SIAM J. FINANCIAL MATH.
Vol. 7, pp. 382–417

Modeling the Variance Risk Premium of Equity Indices: The Role of Dependence and Contagion*

Andrea Granelli† and Almut E. D. Veraart†

Abstract. The variance risk premium (VRP) refers to the premium demanded for holding assets whose variance is exposed to stochastic shocks. This paper identifies a new modeling framework for equity indices and presents for the first time explicit analytical formulas for their VRP in a multivariate stochastic volatility setting, which includes multivariate non-Gaussian Ornstein–Uhlenbeck processes and Wishart processes. Moreover, we propose to incorporate contagion within the equity index via a multivariate Hawkes process and find that the resulting dynamics of the VRP represent a convincing alternative to the models studied in the literature up to date. We show that our new model can explain the key stylized facts of both equity indices and individual assets and their corresponding VRP, while some popular (multivariate) stochastic volatility models may fail.

Key words. variance risk premium, quadratic variation, stochastic volatility, Lévy processes, leverage effect, Hawkes process, self-excitement, contagion, change of measure

AMS subject classifications. 91G99, 60G51, 60G55

DOI. 10.1137/15M1011822

1. Introduction.

1.1. Definition and properties of the variance risk premium.

The variance risk premium (VRP) is a concept of great importance in finance. It can be described as the premium that investors demand for holding assets whose variance is exposed to stochastic shocks and is likely to change over time. It is commonly defined in the literature as the difference between the *expected* risk neutral and physical quadratic variations.

A proper understanding of the VRP is crucial in many areas of finance, like risk management and asset allocation. It has also recently been highlighted how the VRP can predict aggregate stock market returns in separate economies and even across countries (see [9]).

The importance of the VRP is also evident in derivatives pricing. Vanilla options on equity indices are typical examples of derivatives products which are positively exposed to a marketwide increase in variance. To understand whether the VRP is priced in the market, in the sense that investors consider this exposure valuable, it is natural to analyze variance swaps written on equity indices. Historical data show that, on average, implied variance has been higher than realized variance (see [12]), yielding a *negative variance risk premium*. The very existence of a market for variance swaps is a proof that the VRP is present and priced

*Received by the editors March 9, 2015; accepted for publication (in revised form) April 11, 2016; published electronically June 14, 2016.
http://www.siam.org/journals/sifin/7/M101182.html
†Department of Mathematics, Imperial College London, 180 Queen's Gate, SW7 2AZ London, UK, (a.granelli12@imperial.ac.uk, a.veraart@imperial.ac.uk). The first author received financial support from the Department of Mathematics, Imperial College London. The second author received financial support through a Marie Curie FP7 integration grant within the 7th European Union Framework Programme.

in the market. In addition, the market volume for variance swaps has increased steeply over the past few years [11], indicating the need for a deeper understanding of the VRP.

The work in [12] implies that investors who have long positions in variance accept earning lower returns than what could be justified by other market factors, because they are hedged against an increase in the variance of the stock market. Thus, an increase in variance is perceived as an unfavorable event by investors.

1.2. Main contribution of the present work. Most of the work in the literature about the VRP is of an empirical nature. This aim of this paper is instead to provide a unified framework to study the VRP from a theoretical perspective. We aim to fill the gap in the literature by proposing a stochastic model whose aim is to reproduce the empirical features of the VRP.

There are two fundamental characteristics of the VRP that have been found empirically. The first one is that it is stochastic. As we will show, this imposes constraints on the specifications that a stochastic model for an index must have. The second fundamental characteristic was highlighted in [16]: not only is the VRP present and negative for indices, but also it is not present, or is weakly positive, for the individual stocks within the index. Thus, they highlight the existence of a premium originated by the dependence between the stocks, the so-called correlation risk premium.

From a modeling point of view, it becomes crucial then to devise a model that incorporates dependence between the different assets. The vast majority of the literature focuses on univariate models for the index, but then the modeling of correlation does not achieve a sufficient level of complexity to justify and explain the presence of the correlation risk premium.

Our work proposes to model an equity index via a multivariate stochastic price model, which is both parsimonious and powerful enough to account for different forms of dependence between the assets within the index.

We show that our model replicates the empirical findings and in addition produces analytical and explicit formulas for the VRP which very rarely have appeared in the literature. We give mathematical proofs that establish why the dependence between the assets is the main driver of the VRP and find which classes of processes are a priori more appropriate to be used in this framework.

1.2.1. Description and properties of the model. In building our model we followed the findings of [30], who prescribes that the VRP should be driven by both a stochastic volatility component and jumps in the prices.

We first indicate two popular alternatives for the stochastic volatility component of the index: the multivariate non-Gaussian Ornstein–Uhlenbeck (OU) process and the Wishart process. Both are very popular choices in the stochastic volatility literature, and they both account for cross-correlation between the stocks. We obtain the surprising result that the first of these two models is not capable of producing sufficiently complex dynamics of the corresponding VRP. In fact, it is highlighted in [30], [10], [9], and [15] that the VRP exhibits stochastic fluctuations, while the aforementioned stochastic volatility model implies deterministic dynamics, under the widely used structure preserving change of measure. We will also show that the result for the Wishart model is instead heavily dependent on the particular change of measure used. In addition, these two models do not allow us to split between individual and correlated stocks in the final VRP formulas.

In order to replicate the empirical fact that individual stocks do not create VRP, while pairs do, we suggest a possible alternative model, where the stochastic volatility matrix is *diagonal* and consists of non-Gaussian OU components, in the presence of correlated Brownian motion. This choice is motivated by the need of having a jump component in the variances, as proven in [30], but at the same time introducing a model which is analytically tractable and allows for an explicit derivation of results. In particular, we obtain deterministic dynamics for the VRP of single stocks while we show that the contribution brought by the dependence between stocks exhibits stochastic fluctuations.

The second driver we make use of is a multivariate jump process. The importance of a jump component in the price is well-known and documented. The presence of jumps offers a solution to the problem of calibrating option prices to market data, being capable of describing smiles and skews in volatility and performing well across different maturities, unlike diffusion-based local volatility models. Moreover, they perform well if used to fit historical price data (see, e.g., [13, pp. 13–14]). We first discuss the contribution of a pure-jump Lévy process. The model we propose, consisting of the diagonal non-Gaussian OU stochastic volatility and the multivariate Lévy process, has all the features that are observed empirically in [16] and [30].

It is to be stressed, though, that the empirical study performed in [16] used data from 1993 until 2003. The more recent financial crisis has highlighted the need for more sophisticated stochastic models to be used in mathematical finance. While Lévy processes belong to the classical candidates when modeling jumps in asset prices, we also focus on an alternative class of jump processes which has recently attracted a lot of attention in financial applications: the class of Hawkes processes.

Hawkes processes are counting processes that allow for self- and mutual excitation, thanks to the peculiar form of their intensities and the stochastic differential equations that they satisfy. They provide our model with a different sort of dependence, which has not yet been explored in the modeling of equity indices: They account for the *contagion* effect between the stocks. Hawkes processes have been used in the financial literature to model default times (see, e.g., [19]) and, more recently, to model stock returns during crises; see [1], where Hawkes processes are successfully used to model jump clusters, financial contagion, and self-excitement across six large scale economies. We will see that the contagion effect is a device through which the VRP obtains stochastic dynamics.

The Hawkes process was first introduced in [21] in its most basic form and subsequently used, for example, for the modeling of earthquakes, thanks to its peculiar characteristic of being self-exciting. Every component of the intensity $\mathbf{\Lambda}_t$ is affected by all the other ones and in turn affects them: When a component of the Hawkes process exhibits a jump, the corresponding intensity increases. This is the *self-excitation* property of the Hawkes process. This, in turn, triggers an increase also in the other elements of the intensity vector, causing a boost in the probability of subsequent jumps in the other components. This property is called *mutual excitation*. All these jumps subsequently affect all the intensities, feeding this circle and thus creating a large probability of encountering jump clusters. In a model where jumps are intrinsically *rare events*, this scenario would be extremely unlikely. In absence of jumps, the intensities quickly revert to their original value.

While Hawkes processes have been shown to be a good model to describe the interactions between different economies, we propose to use them to model stocks in a single market, using the intuition that the contagion effect studied in [1] across markets can also appear within one market.

The introduction of the Hawkes process brings a substantial contribution to the VRP. We show that the Hawkes intensity process can be identified as a primary driver for the VRP of the index.

1.3. Mathematical contributions.

There are also several mathematical stand-alone contributions in this paper. Along with the results we obtain concerning the dynamics of the risk premium, we employ a technique to overcome the mathematical intractability of computing conditional expectations of square roots of processes with given dynamics. We do so by using an integral representation that leads to an analytic, exact formula, which we use to read off properties of the VRP. We also give some examples where we apply our results to explicitly compute the premium in the case of the Gamma process. Such a technique has, however, wider application and could be useful for scopes beyond the ones studied in this paper.

Second, we provide useful explicit formulas for the conditional expectation of the Hawkes intensity process, in both the univariate and the multivariate case. These formulas complement the existing literature on the distributional properties of this class of processes; see, for example, [19].

We also provide a proof of the existence of an equivalent martingale measure. Indeed, all our analytical results lay upon the fact that our model admits a *structure preserving* equivalent martingale measure. A risk neutral martingale measure $\mathbb{Q}$ will be called *structure preserving* if both the stock price model and the stochastic volatility have the same features under both the physical measure $\mathbb{P}$ and $\mathbb{Q}$: They preserve their probabilistic properties and follow dynamics driven by the same classes of processes, although different values for the parameters are allowed. The result of an existence of such a measure for a univariate model with similar specifications, but without the Hawkes process, was provided for the first time in [26] and then extended for the multivariate case in [25]. In our work we solve the problem of finding the class of measures under which the Hawkes process maintains its characteristics, drawing upon results in [29], and therefore we have a characterization of the class of structure preserving changes of measure for our multivariate model.

The rest of this article is structured as follows. Section 2 gives the detailed description of the model we use and derives dynamics of important quantities to be used subsequently; sections 3 and 4 contain, respectively, the derivations of the contribution to the total VPR given by the diffusion and the jump components. Finally, section 5 summarizes our main results. In Appendix E we explicitly describe how to construct a *structure preserving* change of measure for our model. The appendix also contains other selected proofs.

2. Model assumptions and properties.

In this section, we present the model assumptions in detail and present some important properties of our modeling framework. In what follows we will let $\mathbb{R}^+ := [0, \infty)$. Further, for a subset E of $\mathbb{R}^n$, $\mathcal{B}(E)$ will denote the class of Borel subsets of E.

We will consider a stochastic basis $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{0\leq t\leq T}, \mathbb{P})$ on a fixed time horizon $[0, T]$, with the filtration $(\mathscr{F}_t)$ satisfying the usual assumptions of completeness and right continuity.

Suppose the equity index is composed of D stocks, whose log-prices are denoted by $X^{(1)}, \dots, X^{(D)}$. We model them simultaneously via a D-dimensional semimartingale $\mathbf{X}$ whose differential dynamics are

$$d\mathbf{X}_t = (\boldsymbol{\mu}_t + \boldsymbol{\beta}(\boldsymbol{\Sigma}_t))dt + \boldsymbol{\Sigma}_t^{\frac{1}{2}} d\mathbf{W}_t + d\mathbf{J}_t, \tag{1}$$

where the individual components will be specified in the following. We will think of $(\mathscr{F}_t)$ as the augmentation of the filtration generated by the process $\mathbf{X}$ up to time t.

Assumption 1. The price of the ith stock is given by $S_t^{(i)} = \exp(X_t^{(i)})$.

Assumption 2. The vector process $\boldsymbol{\mu}_t$ is D-dimensional, and the function $\boldsymbol{\beta} : \mathbb{R}^{D\times D} \to \mathbb{R}^D$ is linear.

Assumption 3. The process $\mathbf{W}_t = (W_t^{(1)}, \dots W_t^{(i)}, \dots W_t^{(D)})^\top$ is a D-variate Brownian motion.

Assumption 4. The process $\boldsymbol{\Sigma}^{\frac{1}{2}}$ is the stochastic volatility matrix. It is defined to be the square root of the symmetric, positive-definite stochastic matrix $\boldsymbol{\Sigma}$. That means that we define $\boldsymbol{\Sigma}^{\frac{1}{2}}$ to be the *unique symmetric positive definite* matrix such that $\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}} = \boldsymbol{\Sigma}$ (see [28, p. 115] for a proof of existence and uniqueness).

In this work we will look principally at two different methods to model $\boldsymbol{\Sigma}$, or its square root $\boldsymbol{\Sigma}^{\frac{1}{2}}$, which we will call *constant* or *dynamic correlation models*, similar to the terminology introduced in [8] and then generalized in subsequent work (e.g., [18]).

Dynamic correlation models. In this first class of model, the stochastic matrix $\boldsymbol{\Sigma}$ is modeled through multivariate dynamics. We call it a *dynamic correlation model*, because the dependence between the components is explicitly modeled through the off-diagonal elements of the matrix, which are different from 0 and stochastic. The Brownian motion $\mathbf{W}$ is assumed to be standard and consisting of independent components. First, we will propose a full multivariate model for $\boldsymbol{\Sigma}$, the *multivariate non-Gaussian OU* process, as appearing in [25]. The derivation of the results is provided in section 3.1.1. Second, we compute the VPR for another popular model in the literature of stochastic volatility modeling: the *Wishart* model. See section 3.1.2.

Constant correlation models. We consider the case where $\boldsymbol{\Sigma}$ will be a *diagonal* matrix of OU processes. The model is therefore called a *constant correlation model*, since the off-diagonal elements of $\boldsymbol{\Sigma}$ are zero. Nonetheless, we allow for dependence both in the background driving Lévy processes (BDLPs) of each component and in the Brownian motion $\mathbf{W}$, which will be correlated. Details are given in section 3.2.

Assumption 5. The process $\mathbf{J}_t = (J_t^{(1)}, \dots J_t^{(i)}, \dots J_t^{(D)})^\top$ is a D-dimensional pure-jump process. For each component $J^{(i)}$, we define its jump measure $M^{(i)}(dt, dx)$ on $\mathcal{B}\big((0,T)\big) \times \mathcal{B}(\mathbb{R})$ via the following: for $0 < s < t < T$ and $A \in \mathcal{B}(\mathbb{R})$ we set $M^{(i)}((s,t), A) = \#\{l \in (s,t) | \Delta J_l^{(i)} \in A\}$.

Remark 6. The drift process $\boldsymbol{\mu}$ will play no role in the following discussion, and it is left unspecified. We will later consider the case when $\mathbf{J}$ is a counting process; in that situation, we can think that $\boldsymbol{\mu}$ contains the compensator of $\mathbf{J}$.

Assumption 7. The processes $\mathbf{N}_t, \mathbf{J}_t, \mathbf{W}_t$ are independent.

2.1. Dynamics of the price process. An application of Itô's lemma for semimartingales allows us to describe the dynamics of the stock price $S_t^{(i)} = \exp(X_t^{(i)})$:

$$d\exp(X_t^{(i)}) = dS_t^{(i)} = S_{t-}^{(i)}\Bigg[\Bigg(\mu_t^{(i)} + \Bigg(\beta^{(i)}(\boldsymbol{\Sigma}_t) + \frac{1}{2}\Sigma_t^{(ii)}\Bigg)\Bigg)\, dt + \sum_j \Sigma_t^{\frac{1}{2}(ij)}\, dW_t^{(j)} + dJ_t^{(i)} + \int_{\mathbb{R}} (e^x - 1 - x)\, M^{(i)}(dt, dx)\Bigg]. \tag{2}$$

If we write $a_t^{(i)} = \mu_t^{(i)} + (\beta^{(i)}(\boldsymbol{\Sigma}_t) + \frac{1}{2}\Sigma_t^{(ii)})$, then the above expression becomes

$$\frac{dS_t^{(i)}}{S_{t-}^{(i)}} = a_t^{(i)}\, dt + \sum_j \Sigma_t^{\frac{1}{2}(ij)}\, dW_t^{(j)} + dJ_t^{(i)} + \int_{\mathbb{R}} (e^x - 1 - x) M^{(i)}(dt, dx). \tag{3}$$

2.2. Definition and dynamics of the value-weighted index process. As we mentioned in the introduction, the aim of this study is to model the VRP of equity *indices*. Let us briefly explain how a value-weighted index can be constructed. Here we follow [16], where an empirical study based on the S&P 100, which is a market capitalization-weighted index with quarterly rebalancing, is performed.

Define a value-weighted index as a derivative whose value depends on all the stocks in the market. The weight each stock gets in the index is the ratio between its value and the total value of the stocks in the market. To obtain a market capitalization-weighted index, we can view each single stock $X^{(i)}$ as the (logarithmic) market capitalization process of the ith company.

We will show that the dynamics of the value-weighted index implied by our model give rise to a VRP with the empirical properties found in the literature.

