

Joint dynamics for the underlying asset and its implied volatility surface: A new methodology for option risk management^{*†}

Pascal François^a, Rémi Galarneau-Vincent^b, Geneviève Gauthier^{‡c}, and
Frédéric Godin^d

^aHEC Montréal, Department of Finance, Montreal, Canada.

^bHEC Montréal, Department of Decision Sciences, Montreal, Canada.

^cHEC Montréal, Department of Decision Sciences and GERAD, Montreal, Canada.

Associate member of the Oxford Man Institute of Quantitative Finance, Oxford, UK.

^dConcordia University, Department of Mathematics and Statistics, Montreal, Canada.

July 13, 2024

Abstract

This paper develops a dynamic joint model of the implied volatility (IV) surface and its underlying asset which is tractable and seamless to estimate. It combines an asymptotically well-behaved, parametric IV surface representation with a two-component variance, and non-Gaussian asymmetric GARCH specification for the underlying asset returns. Estimated on S&P 500 index return and option data for the 1996-2020 period, the model captures the IV surface movements well and uses them to obtain an improved fit on index returns. It also proves to be an effective risk management tool, producing reliable Value-at-Risk estimates for straddle and strangle positions, and accurate forecasts of the VIX distribution.

JEL classification: C22, G13, G17.

Keywords: Implied volatility, Dynamic factor model, Risk management, VIX.

^{*}Declarations of interest: none. François is supported by a professorship funded by HEC Montréal. Galarneau-Vincent is supported by the TMX Group, Hydro-Québec, Letko-Brosseau, Manulife, IVADO, and MITACS. Gauthier is supported by NSERC (RGPIN-2019-04029), and a professorship funded by HEC Montréal and the HEC Montréal Foundation. Godin is supported by NSERC (RGPIN-2024-04593).

[†]We would like to thank the discussant Gustavo Freire at the 2023 Eurofidai conference and two anonymous referees for their extremely valuable feedback.

[‡]Corresponding author.

Email addresses: pascal.francois@hec.ca (Pascal François), remi.galarneau-vincent@hec.ca (Rémi Galarneau-Vincent), genevieve.gauthier@hec.ca (Geneviève Gauthier), frederic.godin@concordia.ca (Frédéric Godin).

1 Introduction

The Black & Scholes (1973) model revolutionized the practice of option trading. The model not only provided the first arbitrage-based, explicit option pricing formula, but also highlighted the bijective relationship between the option premium and its implied volatility (IV). Since then, the IV surface has become the standard representation of option market prices. It is also the output to which the performance of option pricing models is benchmarked (Andersen et al., 2015). A rich option pricing literature has developed over the last four decades and extended the original Black and Scholes framework. Most contributions to this literature build upon the standard option pricing approach, which, despite many different modelling assumptions, typically proceeds in two steps. First, it posits the dynamics of the underlying asset return under the historical probability measure. Various additional state variables can be specified (such as volatility, jumps, interest rates, and convenience yields) to augment the model’s realism.¹ Next, the standard approach establishes the rules for the arbitrage-free valuation of contingent claims, which implies characterizing a change of probability measure.

In this paper, we opt for a radically different route to model the dynamics of option prices. We build on an early, yet underdeveloped literature that suggests using the IV surface not as a model output but rather as an input. We propose a dynamic extension of the parametric IV surface of François et al. (2022) that we couple with an asymmetric GARCH process with non-Gaussian innovations for the underlying asset return. Our approach (labelled

¹A non-exhaustive list of modelling innovations includes: GARCH processes in discrete time (Glosten et al. (1993), Duan (1995), Heston & Nandi (2000)), stochastic volatility (Hull & White (1987), Heston (1993), Bates (1996), Duffie et al. (2000)), jumps in returns and in volatility (Merton (1976), Broadie et al. (2007), Bollerslev & Todorov (2011)), two-factor volatility (Bates (2000), Christoffersen et al. (2008), Andersen et al. (2015)), and non-normal innovations (Barndorff-Nielsen (1998), Carr & Wu (2004), Christoffersen et al. (2010)).

the JIVR model, which stands for Joint Implied Volatility and Return) forecasts the future distributions of S&P 500 index straddle positions and of the VIX very accurately. It succeeds in doing so because using the IV surface integrates market information as a whole and helps properly assess higher-order moments and capture tail risk. A noteworthy merit of the JIVR model is the ease with which it is implemented. Estimating parameters is fast and only requires standard techniques such as least-square regressions and maximum likelihood.² First attempts to directly model the IV surface include Zhu & Avellaneda (1998), Schönbucher (1999), Fengler (2006) and Daglish et al. (2007). These early works, sometimes referred to as IV market models, assume diffusion processes for implied volatilities, and they highlight the difficulty of deriving constraints on the risk-neutral drift to prevent arbitrage. Carr & Wu (2016) also emphasizes the problematic fit with the current shape of the IV surface. For that reason, Carr & Wu (2016) restrict the modelling of the IV surface to near-term dynamics. They suggest using their framework in conjunction with a parametric specification for the underlying asset return. In the same spirit, Carr & Wu (2020) limit the diffusion modelling of the IV to the management of the instantaneous P&L of an option position. Aside from diffusions, a related approach consists in extracting the IV surface explanatory factors in a non-parametric fashion (Cont & Da Fonseca (2002), Israelov & Kelly (2017), Cont & Vuletić (2023)). This method, however, only applies to dense regions of the IV surface. Consequently, it rejects some peripheral options (deep out-the-money and long-maturity) that are very informative about higher-order moments. By contrast, the JIVR model works with an asymptotically well-behaved, parametric IV surface representation that allows for

²This is in sharp contrast with the most recent models of the standard approach that work with several latent variables (e.g., volatility components) and must therefore rely on heavy filtering techniques for estimation (see Bates (2022), for a recent review).

reliable interpolation and extrapolation.

Most importantly, we show that a dynamic IV surface can be consistently incorporated with a model for underlying asset returns. Focusing on S&P 500 index options, we opt for an asymmetric GARCH with non-Gaussian innovations to capture the large variations observed in returns and in the characteristics of the IV surface. The variance has two components, as suggested by Christoffersen et al. (2013). Oh & Park (2022) show that the adequate estimation of a two-factor variance process requires additional sources of information from the derivatives market. In contrast to the literature, our framework exploits the available forward-looking information by connecting one of the variance factors to the 1-month, at-the-money (ATM) IV level. This approach displays better fitting performance compared to a conventional one-factor NGARCH while preserving the stability of parameters. Furthermore, the first coefficient of the IV surface, representing the long-term ATM implied volatility, is shown to have a volatility that is proportional to the 1-month ATM implied volatility level – a result in support of Carr & Wu (2016). The other IV factors follow a GARCH-type process. To complete the JIVR model, a Gaussian copula captures the dependence structure between the S&P 500 log-returns and the IV factors.

For any pricing model, the issue of arbitrage and absence thereof is of primary importance. In the standard approach in which the dynamics of state variables are posited under the physical probability measure, this issue is typically tackled by imposing a drift restriction under the equivalent martingale measure. This constitutes a preliminary step before pricing and, consequently, the model-implied option prices are ipso facto consistent with the absence of arbitrage. They may, however, imperfectly match with the option prices currently observed on the market, which, in turn, calls for calibrating the model before implementation.

Our paper belongs to an alternative stream of option pricing (the IV market models) in which the implied volatility surface is not the output of arbitrage-free, model-generated option prices. Rather, its current observation serves as an input for modelling its future behavior. Under this approach, the absence of arbitrage is not posited: It has to be checked given the assumed dynamics for the implied volatility factors. Such a verification has, for a long time, represented a major challenge for this stream of literature.³

We argue that our paper provides a significant contribution in this matter for two reasons. First, we develop a fully operational, easy-to-implement implied volatility surface forecasting method in which the dynamics of the underlying asset price discounted with the risk-free rate is guaranteed to be a martingale under some equivalent measure (thereby complying with the absence of arbitrage). Second, we admittedly cannot enforce a similar constraint among options (this remains computationally very challenging). However, we can tackle the issue of potentially remaining arbitrage opportunities empirically. In this regard, the parametric specification we rely upon yields smoothed implied volatility surfaces which clean out almost all static arbitrage opportunities present in the original data observations.⁴ We acknowledge that future work is needed on the theoretical front to further preclude the possibility of arbitrage. However, we argue that our paper significantly alleviates what has long been

³As Carr and Wu (2020) point it out: “[...] The strand of the practitioner literature that attempts to directly model the implied volatility dynamics [...], takes the observed implied volatility as given while specifying the continuous martingale component of the volatility surface. From these two inputs, they seek to derive the no-arbitrage restrictions on the risk-neutral drift of the implied volatility dynamics. The approach is analogous to the Heath, Jarrow, and Morton’s (1992) model on forward interest rates and can in principle be used to price derivatives written on the implied volatility surface. What prevents these attempts from achieving their objective is that knowledge about the shape of the current implied volatility surface places constraints on the specification of the continuous martingale component for its future dynamics.”

⁴Among the nearly 3.5 million calendar and butterfly spreads checked in the OptionMetrics S&P500 index option price observations between January 4, 1996, and December 31, 2019, François et al. (2022) show that 0.56% of them represent arbitrage opportunities in the data while only 0.01% remain in the smoothed implied volatility surfaces.

considered as a major obstacle to the development of that stream of literature.

While our framework for joint underlying asset returns and IV surface dynamics potentially has many relevant applications, we focus on the risk management of volatility strategies. Two validation tests are considered. First, we perform the backtests of the Value-at-Risk (VaR) for S&P 500 index straddles and strangles from January 2, 1996, to December 31, 2020. In a large majority of cases, our 5-day VaR estimates successfully pass the coverage test on both tails of the distribution (the test is statistically stringent due to the high number of observations). In all other cases, the VaR violation frequencies remain economically sensible. Second, we forecast the distribution of the VIX index. As a benchmark, we use a GARCH model with non-normal innovations directly applied to the VIX. Using an expanding window starting in 2014, the yearly comparison of log-likelihoods documents the superior performance of the JIVR model. We further compare the JIVR point forecasts of the 1-day ahead VIX with several GARCH specifications studied in Hansen et al. (2024). The RMSEs obtained from our model are 1.5 to 2.4 times lower than those of the benchmarks. Overall, the two aforementioned tests show the ability of the JIVR model to adequately manage volatility positions through the accurate forecasting of IV surface.

The rest of the paper is organized as follows. Section 2 presents the data. Section 3 reviews the parametric specification that serves as a building block for our dynamic IV surface model. Section 4 describes and assembles the components of the JIVR model, which is estimated in Section 5. Section 6 explains the risk management applications. Section 7 concludes.

2 Data

The OptionMetrics database provides the dataset, which includes daily quoted bid and ask prices of European call and put options on the S&P 500 index (SPX options) from the CBOE. The dataset extends from January 4, 1996, to December 31, 2020. On any given day t , the data includes the option strike price K , its maturity date, and the associated underlying asset forward price $F_{t,\tau}$, with τ denoting the time-to-maturity. The OptionMetrics database also includes the zero-coupon yield curve and dividend yields.⁵

Option exclusion filters are applied to the dataset, which mostly follow the Bakshi et al. (1997) guidelines. More precisely, we exclude all in-the-money options as well as options with any of the following characteristics: a time-to-maturity shorter than six trading days, a price lower than \$3/8, a bid price of \$0, or a bid-ask spread larger than 175% of the option mid-price.⁶ The final dataset includes 6,292 days and a total of 3,814,217 option quotes.⁷

For an option with strike price K and time-to-maturity $\tau = T - t$, the moneyness is defined as

$$M_{t,\tau} = \frac{1}{\sqrt{\tau}} \log \frac{F_{t,\tau}}{K}. \quad (1)$$

According to that definition, OTM calls (puts) are associated with a negative (positive) value for M . Implied volatilities are calculated by inverting the Black & Scholes (1973)

⁵The OptionMetrics forward price is computed as

$$F_{t,\tau} = S e^{(r_{t,\tau} - q_t)\tau},$$

where $r_{t,\tau}$ is the time- t continuously compounded risk-free rate for time-to-maturity τ , and q_t is the S&P 500 dividend yield.

