

# VOLATILITY IS (MOSTLY) PATH-DEPENDENT

JULIEN GUYON

*Bloomberg L.P., Quantitative Research*

JORDAN LEKEUFACK

*University of California, Berkeley, Department of Statistics  
Bloomberg L.P., Quantitative Research*

**ABSTRACT.** We learn from data that volatility is mostly path-dependent: up to 90% of the variance of the implied volatility of equity indexes is explained endogenously by past index returns, and up to 65% for (noisy estimates of) future daily realized volatility. The path-dependency that we uncover is remarkably simple: a linear combination of a weighted sum of past daily returns and the square root of a weighted sum of past daily squared returns with different time-shifted power-law weights capturing both short and long memory. This simple model, which is homogeneous in volatility, is shown to consistently outperform existing models across equity indexes and train/test sets for both implied and realized volatility. It suggests a simple continuous-time path-dependent volatility (PDV) model that may be fed historical or risk-neutral parameters. The weights can be approximated by superpositions of exponential kernels to produce Markovian models. In particular, we propose a 4-factor Markovian PDV model which captures all the important stylized facts of volatility, produces very realistic price and volatility paths, and jointly fits SPX and VIX smiles remarkably well. We thus show, for the first time, that a continuous-time Markovian parametric stochastic volatility (actually, PDV) model can practically solve the joint SPX/VIX smile calibration problem. This article is dedicated to the memory of Peter Carr whose works on volatility modeling have been so inspiring to us.

**Keywords.** Volatility modeling, path-dependent volatility, endogeneity, empirical PDV model, 4-factor Markovian PDV model, joint S&P 500/VIX smile calibration, stochastic volatility

## 1. INTRODUCTION

It is crucial for the pricing, hedging, and risk-management of portfolios of derivatives to understand and correctly capture the joint dynamics of the underlying assets and their implied volatilities. For example pricing and hedging portfolios of S&P 500 (SPX) derivatives requires to precisely understand the joint dynamics of the SPX and VIX indices. It is also very useful to build accurate predictors of future realized volatility. Such forecasts are used for instance in risk management, derivative hedging, market making, market timing, and portfolio selection. While stochastic volatility (SV) models view volatility as an extra stochastic process driven by its own sources of randomness (possibly correlated with those driving the dynamics of the asset price), it has been recently observed [33, 34, 11, 21, 4] that financial markets exhibit a clear pattern of *path-dependent volatility* (PDV): the implied volatility and future realized volatility depend on the path followed by the asset price in the recent past. In some sense, while SV feeds the volatility level in the asset returns, PDV does quite the opposite: it feeds returns into volatility. [21] focuses on the implied volatility, while in [33, 34, 11, 4], the focus is on the future realized volatility, and PDV is expressed as “feedback from the return to the volatility” and shown as an example of time reversal asymmetry in finance. The joint calibration of traditional flexible parametric SV models, such as the 2-factor Bergomi model [2], to at-the-money SPX skews and VIX implied volatilities also points to PDV models, by automatically selecting fully correlated Brownian

---

*E-mail addresses:* jguyon2@bloomberg.net, julien.guyon@polytechnique.org, jordan.lekeufack@berkeley.edu.

*Date:* July 27, 2022.

This work was initiated in 2020 while Jordan Lekeufack was an intern and then a contractor with the Quantitative Research group at Bloomberg L.P., and finished while he was a Bloomberg Quantitative Finance Ph.D. fellow and intern in 2022.

motions [24]. In Section 2 we provide extra philosophical, intuitive, theoretical, and empirical arguments supporting the use and relevance of PDV models.

In this article we aim to learn precisely how much of volatility is path-dependent, and how it depends on past asset returns. We thus aim to empirically explain volatility as an *endogenous* factor as best as we can. Our empirical study covers (a) learning implied volatility from past asset returns (e.g., learn the VIX from the SPX path), and (b) learning future realized volatility from past asset returns. We prove that volatility is mostly path-dependent. A surprisingly simple linear PDV model explains up to 90% of the variance of the implied volatility (depending on the equity index), and up to 65% of the variance of the (more noisy) future daily realized volatility, even on our very challenging test set (2019–2022). This proves that volatility is mostly endogenous. The *explicit* path-dependency that we uncover is remarkably simple: a linear combination of a weighted sum  $R_1$  of past daily returns and the square root of a weighted sum  $R_2$  of past daily *squared* returns with different time-shifted power-law (TSPL) weights.

We thus uncover a *historical PDV* or *empirical PDV* or  $\mathbb{P}$ -*PDV* model. Its continuous-time version is a diffusion process that can also be fed risk-neutral parameters and turned into an *implied PDV* or *risk-neutral PDV* or  $\mathbb{Q}$ -*PDV* model if needed to calibrate to option prices. For practical engineering purposes, our historical PDV model can easily be made Markovian by approximating the TSPL kernels by superpositions of two or more exponential kernels; mixing two exponential kernels is usually enough (see Figure 4.3). Alternatively, one can directly conduct the empirical study with combinations of exponential kernels rather than TSPL kernels (see Appendix A). In particular, we propose a 4-factor Markovian PDV model which captures all the important stylized facts of volatility, produces remarkably realistic price and volatility paths, and produces remarkably realistic SPX and VIX smiles as well. In particular the model captures volatility clustering as a result of the combined mean reversion of the *observable* factors  $R_1$  and  $R_2$ ; the leverage effect and strong negative SPX skews; the weak and strong Zumbach effects; “rough-like” (but actually not rough) volatility paths; some spurious roughness in daily realized volatility; and strong positive VIX skews despite a constant (lognormal) instantaneous volatility of instantaneous volatility. Remarkably, our model can even jointly calibrate to SPX and VIX smiles (as well as VIX futures) with a very good accuracy (see Figure 4.8). To the best of our knowledge, it is the first time that a *Markovian parametric SV* (here, PDV) model is shown to jointly fit SPX and VIX smiles so accurately, i.e., to practically solve the joint SPX/VIX smile calibration problem.

An SV component may finally be added to the model if needed to account for the small remaining exogenous part of volatility, resulting in a Path-Dependent Stochastic Volatility (PDSV) model. A statistical study of the residuals of the empirical PDV model suggests a simple form for the extra SV component. Given the strong endogeneity of volatility, we believe that this is how volatility modeling should be approached: first explain volatility *explicitly* in an *endogenous* way as best as we can, then model the remaining (smaller) exogenous part.

The remainder of this article is structured as follows. Section 2 introduces and motivates PDV. In Section 3 we learn PDV from empirical data. In Section 4 we investigate the continuous-time limit of our empirical PDV model, present the 4-factor Markovian PDV model, analyze some of its properties, compare it with other models, and show that it practically solves the joint SPX/VIX smile calibration problem. Section 5 studies the residuals and suggests a simple PDSV model. Finally Section 6 concludes.

## 2. PATH-DEPENDENT VOLATILITY

PDV models are models in which the volatility of an asset is a deterministic function of the path followed by the asset prices so far. In continuous time, assuming zero interest rates, repos, and dividends for simplicity, they read

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

where  $W$  is a standard Brownian motion. While the volatility  $\sigma$  drives the dynamics of the asset price  $S$ , there is a feedback loop from past prices to volatility. PDV models are thus pure feedback models in which the volatility is fully *endogenous*. Continuous-time PDV models have been considered for derivatives pricing and hedging [29, 21]. In particular Guyon [21] has shown that they combine benefits from Dupire’s local volatility (LV) model [15] and SV models: like the LV model, they are complete models that can be exactly calibrated to the whole surface of implied volatilities, but like SV models they generate much richer spot-vol dynamics than the LV model. While continuous-time PDV models have so far received much less

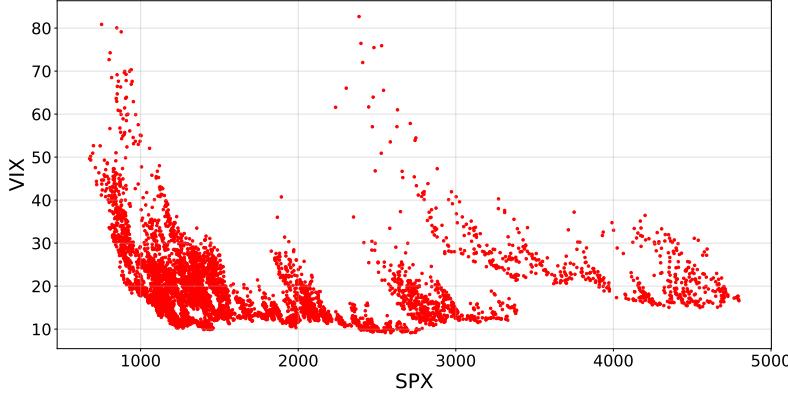


FIGURE 2.1. VIX vs SPX, January 1995–May 2022

attention than the LV or SV models, a recent shift has been observed: five recent volatility modeling papers [4, 20, 24, 14, 30] all end up suggesting continuous-time PDV (or PDSV) models. In discrete time the most famous examples are probably the ARCH [18] and GARCH [6] econometric models and their many descendants. Recent econophysics articles [33, 34, 11, 4] investigate PDV (termed as “volatility feedback effect”) using ARCH-type models, in particular QARCH [31] and FIGARCH [1] models. Our empirical model also combines features from QARCH (for the leverage effect) and FIGARCH (for the long memory of volatility), but with some important differences that will be discussed in Section 3.2.

In the rest of this section we motivate PDV from an array of different points of view: philosophical, intuitive, financial, theoretical, and empirical.

**2.1. A philosophical argument.** Time goes only in one direction, and most natural and social phenomena show hysteresis, i.e., are path-dependent: there is in general absolutely no reason why the future should depend on the past only through the present. This Markovian assumption usually is not a fundamental property; it is often made just for simplicity and ease of computation. For instance, one may assume for simplicity that the price of an option depends only on current time  $t$  and current asset price  $S_t$ :  $P(t, S_t)$ . In fact, most of the time, the present does not capture all the information from the past, and it is much more reasonable to assume that the option price  $P(t, (S_u)_{u \leq t})$  depends on all past asset prices. For instance, if the asset price has recently fallen quickly, the asset volatility is likely high, which impacts the option price.

**2.2. An intuitive argument.** A simple prediction exercise best explains path-dependent volatility. Assume that the SPX value is currently 4,000. What is your best guess of the VIX value in one year if the SPX is worth 5,000 at that time? That would mean that the SPX gained 25% in one year, and due to the leverage effect (the negative link between an asset price and its volatility), a natural guess is a low VIX value, say, 12. Now, if we add the information that two weeks before the one year horizon, the SPX was worth 5,500, would you update your best guess for the VIX? You certainly would, as this means that the market crashed, with the SPX losing 9% of its value in just two weeks, in which case the VIX always shoots up, as SPX puts are more in demand and get more expensive out of fear of a deeper market crash; a best guess could be around 40 for example. This immediately shows that volatility is best explained by the *path* of the asset price, rather than by its current value; or, stated otherwise, by PDV models rather than by the LV model.<sup>1</sup> Figure 2.1 confirms that volatility is not well explained by the asset price.

**2.3. A financial and scaling argument.** In volatility modeling, the two basic quantities that possess a natural scale are the *volatility levels* and the *asset returns*. Therefore a good model should relate these two quantities. PDV models do just that, by explaining the current volatility level by past asset returns. Asset prices clearly do not have a natural scale; they can be one cent or one thousand dollars. But whatever the magnitude of the asset price, the magnitude of asset returns stays the same, say one or a few percent a day. This is precisely due to the fact that volatility stays in its constant, natural range of values regardless of the

<sup>1</sup>It is well known that, while the LV model is the simplest Markovian model consistent with the market prices of vanilla options, it does not generate realistic joint spot/vol dynamics.

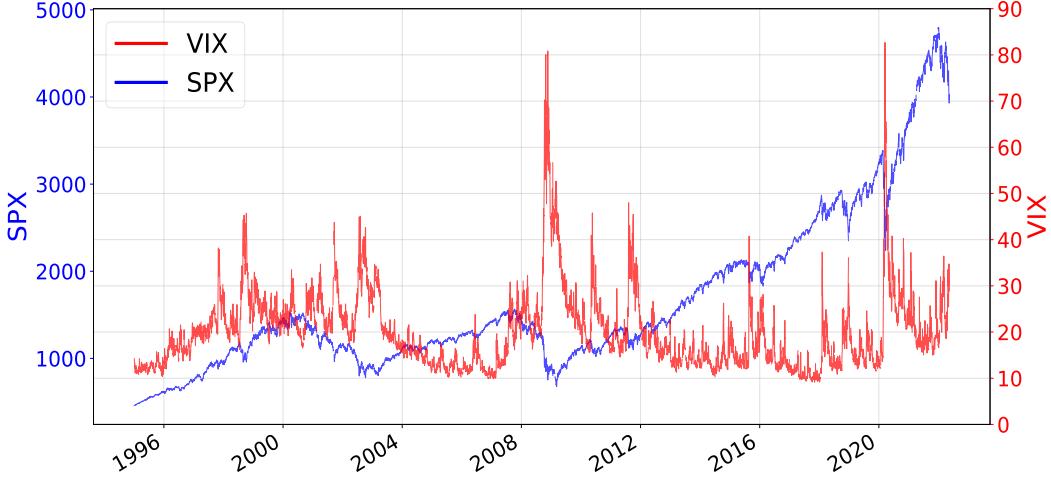


FIGURE 2.2. SPX and VIX time series, Jan 1, 1995–May 15, 2022

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

TABLE 1. Comparison of the LV, SV, and PDV paradigms

asset price. For instance over the past 20 years the VIX index has been “stochastically constant,” staying roughly in the [10, 80] corridor regardless of the SPX value (see Figure 2.2). The VIX spikes are precisely explained by recent large negative SPX returns, irrespective of the value of the SPX index itself.

Let us compare the PDV paradigm with the LV and SV approaches. The comparison is summarized in Table 1. On the one hand, the LV model explains the *volatility level* by the *asset level*, which does not make much financial sense, as we have explained in the previous section. By accurately capturing the spot-vol dynamics observed in the market, well chosen PDV models need not be recalibrated as often as the LV model. On the other hand, SV models relate the *volatility change* (or return) to the simultaneous *asset return* by correlating the Brownian motions that drive the asset price and volatility dynamics (or simply ignore any feedback effect from asset returns to volatility in the case of zero spot-vol correlation). While this approach makes more sense than LV by using asset returns rather than asset levels as explanatory variables, it may not be the most convenient one, as it leads to a path-dependency that is complicated, is not explicitly chosen, and may not reflect the feedback effect of prices on volatility as accurately as an explicitly chosen PDV can, as we now explain.

**2.4. PDV versus SV.** A typical one-factor SV model reads as follows

$$\begin{aligned} \frac{dS_t}{S_t} &= \sigma_t dW_t, \quad \sigma_t = f(t, Y_t) \\ dY_t &= \mu(t, Y_t) dt + \nu(t, Y_t) \left( \rho dW_t + \sqrt{1 - \rho^2} dW_t^\perp \right) \end{aligned}$$

where  $W$  and  $W^\perp$  are two independent Brownian motions,  $\rho \in [-1, 1]$  is the “spot-vol” correlation, and  $f$  is a positive function. The instantaneous volatility reads as  $\sigma_t = f(t, Y_t)$  where

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

Three cases must be distinguished depending on the value of  $\rho$ :

- $\rho = 0$ : the *SV model is strictly path-independent*. The asset price is a slave process with absolutely no feedback on volatility. The instantaneous volatility is a function of the Brownian path  $(W_u^\perp)_{0 \leq u \leq t}$ .

only:

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

- $\rho \notin \{-1, 0, 1\}$ : the SV model is partially path-dependent. There is partial feedback from asset returns to volatility through the “spot-vol” correlation. The instantaneous volatility is a function of both the past asset prices  $(S_u)_{0 \leq u \leq t}$  and the independent random shocks  $(W_u^\perp)_{0 \leq u \leq t}$ :

$$(2.1) \quad \sigma_t = \varphi \left( t, \left( \frac{dS_u}{S_u} \right)_{0 \leq u \leq t}, (dW_u^\perp)_{0 \leq u \leq t} \right) = \psi \left( t, (S_u)_{0 \leq u \leq t}, (W_u^\perp)_{0 \leq u \leq t} \right).$$

- $\rho \in \{-1, 1\}$ : the SV model is fully path-dependent. There is pure feedback from asset returns to volatility:

$$(2.2) \quad \sigma_t = \varphi \left( t, \left( \frac{dS_u}{S_u} \right)_{0 \leq u \leq t} \right) = \psi(t, (S_u)_{0 \leq u \leq t}).$$

SV models that are driven by a single Brownian motion are thus PDV models. Examples include the “complete models with stochastic volatility” of Hobson and Rogers [29], the path-dependent 2-factor Bergomi model of Guyon [24], the quadratic rough Heston model (QRHM) of Gatheral, Jusselin and Rosenbaum [20], and one version of the EWMA Heston model of Parent [30]. Those last three papers are very recent and show that PDV models are gaining much traction.

However, the above functions  $\varphi, \psi$  are complicated and implicit. The path-dependency produced by SV models when  $\rho \neq 0$  is not easily described nor explicitly chosen. By contrast, PDV models aim to explain volatility in an endogenous way *as explicitly as possible*: one directly learns or chooses the PDV functions  $\varphi$  or  $\psi$ . Our approach consists of (1) learning the pure PDV function  $\varphi$  in (2.2) explicitly from market data, and (2) possibly enhance the pure PDV model into a PDSV model as in (2.1), by including independent random shocks in the volatility dynamics, as indicated by a statistical study of the residuals of the PDV learning procedure, to account for the (smaller) exogenous part of the volatility dynamics. By directly choosing the PDV function  $\varphi$ , we easily and naturally capture rapid, very large changes in volatility and the path-dependent nature of the volatility of volatility (“vol of vol”), also known as “strong Zumbach effect” (see Section 2.8), while maintaining volatility in its natural range. It is, in our opinion, a more natural and direct way of modeling volatility.

For tractability and ease of computation and simulation, PDV models can be engineered to be Markovian in small dimension, exactly like traditional SV models are; see [29] and Sections 4.1 and 4.2. They are then amenable to PDE methods, can be quickly simulated, and reduce to SV models driven by one single Brownian motion, but in that case the SV models are not merely postulated: in our case their particular form will result from a solid empirical PDV foundation. In particular Markovian PDV models offer a natural *explanation* for the mean reversion of volatility (see Sections 4.1 and 4.2), whereas SV models simply *postulate* it.

**2.5. Joint calibration of SV models to SPX and VIX smiles.** The joint calibration of traditional parametric SV models to SPX and VIX smiles leads to PDV models. Indeed, when such SV models are calibrated to VIX implied volatilities, in order to fit the very large negative skew of short-term SPX options as best as possible, one must use extremal  $-1$  spot-vol correlation(s). This means that the calibrated SV model actually degenerates into a PDV model.

For instance, Guyon [24] considers the 2-factor Bergomi model. Although the model is one the most flexible among popular SV models, Guyon reports that the joint calibration of the model to the term-structures of SPX at-the-money (ATM) skew and VIX future-implied VIX<sup>2</sup> volatility selects  $-1$  spot-vol correlations, together with very large vol of vol and fast mean reversion (two to three times larger than those reported in [2, 3]), thus producing a (Markovian) pure PDV model with rough-like paths:

$$\begin{aligned} \frac{dS_t}{S_t} &= \sqrt{\xi_t^u} dW_t, \quad \xi_t^T = \xi_0^T f^T(t, X_t^1, X_t^2), \quad T \geq t, \\ X_t^i &= - \int_0^t e^{-\lambda_i(t-u)} dW_u = - \int_0^t e^{-\lambda_i(t-u)} \frac{1}{\sqrt{\xi_u^u}} \frac{dS_u}{S_u}. \end{aligned}$$

In this fully correlated case, the Ornstein-Uhlenbeck processes  $X_t^1$  and  $X_t^2$  are path-dependent variables: they depend only on past asset returns (but in a very complicated way). Using the exponential convolution

kernel  $e^{-k(t-u)}$  guarantees that  $(X_t^1)$  and  $(X_t^2)$ , hence the full vector  $(S_t, X_t^1, X_t^2)$  are Markovian processes, as  $dX_t^i = -\lambda_i X_t^i dt - dW_t$ .

**2.6. An information-theoretical/financial economics argument.** Contrary to SV models, PDV models do not require adding extra sources of randomness (usually Brownian motions) to generate rich spot-vol dynamics: they explain volatility in a purely *endogenous* way. As a result PDV models, unlike SV models, are complete models in which derivatives have a unique, unambiguous price, independent of any preferences or utility functions.<sup>2</sup> In PDV models, 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. While reality is a bit more complex, 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.

**2.7. Path-dependent volatility is generic for option pricing.** In fact, PDV models are 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. Indeed, Brunick and Shreve [10, Corollary 3.11] have proved that, 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

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

In particular, 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.

**2.8. Empirical evidence.** Empirical evidence of path-dependency in the volatility has already been reported in several works, including much of the GARCH literature and [33, 34, 11, 21, 4, 22, 30]. The observed level of the so-called “weak Zumbach effect” [33], which states that “past large-scale realized volatilities are more correlated with future small-scale realized volatilities than vice versa” [4], is most easily captured by PDV models, even though some weak Zumbach effect can also be produced by SV models via nonzero spot-vol correlation [17, 30]. The so-called “strong Zumbach effect”, which states that “conditional dynamics of volatility with respect to the past depend not only on past volatility trajectory but also on the historical price path” [20], can be rephrased as: “there is some price-path-dependency in the volatility dynamics”. The study of those effects was initially motivated by that of time reversal asymmetry in finance [33, 34], which goes beyond the obvious and well-known leverage effect—the fact that past returns affect (negatively) future realized volatilities but not the other way around, which is easily captured by SV models via a negative spot-vol correlation.

In this article, building on [11, 4, 22, 20, 30], which we discuss in the next section, we aim to answer the two crucial questions:

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

We will cover both implied volatility and future realized volatility.

