent Volatility

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

- 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



- 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** $\longrightarrow P(t, (S_u, u \leq t))$

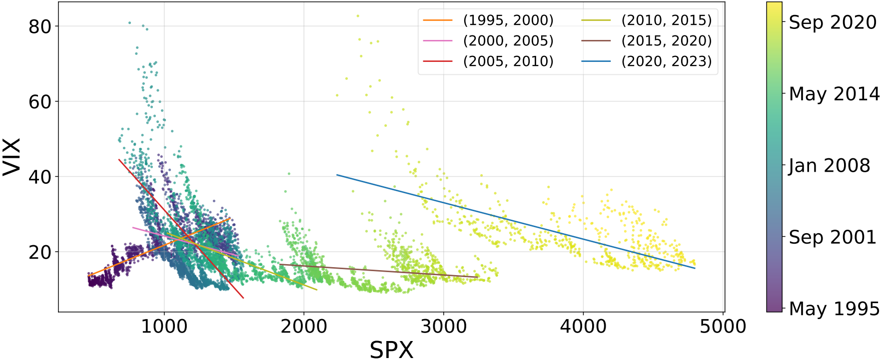
An intuitive argument: a simple quizz

	March 07, 2023		March 07, 2024
SPX	4,000		5,400
VIX			?

An intuitive argument: a simple quizz

	March 07, 2023	February 07, 2024	March 07, 2024
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., Quantitative Finance, '20)
- *The VIX Future in Bergomi Models: Fast Approximation Formulas and Joint Calibration with S&P 500 Skew* (G., SIAM Journal on Financial Mathematics, '22)

Joint calibration of SV models to SPX and VIX smiles

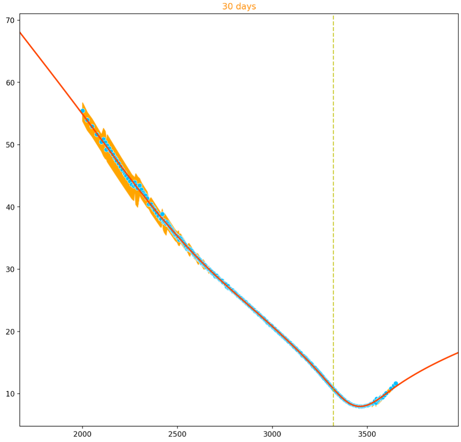


Figure: SPX smile as of January 22, 2020, $T = 30$ days

Joint calibration of SV models to SPX and VIX smiles

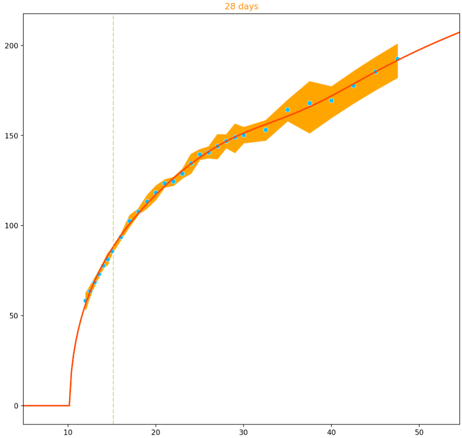


Figure: VIX smile as of January 22, 2020, $T = 28$ days

Joint calibration of SV models to SPX and VIX smiles

- ATM skew:

$$
\begin{array}{rcll}
\text{Definition:} & \mathcal{S}_T & = & \left.\dfrac{d\sigma_{\mathrm{BS}}(K,T)}{\frac{dK}{K}}\right|_{K=F_T} \\
\text{SPX, small } T\text{:} & \mathcal{S}_T & \approx & -1.5 \\
\text{Classical one-factor SV model:} & \mathcal{S}_T & \underset{T\to 0}{\longrightarrow} & \dfrac{1}{2} \times \text{spot-vol correl} \times \text{vol of vol}
\end{array}
$$

- 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)**

$\mathcal{S}_0 \approx -1.5$: see *Does the Term-Structure of Equity At-the-Money Skew Really Follow a Power Law?* (El Amrani, G., '22)

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**.

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.
- Check how the models perform on the test set.
- Training set: 2000–18; 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

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.

[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.

Trend features

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

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

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:

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

with, e.g., $\varphi(r) = r^+$ or r^3 or $(r^-)^2$.