We first consider the sum of the stock prices, denoted by I_t. Let $\mathbf{1}$ be a D-dimensional vector of 1-s; then $I_t = \mathbf{1}^\top \mathbf{S}_t$. We can write its dynamics using a multivariate notation:

$$dI_t = d(\mathbf{1}^\top \mathbf{S}_t) = \mathbf{S}_{t-}^\top \mathbf{a}_t\, dt + \mathbf{S}_{t-}^\top \boldsymbol{\Sigma}_t^{\frac{1}{2}} d\mathbf{W}_t + \mathbf{S}_{t-}^\top\, d\mathbf{J}_t + \int_{\mathbb{R}} \mathbf{S}_{t-}^\top (e^x - 1 - x)\, \mathbf{M}(dx, dt). \tag{4}$$

In the value-weighted index, the weight $w_t^{(i)}$ is the ratio between the value of the stock $S_t^{(i)}$ and the value of the index process I_t. We will call $\mathbf{w}_t$ the D-dimensional vector of weights:

$$\mathbf{w}_t = \left(\frac{S_t^{(1)}}{I_{t-}} \cdots \frac{S_t^{(i)}}{I_{t-}} \cdots \frac{S_t^{(D)}}{I_{t-}}\right)^\top = \frac{\mathbf{S}_t}{\mathbf{1}^\top \mathbf{S}_{t-}}.$$

And finally we look at the log-returns of the process I_t, which we define to be $Y_t := \mathscr{L}og(I_t)$. Here $\mathscr{L}og$ is the stochastic logarithm. We can formally write $dY_t = \frac{d(\mathbf{1}^\top \mathbf{S}_t)}{\mathbf{1}^\top \mathbf{S}_{t-}}$. So, if we now formally divide (4) by the total value $\mathbf{1}^\top \mathbf{S}_{t-}$ we get

$$dY_t = \frac{d(\mathbf{1}^\top \mathbf{S}_t)}{\mathbf{1}^\top \mathbf{S}_{t-}} = \mathbf{w}_t^\top \mathbf{a}_t\, dt + \mathbf{w}_t^\top \boldsymbol{\Sigma}_t^{\frac{1}{2}} d\mathbf{W}_t + \mathbf{w}_t^\top d\mathbf{J}_t + \int_{\mathbb{R}} \mathbf{w}_t^\top (e^x - 1 - x)\mathbf{M}(dt, dx). \tag{5}$$

The model described so far is too general to allow formal derivation of results. In order to simplify the setting we introduce the following assumption.

Assumption 8. The index weights are assumed to be constant over time, i.e., $\mathbf{w}_t \equiv \mathbf{w}$ over the time horizon we consider.

The previous assumption draws from the empirical study performed in [16], where it is stated that already when $D = 100$, like, for example, the S&P100, differences in price due to rebalancing of the weights are negligible. Alternatively, this assumption could also be justified if we used a different point of view when looking at the model (1): We could interpret the vector $\mathbf{X}$ as a vector of factors to which the index price is exposed, instead of a vector of logarithmic stock prices. Although not strictly equivalent, this interpretation would not affect the theoretical results developed in this work, and it would make the assumption of constant weights even more natural.

In conclusion, the formula that we will be using throughout the paper is the following:

$$dY_t = \mathbf{w}^\top \mathbf{a}_t\, dt + \mathbf{w}^\top \boldsymbol{\Sigma}_t^{\frac{1}{2}} d\mathbf{W}_t + \mathbf{w}^\top d\mathbf{J}_t + \int_{\mathbb{R}} \mathbf{w}^\top (e^x - 1 - x)\mathbf{M}(dt, dx). \tag{6}$$

2.3. Definition of the VRP. The quantity we aim to study is the *VRP* of the value weighted index Y. We denote by $[Y]_t^{t+h} := [Y]_{t+h} - [Y]_t$ the quadratic variation of the process Y accumulated over the time interval $[t, t+h]$, for $t \geq 0$, $h > 0$. $\mathrm{VRP}_{t,h}$ is then defined to be

$$\mathrm{VRP}_{t,h} := \frac{1}{h}\mathbb{E}^{\mathbb{P}}\left[[Y]_t^{t+h}\middle|\mathscr{F}_t\right] - \frac{1}{h}\mathbb{E}^{\mathbb{Q}}\left[[Y]_t^{t+h}\middle|\mathscr{F}_t\right] \tag{7}$$

for some equivalent change of measure $\mathbb{P} \to \mathbb{Q}$ and where $\mathscr{F}_t = \sigma\{\mathbf{X}_l, l \in [0,t]\}$ is the filtration generated by the process $\mathbf{X}$ (see [30]). In order to compute it, we need to be able to first find the quadratic variation of Y and then take its conditional expectation.

2.4. Quadratic variation of multivariate Brownian integrals with stochastic volatility and correlation. This section contains some necessary technical results that we will be using when dealing with the VRP of the stochastic volatility integrals.

The first step in computing the VRP consists of evaluating the diffusion term in the quadratic variation of Y as in (6): $[\int_0^\cdot \mathbf{w}^\top \boldsymbol{\Sigma}_s^{\frac{1}{2}}\, d\mathbf{W}_s]_t^{t+h}$. The following proposition gives us an elegant expression.

Proposition 9. *Let $\mathbf{B}$ be a D-standard Brownian motion and $\mathbf{G}$ a $D \times D$ (stochastic) matrix. The quadratic variation of the stochastic integral $\int_0^\cdot \mathbf{w}^\top \mathbf{G}_s\, d\mathbf{B}_s$ computed between times t and $t+h$ has the following expression:*

$$\left[\int_0^\cdot \mathbf{w}^\top \mathbf{G}_s\, d\mathbf{B}_s\right]_t^{t+h} = \mathbf{w}^\top \left(\int_t^{t+h} \mathbf{G}_s \mathbf{G}_s^\top\, ds\right)\mathbf{w}. \tag{8}$$

We will see that our proposed method to model the stochastic volatility consists of having a correlated Brownian motion (see section 3.2). Since this situation does not directly follow from the previous proposition, we state here the result that we will be using.

Corollary 10. *Let $\mathbf{W}$ be a Brownian motion with (stochastic) correlation matrix $\boldsymbol{\rho}_t$, and let $\boldsymbol{\Pi}_t$ such that $\boldsymbol{\Pi}_t\boldsymbol{\Pi}_t^\top = \boldsymbol{\rho}_t$, for all $t > 0$. The quadratic variation of the stochastic integral $\int_0^\cdot \mathbf{w}^\top \boldsymbol{\Sigma}_s^{\frac{1}{2}}\, d\mathbf{W}_s$ computed between times t and $t+h$ has the following expression:*

$$\left[\int_0^\cdot \mathbf{w}^\top \boldsymbol{\Sigma}_s^{\frac{1}{2}}\, d\mathbf{W}_s\right]_t^{t+h} = \mathbf{w}^\top \left(\int_t^{t+h} \boldsymbol{\Sigma}_s^{\frac{1}{2}} \boldsymbol{\rho}_s \boldsymbol{\Sigma}_s^{\frac{1}{2}}\, ds\right)\mathbf{w}. \tag{9}$$

The VRP is the wedge between the conditional expectations of (9) under $\mathbb{P}$ and $\mathbb{Q}$, normalized by h. When taking the conditional expectation with respect to $(\mathscr{F}_t)$, using conditional Fubini, one gets

$$\mathbb{E}\left[\left[\int_t^{t+h} \mathbf{w}^\top \boldsymbol{\Sigma}_s^{\frac{1}{2}}\, d\mathbf{W}_s\right] \Big| \mathscr{F}_t\right] = \mathbf{w}^\top \left(\int_t^{t+h} \mathbb{E}\left[\boldsymbol{\Sigma}_s^{\frac{1}{2}} \boldsymbol{\rho}_s \boldsymbol{\Sigma}_s^{\frac{1}{2}} \Big| \mathscr{F}_t\right] ds\right) \mathbf{w}. \tag{10}$$

Remark 11. Note that formula (10) *does not* depend on our choice of the model for $\boldsymbol{\Sigma}$ but will be true in any multivariate stochastic volatility model.

In order to stress that the stochastic volatility only provides a partial contribution to the total VRP, we will give its contribution a name.

Definition 12. *We define the* diffusive variance risk premium $DVRP_{t,h}$*, at time* t*, over the time span* $[t, t+h]$*, to be the difference*

$$DVRP_{t,h} = \frac{1}{h}\mathbb{E}^{\mathbb{P}}\left[\left[\int_t^{t+h} \mathbf{w}^\top \boldsymbol{\Sigma}_s^{\frac{1}{2}}\, d\mathbf{W}_s\right] \Big| \mathscr{F}_t\right] - \frac{1}{h}\mathbb{E}^{\mathbb{Q}}\left[\left[\int_t^{t+h} \mathbf{w}^\top \boldsymbol{\Sigma}_s^{\frac{1}{2}}\, d\mathbf{W}_s\right] \Big| \mathscr{F}_t\right].$$

We will be able to derive the left-hand side of (10) once we have an explicit expression for $\boldsymbol{\Sigma}$. In the following, we will study some of the most widely used stochastic volatility models and we will show that different model choices will lead to very different dynamics of the corresponding DVRP.

3. Brownian and stochastic volatility component.

3.1. Full multivariate modeling. We now introduce our multivariate models for the stochastic volatility matrix. It is to be stressed that, to the best of the authors' knowledge, this is the first time in the literature that the VRP is studied with a multivariate stochastic volatility model, all the previous studies focusing exclusively on univariate specifications for the equity index.

Since we now introduce correlation in $\boldsymbol{\Sigma}$, it is natural to assume that $\mathbf{W}$ is a standard Brownian motion, and hence, Proposition 9 gives the correct representation for the quadratic variation of the index return Y.

3.1.1. Multivariate non-Gaussian OU model. In this section we propose a full multivariate model for the matrix $\boldsymbol{\Sigma}$: the multivariate non-Gaussian OU model, as defined in [25].

Let $\mathbb{M}_D(K)$ be the set of D-dimensional matrices over the field K, $\mathbb{S}_D$ be the subalgebra of $\mathbb{M}_D(K)$ of symmetric real matrices, and $\mathbb{S}_D^+$ be the cone of all symmetric positive semidefinite matrices. Finally, for $\mathbf{A} \in \mathbb{M}_D(K)$, $\sigma(\mathbf{A})$ is the set of eigenvalues of $\mathbf{A}$.

Definition 13. *An* $\mathbb{S}_D^+$*-valued Lévy process* $\mathbf{L}$ *is called a matrix (Lvy) subordinator if* $\mathbf{L}_t - \mathbf{L}_s$ *belongs to* $\mathbb{S}_D^+$ *for all* $t > s > 0$.

Definition 14. *The non-Gaussian OU model is defined by*

$$d\boldsymbol{\Sigma}_t = (\mathbf{A}\boldsymbol{\Sigma}_t + \boldsymbol{\Sigma}_t\mathbf{A}^\top)\, dt + d\mathbf{L}_t, \tag{11}$$

where $\mathbf{L}$ *is a Lévy subordinator with Lévy measure* $\boldsymbol{\nu}$ *and* $\mathbf{A} \in \mathbb{M}_D(\mathbb{R})$ *such that* $0 \notin \sigma(\mathbf{A}) + \sigma(\mathbf{A})$ *(i.e., it is impossible to write* 0 *as the sum of two eigenvalues of* $\mathbf{A}$*).*

Recall formula (10). In this setting where the Brownian motion $\mathbf{W}$ is assumed to be standard, it reads

$$\mathbb{E}\left[\left[\int_t^{t+h} \mathbf{w}^\top \boldsymbol{\Sigma}_s^{\frac{1}{2}}\, d\mathbf{W}_s\right] \Big| \mathscr{F}_t\right] = \mathbf{w}^\top \left(\int_t^{t+h} \mathbb{E}\left[\boldsymbol{\Sigma}_s \Big| \mathscr{F}_t\right] ds\right) \mathbf{w}. \tag{12}$$

In order then to compute the VRP, we need an explicit expression for the conditional expectation of $\boldsymbol{\Sigma}$. The following proposition provides an explicit answer to the problem. Its proof can be found in the appendix.

Proposition 15. *The conditional expectation of the non-Gaussian multivariate OU process $\boldsymbol{\Sigma}$ admits the explicit representation:*

$$\mathbb{E}\left[\boldsymbol{\Sigma}_s \Big| \mathscr{F}_t\right] = e^{\mathbf{A}(s-t)} \boldsymbol{\Sigma}_t\, e^{\mathbf{A}^\top (s-t)} + \int_t^s e^{\mathbf{A}(s-l)} \cdot \mathbf{k} \cdot e^{\mathbf{A}^\top (s-l)}\, dl,$$

where $e^{\mathbf{A}}$ denotes the matrix exponential and $\mathbf{k}$ is the matrix of the first moments of $\mathbf{L}$, i.e., $k^{(ij)} := \mathbb{E}[L^{(ij)}]$.

Remark 16. With exactly the same reasoning and proof, one can prove that the explicit (strong) solution of the SDE defining the process $\boldsymbol{\Sigma}$, $d\boldsymbol{\Sigma}_t = (\mathbf{A}\boldsymbol{\Sigma}_t + \boldsymbol{\Sigma}_t \mathbf{A}^\top)\, dt + d\mathbf{L}_t$, is given by

$$\boldsymbol{\Sigma}_t = e^{\mathbf{A}t} \boldsymbol{\Sigma}_0\, e^{\mathbf{A}^\top t} + \int_0^t e^{\mathbf{A}(t-l)}\, d\mathbf{L}_l\, e^{\mathbf{A}^\top (t-l)}.$$

Proposition 15 provides the key step in computing the risk premium.

Now, looking back at (12), and noting that the only quantity that changes between $\mathbb{P}$ and $\mathbb{Q}$ is the moment matrix of $\mathbf{L}$, we obtain the final formula:

$$\mathrm{DVRP}_{t,h} = \frac{1}{h}\left(\mathbf{w}^\top \left\{\int_t^{t+h} \left[\int_t^s e^{\mathbf{A}(s-l)} \cdot \left(\mathbf{k}^{\mathbb{P}} - \mathbf{k}^{\mathbb{Q}}\right) \cdot e^{\mathbf{A}^\top (s-l)}\, dl\right] ds\right\} \mathbf{w}\right). \tag{13}$$

We immediately observe that the structure we imposed on $\boldsymbol{\Sigma}$ prevents the VRP from exhibiting stochastic dynamics. This is a surprising result in that this model could have been expected to possess richer dynamics, since the stochastic volatility matrix accounts for dependence between its components. Nevertheless, taking the conditional expectations of such a fully specified model led us to solve a multidimensional ODE with deterministic drivers, and all the unpredictability was lost.

Finally, we observe that as a consequence of the time integrals present in the expression (13) it holds that the diffusion risk premium, at any time t, decays to zero if the time span considered shrinks to zero:

$$\lim_{h \downarrow 0} \mathrm{DVRP}_{t,h} = 0.$$

Hence, over small time intervals, the DVRP becomes negligible. This behavior is not shared with the Hawkes process, as we will see after Corollary 40.

3.1.2. Wishart model. Another popular model for the stochastic volatility matrix is the Wishart model. We follow the definition in [14].

Definition 17. *Let* $\mathbf{\Omega}, \mathbf{M}, \mathbf{Q} \in M_D(\mathbb{R})$, *with* $\det(\mathbf{Q}) \neq 0$, *and* $\mathbf{B}$ *a* D*-dimensional square Brownian motion matrix, whose* $D \times D$ *components are standard Brownian motions independent of the process* $\mathbf{X}$. *The Wishart model is defined to be*

$$d\mathbf{\Sigma}_t = \left(\mathbf{\Omega}\mathbf{\Omega}^\top + \mathbf{M}\mathbf{\Sigma}_t + \mathbf{\Sigma}_t\mathbf{M}^\top\right) dt + \mathbf{\Sigma}_t^{\frac{1}{2}}\, d\mathbf{B}_t\, \mathbf{Q} + \mathbf{Q}^\top (d\mathbf{B}_t)^\top \mathbf{\Sigma}_t^{\frac{1}{2}}. \tag{14}$$

Recall that, thanks to formula (12), we need an expression for the conditional expectation of $\mathbf{\Sigma}$.

Lemma 18. *If* $\mathbf{\Sigma}$ *is defined as in* (14), *then we have, for almost all* $s \geq t$,

$$\mathbb{E}\left[\mathbf{\Sigma}_s \middle| \mathscr{F}_t\right] = e^{\mathbf{M}(s-t)}\mathbf{\Sigma}_t\, e^{\mathbf{M}^\top(s-t)} + \int_t^s e^{\mathbf{M}(s-l)}\mathbf{\Omega}\mathbf{\Omega}^\top e^{\mathbf{M}^\top(s-l)}\, dl. \tag{15}$$

As a consequence of the previous lemma, we have

$$\begin{aligned} &\frac{1}{h}\mathbf{w}^\top \left(\int_t^{t+h} \mathbb{E}\left[\mathbf{\Sigma}_s \middle| \mathscr{F}_t\right] ds\right) \mathbf{w} \\ &\quad = \frac{1}{h}\left(\mathbf{w}^\top \left\{\int_t^{t+h} \left[e^{\mathbf{M}(s-t)}\mathbf{\Sigma}_t\, e^{\mathbf{M}^\top(s-t)} + \int_t^s e^{\mathbf{M}(s-l)}\mathbf{\Omega}\mathbf{\Omega}^\top e^{\mathbf{M}^\top(s-l)}\, dl\right] ds\right\}\mathbf{w}\right). \end{aligned} \tag{16}$$

Now we need to perform an equivalent, structure preserving change of measure from $\mathbb{P}$ to $\mathbb{Q}$. By the converse of Girsanov's theorem, we have $d\mathbf{B}_t = d\mathbf{B}_t^{\mathbb{Q}} + \mathbf{K}_t\, dt$ for some $\mathbf{K}_t$, where $\mathbf{B}^{\mathbb{Q}}$ is a $\mathbb{Q}$-Brownian motion.

If we pick $\mathbf{K}_t = \mathbf{\Sigma}_t^{\frac{1}{2}}\mathbf{H}$, where $\mathbf{H} \in M_D(\mathbb{R})$, then the dynamics (14) will be transformed into

$$d\mathbf{\Sigma}_t = \left(\mathbf{\Omega}\mathbf{\Omega}^\top + \widetilde{\mathbf{M}}\mathbf{\Sigma}_t + \mathbf{\Sigma}_t\widetilde{\mathbf{M}}^\top\right) dt + \mathbf{\Sigma}_t^{\frac{1}{2}}\, d\mathbf{B}_t^{\mathbb{Q}}\mathbf{Q} + \mathbf{Q}^\top (d\mathbf{B}_t^{\mathbb{Q}})^\top\, \mathbf{\Sigma}_t^{\frac{1}{2}},$$

where $\widetilde{\mathbf{M}} := \mathbf{M} + \mathbf{Q}^\top\mathbf{H}^\top$.