### 3. EMPIRICAL STUDY: LEARNING PATH-DEPENDENT VOLATILITY

We run our statistical models on 22 years of data: the training set starts on January 1, 2000 and ends on December 31, 2018, while the test set runs from January 1, 2019 to May 15, 2022. This test set is particularly interesting and challenging, since it contains very different volatility regimes: pre-Covid-19 crisis, quiet financial times (January 2019–March 2020); the Covid-19 crisis of March 2020; and then its aftermath.

**3.1. Features.** A first, crucial task for learning PDV is to identify path-dependent features that have the potential to explain most of the variability in volatility. We focus on two main types of features:

---

<sup>2</sup>SV models become complete if the volatility is allowed to depend on prices of derivatives on the underlying asset. However, writing joint dynamics of an asset price and the prices of options written on this asset is notoriously very difficult.

- **Trend features** are features that capture a *recent trend* in the asset price in order to learn the *leverage effect*, i.e., the fact that volatility tends to rise when asset prices fall. The most important example of a trend feature is a weighted sum of past daily returns

$$(3.1) \quad R_{1,t} := \sum_{t_i \leq t} K_1(t - t_i) r_{t_i}$$

where

$$(3.2) \quad r_{t_i} := \frac{S_{t_i} - S_{t_{i-1}}}{S_{t_{i-1}}}$$

denotes the daily return between day  $t_{i-1}$  and day  $t_i$ , and  $K_1 : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$  is a convolution kernel that puts more or less weight on past daily returns based on the *lag*  $t - t_i$ , i.e., on how far in the past the daily return was observed. The kernel  $K_1$  typically decreases towards zero: the impact of a given daily return fades away over time. Other examples of trend features include the weighted sum of negative parts of past daily returns

$$(3.3) \quad N_t := \sum_{t_i \leq t} K_N(t - t_i) r_{t_i}^- \quad (\text{or more generally } N_t^\varphi := \sum_{t_i \leq t} K_N(t - t_i) \varphi(r_{t_i}))$$

for a nonlinear function  $\varphi$  able to capture a trend, such as  $\varphi(r) = r^+$  or  $r^3$  or  $(r^-)^2$ ; and the spot-to-moving-average ratios

$$(3.4) \quad U_t := \frac{S_t}{A_t}, \quad A_t := \sum_{t_i \leq t} K_A(t - t_i) S_{t_i}.$$

- **Activity features or volatility features** are features that capture recent activity (volatility) in the asset price, regardless of its trend, in order to learn *volatility clustering*, i.e., the fact that periods of large volatility tend to be followed by periods of large volatility, and periods of small volatility tend to be followed by periods of small volatility. Volatility clustering also materializes in the fact that implied volatility tends to be larger when historical volatility is larger. The most important example of a volatility feature is a weighted sum of past squared daily returns

$$(3.5) \quad R_{2,t} := \sum_{t_i \leq t} K_2(t - t_i) r_{t_i}^2.$$

The kernel  $K_2$  also typically decreases towards zero. For simplicity we denote

$$(3.6) \quad \Sigma_t := \sqrt{R_{2,t}},$$

which is the  $K_2$ -weighted historical volatility. Higher even moments of past daily returns may also be considered.

Note that since all those features depend only the *returns*  $r_{t_i}$ , they are scale-invariant, a desirable property that we discussed in Section 2.3.

### 3.2. The model.

In the next sections we show that a simple linear model of the form

$$(3.7) \quad \text{Volatility}_t = \beta_0 + \beta_1 R_{1,t} + \beta_2 \Sigma_t, \quad \beta_0 > 0, \quad \beta_1 < 0, \quad \beta_2 \in (0, 1)$$

explains a very large part of the variability observed in  $\text{Volatility}_t$ , where  $\text{Volatility}_t$  denotes either some implied volatility (e.g., the VIX) observed at  $t$ , or some future realized volatility right after  $t$  (e.g., realized over day  $t+1$ ). The model states that the volatility—not its square, the variance—is simply an affine combination of the trend feature  $R_{1,t}$  and the historical volatility  $\Sigma_t$ . Note that  $\beta_1 < 0$  produces the leverage effect: the smaller  $R_{1,t}$ , the larger  $\text{Volatility}_t$ ;  $\beta_2 > 0$  produces volatility clustering, like in GARCH models: the larger  $\Sigma_t$ , the larger  $\text{Volatility}_t$ ; and  $\beta_2 < 1$  guarantees stability, also like in GARCH models. Importantly, both factors  $R_{1,t}$  and  $\Sigma_t$  are needed to satisfactorily explain the volatility (see Figure 3.7); we find that a simple linear model does the job, explaining a very large part of the variability observed in the volatility. Our experiments with nonlinear models (parameterized by neural networks) and extra trend features  $N_t^\varphi$  and  $U_t$  did not lead to significantly better results.

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  $\tau$ . In this sense, volatility has short memory.
- (2) Nevertheless, volatility also has long memory: the weights  $K_n(\tau)$  decrease slowly for large  $\tau$ , much more slowly than exponential kernels. The persistence of volatility has been extensively documented in the literature, dating back to [1, 9, 27, 12].

The two above facts were checked by running a multivariate lasso regression with variables  $R_{1,t}^{(\lambda_j)}$  and  $\sqrt{R_{2,t}^{(\mu_k)}}$ , where

$$(3.8) \quad R_{n,t}^{(\lambda)} := \sum_{t_i \leq t} K^{(\lambda)}(t - t_i) r_{t_i}^n, \quad K^{(\lambda)}(\tau) := \lambda e^{-\lambda\tau}, \quad \lambda > 0.$$

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

$$(3.9) \quad K(\tau) = K_{\alpha,\delta}(\tau) := Z_{\alpha,\delta}^{-1}(\tau + \delta)^{-\alpha}, \quad \tau \geq 0, \quad \alpha > 1, \quad \delta > 0,$$

with only two parameters  $\alpha > 1$ ,  $\delta > 0$ . The time shift  $\delta$ , typically a few days or weeks (see Table 3 and Table 5), guarantees that  $K_{\alpha,\delta}(\tau)$  does not blow up when the lag  $\tau$  vanishes. If we force  $\delta$  to be 0, we recover the power-law kernel of rough volatility models. However, our empirical tests all select positive  $\delta$ . Note that for the purpose of uniquely identifying the linear coefficients  $\beta_j$  in (3.7), we normalize the kernels  $K$  so that  $\sum_{i=0}^{\infty} K(i\Delta t)\Delta t = 1$ , where  $\Delta t = \frac{1}{252}$  represents one business day. In the continuous-time limit,

$$(3.10) \quad Z_{\alpha,\delta} = \int_0^{\infty} (\tau + \delta)^{-\alpha} d\tau = \frac{\delta^{1-\alpha}}{\alpha - 1}.$$

Other examples of two-parameter kernels that produce a power-law-like decay (away from very short maturities) but do not blow up for vanishing lag  $\tau$  are

$$(3.11) \quad K(\tau) = K_{\alpha,\kappa}(\tau) := I((\kappa\tau)^{\alpha}) \quad \text{or} \quad I(\kappa\tau)^{\alpha}, \quad \tau \geq 0, \quad I(x) := \frac{1 - e^{-x}}{x}$$

with  $\alpha > 1$ ,  $\kappa > 0$ , as suggested in [25].

**3.3. Comparison with other models.** Similar models have been proposed in the past. For instance, (the infinite-memory generalization of) the quadratic ARCH (QARCH) model of Sentana [31] expresses the instantaneous variance Volatility $_t^2$  as a general quadratic form of the past daily returns:

$$(3.12) \quad \text{Volatility}_t^2 = \beta_0 + \beta_1 R_{1,t} + \beta_2 R_{2,t}^Q, \quad R_{2,t}^Q := \sum_{t_i, t_j \leq t} K_2^Q(t - t_i, t - t_j) r_{t_i} r_{t_j}.$$

Chicheportiche and Bouchaud [11] have shown that in the US stock market the off-diagonal entries of the bidimensional kernel  $K_2^Q$  are one order of magnitude smaller than the diagonal entries. The corresponding diagonal QARCH model reads

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

with  $K_2(\tau) := K_2^Q(\tau, \tau)$ . Chicheportiche and Bouchaud do not provide explicit parametric forms for  $K_1$  and  $K_2$  but report power-law-like decays in [11, Figure 4]. The ZHawkes process of Blanc et al. [4]

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

replaces the linear dependence of the variance Volatility $_t^2$  on the trend feature  $R_{1,t}$  by a quadratic dependence. See [14] for a microstructural foundation of this model. The discrete-time version of the QRHM [20] with null  $\theta_0$  can be written as

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

with a Mittag-Leffler kernel  $K_1$  such that  $K_1(\tau) \underset{\tau \rightarrow 0}{\sim} \tau^{\alpha-1}$  and  $K_1(\tau) \underset{\tau \rightarrow +\infty}{\sim} \tau^{-\alpha-1}$  with  $\alpha \in (\frac{1}{2}, 1)$ ; and the discrete-time version of the threshold EWMA Heston model [30, Section 2.3] reads as

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

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

Our model differs from the above models in several ways:

	Train		Test	
	RMSE	$r^2$	RMSE	$r^2$
<b>VIX</b>	0.020	0.946	0.035	0.855
<b>VIX9D</b>	0.023	0.876	0.034	0.914
<b>VSTOXX</b>	0.026	0.929	0.029	0.913
<b>IVI</b>	0.023	0.925	0.030	0.870
<b>VDAX-NEW</b>	0.025	0.934	0.027	0.918
<b>Nikkei 225 VI</b>	0.030	0.890	0.031	0.800

TABLE 2. RMSE and  $r^2$  scores for our model (3.7) for various implied volatility indexes

	$\beta_0$	$\alpha_1$	$\delta_1$	$\beta_1$	$\alpha_2$	$\delta_2$	$\beta_2$
<b>VIX</b>	0.057	1.06	0.020	-0.095	1.60	0.052	0.82
<b>VIX9D</b>	0.045	1.00	0.011	-0.12	1.25	0.011	0.88
<b>VSTOXX</b>	0.032	3.96	0.13	-0.036	1.90	0.089	0.97
<b>IVI</b>	0.022	2.26	0.081	-0.058	1.6	0.063	0.99
<b>VDAX-NEW</b>	0.036	5.54	0.16	-0.024	2.21	0.103	0.92
<b>Nikkei 225 VI</b>	0.055	0.78	0.008	-0.069	2.09	0.077	0.86

TABLE 3. Optimal parameters of our model (3.7) for various implied volatility indexes

- (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  $\Sigma_t$  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-blown-up power-law-like decays*.
- (5) Compared with (M3) and (M4), we empirically prove the importance of including the historical volatility factor  $\Sigma_t$ . In particular Figures 3.7 and 3.8 show that the models that ignore the historical volatility feature  $\Sigma$  perform poorly compared to those that do not ignore it.
- (6) Compared with (M2), we argue that it is not necessary to include a quadratic factor  $R_{1,t}^2$ , as the quadratic-like (decreasing then increasing) dependence of the volatility (resp. variance) on  $R_{1,t}$  is already captured by the factor  $\Sigma_t$  (resp.  $R_{2,t}$ ), see Section 3.4.

We will compare the performances of all these models in Section 3.6.

**3.4. Learning implied volatility.** In this section we learn the *implied* volatility from past asset returns. In the next section we will cover the future daily *realized* volatility. While in principle daily realized volatility is a better proxy for the (hypothetical) instantaneous volatility than implied volatility, realized volatilities are very noisy. The main benefit of working with implied volatilities is that they are well defined measures of volatility calculated from the prices of derivatives that are liquidly traded on an exchange. In our empirical study we will only consider *short-term* implied volatilities (with a maturity not exceeding 30 days) which are still expected to be reasonable proxies for the instantaneous volatility [16].

We consider the following popular implied volatility indexes: VIX, VSTOXX, IVI, VDAX-NEW, Nikkei 225 VI, representing the 30-day implied volatilities of SPX, EURO STOXX 50, FTSE 100, DAX, and Nikkei 225 respectively, as well as VIX9D (the “9-day VIX”). For each model, we minimize the mean squared error (MSE) on the training set using the *trust region reflective algorithm* [8] of the function `least_squares` from the `scipy` [32] module in Python. In order to start from reasonable initial TSPL parameter guesses  $(\alpha_1, \delta_1, \alpha_2, \delta_2)$ , we first perform a multivariate lasso regression using the variables (3.8). For both  $K_1$  and  $K_2$ , we then find the TSPL that best fits the linear combination of exponential kernels produced by lasso. Note that to compute  $R_{1,t}$  and  $\Sigma_t$ , we use the previous 1,000 business days (about 4 years). Given the shapes of the kernels  $K_1, K_2$  (see Figures 3.1 and 3.4), 4 years is a very safe cutoff lag.

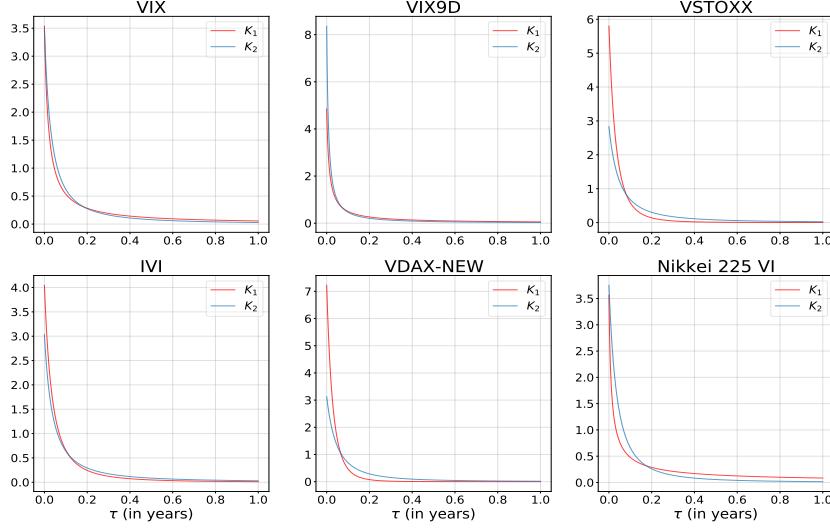


FIGURE 3.1. Optimal kernels  $K_1$  and  $K_2$  in our model (3.7) for various implied volatility indexes

To evaluate model performance, we use the R squared ( $r^2$ ) score and the root mean squared error (RMSE). The scores of our model (3.7) are reported in Table 2. The  $r^2$  value expresses how much of the variance of the volatility is explained by the model. For example, our model explains 94.6% of the variation of the VIX on the training data set (in sample) and 85.5% of it on the test data set (out of sample); and it explains 92.9% of the variation of the VSTOXX on the training set and 91.3% of it on the test set. The train RMSEs are about only 2–2.5 volatility points (3 for the Nikkei index), while the test RMSEs are about 3 volatility points. The remarkably high  $r^2$  scores and low RMSEs demonstrate the *high degree of endogeneity of volatility*.

The optimal parameters are reported in Table 3 and the corresponding kernels  $K_1, K_2$  in Figure 3.1. In particular we note that

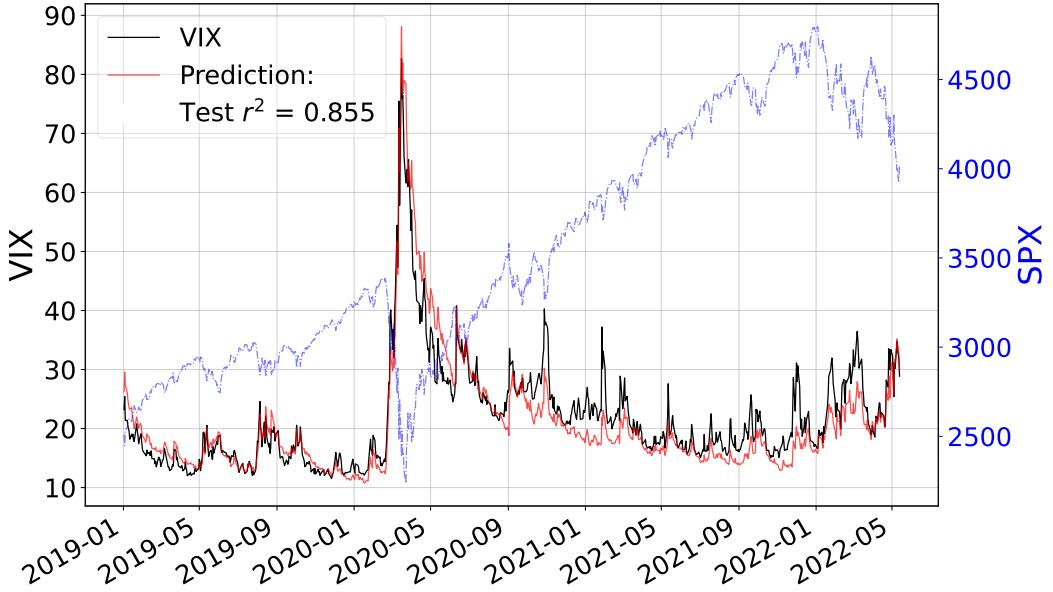
- the optimal kernels  $K_1$  and  $K_2$  are quite similar;
- the optimal time shifts  $\delta_1, \delta_2$  are positive, ranging from 0.01 (2.5 business days) to 0.16 (40 business days); the larger  $\delta_i$ , the larger  $\alpha_i$ ;
- The  $\beta_0$ s are positive, between 2 and 6%; the  $\beta_1$ s are negative; the  $\beta_2$ s take values in  $(0.8, 1)$ , and are close to 1 for the IVI and the VSTOXX. (Note that we do not impose any constraint on the  $\beta$ s when running the optimization.)

Our model accurately predicts the current VIX based on the SPX path only, as we can see in Figure 3.2 (top) which compares the time series of our VIX prediction  $\widehat{\text{VIX}}$  with the time series of the actual VIX on the test set. *It is crucial to note that at each point in time our prediction of the current VIX value completely ignores past VIX values, in particular the values taken by the VIX in the recent past; our prediction is based on past SPX values only (the dashed blue line in Figure 3.2 (top)).* One would of course build a more accurate prediction of the current VIX value by also taking recent past VIX values into account. However, since our goal is to measure how much of the VIX is explained by past SPX returns only, we refrain from doing so. Our PDV model (3.7) accurately captures:

- the small pre-Covid-19 VIX variations;
- the very fast increase in VIX values in March 2020, at the beginning of the Covid-19 crisis;
- the slow decay of the VIX after March 2020, with the VIX staying quite large despite the fact that the SPX index performed extraordinarily well between April 2020 and December 2021.

Figure 3.2 (top) reveals some peculiarities of the Covid-19 crisis, compared to the past crises that belong to the training data set (2000–2018):

- The VIX index decreased faster than expected (based on the SPX bounce) in April and May 2020.
- However, the VIX stayed slightly higher than expected since October 2020. In particular, moderate SPX drops produced larger VIX spikes than expected, possibly a sign of market nervousness as the Covid-19 crisis continued to unfold and bring uncertainty about the future of the economy.