- Another example: spot-to-moving-average ratio

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

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

- 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

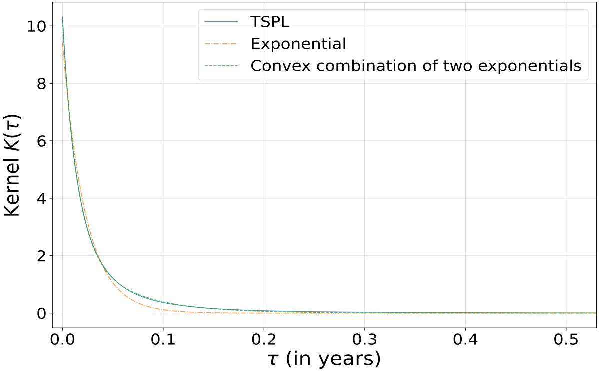


Figure: Typical kernel shapes

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

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

- 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):

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

  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):

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

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

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

- ZHawkes process (BDB '16):

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

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

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

  with Mittag-Leffler kernel K_1.
- Discrete-time version of the threshold EWMA Heston model (Parent '21):

$$\text{Volatility}_t = \beta_0 + \beta_1 (\beta_2 - R_{1,t})_+ \tag{M4}$$

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

Our model differs in several ways

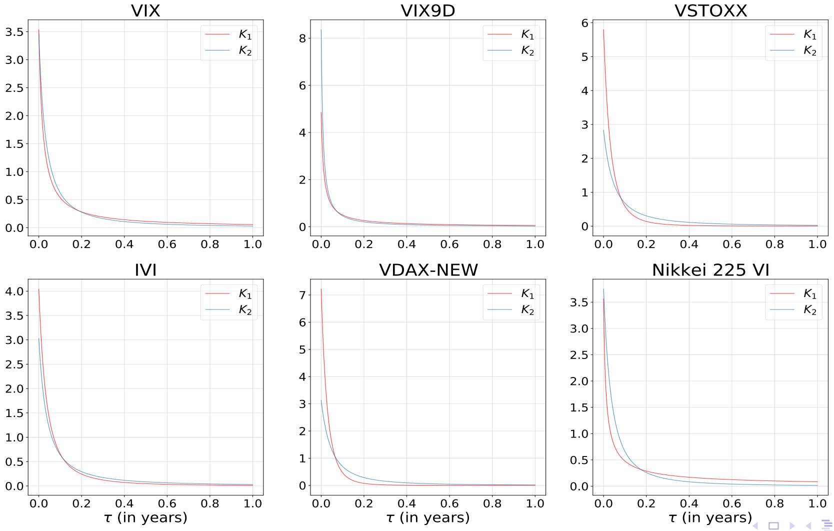
1. Models (M1)-(M3), like almost all ARCH models, model the **square** of the volatility, the variance. Instead, **we directly model the volatility itself**.
2. We use the square root Σ_t of $R_{2,t}$ rather than $R_{2,t}$ itself as one of the linear factors.
3. As a consequence, all the terms in our linear model are **homogeneous to a volatility** (or asset return), whereas (M1) and (M3) mix **heterogeneous** linear factors in volatility and variance (or return and squared return), and all the terms in the linear model (M2) are homogeneous to a variance.
4. We use new, explicit parametric forms for the kernels K_1 and K_2, capturing **non-blowing-up power-law-like decays**.
5. Compared with (M3) and (M4), we empirically prove the **importance of including the historical volatility factor Σ_t**.
6. 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	1.06	0.020	-0.095	1.60	0.052	0.82
VIX9D	0.045	1.00	0.011	-0.12	1.25	0.011	0.88
VSTOXX	0.032	3.96	0.13	-0.036	1.90	0.089	0.97
IVI	0.022	2.26	0.081	-0.058	1.6	0.063	0.99
VDAX-NEW	0.036	5.54	0.16	-0.024	2.21	0.103	0.92
Nikkei 225 VI	0.055	0.78	0.008	-0.069	2.09	0.077	0.86

Table: Optimal parameters of our model for various implied volatility indexes.

Results: Implied volatility



Results: Implied volatility

	Train		Test	
	RMSE	r^2	**RMSE**	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	0.023	0.925	0.030	0.870
VDAX-NEW	0.025	0.934	0.027	0.918
Nikkei 225 VI	0.030	0.890	0.031	0.800

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

Results: Implied volatility

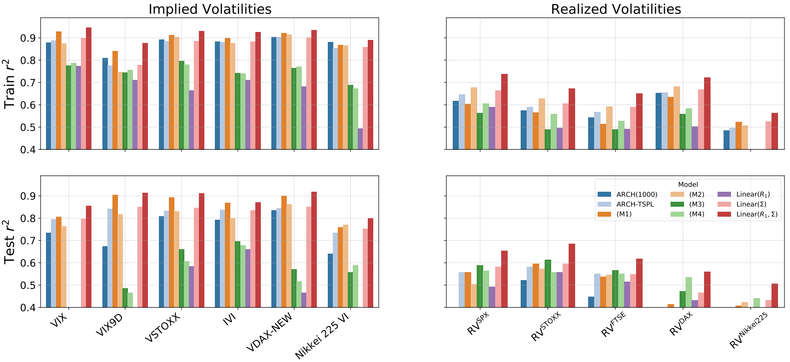


Figure: Comparison of r^2 scores for ARCH benchmarks, models (M1)-(M4), and our linear models. Top: r^2 score on train set. Bottom: r^2 score on test set. Left: Implied volatilities. Right: Realized volatilities.