Instead, choosing $\mathbf{K}_t = \mathbf{\Sigma}_t^{-\frac{1}{2}}\mathbf{H}$ yields the following expression:

$$d\mathbf{\Sigma}_t = \left(\widetilde{\mathbf{\Omega}}\widetilde{\mathbf{\Omega}}^\top + \mathbf{M}\mathbf{\Sigma}_t + \mathbf{\Sigma}_t\mathbf{M}^\top\right) dt + \mathbf{\Sigma}_t^{\frac{1}{2}}\, d\mathbf{B}_t^{\mathbb{Q}}\, \mathbf{Q} + \mathbf{Q}^\top (d\mathbf{B}_t^{\mathbb{Q}})^\top \mathbf{\Sigma}_t^{\frac{1}{2}},$$

where $\widetilde{\mathbf{\Omega}}\widetilde{\mathbf{\Omega}}^\top = \mathbf{\Omega}\mathbf{\Omega}^\top + \mathbf{H}\mathbf{Q} + \mathbf{Q}^\top\mathbf{H}^\top$. The crucial difference between the two choices is whether the matrix $\mathbf{M}$ changes under the two measures. If it does, then the DVRP obtained will be stochastic, as a consequence of formula (15). An explicit formula for the DVRP in this situation follows by taking differences of (16) under the two measures, as the next example shows in a numerical example.

Example 19 (two-dimensional Wishart). In this example we will explicitly give a numerical example of the DVRP for the simple case of a Wishart model with two stocks. We specify the model in (14) with the following choices:

$$\mathbf{M} = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}, \qquad \mathbf{Q} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \qquad \mathbf{\Omega}\mathbf{\Omega}^\top = \begin{pmatrix} 1 & 1 \\ 1 & 3 \end{pmatrix}.$$

As a possible choice of measure change we use $\mathbf{K}_t = \mathbf{\Sigma}_t^{\frac{1}{2}}\mathbf{H}$, with $\mathbf{H} = \left(\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix} \right)$, whence $\widetilde{\mathbf{M}} = \left(\begin{smallmatrix} 0 & 2 \\ 1 & 1 \end{smallmatrix} \right)$. The (i,j)-element of $\mathbf{\Sigma}$ will be denoted by $\sigma_{i,j} := \sigma_{i,j;t}$ and, for ease of notation, we will not stress its dependence on time.

The quantities appearing in (16) evaluate to

$$e^{\mathbf{M}(s-t)}\mathbf{\Sigma}_t e^{\mathbf{M}^\top(s-t)} = \begin{pmatrix} e^{2s-2t}(\sigma_{1,1}+\sigma_{2,1}+\sigma_{1,2}+\sigma_{2,2}) & e^{s-t}(\sigma_{1,1}+\sigma_{2,1}) + e^{3s-3t}(\sigma_{1,2}+\sigma_{2,2}) \\ (\sigma_{1,1}+\sigma_{1,2})e^{s-t} + (\sigma_{2,1}+\sigma_{2,2})e^{3s-3t} & \sigma_{1,1} + (\sigma_{1,2}+\sigma_{2,1})e^{2s-2t} + \sigma_{2,2}e^{4s-4t} \end{pmatrix}$$

and

$$\int_t^s e^{\mathbf{M}(s-l)}\mathbf{\Omega}\mathbf{\Omega}^\top e^{\mathbf{M}^\top(s-l)}\,dl = \begin{pmatrix} 3\left(e^{2(s-t)}-1\right) & 2\left(e^{s-t}-1\right) + \frac{4}{3}\left(e^{3(s-t)}-1\right) \\ 2\left(e^{s-t}-1\right) + \frac{4}{3}\left(e^{3(s-t)}-1\right) & (s-t) + \left(e^{2(s-t)}-1\right) + \frac{3}{4}\left(e^{4(s-t)}-1\right) \end{pmatrix}.$$

Under $\mathbb{Q}$ we need to do similar computations using our new matrix of coefficients $\widetilde{\mathbf{M}}$. Integrating our results from t to $t+h$, multiplying by the weights, and dividing by h, we obtain the final formula:

$$\begin{aligned} DVRP_{t,h} = (w_1)^2 \Bigg[&(\sigma_{1,1}+\sigma_{2,1}+\sigma_{1,2}+\sigma_{2,2})\frac{1}{2h}\left(e^{2h}-1\right) + \frac{1}{h}\left(e^{2h}-1-2h\right) - \sigma_{1,1} \\ &- \frac{\sigma_{2,1}+\sigma_{1,2}}{2h}\left(e^{2h}-1\right) - \frac{\sigma_{2,2}}{4h}(e^{4h}-1) - \frac{h}{2} - \frac{3}{16h}\left(e^{4h}-4h-1\right)\Bigg] \\ &+ (w_2)^2\Bigg[\sigma_{1,1} + \frac{\sigma_{2,1}+\sigma_{1,2}}{2h}\left(e^{2h}-1\right) + \frac{\sigma_{2,2}}{4h}(e^{4h}-1) - \frac{1}{h}\left(e^{2h}-2h-1\right) \\ &+ \frac{3}{16h}\left(e^{4h}-4h-1\right) + \frac{h}{2} - (\sigma_{1,1}+\sigma_{2,1}+\sigma_{1,2}+\sigma_{2,2})\frac{1}{2h}\left(e^{2h}-1\right)\Bigg]. \end{aligned}$$

The formula makes it clear that with this choice of model, it is not possible to disentangle the effects of single stocks from that of correlated stocks, as we observe the appearance of different components of the matrix $\mathbf{\Sigma}$ in all terms of the weighted sum.

Example 20 (the Heston model). The Wishart model reduces to the Heston model if $D=1$. We recall the definition of the Heston model:

$$\begin{cases} dS_t = \mu S_t\,dt + \sqrt{v_t}S_t\,dW_t^S, \\ dv_t = k(\vartheta - v_t)\,dt + \zeta\sqrt{v_t}\,dW_t^v. \end{cases}$$

The former is equivalent to $dX_t = d\left(\log(S_t)\right) = (\mu - v_t)\,dt + \sqrt{v_t}\,dW_t^S$; thus we see that we can obtain the Heston specification by putting $\beta\colon \mathbb{R}\to\mathbb{R}; \beta(x) := x, J\equiv 0$ in our model in (1).

In this one-dimensional case, our index I consists only of the stock S. The return process of the index is $Y_t = \mu t + \int_0^t \sqrt{v_u}\,dW_u^S$, from which it immediately follows that

$$\mathbb{E}\left[[Y]_t^{t+h}\middle|\mathscr{F}_t\right] = \int_t^{t+h}\mathbb{E}^{\mathbb{P}}\left[v_u\middle|\mathscr{F}_t\right]\,du.$$

From the representation $v_u = k\int_0^u(\vartheta - v_s)\,ds + \zeta\int_0^u\sqrt{v_s}\,dW_s^v$, we obtain

$$\mathbb{E}\left[v_u\middle|\mathscr{F}_t\right] = v_t + k\int_t^u\left(\vartheta - \mathbb{E}\left[v_s\middle|\mathscr{F}_t\right]\right)\,ds = v_t + k\vartheta(u-t) - k\int_t^u\mathbb{E}\left[v_s\middle|\mathscr{F}_t\right]\,ds.$$

The unique solution is: $\mathbb{E}[v_u|\mathscr{F}_t] = v_t e^{-k(u-t)} + \vartheta\left(1 - e^{-k(u-t)}\right)$, and thus

$$\mathbb{E}\left[[Y]_t^{t+h}\middle|\mathscr{F}_t\right] = \frac{1}{k}(\vartheta - v_t)\left(e^{-kh} - 1\right) + \vartheta h. \tag{17}$$

Now we must find an equivalent martingale measure $\mathbb{Q}$. We will restrict ourselves to the class of structure preserving measures $\mathbb{Q}$. By the converse of Girsanov's theorem, one can write $dW_t^v = d\tilde{W}_t^v + K_t\,dt$, where $\tilde{W}$ is Brownian motion under $\mathbb{Q}$, and hence

$$dv_t = k(\vartheta - v_t)dt + \zeta\sqrt{v_t}\left(d\tilde{W}_t^v + K_t\,dt\right) = \left(k(\vartheta - v_t) + \zeta K_t\sqrt{v_t}\right)dt + \zeta\sqrt{v_t}d\tilde{W}_t^v.$$

Due to the particular structure of the Heston model, we can specify the structure preserving Girsanov tranformation in more than one way.

The most common choice in the literature is to ask

$$K_t = \sqrt{v_t}H \tag{18}$$

for some constant H, obtaining

$$\begin{aligned} dv_t &= \left(k(\vartheta - v_t) + \zeta v_t H\right)\,dt + \zeta\sqrt{v_t}d\tilde{W}_t^v = \left(v_t(\zeta H - k) + k\vartheta\right)\,dt + \zeta\sqrt{v_t}d\tilde{W}_t^v \\ &= (k - \zeta H)\left(\frac{k\vartheta}{k - \zeta H} - v_t\right)\,dt + \zeta\sqrt{v_t}d\tilde{W}_t^v = \tilde{k}(\tilde{\vartheta} - v_t) + \zeta\sqrt{v_t}d\tilde{W}_t^v. \end{aligned}$$

In this situation, the DVRP takes the form

$$\mathrm{DVRP}_{t,h} = h(\vartheta - \tilde{\vartheta}) + \frac{\vartheta}{k}\left(e^{-kh} - 1\right) - \frac{\tilde{\vartheta}}{\tilde{k}}\left(e^{-\tilde{k}h} - 1\right) + \left[\frac{(1 - e^{-kh})}{k} - \frac{\left(1 - e^{-\tilde{k}h}\right)}{\tilde{k}}\right]v_t.$$

Remark 21. Much of the literature regarding the VRP has focused on the one-dimensional case. The question whether it is possible to obtain stochastic dynamics if $D = 1$ has attracted much attention. Although the Heston model gives an interesting result in itself, we want to remark that the focus of the paper is fully multivariate, and we are mainly concerned with how the dependence between the components generates a stochastic VRP.

3.2. Using a diagonal matrix of non-Gaussian OU processes. The previous two sections presented results for the VRP using popular multivariate stochastic volatility models. Recall that the study performed in [16] shows that individual stocks do not create VRP, while pairs of correlated stocks do. A fully multivariate stochastic volatility model does not allow us to disentangle the two contributions. We now present a different model to solve this issue.

In this section we will model the stochastic matrix $\mathbf{\Sigma}$ via a *diagonal matrix* of non-Gaussian OU processes (as in [4]) and introduce correlation in the Brownian motion $\mathbf{W}$.

In detail, we define the *variance matrix* $\mathbf{\Sigma}$ to be: $\mathbf{\Sigma}_t = \mathrm{diag}(\sigma_t^{2(1)}, \dots \sigma_t^{2(D)})$. In order to define each component of $\mathbf{\Sigma}$, consider a vector $\mathbf{\Lambda} \in \mathbb{R}^D$ with all components strictly positive. Consider a multivariate Lévy process $\mathbf{L} := (L^{(1)}_{\lambda^{(1)}t}, \dots, L^{(i)}_{\lambda^{(i)}t}, \dots, L^{(D)}_{\lambda^{(D)}t})^\top$, with Lévy measure $\boldsymbol{\nu}$ and with all components being subordinators with zero drift. Each element $\sigma_t^{2(i)}$ follows the Lévy driven SDE:

$$d\sigma_t^{2(i)} = -\lambda^{(i)}\sigma_t^{2(i)}\,dt + dL^{(i)}_{\lambda^{(i)}t}. \tag{19}$$

In this context, we will call $\mathbf{L}$ a BDLP. Note that we are using a particular form of the multivariate non-Gaussian OU process studied in section 3.1.1, when $\mathbf{A}$ is a diagonal matrix. Also observe that different components of $\mathbf{\Sigma}$ need not be independent, since their BDLPs come from a multivariate Lévy process.

We furthermore assume that the multivariate Brownian motion $\mathbf{W}$ is correlated, with two components $W^{(i)}$ and $W^{(j)}$ satisfying

$$[W^{(i)}, W^{(j)}]_t = \int_0^t \rho_s^{(ij)}\,ds. \tag{20}$$

That is the same as writing $\mathbf{W}_t = \mathbf{\Pi}_t\mathbf{B}_t$, where $\mathbf{B}_t$ is a multivariate Brownian motion with independent components and the matrix $\mathbf{\Pi}_t$ can be given by any decomposition of the correlation matrix $\boldsymbol{\rho}_t$: $\mathbf{\Pi}_t^\top\mathbf{\Pi}_t = \mathbf{\Pi}_t\mathbf{\Pi}_t^\top = \boldsymbol{\rho}_t$, where $\rho_t^{(ii)} = 1$, for all i. We can take, for example, the Choleski decomposition of $\boldsymbol{\rho}_t$ in which $\mathbf{\Pi}_t$ is lower triangular.

Assumption 22. We furthermore assume that $\rho_t^{(ij)}$ is a *deterministic* function of time. Although this seems to be a restrictive assumption, our stock price model embeds other forms of stochastic dependence between the stock prices, like the general Lévy measure $\boldsymbol{\nu}$, as well as the multivariate Hawkes process, as we shall see later. Note further that the assumption of having a deterministic, sometimes even constant correlation, but having stochastic volatility is a common assumption in multivariate models (see, e.g., the influential work in [8] on constant conditional correlation models in the time series literature), since it significantly simplifies inference in multivariate models.

We now need to find the quadratic variation of the diffusion component of the index return Y, when the volatility structure is defined as above. To this end, if we specialize the result in

Corollary 10 for our choice of $\boldsymbol{\Sigma}^{\frac{1}{2}}$, we obtain

$$\begin{aligned}\left[\int_0^{\cdot} \mathbf{w}^\top \boldsymbol{\Sigma}_s^{\frac{1}{2}}\, d\mathbf{B}_s\right]_t^{t+h} &= \mathbf{w}^\top \left(\int_t^{t+h} \boldsymbol{\Sigma}_s^{\frac{1}{2}} \boldsymbol{\rho}_s \boldsymbol{\Sigma}_s^{\frac{1}{2}}\, ds\right) \mathbf{w} \\ &= \sum_{i=1}^{D} (w^{(i)})^2 \int_0^t (\sigma_s^{(i)})^2\, ds + \sum_{i=1}^{D}\sum_{i\neq j} w^{(i)} w^{(j)} \int_0^t \sigma_s^{(i)} \sigma_s^{(j)} \rho_s^{(ij)}\, ds. \end{aligned} \tag{21}$$

Formula (21) is identical to (2) in [16]. For this choice of model for $\boldsymbol{\Sigma}$, we can decompose the risk premium into the contribution from the single stocks and the contribution from two correlated stocks.

3.2.1. Contribution of the single stocks.

We begin with the analysis of the contribution of the single stocks:

$$\sum_{i=1}^{D} (w^{(i)})^2 \int_0^t (\sigma_s^{(i)})^2\, ds.$$

For our proposed constant correlation model, we will decompose the diffusion risk premium into two components: the DVRP for individual stocks, denoted by IDVRP, and the DVRP from pairs of correlated stocks, called CDVRP. We give here the definition of the former. For the latter, see Definition 26.

Definition 23. *We call* DVRP for individual stocks *at time t, for the ith stock in the index and over the time span $[t, t+h]$, denoted by $IDVRP_{t,h}^{(i)}$, the process defined for $t \geq 0$:*

$$IDVRP_{t,h}^{(i)} = \frac{1}{h}\mathbb{E}\left[\left[\int_0^{\cdot} \sigma_l^{(i)}\, dW_l^{(i)}\right]_t^{t+h} \Big| \mathscr{F}_t\right] - \frac{1}{h}\mathbb{E}^{\mathbb{Q}}\left[\left[\int_0^{\cdot} \sigma_l^{(i)}\, dW_l^{(i)}\right]_t^{t+h} \Big| \mathscr{F}_t\right].$$

The IDVRP represents the component of the total risk premium which is attributable to each single stock, only through the quadratic variation of its stochastic volatility.

An application of conditional Fubini to (21) shows that we can derive the IDVRP once we obtain an expression for $\mathbb{E}[\big(\sigma^{(i)}\big)_s^2 | \mathscr{F}_t]$ for $s \geq t$.

3.2.2. Diffusion variance risk premium.

The power of the non-Gaussian OU model assumption for σ^2 is manifested in the following proposition.

Proposition 24. *Let $u \geq t$. Then the conditional expectation of $\sigma_u^{2(i)}$ with respect to $(\mathscr{F}_t)$ is given by*

$$\mathbb{E}\left[\sigma_u^{2(i)} \Big| \mathscr{F}_t\right] = k_1^{(i)} + (\sigma_t^{2(i)} - k_1^{(i)}) e^{-\lambda^{(i)}(u-t)}, \tag{22}$$

where $k_1^{(i)}$ is defined as $\mathbb{E}[L_1^{(i)}] = k_1^{(i)} = \int_{\mathbb{R}} x\, \nu^{(i)}(dx)$, and $\nu^{(i)}$ is the Lévy measure of the process $L^{(i)}$.

The previous result gives us immediately upon integration an expression for the DVRP.