⁶This last criterion is similar to that of Azzone & Baviera (2022). When the ratio of the bid-ask spread over the mid-price is large, the latter induces implied volatilities largely deviating from the rest of the IV surface. Options excluded due to this criterion represent a tiny proportion (0.3%) of the total number of options in the dataset.

⁷The IV surface on October 9, 2006, is removed from the dataset because it is very erratic and most likely due to unreliable data on that day.

formula, using the mid-quote price as the observed price.

Table 1: Descriptive statistics of the SPX options implied volatilities

	Calls		Puts			All
	$M \leq -0.2$	$-0.2 < M \leq 0$	$0 < M \leq 0.2$	$0.2 < M \leq 0.8$	$M \geq 0.8$	
Mean (%)	18.89	15.85	20.83	29.53	47.69	25.37
Standard deviation (%)	7.57	5.89	6.27	7.25	12.69	11.74
Number of contracts	334,482	839,841	841,813	1,439,416	358,665	3,814,217

	Days-to-maturity					All
	$\tilde{\tau} \leq 30$	$30 < \tilde{\tau} \leq 90$	$90 < \tilde{\tau} \leq 180$	$180 < \tilde{\tau} \leq 365$	$\tilde{\tau} \geq 365$	
Mean (%)	27.27	25.54	25.99	25.36	23.00	25.37
Standard deviation (%)	15.48	12.75	12.22	10.77	8.72	11.74
Number of contracts	329,083	1,115,684	722,542	738,335	908,573	3,814,217

Descriptive statistics of the daily SPX options implied volatility (IV) surfaces from January 4, 1996, to December 31, 2020, grouped by buckets of moneyness and time-to-maturity. The moneyness M is defined in Equation (1), and $\tilde{\tau}$ represents the time-to-maturity of the option in days.

Table 1 provides a brief description of the sample IVs. Detailed statistics are reported for buckets of moneyness and maturity. The average IV increases with M , reflecting the well-known smile phenomenon, except for deep out-of-the-money (DOTM) calls ($M < -0.2$), where index option smiles typically exhibit a smirk. On average, the term structure of IVs slightly (but not monotonically) decreases. The standard deviation of the IV decreases with the time-to-maturity, indicating a time-varying time-to-maturity slope.

3 Factor-based representation of volatility surfaces

This section recalls the static parametric volatility surface representation model of François et al. (2022) on which the subsequent dynamic model is based. The factors were chosen so that the surface is doubly continuously differentiable in the direction of moneyness, which is a necessary condition for the existence of a risk-neutral density function for the underlying asset price. On any day t , the implied volatility surface (i.e., IVs for any combination of

moneyiness M and time-to-maturity τ) is described with a five-factor model:

$$\begin{aligned} \sigma(M, \tau, \beta_t) = & \underbrace{\beta_{t,1}}_{\text{Long-term ATM IV}} + \beta_{t,2} \underbrace{e^{-\sqrt{\tau/T_{\text{conv}}}}}_{\text{Time-to-maturity slope}} + \beta_{t,3} \underbrace{\left(M \mathbb{1}_{\{M \geq 0\}} + \frac{e^{2M} - 1}{e^{2M} + 1} \mathbb{1}_{\{M < 0\}} \right)}_{\text{Moneyiness slope}} \\ & + \beta_{t,4} \underbrace{\left(1 - e^{-M^2} \right) \log(\tau/T_{\text{max}})}_{\text{Smile attenuation}} + \beta_{t,5} \underbrace{\left(1 - e^{(3M)^3} \right) \log(\tau/T_{\text{max}}) \mathbb{1}_{\{M < 0\}}}_{\text{Smirk}}, \quad \tau \in [T_{\text{min}}, T_{\text{max}}] \end{aligned} \quad (2)$$

where $\beta_t = (\beta_{t,1}, \beta_{t,2}, \beta_{t,3}, \beta_{t,4}, \beta_{t,5})$ are referred to as the factor coefficients. These factors represent the long-term at-the-money (ATM) level, the time-to-maturity slope, the moneyiness slope, the smile attenuation over long maturities, and the smirk, respectively.

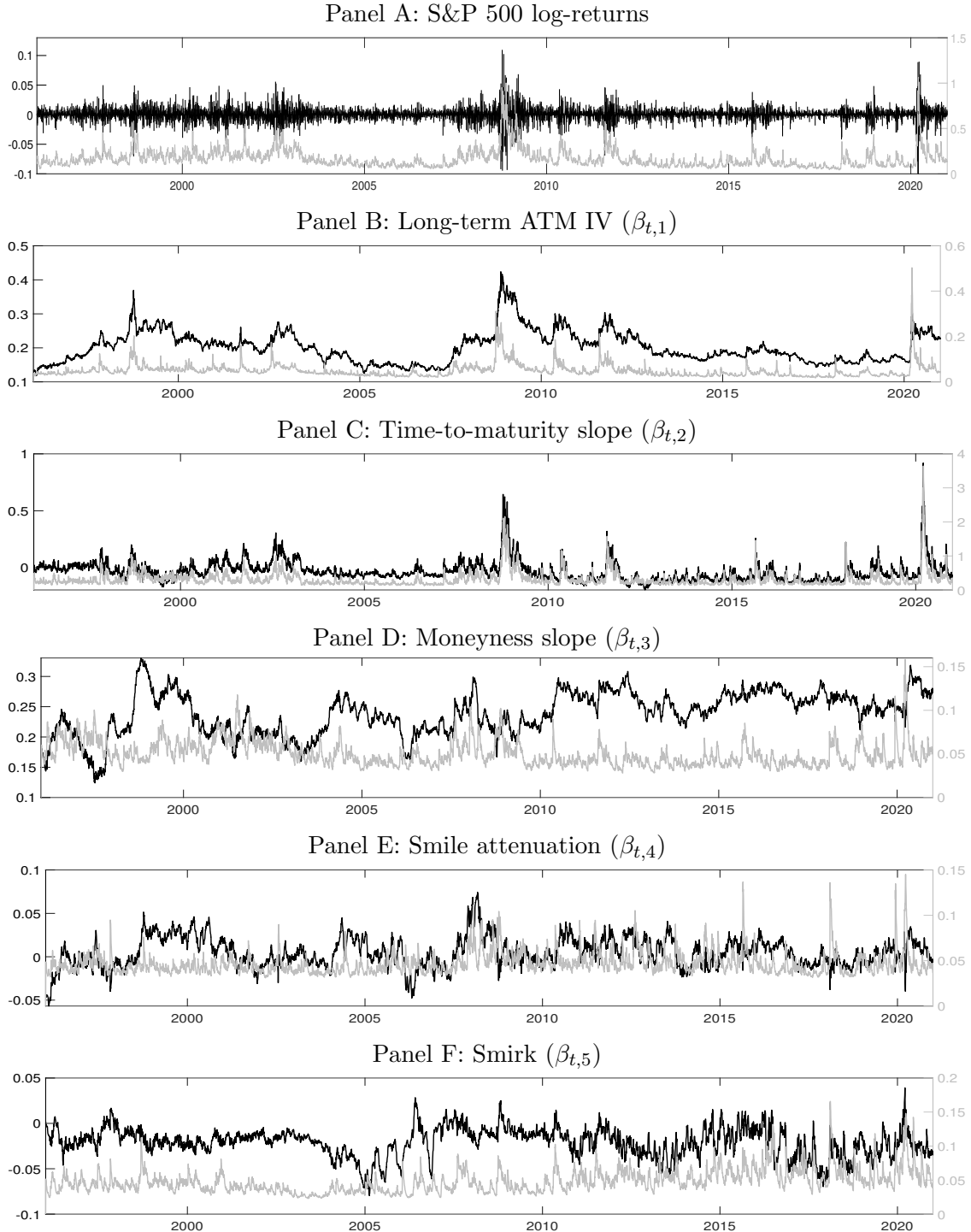
The model is fitted daily (i.e., for each t) to the option prices by minimizing the sum of squared IV differences between the model and the observed prices while incorporating prior information to maintain the financial interpretability of the coefficients.⁸

The black line in Panel A of Figure 1 represents the S&P 500 log-returns. The other five panels contain the time series of estimated coefficients $\beta_{t,1}, \dots, \beta_{t,5}$ (black lines). The time-to-maturity slope in Equation (2) represents the short-term ATM implied volatility minus the long-term ATM implied volatility.⁹ Thus, the coefficient β_2 is negative (resp. positive) when the short-term implied volatility is lower (resp. greater) than the long-term implied volatility. As expected, Panels B and C of Figure 1 show that the long-term level and the slope increased sharply during the 2008 subprime crisis and the COVID-19 pandemic.

Table 2 presents the summary statistics for fitted factor coefficients. The long-term level coefficient $\beta_{t,1}$ varies between 0.12 and 0.42 and displays a mean of 0.2, which is consistent with expectations for a long-term volatility level. The time-to-maturity slope $\beta_{t,2}$ ranges

⁸Following François et al. (2022), the model horizon T_{max} is set to 5 years and T_{conv} to 0.25 to capture the fast convexity change in the IV term structure. $T_{\text{min}} = \frac{6}{252}$ corresponds to the smallest time-to-maturity in our sample. The parameters are estimated by means of least-square regressions with a Bayesian adjustment.

⁹Due to the fact that $\sigma_t(0, 0, \beta_t) - \lim_{\tau \rightarrow \infty} \sigma_t(0, \tau, \beta_t) = \beta_{t,2}$.



On a two-scale graph, Panel A presents the S&P 500 log-returns (black line) and its estimated annualized volatility (grey line) obtained from Equation (3). The other panels display the daily factor coefficient estimates (black line) as well as their associated estimated volatility (grey line) computed from Equations (5) or (6). The daily sample extends from January 4, 1996, to December 31, 2020.

Figure 1: S&P 500 daily returns, daily IV surface coefficients and their volatilities

Table 2: Summary statistics of the factor coefficients

	Min	Q1	Median	Q3	Max	Mean	Std	Skew	Kurt
Long-term level ($\beta_{t,1}$)	0.12	0.17	0.19	0.23	0.42	0.20	0.05	1.02	4.54
TmT Slope ($\beta_{t,2}$)	-0.20	-0.09	-0.05	0.00	0.92	-0.03	0.09	2.66	17.08
Moneyness Slope ($\beta_{t,3}$)	0.12	0.21	0.24	0.27	0.33	0.24	0.04	-0.38	2.87
Smile attenuation ($\beta_{t,4}$)	-0.06	-0.01	0.00	0.02	0.07	0.00	0.02	0.19	3.71
Smirk ($\beta_{t,5}$)	-0.08	-0.03	-0.02	-0.01	0.04	-0.02	0.02	-0.44	3.48
Short-term vol. ($\beta_{t,1} + \beta_{t,2}$)	0.02	0.09	0.15	0.21	1.21	0.17	0.11	2.40	13.14

Summary statistics of the factor coefficient estimates. The (ATM) long-term level is $\beta_{t,1}$. TmT slope is the term structure slope $\beta_{t,2}$, that is, the difference between the short-term and the long-term ATM implied volatilities. The moneyness slope corresponds to $\beta_{t,3}$ and the smile attenuation to $\beta_{t,4}$. The moneyness smirk for call options is captured by $\beta_{t,5}$. The last row (short-term volatility) is obtained by summing the long-term level factor with the time-to-maturity slope factor ($\beta_{t,1} + \beta_{t,2}$).

from -0.2 to 0.92, and its skewness is strongly positive at 2.66, indicating that it decreases slightly during periods of calm, while sharply increasing during periods of turmoil. The three quartiles for the time-to-maturity slope are all negative, showing that the IV term structure is rarely increasing.