Results: Implied volatility

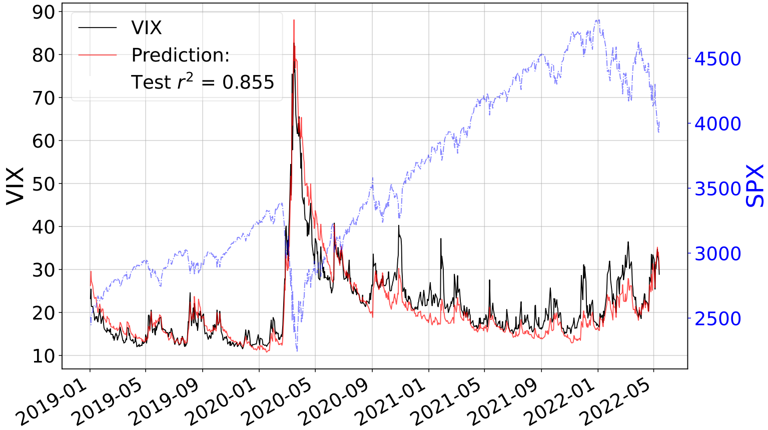


Figure: Time series of VIX and predicted values on test set. The dashed blue line represents the SPX time series.

Results: Implied volatility

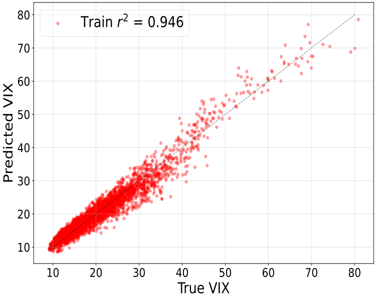


Figure: Predicted VIX vs true VIX on train/test set.

Results: Implied volatility

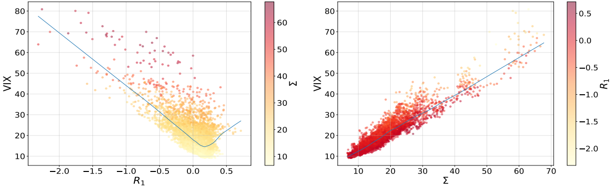


Figure: VIX vs features on the train data set.