Proposition 25. *The DVRP originated by each stock* $S^{(i)}$ *is given by*

$$IDVRP^{(i)}_{t,h} = \left(k_1^{\mathbb{P}(i)} - k_1^{\mathbb{Q}(i)}\right)\left(1 + \frac{e^{-\lambda^{(i)}h} - 1}{\lambda^{(i)}h}\right). \tag{23}$$

Let us make a few comments on what formula (23) tells us. We immediately observe that, in line with the empirical findings in [16] and [7], the DVRP of single stocks is deterministic. Since the function $f(x) = 1 + \frac{e^{-x}-1}{x}$ is positive for $x > 0$, the sign of each stock's DVRP is given by the sign of the difference $(k_1^{\mathbb{P}(i)} - k_1^{\mathbb{Q}(i)})$ of the first moments of the BDLP between $\mathbb{P}$ and $\mathbb{Q}$.

3.2.3. Correlation risk premium. We will now consider the contribution to the VRP originated by the correlation between the stocks. Looking back at formula (21), we look now at the contribution of the quadratic covariation between two stocks in the index Y:

$$\sum_{i=1}^{D}\sum_{i\neq j} w^{(i)}w^{(j)}\int_0^t \sigma_s^{(i)}\sigma_s^{(j)}\rho_s^{(ij)}\,ds.$$

We now give the definition of the correlated VRP, similarly to Definition 23.

Definition 26. *We call* correlated (diffusive) VRP *for the stocks* $(X^{(i)}, X^{(j)})$ *at time* t*, over the time span* $[t, t+h]$*, the process defined for* $t \geq 0$*:*

$$CDVRP^{(i,j)}_{t,h} := \frac{1}{h}\mathbb{E}^{\mathbb{P}}\left[\int_t^{t+h}\rho_s^{(i,j)}\sigma_s^{(i)}\sigma_s^{(j)}\,ds\middle|\mathscr{F}_t\right] - \frac{1}{h}\mathbb{E}^{\mathbb{Q}}\left[\int_t^{t+h}\rho_s^{(i,j)}\sigma_s^{(i)}\sigma_s^{(j)}\,ds\middle|\mathscr{F}_t\right].$$

Another application of conditional Fubini shows that the quantity of interest in order to compute the CDVRP is

$$\mathbb{E}\left[\rho_s^{(ij)}\sigma_s^{(i)}\sigma_s^{(j)}\middle|\mathscr{F}_t\right].$$

Assumption (22) that $\rho_t^{(ij)}$ is deterministic will be fundamental. Indeed, it allows us to write

$$\mathbb{E}^{\mathbb{P}}\left[\rho_s^{(ij)}\sigma_s^{(i)}\sigma_s^{(j)}\middle|\mathscr{F}_t\right] = \rho_s^{(ij)}\mathbb{E}^{\mathbb{P}}\left[\sigma_s^{(i)}\sigma_s^{(j)}\middle|\mathscr{F}_t\right].$$

The DVRP was greatly simplified thanks to the choice of directly modeling the square volatilities σ^2. The correlation risk premium appears more complicated in this setting. The natural way to proceed would be to compute the dynamics of $\sigma^{2(i)}\sigma^{2(j)}$ via Itô's formula. Unfortunately, this would lead to the introduction of square roots, and explicit evaluation of conditional expectations would become infeasible.

We propose a solution to this technical problem. We can employ an integral representation of the square root and subsequently interchange the integral sign with the conditional expectation operator. What we use is the following formula, which is proved, for example, in [2, pp. 80–81]:

$$\sqrt{x} = \frac{1}{2}\frac{1}{\Gamma\left(\frac{1}{2}\right)}\int_0^\infty \frac{(1-e^{-xy})}{y^{\frac{3}{2}}}\,dy. \tag{24}$$

Applying the formula twice, with $x = \sigma_u^{2(i)}, y = \sigma_u^{2(j)}$, and recalling that $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, one gets

$$\begin{aligned}\sigma_u^{(i)}\sigma_u^{(j)} &= \frac{1}{4\pi}\left(\int_0^\infty \frac{1-e^{-\sigma_u^{2(i)}y}}{y^{\frac{3}{2}}}\,dy\right)\left(\int_0^\infty \frac{1-e^{-\sigma_u^{2(j)}x}}{x^{\frac{3}{2}}}\,dx\right)\\ &= \frac{1}{4\pi}\int_0^\infty dy\left[\int_0^\infty \frac{\left(1-e^{-\sigma_u^{2(i)}y}\right)}{y^{\frac{3}{2}}}\frac{\left(1-e^{-\sigma_u^{2(j)}x}\right)}{x^{\frac{3}{2}}}\,dx\right] \qquad (25)\\ &= \frac{1}{4\pi}\int_0^\infty dy\left[\int_0^\infty \frac{1}{(xy)^{\frac{3}{2}}}\left(1-e^{-\sigma_u^{2(i)}y}-e^{-\sigma_u^{2(j)}x}+e^{-\sigma_u^{2(i)}y-\sigma_u^{2(j)}x}\right)dx\right].\end{aligned}$$

The equality on line (25) follows from Tonelli's theorem, since the integrands are positive, seen as functions of x and y. The conditional expectation of the product $\sigma_u^{(i)}\sigma_u^{(j)}$ now reduces to the computation of several Laplace transforms of the processes involved.

We can summarize with the following.

Theorem 27. *The correlation risk premium at time t for the stocks $(X^{(i)}, X^{(j)})$, over the time span $[t, t+h]$, has the following expression:*

$$\text{(26)}\qquad CDVRP_{t,h}^{(i,j)} = \frac{1}{4h\pi}\int_t^{t+h}\rho_s^{(i,j)}\left(\int_0^\infty dy\int_0^\infty \frac{1}{(xy)^{\frac{3}{2}}}H\left(\sigma_t^{2(i)}, \sigma_t^{2(j)}, x, y, t, s\right)dx\right)ds.$$

An explicit expression for the deterministic function H appears in (41) in the appendix.

The integral representation allows us to observe a few notable facts.

The $\text{CDVRP}^{(i,j)}$ is random, since two of the arguments of H are the random variables $\sigma_t^{2(i)}, \sigma_t^{2(j)}$. It is not deterministic precisely because the Laplace exponents of the components of the Lévy process $\mathbf{L}$ are different between $\mathbb{P}$ and $\mathbb{Q}$, as an analysis of formula (41) reveals. As a consequence, the correlation risk premium between two stocks exhibits stochastic dynamics. Moreover, if s tends to t, then H tends to zero almost surely in (26). Hence we again obtain that the correlation risk premium gives no contribution to the total VRP if the time span considered shrinks to zero: that is,

$$\lim_{h\downarrow 0}\text{CDVRP}_{t,h}^{(i,j)} = 0.$$

We conclude that this choice of modeling the stochastic volatility via a constant correlation model gives us the best results. In accordance with the empirical studies performed in the literature (see, for example, [16]), it implies deterministic dynamics for all the individual risk premia and stochastic dynamics for the correlation risk premia. In addition, it appears as much simpler than compared to the previous cases. Here, in order to correctly specify the model, we just need a D-dimensional Lévy process instead of a $D \times D$-dimensional Lévy matrix, a D-dimensional vector of parameters $\boldsymbol{\lambda}$, and the deterministic correlation matrix $\boldsymbol{\rho}$ of the Brownian motion.

Remark 28. We observe that our model is not the only choice one can make to obtain stochastic dynamics for the correlation risk premium. For example, choosing a diagonal matrix $\mathbf{A}$ in definition (11) leads to obtain diagonal elements for $\mathbf{\Sigma}$ which are univariate non-Gaussian OU processes (see [25]). Hence, combining such a model with a correlated Brownian motion would still produce deterministic IDVRP and stochastic CDVRP. Nonetheless, our model is still analytically tractable and allows for explicit description of these quantities only up to a deterministic double integral, which can be approximated by standard techniques.

The next example will illustrate our results in a situation where Gamma subordinators are used as drivers of the stochastic volatility.

Example 29 (OU-$\Gamma(\nu, r)$). As an example, we present an explicit calculation using independent Gamma processes as BDLPs. Let X be a Gamma process with parameters (ν, r). Then its characteristic triplet is

$$\left(\frac{r}{\nu}\left(1-e^{-\nu}\right),\, 0,\, \frac{r}{x}e^{-\nu x}1_{\{x>0\}}\, dx\right).$$

Since $r\int_1^\infty \log(x)\frac{e^{-\nu x}}{x} < \infty$, there exists a non-Gaussian OU process whose BDLP is a Gamma process.

Let us now consider the same SDEs as in (19), with the parameters $\lambda^{(i)}, \lambda^{(j)}$ set to 1:

$$\begin{cases} d\sigma_t^{2(i)} = -\sigma_t^{2(i)}\, dt + dL_t^{(i)}, \\ d\sigma_t^{2(j)} = -\sigma_t^{2(j)}\, dt + dL_t^{(j)}, \end{cases}$$

where $L^{(i)}$ and $L^{(j)}$ are Gamma *independent* processes with parameters $(\xi^{(i)}, r^{(i)})$ and $(\xi^{(j)}, r^{(j)})$, respectively. We will also need the following useful result: If $Y_u^{(k)}$ denotes the Lévy–Itô integral,

$$Y_u^{(k)} := \int_t^u e^{(s-u)}\, dL_s^{(k)},$$

then its Laplace transform can be explicitly computed via the following formula:

$$\mathbb{E}^{\mathbb{P}}\left[e^{-yY_u^{(k)}}\right] = e^{\int_t^u \varphi_{L^{(k)}}\left(ye^{-(u-s)}\right) ds}. \tag{27}$$

Let now $k \in \{i, j\}$. The Laplace exponent of $L^{(k)}$ is

$$\varphi_{L^{(k)}}(y) = r^{(k)} \log\left(\frac{\xi^{(k)}}{\xi^{(k)} + y}\right),$$

hence the Laplace exponent of the process $Y_u^{(k)} = \int_t^u e^{(s-u)}\, dL_s^{(k)}$ follows from formula (27):

$$\mathbb{E}\left[e^{-yY_u^{(k)}}\right] = e^{r^{(k)}\int_t^u \log\left(\frac{\xi^{(k)}}{\xi^{(k)} + ye^{(s-u)}}\right) ds}.$$

The independence assumption gives us an explicit expression for the joint Laplace transform:

$$\mathbb{E}\left[e^{-yY_u^{(i)}-xY_u^{(j)}}\right] = e^{r^{(i)}\int_t^u \log\left(\frac{\xi^{(i)}}{\xi^{(i)}+ye^{(s-u)}}\right)ds + r^{(j)}\int_t^u \log\left(\frac{\xi^{(j)}}{\xi^{(j)}+xe^{(s-u)}}\right)ds}$$
$$= e^{\int_t^u \log\left[\left(\frac{\xi^{(i)}}{\xi^{(i)}+ye^{(s-u)}}\right)^{r^{(i)}}\left(\frac{\xi^{(j)}}{\xi^{(j)}+xe^{(s-u)}}\right)^{r^{(j)}}\right]ds}.$$

Suppose now that we perform an Esscher change of measure, via a density given by

$$\left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_t = \frac{e^{\vartheta(L_t^{(i)}+L_t^{(j)})}}{\mathbb{E}\left[e^{\vartheta(L_t^{(i)}+L_t^{(j)})}\right]},$$

for some $\vartheta < \min(\xi^{(i)}, \xi^{(j)})$. Under $\mathbb{Q}$, the Lévy measure of the Gamma process $L^{(k)}$ becomes $\frac{r^{(k)}}{x}e^{-(\xi^{(k)}-\vartheta)}\,dx$. It follows that, under $\mathbb{Q}$,

$$\mathbb{E}^{\mathbb{Q}}\left[e^{-yY_u^{(k)}}\right] = e^{r^{(k)}\int_t^u \log\left(\frac{\xi^{(k)}-\vartheta}{\xi^{(k)}-\vartheta+ye^{(s-u)}}\right)ds}.$$

An explicit expression for $H(\sigma_t^{2(i)}, \sigma_t^{2(j)}, x, y, t, u)$ is readily found to be given by

$$\begin{aligned}
&H(\sigma_t^{2(i)}, \sigma_t^{2(j)}, x, y, t, u)\\
&= -\exp\left(-ye^{-(u-t)}\sigma_t^{2(i)}\right)\left[e^{r^{(i)}\int_t^u \log\left(\frac{\xi^{(i)}}{\xi^{(i)}+ye^{(s-u)}}\right)ds} - e^{r^{(i)}\int_t^u \log\left(\frac{\xi^{(i)}-\vartheta}{\xi^{(i)}-\vartheta+ye^{(s-u)}}\right)ds}\right]\\
&\quad -\exp\left(-xe^{-(u-t)}\sigma_t^{2(j)}\right)\left[e^{r^{(j)}\int_t^u \log\left(\frac{\xi^{(j)}}{\xi^{(j)}+xe^{(s-u)}}\right)ds} - e^{r^{(j)}\int_t^u \log\left(\frac{\xi^{(j)}-\vartheta}{\xi^{(j)}-\vartheta+xe^{(s-u)}}\right)ds}\right]\\
&\quad +\exp\left(-ye^{-(u-t)}\sigma_t^{2(i)} - xe^{-(u-t)}\sigma_t^{2(j)}\right)\\
&\quad \times\left[e^{\int_t^u\left[r^{(i)}\log\left(\frac{\xi^{(i)}}{\xi^{(i)}+ye^{(s-u)}}\right)+r^{(j)}\log\left(\frac{\xi^{(j)}}{\xi^{(j)}+xe^{(s-u)}}\right)\right]ds}\right.\\
&\quad \left. - e^{\int_t^u\left[r^{(i)}\log\left(\frac{\xi^{(i)}}{\xi^{(i)}-\vartheta+ye^{(s-u)}}\right)+r^{(j)}\log\left(\frac{\xi^{(j)}}{\xi^{(j)}-\vartheta+xe^{(s-u)}}\right)\right]ds}\right].
\end{aligned}$$

Figures 1 and 2 graphically illustrate the behavior of the risk premiuim in this setting.

Remark 30 (multifactor stochastic volatility model). The model and the results in the previous section can be generalized to allow for multifactor volatility models. Indeed, it is well known that a single non-Gaussian OU process often does not perform well in empirical studies. In particular, autocorrelation functions of realized volatilities exhibit a slower decay than can be justified by a single OU process. One way to overcome this problem is to employ a (finite) superposition of such processes, as explained in [5]. It is possible to obtain explicit formulas in this setting.

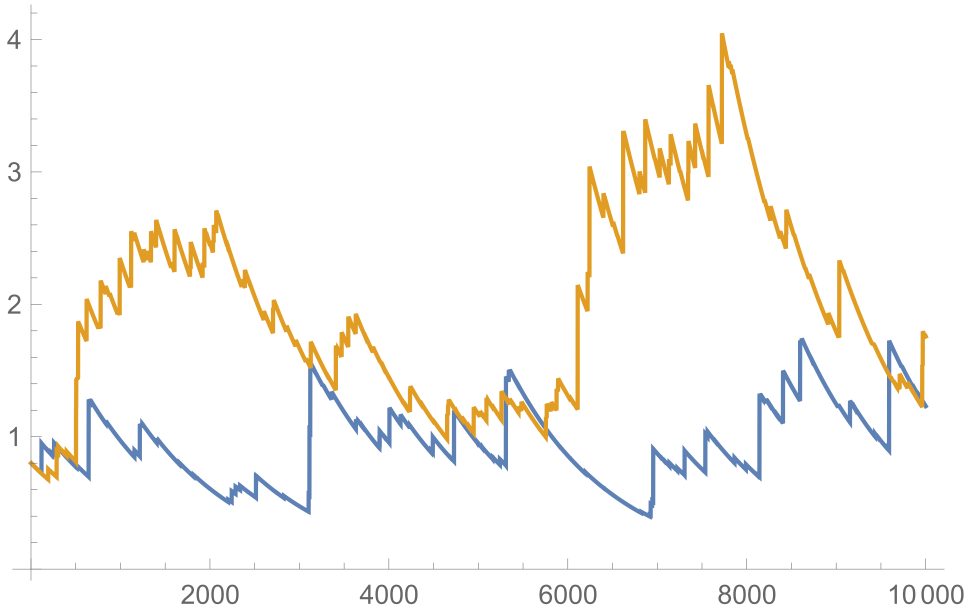

Figure 1. *Simulated paths of two non-Gaussian OU processes with Gamma processes as BDLPs, with* $\sigma_0^{2(i)} = \sigma_0^{2(j)} = 0.8, r^{(i)} = 2, r^{(j)} = 3, \nu^{(i)} = 1, \nu^{(j)} = 1$. *The blue line represents the path of* $\sigma^{2(i)}$, *the orange* $\sigma^{2(j)}$. *The scale on the time axis represents the simulation step, in such a way that* $t = 10000$ *corresponds to the final time* $T = 8.78$.

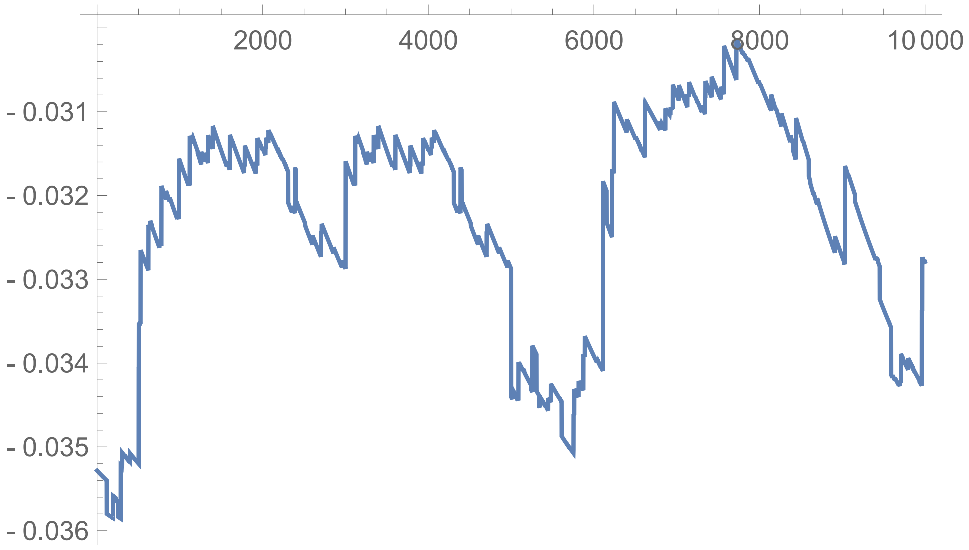

Figure 2. *Simulation of the VRP generated by the paths of the stochastic volatilities as in Figure 1, computed with Theorem 27, with* $h = T/365$. *For the change of measure we picked* $\vartheta = \frac{1}{2}$.