Table 3: Correlation matrix of factor coefficient variations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) S&P 500 log-returns	1	-0.54	-0.71	0.06	-0.21	-0.29	-0.77
(2) Long-term level ($\Delta\beta_{t,1}$)	-0.62	1	0.16	-0.06	0.24	0.15	0.29
(3) TmT Slope ($\Delta\beta_{t,2}$)	-0.76	0.33	1	-0.03	0.13	0.31	0.98
(4) Moneyness Slope ($\Delta\beta_{t,3}$)	0.11	-0.06	-0.09	1	0.27	0.13	-0.03
(5) Smile attenuation ($\Delta\beta_{t,4}$)	-0.14	0.18	0.06	0.24	1	-0.05	0.16
(6) Smirk ($\Delta\beta_{t,5}$)	-0.24	0.15	0.33	0.11	-0.06	1	0.33
(7) Short-term vol. ($\Delta(\beta_{t,1} + \beta_{t,2})$)	-0.81	0.46	0.99	-0.09	0.08	0.33	1

Pearson (below the diagonal) and Spearman (bold numbers above the diagonal) correlations between the S&P 500 log-returns and the coefficients variations $\Delta\beta_{t,i} = \beta_{t,i} - \beta_{t-1,i}$ from January 4, 1996, to December 31, 2020.

Table 3 displays the sample correlation matrix applied to the S&P 500 log-returns and the the daily factor coefficient estimate variations $\Delta\beta_t = \beta_{t,i} - \beta_{t-1,i}$, $i = 1, \dots, 5$. The S&P 500 log-returns are strongly negatively correlated with variations of the long-term volatility coefficient $\Delta\beta_1$ and with those of the time-to-maturity slope $\Delta\beta_2$, but even more so with

those of the short-term ATM implied volatility $\Delta(\beta_{t,1} + \beta_{t,2})$. This is a manifestation of the leverage effect generating higher short-term volatility associated with negative S&P 500 log-returns. Thus, negative index returns generally impact long-term volatility, but not as much as short-term volatility. The almost perfect correlation between $\Delta\beta_2$ and $\Delta(\beta_1 + \beta_2)$ highlights the stability of long-term IVs.

4 The IV surface dynamics

We develop a dynamic extension of the five-factor representation of the IV surface introduced by (François et al., 2022). Each factor is driven by an autoregressive NGARCH process with non-Gaussian innovations. The whole IV surface dynamics is anchored with the underlying asset price process by connecting the very short ATM IV with a fraction of the instantaneous asset return volatility (the proportional gap accounts for the presence of a volatility risk premium). The underlying asset return drift, which captures the equity risk premium, is consistent with no-arbitrage theory. Finally, the model specifies a dependence structure between the underlying asset and the factor coefficients. The contribution of all model components to the accurate modelling of IV surface dynamics is tested and validated in a robustness check (see Appendix B).

Three ingredients are required for the model to be fully characterized: (i) the physical dynamics for the underlying S&P 500 log-returns with time-varying volatility, (ii) the physical dynamics for each of the five-factor coefficients, and (iii) a dependence structure between the underlying asset and the coefficients.

4.1 S&P 500 log-returns

In the spirit of Christoffersen et al. (2008)'s GARCH volatility component model, in which the conditional volatility is mean-reverting around a long-run component, the log-return dynamics follow an adaptation of an NGARCH(1,1)-NIG model where the variance is anchored in the 1-month ATM implied volatility,¹⁰ thereby incorporating the forward-looking information of the option data. The excess log-return¹¹ $R_{t+1} = \log \frac{S_{t+1}}{S_t} - r_t + q_t$ satisfies

$$\begin{aligned} R_{t+1} &= \xi_{t+1} - \psi(\sqrt{h_{t+1,R}\Delta}) + \sqrt{h_{t+1,R}\Delta}\epsilon_{t+1,R}, \\ h_{t+1,R} &= V_t + \kappa_R(h_{t,R} - V_t) + a_R h_{t,R}(\epsilon_{t,R}^2 - 1 - 2\gamma_R \epsilon_{t,R}), \\ \sqrt{V_t} &= \omega_R \sigma\left(0, \frac{1}{12}, \beta_t\right), \end{aligned} \tag{3}$$

where $\Delta = \frac{1}{252}$ represents the daily time step. The sequence of innovations $\{\epsilon_{t,R}\}_{t=1}^T$ is constituted of independent standardized NIG random variables with two parameters ζ and ϕ which influence the skewness and the kurtosis of the distribution.¹² The convexity correction $\psi(\sqrt{h_{t+1,R}\Delta})$ is derived from the cumulant generating function ψ of the standardized NIG

¹⁰As shown by Ledoit & Santa-Clara (1998) and Yan (2011), the very short end of the IV surface at the ATM point converges to the instantaneous volatility under the Equivalent Martingale Measure, i.e. ω_R should be very close to 1 in the absence of volatility risk premium. As shown in Christoffersen et al. (2008), the parameter ω_R should be different than 1. Note that in our framework, $\sigma\left(0, \frac{1}{12}, \beta_t\right) = \left(\beta_{t,1} + \beta_{t,2} \exp\left(-\sqrt{\frac{1}{12} \frac{1}{T_{\text{conv}}}}\right)\right)$.

¹¹ S denotes the index level, r is the daily risk-free rate and q stands for the daily dividend yield.

¹²Their expectation, variance, skewness and excess kurtosis are, respectively:

$$\mathbb{E}[\epsilon_{t,R}] = 0, \quad \mathbb{E}[\epsilon_{t,R}^2] = 1, \quad \mathbb{E}[\epsilon_{t,R}^3] = \frac{3\zeta}{\phi^2} \quad \text{and} \quad \mathbb{E}[\epsilon_{t,R}^4] - 3 = 3\left(\frac{\phi^2 + 5\zeta^2}{\phi^4}\right).$$

distribution, which is described in Appendix F.¹³ Because $E[S_{t+1}|\mathcal{F}_t] = S_t \exp(\xi_{t+1})$,

$$\xi_{t+1} = \psi\left(-\lambda\sqrt{h_{t+1,R}\Delta}\right) - \psi\left((1-\lambda)\sqrt{h_{t+1,R}\Delta}\right) + \psi\left(\sqrt{h_{t+1,R}\Delta}\right). \quad (4)$$

is interpreted as the equity risk premium.¹⁴ As shown in Appendix A, Equation (4) is consistent with the no-arbitrage principle for the underlying asset.

According to Equation (3), the conditional annualized daily variance¹⁵ of the S&P 500 log-returns, $h_{t,R}$, exhibits mean-reverting behaviour around a fraction of the 1-month ATM squared implied volatility. The conditional variance noise term $(\epsilon_{t,R}^2 - 1 - 2\epsilon_{t,R}\gamma_R)$ is centered around 0. Results in Appendix B substantiate the largely superior fitting performance of the standalone S&P 500 log-returns representation stemming from this novel characterization of the variance process over a NGARCH-NIG(1,1) model.

4.2 Factor coefficient dynamics

The second model component specifies the long-term ATM surface level dynamics for $\beta_{t,1}$. As in Carr & Wu (2016), we assume that the volatility of the implied volatility (the volvol) is proportional to the implied volatility level. That assumption is substantiated by Figure 2, which reports the time series of the 1-month ATM implied volatility obtained from Equation (2) and a proxy of the volvol consisting of the sample standard deviation of $\Delta\beta_1$ computed

¹³For $-\sqrt{\zeta^2 + \phi^2} - \zeta < z < \sqrt{\zeta^2 + \phi^2} - \zeta$, the cumulant generating function is given by

$$\psi(z) = \frac{\phi^2}{\phi^2 + \zeta^2} \left(-\zeta z + \phi^2 - \phi\sqrt{\phi^2 + \zeta^2 - (\zeta + z)^2} \right).$$

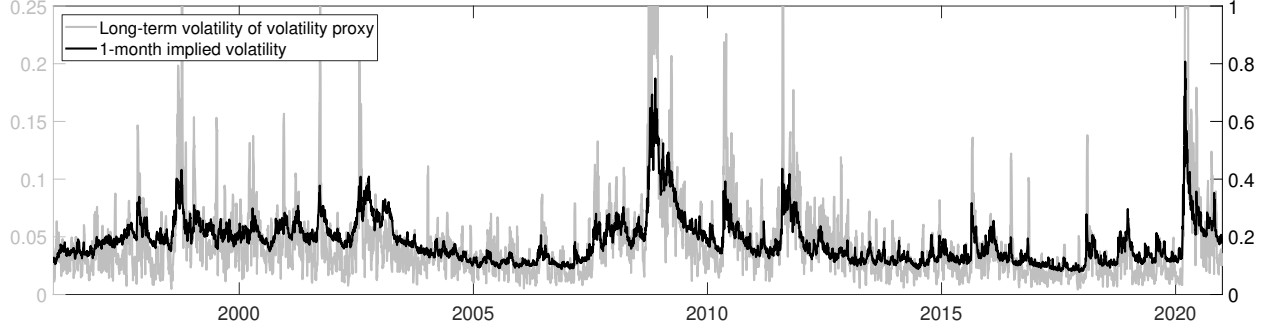
¹⁴If the innovations had a Gaussian distribution, we would have $\psi(z) = \frac{z^2}{2}$ and ξ_{t+1} would become $\lambda h_{t+1,R}\Delta$, with λ being the price of risk.

¹⁵The conditional variance dynamics can be rewritten as

$$h_{t+1,R} = (1 - \kappa_R)V_t + (\kappa_R - a_R(\gamma_R^2 + 1))h_t + a_R h_t (\epsilon_{t,R} - \gamma_R)^2.$$

Therefore, the conditional variance stays nonnegative if $V_t > 0$, $0 \leq \kappa_R \leq 1$ and $|\gamma_R| \leq \sqrt{\frac{\kappa_R - a_R}{a_R}}$.

with a 5-day rolling window. On a two-scale graph, Figure 2 highlights the similarity of the two time series although the volvol proxy is noisier.



The long-term volatility of volatility proxy corresponds to the sample standard deviation of $\Delta\beta_1$ over the last five days, annualized (grey line, left scale). The 1-month implied volatility is $\sigma\left(0, \frac{1}{12}, \beta_t\right) = \left(\beta_{t,1} + \beta_{t,2} \exp\left(-\sqrt{\frac{1}{12} \frac{1}{T_{\text{conv}}}}\right)\right)$ (black line, right scale).

Figure 2: Comparing the 1-month ATM IV to the long-term IV volatility proxy

In line with the evidence in Figure 2, we propose a volatility process $\sqrt{h_{t,1}}$ for the long-term level ($\beta_{t,1}$) that is anchored to the underlying asset 1-month implied volatility. More precisely, the conditional variance $h_{t,1}$ of the ATM long-term level of the IV surface ($\beta_{t,1}$) exhibits a mean-reverting behavior driven by a fraction ω_1 of the squared one-month implied volatility level, to which we add a noise term. The ATM long-term level of the IV surface ($\beta_{t,1}$) evolution is therefore modelled with an AR-NGARCH(1,1)-NIG model:

$$\begin{aligned}\beta_{t+1,1} &= \alpha_1 + \sum_{j=1}^5 \theta_{1,j} \beta_{t,j} + \sqrt{h_{t+1,1}} \Delta\epsilon_{t+1,1}, \\ h_{t+1,1} &= U_t + \kappa_1 (h_{t,1} - U_t) + a_1 h_{t,1} (\epsilon_{t,1}^2 - 1 - 2\epsilon_{t,1} \gamma_1), \\ \sqrt{U_t} &= \omega_1 \sigma\left(0, \frac{1}{12}, \beta_t\right).\end{aligned}\tag{5}$$

The evolution of the four other daily coefficients of Equation (2) is represented by an

AR-NGARCH(1,1)-NIG process. For $i = 2, 3, 4, 5$,

$$\begin{aligned}\beta_{t+1,i} &= \alpha_i + \sum_{j=1}^5 \theta_{i,j} \beta_{t,j} + \nu \beta_{t-1,2} \mathbb{1}_{\{i=2\}} + \sqrt{h_{t+1,i}} \Delta \epsilon_{t+1,i} \\ h_{t+1,i} &= \sigma_i^2 + \kappa_i (h_{t,i} - \sigma_i^2) + a_i h_{t,i} (\epsilon_{t,i}^2 - 1 - 2\epsilon_{t,i} \gamma_i).\end{aligned}\tag{6}$$

A second-order lag for the time-to-maturity slope coefficient is included in the specification to capture the auto-correlation present in its level and its variations. The IV surface coefficients exhibit strong autocorrelation, parameters $\theta_{i,i}$, $i = 1, \dots, 5$ are expected to be close to 1.