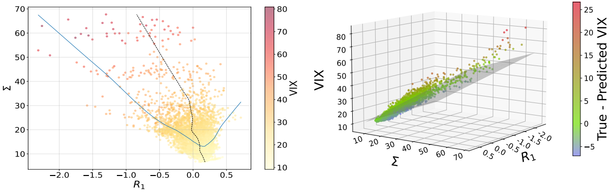


Figure: Σ vs R_1 on the train data set and 3D scatter plot of VIX vs R_1 and Σ

Results: Implied volatility

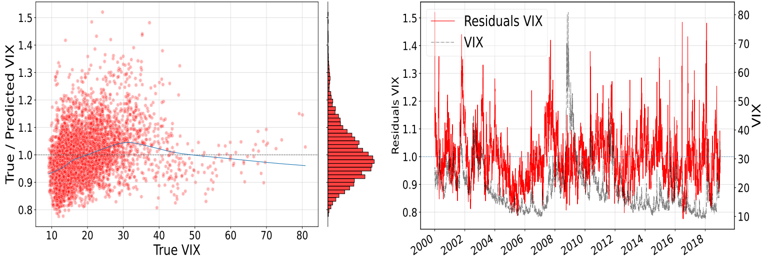
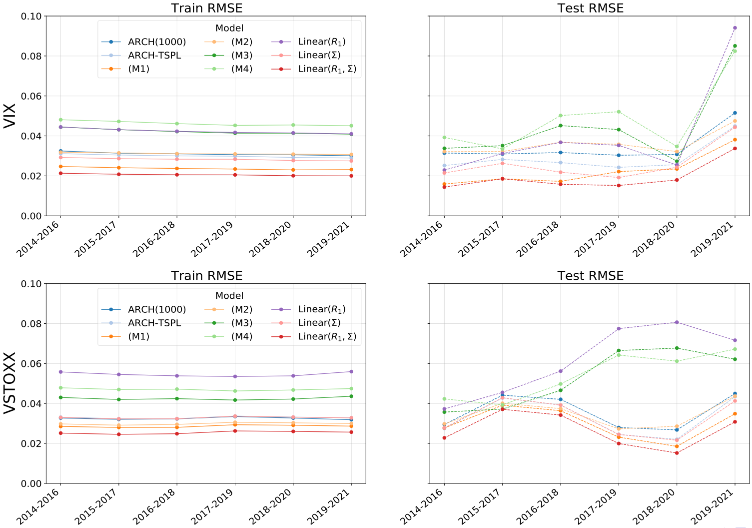
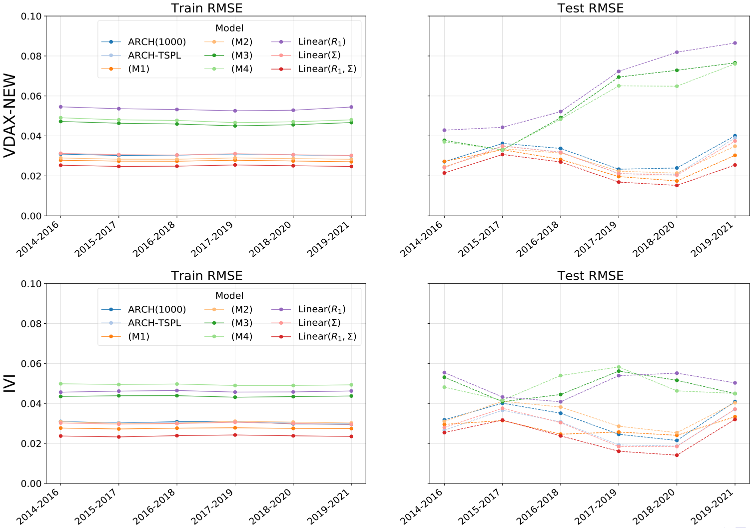


Figure: Residuals plots for VIX predictions.

Varying the test set



Varying the test set

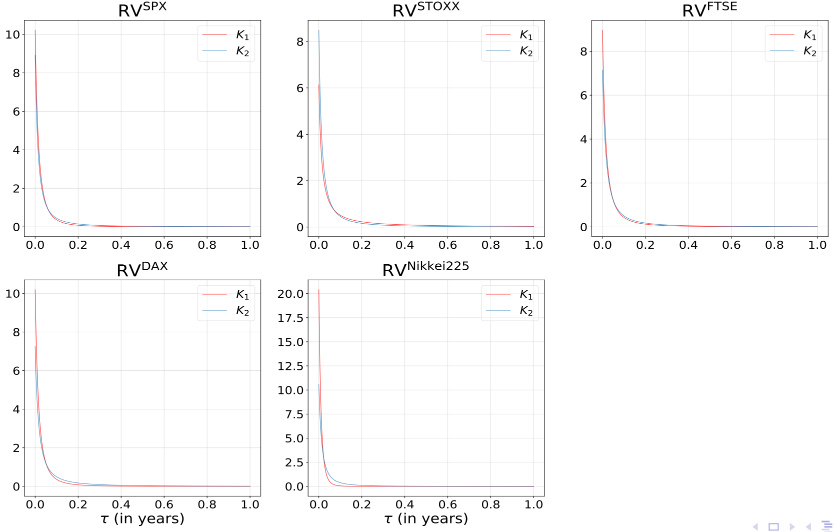


Results: Realized volatility

	β_0	α_1	δ_1	β_1	α_2	δ_2	β_2
SPX	0.018	2.82	0.044	-0.042	1.86	0.025	0.71
STOXX	0.023	1.31	0.017	-0.062	1.79	0.024	0.70
FTSE	0.017	2.22	0.034	-0.043	1.84	0.031	0.76
DAX	0.001	2.87	0.045	-0.030	1.80	0.029	0.81
NIKKEI	0.032	6.30	0.063	-0.011	2.30	0.030	0.51

Table: Optimal parameters of our model for the daily realized volatility of various indexes.

Results: Realized volatility



Results: Realized volatility

	Train		Test	
	RMSE	r^2	RMSE	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: RMSE and r^2 scores for our model for the daily realized volatility of various indexes.