4. Jump component.

We proceed now with our analysis of the VRP due to the presence of jumps in our model. Not only are there many reasons to introduce a jump component in the stock prices (including, in particular, volatility smile fitting, heavy-tailed distributions of returns, inconsistency of prices for deeply out-of-the-money options; see [13, pp. 13–14] for an overview of the benefits of introducing jumps), but also the work in [30] presents empirical evidence according to which jumps in the prices are necessary to properly give an accurate description of the VRP.

Similarly to the previous sections, the *jump variance risk premium* (JVRP) is the component of the VRP originated from the jump process.

4.1. Lévy processes. The most natural choice when modeling jumps in prices are Lévy processes. In this section we provide the analysis of the VRP when the jump process $\mathbf{J}$ is a pure-jump Levy process.

Assumption 31. The process $\mathbf{J}_t = (J_t^{(1)}, \dots J_t^{(i)}, \dots J_t^{(D)})^\top$ is a D-dimensional pure-jump Lévy process with Lévy triplet $(0, 0, \boldsymbol{\eta})$ with respect to the truncation function $1_{\{|\mathbf{x}|\leq 1\}}$. We assume that all the components have moments of the second order, so that $\int_{|x|\geq 1} x^2 \eta^{(i)}(x)\, dx < \infty$. For each component $J^{(i)}$, we define its jump measure $M^{(i)}(dt, dx)$ on $\mathcal{B}\big((0,T)\big) \times \mathcal{B}(\mathbb{R})$ via the following: for $0 < s < t < T$ and $A \in \mathcal{B}(\mathbb{R})$: $M^{(i)}((s,t), A) = \#\{l \in (s,t) | \Delta J_l^{(i)} \in A\}$.

Remark 32. The presence of a multivariate Lévy process with a general Lévy measure allows for a wealth of different possibilities to model the dependence between the assets, like the use of Lévy copulas. See, for example, [13] for such an approach.

We sum up our results in the following.

Theorem 33. *The total contribution to the VRP from a D-dimensional Lévy process $\mathbf{J}$ is*

$$\sum_{i=1}^{D} \left(w^{(i)}\right)^2 \int_{\mathbb{R}} (e^x - 1)^2 \left(d\eta^{\mathbb{P}(i)}(dx) - d\eta^{\mathbb{Q}(i)}(dx)\right) + \sum_{i,j=1, i\neq j}^{D} \left(\omega^{(i)}\omega^{(j)}\right) \int_{\mathbb{R}^2} (e^x - 1)(e^y - 1) \left(d\eta^{\mathbb{P}(i,j)}(dx, dy) - d\eta^{\mathbb{Q}(i,j)}(dx, dy)\right). \tag{28}$$

The Lévy risk premium is simple to analyze: The first term in (28) accounts for the individual stock prices, while the second one accounts for the dependence between the stocks.

We observe that because of the independent increments of $\mathbf{J}$, the Lévy contribution is constant, and it does not even depend on the time span h, or the instant of time t. In particular, if h tends to zero, or infinity, the risk premium does not go to zero.

The theoretical features that our model presents so far are in line with the empirical findings in [16] and [30]: the stock price model features stochastic volatility and jumps, and even with the addition of the Lévy contributions, we obtain the required result that single stocks do not exhibit significant VRP, while pairs of correlated stocks do.

4.2. Hawkes component and the financial crisis. As we mentioned in the introduction, the empirical study in [16] was performed before the financial crisis in 2008. We now ask ourselves what would be the effect of introducing another source of jumps that accounts for the correlation that is observed within a stock market in a period of financial crisis. For this purpose, we employ the class of Hawkes processes.

Hawkes processes have been used in the literature to model default times (see, e.g., [19]) and stock returns in different economies during crises; see [1]. They are used as a model for the contagion effect between defaulting firms, or different large-scale economies. In the setting of market microstructure, they have also been used for the modeling of stock prices; see, for example, [3]. We propose to introduce them as a model of the contagion effect between the stocks within a single index.

Assumption 34. The process $\mathbf{N}_t = \left(N_t^{(1)}, \dots N_t^{(D)}\right)^\top$ is a D-dimensional Hawkes process with vector of intensities $\mathbf{\Lambda}_t = \left(\lambda_t^{(1)}, \dots \lambda_t^{(D)}\right)^\top$. The Hawkes process is a counting process, and it is defined through its intensity process. A rigorous mathematical formulation of its behavior is the following. We fix a vector $\mathbf{\Lambda}_0 = (\lambda_0^{(1)}, \dots, \lambda_0^{(D)})$ of positive real numbers and let, for any $i \in \{1, \dots D\}$,

$$\lambda_t^{(i)} = \lambda_0^{(i)} + \sum_{j=1}^{D} \int_0^t g^{(i,j)}(t-l)\, dN_l^{(j)},$$

where the deterministic functions $g^{(i,j)}(x)$ account for the time decay of a shock originated by a jump of $\mathbf{N}$. We will work with the classical Hawkes model, for which $g^{(i,j)}(x) = \alpha^{(i,j)} e^{-\beta^{(i)} x}$:

$$\lambda_t^{(i)} = \lambda_0^{(i)} + \sum_{j=1}^{D} \alpha^{(i,j)} \int_0^t e^{-\beta^{(i)}(t-l)}\, dN_l^{(j)},$$

where all the numbers $\alpha^{(i,j)}, \beta^{(i)}$ are positive. The parameter $\alpha^{(i,j)}$ represents the impact on $\lambda^{(i)}$ of a jump in the jth component $N^{(j)}$, and $\beta^{(i)}$ is the rate at which $\lambda^{(i)}$ reverts to $\lambda_0^{(i)}$.

There is an equivalent formulation of these properties that makes use of the stochastic differential equations that every $\lambda^{(i)}$ satisfies. Indeed, by the integration by parts formula, one can prove that

$$d\lambda_t^{(i)} = -\beta^{(i)}(\lambda_t^{(i)} - \lambda_0^{(i)})\, dt + \sum_{j=1}^{D} \alpha^{(i,j)}\, dN_t^{(j)}. \tag{29}$$

From this it is actually clear that $\lambda^{(i)}$ reverts to $\lambda_0^{(i)}$ *exponentially* at rate $\beta^{(i)}$. For other probabilistic properties of the Hawkes process, like its connection with the Markov property and expressions for its moments and transforms, see [19]. We will use some of those results later in the paper.

Remark 35. It is a consequence of our assumption that $\mathbf{N}$ is a multivariate point process, that any two components of the Hawkes process $\mathbf{N}$ are instantaneously *uncorrelated*, or, equivalently, they never jump together; that means, if $i \neq j$,

$$[N^{(i)}, N^{(j)}]_t = \sum_{0 \le s \le t} \Delta N_s^{(i)} \Delta N_s^{(j)} = 0 \qquad \text{a.s.}$$

The intuition here is that there is an infinitesimal time delay between a jump in $N^{(i)}$ and in $N^{(j)}$ because the value of the predictable intensity $\lambda_t^{(j)}$ affects the behavior of the right continuous process $N_t^{(j)}$ from immediately after time t.

In this section we analyze the VRP when in formula (1), $\mathbf{J}$ is a pure-jump process whose jumps come from an underlying D-dimensional Hawkes process $\mathbf{N}_t$ with intensity $\boldsymbol{\lambda}_t$. The jump sizes have fixed probability distributions $Z^{(i)}$ on $\mathbb{R}$ that have no mass at zero:

$$J_t^{(i)} = \int_{[0,t]\times\mathbb{R}} x\, M^{(i)}(ds, dx).$$

In this situation, the jump measure $M^{(i)}(dt, dx)$ is compensated by $\lambda_t^{(i)}\, dt \otimes Z^{(i)}(dx)$.

Example 36. If we choose $Z^{(i)}(dx) = \delta_1(dx)$, where δ_1 is the Dirac measure at 1, then we have $J_t^{(i)} = N_t^{(i)}$, i.e., we recover the original Hawkes process.

Applying Itô's formula, we obtain that the contribution to the ith stock dynamics is given by

$$\begin{aligned}(30)\quad &\int_0^t \int_{\mathbb{R}} (e^x - 1)\; M^{(i)}(ds, dx) \\ &= \int_0^t \int_{\mathbb{R}} (e^x - 1)\; \tilde{M}^{(i)}(ds, dx) + \int_0^t \int_{\mathbb{R}} (e^x - 1)\, Z^{(i)}(dx) \otimes \lambda_s^{(i)}\, ds.\end{aligned}$$

Assumption 37. We will assume that the compensated integral

$$\int_{[0,t]\times\mathbb{R}} (e^x - 1)^2\, \tilde{M}^{(i)}(ds, dx)$$

is a martingale for all $i \in \{1, \dots D\}$.

Remark 38. Since $J^{(i)}$ is a local martingale, a sufficient condition for Assumption (37) to be satisfied is that

$$\mathbb{E}\left[\left[\int_{[0,\cdot]\times\mathbb{R}} (e^x - 1)^2\, \tilde{M}^{(i)}(ds, dx)\right]_t\right] < \infty \tag{31}$$

for all $t > 0$ (see [27]). We have that the quantity in (31) equals

$$\mathbb{E}\left[\int_{[0,t]\times\mathbb{R}} (e^x - 1)^4\, M^{(i)}(ds, dx)\right] = \int_{\mathbb{R}} (e^x - 1)^4\, Z^{(i)}(dx) \int_0^t \mathbb{E}\left[\lambda_s^{(i)}\right]\, ds.$$

As a consequence of formula (34) below, $\int_0^t \mathbb{E}\big[\lambda_s^{(i)}\big]\, ds < \infty$ for all $t > 0$. It is then sufficient to impose conditions on the jump distributions $Z^{(i)}$ so that

$$\int_{\mathbb{R}} (e^x - 1)^4\, Z^{(i)}(dx) < \infty.$$

A possible choice is the Gaussian distribution $Z^{(i)}(dx) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx$.

We will start with the univariate case. Some of the results derived for one stock will be the building blocks to solve the multivariate case.

4.2.1. Univariate case. In the univariate setting, the JVRP at time t, over the time span h, originated by a Hawkes process is

$$\text{(32)}\quad \mathrm{JVRP}_{t,h} := \frac{1}{h}\mathbb{E}^{\mathbb{P}}\left[\left[\int_0^t \int_{\mathbb{R}} (e^x - 1)\, M^{\mathbb{P}}(du, dx)\right]_t^{t+h} \Big| \mathscr{F}_t\right] - \frac{1}{h}\mathbb{E}^{\mathbb{Q}}\left[\left[\int_0^t \int_{\mathbb{R}} (e^x - 1)\, M^{\mathbb{Q}}(du, dx)\right]_t^{t+h} \Big| \mathscr{F}_t\right].$$

Using a martingale decomposition as in (30), we see that we need to compute

$$\text{(33)}\quad \mathrm{JVRP}_{t,h} = \frac{1}{h}\mathbb{E}^{\mathbb{P}}\left[\int_t^{t+h} \lambda_u^{\mathbb{P}}\, du \Big| \mathscr{F}_t\right] \int_{\mathbb{R}} (e^x - 1)^2\, Z^{\mathbb{P}}(dx) - \frac{1}{h}\mathbb{E}^{\mathbb{Q}}\left[\int_t^{t+h} \lambda_u^{\mathbb{Q}}\, du \Big| \mathscr{F}_t\right] \int_{\mathbb{R}} (e^x - 1)^2\, Z^{\mathbb{Q}}(dx).$$

In what follows, we will denote $\kappa^{\mathbb{P}} := \int_{\mathbb{R}} (e^x - 1)^2\, Z^{\mathbb{P}}(dx)$ and similarly for $\mathbb{Q}$.

The crucial quantity we need to compute for our analysis is $\mathbb{E}\left[\lambda_u | \mathscr{F}_t\right]$. Since we could not find an explicit computation in the literature, we state it in the following lemma. Its proof can be found in the appendix.

Lemma 39. *Let N be a univariate Hawkes process with intensity $\lambda_u = \lambda_0 + \alpha \int_0^u e^{-\beta(u-l)} dN_l$. Then, for Lebesgue-a.e. $u \geq t$, it holds that*

$$\text{(34)}\quad \mathbb{E}\left[\lambda_u \Big| \mathscr{F}_t\right] = \frac{-\beta\lambda_0 + e^{(u-t)(\alpha-\beta)}(\beta\lambda_0 + (\alpha - \beta)\lambda_t)}{\alpha - \beta}.$$

We note that the result of Lemma 39 is important for our purposes. Indeed formula (34) shows the stochastic nature of $\mathbb{E}\left[\lambda_u | \mathscr{F}_t\right]$: At time t, this quantity will depend on λ_t. With a view on the operations we are going to perform next, and noting that stochastic jump intensities *do depend* on the probability measure (as opposed to the process σ^2 in the diffusion case), this is a promising result that will lead us to obtain a stochastic dynamic for the Hawkes VRP.

Exploiting Lemma 39, one can obtain the following expression for the Hawkes risk premium.

Corollary 40. *For the univariate case, the Hawkes VRP at time t, over the time span h, has the following form:*

$$\text{(35)}\quad JVRP_{t,h} = \frac{1}{h}\int_t^{t+h} \left(\kappa^{\mathbb{P}}\mathbb{E}^{\mathbb{P}}\left[\lambda_s^{\mathbb{P}} \Big| \mathscr{F}_t\right] - \kappa^{\mathbb{Q}}\mathbb{E}^{\mathbb{Q}}\left[\lambda_s^{\mathbb{Q}} \Big| \mathscr{F}_t\right]\right) dl.$$

It is of course possible to substitute the expression (34) into (35) to obtain an explicit representation of the Hawkes VRP. A few observations are in order. First we observe that the risk premium, even for a single stock, is *stochastic*, as we can infer from an analysis of formula (34): the conditional expectation $\mathbb{E}\left[\lambda_s | \mathscr{F}_t\right]$ depends on the stochastic intensity λ_t multiplied by a factor which depends on α, β, that change from $\mathbb{P}$ to $\mathbb{Q}$ (see, for example, section D). Even if the law of Z is the same under $\mathbb{P}$ and $\mathbb{Q}$, still the difference will not cancel out. Since the quadratic variation of the Hawkes process is not absolutely continuous with respect to the Lebesgue measure, the Hawkes VRP *may not* go to zero if h tends to zero. Indeed, apart from a set of times t with zero Lebesgue measure, the limit $h \to 0$ is equal to $\lambda_t^{\mathbb{P}}\kappa^{\mathbb{P}} - \lambda_t^{\mathbb{Q}}\kappa^{\mathbb{Q}}$. This is in contrast of course with the diffusion case, for which the limit is zero, for a.e. time t. As opposed to the Lévy case, this *instantaneous* risk premium depends on the time t.

4.2.2. Multivariate case.

In this section we will state the final result for the multivariate Hawkes integral. The interested reader can find the proofs in the appendix.

Proposition 41. *The Hawkes contribution to the VRP has the form*

$$JVRP_{t,h} = \frac{1}{h}\sum_{i=1}^{D}\left(\omega^{(i)}\right)^2 \int_t^{t+h} \left(\kappa^{\mathbb{P}(i)}\mathbf{V}_l^{\mathbb{P}(i)} - \kappa^{\mathbb{Q}(i)}\mathbf{V}_l{}^{\mathbb{Q}(i)}\right) dl. \tag{36}$$

An explicit expression for the matrices $\mathbf{V}^{\mathbb{P}}$ and $\mathbf{V}^{\mathbb{Q}}$ can be found in (48) in the appendix.

The observations from the univariate case carry over to the multivariate case. The most relevant thing to note is that the multivariate Hawkes risk premium is stochastic, since it depends on the stochastic intensity $\mathbf{\Lambda}_t$, through the matrices $\mathbf{V}^{\mathbb{P}}$ and $\mathbf{V}^{\mathbb{Q}}$. As seen from the expression (48), in the multivariate case, the contribution from a single stock depends on all the other intensities. Indeed, the vector $\mathbf{\Lambda}_t$, is multiplied by the matrix $e^{h(-\mathbf{A}^{\beta}\mathbf{A}^{-1}+\mathbf{A})}$.