4.3 Dependence structure

Specifying a dependence structure completes the modelling framework. A Gaussian copula captures the dependence among the NIG innovations $(\epsilon_{t,R}, \epsilon_{t,1}, \dots, \epsilon_{t,5})$. Interactions between the IV surface coefficients are captured both through auto-regressive parameters $\theta_{i,j}$, $i \neq j$, and through the dependence between the innovations $\epsilon_{t,i}$, $i \in \{R, 1, \dots, 5\}$.

Equations (3)-(6) coupled with the dependence structure of the Gaussian copula comprehensively describe the dynamics of the joint implied volatility and return (JIVR) model.

5 Estimation

The parameters from the dynamic model presented in Section 4 are estimated through a two-step approach.¹⁶ In the first step, parameters of the marginal processes $\{R_t\}, \{\beta_{t,1}\}, \dots, \{\beta_{t,5}\}$ are estimated separately by maximum likelihood.¹⁷ In the second step, parameters of the

¹⁶Although one-step estimation can sometimes produce more accurate results, see Newey & McFadden (1994) for instance, the large number of parameters in our framework makes such an approach extremely challenging from a computational standpoint. As such, the two-step approach is preferred.

¹⁷The backward parameter selection algorithm with the Bayesian Information Criterion (BIC) is implemented. Such an iterative procedure is detailed, for instance, in James et al. (2013).

Gaussian copula are estimated from the model residuals obtained in the first step.¹⁸

Table 4: Cramér-von Mises goodness-of-fit tests

	P-value
S&P500 log-returns	31.6%
Long-term level ($\beta_{t,1}$)	65.2%
TmT Slope ($\beta_{t,2}$)	80.4%
Moneyness Slope ($\beta_{t,3}$)	63.0%
Smile attenuation ($\beta_{t,4}$)	62.5%
Smirk ($\beta_{t,5}$)	32.7%

The table presents p -values of the Cramér-von Mises test (see Appendix C) applied to residuals of the AR-NGARCH-NIG models displayed in Equations (3)-(6) over the whole period ranging from January 4, 1996, to December 31, 2020.

To test the statistical adequacy of the model, Cramér-von Mises tests (see Appendix C) are applied to the residuals of each marginal process. The null hypothesis is that the residuals have a NIG distribution. Table 4 presents the p -values of the tests, with the null hypothesis never being rejected.

Table 5 displays the estimated model parameters. Regarding the S&P 500 dynamics (3), the constant ω linking the 1-month implied ATM volatility to the physical instant volatility is estimated at around 0.98, implying that the physical volatility factor is, on average, smaller than the 1-month implied ATM volatility. The negative skewness and the positive excess kurtosis of the NIG innovation distributions indicate the presence of extreme return movements. As a result, the speed of reversion of the S&P 500 conditional variance ($\kappa_R = 89\%$) is not as close to 1 as it would have been if noises were Gaussian. Indeed, a smaller persistence makes scenarios with prolonged extreme volatility due to a single large S&P 500 return less likely. As expected, the asymmetry parameter γ_R is positive, implying that the S&P 500

¹⁸The Gaussian copula is estimated by converting the residuals whose marginals are approximately NIG into pseudo-residuals with approximately standard Gaussian marginals. This is done through the successive application of the NIG cdf and the Gaussian inverse cumulative distribution function to original residuals. A correlation matrix is computed from the set of residuals with Gaussian marginals, which corresponds to the Gaussian copula parameters.

variance reacts more strongly to negative return shocks than to positive shocks.

The grey lines in Figure 1 display the time series of the estimated S&P 500 log-returns annualized volatility ($\sqrt{h_R}$) as well as the time series of the factor coefficients annualized volatility ($\sqrt{h_i}$ for $i = 1, 2, 3, 4, 5$). The volatility time series of the log-returns and the long-term level factor coefficient closely follow the IV surface level, which is consistent with the specification of the respective variance processes. As expected, the volatility sharply increases during periods of market turmoil, such as the 2008 financial crisis or during the COVID pandemic, and is relatively low during periods where the market is calm. Interestingly, the time-to-maturity slope volatility closely follows the time-to-maturity slope level itself. The volatility time series of the other three factor coefficients do not exhibit any clear pattern related to the underlying log-returns volatility or the IV surface level.

For all five factor coefficients (β_i) displayed in Table 5, large values for the auto-regressive parameter $\theta_{i,i}$ imply strong persistence in their dynamics. Moreover, the speed of reversion parameters κ_i are quite high, indicating that the volatilities of implied volatility coefficients are also persistent. For the long-term implied volatility level β_1 , the asymmetry parameter γ_1 is negative, which is expected because a positive shock on the long-term implied volatility is a sign of market uncertainty and increases the variability of the volatility surface. Because β_2 represents the difference between the short- and the long-term implied volatility levels, β_2 increases during financial turmoil. Again, positive shocks on β_2 have a larger impact on the variability of the implied volatility surface than negative ones, resulting in a negative, yet non-statistically significant asymmetry parameter γ_2 .

Table 5: JIVR model parameter estimates

	β_1	β_2	β_3	β_4	β_5	λ	S&P500
α	0.0009* (0.0002)	0.0084* (0.0009)	0.0008* (0.0003)	-0.0014* (0.0003)	0.0007* (0.0002)		2.7113* (0.0002)
θ_1	0.9963* (0.0009)	-0.0139* (0.0030)		0.0028* (0.0008)			
θ_2	0.0037* (0.0005)	0.8778* (0.0119)	0.0013* (0.0006)				
θ_3		-0.0326* (0.0039)	0.9971* (0.0011)	0.0037* (0.0011)	-0.0042* (0.0009)		
θ_4				0.9803* (0.0028)			
θ_5		-0.0478* (0.0073)			0.9860* (0.0023)		
ν		0.0894* (0.0121)					
σ		0.3803*	0.0522*	0.0486*	0.0515*		
ω	0.2676* (0.0064)						0.9773* (0.0009)
κ	0.8382* (0.0279)	0.9658* (0.0032)	0.9743* (0.0054)	0.9454* (0.0110)	0.9808* (0.0041)		0.8890* (0.0100)
a	0.1342* (0.0150)	0.0983* (0.0072)	0.0926* (0.0101)	0.1022* (0.0113)	0.1005* (0.0100)		0.0561* (0.0041)
γ	-0.1118* (0.0081)	-1.4828 (1.1121)	0.0968 (0.0534)	0.0606 (0.0624)	-0.1030 (0.0563)		2.5078* (0.1125)
β_{NIG}	0.1438* (0.0375)	0.8529* (0.0008)	0.0291* (0.0007)	-0.1591* (0.0004)	0.0927* (0.0004)		-0.6413* (0.0004)
γ_{NIG}	1.3511* (0.0717)	1.5389* (0.0828)	2.2848* (0.1880)	1.4500* (0.0766)	1.4285* (0.0719)		2.0397* (0.0719)
Skew	0.24	1.08	0.02	-0.23	0.14		-0.4625
Ex. Kurt	1.74	3.21	0.58	1.51	1.50		1.0775
Log. Lkhd.	-28,314	-16,940	-27,532	-28,040	-27,874		-20,673
Log. Lkhd. All	-28,322	-16,897	-27,535	-28,042	-27,879		

JIVR model parameters estimated over the whole daily sample ranging from January 4, 1996, to December 31, 2020. The standard errors are displayed under the estimates in parentheses. The model is regularized using a backward selection method with the BIC criterion. The log-likelihoods (Log. Lkhd.) and the log-likelihood of the model where no parameter from the θ matrix is set to 0 (Log. Lkhd. All) are reported. The skewness (Skew) and excess kurtosis (Ex. Kurt) of the NIG distributions for residuals are shown for all coefficients and the S&P 500 log-returns. Parameters with a star superscript (*) are significantly different from 0 at the 5% confidence level.

Table 6: Gaussian copula

	(1)	(2)	(3)	(4)	(5)	(6)
(1) S&P 500 log-returns (Y_t)	1.00					
(2) Long-term level ($\beta_{t,1}$)	-0.55	1.00				
(3) TmT Slope ($\beta_{t,2}$)	-0.69	0.13	1.00			
(4) Moneyness Slope ($\beta_{t,3}$)	0.03	-0.03	-0.01	1.00		
(5) Smile attenuation ($\beta_{t,4}$)	-0.22	0.25	0.12	0.28	1.00	
(6) Smirk ($\beta_{t,5}$)	-0.34	0.17	0.37	0.13	-0.05	1.00

Estimated Gaussian copula parameters for the innovations of the JIVR model $\{\epsilon_{t,R}, \epsilon_{t,1}, \dots, \epsilon_{t,5}\}$. The Gaussian copula is estimated on the residuals extracted from estimated models illustrated in Equations (3)-(6) over the whole sample ranging from January 4, 1996, to December 31, 2020.

Table 6 presents estimates for the Gaussian copula parameter matrix. Results indicate that the log-return innovations are strongly negatively associated with shocks on the first two coefficients, i.e., the long-term level and time-to-maturity slope.

Appendix B reports the performance of several nested models which remove some of the components from the JIVR model. Although such nested models are more parsimonious, Table 11 shows that the JIVR model exhibits higher performance in terms of out-of-sample likelihood. As such, attempts to reduce the complexity of our model would come at the expense of sacrificing predictive power. Note that the log-likelihood performance metric is preferred to the root-mean-square error of implied volatilities used for instance in Huang et al. (2017) since the objective of the present study is distributional (rather than point) forecasting.

6 Risk management applications

We test the ability of the JIVR model to accurately estimate risk metrics for option portfolios on real data. In a first set of numerical experiments, we examine standard positions in

volatility management. Then, we compare VIX index forecasts produced by the JIVR model to those provided by an approach directly modelling the VIX time series. The goal is to assess how well the model does in processing the information from remote areas of the IV surface to capture higher moments and tail behavior.

In the applications presented below, backtesting and forecasting procedures rely on an expanding window methodology. More precisely, the total sample period is divided by year. For each iteration N , where $N = 2007, \dots, 2020$, the model is first estimated over the training sample, which covers years 1996 to $N - 1$. Then, for each day t in year N , multiple d -day-ahead IV surface predictions are generated using the estimated model, the latter being denoted by m_N . Such predictions are then used to calculate daily outcomes. Algorithm 1 summarizes the procedure.

Algorithm 1 The expanding window

```

for  $N = 2007 : 2020$  do                                 $\triangleright$  2007 to 2020 is the out-of-sample period.
    Compute  $m_N$ , the estimated model over the training period 1996 to  $N - 1$ .
    for  $t = 1 : D_N$  do                                     $\triangleright D_N$  is the number of trading days in year  $N$ 
        for  $i = 1 : s$  do                                     $\triangleright s$  is the number of simulations
            Simulate  $\tilde{\beta}_{t+d}^{\{i\}}$  and  $\tilde{Y}_{t+d}^{\{i\}}$  from model  $m_N$  using the information set at time  $t$ .
        end for
    end for
end for

```

6.1 Straddle and strangle positions

Straddles and strangles are standard option strategies that can be used to take positions in the underlying asset volatility. To evaluate the risk of such strategies, we consider the Value-at-Risk (VaR), a popular risk metric used by practitioners. The accuracy of VaR estimates is evaluated through a standard backtesting procedure. VaR estimates are produced for the six

strategies being considered (1-month, 3-month and 6-month straddles and strangles) over a time horizon of five trading days. Straddles positions considered are at-the-money, whereas the moneyness of options included in strangle positions is respectively $M = -0.1$ for the call and $M = 0.1$ for the put.