Results: Realized volatility

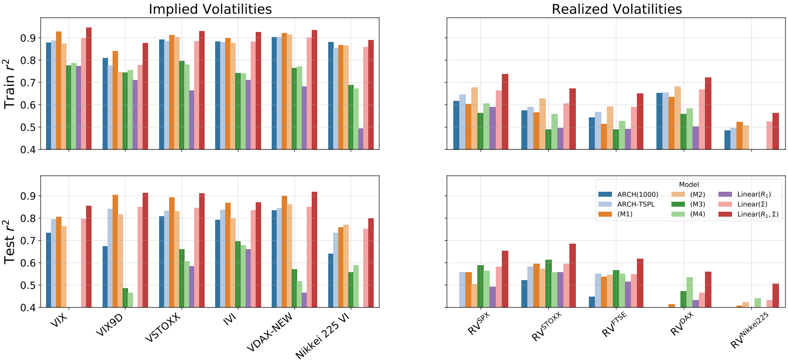


Figure: Comparison of r^2 scores for ARCH benchmarks, models (M1)-(M4), and our linear models. Top: r^2 score on train set. Bottom: r^2 score on test set. Left: Implied volatilities. Right: Realized volatilities.

Results: Realized volatility

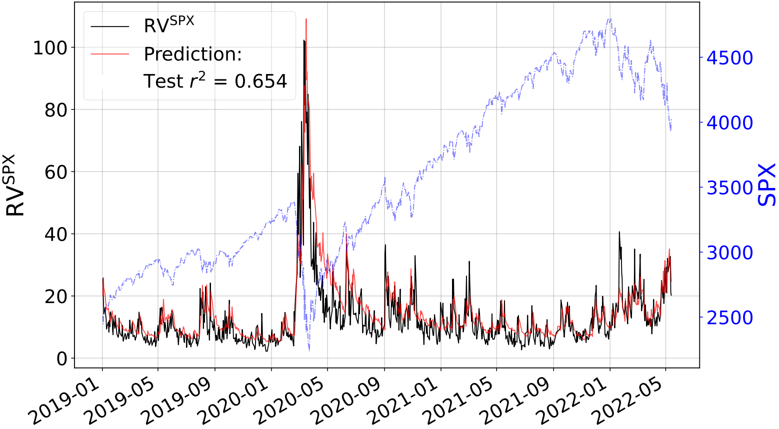


Figure: Time series of VIX and predicted values on test set. The dashed blue line represents the SPX time series.

Results: Realized volatility

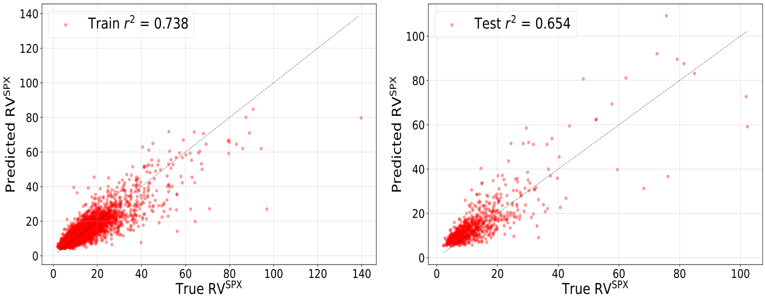


Figure: Predicted VIX vs true VIX on train/test set.

Results: Realized volatility

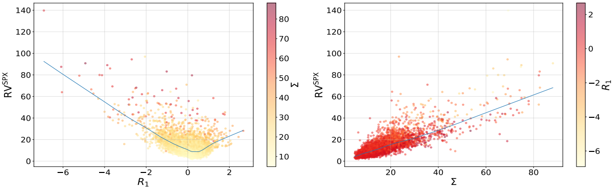


Figure: RV^{SPX} vs features on the train data set.

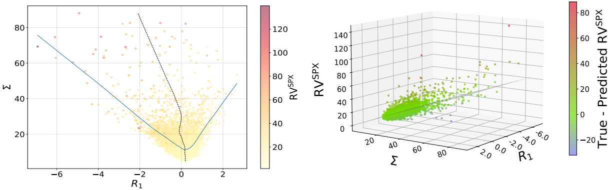


Figure: Σ vs R_1 on the train data set and 3D scatter plot of RV^{SPX} vs R_1 and Σ.

Results: Realized volatility

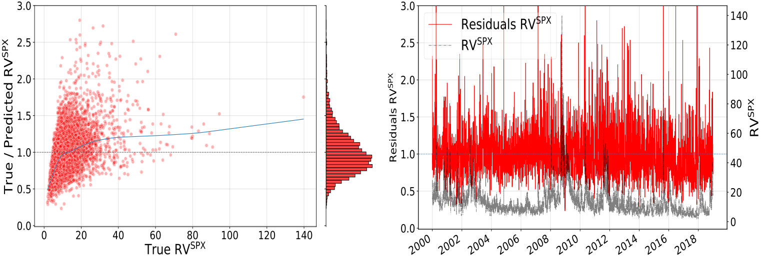


Figure: Residuals plots for RV^{SPX} predictions.