In the limit when h tends to zero, we have

$$\lim_{h\downarrow 0} \mathrm{JVRP}_{t,h} = \sum_{i=1}^{D}\left(w^{(i)}\right)^2\left(\kappa^{\mathbb{P}(i)}\lambda_t^{\mathbb{P}(i)} - \kappa^{\mathbb{Q}(i)}\lambda_t^{\mathbb{Q}(i)}\right).$$

Example 42. In this example we show how to explicitly compute the Hawkes risk premium in the easy case with two stocks, only driven by a two-dimensional Hawkes process. Let the dynamics of the stocks be given by

$$\begin{cases} dX_t^{(1)} = \int_{\mathbb{R}} xM^{(1)}(dt,dx) \\ dX_t^{(2)} = \int_{\mathbb{R}} xM^{(2)}(dt,dx) \end{cases} \iff \begin{cases} dS_t^{(1)} = S_{t-}^{(1)} \int_{\mathbb{R}}(e^x-1)\, M^{(1)}(dt,dx) \\ dS_t^{(2)} = S_{t-}^{(2)} \int_{\mathbb{R}}(e^x-1)\, M^{(2)}(dt,dx), \end{cases}$$

and the two Hawkes intensities, where, for simplicity, we set the β coefficients equal to 1, and the matrix of the α coefficients equal to $\mathbf{A} = \begin{pmatrix} \frac{1}{3}, & \frac{1}{3} \\ 0, & \frac{1}{2} \end{pmatrix}$:

$$\begin{cases} d\lambda_t^{(1)} = -\left(\lambda_t^{(1)} - \lambda_0^{(1)}\right) dt + \frac{1}{3}\, dN_t^{(1)} + \frac{1}{3}\, dN_t^{(2)}, \\ d\lambda_t^{(2)} = -\left(\lambda_t^{(2)} - \lambda_0^{(2)}\right) dt + \frac{1}{2}\, dN_t^{(2)}. \end{cases}$$

Next, we write the index process I_t and derive its return process Y:

$$\begin{aligned} I_t &= S^{(1)}_t + S^{(2)}_t = \int_0^t S^{(1)}_{s-} \int_{\mathbb{R}} (e^x - 1)\, M^{(1)}(ds, dx) + \int_0^t S^{(2)}_{s-} \int_{\mathbb{R}} (e^x - 1)\, M^{(2)}(ds, dx),\\ Y_t &= \mathscr{L}og(I_t) = \int_0^t \frac{S^{(1)}_{s-}}{I_s} \int_{\mathbb{R}} (e^x - 1)\, M^{(1)}(ds, dx) + \int_0^t \frac{S^{(2)}_{s-}}{I_s} \int_{\mathbb{R}} (e^x - 1)\, M^{(2)}(ds, dx)\\ &= w^{(1)} \int_0^t \int_{\mathbb{R}} (e^x - 1)\, M^{(1)}(ds, dx) + w^{(2)} \int_0^t \int_{\mathbb{R}} (e^x - 1)\, M^{(2)}(ds, dx), \end{aligned}$$

thanks to our assumption of constant weights. Thus

$$[Y]_t = \left(w^{(1)}\right)^2 \int_0^t \int_{\mathbb{R}} (e^x - 1)^2\, M^{(1)}(ds, dx) + \left(w^{(2)}\right)^2 \int_0^t \int_{\mathbb{R}} (e^x - 1)^2\, M^{(2)}(ds, dx).$$

Let us now assume that the change of measure is as simple as possible, namely, that the two intensities are scaled by the same factor Γ (see also the next section for a more detailed discussion on this matter) and that $\kappa^{\mathbb{P}}$ and $\kappa^{\mathbb{Q}}$ are both equal to 1. Then, applying Proposition 41, we obtain

$$\begin{aligned} \text{(37)}\quad \mathbf{V}^{\mathbb{P}}_u(t) - \mathbf{V}^{\mathbb{Q}}_u(t) = (1-\Gamma) \times \Bigg[&\begin{pmatrix} e^{\frac{u-t}{3}} \left(\lambda^{(1)}_t + \lambda^{(2)}_t\right) \\ \lambda^{(1)}_t + e^{\frac{u-t}{2}} \lambda^{(2)}_t \end{pmatrix} \\ &+ \begin{pmatrix} 3(\lambda^{(1)}_0 + \lambda^{(2)}_0)\left(e^{\frac{u-t}{3}} - 1\right) + \lambda^{(1)}_0 (u-t) e^{\frac{u}{3}} + 2\lambda^{(2)}_0 \left(e^{\frac{u-t}{2}} - 1\right) \\ 3(\lambda^{(1)}_0 + \lambda^{(2)}_0)\left(e^{-\frac{t}{3}} - e^{-\frac{u}{3}}\right) + \lambda^{(1)}_0 (u-t) e^{\frac{u}{2}} + 2\lambda^{(2)}_0 \left(e^{\frac{u-t}{2}} - 1\right) \end{pmatrix} \Bigg]. \end{aligned}$$

If, for ease of exposition, we set $\tilde{\omega}_i = \left(\omega^{(i)}\right)^2$, we finally obtain

$$\begin{aligned} \text{(38)}\qquad \mathrm{JVRP}_{t,h} &= \frac{1-\Gamma}{h} \int_t^{t+h} \Bigg[\tilde{\omega}_2 \left(\lambda^{(1)}_t + \lambda^{(2)}_t e^{\frac{u-t}{2}}\right) + \tilde{\omega}_1 \left(\lambda^{(1)}_t + \lambda^{(2)}_t\right) e^{\frac{u-t}{3}} \\ &\quad + \lambda^{(1)}_0 (u-t) \left(\tilde{\omega}_1 e^{\frac{u}{3}} + \tilde{\omega}_2 e^{\frac{u}{2}}\right) + 2\lambda^{(2)}_0 \left(e^{\frac{u-t}{2}} - 1\right)(\tilde{\omega}_1 + \tilde{\omega}_2) \\ &\quad + 3(\lambda^{(1)}_0 + \lambda^{(2)}_0)\left(\tilde{\omega}_1 \left(e^{\frac{u-t}{3}} - 1\right) + \tilde{\omega}_2 \left(e^{-\frac{t}{3}} - e^{-\frac{u}{3}}\right)\right) \Bigg] du \\ &= \frac{1-\Gamma}{h} \Bigg[3(\lambda^{(1)}_0 + \lambda^{(2)}_0)\left(\tilde{\omega}_1 \left(3(e^{\frac{h}{3}} - 1) - h\right) + \tilde{\omega}_2 e^{-\frac{t}{3}} \left(h + 3e^{-\frac{h}{3}} - 3\right)\right) \\ &\quad + \lambda^{(1)}_0 \left(\tilde{\omega}_1 3 \left(e^{\frac{h}{3}}(h-3) + 3\right) e^{\frac{t}{3}} + \tilde{\omega}_2 2 \left(e^{\frac{h}{2}}(h-2) + 2\right) e^{\frac{t}{2}}\right) \\ &\quad + \tilde{\omega}_1 (\lambda^{(1)}_t + \lambda^{(2)}_t) 3 \left(e^{\frac{h}{3}} - 1\right) + 2\lambda^{(2)}_0 (\tilde{\omega}_1 + \tilde{\omega}_2)\left(-h + 2e^{\frac{h}{2}} - 2\right) \\ &\quad + \tilde{\omega}_2 \left(\lambda^{(1)}_t h + \lambda^{(2)}_t 2 \left(e^{\frac{h}{2}} - 1\right)\right) \Bigg]. \end{aligned}$$

Figures 3 and 4 show a numerical simulation for $JVRP_{t,h}$ and the corresponding realizations of the Hawkes intensities.

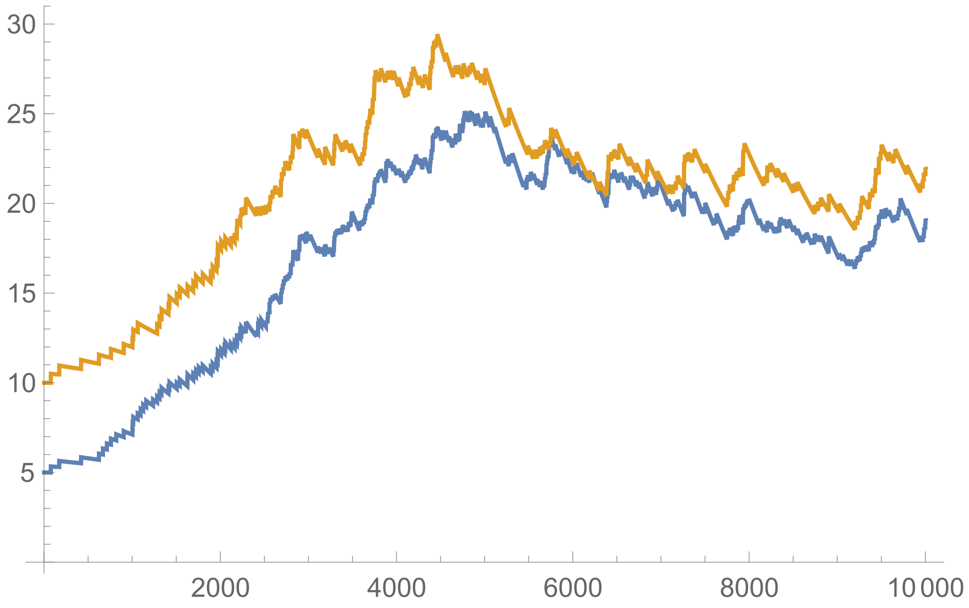

Figure 3. *Simulated paths of the intensities from a bivariate Hawkes processes as in Example 42, with* $\lambda_0^{(1)} = 5, \lambda_0^{(2)} = 10, \beta^{(1)} = \beta^{(2)} = 1, \alpha^{(1,1)} = \frac{1}{3}, \alpha^{(1,2)} = \frac{1}{3}, \alpha^{(2,1)} = 0, \alpha^{(2,2)} = \frac{1}{2}$. *As in Figure 1,* $t = 10000$ *corresponds to* $T = 8.78$.

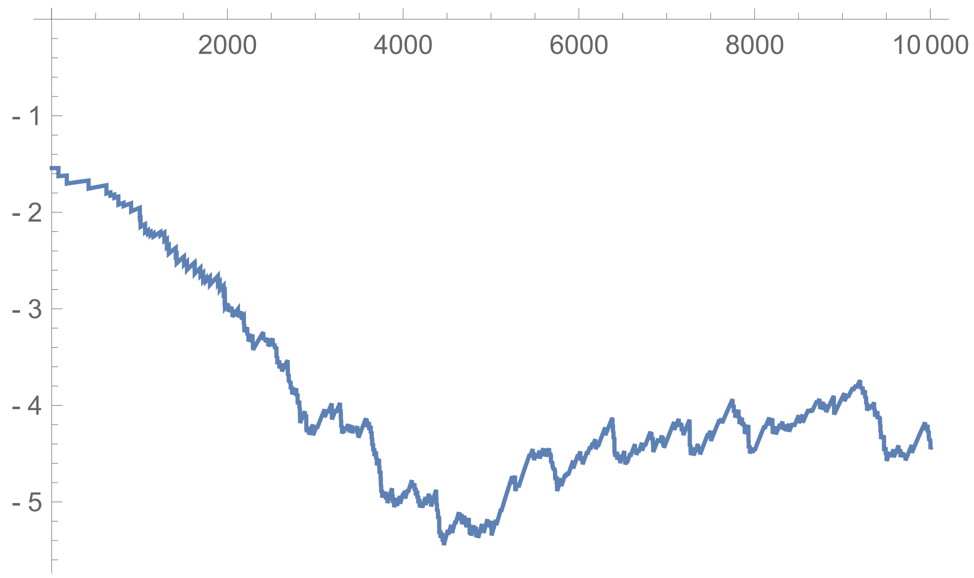

Figure 4. *Simulated path of the Hawkes risk premium, computed as in formula (42), with* $h = T/365$ *and* $\Gamma = 1.2$.

We observe that the introduction of a full multivariate jump process with a stochastic intensity has greatly increased the complexity of the dynamics of the VRP. In particular, we observe that this choice of modeling has led us to obtain stochastic dynamics for the jump component. This is an interesting result in its own, and it complements the existing literature, as in [7], the authors create a jump model with stochastic VRP either by introducing a stochastic volatility term in the Lévy integral, $d\sigma_t^2 = -\lambda\sigma_t^2\, dt + v_{\lambda t-}\, dL_{\lambda t}$, where v is a stationary, positive, càdlàg stochastic process, or by time-changing the Lévy process L with an independent non-Lévy time change (otherwise we would still obtain a Lévy process): $dL_{\tau_{\lambda t}}$, where $\tau_t = \int_0^t \xi_s\, ds$, and ξ is a positive, stationary, càdlàg process.

We finally note that this fully multivariate jump model does not allow us to distinguish anymore between the contribution of single and correlated stocks, as the weight $w^{(i)}$ is multiplied by all the components of the intensity vector.

5. Summary of results. In this work we provided new insights on the dynamics of the VRP by proposing a new multivariate stochastic model for the dynamics of an equity index. We proved that it is possible to devise a model consistent with the empirical findings that the VRP of single stocks is negligible, while the VRP originated by the correlation between the stocks exhibits stochastic fluctuations. To achieve this, we can use the diagonal non-Gaussian OU model, adding the Lévy jumps.

Moreover, we provided proofs that in that model the dependence between the assets is the main driver of the VRP, and we explained the mathematical reasons behind this phenomenon. We analyzed in detail various alternative methods of embedding a correlation structure within the index and found that popular stochastic volatility models, like the Wishart model, or the multivariate non-Gaussian OU model, do not match the empirical findings, and may even imply deterministic dynamics for the VRP.

In order to derive the contribution of the dependency between two stocks in the index, Fourier methods were employed, leading us to obtain an integral representation of the risk premium. This latter formula shows a stochastic behavior of the correlation risk premium, which can therefore be interpreted as the main factor driving the component of the diffusive VRP of the index due to the diffusion part of the stock prices.

As an answer to the need for a more sophisticated mathematical process that could be used to model stock prices during crises, we propose the class of the self- and mutual exciting Hawkes processes. For the first time in the literature we employ this class of processes in the modeling of an equity index, to allow for a contagion effect within the stocks in the index. We obtain that the stochastic Hawkes intensity drives the jump contribution to the VRP, but it then becomes impossible to split the contribution of single stocks from that of correlated pairs.

In the appendix, we establish the existence of a risk-neutral structure preserving equivalent measure for our model, thus ensuring absence of arbitrage. In particular, we characterize the class of measure transformations that preserve the self-affecting structure of the Hawkes process.

We conclude that our work provides a state-of-the-art multivariate modeling framework to analyze the VRP which is theoretically sound and also produces explicit and analytically tractable formulas.

Appendix A. Background results.

A.1. Point processes and intensities.

Definition 43 (stochastic intensity). *Let N_t be a point process adapted to some filtration $\mathscr{F}_t$, and let λ_t be a nonnegative $\mathscr{F}_t$-progressively measurable process such that for all $t \geq 0$, $\int_0^t \lambda_s\, ds < \infty, \mathbb{P}$-a.s.*

If for all nonnegative $\mathscr{F}_t$-predictable process C_t, the equality $\mathbb{E}[\int_0^\infty C_s dN_s] = \mathbb{E}\left[\int_0^\infty C_s \lambda_s\, ds\right]$ is verified, then we say that N_t admits the $(\mathbb{P}, \mathscr{F}_t)$-intensity λ_t.

We will sometimes write $\lambda^{\mathbb{P}}$ to stress the dependence on the probability measure.

We provide here the statement of Theorem 2.2 in [29] that allowed us to perform the change of measure in section E. We employ the same notation as in Definition 49.

Theorem 44. *It holds that the process*

$$\mathscr{E}\left(\sum_{i=1}^{D} \int_0^t \left(\psi_t^{(i)} - 1\right) d\tilde{N}_t^{(i)}\right)$$

is a martingale if there is an $\varepsilon > 0$ such that whenever $0 \leq u \leq t$ with $|t-u| \leq \varepsilon$, one of the following two conditions is satisfied:

$$\mathbb{E}\left[\exp\left(\sum_{i=1}^{D} \int_u^t \left(\psi_s^i \log \psi_s^i - \left(\psi_s^i - 1\right)\right) \lambda_s^i\, ds\right)\right] < \infty \qquad \textit{or} \tag{39}$$

$$\mathbb{E}\left[\exp\left(\sum_{i=1}^{D} \int_u^t \lambda_s^i\, ds + \int_u^t \max\left(0, \log \psi_s^i\right)\, dN_s^i\right)\right] < \infty. \tag{40}$$

Appendix B. Proofs of selected results in section 3. What follows is an expression for H as in Theorem 27.

$$\begin{aligned}
&H(\sigma_t^{2(i)}, \sigma_t^{2(j)}, x, y, t, u) \\
&:= -\exp\left(-ye^{\lambda^{(i)}(u-t)\sigma_t^{2(i)}}\right) \\
&\times \left[\exp\left(\lambda^{(i)} \int_t^u \varphi_{L^{(i)}}^{\mathbb{P}}\left(ye^{-\lambda^{(i)}(u-s)}\right) ds\right) - \exp\left(\lambda^{(i)} \int_t^u \varphi_{L^{(i)}}^{\mathbb{Q}}\left(ye^{-\lambda^{(i)}(u-s)}\right) ds\right)\right] \\
&- \exp\left(-ye^{\lambda^{(i)}(u-t)\sigma_t^{2(i)}}\right) \\
&\times \left[\exp\left(\lambda^{(i)} \int_t^u \varphi_{L^{(i)}}^{\mathbb{P}}\left(ye^{-\lambda^{(i)}(u-s)}\right) ds\right) - \exp\left(\lambda^{(i)} \int_t^u \varphi_{L^{(i)}}^{\mathbb{Q}}\left(ye^{-\lambda^{(i)}(u-s)}\right) ds\right)\right] \\
&+ \exp\left(-ye^{-\lambda^{(i)}(u-t)\sigma_t^{2(i)}} - xe^{-\lambda^{(j)}(u-t)\sigma_t^{2(j)}}\right) \\
&\times \left[\exp\left(\int_t^u \varphi_{(L^{(i)},L^{(j)})}^{\mathbb{P}}\left(ye^{-\lambda^{(i)}(u-s)}, xe^{-\lambda^{(j)}(u-s)}\right) ds\right)\right. \\
&\left. - \exp\left(\int_t^u \varphi_{(L^{(i)},L^{(j)})}^{\mathbb{Q}}\left(ye^{-\lambda^{(i)}(u-s)}, xe^{-\lambda^{(j)}(u-s)}\right) ds\right)\right],
\end{aligned} \tag{41}$$

where $\varphi_{L^{(i)}}$ is the Laplace exponent of $L^{(i)}$, i.e., the logarithm of the Laplace transform of the random variable $L_1^{(i)}$ and similarly for the multivariate Lévy process $(L_{\lambda^{(i)}}^{(i)}, L_{\lambda^{(j)}}^{(j)})$: We denote by $\varphi_{(L^{(i)},L^{(j)})}$ its Laplace exponent.