(A) Time series of VIX and predicted VIX values on test set. The dashed blue line represents the SPX time series

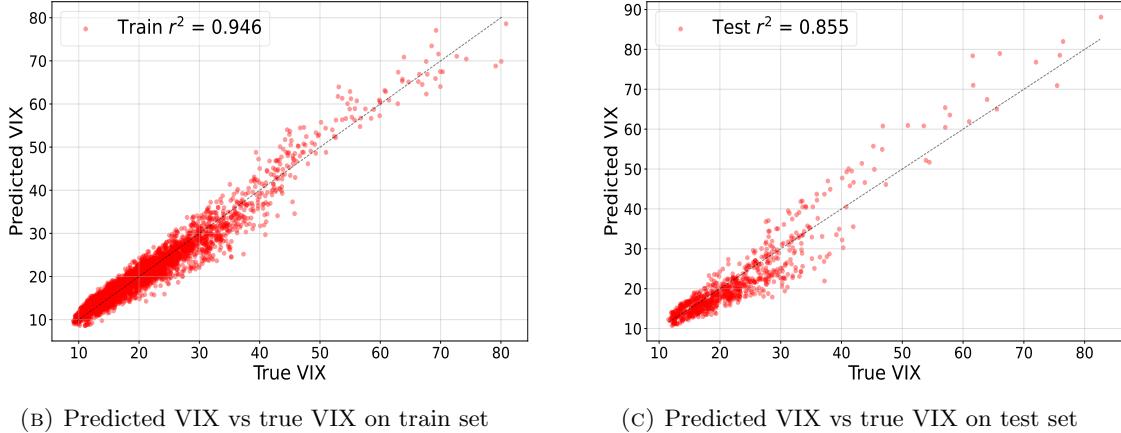


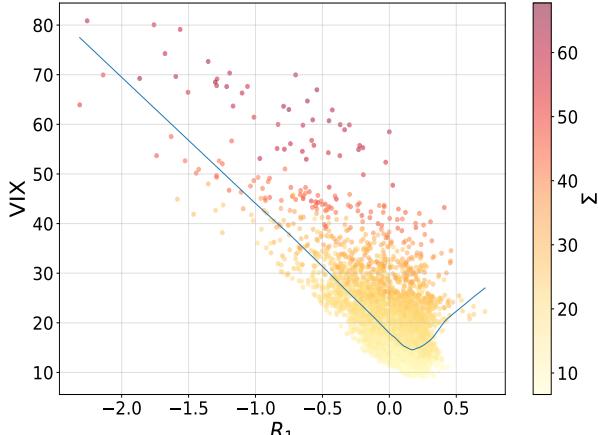
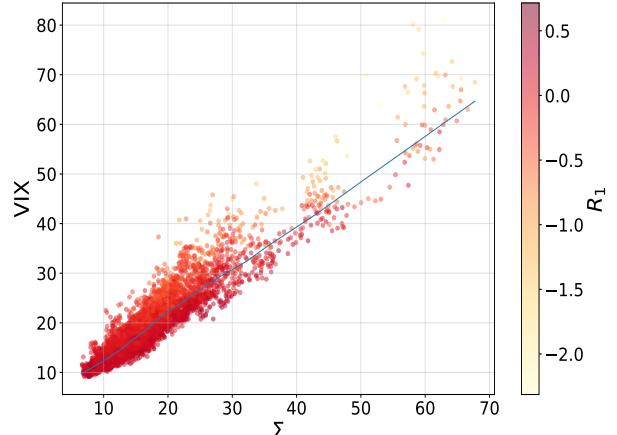
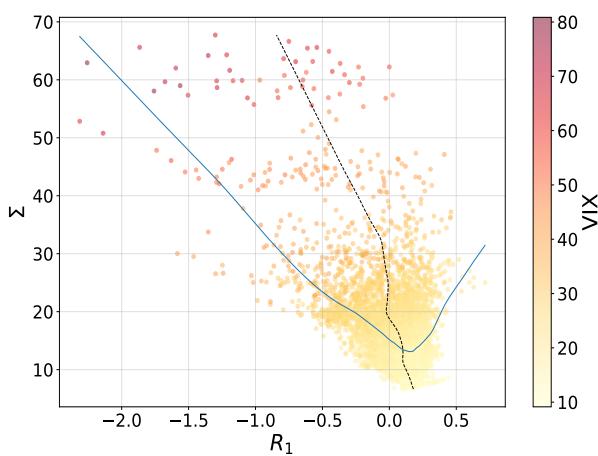
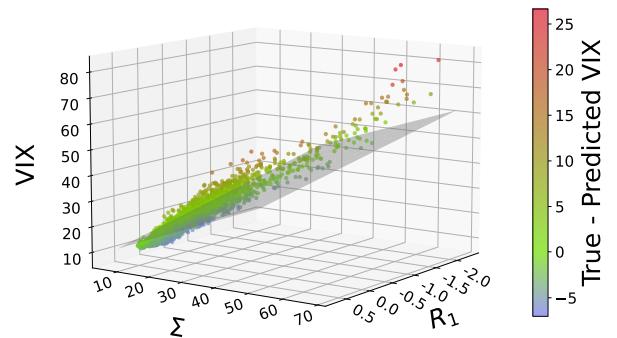
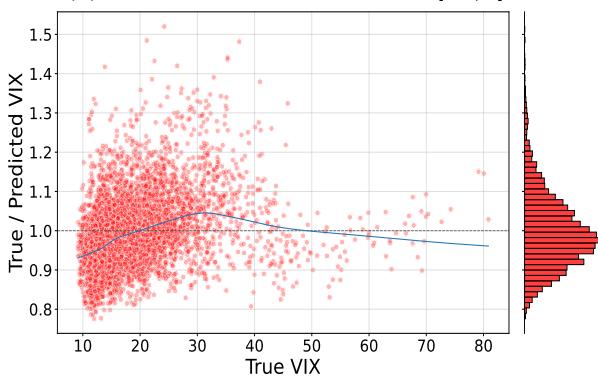
FIGURE 3.2. Top: VIX prediction time series of our model (3.7). Bottom: prediction against true value on train set (Jan 1, 2000–Dec 31, 2018, left) and on test set (Jan 1, 2019–May 15, 2022, right)

Figure 3.3 helps to visualize the model, its relevance, and its performance. Figure B shows that the VIX is mostly a linear function of the historical volatility  $\Sigma$ . The color scale shows that the trend  $R_1$  contributes significantly as well: for a given  $\Sigma$ , the smaller  $R_1$ , the larger the VIX, roughly linearly. Figure A shows that the VIX is mostly a decreasing function of  $R_1$ , but also that:

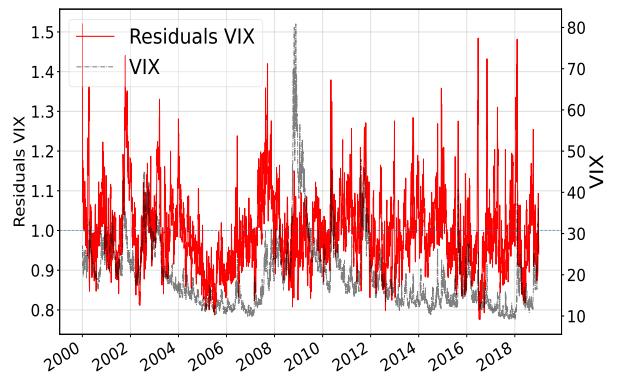
- (1)  $\Sigma$  significantly contributes to the VIX as well, more than  $R_1$  itself;
- (2) the VIX tends to increase with  $R_1$  for large  $R_1$ .

To account for this last observation, one could think of adding an  $R_1^2$  feature in the linear model. However, Figure 3.3 (C) shows that this is not needed when the linear feature  $\Sigma$  is present.<sup>3</sup> Indeed, it shows that a linear model in  $\Sigma$  already captures a quadratic-like dependence on  $R_1$ : the empirical conditional expectation  $\mathbb{E}[\Sigma|R_1]$  looks like the empirical conditional expectation  $\mathbb{E}[VIX|R_1]$ , a decreasing then increasing function of  $R_1$ . This nonmonotonic shape of  $\mathbb{E}[\Sigma|R_1]$  is not surprising, as large negative (resp. positive) values of  $R_{1,t}$  imply large negative (resp. positive) recent SPX returns, hence large positive recent SPX squared returns,

<sup>3</sup>We also verified empirically that an  $R_1^2$  term is not necessary, as adding it does not improve the train RMSE or  $r^2$ .

(A) VIX vs  $R_1$ (B) VIX vs  $\Sigma$ (C)  $\Sigma$  vs  $R_1$ . Dashed black line:  $\mathbb{E}[R_1|\Sigma]$ (D) 3D scatter plot of VIX vs  $R_1$  and  $\Sigma$ . The grey plane represents the linear plane generated by our model (3.7). The color codes for the vertical distance from each point to the plane, i.e., the residual

(E) (Ratio) residuals vs VIX. The plot on the right displays the distribution of the residuals



(F) (Ratio) residuals time series

FIGURE 3.3. VIX, residuals, and features scatter plots on the train set. On a scatter plot displaying  $Y$  vs  $X$ , the blue line represents  $\mathbb{E}[Y|X]$ . VIX and  $\Sigma$  are expressed in percentage

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

TABLE 4. RMSE and  $r^2$  scores for our model (3.7) for the daily realized volatility of various indexes

	$\beta_0$	$\alpha_1$	$\delta_1$	$\beta_1$	$\alpha_2$	$\delta_2$	$\beta_2$
<b>SPX</b>	0.018	2.82	0.044	-0.042	1.86	0.025	0.71
<b>STOXX</b>	0.023	1.31	0.017	-0.062	1.79	0.024	0.70
<b>FTSE</b>	0.017	2.22	0.034	-0.043	1.84	0.031	0.76
<b>DAX</b>	0.001	2.87	0.045	-0.030	1.80	0.029	0.81
<b>NIKKEI</b>	0.032	6.30	0.063	-0.011	2.30	0.030	0.51

TABLE 5. Optimal parameters of our model (3.7) for the daily realized volatility of various indexes

i.e., large  $\Sigma_t$ , while small values of  $R_{1,t}$  are consistent with both small and large values of  $\Sigma_t$ . Note that, by contrast, the empirical conditional expectation  $\mathbb{E}[R_1|\Sigma] \approx 0$ : given the historical volatility  $\Sigma$ , on average, the historical trend is close to zero. More precisely, it is slightly positive for small  $\Sigma$ , and negative, with a linear trend, for large  $\Sigma$ : a small historical volatility corresponds to slightly positive weighted past returns on average, whereas a large historical volatility is associated with negative weighted past returns on average. Figures 3.3 (A)(B)(C) are actually the three two-dimensional projections of the three-dimensional scatter plot in Figure 3.3 (D), which perhaps offers the best visualization of our model: the VIX roughly lies in the plane defined by our model

$$(3.13) \quad \text{VIX} = \beta_0 + \beta_1 R_1 + \beta_2 \Sigma, \quad \beta_0 > 0, \quad \beta_1 < 0, \quad \beta_2 \in (0, 1).$$

This model captures the quadratic-like behavior of the VIX as a function of  $R_1$  only, as

$$\mathbb{E}[\text{VIX}|R_1] = \beta_0 + \beta_1 R_1 + \beta_2 \mathbb{E}[\Sigma|R_1]$$

and the linear behavior of the VIX as a function of  $\Sigma$  only, as

$$\mathbb{E}[\text{VIX}|\Sigma] = \beta_0 + \beta_1 \mathbb{E}[R_1|\Sigma] + \beta_2 \Sigma \approx \beta_0 + \beta_2 \Sigma.$$

Therefore models (M3)-(M4) can be seen as approximations of the projection of our model onto models where the volatility (here, the VIX) depends only on  $R_1$ :  $\text{Volatility}_t = f(R_{1,t})$ .

Finally, Figure 3.3 (E) displays the ratio residuals  $\frac{\text{VIX}}{\mathbb{E}[\text{VIX}]}$  against the true VIX. It shows that the ratio is roughly centered around 1, with little dependence on the value of the VIX, and is approximately Gaussian. Figure 3.3 (F) displays the time series of the ratio residuals and shows that the process of the ratio looks stationary.

**3.5. Learning realized volatility.** In this section we now learn *next day realized* volatility from past asset returns. We use the Oxford-Man Institute volatility database [28] and train our statistical model on SPX, EURO STOXX 50, FTSE 100, DAX, and Nikkei 225. Realized volatility (RV) is much more noisy than implied volatility. Several observers using different measures of RV may come up with quite different estimates. The Oxford-Man Institute RV is based on 5-minute returns per trading day; there are 78 5-minute returns per trading day for the SPX. As explained in Section 4.6, the small sample size of 78 means that a significant part of the RV estimate is just white noise, and it creates spurious roughness when we try to measure the roughness of volatility paths. By contrast, implied volatility is a well-defined quantity resulting from supply and demand, at least for assets having a liquid options market like the equity indexes considered here. Therefore we do not want our model to exactly learn RV estimates, and we do not expect our model to explain as much variability of RV as of implied volatility. The prediction scores are indeed lower, as we can see in Table 4. Still, past asset returns explain about 70% of the variability of next day RV on the train

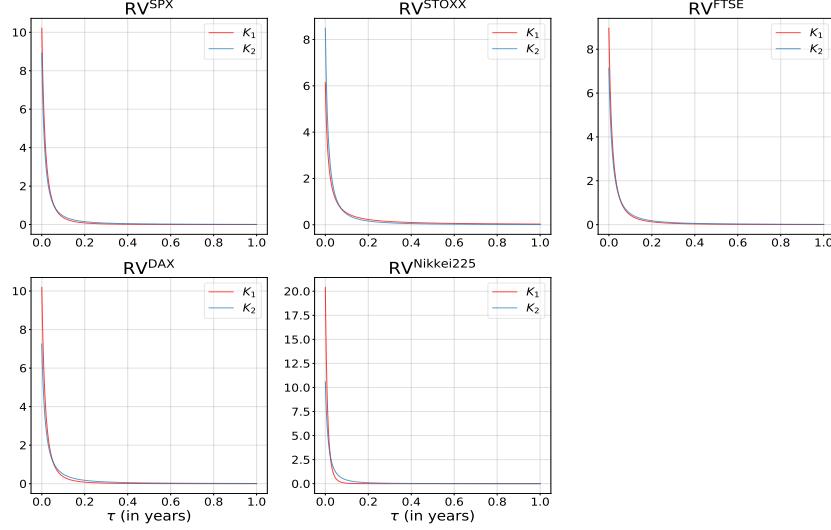


FIGURE 3.4. Optimal kernels  $K_1$  and  $K_2$  in our model (3.7) for the daily realized volatility of various indexes

$n$	3				5			
	Train		Test		Train		Test	
	RMSE	$r^2$	RMSE	$r^2$	RMSE	$r^2$	RMSE	$r^2$
<b>SPX</b>	0.041	0.785	0.060	0.641	0.040	0.783	0.064	0.572
<b>STOXX</b>	0.048	0.744	0.054	0.738	0.046	0.752	0.056	0.705
<b>FTSE</b>	0.042	0.743	0.052	0.709	0.040	0.755	0.051	0.697
<b>DAX</b>	0.047	0.779	0.056	0.553	0.046	0.783	0.058	0.490
<b>Nikkei 225</b>	0.043	0.612	0.046	0.505	0.042	0.610	0.046	0.473

TABLE 6. RMSE and  $r^2$  scores for our model (3.7) for the  $n$ -day realized volatility of various indexes,  $n \in \{3, 5\}$

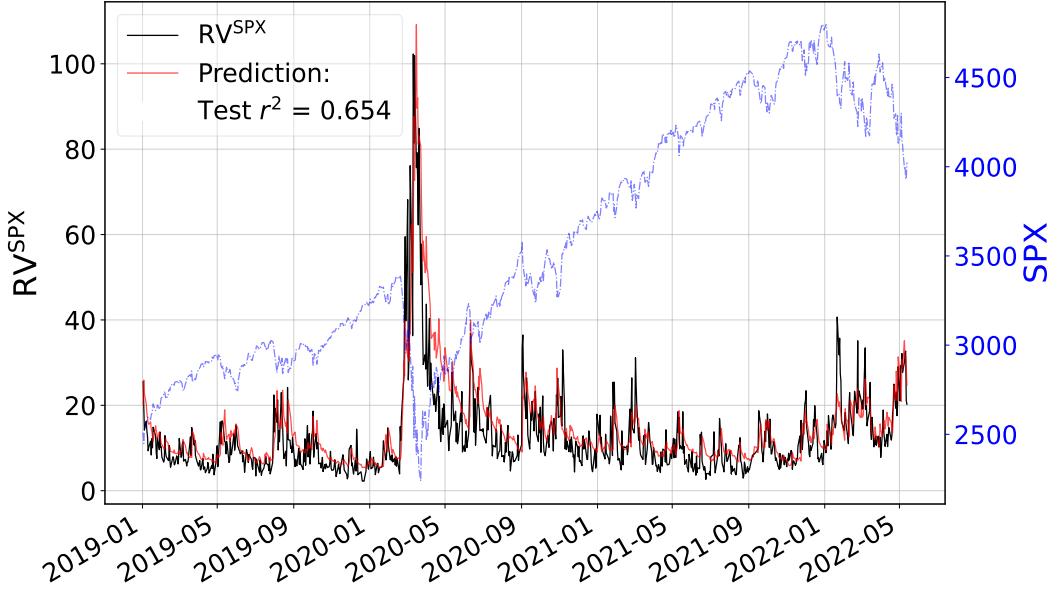
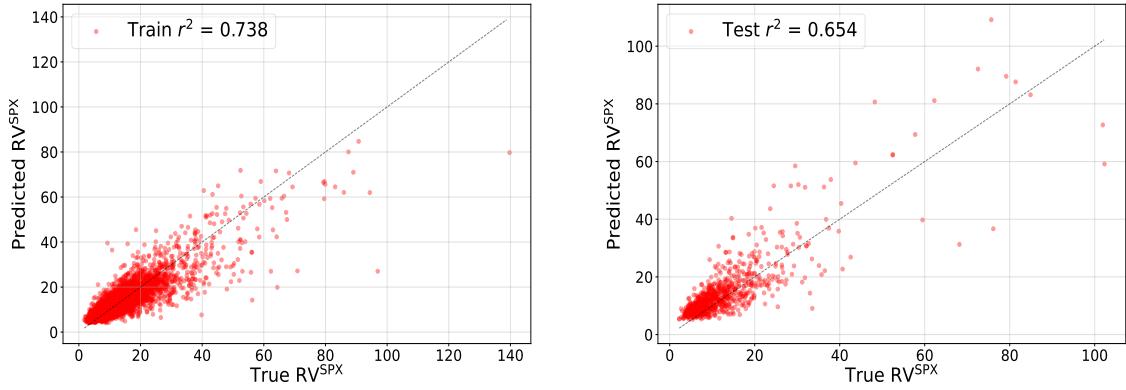
set, and about 60% on the test set. Compared to implied volatilities, the lower  $r^2$  and larger RMSE values are due to the noise in estimating RV. In Table 6 we also report the  $r^2$  scores and RMSEs for  $n$ -day future realized volatilities when  $n = 3$  and  $n = 5$ . While the train  $r^2$  are larger for those larger values of  $n$ , the test  $r^2$  stay roughly the same.

In Table 5 and Figure 3.4 we report the optimal parameters of our model for RV and the corresponding optimal kernels  $K_1, K_2$  for the various equity indexes considered. We note that:

- compared to implied volatility parameters, the  $\beta$ s are smaller in absolute value, and the kernels  $K_1, K_2$  decrease faster;
- the parameters  $(\alpha_2, \delta_2, \beta_2)$  are remarkably similar across indexes (except maybe for the Nikkei);
- the time shifts  $\delta$  lie between 0.017 (4 business days  $\approx$  1 week) and 0.06 (15 business days  $\approx$  3 weeks).

Figure 3.5 (top) shows that (a) RV is much more noisy than implied volatility, (b) it decays much faster than implied volatility after large shocks, and (c) our simple linear model accurately predicts (a denoised version of) RV. Finally, Figure 3.6 is the equivalent of Figure 3.3 for the RV of the SPX index and brings similar qualitative conclusions, up to the extra measurement noise in RV.

*Remark 1.* As an alternative to the TSPL power-law kernel, one can use a convex combination of two exponential kernels as in (4.9). Figure 4.3 shows that the two parametrizations produce very similar kernels. We report the optimal parameters and scores of our linear model when we use this alternate kernel in Appendix A for both implied and realized volatility. Note that, as expected, both kernels lead to very similar scores and  $\beta$ s.

(A) Time series of  $RV^{SPX}$  and predicted  $RV^{SPX}$  values. The dashed blue line represents the SPX time seriesFIGURE 3.5. Top:  $RV^{SPX}$  prediction time series of our model (3.7). Bottom: prediction against true value on train set (Jan 1, 2000–Dec 31, 2018, left) and on test set (Jan 1, 2019–May 15, 2022, right)

**3.6. Comparing model scores.** In this section we compare the performances of our “ $\text{Linear}(R_1, \Sigma)$ ” model (3.7); of its simpler versions “ $\text{Linear}(R_1)$ ”, where we enforce  $\beta_2 = 0$  in (3.7), and “ $\text{Linear}(\Sigma)$ ”, where we enforce  $\beta_1 = 0$  in (3.7); and of Models (M1)–(M4), all equipped with TSPL kernels  $K_1, K_2$ . We also compare our model with two ARCH benchmarks. The first one, denoted by ARCH(1000), is an ARCH model using the  $p = 1,000$  previous daily returns; we use lasso to fit it for better generalization. The second one, denoted by ARCH-TSPL, is an ARCH model where the weights follow a TSPL kernel (3.9). It is thus identical to (M1) with  $\beta_1 = 0$ . We compare the scores of our linear models against (M1)–(M4) and the benchmark ARCH models in Figure 3.7 ( $r^2$  scores) and Figure 3.8 (RMSE scores) for the five major equity indexes and their implied volatility indexes considered above.<sup>4</sup>

Figures 3.7 and 3.8 show that our model consistently outperforms all the other models across equity indexes for both implied and realized volatilities. As already mentioned, this remarkable result might be due to a

<sup>4</sup>Note that despite the fact that the model ARCH-TSPL is included in ARCH(1000), it has a larger train  $r^2$  for some indexes. This happens because in Figure 3.7 and Figure 3.8, we compare the prediction of volatilities, while those models are optimized on the prediction of variances. The same remark applies to ARCH-TSPL and (M1).