The 4-Factor 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:

$$
\begin{aligned}
\frac{dS_t}{S_t} &= \sigma(R_{1,t}, R_{2,t})\, dW_t, \qquad \sigma(R_1, R_2) = \beta_0 + \beta_1 R_1 + \beta_2 \sqrt{R_2}, \\
dR_{1,t} &= \lambda_1 \left(\frac{dS_t}{S_t} - R_{1,t}\, dt \right) = \lambda_1 \big(\sigma(R_{1,t}, R_{2,t})\, dW_t - R_{1,t}\, dt \big), \\
dR_{2,t} &= \lambda_2 \left(\left(\frac{dS_t}{S_t} \right)^2 - R_{2,t}\, dt \right) = \lambda_2 \left(\sigma(R_{1,t}, R_{2,t})^2 - R_{2,t} \right) dt.
\end{aligned}
$$

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

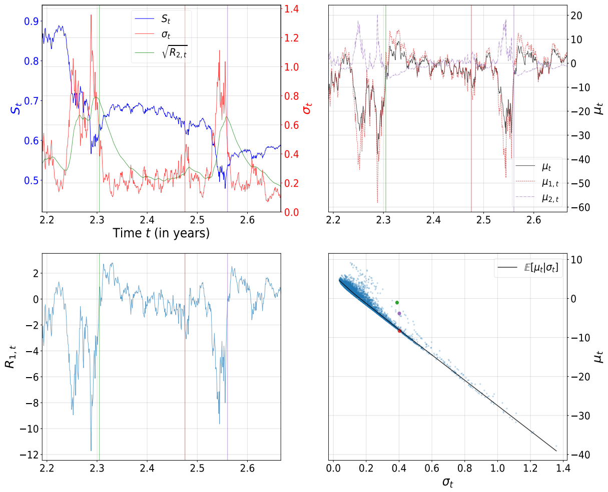
The 2-Factor PDV model

- 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:
- Nonnegativity of volatility guaranteed if $\lambda_2 < 2\lambda_1$.

Drift of the instantaneous volatility in the 2-factor PDV model



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

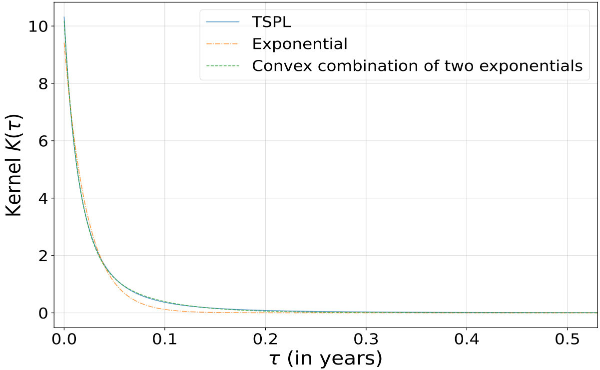


Figure: TSPL kernel K_1 and its approximations by an exponential and by a linear combination of two exponentials.

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

$$d\sigma_t = \beta_1 \bar{\lambda}_1 \sigma_t \, dW_t + \left(-\beta_1 \bar{\lambda}_1 \bar{R}_{1,t} + \frac{\beta_2 \bar{\lambda}_2}{2} \frac{\sigma_t^2 - \bar{R}_{2,t}}{\sqrt{R_{2,t}}} \right) dt$$

where $\lambda_{n,t} = (1-\theta_n)\lambda_{n,0} + \theta_n \lambda_{n,1},$

$$\bar{R}_{n,t} = \frac{((1-\theta_n)\lambda_{n,0} R_{n,0,t} + \theta_n \lambda_{n,1} R_{n,1,t})}{\bar{\lambda}_n} \quad , n \in \{1, 2\}$$

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 $(\sigma_u)_{u \le t}$ 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

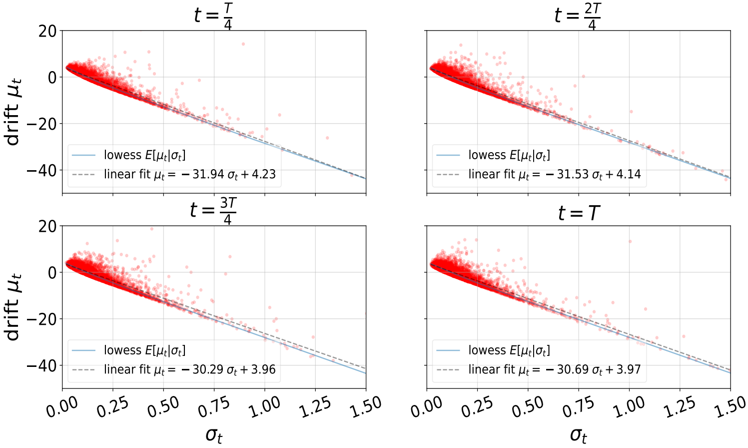


Figure: Drift of σ_t vs σ_t for different maturities and for $N = 10$k paths, $T = 1$ year.