In Example 29 we make use of the following relation, which is the reverse implication of (27):

$$\exp\left(\int_t^u \varphi^{\mathbb{P}}_{(L^{(i)},L^{(j)})}\left(ye^{-\lambda^{(i)}(u-s)}, xe^{-\lambda^{(j)}(u-s)}\right) ds\right) = \mathbb{E}\left[e^{-y\int_t^u e^{-\lambda^{(i)}(u-s)}\, dL^{(i)}_{\lambda^{(i)}s} - x\int_t^u e^{-\lambda^{(j)}(u-s)}\, dL^{(j)}_{\lambda^{(j)}s}}\right].$$

B.1. Proof for the multivariate non-Gaussian OU model.

Proof of Proposition 15. Recall the dynamics (11) of $\mathbf{\Sigma}$:

$$d\mathbf{\Sigma}_t = (\mathbf{A}\mathbf{\Sigma}_t + \mathbf{\Sigma}_t\mathbf{A}^\top)\, dt + d\mathbf{L}_t.$$

Conditioning upon $\mathscr{F}_t$ and applying the conditional Fubini, one gets the following:

$$\mathbb{E}\left[\mathbf{\Sigma}_s \middle| \mathscr{F}_t\right] = \mathbf{\Sigma}_t + \int_t^s \left(\mathbf{A}\mathbb{E}\left[\mathbf{\Sigma}_l\middle|\mathscr{F}_t\right] + \mathbb{E}\left[\mathbf{\Sigma}_l\middle|\mathscr{F}_t\right]\mathbf{A}^\top\right) dl + (s-t)\mathbf{k}.$$

Taking derivatives with respect to s, we see that the problem of finding an explicit expression for $\mathbb{E}[\mathbf{\Sigma}_s|\mathscr{F}_t]$ boils down to solving the following matrix system of linear ODEs:

$$\begin{cases} \frac{d}{ds}\mathbb{E}\left[\mathbf{\Sigma}_s\middle|\mathscr{F}_t\right] = \left(\mathbf{A}\mathbb{E}\left[\mathbf{\Sigma}_s\middle|\mathscr{F}_t\right] + \mathbb{E}\left[\mathbf{\Sigma}_s\middle|\mathscr{F}_t\right]\mathbf{A}^\top\right) + \mathbf{k}, \\ \mathbb{E}\left[\mathbf{\Sigma}_t\middle|\mathscr{F}_t\right] = \mathbf{\Sigma}_t, \end{cases} \tag{42}$$

where $\mathbf{k}$ is the matrix of the first moments of $\mathbf{L}$, i.e., $\mathbb{E}[L^{(ij)}] = k^{(ij)}$.

Using the operator vec: $M_D(\mathbb{R}) \to \mathbb{R}^{D^2}$ that stacks the column of a $D \times D$ matrix into a single vector belonging to $\mathbb{R}^{D^2}$, we can rewrite the previous system in the following form (see [22, p. 440]):

$$\begin{cases} \frac{d}{ds}\operatorname{vec}\left(\mathbb{E}\left[\mathbf{\Sigma}_s\middle|\mathscr{F}_t\right]\right) = (Id_D \otimes \mathbf{A} + \mathbf{A} \otimes Id_D)\operatorname{vec}\left(\mathbb{E}\left[\mathbf{\Sigma}_s\middle|\mathscr{F}_t\right]\right) + \operatorname{vec}(\mathbf{k}), \\ \operatorname{vec}\left(\mathbb{E}\left[\mathbf{\Sigma}_t\middle|\mathscr{F}_t\right]\right) = \operatorname{vec}(\mathbf{\Sigma}_t), \end{cases}$$

where $\otimes: M_n(\mathbb{R}) \times M_m(\mathbb{R}) \to M_{nm}(\mathbb{R})$ is the Kronecker product. If we denote by $\tilde{\mathbf{A}}$ the linear operator $Id_D \otimes \mathbf{A} + \mathbf{A} \otimes Id_D$, then a solution to the system is given by

$$\operatorname{vec}\left(\mathbb{E}\left[\mathbf{\Sigma}_s\middle|\mathscr{F}_t\right]\right) = e^{\tilde{\mathbf{A}}(s-t)}\operatorname{vec}(\mathbf{\Sigma}_t) + \int_t^s e^{\tilde{\mathbf{A}}(s-l)} \cdot \operatorname{vec}(\mathbf{k})\, dl.$$

In [22] it is proved that $e^{t\tilde{\mathbf{A}}}\mathbf{\Sigma} = e^{t\mathbf{A}}\mathbf{\Sigma}e^{t\mathbf{A}^\top}$; see also [6]. We can thus go back to the matrix notation and finally obtain,

$$\mathbb{E}\left[\mathbf{\Sigma}_s\middle|\mathscr{F}_t\right] = e^{\mathbf{A}(s-t)}\mathbf{\Sigma}_t\, e^{\mathbf{A}^\top(s-t)} + \int_t^s e^{\mathbf{A}(s-l)} \cdot \mathbf{k} \cdot e^{\mathbf{A}^\top(s-l)}\, dl. \qquad \blacksquare$$

B.2. Proof for the Wishart model.

Proof of Lemma 18. Recall that formula (10) holds regardless of the chosen model for $\mathbf{\Sigma}$. We therefore proceed to compute the conditional expectation of $\mathbf{\Sigma}$ in the Wishart model.

All the D^2 stochastic integrals in $\int_0^t \mathbf{\Sigma}_s^{\frac{1}{2}}\, d\mathbf{W}_s$ are local martingales. A sufficient condition for them to be true martingales is that, for $t \geq 0$,

$$\mathbb{E}\left[\left[\int^{\cdot} \mathbf{\Sigma}_s^{\frac{1}{2}}\, d\mathbf{W}_s\right]_t\right] = \mathbb{E}\left[\int_0^t \mathbf{\Sigma}_s\, ds\right] < \infty; \tag{43}$$

see [24]. For all $t \geq 0$, the moment generating function of the *integrated Wishart process* $\int_0^t \mathbf{\Sigma}_s\, ds$ is defined on a neighborhood of 0; hence, the integrated Wishart process admits moments of all order and condition (43) is satisfied for $t \geq 0$. An explicit expression for the transform can be found in [20], while a detailed discussion for the integrated univariate square root process appears in [17].

Conditioning (14) upon $(\mathscr{F}_t)$ and applying the conditional Fubini leads to

$$\mathbb{E}\left[\mathbf{\Sigma}_s\middle|\mathscr{F}_t\right] = \mathbf{\Sigma}_t + \int_t^s \left(\mathbf{\Omega}\mathbf{\Omega}^\top + \mathbf{M}\,\mathbb{E}\left[\mathbf{\Sigma}_l\middle|\mathscr{F}_t\right] + \mathbb{E}\left[\mathbf{\Sigma}_l\middle|\mathscr{F}_t\right]\mathbf{M}^\top\right) dl,$$

exploiting the fact that the stochastic integrals $\int_0^{\cdot} \Sigma_s^{\frac{1}{2}(ij)}\, dW_s^{(j)}$ are martingales.

Taking, as usual, derivatives with respect to s (which we can, Lebesgue-a.e.) we obtain the following system of ODEs:

$$\begin{cases} \frac{d}{ds}\mathbb{E}\left[\mathbf{\Sigma}_s\middle|\mathscr{F}_t\right] = \mathbf{\Omega}\mathbf{\Omega}^\top + \mathbf{M}\,\mathbb{E}\left[\mathbf{\Sigma}_s\middle|\mathscr{F}_t\right] + \mathbb{E}\left[\mathbf{\Sigma}_s\middle|\mathscr{F}_t\right]\mathbf{M}^\top, \\ \mathbb{E}\left[\mathbf{\Sigma}_t\middle|\mathscr{F}_t\right] = \mathbf{\Sigma}_t. \end{cases} \tag{44}$$

We immediately see that it is identical to the system of ODEs in (42); therefore we can draw the same conclusions we drew there. In particular, we will have the solution

$$\mathbb{E}\left[\mathbf{\Sigma}_s\middle|\mathscr{F}_t\right] = e^{\mathbf{M}(s-t)}\mathbf{\Sigma}_t\, e^{\mathbf{M}^\top(s-t)} + \int_t^s e^{\mathbf{M}(s-l)}\mathbf{\Omega}\mathbf{\Omega}^\top e^{\mathbf{M}^\top(s-l)}\, dl. \qquad \blacksquare$$

Appendix C. Proofs of the results in section 4.2.

C.1. Proof for the univariate conditional expectation of the Hawkes intensity.

Proof of Lemma 39. Let $u \geq t$; then

$$\begin{aligned} \mathbb{E}\left[\lambda_u\middle|\mathscr{F}_t\right] &= \mathbb{E}\left[\lambda_0 + \alpha\int_0^u e^{-\beta(u-l)}\, dN_l\middle|\mathscr{F}_t\right] \\ &= \lambda_0 + \alpha\int_0^t e^{-\beta(u-l)}\, dN_l + \alpha\int_t^u e^{-\beta(u-l)}\mathbb{E}\left[\lambda_l\middle|\mathscr{F}_t\right] dl. \end{aligned} \tag{45}$$

A differentiation with respect to u gives that the following holds Lebesgue-a.e.:

$$\frac{d}{du}\mathbb{E}\left[\lambda_u\middle|\mathscr{F}_t\right] = -\alpha\beta\int_0^t e^{-\beta(u-l)}\, dN_l - \alpha\beta\int_t^u e^{-\beta(u-l)}\mathbb{E}\left[\lambda_l\middle|\mathscr{F}_t\right] dl + \alpha\mathbb{E}\left[\lambda_u\middle|\mathscr{F}_t\right].$$

We now note that, from (45),

$$-\alpha\beta \int_0^t e^{-\beta(u-l)}\, dN_l - \alpha\beta \int_t^u e^{-\beta(u-l)} \mathbb{E}\left[\lambda_l \middle| \mathscr{F}_t\right] dl = -\beta\left(\mathbb{E}\left[\lambda_u \middle| \mathscr{F}_t\right] - \lambda_0\right),$$

so we have that, as already seen for the other processes, $\mathbb{E}[\lambda_u|\mathscr{F}_t]$ satisfies pathwise

$$\frac{d}{du}\mathbb{E}\left[\lambda_u \middle| \mathscr{F}_t\right] = -\beta\left(\mathbb{E}\left[\lambda_u \middle| \mathscr{F}_t\right] - \lambda_0\right) + \alpha\mathbb{E}\left[\lambda_u \middle| \mathscr{F}_t\right] = (\alpha-\beta)\mathbb{E}\left[\lambda_u \middle| \mathscr{F}_t\right] + \beta\lambda_0, \tag{46}$$

with the initial condition: $\mathbb{E}[\lambda_t|\mathscr{F}_t] = \lambda_t := \lambda_0 + \alpha\int_0^t e^{-\beta(t-l)}\, dN_l$. The only solution is then given by the following formula:

$$\mathbb{E}\left[\lambda_u \middle| \mathscr{F}_t\right] = \frac{-\beta\lambda_0 + e^{(u-t)(\alpha-\beta)}(\beta\lambda_0 + (\alpha-\beta)\lambda_t)}{\alpha-\beta}. \tag{47}$$

■

C.2. Notation and results for the conditional expectation of the multivariate Hawkes process. Our aim is to compute $\mathbb{E}[\lambda_u^{(i)}|\mathscr{F}_t]$, for $u \geq t$ and for each $i \in \{1, 2 \ldots, D\}$; in order to do so, it is convenient to stack all these quantities into a D-dimensional vector $\mathbf{V}_u(t)$, whose generic ith component is $V_u^{(i)}(t) = \mathbb{E}[\lambda_u^{(i)}|\mathscr{F}_t]$. We will need some more notation:

$$U_u^{(i)}(t) = \int_t^u \mathbb{E}\left[e^{-\beta^{(i)}(u-l)}\, \lambda_l^{(i)} \middle| \mathscr{F}_t\right] dl, \qquad A^{(i,j)} = \alpha^{(i,j)}, \qquad A^{\beta(i,j)} = \beta^{(j)}\alpha^{(i,j)},$$

$$K_u^{(i)}(t) = \int_0^t e^{-\beta^{(i)}(u-l)}\, dN_l^{(i)}, \quad \mathbf{\Lambda}_0 = \left(\lambda_0^{(1)}, \ldots, \lambda_0^{(D)}\right)^\top, \quad \mathbf{\Lambda}_t = \left(\lambda_t^{(1)}, \ldots, \lambda_t^{(D)}\right)^\top.$$

We can now establish the following result.

Proposition 45. *If the matrix of weights $\mathbf{A}$ is invertible, then the vector of conditional expectations whose ith component is $\mathbb{E}^{\mathbb{P}}[\lambda_u^{(i)}|\mathscr{F}_t]$ has the form*

$$\mathbf{V}_u(t) = \int_t^u \left[e^{(u-s)(-\mathbf{A}^\beta\mathbf{A}^{-1}+\mathbf{A})}\mathbf{A}^\beta\mathbf{A}^{-1}\mathbf{\Lambda}_0\right] ds + e^{(u-t)(-\mathbf{A}^\beta\mathbf{A}^{-1}+\mathbf{A})}\mathbf{\Lambda}_t, \tag{48}$$

where $e^{\mathbf{A}}$ denotes the matrix exponential.

Appendix D. Hawkes change of intensity. In this section we discuss how to change a given intensity vector of a Hawkes process in such a way to obtain another Hawkes intensity via an equivalent measure change. The following theorem constitutes the fundamental theoretical motivation that justifies our derivation of the Hawkes risk premium result.

Theorem 46 (characterization of Hawkes structure preserving changes of measure). *Let N be a Hawkes process under $\mathbb{P}$ with intensity $\lambda^{\mathbb{P}}$, satisfying*

$$d\lambda_t^{\mathbb{P}} = \beta(\lambda_0^{\mathbb{P}} - \lambda_t^{\mathbb{P}})\, dt + \alpha dN_t,$$

and let $\lambda^{\mathbb{Q}}$ be another Hawkes intensity, that satisfies the SDE

$$d\lambda_t^{\mathbb{Q}} = \tilde{\beta}(\lambda_0^{\mathbb{Q}} - \lambda_t^{\mathbb{Q}})\, dt + \tilde{\alpha} dN_t. \tag{49}$$

Here $\alpha, \tilde{\alpha}, \beta, \tilde{\beta}, \lambda_0^{\mathbb{P}}, \lambda_0^{\mathbb{Q}}$ are all strictly positive real numbers. Then, there exists a probability measure $\mathbb{Q}$, equivalent to $\mathbb{P}$, under which N is a Hawkes process with intensity $\lambda^{\mathbb{Q}}$. The density process $\frac{d\mathbb{Q}}{d\mathbb{P}}$ is

$$\frac{d\mathbb{Q}}{d\mathbb{P}}\bigg|_{\mathscr{F}_t} = \mathscr{E}\left(\int_0^t (H_s - 1)\, d\tilde{N}_s^{\mathbb{P}}\right), \tag{50}$$

where $H_t := \frac{\lambda_t^{\mathbb{Q}}}{\lambda_t^{\mathbb{P}}}$ and has the dynamics

$$dH_t = \left[\frac{1}{\lambda_{t-}^{\mathbb{P}}}\left(\tilde{\beta}(\lambda_0^{\mathbb{Q}} - \lambda_t^{\mathbb{Q}}) - \beta\frac{\lambda_{t-}^{\mathbb{Q}}}{\lambda_{t-}^{\mathbb{P}}}(\lambda_0^{\mathbb{P}} - \lambda_t^{\mathbb{P}})\right)\right] dt + \frac{1}{\lambda_{t-}^{\mathbb{P}}}\left[\tilde{\alpha} - \frac{\lambda_{t-}^{\mathbb{Q}}\alpha}{\lambda_t^{\mathbb{P}}} - \frac{\alpha\tilde{\alpha}}{\lambda_t^{\mathbb{P}}}\right] dN_t. \tag{51}$$

Remark 47. We wish to stress that we use the same notation for the process N under the two measures $\mathbb{P}$ and $\mathbb{Q}$, although the process possesses different characteristics under the two measures. The relation between the characteristic triplets of a semimartingale under two different (locally absolutely continuous) measures is provided by the Girsanov's theorem for semimartingales (see [23] for a reference). In particular, the compensator of N under $\mathbb{P}$ is $\int_0^t \lambda_s^{\mathbb{P}}\, ds$, while under $\mathbb{Q}$ it is $\int_0^t \lambda_s^{\mathbb{Q}}\, ds$. We will therefore denote $\tilde{N}^{\mathbb{P}} := N_t - \int_0^t \lambda_s^{\mathbb{P}}\, ds$ and $\tilde{N}^{\mathbb{Q}} := N_t - \int_0^t \lambda_s^{\mathbb{Q}}\, ds$.

The next example shows a concrete and easy situation where it is possible to obtain another Hawkes intensity from a given one. Its proof follows later in the appendix.