Daily out-of-sample d -day-ahead VaR estimates for various confidence levels over the years 2007 to 2020 are produced according to the expanding window approach.¹⁹ To forecast the strategy return distribution for a horizon of d days, i.e., a return between times t and $t + d$, one must first compute the current price of the strategy at time t using the fitted IV surface. The return $\frac{V_{t+d}^{\{i\}} - V_t}{V_t}$ is then computed for each simulated²⁰ scenario i , where V_t denotes the time- t strategy value. The return (not price) VaR is estimated to allow for comparisons across straddles and strangles.

VaR coverage tests are conducted to assess the backtesting performance (Kupiec et al., 1995). A VaR breach occurs when an observation falls below (above) the return quantile associated with the 1% or 5% (95% or 99%) confidence level, addressing potential losses for investors taking long (short) positions. The VaR coverage test is a likelihood ratio test that determines if the proportion of realized VaR breaches is significantly different from the VaR confidence level. More details are provided in Appendix D.

Table 7 exhibits the proportion of observed daily VaR breaches over the out-of-sample period (from 2007 to 2020) for each confidence level. Additionally, it is possible to perform

¹⁹The risk-free rate and the dividend yield are kept constant during the simulations over d trading days. Under these assumptions, the forward price is simulated as follows:

$$\tilde{F}_{t+d,\tau-d}^{\{i\}} = F_{t,\tau} e^{\sum_{u=1}^d \tilde{R}_{t+u}^{\{i\}}}$$

where $F_{t,\tau}$ is the forward price at time t with maturity τ and $\tilde{R}_{t+u}^{\{i\}}$ is the simulated daily excess log-return for path i .

²⁰The Monte Carlo simulation is based on 75,000 paths.

Table 7: VaR coverage test for straddles and strangles

		Straddles			Strangles		
	Time-to-maturity	1	3	6	1	3	6
1%	% of VaR breaches	0.57%	0.77%	1.13%	0.71%	0.77%	0.99%
	<i>p</i> -values	(0.50%)	(14.52%)	(43.12%)	(6.72%)	(14.52%)	(96.62%)
5%	% of VaR breaches	2.87%	4.28%	4.48%	3.15%	4.20%	4.40%
	<i>p</i> -values	(0%)	(4.57%)	(15.14%)	(0%)	(2.49%)	(9.39%)
95%	% of VaR breaches	5.30%	4.68%	5.25%	4.96%	4.85%	5.22%
	<i>p</i> -values	(41.05%)	(37.97%)	(50.22%)	(92.30%)	(68.35%)	(55.19%)
99%	% of VaR breaches	1.33%	1.16%	1.70%	1.33%	1.28%	1.59%
	<i>p</i> -values	(5.84%)	(34.27%)	(0.01%)	(5.84%)	(11.35%)	(0.12%)

For strangles, the moneyness of the call option is $M = -0.1$ and that of the put option $M = 0.1$. For straddles, both options are at-the-money. Time-to-maturity (TmT) is in months. The rows (1%, 5%, 95%, and 99%) represent the VaR confidence levels. The distribution forecast horizon is 5 days ahead. The backtest period extends from January 2, 2007, to December 31, 2020. The VaR coverage test is described in Section D in which the number of VaR estimates is $N \approx 14 \times 250 = 3500$. Values in parentheses represent the *p*-values of the tests.

tests to verify whether the differences are statistically significant. The *p*-values for the VaR coverage test are provided in parentheses. Among the 24 tests conducted, 17 have a *p*-value above 5%, while 7 tests conclude that the proportion of VaR breaches is statistically different from the VaR confidence level.

The size of the out-of-sample period makes these tests very powerful, and even slight deviations from the VaR confidence level are diagnosed as departures. However, in all cases, even when the *p*-values are small, the proportion of VaR breaches is not economically far from its theoretical counterpart. Specifically, the proportions of VaR_{1%} breaches lie within (0.57%, 1.70%), which is economically close to 1%. Similarly, for the VaR_{5%}, the proportions of VaR breaches range within (2.87%, 5.30%). Considering that our out-of-sample test includes the 2008 global financial crisis and the 2020 COVID-19 crisis, we believe that the model can be effectively used as a risk management tool.

6.2 Forecasting the VIX index distribution

The VIX index has high practical importance since it encompasses information related to market perceptions about the future volatility of the S&P 500 index over a 30-day horizon.

The Chicago Board Options Exchange (CBOE) computes the VIX index (VIX_t^{index}) from a portfolio of available put and call options as explained in detail in CBOE (2014). For a given time-to-maturity τ ,

$$VIX_{t,\tau} = 100 \sqrt{\left(\frac{2}{\tau} \sum_{i=1}^{N_{t,\tau}} \frac{\Delta K_{i,\tau}}{K_{i,\tau}^2} e^{r\tau} P_{t,\tau}(K_{i,\tau}) - \frac{1}{\tau} \left(\frac{F_{t,\tau}}{K_{j,\tau}} - 1 \right)^2 \right)} \quad (7)$$

where $K_{1,\tau} < \dots < K_{j,\tau} \leq F_{t,\tau} < K_{j+1,\tau} < \dots < K_{N_{t,\tau}}$ are the strike prices of quoted options with maturity τ , $P_{t,\tau}(K_i)$ is the time t out-of-the-money option price.²¹ The strike price variations are $\Delta K_{1,\tau} = K_{2,\tau} - K_{1,\tau}$, $\Delta K_{i,\tau} = \frac{1}{2}(K_{i+1,\tau} - K_{i-1,\tau})$ for $1 < i < N_{t,\tau}$ and $\Delta K_{N,\tau} = K_{N_{t,\tau},\tau} - K_{N_{t,\tau}-1,\tau}$. The $(VIX_{t,T}^{\text{index}})^2$ of maturity $T = 30$ days is a linear interpolation between VIX_{t,τ_1}^2 and VIX_{t,τ_2}^2 , where $\tau_1 \leq T \leq \tau_2$ are the two nearest available time-to-maturities surrounding T :

$$VIX_{t,T}^{\text{index}} = \sqrt{\frac{\tau_2 - T}{\tau_2 - \tau_1} VIX_{t,\tau_1}^2 + \frac{T - \tau_1}{\tau_2 - \tau_1} VIX_{t,\tau_2}^2}. \quad (8)$$

We use the JIVR model to generate forecasts for the VIX variation

$$\Delta VIX_{t,t+d}^{\text{JIVR}} = VIX_{t+d}^{\text{JIVR}} - VIX_t^{\text{JIVR}}$$

over a prediction horizon of d days. More precisely, $VIX_{t,\tau}^{\text{JIVR}}$ is obtained from Equation (8) by replacing the OTM option quoted prices by the ones obtained from the fitted IV surface that day. This step requires the identification of the moneyness levels $M_1, \dots, M_{N_{t,\tau}}$

²¹These prices are obtained from the bid-ask spread mid-points. They correspond to a put option for $K_{i,\tau} \leq F_{t,\tau}$, and a call option for $K_{i,\tau} > F_{t,\tau}$.

corresponding to the available strike prices $K_1, \dots, K_{N_{t,\tau}}$. From Monte Carlo simulations, the JIVR model generates scenarios for the IV surface and the underlying asset log-return for a horizon of d days. Using the same available moneyness levels and maturities as for time t , the VIX forecast $\text{VIX}_{t+d}^{\text{JIVR}}$ is computed from the corresponding option prices on the predicted IV surfaces. Lastly, the $\text{VIX}_{t+d,T}^{\text{index}}$ forecast is

$$\text{VIX}_{t+d,T}^{\text{index}} = \text{VIX}_{t,T}^{\text{index}} + \Delta \text{VIX}_{t,t+d}^{\text{JIVR}}.$$

This approach is compared to a direct modelling of the VIX index time series through:

$$\text{VIX}_t = \alpha_{\text{VIX}} + \beta_{\text{VIX}} \text{VIX}_{t-1} + \sqrt{h_{t,\text{VIX}}} \Delta \epsilon_{t,\text{VIX}} \quad (9)$$

$$h_{t+1}^{\text{VIX}} = \sigma_{\text{VIX}}^2 + \kappa (h_t^{\text{VIX}} - \sigma_{\text{VIX}}^2) + ah_t^{\text{VIX}} (\epsilon_{t,\text{VIX}}^2 - 1 - 2\epsilon_{t,\text{VIX}}\gamma) \quad (10)$$

where the variance process $\{h_{t+1}^{\text{VIX}}\}$ exhibits mean-reversion around a fixed parameter σ_{VIX}^2 . The sequence of innovations $\{\epsilon_{t,\text{VIX}}\}_{t=1}^T$ is constituted of independent standardized NIG random variables.

To compare the log-likelihoods of both models, we introduce the average likelihood ratio (ALR). The ALR is defined as the geometric average of the ratio of likelihood scores for observations of one model over those of another,

$$\begin{aligned} \text{ALR} &= \left(\frac{\mathcal{L}^{(1)}(O_{1:T}|\theta^{(1)})}{\mathcal{L}^{(2)}(O_{1:T}|\theta^{(2)})} \right)^{\frac{1}{T}} \\ &= \exp \left(\frac{1}{T} \left(\sum_{t=1}^T \log \mathcal{L}_t^{(1)}(O_t|O_{1:t-1}, \theta^{(1)}) - \log \mathcal{L}_t^{(2)}(O_t|O_{1:t-1}, \theta^{(2)}) \right) \right) \end{aligned} \quad (11)$$

where T is the total number of days in the set of out-of-sample folds, $\mathcal{L}^{(1)}$ and $\mathcal{L}^{(2)}$ are the likelihoods of the first and second models respectively, $O_{1:T}$ is the time series of out-of-sample observations including all out-of-sample folds, and $\theta^{(1)}$ and $\theta^{(2)}$ represent the model

parameter sets considered when computing $\mathcal{L}^{(1)}$ and $\mathcal{L}^{(2)}$. In our case, model 1 corresponds to the JIVR model, while model 2 corresponds to the direct model. The ALR indicates how much more or less likely an observation is, on average, in one model versus the other in relative terms.

Out-of-sample yearly log-likelihoods are computed using the expanding window methodology described by Algorithm 1.²² However, the out-of-sample period is reduced to between 2014 and 2020, which corresponds to the period where reported VIX values are computed with the most recent calculation method published in CBOE (2014).

Table 8: Out-of-sample performance for VIX distribution forecasting

	JIVR	Direct approach	ALR	<i>p</i> -values
2014	-346.58	-352.33	1.02	22.9%
2015	-434.79	-430.55	0.98	75.8%
2016	-372.33	-379.62	1.03	10.0%
2017	-251.82	-269.93	1.07	2.8%
2018	-436.39	-445.06	1.04	16.5%
2019	-359.79	-362.01	1.01	24.5%
2020	-519.20	-537.27	1.07	1.8%
Total	-2,720.90	-2,776.77	1.03	0.4%

Log-likelihoods for each of the out-of-sample years as well as for the aggregated out-of-sample period (Total). To compute log-likelihoods, a prediction horizon of $d = 1$ is used. The log-likelihood for each year is computed using both the JIVR model and the direct model. The respective parameters of both models are estimated over previous years' observations. Bold numbers highlight which of the two models outperforms the other, either for a specific year or in aggregate. The last column of the table (*p*-values) corresponds to the Diebold & Mariano (1995) test *p*-values. The Diebold & Mariano (1995) test, described in Appendix E, verifies if the predictive accuracy of two models is equal for a specified performance metric (log-likelihood). If the null hypothesis of the test is rejected, then one model statistically outperforms the other.