The 4-Factor PDV model: sample paths

	β_0	β_1	$\beta_{1,2}$	$\lambda_{1,0}$	$\lambda_{1,1}$	θ_1	β_2	$\lambda_{2,0}$	$\lambda_{2,1}$	θ_2
Realistic path	0.04	-0.11	–	55	10	0.25	0.65	20	3	0.5
Implied	0.006	-0.157	0.078	70	30.5	0.21	0.683	10.6	5.2	0.7

Table: Different sets of parameters of the 4-factor Markovian PDV Model.

The 4-Factor PDV model: sample paths

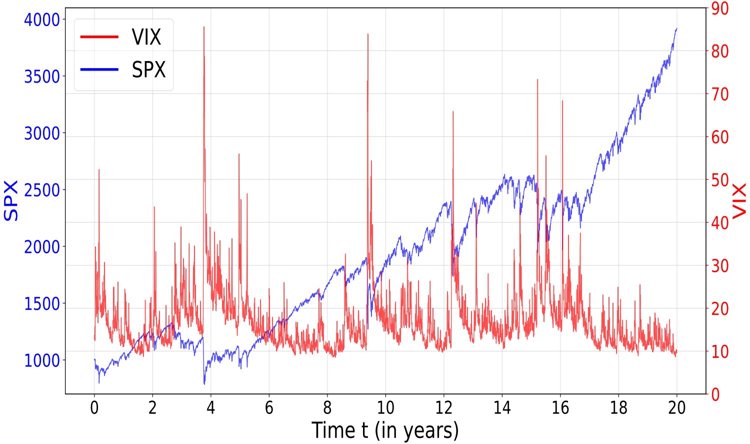
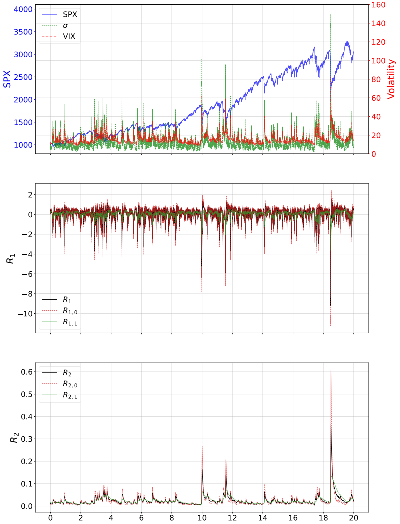


Figure: SPX and VIX time series on a typical path of 20 years.

The 4-Factor PDV model: sample paths



The 4-Factor PDV model: scatter plots

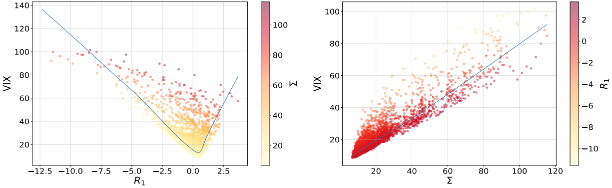


Figure: VIX vs features on 5 simulated paths of 20 years.

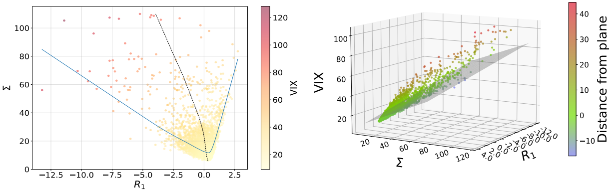


Figure: Σ vs R_1 and 3D scatter plot of VIX vs R_1 and Σ.

The 4-Factor PDV model: scatter plots

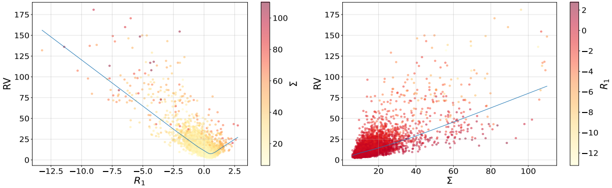


Figure: RV vs features on 5 simulated paths of 20 years.