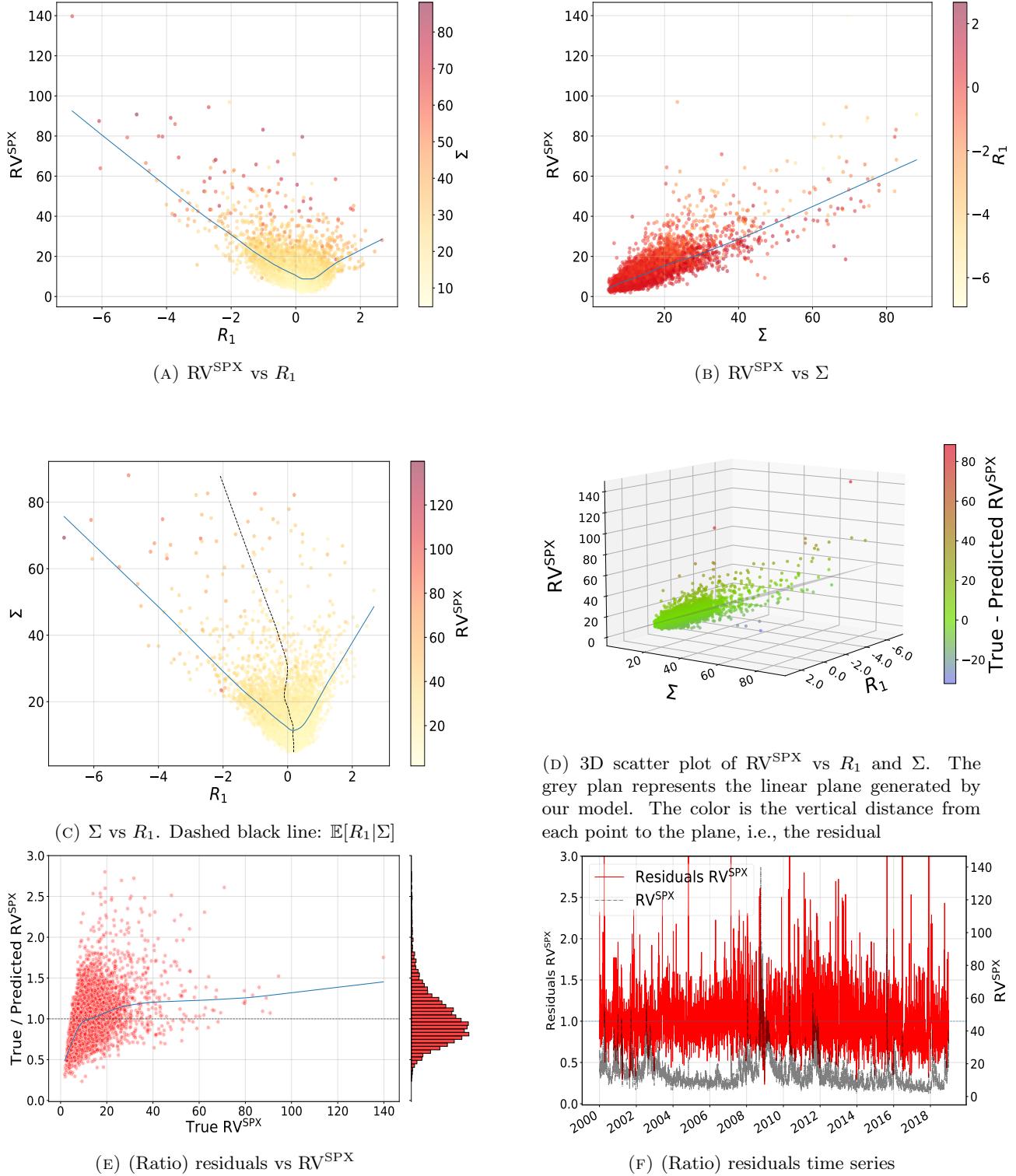


FIGURE 3.6.  $RV^{SPX}$ , residuals, and features scatter plots on the train set. On a scatter plot displaying  $Y$  vs  $X$ , the blue line represents  $\mathbb{E}[Y|X]$ .  $RV^{SPX}$  and  $\Sigma$  are expressed in percentage

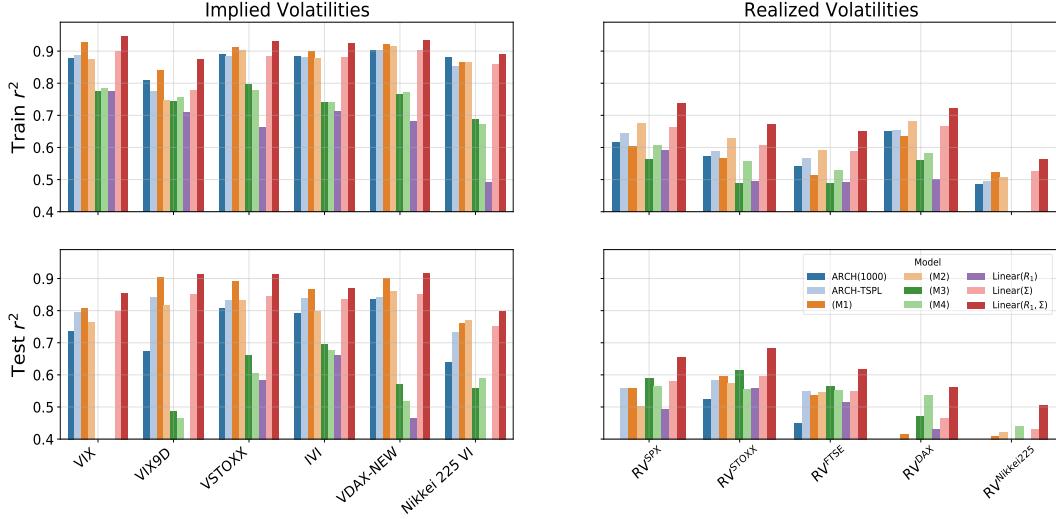


FIGURE 3.7. Comparison of  $r^2$  scores for the different models (M1)-(M4), ARCH models, and our linear models. Top:  $r^2$  scores on the train set. Bottom:  $r^2$  scores on the test set. Left: Implied volatilities. Right: Realized volatilities

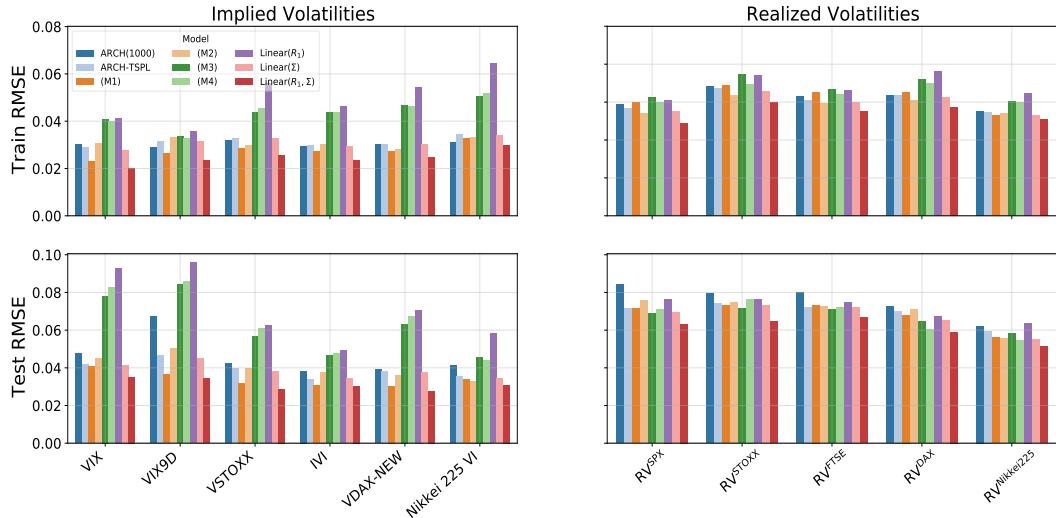


FIGURE 3.8. Comparison of RMSE scores for the different models (M1)-(M4), ARCH models, and our linear models. Top: RMSE scores on the train set. Bottom: RMSE scores on the test set. Left: Implied volatilities. Right: Realized volatilities

*homogeneity in volatility* property: all the terms in our linear model (3.7) have the dimension of 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) and the ARCH benchmarks are homogeneous to a *variance*. One possible explanation for the overperformance of volatility-homogeneous models is that, when reacting to market movements, traders probably compare homogeneous quantities, and they probably think more naturally in terms of returns or volatility than in terms of squared returns or variance. Note that the second best performing model for learning implied volatilities seems to be Model (M1).

Like our model, Model (M4) is homogeneous in volatility, but it ignores an important explanatory variable: the historical volatility  $\Sigma$ . Figures 3.7 and 3.8 show that it is important to combine the volatility feature  $\Sigma$  with the trend feature  $R_1$ : if we ignore one of these two features, the performance of our model drops significantly. The performance drops the most when we ignore  $\Sigma$ . This is also well illustrated by Figure 3.3

(D): the VIX is very well captured by an affine function of  $\Sigma$  and  $R_1$  and depends more on  $\Sigma$  than on  $R_1$ , but if we ignore the dependence on  $R_1$ , we miss an important “angle” and explanatory factor. Note that the models that ignore the volatility features  $\Sigma$  or  $R_2$  (Models (M3), (M4), and Linear( $R_1$ )) perform particularly poorly when we learn the path-dependence of implied volatilities.

Appendix B shows that our results are robust with respect to the test set. In particular, our model consistently outperforms the other models when we vary the test set.

#### 4. THE CONTINUOUS-TIME EMPIRICAL PATH-DEPENDENT VOLATILITY MODEL

We now consider the continuous-time limit of Model (3.7), where we identify Volatility $_t$  with the instantaneous volatility  $\sigma_t$ :

$$(4.1) \quad \begin{aligned} \frac{dS_t}{S_t} &= \sigma_t dW_t, \\ \sigma_t &= \sigma(R_{1,t}, R_{2,t}), \\ \sigma(R_1, R_2) &= \beta_0 + \beta_1 R_1 + \beta_2 \sqrt{R_2}, \\ R_{1,t} &= \int_{-\infty}^t K_1(t-u) \frac{dS_u}{S_u} = \int_{-\infty}^t K_1(t-u) \sigma_u dW_u, \\ R_{2,t} &= \int_{-\infty}^t K_2(t-u) \left( \frac{dS_u}{S_u} \right)^2 = \int_{-\infty}^t K_2(t-u) \sigma_u^2 du. \end{aligned}$$

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 cannot be written as functions of  $(R_{1,t}, R_{2,t})$ .

##### 4.1. A (too) simple Markovian approximation: the 2-factor Markovian PDV model.

4.1.1. *The simplest Markovian version of Model (4.1).* The simplest kernels yielding a Markovian model are the (normalized) exponential kernels  $K_1(\tau) := K^{(\lambda_1)}(\tau) := \lambda_1 e^{-\lambda_1 \tau}$  and  $K_2(\tau) := K^{(\lambda_2)}(\tau) := \lambda_2 e^{-\lambda_2 \tau}$ ,  $\lambda_1, \lambda_2 > 0$ . In this case,  $K'_1 = -\lambda_1 K_1$  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:

$$(4.2) \quad \begin{aligned} \frac{dS_t}{S_t} &= \sigma(R_{1,t}, R_{2,t}) dW_t, \quad \sigma(R_1, R_2) = \beta_0 + \beta_1 R_1 + \beta_2 \sqrt{R_2}, \\ dR_{1,t} &= \lambda_1 \left( \frac{dS_t}{S_t} - R_{1,t} dt \right) = \lambda_1 (\sigma(R_{1,t}, R_{2,t}) dW_t - R_{1,t} dt), \\ dR_{2,t} &= \lambda_2 \left( \left( \frac{dS_t}{S_t} \right)^2 - R_{2,t} dt \right) = \lambda_2 (\sigma(R_{1,t}, R_{2,t})^2 - R_{2,t} dt). \end{aligned}$$

We call this model the *2-factor Markovian PDV model*, as the instantaneous volatility is a deterministic function of the two Markovian factors  $(R_{1,t}, R_{2,t})$ . The model is fully described by the three-dimensional Markovian vector  $(S_t, R_{1,t}, R_{2,t})$ .

As explained in Section 3.2, 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. In the next section we will capture this mix of short and long memory in a Markovian way by choosing  $K_1$  and  $K_2$  to be convex combinations of two exponential kernels. Model (4.2) is therefore not fully satisfying, but its main merit is its simplicity, and analyzing it provides interesting qualitative insights on the dynamics

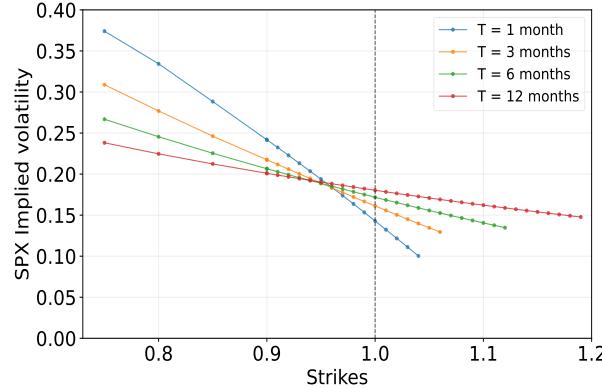
$$(4.3) \quad d\sigma_t = \left( -\beta_1 \lambda_1 R_{1,t} + \frac{\beta_2 \lambda_2 \sigma_t^2 - R_{2,t}}{2 \sqrt{R_{2,t}}} \right) dt + \beta_1 \lambda_1 \sigma_t dW_t$$

$\beta_0$	$\beta_1$	$\lambda_1$	$\beta_2$	$\lambda_2$
0.08	-0.08	62	0.5	40

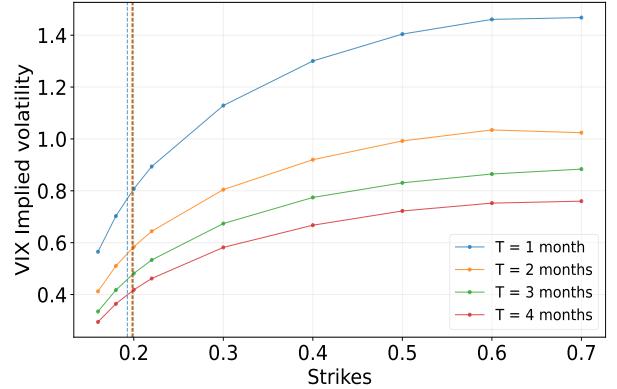
TABLE 7. Parameters used for the simulation of the 2-factor Markovian PDV Model

of the volatility

$$(4.4) \quad \sigma_t = \beta_0 + \beta_1 R_{1,t} + \beta_2 \sqrt{R_{2,t}}.$$



(A) SPX smiles for different maturities



(B) VIX smiles for different maturities

FIGURE 4.1. SPX smiles and VIX smiles in the 2-factor PDV model using the parameters in Table 7 and initial values  $R_{1,0} = -0.044$ ,  $R_{2,0} = 0.007$ . We use 100,000 paths to compute SPX option prices; 20,000 paths and 5,000 nested paths to compute VIX option prices via nested Monte Carlo; and 5 time steps per day

4.1.2. *The volatility of volatility.* Model (4.2) naturally produces a *multiplicative* dynamics for the volatility. Indeed, (4.3) expresses  $\sigma_t$  as a multiplicative Brownian motion with drift: the instantaneous (lognormal) volatility of the instantaneous volatility is constant, equal to  $|\beta_1 \lambda_1|$ . A constant vol of vol might seem inconsistent with the positive VIX skew observed in the market. In fact, Figure 4.1 shows that, when it is combined with a large mean reversion of volatility (as observed empirically), a constant vol of vol can actually produce a pronounced positive VIX skew. An explanation for this is given in Section 4.2 when we consider the 4-factor PDV Markovian model.

*Remark 2.* In order to add a degree of freedom for handling the VIX skew, one could slightly modify our model to get a supermultiplicative process, i.e., an instantaneous (lognormal) volatility of the instantaneous volatility that increases with the instantaneous volatility. For instance one may add a linear factor  $R_{1,t}^2$ , even though the data does not seem to suggest that. Our model would thus read like (4.1) but with

$$(4.5) \quad \sigma(R_1, R_2) = \beta_0 + \beta_1(R_1 - \bar{r})^2 + \beta_2 \sqrt{R_2}, \quad \beta_0, \beta_1, \beta_2, \bar{r} > 0.$$

In this case, the dynamics of  $\sigma_t$ , up to drift terms, reads

$$(4.6) \quad d\sigma_t = \dots dt + 2\beta_1 \lambda_1 (R_{1,t} - \bar{r}) \sigma_t dW_t = \dots dt \pm 2\beta_1 \lambda_1 \sqrt{\sigma_t - \beta_0 - \beta_2 \sqrt{R_{2,t}}} \sigma_t dW_t.$$

Interestingly, when  $\beta_2 = 0$ ,  $\beta_0$  represents the minimum value of the volatility, and the functional shape of the (lognormal) volatility of volatility  $\sigma \mapsto \sqrt{\sigma - \beta_0}$ , a shifted square root, resembles a lot the increasing and concave shape of VIX smiles; the same is true on average when  $\beta_2 > 0$ . However, again, Figure 4.1 shows that adding this extra degree of freedom is not necessary to generate a strong positive VIX skew.

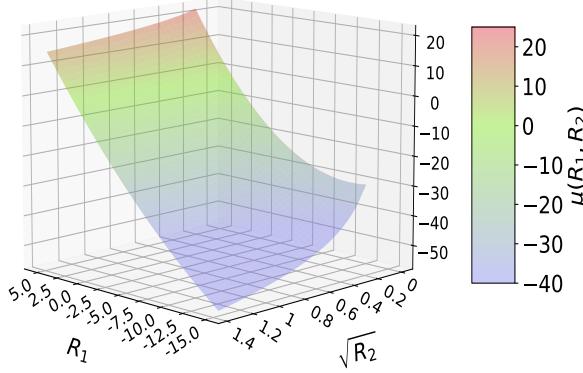


FIGURE 4.2. 3D surface of the drift  $\mu(R_1, R_2)$  of the instantaneous volatility in the 2-factor Markovian PDV model

4.1.3. *The drift of volatility: volatility clustering via mean reversion.* Even though Model (4.2) is simple, the drift of volatility is already quite complex in that, by contrast with the volatility of volatility, it does not only depend on the volatility itself, but on its components  $R_1$  and  $R_2$  separately, unless  $\beta_1 = 0$  or  $\beta_2 = 0$ .

In order to build some intuition on this drift, let us first consider the two extreme cases  $\beta_2 = 0$  and  $\beta_1 = 0$ . When  $\beta_2 = 0$ , the dynamics of  $\sigma_t = \beta_0 + \beta_1 R_{1,t}$  simply reads as the traditional autonomous diffusion

$$d\sigma_t = \lambda_1(\beta_0 - \sigma_t) dt + \beta_1 \lambda_1 \sigma_t dW_t.$$

The instantaneous volatility  $\sigma_t$  mean reverts towards  $\beta_0$  with constant speed  $\lambda_1$ . When  $\beta_1 = 0$ , the dynamics of  $\sigma_t = \beta_0 + \beta_2 R_{2,t} \geq \beta_0$  simply reads as the ordinary differential equation

$$d\sigma_t = \frac{\lambda_2(1-\beta_2^2)}{2} \frac{\sigma_t - \gamma}{\sigma_t - \beta_0} (\sigma^* - \sigma_t) dt, \quad \gamma := \frac{\beta_0}{1+\beta_2} < \beta_0 < \frac{\beta_0}{1-\beta_2} =: \sigma^*.$$

The instantaneous volatility  $\sigma_t$  mean reverts (converges) towards  $\sigma^*$  with the volatility-dependent speed  $\frac{\lambda_2(1-\beta_2^2)}{2} \frac{\sigma_t - \gamma}{\sigma_t - \beta_0}$ . Therefore, in both cases, the volatility mean reverts.

In the general case where  $\beta_1 \neq 0$  and  $\beta_2 \neq 0$ , the drift  $\mu_t$  of  $\sigma_t$  is not simply a function of  $\sigma_t$  but a function of the two factors  $R_{1,t}$  and  $R_{2,t}$ ,  $\mu_t = \mu(R_{1,t}, R_{2,t})$ , with

$$(4.7) \quad \mu(R_1, R_2) := -\beta_1 \lambda_1 R_1 + \frac{\beta_2 \lambda_2}{2} \frac{(\beta_0 + \beta_1 R_1 + \beta_2 \sqrt{R_2})^2 - R_2}{\sqrt{R_2}}.$$

Figure 4.2 shows the graph of the function  $\mu$ . The bottom right plot of Figure C.1 shows the scatter plot of  $\mu_t$  vs  $\sigma_t$  for the parameters in Table 7. The scatter plot shows that the model naturally generates volatility clustering via a clear trend of mean reversion. Rather than *postulating* it, like most continuous-time SV models do, Model (4.2) thus offers an *explanation* for the mean reversion of volatility.

Figure C.1 illustrates the rich drift dynamics on one sample path. Assume that the asset price falls abruptly after quiet times; so does  $R_{1,t}$ ;  $R_{2,t}$  increases, and the volatility  $\sigma_t$  spikes “doubly” due to the changes in  $R_1$  and  $R_2$ , since  $\beta_1 < 0$  and  $\beta_2 > 0$ . Due to the fast mean reversion of  $R_1$  towards zero, the volatility tends to decrease quite fast: the first part  $\mu_{1,t} := -\beta_1 \lambda_1 R_{1,t}$  of the drift (4.7) is negative and large. However, the spike in volatility means that  $\sigma_t^2$  exceeds its moving average  $R_{2,t}$  ( $\sigma_t^2$  increases faster than its moving average), so the second part  $\mu_{2,t} := \frac{\beta_2 \lambda_2}{2} \frac{\sigma_t^2 - R_{2,t}}{\sqrt{R_{2,t}}}$  of the drift (4.7) is initially positive and “fights against” the first term (as long as  $\sigma_t^2 > R_{2,t}$ ). This explains why after a very rapid volatility spike the volatility (decays but) can stay quite high for some time, even if the asset price recovers quickly: the memory of past volatility (squared returns) competes with the memory of (signed) past returns. This pattern of volatility increasing

	$\beta_0$	$\beta_1$	$\beta_{1,2}$	$\lambda_{1,0}$	$\lambda_{1,1}$	$\theta_1$	$\beta_2$	$\lambda_{2,0}$	$\lambda_{2,1}$	$\theta_2$
Parameter set 1 ( <i>Historical</i> )	0.04	-0.13	—	55	10	0.25	0.65	20	3	0.5
Parameter set 2 ( <i>Implied</i> )	0.006	-0.157	0.078	70	30.5	0.21	0.683	10.6	5.2	0.7
Parameter set 3	0.2	-0.023	—	70	25	0.51	0.11	40	20	0.5

TABLE 8. Various sets of parameters of the 4-factor Markovian PDV Model. The parameter  $\beta_{1,2}$  is introduced in (4.19)

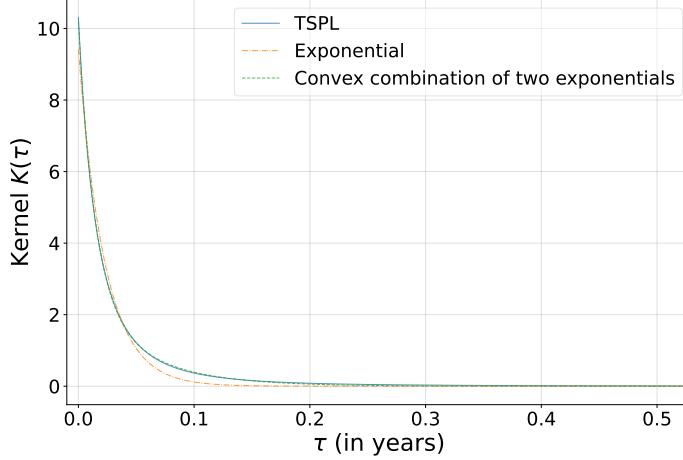


FIGURE 4.3. TSPL Kernel  $K_1$  for the prediction of the RV<sup>SPX</sup> and the best fits by one exponential and by a convex combination of exponentials. The convex combination is a much better fit

very fast (due to the combined effect of  $R_1$  and  $R_2$  when the asset price drops) and decreasing more slowly (due to the opposite effects of  $R_1$  and  $R_2$  when the market recovers) is well illustrated in the four plots of Figure C.1. It is also well illustrated by Figure D.1 in the case of the 4-factor PDV model (see Section 4.2).