Table 8 exhibits the log-likelihood, the ALR, and the *p*-values from the Diebold & Mariano (1995) test computed for each out-of-sample year as well as for the whole out-of-sample

²²The log-likelihood cannot be computed directly from the simulated VIX distribution generated by the JIVR model. To circumvent this issue, a kernel density estimate (`ksdensity` function from the MATLAB software with the default bandwidth) is applied to the simulated VIX values to obtain a density estimate. The latter is then used to compute the log-likelihood for the JIVR model.

period. The null hypothesis of Diebold & Mariano (1995) test, described in Appendix E, assumes that the predictive accuracy of the models is statistically equal. For a significance level of 5%, a p -value under 2.5% (resp. over 97.5%) indicates that the JIVR model (resp. direct model) significantly outperforms the direct model (JIVR model). Results show that the JIVR model largely outperforms the direct model, with ARL above 1 for six out of the seven out-of-sample years and for the entire aggregated period (2014–2020). The Diebold & Mariano (1995) test p -values reveal that the JIVR model statistically outperforms the direct model on the entire sample.

6.3 One-day ahead VIX point forecasting

The likelihood criterion was favored over out-of-sample RMSEs because we are addressing the issue of distributional forecasting rather than point forecasting. When it comes to 1-day ahead VIX prediction errors, we have numerous models to benchmark against, as tested in Hansen et al. (2024). Table 9 presents the VIX prediction results from Hansen et al. (2024), to which we append the final column showcasing the performance of our model. The JIVR RMSE of 1.451 is notably 1.5 times better than the best model evaluated in Table 5 of Hansen et al. (2024). Figure 3 displays the one-day ahead VIX forecasts and the realized VIX over the January 3, 2007 – December 30, 2018 time period.

Table 9: One-day ahead VIX forecast RMSE

	RG	EG	G	HN	HNvd	JIVR
January 3, 2007 – December 30, 2018	2.870	3.259	3.278	4.298	3.903	1.894
January 3, 2007 – December 30, 2012	3.383	3.972	3.944	5.497	4.866	2.247
January 3, 2013 – December 30, 2018	2.242	2.339	2.435	2.591	2.603	1.452

The first five columns are extracted from table 5 of Hansen et al. (2024). RG stands for the realized GARCH model, EG corresponds to the EGARCH, G is for GARCH, HN stands for Heston-Nandi GARCH, HNvd represents the Heston-Nandi GARCH with the variance-dependent SDF, and JIVR is our Joint Implied Volatility and Returns model.

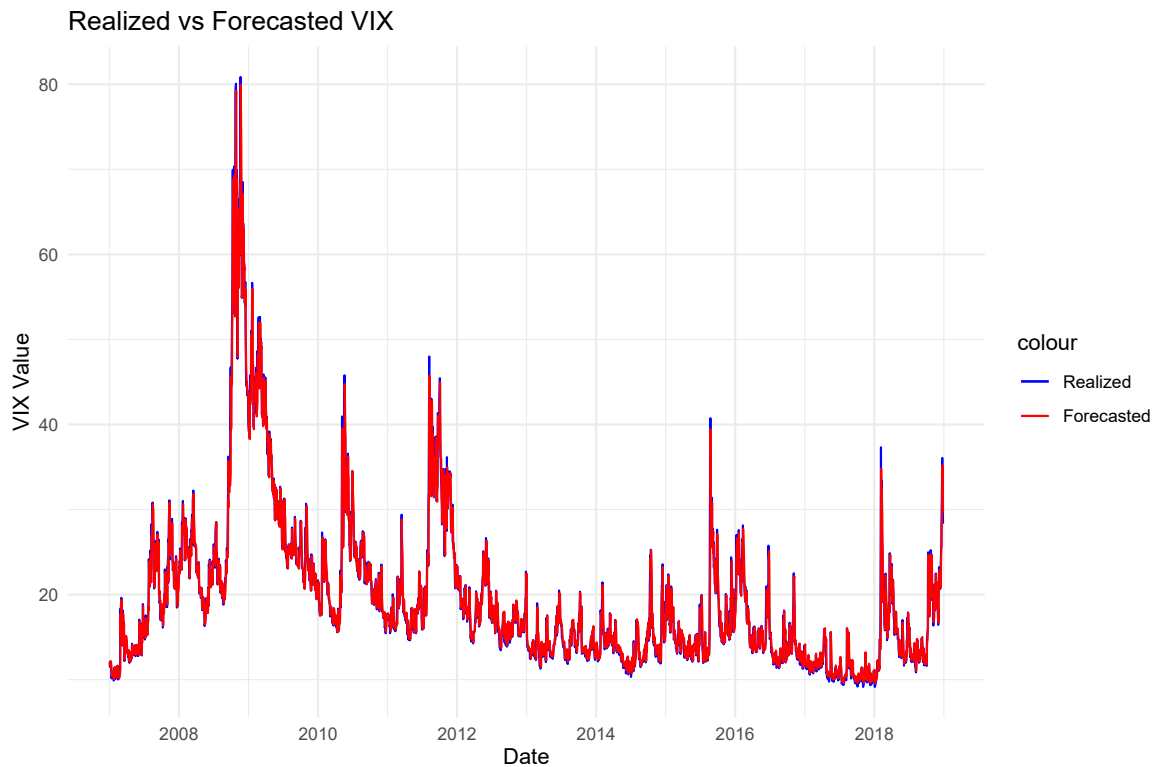


Figure 3: Comparing the realized VIX with JIVR model forecasts

7 Conclusion

This study develops the JIVR model, a characterization of the joint dynamics of the S&P 500 index and of its associated implied volatility surface. The approach is reminiscent of the dynamic Nelson-Siegel model of Diebold & Li (2006); the current implied volatility surface is used as an input to the model, thereby greatly enhancing the ability of the approach to depict current market conditions accurately. The parametric model of François et al. (2022) is leveraged to decompose the implied volatility surface into contributions from five economically interpretable factors. The parametric model has been shown in François et al. (2022) to properly capture the implied volatility shape while supporting extrapolation beyond observable areas of the surface and leaving very limited room for static arbitrage when applied to real data.

The JIVR model relies on joint NGARCH-type dynamics with fat-tailed and asymmetric NIG innovations to represent the evolution of the five IV surface factors and of the S&P 500 log-returns. The NGARCH processes for log-returns and the long-term level of the IV surface both include two variance components and are anchored in respective proportions of the 1-month ATM IV. This novel characterization effortlessly integrates information from the IV surface into the variance dynamics of the S&P 500 log-returns and of the long-term IV factor. Such model specification proves consistent with the Carr & Wu (2016) postulate expressing a proportionality relationship between the implied volatility level and the volvol. All other IV factors are represented by regular NGARCH-NIG processes. The dependence structure between innovations of all factors and of the underlying return is captured by a Gaussian copula.

The estimation of the model is performed through least-squares and conventional maximum likelihood procedures and is quite seamless. Cramér-von Mises goodness-of-fit tests applied to the residuals illustrate the appropriateness of the model specification. The JIVR model is shown to provide a vastly superior fit to observations than conventional GARCH processes estimated independently on each of the factors and on the underlying returns. In particular, the marginal specification of the S&P 500 returns process extracted from the JIVR model exhibits significantly higher performance than a standalone GARCH process, thereby highlighting the substantial added value of information borrowed from the IV surfaces when forecasting the S&P 500 return distribution.

Three exercises are conducted to demonstrate the capability of the JIVR model to accurately generate scenarios for the future implied volatility surface and the underlying return. The first exercise consists in assessing the risk of straddle and strangle positions with various times-to-maturity and a 5-day forecasting horizon. Proportions of VaR breaches confirm the overall accuracy of VaR estimates provided by the model. The second exercise compares the predictive performance of the JIVR model with that of a conventional time series counterpart model to forecast the VIX index distribution. Diebold-Mariano tests applied on out-of-sample likelihood scores confirm that the outperformance provided by the JIVR model is statistically significant. The third exercise reports the RMSEs for the 1-day ahead VIX point forecasts obtained from the JIVR model and several GARCH-type benchmarks. Our approach yields RMSEs lower by a factor of at least 1.5.

Early attempts at jointly modelling the underlying asset return and the corresponding IV surface dynamics have been facing implementation challenges that were initially deemed unsurmountable. The JIVR model proposed here shows that effective S&P 500 index op-

tions risk management is feasible when asset returns with non-Gaussian NGARCH processes and two-factor volatility are combined with a robust, parametric IV surface specification. More importantly, our contribution suggests that this stream of literature deserves further investigation, as it may have much more applicability than originally thought. Further tests involving option pricing and option replication, applied to other types of underlying assets, will help determine the exact potential of this approach.

References

- Andersen, T.G., Fusari, N. & Todorov, V. (2015). The risk premia embedded in index options. *Journal of Financial Economics*, 117(3):558–584.
- Azzone, M. & Baviera, R. (2022). Additive normal tempered stable processes for equity derivatives and power-law scaling. *Quantitative Finance*, 22(3):501–518.
- Bakshi, G., Cao, C. & Chen, Z. (1997). Empirical performance of alternative option pricing models. *The Journal of Finance*, 52(5):2003–2049.
- Barndorff-Nielsen, O.E. (1998). Processes of normal inverse Gaussian type. *Finance and Stochastics*, 2(1):41–68.
- Barndorff-Nielsen, O.E., Mikosch, T. & Resnick, S.I. (2001). *Lévy processes: Theory and applications*. Springer Science & Business Media.
- Bates, D.S. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *The Review of Financial Studies*, 9(1):69–107.
- Bates, D.S. (2000). Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics*, 94(1-2):181–238.
- Bates, D.S. (2022). Empirical option pricing models. *Annual Review of Financial Economics*, 14:369–389.
- Black & Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654.

- Bollerslev, T. & Todorov, V. (2011). Tails, fears, and risk premia. *The Journal of Finance*, 66(6):2165–2211.
- Broadie, M., Chernov, M. & Johannes, M. (2007). Specification and risk premiums: The information in S&P 500 futures options. *Journal of Finance*, 62(3):1453–1490.
- Carr, P. & Wu, L. (2004). Time-changed Lévy processes and option pricing. *Journal of Financial Economics*, 71(1):113–141.
- Carr, P. & Wu, L. (2016). Analyzing volatility risk and risk premium in option contracts: A new theory. *Journal of Financial Economics*, 120(1):1–20.
- Carr, P. & Wu, L. (2020). Option profit and loss attribution and pricing: A new framework. *The Journal of Finance*, 75(4):2271–2316.
- CBOE (2014). The CBOE volatility index-VIX. *White Paper*.
- Christoffersen, P., Elkamhi, R., Feunou, B. & Jacobs, K. (2010). Option valuation with conditional heteroskedasticity and nonnormality. *The Review of Financial Studies*, 23(5):2139–2183.
- Christoffersen, P., Jacobs, K. & Ornathanalai, C. (2013). GARCH option valuation: theory and evidence. *The Journal of Derivatives*, 21(2):8–41.
- Christoffersen, P., Jacobs, K., Ornathanalai, C. & Wang, Y. (2008). Option valuation with long-run and short-run volatility components. *Journal of Financial Economics*, 90(3):272–297.