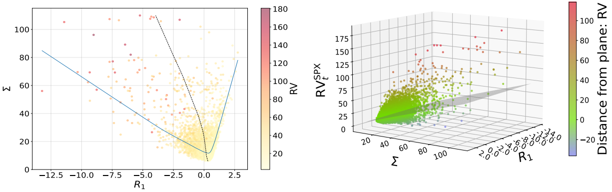


Figure: Σ vs R_1 and 3D scatter plot of RV vs R_1 and Σ.

The 4-Factor PDV model: very realistic smiles

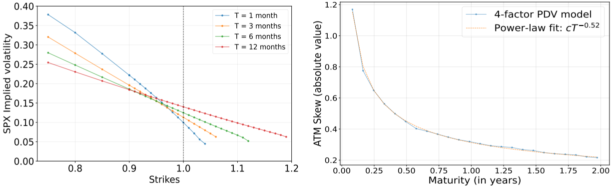


Figure: Model SPX smiles and term-structure of ATM skew.

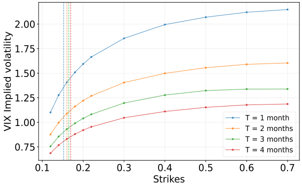


Figure: Model VIX smiles.

The 4-Factor PDV model: joint SPX/VIX calibration

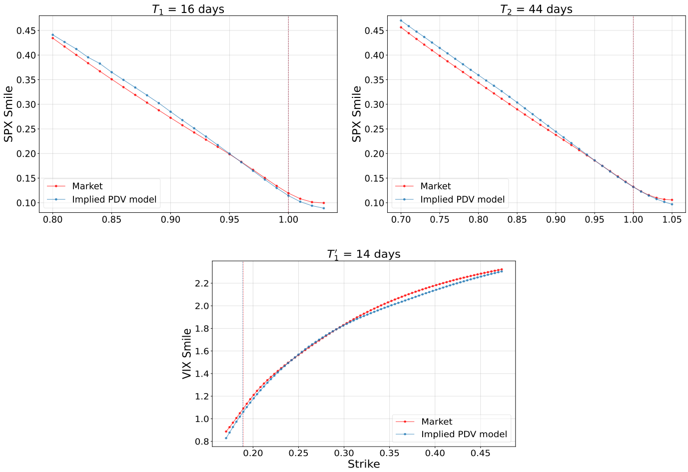


Figure: Calibration as of June 2, 2021.

The 4-Factor PDV model: joint SPX/VIX calibration

	β_0	β_1	$\beta_{1,2}$	$\lambda_{1,0}$	$\lambda_{1,1}$	θ_1	β_2	$\lambda_{2,0}$	$\lambda_{2,1}$	θ_2
Realistic path	0.04	-0.11	–	55	10	0.25	0.65	20	3	0.5
Implied	0.006	-0.157	0.078	70	30.5	0.21	0.683	10.6	5.2	0.7

Table: Different sets of parameters of the 4-factor Markovian PDV Model

Conclusion

- **Volatility is (mostly) path-dependent, endogenous**: it is very well explained by recent past asset returns only.
- A very simple path-dependent volatility model accurately explains the current VIX or RV value by recent SPX returns:

$$\text{Volatility}_t = \beta_0 + \beta_1 R_{1,t} + \beta_2 \sqrt{R_{2,t}}, \qquad \beta_0 > 0, \;\; \beta_1 < 0, \;\; \beta_2 \in (0,1)$$

- It mixes recent past trend and recent past volatility.
- Multi-scale trading memory: different time scales of path-dependence are needed $\iff$ various time horizons of investors/traders

Conclusion

- **The 4-Factor PDV model is the natural Markovian continuous-time version**: 2 times scales for recording the past trend + 2 time scales for recording the past volatility.
- The model is Markovian, so extremely easy to simulate, but also (1) rough-like (the kernels K_1, K_2 do not blow up at zero maturity but look like rough kernels),
  (2) generates some spurious roughness caused by noisy estimation of RV.
- Unlike in classical SV models, the 4 factors are observable: averages of past returns and past squared returns, as opposed to averages of past dW.
- The 9 parameters all have a clear financial interpretation.
- The 4-Factor PDV model seems to be the first parametric Markovian model to practically solve the joint calibration problem.
- The model exhibits a jumpy behavior in the absence of actual jumps.

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.,

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

  where a_t is some stochastic volatility, for instance: **PDSV**
- The ratio residuals $\frac{\mathrm{VIX}_t}{f(S_u, u \le t)}$ 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*, SIAM Journal on Financial Mathematics 13(4):1418–1485, 2022.
- **Guyon, J., Lekeufack, J.: Volatility is (mostly) path-dependent, SSRN 4174589, 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*, Quantitative Finance 23(1):71–93, 2023.
- 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.