**4.1.4. Price-path-dependence of volatility dynamics: strong Zumbach effect.** Note that (4.3) describes volatility dynamics that are *price-path-dependent*: the drift of  $\sigma_t$  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  $R_{1,t}$ . This has been coined the “strong Zumbach effect” in [14, 20]. Models that ignore the feedback effect of the trend  $R_{1,t}$  and capture the feedback effect of past returns only through  $R_{2,t}$  fail to produce strong Zumbach effect since  $R_{2,t}$  is a function of just the past values  $(\sigma_u)_{u \leq t}$ . Note, however, that in our continuous-time model (4.1), the diffusion coefficient of  $\sigma_t$  depends only on  $\sigma_t$ ; and for the quadratic version (4.5), the diffusion coefficient of  $\sigma_t$  depends on all the past values  $(\sigma_u)_{u \leq t}$  through  $\sigma_t$  and  $R_{2,t}$ , as (4.6) shows. In this last case, while the drift of volatility is price-path-dependent, the volatility of volatility is merely path-dependent: it can be written as a function of just the past values  $(\sigma_u)_{u \leq t}$  of the volatility itself.

**4.1.5. Nonnegativity of volatility.** From (4.3), if  $\sigma_t$  reaches zero, its (normal) volatility  $\beta_1 \lambda_1 \sigma_t$  vanishes, and since  $\sigma_t = \beta_0 + \beta_1 R_{1,t} + \beta_2 \sqrt{R_{2,t}} = 0$ , (i)  $R_{1,t} > 0$ , since  $\beta_0, \beta_2 > 0$  and  $\beta_1 < 0$ , and (ii) the drift of  $\sigma_t$  is equal to

$$(4.8) \quad -\beta_1 \lambda_1 R_{1,t} - \frac{\beta_2 \lambda_2}{2} \sqrt{R_{2,t}} = -\beta_1 \lambda_1 R_{1,t} + \frac{\lambda_2}{2} (\beta_0 + \beta_1 R_{1,t}) = \frac{\beta_0 \lambda_2}{2} + \beta_1 R_{1,t} \left( \frac{\lambda_2}{2} - \lambda_1 \right).$$

Therefore, if  $\lambda_2 < 2\lambda_1$ , the drift of  $\sigma_t$  is positive, which pushes  $\sigma_t$  back up: the volatility stays nonnegative. Market data shows that  $K_1$  and  $K_2$  decrease approximately at the same speed, which means in this simple Markovian version that  $\lambda_2 \approx \lambda_1$ : the sufficient condition for nonnegativity,  $\lambda_2 < 2\lambda_1$ , is thus satisfied.

**4.2. A better Markovian approximation: the 4-factor Markovian PDV model.** While a TSPL kernel  $\tau \mapsto Z_{\alpha,\delta}^{-1}(\tau + \delta)^{-\alpha}$  cannot be well approximated by a single exponential kernel, Figure 4.3 illustrates

that it can be very well approximated over a large range of lags  $\tau$  by a convex combination of two exponential kernels,

$$(4.9) \quad K_{\theta, \lambda_0, \lambda_1} : \tau \mapsto (1 - \theta)\lambda_0 e^{-\lambda_0 \tau} + \theta\lambda_1 e^{-\lambda_1 \tau}, \quad \text{where } \theta \in [0, 1], \lambda_0 > \lambda_1 > 0.$$

The very large weights given to recent returns are captured by a very large  $\lambda_0$ , while the long memory, i.e., the slow decay of the weights for large lags  $\tau$ , is produced by a small  $\lambda_1$ ;  $\theta$  is a mixing factor. Note that when  $\theta = 0$ , the kernel is simply  $K^{(\lambda_0)}$ , and when  $\theta = 1$ , the kernel is simply  $K^{(\lambda_1)}$ , hence the 0–1 notation.

We therefore 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  $K_2$ , respectively, and, for  $n \in \{1, 2\}$  and  $j \in \{0, 1\}$ , we denote

$$(4.10) \quad R_{n,j,t} := \int_{-\infty}^t \lambda_{n,j} e^{-\lambda_{n,j}(t-u)} \left( \frac{dS_u}{S_u} \right)^n.$$

This yields the following Markovian version of Model (4.1):

$$(4.11) \quad \begin{aligned} \frac{dS_t}{S_t} &= \sigma_t dW_t \\ \sigma_t &= \sigma(R_{1,t}, R_{2,t}) \\ \sigma(R_1, R_2) &= \beta_0 + \beta_1 R_1 + \beta_2 \sqrt{R_2} \\ R_{1,t} &= (1 - \theta_1)R_{1,0,t} + \theta_1 R_{1,1,t} \\ R_{2,t} &= (1 - \theta_2)R_{2,0,t} + \theta_2 R_{2,1,t} \\ dR_{1,j,t} &= \lambda_{1,j} \left( \frac{dS_t}{S_t} - R_{1,j,t} dt \right) = \lambda_{1,j} (\sigma(R_{1,t}, R_{2,t}) dW_t - R_{1,j,t} dt), \quad j \in \{0, 1\} \\ dR_{2,j,t} &= \lambda_{2,j} \left( \left( \frac{dS_t}{S_t} \right)^2 - R_{2,j,t} dt \right) = \lambda_{2,j} (\sigma(R_{1,t}, R_{2,t})^2 - R_{2,j,t}) dt, \quad j \in \{0, 1\}. \end{aligned}$$

Note that the vectors  $(R_{1,0,t}, R_{1,1,t}, R_{2,0,t}, R_{2,1,t})$  and  $(S_t, R_{1,0,t}, R_{1,1,t}, R_{2,0,t}, R_{2,1,t})$  are Markovian, in dimensions 4 and 5 respectively. Therefore we call this model the *4-factor Markovian PDV model*. This makes Model (4.11) an easy-to-simulate version of Model (4.1) with kernels

$$\begin{aligned} K_1(\tau) &= (1 - \theta_1)\lambda_{1,0} e^{-\lambda_{1,0}\tau} + \theta_1\lambda_{1,1} e^{-\lambda_{1,1}\tau} \\ K_2(\tau) &= (1 - \theta_2)\lambda_{2,0} e^{-\lambda_{2,0}\tau} + \theta_2\lambda_{2,1} e^{-\lambda_{2,1}\tau} \end{aligned}$$

that possess the three most important kernel features inferred from data: finiteness at  $\tau = 0$ , fast decay for small lags  $\tau$ , and slow decay for large lags.

The nine parameters of the model have the following roles:

- $\lambda_{1,0}$  captures the dependence of  $R_1$  on recent returns; the higher  $\lambda_{1,0}$ , the larger the weight of recent returns;
- $\lambda_{1,1} < \lambda_{1,0}$  captures the dependence of  $R_1$  on older returns; the smaller  $\lambda_{1,1}$ , the larger the weight of old returns;
- $\lambda_{2,0}$  captures the dependence of  $R_2$  on recent squared returns; the higher  $\lambda_{2,0}$ , the larger the weight of recent squared returns;
- $\lambda_{2,1} < \lambda_{2,0}$  captures the dependence of  $R_2$  on older squared returns; the smaller  $\lambda_{2,1}$ , the larger the weight of old squared returns;
- $\theta_1$  and  $\theta_2$  mix the dependence on recent and older returns (or squared returns) to form, respectively, the summary random variables  $R_1$  and  $R_2$ .
- $\beta_1$  and  $\beta_2$  are the (linear) sensitivities of the volatility to  $R_1$  and  $\sqrt{R_2}$ , respectively, and  $\beta_0$  the constant term of the linear model.

The dynamics of the instantaneous volatility reads

$$(4.12) \quad d\sigma_t = \beta_1 ((1 - \theta_1)\lambda_{1,0} + \theta_1\lambda_{1,1}) \sigma_t dW_t + \left\{ -\beta_1 ((1 - \theta_1)\lambda_{1,0}R_{1,0,t} + \theta_1\lambda_{1,1}R_{1,1,t}) \right. \\ \left. + \frac{\beta_2}{2} \frac{((1 - \theta_2)\lambda_{2,0} + \theta_2\lambda_{2,1}) \sigma_t^2 - ((1 - \theta_2)\lambda_{2,0}R_{2,0,t} + \theta_2\lambda_{2,1}R_{2,1,t})}{\sqrt{R_{2,t}}} \right\} dt$$

and satisfies similar qualitative properties as dynamics (4.3):

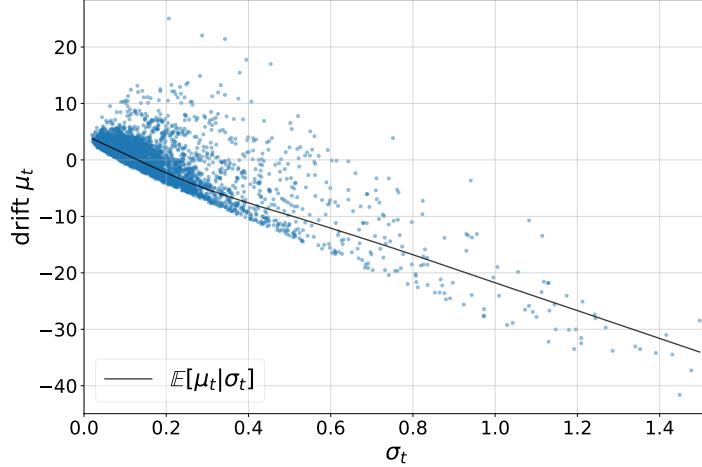


FIGURE 4.4. Drift of the volatility  $\mu_t$  vs  $\sigma_t$  for the 4-factor PDV model using Parameter set 1 in Table 8

- The drift of  $\sigma_t$  produces volatility clustering via a clear trend of mean reversion of volatility, as illustrated by Figure 4.4.
- The lognormal volatility of  $\sigma_t$  is constant.
- The dynamics of  $(\sigma_t)$  are price-path-dependent: the drift of  $\sigma_t$  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  $R_{1,1,t}$ .

Our simulations (see Section 4.4) show that for realistic parameter values, the volatility in the 4-factor PDV model always stays nonnegative. Like for the 2-factor model (see Remark 2), the quadratic-in- $R_1$  extension (4.5) of the model leads to a lognormal volatility of volatility of the form  $\sqrt{\sigma - \beta_0 - \beta_2 \sqrt{R_2}}$ , similar in average to the shape of VIX smiles. However, Figures 4.7 and 4.8 show that, due to the large mean reversion of volatility, Model (4.11) generates positive VIX skew and can even calibrate to market VIX smiles very accurately. In particular, the very large mean reversion makes it very unlikely that the VIX take values smaller than, say, 10%, even though the instantaneous volatility can take much lower values (see Figure D.1). This naturally produces a VIX implied volatility that almost vanishes for strikes below, say, 10%, and thus a very positive VIX skew.

Note that a constant lognormal volatility of volatility is rather in line with market data: Figure 4.5 shows that VIX 1-day returns are more stochastically constant than SPX 1-day returns. VIX 1-day returns also tend to be larger in absolute value: the vol of vol is larger than the volatility itself. In our 4-factor PDV model, this translates into the fact that  $|\beta_1|((1 - \theta_1)\lambda_{1,0} + \theta_1\lambda_{1,1})$  is large. Note that  $\lambda_{1,0}$  is typically very large, say, in [40, 100], to account for the very large fast mean reversion.

**4.3. Comparison with some SV models.** It is instructive to compare our model with the path-dependent versions of popular SV models, with spot-vol correlations set to  $-1$ . Where  $\sigma_t > 0$ , the path-dependent version of the Heston dynamics reads

$$(4.13) \quad d\sigma_t = -\frac{\lambda(\sigma_t^2 - m) + \frac{\omega^2}{4}}{2\sigma_t} dt - \frac{\omega}{2} dW_t.$$

Therefore the lognormal volatility of volatility  $\frac{\omega}{2\sigma_t}$  decreases with  $\sigma_t$ , in contrast with the increasing smiles of volatility of volatility observed in the market. Moreover, the mean reversion of  $\sigma_t^2$  to the constant  $m$  is merely postulated, and no strong Zumbach effect is captured.

Closer to our model are (the PDV versions of) the Bergomi and rough Bergomi models. Both the Bergomi and rough Bergomi models postulate that the instantaneous volatility  $\sigma_t = f(t)e^{-\frac{\omega}{2}X_t}$  is a nonlinear function (an exponential) of a random variable of the type

$$(4.14) \quad X_t := \int_0^t K(t-u) dW_u$$

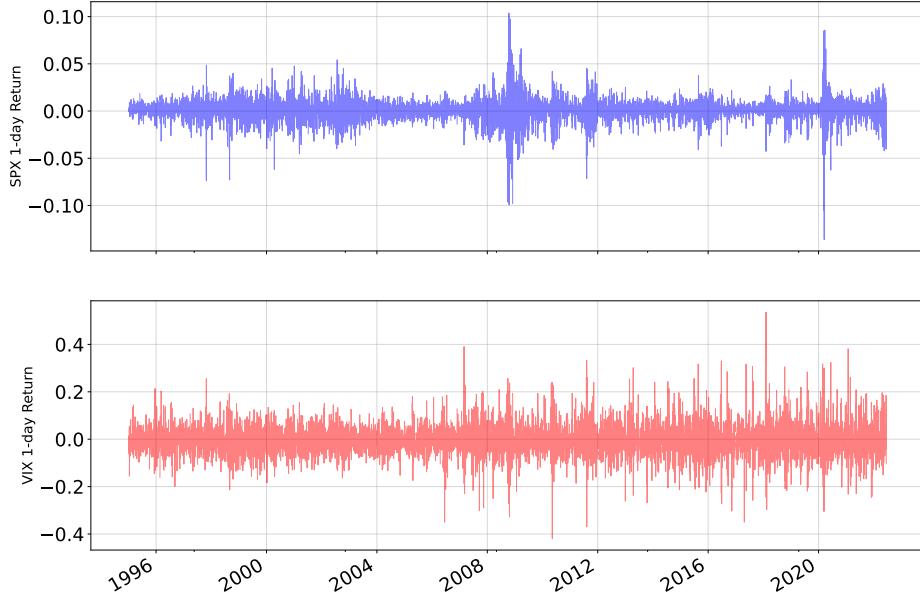


FIGURE 4.5. SPX (top) and VIX (bottom) daily returns, Jan 1, 1995–May 15, 2022. Note the different scales on the  $y$ -axis

with  $K$  a single exponential kernel for the one-factor Bergomi model, a linear combination of two exponential kernels for the 2-factor Bergomi model, and a power-law kernel for the rough Bergomi model. Rather, in our model, if we momentarily ignore the  $R_2$  term ( $\beta_2 = 0$ ), we write  $\sigma_t$  as a linear function of a random variable of the type

$$(4.15) \quad R_{1,t} := \int_{-\infty}^t K_1(t-u) \frac{dS_u}{S_u}.$$

That is, we explain volatility *linearly* by an average of *observable* quantities (the past returns  $dS_u/S_u$ ) rather than *nonlinearly* by an average of the *unobservable* quantities  $dW_u$ —only the product  $\sigma_u dW_u$  is observed;  $\sigma_u$  and the innovation  $dW_u$  are not separately observed. *Our observation-driven model thus reflects that traders can only respond to what they can observe, and that they do it in a linear rather than nonlinear way.* While both our approach and the Bergomi (exponential Ornstein-Uhlenbeck) approach lead to a constant lognormal volatility of volatility, they are thus different in spirit and lead to different volatility drifts, even when  $\beta_2 = 0$ . In particular, in our model the drift of the instantaneous volatility depends explicitly on the past returns  $dS_u/S_u$ , which is not the case in the Bergomi model.<sup>5</sup>

Another important difference between our model and the Bergomi and rough Bergomi models is that, in the spirit of GARCH models, our model captures the feedback effect of past *squared* returns onto volatility via the  $R_2$  (or  $\Sigma$ ) term. Since  $\beta_2 > 0$ , this term generates volatility clustering, which materializes in an extra term in the drift of  $\sigma_t$  which can be quite large (see Figure C.1 in the case of the 2-factor model), impacts the dynamics of the volatility by mitigating the mean reversion of  $R_1$  (slow decay of volatility after it spikes), and is absent of all popular continuous-time SV models.

Let us finally compare our model to more recent continuous-time PDV models. The QRHM [20] postulates that

$$(4.16) \quad \begin{aligned} \sigma_t &= \sqrt{a(Z_t - b)^2 + c} \\ Z_t &= \int_0^t K(t-u)\theta_0(u) du + \eta \int_0^t K(t-u) \frac{dS_u}{S_u} \end{aligned}$$

<sup>5</sup>A similar comparison with the dynamics of instantaneous volatility in the rough Bergomi model is not directly possible since in this model the instantaneous volatility is not a semimartingale.

where  $a, b, c > 0$ ,  $\theta_0$  is a deterministic function, and  $K$  is the Mittag-Leffler function. When  $\theta_0 \equiv 0$ , if we denote  $K_1 := \eta K$ , we have  $Z_t = R_{1,t}$ , up to the additive constant  $\int_{-\infty}^0 K_1(t-u) \frac{dS_u}{S_u}$ , and the model reads as Model (4.5) with  $\beta_2 = 0$  and  $\sigma_t$  replaced by  $\sigma_t^2$ . In the general case where  $\theta_0 \not\equiv 0$ ,  $Z_t$  reads as  $R_{1,t} + \varphi(t)$  for a deterministic function  $\varphi$ , i.e.,  $b = b(t)$  is allowed to be time-dependent. Since in this work we look for a persistent, stationary, time-independent relationship between past returns and volatility, we do not allow time-dependent parameters in our model. The main differences between Model (4.16) and our generic model (4.1) is that, based on our empirical study in Section 3, (a) we include the feedback effect from past squared returns (or past volatility) via  $R_2$ , (b) we do not include an  $R_1^2$  factor, and (c) the linear combination of factors defines the volatility rather than the variance. Note that an  $R_1^2$  term is not needed to generate positive VIX skew. Our model generates positive VIX skew despite the constant vol of vol (see Figures 4.7 and 4.8).<sup>6</sup>

A similar model, one particular threshold version of the EWMA Heston model of Parent [30], postulates that

$$(4.17) \quad \sigma_t = a(b - R_{1,t})_+ + c$$

with  $K_1$  a single exponential kernel. This model is a similar, asymmetric version of (4.16), with a kernel that does not account for the long memory of volatility, and our above remarks apply as well. Note that in this model the volatility distribution has an unrealistic Dirac mass at  $\sigma_t = c$ .

Finally, the low-frequency limit of the ZHawkes model of [4, Section 4.2] corresponds to

$$(4.18) \quad \sigma_t = \sqrt{\beta_0 + \beta_1 R_{1,t}^2 + \beta_2 R_{2,t}}$$

with (single) exponential kernels  $K_1$  and  $K_2$ . While this model is similar to ours in spirit, our empirical study (Section 3) has shown that the path-dependence of volatility is better captured by  $\sigma_t = \beta_0 + \beta_1 R_{1,t} + \beta_2 \sqrt{R_{2,t}}$ , and in particular that (a) a linear term in  $R_{1,t}$  must be included to account for the leverage effect, and (b) no parabolic term in  $R_{1,t}$  is needed when the  $R_{2,t}$  factor is present. Moreover, our model captures both the short and long time scales of the feedback effect of price returns on volatility, which single exponential kernels cannot do.

**4.4. Sample paths.** In this section we show that the 4-factor PDV model generates very realistic financial time series for both price and volatility. Figure 4.6 shows the actual time series of the SPX and the VIX as well as an example of a sample path in the model. The VIX in the model is computed using 5,000 nested Monte Carlo paths, and 10 time steps per business day. (The path of the instantaneous volatility is reported in Figure D.1 in the appendix. Note that the instantaneous volatility takes much lower and much higher values than the VIX, due to the large fast mean reversion of the instantaneous volatility.)<sup>7</sup>

The similarity between the actual and simulated time series is particularly striking. In particular:

- There is a succession of periods of positive price trends with low volatility and periods of volatility bursts accompanying big drops in price.
- Within periods of positive price trends, there are many smaller drops in price and simultaneous smaller VIX spikes.
- Volatility has a lot of variability and tends to cluster.
- As described in [7], “volatility fluctuations have no clear characteristic time scale; volatility bursts can last anything between a few hours and a few years.”
- Both the price and volatility series are not invariant upon time reversal.

A particularly important point is the following. While the time reversal asymmetry is obvious in price series, our model is also able to reproduce the more subtle (but still clear to the naked eye) time reversal asymmetry observed in volatility series: whatever the spike amplitude, volatility spikes very fast, and then tends to decrease more slowly; this is particularly clear after large volatility spikes. By contrast, the volatility spikes in the EWMA Heston model look time reversal symmetric, as illustrated in the sample path presented in [30, Figure 4]. And the QRHM sample paths in [20, Figure 2] shows a VIX that, after large spikes,

<sup>6</sup>Note that in [20] the authors justify that their model produces positive VIX skew by looking at the case where  $K$  is an exponential kernel, applying the Itô formula, and claiming that, when  $c = 0$ , “when the volatility is high, the volatility of volatility is also high.” In fact, in Model (4.16), for any non-blown-up kernel  $K$ , the (lognormal) volatility of volatility is constant (in absolute value) when  $c = 0$ ; in this model, it is actually when  $c > 0$  that the absolute lognormal volatility of volatility, proportional to  $\sqrt{1 - c/\sigma^2}$ , increases with the volatility  $\sigma$  in a concave way.