- Cont, R. & Da Fonseca, J. (2002). Dynamics of implied volatility surfaces. *Quantitative Finance*, 2(1):45–60.
- Cont, R. & Vuletić, M. (2023). Simulation of arbitrage-free implied volatility surfaces. *Applied Mathematical Finance*, 30(2):94–121.
- Daglish, T., Hull, J. & Suo, W. (2007). Volatility surfaces: theory, rules of thumb, and empirical evidence. *Quantitative Finance*, 7(5):507–524.
- Diebold, F. & Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–63.
- Diebold, F.X. & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2):337–364.
- Duan, J.C. (1995). The GARCH option pricing model. *Mathematical Finance*, 5(1):13–32.
- Duffie, D., Pan, J. & Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376.
- Fengler, M.R. (2006). *Semiparametric modeling of implied volatility*. Springer Science & Business Media.
- François, P., Galarneau-Vincent, R., Gauthier, G. & Godin, F. (2022). Venturing into uncharted territory: An extensible implied volatility surface model. *Journal of Futures Markets*, 42(10):1912–1940.
- Glosten, L.R., Jagannathan, R. & Runkle, D.E. (1993). On the relation between the expected

- value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801.
- Hansen, P.R., Huang, Z., Tong, C. & Wang, T. (2024). Realized GARCH, CBOE VIX, and the volatility risk premium. *Journal of Financial Econometrics*, 22(1):187–223.
- Heston, S.L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343.
- Heston, S.L. & Nandi, S. (2000). A closed-form GARCH option valuation model. *The Review of Financial Studies*, 13(3):585–625.
- Huang, Z., Wang, T. & Hansen, P.R. (2017). Option pricing with the realized GARCH model: An analytical approximation approach. *Journal of Futures Markets*, 37(4):328–358.
- Hull, J. & White, A. (1987). The pricing of options on assets with stochastic volatilities. *The Journal of Finance*, 42(2):281–300.
- Israelov, R. & Kelly, B.T. (2017). Forecasting the distribution of option returns. *Available at SSRN 3033242*.
- James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). *An introduction to statistical learning : with applications in R*, volume 112. Springer.
- Kupiec, P.H. et al. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3(2):73–84.
- Ledoit, O. & Santa-Clara, P. (1998). Relative pricing of options with stochastic volatility. *University of California-Los Angeles finance working paper*, pages 9–98.

- Merton, R.C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1-2):125–144.
- Newey, W.K. & McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, 4:2111–2245.
- Oh, D.H. & Park, Y.H. (2022). GARCH option pricing with volatility derivatives. *Journal of Banking & Finance*, page 106718.
- Schönbucher, P.J. (1999). A market model for stochastic implied volatility. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 357(1758):2071–2092.
- Yan, S. (2011). Jump risk, stock returns, and slope of implied volatility smile. *Journal of Financial Economics*, 99(1):216–233.
- Zhu, Y. & Avellaneda, M. (1998). A risk-neutral stochastic volatility model. *International Journal of Theoretical and Applied Finance*, 1(2):289–310.

A The underlying asset drift specification

A necessary condition for the absence of arbitrage opportunities is the existence of a martingale measure \mathbb{Q} under which the discounted value of the underlying asset (adjusted for dividends) is a martingale. A widely used approach in the option pricing literature is the conventional Esscher transform, leading to a Radon-Nikodym of the form

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp \left(- \sum_{t=1}^T \left(\lambda \sqrt{h_{t+1,R}\Delta} \epsilon_{t+1,R} + \psi \left(-\lambda \sqrt{h_{t+1,R}\Delta} \right) \right) \right).$$

Because $\exp \left(\log \frac{S_{t+1}}{S_t} - r_{t+1} + q_{t+1} \right) = R_{t+1}$, the martingale condition implies

$$\begin{aligned} 1 &= \mathbb{E}^{\mathbb{Q}} \left[\frac{\exp \left(- \sum_{u=1}^{t+1} (r_u - q_u) \right) S_{t+1}}{\exp \left(- \sum_{u=1}^t (r_u - q_u) \right) S_t} \middle| \mathcal{F}_t \right] \\ &= \mathbb{E}^{\mathbb{P}} \left[\frac{\frac{d\mathbb{Q}}{d\mathbb{P}} \Big|_{\mathcal{F}_{t+1}}}{\frac{d\mathbb{Q}}{d\mathbb{P}} \Big|_{\mathcal{F}_t}} \exp(R_{t+1}) \middle| \mathcal{F}_t \right] \\ &= \mathbb{E}^{\mathbb{P}} \left[e^{-\lambda \sqrt{h_{t+1,R}\Delta} \epsilon_{t+1,R} - \psi(-\lambda \sqrt{h_{t+1,R}\Delta})} e^{\xi_{t+1} - \psi(\sqrt{h_{t+1,R}\Delta}) + \sqrt{h_{t+1,R}\Delta} \epsilon_{t+1,R}} \middle| \mathcal{F}_t \right] \\ &= \exp \left(\xi_{t+1} + \psi \left(\sqrt{h_{t+1,R}\Delta} - \lambda \sqrt{h_{t+1,R}\Delta} \right) - \psi \left(-\lambda \sqrt{h_{t+1,R}\Delta} \right) - \psi(\sqrt{h_{t+1,R}\Delta}) \right). \end{aligned}$$

Obtaining a solution to the above equation for all values of λ leads to (4).

B Model components' contribution to performance

The JIVR model, described in Section 4, encompasses multiple features such as (i) GARCH-type stochastic volatilities, (ii) non-Gaussian innovations following NIG distributions, (iii) the leverage effect captured in the NGARCH's asymmetric variance responses as in Duan (1995), (iv) a two-component stochastic volatility process for the underlying asset log-returns and the long-term level factor β_1 , and (v) a Gaussian copula to capture the dependence structure between the innovations.

To study the contribution of each model feature, we compare the performance of a sequence of nested sub-models with incremental complexity obtained by repeatedly adding a model feature to the previous element of the sequence.

There are six sub-models: (1) *BS*, the Black-Scholes model, (2) *Gaussian GARCH*, (3) *Gaussian NGARCH*, (4) *NIG NGARCH*, (5) the *Indep. JIVR* model where the Gaussian copula is set to the identity matrix and (6) the *JIVR* model. The specifications of the first four models are described in Table 10. The fifth model (*Indep. JIVR*) is described by Equations (3)-(6), but the innovations across the underlying return and IV factor components are assumed to be independent. Finally, the *JIVR model* from Section 4 includes the copula to capture the dependence between the various components.

Table 10: Nested sub-models

	S&P 500 log-returns R_t		Volatility and surface factor coefficients β_i	
	$R_t = (r_{t,1/\Delta} - q_t + \lambda h_{t+1,R}) \Delta - \psi(\sqrt{h_{t+1,R}\Delta}) + \sqrt{h_{t+1,R}\Delta}\epsilon_{t+1,R}$		$\beta_i = \alpha_i + \sum_{j=1}^5 \theta_{i,j}\beta_{t,j} + \sqrt{h_{t+1,i}\Delta}\epsilon_{t+1,i}$	
BS	$h_{t+1,R} = h_R$	$\epsilon_{t,R} \sim \mathcal{N}(0, 1)$	$h_{t+1,i} = h_i$	$\epsilon_{t,R} \sim \mathcal{N}(0, 1)$
Gaussian GARCH	$h_{t+1,R} = \sigma_R^2 + \kappa_R (h_{t,R} - \sigma_R^2) + a_R h_{t,R} (\epsilon_{t,R}^2 - 1)$	$\epsilon_{t,R} \sim \mathcal{N}(0, 1)$	$h_{t+1,i} = \sigma_i^2 + \kappa_i (h_{t,i} - \sigma_i^2) + a_i h_{t,i} (\epsilon_{t,i}^2 - 1)$	$\epsilon_{t,R} \sim \mathcal{N}(0, 1)$
Gaussian NGARCH	$h_{t+1,R} = \sigma_R^2 + \kappa_R (h_{t,R} - \sigma_R^2) + a_R h_{t,R} (\epsilon_{t,R}^2 - 1 - 2\gamma_R \epsilon_{t,R})$	$\epsilon_{t,R} \sim \mathcal{N}(0, 1)$	$h_{t+1,i} = \sigma_i^2 + \kappa_i (h_{t,i} - \sigma_i^2) + a_i h_{t,i} (\epsilon_{t,i}^2 - 1 - 2\gamma_R \epsilon_{t,i})$	$\epsilon_{t,R} \sim \mathcal{N}(0, 1)$
NIG NGARCH	$h_{t+1,R} = \sigma_R^2 + \kappa_R (h_{t,R} - \sigma_R^2) + a_R h_{t,R} (\epsilon_{t,R}^2 - 1 - 2\gamma_R \epsilon_{t,R})$	$\epsilon_{t,R} \sim \text{NIG}^*$	$h_{t+1,i} = \sigma_i^2 + \kappa_i (h_{t,i} - \sigma_i^2) + a_i h_{t,i} (\epsilon_{t,i}^2 - 1 - 2\gamma_R \epsilon_{t,i})$	$\epsilon_{t,i} \sim \text{NIG}^*$

The first row displays the specification for the S&P 500 log-returns and the factor coefficients. The subsequent rows indicate the volatility process and the distribution of the innovations for their respective model. The NIG* refers to the standardized NIG distribution described in Appendix F.

Parameters from all sub-models are estimated through maximum likelihood, as described in Section 5. The log-likelihood performance metric is computed out-of-sample based on the expanding window method described in Section 6. The model performance is measured out-of-sample to ensure that overfitting does not unduly give an advantage to more complex sub-

models. To compare the log-likelihoods across sub-models, the out-of-sample log-likelihood and the corresponding ALR metric defined in Equation (11) are considered.

Panel A of Table 11 exhibits the out-of-sample log-likelihoods for each model. Results are presented for all constituents (i.e. the S&P 500 log-returns and the factor coefficients) individually, and the last row, *Joint*, presents the aggregated log-likelihood for the joint model. Panel B displays the ALR metrics, which are computed based on the log-likelihood of any model with respect to the previous one in the series of nested sub-models. The ALR is described in detail in Section 6.2. Again, results are presented for separate constituents and in aggregate.

First, the results clearly highlight the importance of including a GARCH-type volatility for all five factor coefficients and for the S&P 500 log-returns. The ALRs are all large and far above 1, indicating that the superiority of the performance of the Gaussian GARCH model over that of the BS is unequivocal. The addition of an asymmetric volatility response through the replacement of GARCH with NGARCH processes results in improvements mainly for the S&P 500 log-returns and for the time-to-maturity slope. For other factors, such a modification does not lead to much (if any) improvement in performance. Integrating NIG-distributed innovations positively impacts the fitting performance for the S&P 500 log-returns and for all five factor coefficients. The second-to-last column (*Indep. JIVR*) displays the relative performance of the full specification described in Equations (3)-(6) over that of the NIG-NGARCH model. With an ALR of 1.03 for the S&P 500 log-returns and 1.02 for the long-term level factor, the results highlight a large improvement stemming from the inclusion of the two-component variance process. Lastly, the last column, which depicts the impact of including the Gaussian copula, exhibits an ALR of 2.29. This highlights the

importance of taking the dependence structure into consideration.

Table 11: Out-of-sample model performance

	BS	Gaussian	Gaussian	NIG	JIVR Model	
	AR(1)	GARCH	NGARCH	NGARCH	Indep.	Copula
Panel A: Log-lik.						
S&P 500 log-returns	10,124	11,365	11,466	11,600	11,709	
Long-term level	13,788	15,617	15,620	15,785	15,861	
TmT Slope	7,115	8,681	8,843	9,304	9,319	
Moneyness Slope	15,392	15,674	15,672	15,709	15,709	
Smile attenuation	14,922	15,251	15,250	15,352	15,352	
Smirk	14,441	14,861	14,860	14,995	14,995	
Joint	75,785	81,452	81,713	82,748	82,947	85,873
Panel B: ALR						
S&P 500 log-returns		1.42	1.03	1.04	1.03	
Long-term level		1.68	1.00	1.05	1.02	
TmT Slope		1.56	1.05	1.14	1.01	
Moneyness Slope		1.08	1.00	1.01	1.00	
Smile attenuation		1.10	1.00	1.03	1.00	
Smirk		1.13	1.00	1.04	1.00	
Joint		4.99	1.08	1.34	1.06	2.29

The table exhibits log-likelihoods (Panel A) and the ALR (Panel B) for the aggregated out-of-sample period (2007-2020). The specification of the first four models (*BS*, *Gaussian GARCH*, *Gaussian NGARCH*, *NIG NGARCH*) are presented in Table 10 and the specification of last two models (*JIVR* Indep. and *JIVR* copula) correspond to the specification exhibited in Section 4, i.e. Equation (3)-(6). The last row (Joint) displays the log-likelihood for the joint models assuming independence between the log-returns and the five factors, except for the last column (*JIVR model copula*) where the Gaussian copula captures the dependence. Panel B exhibits the ALR metrics, which use the log-likelihood of the corresponding row and column of Panel A as the numerator input, and the log-likelihood of the corresponding row but the previous column of Panel A as the denominator input.