<sup>7</sup>In all our simulations, we actually cap the instantaneous volatility at 1.5.

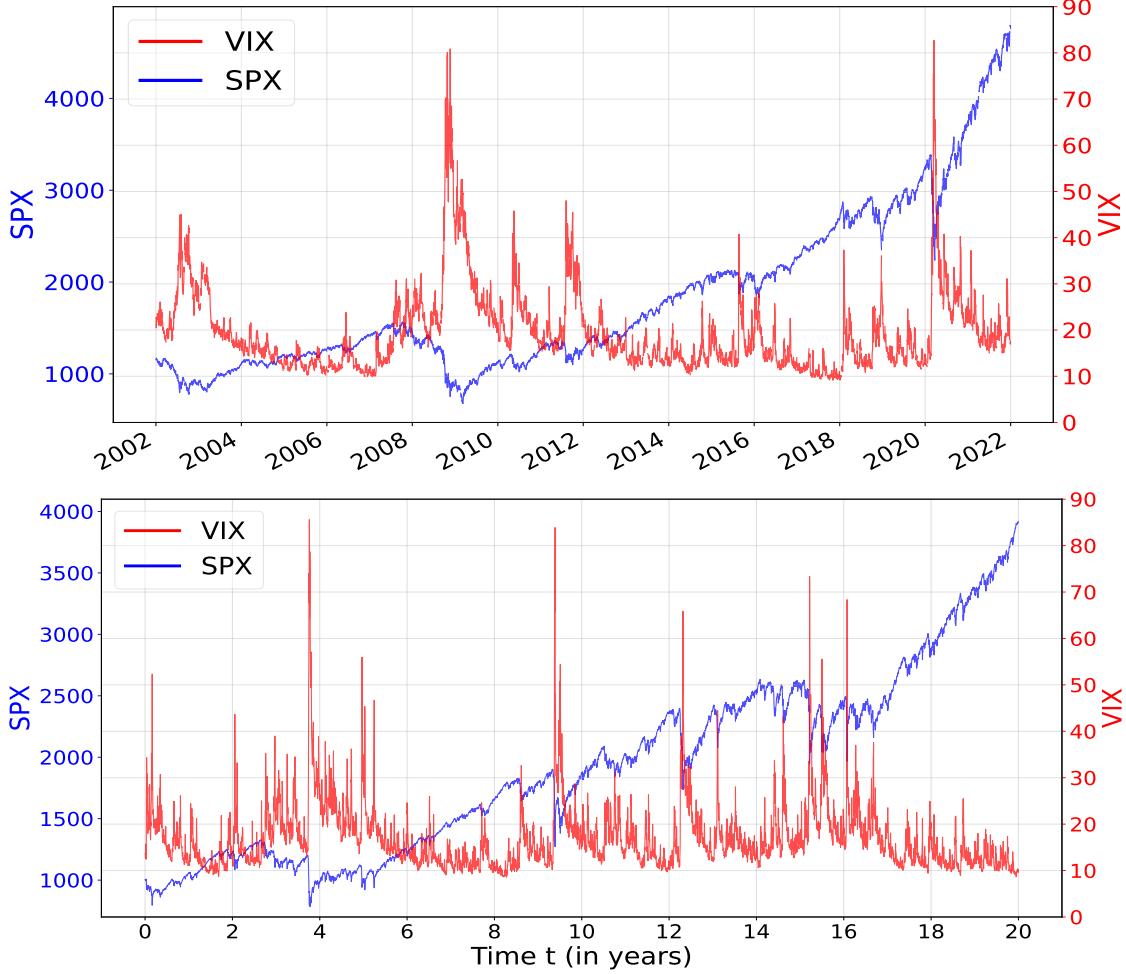


FIGURE 4.6. SPX and VIX time series. Top: actual time series from Jan 1, 2002 to Jan 1, 2022. Bottom: simulation over 20 years of a path in the 4-factor PDV model using Parameter set 1 from Table 8 and initial values  $R_{1,0,0} = 0.078$ ,  $R_{1,1,0} = 0.16$ ,  $R_{2,0,0} = 0.074$ ,  $R_{2,1,0} = 0.016$ . We use 10 time steps per day, and 5,000 nested paths to compute the VIX

decreases faster than in the market (pay a particular attention to the different time and VIX scales in [20, Figures 1 and 2]). Note that our model is better equipped to reproduce this time reversal asymmetry and slow VIX decay because it incorporates the important  $R_2$  factor, which captures the feedback of past squared returns onto volatility and is absent of both the QRHM and the EWMA Heston model.

Interestingly, our model generates a jumpy behavior for the instantaneous volatility and the VIX without including any actual jumps in the dynamics, and, what is more, with a constant instantaneous volatility of instantaneous volatility. The jumpy behavior of the volatility is a consequence of the feedback effect: if the asset experiences a series of very negative random shocks  $dW_t$ , the volatility increases due to the  $R_1$  feedback term, and the later negative shocks  $dW_t$  come multiplied by an even larger volatility, possibly generating a market crash and sudden volatility spikes. The market can then quickly recover when positive random shocks  $dW_t$  are drawn, and come multiplied by a very large volatility.

Finally, in Appendix E we show that the 4-factor PDV model also captures very well the statistical relationship between  $R_1$ ,  $\Sigma$ , and the volatility (VIX or RV) observed in the market. This is clear when comparing Figures E.1 and E.2 with Figures 3.3 and 3.6, respectively.

**4.5. Joint calibration to SPX and VIX smiles.** Another remarkable property of the 4-factor PDV model is that it can generate very realistic smiles of both the SPX and the VIX. For instance, Figure 4.7 shows SPX

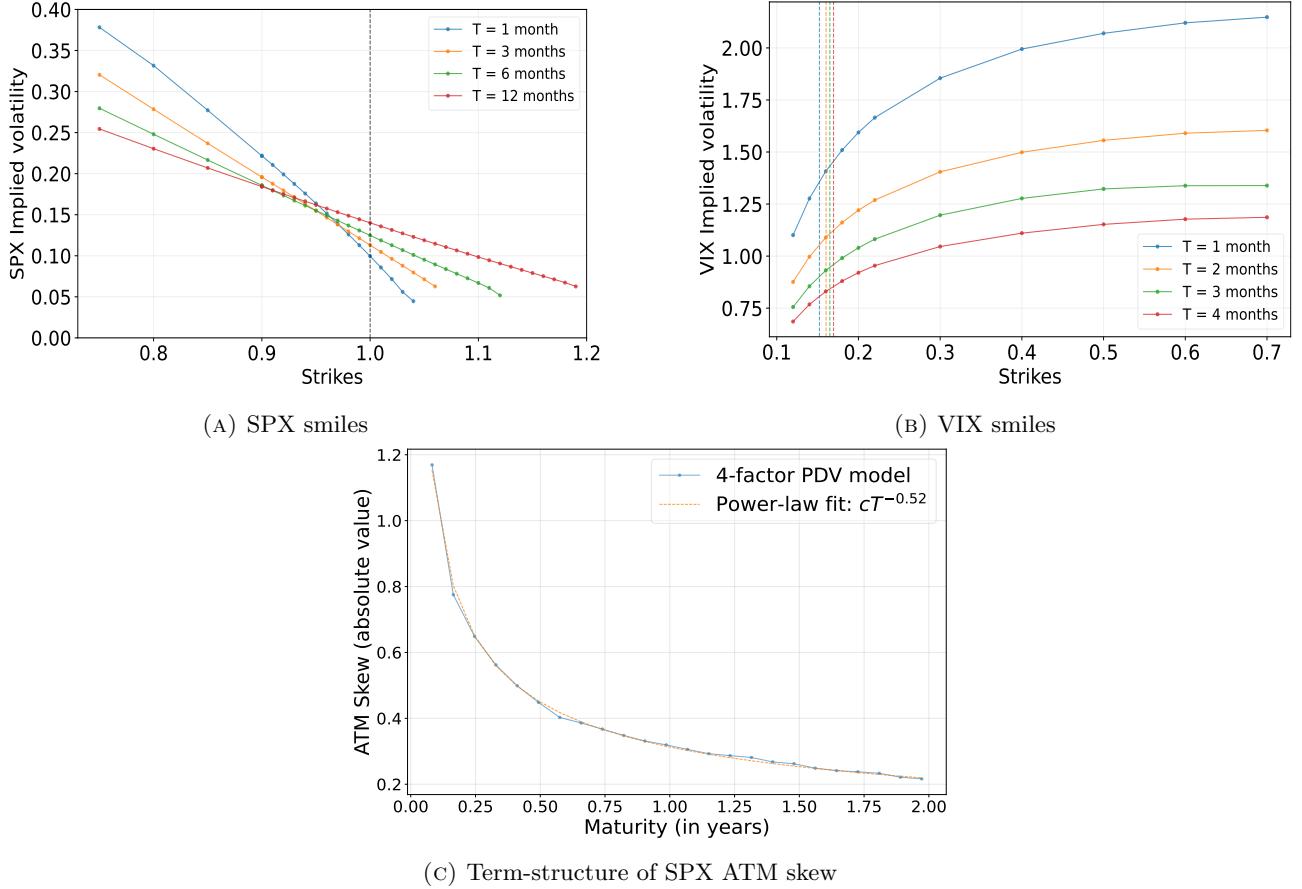


FIGURE 4.7. SPX smiles, VIX smiles and VIX futures, and term-structure of SPX ATM skew in the 4-factor PDV model using Parameter set 1 in Table 8 and initial values  $R_{1,0,0} = 0.168$ ,  $R_{1,1,0} = 0.244$ ,  $R_{2,0,0} = 0.005$ ,  $R_{2,1,0} = 0.03$ . We use 100,000 paths to compute SPX option prices; 20,000 paths and 5,000 nested paths to compute VIX option prices via nested Monte Carlo; and 10 time steps per day. VIX futures are represented with vertical lines on the top-right graph

smiles, VIX smiles, and the term-structure of SPX ATM skew produced by the 4-factor PDV model using Parameter set 1 in Table 8, i.e., the parameters that produced the very realistic sample paths in Figure 4.6.

In fact, we can even jointly calibrate the 4-factor PDV model to market SPX and VIX smiles with a very good accuracy. We thus show, for the first time, that a continuous-time *Markovian parametric* stochastic (actually, path-dependent) volatility model can practically solve the joint SPX/VIX smile calibration problem.

To be precise, we perform the joint calibration on a slightly modified version of the 4-factor PDV model, where the deterministic function  $(R_1, R_2) \mapsto \sigma(R_1, R_2)$  in (4.11) is replaced by

$$(4.19) \quad \sigma(R_1, R_2) = \beta_0 + \beta_1 R_1 + \beta_2 \sqrt{R_2} + \beta_{1,2} R_1^2 \mathbf{1}_{R_1 \geq 0}.$$

By adding a quadratic term in  $R_1^+$  we allow for a better calibration of the increasing SPX smile for out-the-money call options. In Figure 4.8, we have (for now, manually) calibrated the above version of the 4-factor PDV model to the first two monthly maturities of the SPX smile, the first monthly VIX future, and the first monthly maturity of the VIX smile as of June 2, 2021. It is remarkable that we can so accurately fit SPX and VIX smiles jointly (especially for such short maturities) with a parametric Markovian continuous-time model having only 10 parameters. The calibrating (or implied or risk-neutral) parameters are reported in Table 8. We call *risk-neutral PDV model* or *implied PDV model* or  $\mathbb{Q}$ -*PDV model* the 4-factor PDV model when it is fed with these implied parameters.

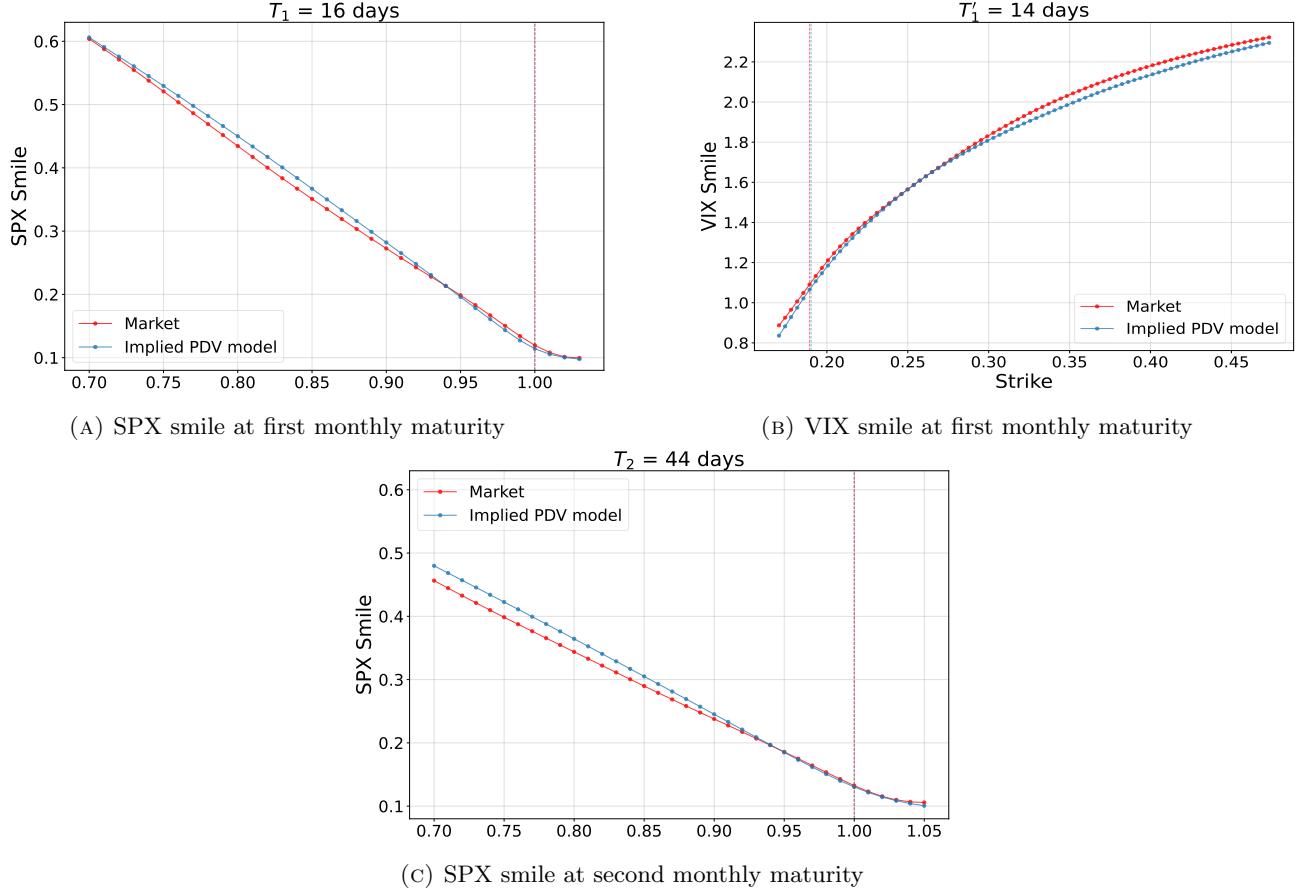


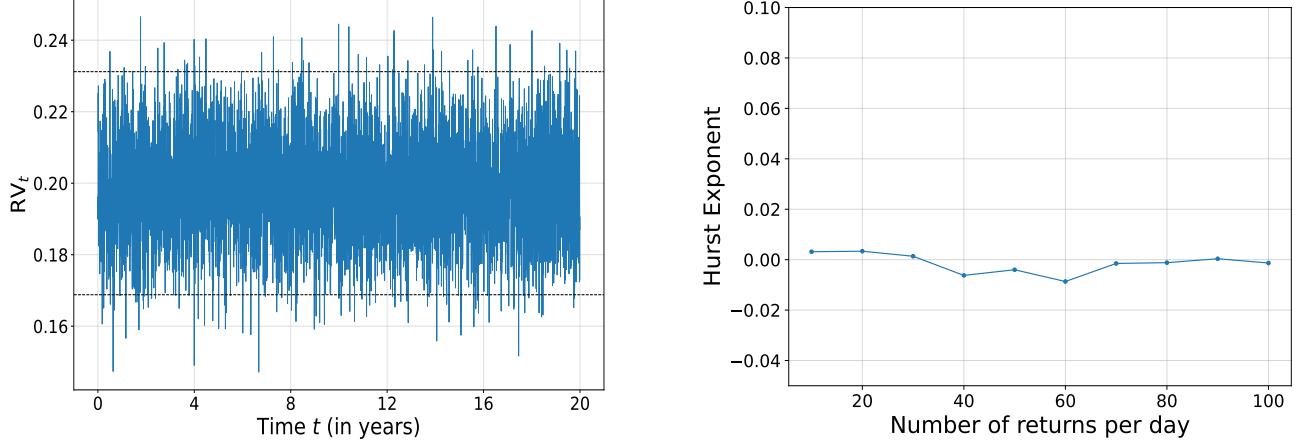
FIGURE 4.8. Calibration of the 4-factor PDV model to market smiles on June 2, 2021. We use 100,000 paths for the computation of SPX smile; 10,000 paths and 5,000 nested paths for the computation of the VIX smile; 10 time steps per day; and  $R_{1,0,0} = 0.303$ ,  $R_{1,1,0} = 0.235$ ,  $R_{2,0,0} = 0.016$ ,  $R_{2,1,0} = 0.02$ . Those initial values are the weighted averages of past returns and past squared returns on the calibration date. The parameters of the calibrated model (“implied PDV model”) are reported in Table 8 (Parameter set 2). Note that the market VIX future is very well calibrated, as the two vertical lines in the top-right graph coincide

Note that traditional parametric SV models, when calibrated to VIX smiles, cannot generate enough SPX ATM skew in absolute value [23, 24]. Interestingly, in our numerical example, the parameters are such that the 4-factor PDV model, while being well calibrated to the VIX smile, produces slightly too large SPX ATM skews in absolute value. This is a sign that a better, automatized calibration procedure should be able to even more precisely fit the SPX and VIX smiles, jointly.

**4.6. Spurious roughness.** While the volatility process ( $\sigma_t$ ) in our 4-factor PDV model is a Brownian diffusion with Hurst exponent  $H = 0.5$ , estimating  $H$  from a sample path using the methodology of [19] leads to an estimate  $\hat{H} < 0.5$ , due to the noisy estimation of  $\sigma_t$  by the realized volatility

$$(4.20) \quad \text{RV}_t := \sqrt{\frac{1}{n\delta t} \sum_{i=1}^n \ln \left( \frac{S_{t+i\delta t}}{S_{t+(i-1)\delta t}} \right)^2}$$

using  $n$  observed asset returns, each over a period of  $\delta t$ . The Oxford-Man institute database uses  $n = 78$  consecutive  $\delta t = 5$ -minute returns per trading day for the SPX index. To explain that the noise in  $\text{RV}_t$  may



(A) Time series of simulated  $RV_t$  over 20 years. The horizontal lines show the 95% confidence interval for the  $\sigma(1 + 0.08\mathcal{N}(0, 1))$  distribution

(B) Estimation  $\hat{H}$  of Hurst exponent as a function of the number of returns per day used in the estimation of the daily realized volatility

FIGURE 4.9. Time series of  $RV_t$  for the Black-Scholes model with volatility  $\sigma = 0.2$  and estimation of the Hurst exponent using the methodology of [19]

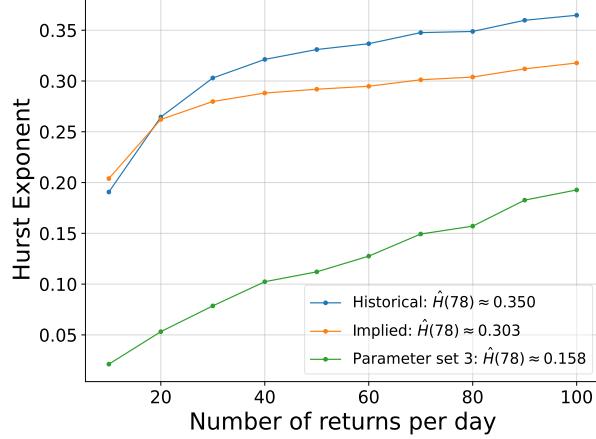


FIGURE 4.10. Estimation  $\hat{H}$  of Hurst exponent (using the methodology of [19]) as a function of the number of returns per day used in the estimation of the daily realized volatility for three sets of parameters for our 4-factor PDV model (see Table 8). The value displayed in the legend is the estimated Hurst exponent using  $n = 78$  returns per day, and 20 years of simulated data

cause spurious roughness, let us take an extreme example. Assume that  $\sigma_t \equiv \sigma$  is constant. Then

$$(4.21) \quad RV_t := \sqrt{\frac{1}{n\delta t} \sum_{i=1}^n \left( \sigma(W_{t+i\delta t} - W_{t+(i-1)\delta t}) - \frac{\sigma^2}{2}\delta t \right)^2}.$$

For two different trading days  $t$  and  $t'$ , the realized volatilities  $RV_t$  and  $RV_{t'}$  are thus independent random variables. This creates spurious roughness in the volatility process, as  $RV_t$  is then essentially a white noise process, for which  $\hat{H} \approx 0$ . In our example, by the central limit theorem,  $RV_t^2$  is approximately a Gaussian random variable with mean  $\sigma^2 + \frac{\sigma^4 \delta t}{4} \approx \sigma^2$  and standard deviation  $\sqrt{\frac{v}{n}}$  where  $v \approx 2\sigma^4$ . For two different trading days  $t$  and  $t'$ , the realized volatilities  $RV_t$  and  $RV_{t'}$  are thus independent random variables with

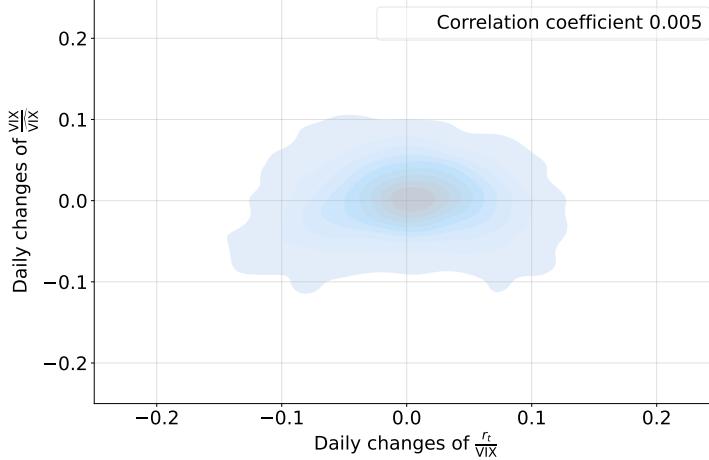


FIGURE 5.1. KDE plot of the daily change of the ratio residual of VIX (to approximate  $da_t$ ) against the daily change of the ratio between the daily SPX return and VIX (to approximate  $dW_t$ ), on the train set

approximate distribution  $\sqrt{\mathcal{N}(\sigma^2, \frac{2\sigma^4}{n})} \approx \sigma \left(1 + \frac{1}{\sqrt{2n}} \mathcal{N}(0, 1)\right)$ . For  $n = 78$ , this is approximately equal to  $\sigma(1 + 0.08 \mathcal{N}(0, 1))$ ;  $RV_t$  is then essentially a white noise process centered on  $\sigma$ , and we indeed numerically get  $\hat{H} \approx 0$ , see Figure 4.9.

In more realistic cases, the realized volatility  $RV_t$  is the true instantaneous volatility process  $\sigma_t$  plus some white noise as above, which adds some spurious roughness. Therefore the Hurst exponent  $\hat{H}$  estimated from  $(RV_t)$  is smaller than the true Hurst exponent  $H = 0.5$  of the instantaneous volatility process  $(\sigma_t)$ . Bolko et al. [5] recently made a similar observation and proposed a corrected estimator; see also [13]. Figure 4.10 shows  $\hat{H}$  as a function of the number  $n$  of returns used in the calculation of the daily realized volatility, with  $n\delta t = 1$  day, for three sets of parameters of the 4-factor PDV model (see Table 8). For  $n = 78$ , for Parameter sets 1 (historical) and 2 (implied), we obtain  $\hat{H} \approx 0.35$  and  $\hat{H} \approx 0.3$ , respectively. Note that with Parameter set 3, we reproduce the value of the SPX historical data  $\hat{H} \approx 0.15$  reported in [19].

## 5. PATH-DEPENDENT STOCHASTIC VOLATILITY

In order to account for the small exogenous part of volatility, we suggest the following path-dependent stochastic volatility (PDSV) model, which is similar to the 4-factor PDV model, except that the instantaneous volatility  $\sigma_t = a_t \sigma(R_{1,t}, R_{2,t})$  is now the product of the PDV  $\sigma(R_{1,t}, R_{2,t})$  and some stochastic volatility ( $a_t$ ):

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

Since the ratio residuals in Figure 3.3 seem to be stationary, mean revert to one, with a quite large volatility and mean reversion, it is natural to model  $(a_t)$  as an Ornstein-Uhlenbeck (OU) process

$$(5.2) \quad da_t = \kappa(1 - a_t) dt + \nu dZ_t$$

with  $Z$  a Brownian motion and  $\kappa, \nu$  positive and large. The value of  $\kappa$  may be inferred from the autocorrelation function of the ratio residuals;  $\frac{1}{\kappa}$  represents the characteristic time of mean reversion of those residuals. The value of  $\nu$  can then be chosen so that the stationary volatility  $\frac{\nu}{\sqrt{2\kappa}}$  of  $(a_t)$  is in line with the observed stationary volatility of the ratio residuals. Since the daily increments of the ratio residuals seem to be approximately uncorrelated with the daily (normalized) SPX returns (see Figure 5.1), it is reasonable to assume that the Brownian motions  $W$  and  $Z$  are independent. Another natural choice for  $(a_t)$  is an exponential-OU process:

$$(5.3) \quad a_t = e^{X_t}, \quad dX_t = -\kappa X_t dt + \nu dZ_t.$$

Note that, even though we model the instantaneous volatility, we prefer to suggest a dynamics for  $(a_t)$  based on the ratio residuals of the implied volatility, e.g., the VIX. This is because the implied volatility is a well-defined quantity, whereas the future daily realized volatility is so noisy that it is unsure how much noise in the ratio residuals comes from the exogenous part of volatility, and how much comes from the daily realized volatility itself. However, since the instantaneous volatility tends to mean revert faster and fluctuate more than the implied volatility, the values of  $\kappa$  and  $\nu$  could be chosen larger than those directly read from the ratio residuals of the implied volatility.

With those choices of  $(a_t)$ , the PDSV model (5.1) is a five-factor (six-dimensional) Markovian model that is very easy and quick to simulate and whose 11 parameters (the 9 parameters of the 4-factor PDV model, plus  $\kappa$  and  $\nu$ ) have a very clear interpretation. In this model, the VIX is a deterministic function of  $R_{1,0,t}, R_{1,1,t}, R_{2,0,t}, R_{2,1,t}$  and  $a_t$  (or  $X_t$ ) that can be estimated by nested Monte Carlo or Least Squares Monte Carlo.

## 6. CONCLUSION

In this article, we have shown empirically that the volatility of equity markets is mostly path-dependent, endogenous. We have considered both implied and future realized volatility, and revealed that a simple linear model linking the volatility to the recent trend in the asset price (a weighted sum  $R_1$  of past daily returns) and the recent realized volatility (the square root  $\Sigma = \sqrt{R_2}$  of a weighted sum of past daily squared returns), with different time-shifted power-law weights, explains a large part of the observed variability in the volatility. Over a very challenging data set that covers the Covid-19 crisis and its aftermath (2019–2022), we obtain test  $r^2$  scores around 85 to 90% for the implied volatility. The scores for the daily realized volatility obviously cannot be as high, due to the large measurement noise in realized volatility. Still, they are quite high, lying around 60%.

We have thus uncovered a very simple path-dependent volatility (PDV) model that accurately explains the current volatility value based on recent daily returns. The model captures many well-known stylized facts about volatility:

- leverage effect;
- volatility clustering and volatility bursts;
- time-reversal asymmetry: weak and strong Zumbach effects;
- multiscale memory encoded in the time-shifted power-law weights, which aggregate the various time horizons of investors and traders.

We have shown that our model learns the volatility better than similar models from the ARCH and econophysics literature, consistently across equity indexes and test sets, for both implied and realized volatilities. This might be because our model is homogeneous in volatility (or returns), while the other models either are homogeneous in variance (or squared returns) or mix heterogeneous volatility/variance linear factors. It seems plausible that, when they react to market movements, traders compare homogeneous quantities, and they probably think more in terms of returns or volatility than in terms of squared returns or variance. Moreover, we have proved that while the most important factor explaining volatility is the historical volatility  $\Sigma = \sqrt{R_2}$ , the recent trend  $R_1$  also has significant explanatory power and must not be ignored. Our model is, in fact, the simplest linear model mixing those two factors that is homogeneous in volatility.

One very nice feature of our empirical PDV model is that it easily passes to the continuous-time limit, which we call the Continuous-Time Empirical PDV Model. This model is easily made Markovian by approximating the time-shifted power-law kernels by convex combinations of two exponential kernels. We thus obtain the 4-factor Markovian PDV model, whose main benefits are the following:

- It faithfully captures the path-dependence of volatility observed empirically. This produces a constant lognormal volatility of volatility and a more complicated drift that describes how the trend of the instantaneous volatility depends on the four factors. In particular, this drift explains why the volatility decays more slowly than it spikes.
- Contrary to stochastic volatility (SV) models, the 4 factors have a clear financial meaning and are *observable quantities*: they are exponentially-weighted moving averages of past returns and of past squared returns (as opposed to moving averages of unobservable past Brownian shocks), with two different averaging time scales for each.
- In contrast with SV models, rather than merely postulating the mean reversion of volatility, the model infers it from and explains it by the mean reversion of the four factors.
- The model can generate very realistic paths of SPX and VIX, and rough-like paths of instantaneous volatility due to the very large volatility of volatility and very large fast mean reversion. Some spurious roughness of realized volatility is also observed due to the large noise in the estimation of the daily realized volatility.
- The model can be very accurately calibrated to SPX smiles, the VIX future, and VIX smiles, jointly.
- Being Markovian in low dimension, the model is very easy to simulate.

Therefore, while in [7] Bouchaud argued that “from a purely engineering point of view, the Quadratic Rough Heston Model (QRHM) [20] is probably hard to beat,” we believe that our 4-factor PDV model improves doubly on the QRHM from both an engineering and a financial point of view:

- (1) From an engineering point of view, our model, which is Markovian, is much easier and faster to simulate than the non-Markovian QRHM, which is particularly important in view of calibration to option prices. Our model only has 9 parameters (or 10 in its version (4.19)), which all have a clear financial interpretation.
- (2) From a financial point of view, it captures all the important stylized facts of volatility already present in the QRHM, but better captures the path-dependency of volatility by adding to the feedback of past returns  $R_1$  onto volatility the important feedback of past *squared* returns  $R_2$ . Our empirical study shows the importance of incorporating both feedback effects jointly. In particular, our model explains the quadratic-like dependence of volatility on  $R_1$  observed in market data as a consequence of the linear dependence of volatility on  $\sqrt{R_2}$ . In some sense, the QRHM can thus be seen as a non-Markovian version of the projection of our 4-factor Markovian PDV model onto the models where the volatility is allowed to depend on  $R_1$  only, rather than on  $R_1$  and  $R_2$  jointly.

In order to account for the (small) exogenous part in volatility, we can enhance the 4-factor PDV model by multiplying the path-dependent volatility by an SV component ( $a_t$ ). We thus obtain what we call a Path-Dependent Stochastic Volatility (PDSV) model. We have suggested simple stationary dynamics for ( $a_t$ ) based on the statistical properties of the residuals of our empirical study.

Compared with local volatility (LV) and SV modeling, we believe and we hope to have convinced the reader that this is a better and more natural way of modeling volatility:

- (1) First model the (large) purely *endogenous* part of volatility *explicitly* as best as we can, using PDV models that depend on *observable* factors, such as past asset returns.
- (2) Then model the (small) exogenous part, if needed, based for instance on a statistical study of the residuals of the PDV model.

This comes in contrast with (a) LV modeling, which ignores the path-dependent nature of volatility, and (b) SV modeling, which starts from the exogenous part, postulates a dynamics for the instantaneous volatility that typically depends on *unobservable* factors (such as moving averages of past Brownian shocks), and can only generate some *implicit*, complicated path-dependency by correlating the Brownian motions that drive the dynamics of the asset price and the stochastic volatility. It is our hope that this PDV/PDSV approach opens an interesting line of future research in volatility modeling and stirs interesting debates and discussions among practitioners and academic researchers.

**Acknowledgements.** We would like to thank the participants of QuantMinds 2021, the Peter Carr Memorial Conference at NYU Tandon, the 11th World Congress of the Bachelier Finance Society, and the 9th International Colloquium on BSDEs and Mean Field Systems, the Research in Options 2022 conference,

and in particular Frédéric Abergel and Lorenzo Torricelli, as well as Mathieu Rosenbaum, for interesting comments.

## REFERENCES

- [1] Baillie, R.T., Bollerslev, T., Mikkelsen, H.-O.: *Fractionally integrated generalized autoregressive conditional heteroskedasticity*, Journal of Econometrics 74:3–30, 1996.
- [2] Bergomi, L.: *Smile dynamics II*, Risk, October, 2005.
- [3] Bergomi, L.: *Stochastic Volatility Modeling*, Chapman & Hall/CRC Financial Mathematics Series, 2016.
- [4] Blanc, P., Donier, B., Bouchaud, J.-P.: *Quadratic Hawkes processes for financial prices*, Quantitative Finance, 17(2):171–188, 2017.
- [5] Bolko, A., Christensen, K., Pakkanen, M., Veliyev, B.: *A GMM approach to estimate the roughness of stochastic volatility*, preprint, 2022.
- [6] Bollerslev, T.: *Generalized autoregressive conditional heteroskedasticity*, Journal of Econometrics, 31:307–327, 1986.
- [7] Bouchaud, J.-P.: *Radical complexity*, Entropy 23:1676, 2021.
- [8] Byrd, R., Schnabel, R., Gerald S., *Approximate solution of the trust region problem by minimization over two-dimensional subspaces*, Math. Programming, 40, pp. 247–263, 1988.
- [9] Breidt, F.J., Crato, N., de Lima, P.: *The detection and estimation of long memory in stochastic volatility*, Journal of Econometrics 83(1–2):325–348, March–April 1998.
- [10] Brunick, G., Shreve, S.: *Mimicking an Itô process by a solution of a stochastic differential equation*, Ann. Appl. Prob., 23(4):1584–1628, 2013.
- [11] Chicheportiche, R., Bouchaud, J.-P.: *The fine-structure of volatility feedback I: Multi-scale self-reflexivity*, Physica A 410:174–195, 2014.
- [12] Comte, F., Renault, E.: *Long memory in continuous-time stochastic volatility models*, Math. Finance 8(4):291–323, 1998.
- [13] Cont, R., Das, P.: *Rough volatility: fact or artefact?*, preprint, 2022.
- [14] Dandapani, A., Jusselin, P., Rosenbaum, M.: *From quadratic Hawkes processes to super-Heston rough volatility models with Zumbach effect*, Quantitative Finance, 21(8):1235–1247, 2021.
- [15] Dupire, B.: *Pricing with a smile*, Risk, January, 1994.
- [16] Durrleman, V.: *From implied to spot volatilities*, Finance and Stochastics 14:157–177, 2010.
- [17] El Euch, O., Gatheral, J., Radoičić, R., Rosenbaum, M.: *The Zumbach effect under rough Heston*, Quantitative Finance 20(2):235–241, 2020.
- [18] Engle, R.: *Autoregressive conditional heteroscedasticity with estimates of variance of United Kingdom inflation*, Econometrica 50:987–1008, 1982.
- [19] Gatheral, J., Jaisson, T., Rosenbaum, M.: *Volatility is rough*, Quantitative Finance 18(6):933–949, 2018.
- [20] Gatheral, J., Jusselin, P., Rosenbaum, M.: *The quadratic rough Heston model and the joint calibration problem*, Risk, May 2020.
- [21] Guyon, J.: *Path-dependent volatility*, Risk, October 2014. Longer version available at [ssrn.com/abstract=2425048](https://ssrn.com/abstract=2425048).
- [22] Guyon, J.: *Path-dependent volatility: practical examples*, presentation at Global Derivatives 2017, Barcelona, May 10, 2017.
- [23] Guyon J.: *Inversion of Convex Ordering in the VIX Market*, Quantitative Finance, 20(10):1597–1623, 2020.
- [24] Guyon, J.: *The VIX Future in Bergomi Models: Fast Approximation Formulas and Joint Calibration with S&P 500 Skew*, to appear in SIAM J. Finan. Math. 2022.
- [25] Guyon, J.: *The Smile of Stochastic Volatility: Revisiting the Bergomi-Guyon Expansion*, preprint, available at [ssrn.com/abstract=3956786](https://ssrn.com/abstract=3956786), 2021.
- [26] Gyurkó, L.G., Kontkowski, M., Lyons, T., Field, J.: *Extracting information from the signature of a financial data stream*, preprint arXiv:1307.7244, 2014.
- [27] Harvey, A.: *Long Memory in Stochastic Volatility*, in Knight, J. and Satchell, S. (eds.), Forecasting Volatility in Financial Markets, Oxford: Butterworth-Haineman, 307–320, 1998.
- [28] Heber, Gerd, Asger Lunde, Neil Shephard, Kevin Sheppard: *Oxford-Man Institute's realized library*, version 0.3, Oxford-Man Institute, University of Oxford, 2009.
- [29] Hobson, D., Rogers, L.C.G.: *Complete models with stochastic volatility*, Mathematical Finance 8(1):27–48, 1998.
- [30] Parent, L.: *The EWMA Heston model*, preprint, 2021.
- [31] Sentana, E.: *Quadratic ARCH models*, Rev. Econ. Stud., 62(4):639–661, 1995.
- [32] Virtanen, Pauli, et al.: *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*, Nature Methods 2020, 261–272.
- [33] Zumbach, G.: *Time reversal invariance in finance*, Quantitative Finance 9(5):505–515, 2009.
- [34] Zumbach, G.: *Volatility conditional on price trends*, Quantitative Finance 10(4):431–442, 2010.

## APPENDICES

APPENDIX A. EMPIRICAL STUDY: THE CASE WHERE  $K_1$  AND  $K_2$  ARE CONVEX COMBINATIONS OF TWO EXPONENTIAL KERNELS

		Train		Test	
		RMSE	$r^2$	RMSE	$r^2$
Implied	<b>VIX</b>	0.020	0.947	0.034	0.868
	<b>VIX9D</b>	0.023	0.872	0.038	0.896
	<b>VSTOXX</b>	0.026	0.929	0.029	0.913
	<b>IVI</b>	0.024	0.924	0.030	0.874
	<b>VDAX-NEW</b>	0.024	0.935	0.027	0.920
	<b>Nikkei 225 VI</b>	0.029	0.893	0.031	0.797
Realized	<b>SPX</b>	0.049	0.740	0.062	0.671
	<b>STOXX</b>	0.060	0.672	0.064	0.648
	<b>FTSE</b>	0.055	0.651	0.067	0.618
	<b>DAX</b>	0.057	0.72	0.060	0.55
	<b>NIKKEI</b>	0.051	0.568	0.052	0.501

TABLE 9. RMSE and  $r^2$  scores for our model (3.7) when  $K_1, K_2$  are of the type (4.9)

		$\beta_0$	$\lambda_{1,0}$	$\lambda_{1,1}$	$\theta_1$	$\beta_1$	$\lambda_{2,0}$	$\lambda_{2,1}$	$\theta_2$	$\beta_2$
Implied	<b>VIX</b>	0.054	52.8	3.79	0.81	-0.078	17.3	1.16	0.43	0.82
	<b>VIX9D</b>	0.042	71.2	3.5	0.78	-0.101	26.9	0.58	0.38	0.88
	<b>VSTOXX</b>	0.039	44.0	12.6	0.65	-0.035	16.0	2.21	0.44	0.93
	<b>IVI</b>	0.030	37.1	4.9	0.72	-0.067	15.8	1.5	0.43	0.94
	<b>VDAX-NEW</b>	0.038	45.7	17.8	0.56	-0.023	15.0	2.1	0.323	0.91
	<b>Nikkei 225 VI</b>	0.066	49.1	0.9	0.87	-0.077	32.3	5.5	0.606	0.80
Realized	<b>SPX</b>	0.020	64.5	3.83	0.67	-0.054	37.6	1.2	0.20	0.67
	<b>STOXX</b>	0.027	57.1	2.3	0.67	-0.062	34.8	1.6	0.2	0.66
	<b>FTSE</b>	0.020	72.4	9.2	0.604	-0.042	36.4	2.86	0.33	0.73
	<b>DAX</b>	0.006	84.3	23.8	0.62	-0.025	31.5	2.7	0.27	0.78
	<b>NIKKEI</b>	0.039	100.8	0.7	0.15	-0.050	52.2	6.9	0.24	0.46

TABLE 10. Optimal parameters of our model (3.7) when  $K_1, K_2$  are of the type (4.9)

## APPENDIX B. EMPIRICAL STUDY: VARYING THE TEST SET

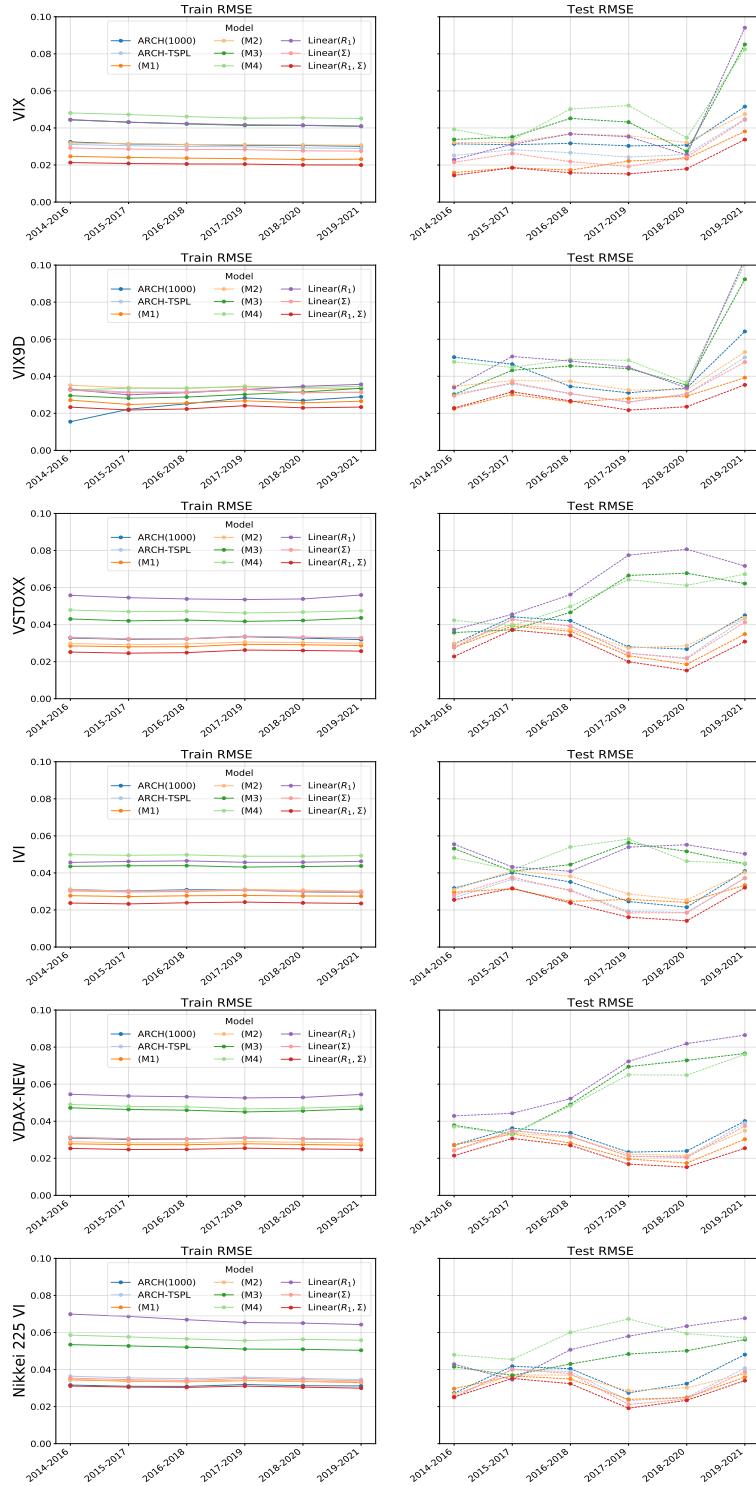


FIGURE B.1. RMSEs for the implied volatility when we vary the test set, for various models. For instance, for the label 2014-2016, the models are trained from Jan 1, 2000 to Dec 31, 2013 and tested over the next two years (Jan 1, 2014 to Jan 1, 2016).