C Cramér-von Mises test

The Cramér-von Mises p -value is computed through a two-step process. The first step involves generating a cumulative distribution function for the Cramér-von Mises statistic

using a bootstrapping scheme.

More precisely, numerous sets of random variables following a NIG distribution with the parameters exhibited in Table 4 are simulated. Parameters of the NIG distribution are fitted over each simulated set. The Cramér-von Mises statistic is computed for all simulated sets using the fitted parameters. This method generates a large number of simulated Cramér-von Mises statistics, which forms the Cramér-von Mises test distribution. The second step consists in estimating the Cramér-von Mises statistic on the residuals. The p -value of the Cramér-von Mises statistic is obtained using the simulated Cramér-von Mises statistic's distribution, as computed in Step 1.

Algorithm 2 Cramér-von Mises test

```

for  $n = 1$  to  $N$  do
  for  $m = 1$  to  $M$  do
    draw  $x_{m,n} \sim NIG(\zeta_{NIG}, \phi_{NIG})$ 
  end for
   $[\tilde{\zeta}_{NIG}^{n*}, \tilde{\phi}_{NIG}^{n*}] = \arg \max_{\tilde{\zeta}_{NIG}^n, \tilde{\phi}_{NIG}^n} \sum_{j=1}^M \log \left( f_{NIG}(x_{j,n}; \tilde{\zeta}_{NIG}^n, \tilde{\phi}_{NIG}^n) \right)$ 
   $CV(n) = \frac{1}{12M} + \sum_{j=1}^M \left( \frac{2j-1}{2M} - F_{NIG}(x_{j,n}; \tilde{\zeta}_{NIG}^{n*}, \tilde{\phi}_{NIG}^{n*}) \right)^2$ 
end for
 $CVS = \frac{1}{12M} + \sum_{j=1}^M \left( \frac{2j-1}{2M} - F_{NIG}(r_j; \zeta_{NIG}, \phi_{NIG}) \right)^2$ 
 $p\text{-value} = \frac{(\sum_{j=1}^N \mathbb{1}_{\{CV(j) > CVS\}})}{N};$ 

```

In Algorithm 2, F_{NIG} refers to the cumulative distribution function of the NIG distribution, f_{NIG} is the density distribution function of the NIG distribution, and r is the vector of residuals.

D VaR coverage tests

In this section, the methodology for the VaR coverage test is presented for the specific case where $\alpha < 0.5$. The principle to compute the test for $\alpha > 0.5$ is identical, *mutatis mutandis*.

The VaR coverage test verifies if the proportion of ex-post VaR breaches are close to the model's α percentile level. The notation VaR_t^α represents the VaR estimate for a specific day t and confidence level α .

The test is performed following the backtesting procedure called the *hit sequence*, which is described in detail in Kupiec et al. (1995). The *hit sequence* of VaR breaches is defined as follows:

$$I_{t+d} = \begin{cases} 1 & \frac{V_{t+d}-V_t}{V_t} < \text{VaR}_{t+d}^\alpha, \\ 0 & \text{otherwise} \end{cases}$$

where V_t is the value of the strategy at time t . If the VaR calculation methodology is well specified, the frequency of VaR breaches (i.e., elements I_{t+d} equal to 1) should be close to α .

More precisely, the *hit sequence* should be composed of independent and identically distributed Bernoulli random variables. The null hypothesis of the test is thus $H_0 : \sum_{t=d+1}^N I_t \sim \text{Binomial}(\alpha, N-d)$. The test is performed as follows:

$$\begin{aligned} p &= \frac{1}{N-d} \sum_{t=d+1}^N I_t, \\ L_1 &= (N-d) ((1-p) \log(1-\alpha) + p \log(\alpha)), \\ L_2 &= (N-d) ((1-p) \log(1-p) + p \log(p)). \end{aligned}$$

According to the likelihood ratio test, $-2(L_1 - L_2) \sim \chi_1^2$ distribution where χ_1^2 is a Chi-squared distribution with one degree of freedom. The p -value of the test is computed as $1 - \text{CDF}_{\chi_1^2}(-2(L_1 - L_2))$.

E Diebold & Mariano (1995) test

In Section 6.2, we use the Diebold & Mariano (1995) test to compare the predictive performance of the JIVR model with that of the direct model from a statistical standpoint. In our particular case, the considered predictive performance measure is the log-likelihood. Let $\mathcal{L}_t^{(1)}$ and $\mathcal{L}_t^{(2)}$ be log-likelihoods at time t of the JIVR model and the direct model, respectively.

We refer to the time-series of $d_t = \mathcal{L}_t^{(1)} - \mathcal{L}_t^{(2)}$ as the loss differential. The sample autocovariance γ_k at lag k is defined as

$$\gamma_k = \frac{1}{N} \sum_{t=k+1}^N (d_t - \bar{d})(d_{t-k} - \bar{d}), \quad \bar{d} = \frac{1}{N} \sum_{t=1}^N d_t.$$

The Diebold & Mariano (1995) statistic is computed as $DM = \frac{\bar{d}}{\sqrt{(\gamma_0 + 2 \sum_{k=1}^{h-1} \gamma_k)/N}}$ where $h = \lfloor N^{1/3} + 1 \rfloor$ and $DM \sim \mathcal{N}(0, 1)$. The p -value of the test is computed as $1 - CDF_{\mathcal{N}}(DM)$.

F Standardized Normal Inverse Gaussian probability density function

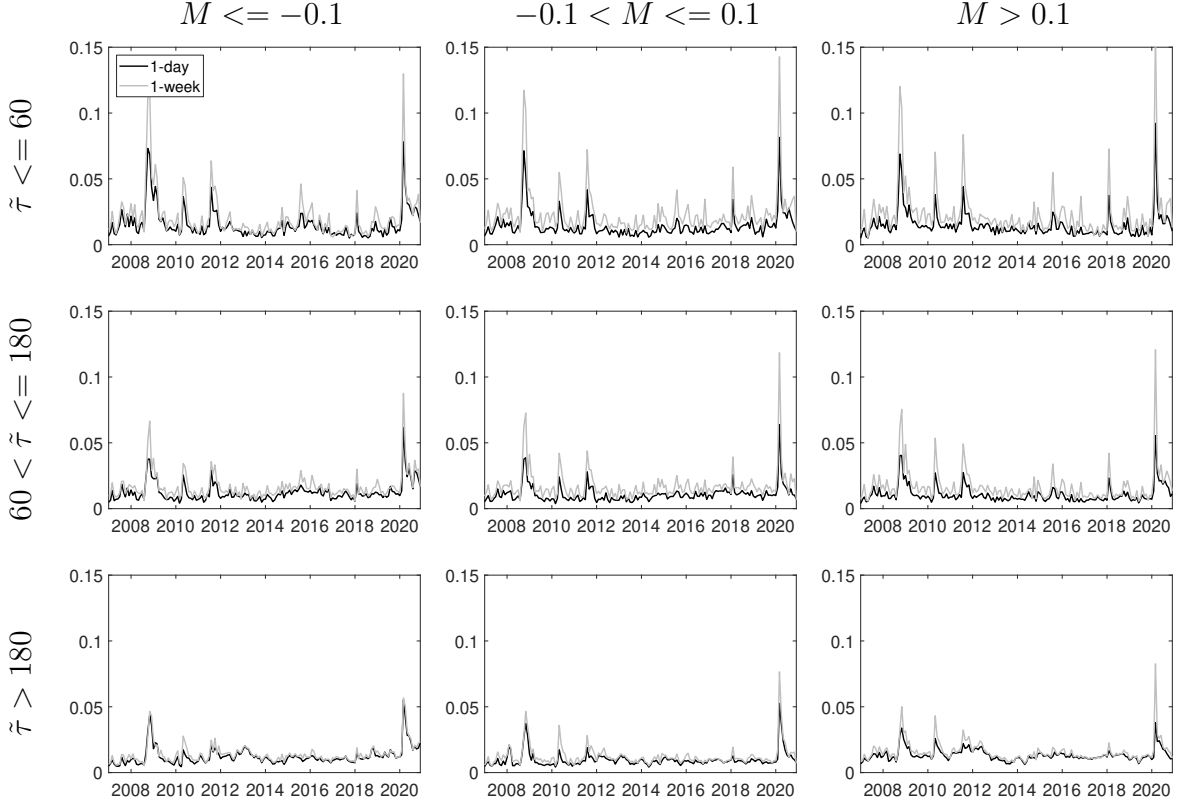
The probability density function of the standardized NIG distribution is defined as:

$$f(x) = \frac{B_1 \left(\sqrt{\frac{\phi^6}{\phi^2 + \zeta^2} + (\phi^2 + \zeta^2) \left(x + \frac{\phi^2 \zeta}{\phi^2 + \zeta^2} \right)^2} \right)}{\pi \sqrt{\frac{1}{\phi^2 + \zeta^2} + \frac{\phi^2 + \zeta^2}{\phi^6} \left(x + \frac{\phi^2 \zeta}{\phi^2 + \zeta^2} \right)^2}} e^{\left(\frac{\phi^4}{\phi^2 + \zeta^2} + \zeta \left(x + \frac{\phi^2 \zeta}{\phi^2 + \zeta^2} \right) \right)}, \quad x \in \mathbb{R},$$

where B_1 denotes the modified Bessel function of the second kind with index 1, which is described in Barndorff-Nielsen et al. (2001).²³

²³This density is obtained by replacing β^{NIG} and γ^{NIG} with ζ and ϕ , respectively, in the common $(\alpha^{NIG}, \beta^{NIG}, \delta^{NIG}, \mu^{NIG})$ -specification of the NIG density and by imposing a null mean and unit variance to express δ^{NIG}, μ^{NIG} in terms of $\alpha^{NIG}, \beta^{NIG}$ (or alternatively $\gamma^{NIG} = \sqrt{(\alpha^{NIG})^2 - (\beta^{NIG})^2}$).

the 2008 Financial crisis and the 2020 COVID-19 pandemic. This result is to be expected because periods of higher volatility imply more considerable variation of the IV surface, which drives the observed volatility surfaces further away from the predicted surface.



Each panel displays the time series of the monthly averaged IVRMSE per bucket computed using the model forecasted IV surfaces for two horizons: one day and one week. For every month, the daily IVRMSE per bucket is computed for the desired horizon and then averaged. The daily IVRMSE per bucket is obtained by computing the RMSE between the observed surface IV and the forecasted IV of all options whose characteristics (moneyness and time-to-maturity) fall in the corresponding bucket.

Figure 6: Out-of-sample monthly RMSE across moneyness and time-to-maturity

To assess the added value of forecasting the IV surface, the IVRMSEs computed from the JIVR model forecasts are compared to those obtained with a base case approach. Such a base case approach relies on the last observed fitted IV surface, i.e. uses the fitted β s at time t as the IV surface forecast at time $t + d$. Table 13 exhibits the average IVRMSE computed over the out-of-sample period for the JIVR model and the base case. Results show