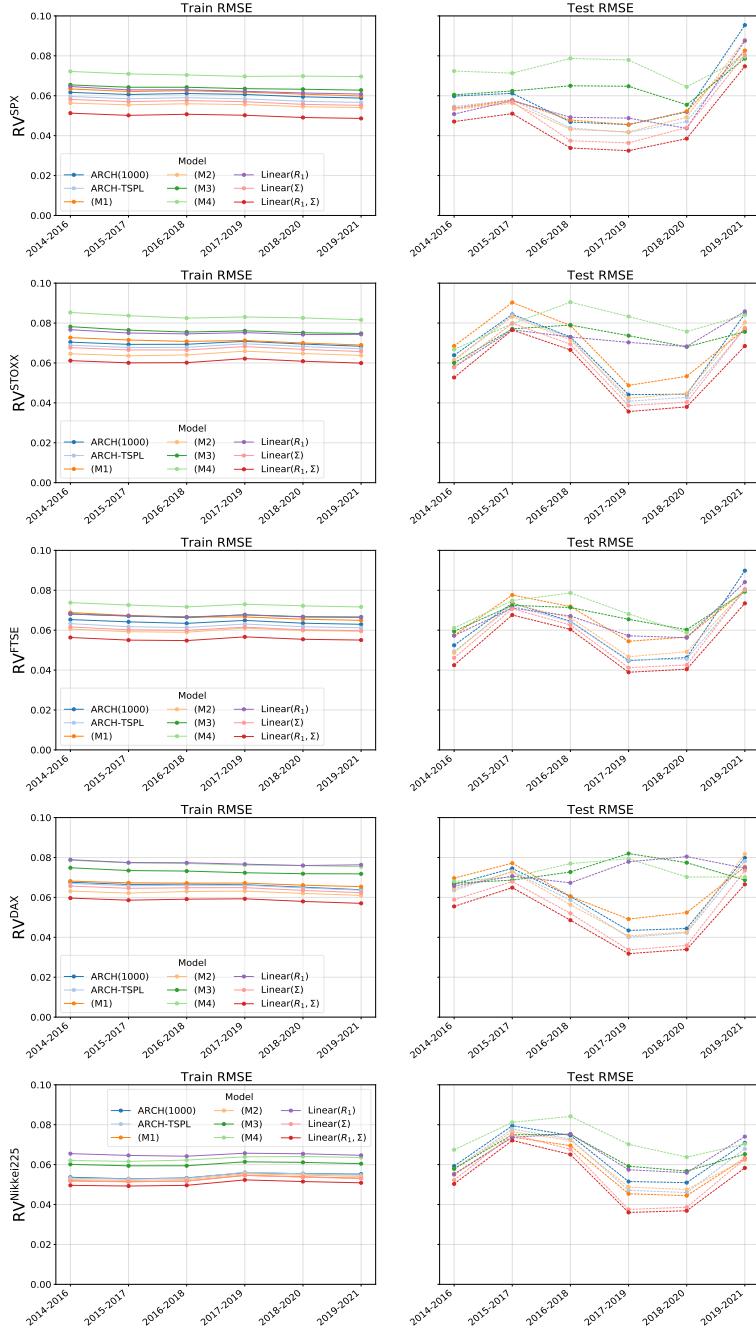


FIGURE B.2. RMSEs for the realized volatility when we vary the test set, for various models. For instance, for the label 2014-2016, the models are trained from Jan 1, 2000 to Dec 31, 2013 and tested over the next two years (Jan 1, 2014 to Jan 1, 2016).

Note that our model consistently outperforms the other models across train/test sets. The corresponding test  $r^2$  may not be as large as those in Table 2, since in some test sets the volatility has less unconditional variance than in the challenging 2019-2022 test set.

## APPENDIX C. DRIFT OF THE INSTANTANEOUS VOLATILITY IN THE 2-FACTOR PDV MODEL

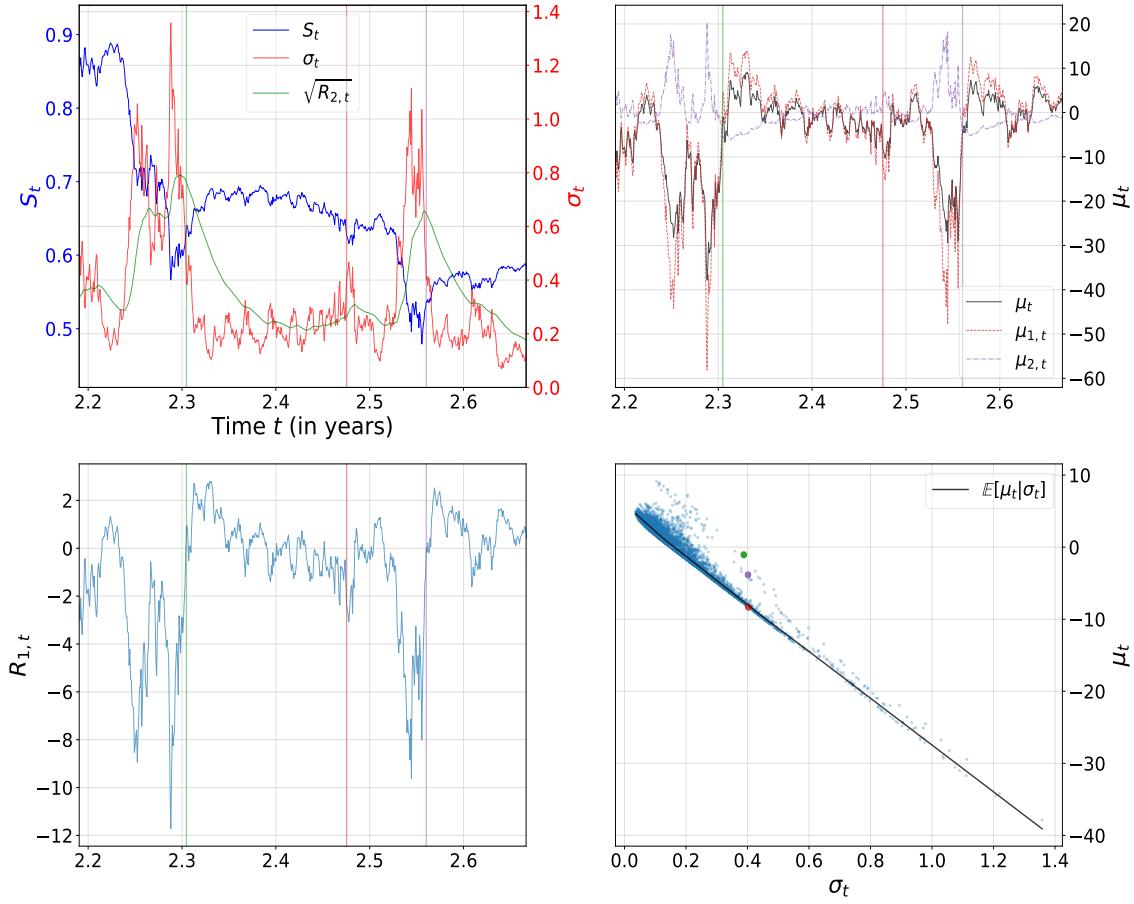


FIGURE C.1. One simulated path over 5 years using the 2-factor PDV model with the parameters in Table 7. Bottom right: scatter plot of the drift  $\mu_t$  of  $\sigma_t$  against  $\sigma_t$  itself. The other three graphs show the time series of various variables within a selected time window. The three colored vertical lines correspond to the time stamp of the three points with the same respective color in the bottom right graph. For those three times the instantaneous volatility  $\sigma_t \approx 0.4$  is about the same, but the drift is very different. After a volatility spike (green and purple),  $\sigma_t$  experiences a less negative drift than the average, i.e., a smaller mean reversion

## APPENDIX D. A SAMPLE PATH IN THE 4-FACTOR PDV MODEL

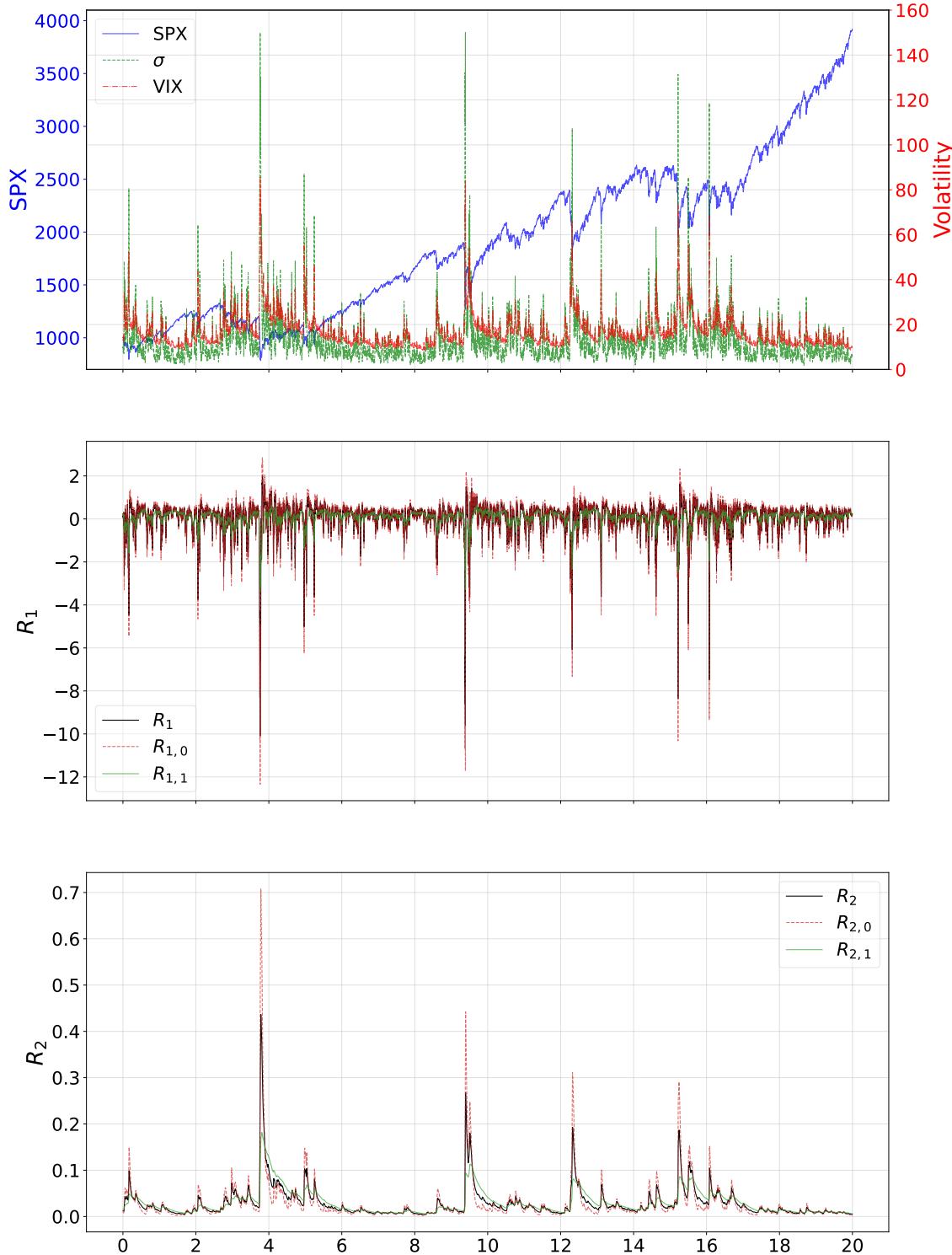


FIGURE D.1. Time series of one simulated path in the 4-factor PDV model using Parameter set 1 in Table 8. Top: underlying price, VIX, and instantaneous volatility  $\sigma$ . Middle:  $R_1, R_{1,0}, R_{1,1}$ . Bottom:  $R_2, R_{2,0}, R_{2,1}$

## APPENDIX E. SCATTER PLOTS OF SIMULATED PATHS IN THE 4-FACTOR PDV MODEL

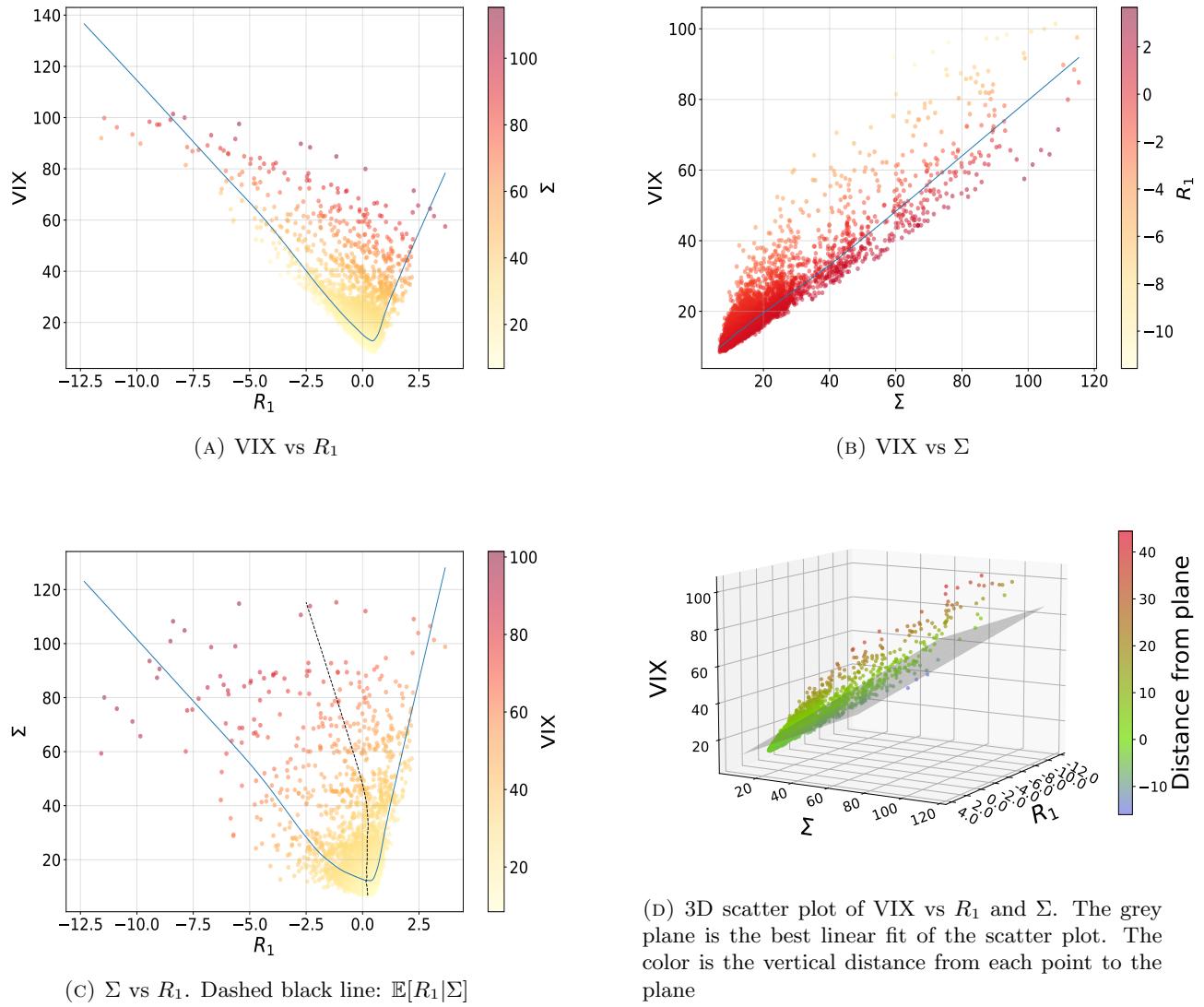


FIGURE E.1. VIX and features scatter plots in the 4-factor Markovian PDV model. We simulate 5 paths of 5 years using Parameter set 1 in Table 8. As initial values, we use  $R_{1,0,0} = 0.078, R_{1,1,0} = 0.16, R_{2,0,0} = 0.007, R_{2,1,0} = 0.016$ . On a plot displaying  $Y$  vs  $X$ , the blue line represents the estimation of  $\mathbb{E}[Y|X]$ .  $\Sigma$  and VIX are expressed in percentage

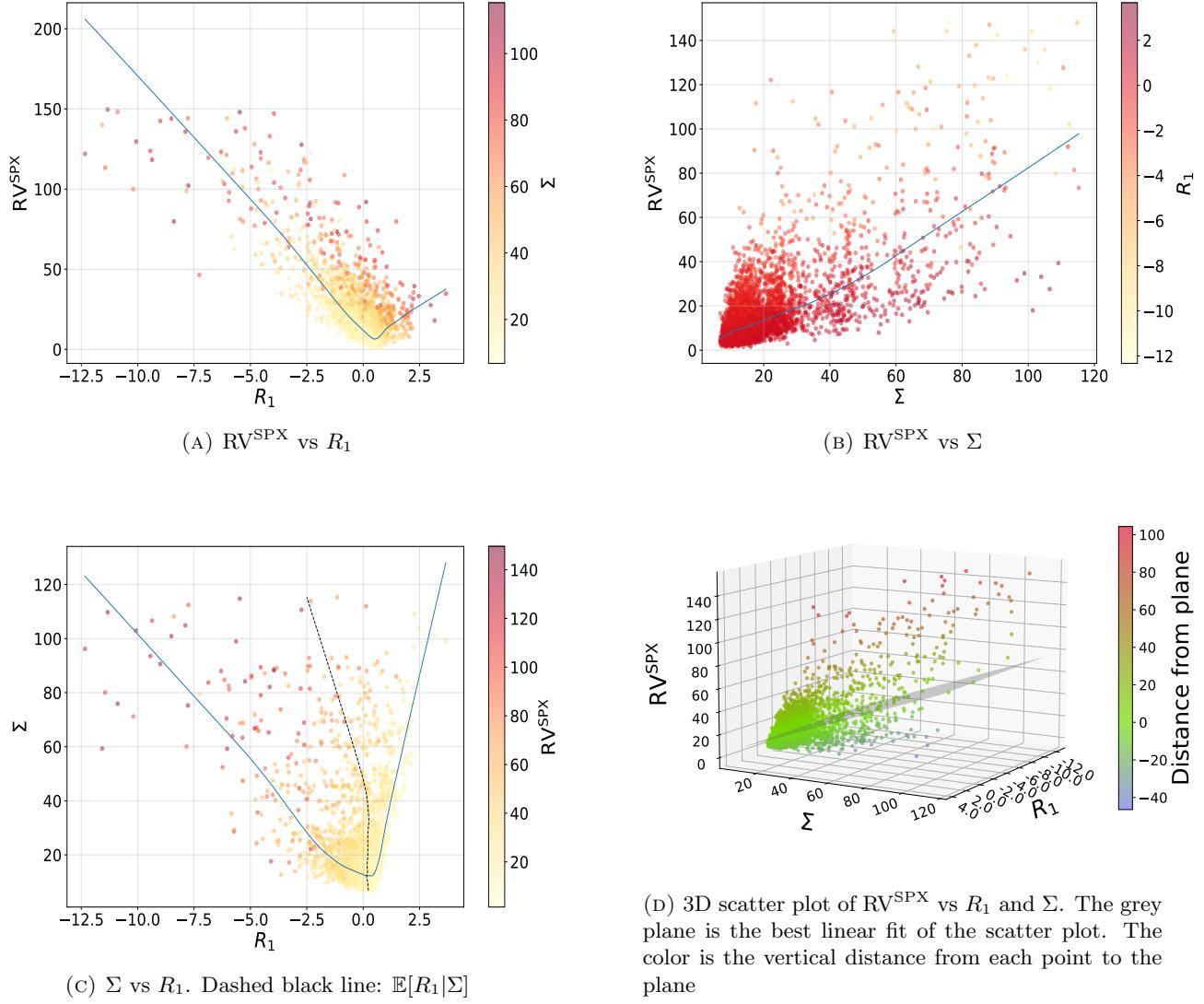


FIGURE E.2.  $RV^{SPX}$  and features scatter plots in the 4-factor Markovian PDV model. We simulate 5 paths of 5 years using Parameter set 1 in Table 8. As initial values, we use  $R_{1,0,0} = 0.078$ ,  $R_{1,1,0} = 0.16$ ,  $R_{2,0,0} = 0.007$ ,  $R_{2,1,0} = 0.016$ . On a plot displaying  $Y$  vs  $X$ , the blue line represents the estimation of  $\mathbb{E}[Y|X]$ .  $\Sigma$  and  $RV_t$  are expressed in percentage