Example 48. Let $\mathbf{\Lambda}$ be a D-variate stochastic intensity process of a Hawkes process, following the SDE in (29). Let $\Gamma > 0$ be a constant and consider the scaled vector $\mathbf{\Lambda}_t^{\mathbb{Q}} = \Gamma\mathbf{\Lambda}_t$. Then, also $\mathbf{\Lambda}^{\mathbb{Q}}$ is the intensity process of a Hawkes process, with parameters $\beta^{(i)\mathbb{Q}} = \beta^{(i)}$, $\lambda_0^{(i)\mathbb{Q}} = \Gamma\lambda_0^{(i)}$, $\alpha^{(i,j)\mathbb{Q}} = \Gamma\alpha^{(i,j)}$.

Proof of Theorem 46. The proof of Theorem 46 is divided into two steps. In the first one we will prove that an equivalent measure exists and that the density is given by the stated stochastic exponential (50). In the second step we will give an explicit representation of the process H in (50).

First step. The existence of the change of measure builds upon a result in [29]. Theorem 2.2 and Corollary 2.3 in that work state sufficient conditions for the change of measure to be given by the exponential martingale (50). As an application of that, they show in Example 4.6 the existence of a change of measure between $\lambda \equiv 1$ and $\lambda^{\mathbb{Q}} = \varphi(\int_0^t h(t-l)\, dN_l)$ for a bounded h, and $\varphi(x) \leq |x|$. In our specification of the Hawkes intensity, we assume $h(t-l) = \alpha \exp(-\beta(t-l))$, which is trivially bounded for $0 \leq l \leq t$, but we have $\varphi(x) = \lambda_0 + x$. It is straightforward to extend that result to a case where $\varphi(x) \leq K + |x|$ for $K > 0$. Indeed, let $h(x) < C$; then

$$\lambda_t^{\mathbb{Q}} = K + \int_0^t h(t-l)\, dN_l < K + CN_t,$$

and the result follows from Example 4.3 in that work. Our existence result is now proven by just observing that, for the same reason, there exists a change of measure bringing an intensity always equal to 1 into $\lambda^{\mathbb{P}}$, and it is therefore possible to build a change of measure bringing $\lambda^{\mathbb{P}}$ into $\lambda^{\mathbb{Q}}$.

Second step. The dynamics of H are obtained through Itô's formula for semimartingales. First, recall Itô's product rule:

$$d\left(H_t\right) = d\left(\lambda_t^{\mathbb{Q}} \cdot \frac{1}{\lambda_t^{\mathbb{P}}}\right) = \lambda_{t^-}^{\mathbb{Q}}\, d\left(\frac{1}{\lambda_t^{\mathbb{P}}}\right) + \frac{1}{\lambda_{t^-}^{\mathbb{P}}} d(\lambda_t^{\mathbb{Q}}) + d(\lambda^{\mathbb{Q}}) d\left(\frac{1}{\lambda_t^{\mathbb{P}}}\right).$$

Since $\Delta\lambda_t^{\mathbb{P}} = \alpha 1_{\{\Delta N_t = 1\}}$, Itô's formula gives

$$\begin{aligned} d\left(\frac{1}{\lambda_t^{\mathbb{P}}}\right) &= -\frac{1}{\left(\lambda_{t^-}^{\mathbb{P}}\right)^2}\left(\beta\left(\lambda_0^{\mathbb{P}} - \lambda_t^{\mathbb{P}}\right) dt + \alpha\, dN_t\right) + \left(\frac{1}{\lambda_t^{\mathbb{P}}} - \frac{1}{\lambda_{t^-}^{\mathbb{P}}} + \frac{1}{\left(\lambda_{t^-}^{\mathbb{P}}\right)^2}\alpha\right) dN_t \\ &= -\frac{1}{\left(\lambda_{t^-}^{\mathbb{P}}\right)^2}\left(\beta\left(\lambda_0^{\mathbb{P}} - \lambda_t^{\mathbb{P}}\right) dt\right) + \left(\frac{1}{\lambda_t^{\mathbb{P}}} - \frac{1}{\lambda_{t^-}^{\mathbb{P}}}\right) dN_t. \end{aligned}$$

Algebra shows that

$$\left(\frac{1}{\lambda_t^{\mathbb{P}}} - \frac{1}{\lambda_{t^-}^{\mathbb{P}}}\right) dN_t = -\frac{\alpha}{\lambda_t^{\mathbb{P}}\lambda_{t^-}^{\mathbb{P}}}\, dN_t,$$

so we have

$$d\left(\frac{1}{\lambda_t^{\mathbb{P}}}\right) = -\frac{1}{\left(\lambda_{t^-}^{\mathbb{P}}\right)^2}\left(\beta\left(\lambda_0^{\mathbb{P}} - \lambda_t^{\mathbb{P}}\right) dt\right) - \frac{\alpha}{\lambda_t^{\mathbb{P}}\lambda_{t^-}^{\mathbb{P}}}\, dN_t.$$

As a consequence

$$d\left(\lambda^{\mathbb{Q}}\right) d\left(\frac{1}{\lambda_t^{\mathbb{P}}}\right) = -\frac{\alpha\tilde{\alpha}}{\lambda_t^{\mathbb{P}}\lambda_{t^-}^{\mathbb{P}}}\, dN_t.$$

An easy rearrangement of the terms in Ito's product rule finally yields

$$d\left(\frac{\lambda_t^{\mathbb{Q}}}{\lambda_t^{\mathbb{P}}}\right) = \left[\frac{1}{\lambda_{t^-}^{\mathbb{P}}}\left(\tilde{\beta}(\lambda_0^{\mathbb{Q}} - \lambda_t^{\mathbb{Q}}) - \beta\frac{\lambda_{t^-}^{\mathbb{Q}}}{\lambda_{t^-}^{\mathbb{P}}}(\lambda_0^{\mathbb{P}} - \lambda_t^{\mathbb{P}})\right)\right] dt + \frac{1}{\lambda_{t^-}^{\mathbb{P}}}\left[\tilde{\alpha} - \frac{\lambda_{t^-}^{\mathbb{Q}}\alpha}{\lambda_t^{\mathbb{P}}} - \frac{\alpha\tilde{\alpha}}{\lambda_t^{\mathbb{P}}}\right] dN_t. \quad \blacksquare$$

Proof of Example 48. The proof is trivial and follows from the SDE that any of the components of $\mathbf{\Lambda}^{\mathbb{Q}}$ satisfies. Let $i \in \{1, \ldots, D\}$; then

$$d\left(\lambda_t^{(i)\mathbb{Q}}\right) = d\left(\Gamma\lambda_t^{(i)}\right) = \Gamma\beta^{(i)}(\lambda_0^{(i)} - \lambda_t^{(i)})\, dt + \Gamma\sum_j \alpha^{(i,j)}\, dN_t^{(j)}.$$

If we call $\beta^{(i)\mathbb{Q}} = \beta^{(i)}$, $\lambda_0^{(i)\mathbb{Q}} = \Gamma\lambda_0^{(i)}$, $\alpha^{(i,j)\mathbb{Q}} = \Gamma\alpha^{(i,j)}$, then the SDE becomes

$$d\left(\lambda_t^{(i)\mathbb{Q}}\right) = \beta^{(i)\mathbb{Q}}\left(\lambda_0^{(i)\mathbb{Q}} - \lambda_t^{(i)\mathbb{Q}}\right) dt + \sum_j \alpha^{(i,j)\mathbb{Q}}\, dN_t^{(j)},$$

thus proving the claim. $\blacksquare$

Appendix E. Existence of a structure preserving change of measure.

E.1. Leverage and multivariate non-Gaussian OU volatility. In this section we state the result of existence of a structure preserving $\mathbb{Q}$, which we previously assumed in the computations of the variance risk premia, in the special case where the volatility matrix has a multivariate non-Gaussian OU specification and in presence of both a Lévy process and the Hawkes process in the dynamics of the stocks.

Throughout this section, we will denote by $\mathscr{E}(X)$ the *stochastic exponential* of the semimartingale X, defined to be the only process Y that satisfies the SDE:

$$dY_t = Y_{t^-}\, dX_t.$$

A structure preserving change of measure for the multivariate non-Gaussian OU model exists and is proved in [25]. With the introduction of a multivariate Hawkes process, we need to extend their result. We introduce the following.

Definition 49. *Assume we are given a D-dimensional point process $\mathbf{N}$ with nonnegative, predictable, and locally bounded stochastic intensity $\mathbf{\Lambda}$, and let $\mathbf{\Lambda}^{\mathbb{Q}}$ be another nonnegative, predictable, and locally bounded process.*

*Then $\mathbf{\Lambda}^{\mathbb{Q}}$ is said to be $\mathbf{\Lambda}$-*compatible *if, for any $i \in \{1, \dots, D\}$, it holds that $\lambda_t^{(i)\mathbb{Q}} = 0$ whenever $\lambda_t^{(i)} = 0$ and the process $\psi_t^{(i)} := \frac{\lambda_t^{(i)\mathbb{Q}}}{\lambda_t^{(i)}}$ is locally integrable.*

The change of intensity described in Theorem 46 meets both the compatibility and the local integrability conditions. Indeed, the natural choice of stopping times to consider are the jump times $\{\tau_k\}_{k\in\mathbb{N}}$ of $\mathbf{N}$. The stopped processes $\psi_{\tau_k}^{(i)}$ are bounded, and hence, for all $i \in \{1, \dots D\}$, $\psi^{(i)}$ is locally integrable.

An application of Theorem 2.2 in [29] (whose statement can also be found in Theorem 44) gives us that the martingale $\mathscr{E}\big(\sum_{i=1}^{D} \int_0^t \big(\psi_t^{(i)} - 1\big)\, d\tilde{N}_t^{(i)}\big)$ defines a new probability under which $\mathbf{N}$ is a point process with Hawkes intensity given by (49). An equivalent martingale measure for our model (1) is then given by the product measure between one for the multivariate non-Gaussian OU model in [25] and the one just described above. We can formally collect these observation in the following.

Theorem 50. *Let a price model be given by $S_t^{(i)} = \exp(X_t^{(i)})$ for $i \in \{1, \dots D\}$, where the dynamics of $X^{(i)}$ are modeled via the multivariate specification*

$$\begin{cases} d\mathbf{X}_t &= (\boldsymbol{\mu}_t + \boldsymbol{\beta}(\mathbf{\Sigma}_t))\, dt + \mathbf{\Sigma}_t^{\frac{1}{2}}\, d\mathbf{W}_t + \boldsymbol{\rho}(d\mathbf{L}_t) + \boldsymbol{\zeta}_{t^-}\, d\mathbf{N}_t, \\ d\mathbf{\Sigma}_t &= (\mathbf{A}\mathbf{\Sigma}_t + \mathbf{\Sigma}_t\mathbf{A}^\top)\, dt + d\mathbf{L}_t, \end{cases}$$

where $\boldsymbol{\rho}\colon \mathbb{R}^{D\times D} \to \mathbb{R}^D$ is linear. Let another Hawkes intensity $\mathbf{\Lambda}^{\mathbb{Q}}$, compatible with $\mathbf{\Lambda}$, be given, and let $\boldsymbol{\psi}$ be such that

$$\psi^{(i)}(\mathbf{x}, t) = 1_{\{\mathbf{x}^{(i)}=1\}} \frac{\lambda^{\mathbb{Q},i}}{\lambda^{\mathbb{P},i}}.$$

Further, let $\boldsymbol{\chi}\colon \mathbb{S}_D^+ \to (0,\infty)$ *be such that, for* $i \in \{1,\dots,D\}$,

$$\int_{\mathbb{S}_D^+} \left(\sqrt{\boldsymbol{\chi}(\mathbf{X})} - 1\right)^2 \boldsymbol{\nu}(d\mathbf{X}) < \infty \quad \text{and} \quad \int_{\|\mathbf{X}\|>1} e^{\rho^{(i)}(\mathbf{X})} \boldsymbol{\nu}^{\boldsymbol{\chi}}(\mathbf{X}) d\mathbf{X} < \infty,$$

where for $B \in \mathscr{B}(\mathbb{S}_D^+)$, $\boldsymbol{\nu}^{\boldsymbol{\chi}}(B) = \int_B \boldsymbol{\chi}(\mathbf{X})\nu(d\mathbf{X})$.

If we define

$$\boldsymbol{\varphi}_t = -\boldsymbol{\Sigma}_t^{-\frac{1}{2}} \left(\boldsymbol{\mu}_t + \boldsymbol{\beta}(\boldsymbol{\Sigma}_t) + \begin{pmatrix} \frac{\Sigma_t^{11}}{2} \\ \vdots \\ \frac{\Sigma_t^{DD}}{2} \end{pmatrix} + \begin{pmatrix} \int_{\mathbb{S}_D^+} e^{\rho^{(1)}(\mathbf{X})} \nu^{\boldsymbol{\chi}}(\mathbf{X}) d\mathbf{X} \\ \vdots \\ \int_{\mathbb{S}_D^+} e^{\rho^{(D)}(\mathbf{X})} \nu^{\boldsymbol{\chi}}(\mathbf{X}) d\mathbf{X} \end{pmatrix} + \begin{pmatrix} \psi^{(1)} \\ \vdots \\ \psi^{(D)} \end{pmatrix} - \mathbf{1}r \right),$$

then the process

$$\vartheta_t = \mathscr{E}\left(\boldsymbol{\varphi} \cdot \boldsymbol{W} + (\boldsymbol{\chi} - 1) \star (\boldsymbol{\mu}^{\mathbf{Z}} - \boldsymbol{\nu}^{\mathbf{Z}}) + (\boldsymbol{\psi} - 1) \star (\boldsymbol{\nu}^{\mathbf{N}} - \boldsymbol{\mu}^{\mathbf{N}})\right)$$

defines a new equivalent martingale probability measure $\mathbb{Q}$ *via* $\frac{d\mathbb{Q}}{d\mathbb{P}} = \vartheta_T$, *under which the log-price process follows:*

$$\begin{cases} dX_t^{(i)} = \left(r - \dfrac{1}{2}\Sigma_t^{ii} - \displaystyle\int_{\mathbb{S}_D^+} \left(e^{\rho^{(i)}(\mathbf{X})} - 1\right) d\boldsymbol{\nu}^{\boldsymbol{\chi}}(d\mathbf{X}) - Z_t^{(i)} \lambda_t^{(i)} \boldsymbol{\psi}_t^{(i)} \right) dt \\ \qquad\qquad + \left(\boldsymbol{\Sigma}_t^{\frac{1}{2}} \, d\mathbf{W}_t^{\mathbb{Q}} \right)^{(i)} + \rho^{(i)}(d\mathbf{Z}_t) + \zeta_{t-}^{(i)} \, d\mathbf{N}_t^{(i)}, \\ d\boldsymbol{\Sigma}_t = (\mathbf{A}\boldsymbol{\Sigma}_t + \boldsymbol{A}^\top \boldsymbol{\Sigma}_t) \, dt + d\mathbf{L}_t, \end{cases}$$

where $\mathbf{W}^{\mathbb{Q}}$ *is a* $\mathbb{Q}$*-Brownian motion,* $\mathbf{L}$ *is an independent Lévy process with Lévy measure* $\boldsymbol{\nu}^{\boldsymbol{\chi}}$, *and* $\mathbf{N}$ *is an independent Hawkes process with intensity* $\boldsymbol{\Lambda}^{\mathbb{Q}}$.

Acknowledgments. We wish to thank Dan Crisan and Alexander Sokol for helpful comments on an earlier draft of this paper.

REFERENCES

[1] Y. Ait-Sahalia, J. Cacho-Diaz, and R. J. Laeven, *Modeling Financial Contagion Using Mutually Exciting Jump Processes*, Tech. report, National Bureau of Economic Research, 2010.

[2] D. Applebaum, *Lévy Processes and Stochastic Calculus*, Cambridge Stud. Adv. Math. 93, Cambridge University Press, Cambridge, UK, 2004.

[3] E. Bacry, S. Delattre, M. Hoffmann, and J. F. Muzy, *Modelling microstructure noise with mutually exciting point processes*, Quant. Finance, 13 (2013), pp. 65–77.

[4] O. E. Barndorff-Nielsen and N. Shephard, *Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics*, J. R. Stat. Soc. Ser. B Stat. Methodol., 63 (2001), pp. 167–241.

[5] O. E. Barndorff-Nielsen and N. Shephard, *Econometric analysis of realized volatility and its use in estimating stochastic volatility models*, J. R. Stat. Soc. Ser. B Stat. Methodol., 64 (2002), pp. 253–280.

[6] O. E. Barndorff-Nielsen and R. Stelzer, *Positive-definite matrix processes of finite variation*, Probab. Math. Statist., 27 (2007), pp. 3–43.

[7] O. E. Barndorff-Nielsen and A. E. D. Veraart, *Stochastic volatility of volatility and variance risk premia*, J. Financial Econometrics, 11 (2013), pp. 1–46.
[8] T. Bollerslev, *Modelling the coherence in short-run nominal exchange rates: A multivariate generalised ARCH model*, Rev. Econom. Statist., 72 (1990), pp. 498–505.
[9] T. Bollerslev, J. Marrone, L. Xu, and H. Zhou, *Stock return predictability and variance risk premia: Statistical inference and international evidence*, J. Financial Quantitative Anal., 49 (2014), pp. 633–661.
[10] T. Bollerslev, G. Tauchen, and H. Zhou, *Expected stock returns and variance risk premia*, Rev. Financial Studies, 22 (2009), pp. 4463–4492.
[11] O. Bondarenko, *Variance trading and market price of variance risk*, J. Econometrics, 180 (2014), pp. 81–97.
[12] P. Carr and L. Wu, *Variance risk premiums*, Rev. Financial Studies, 22 (2009), pp. 1311–1341.
[13] R. Cont and P. Tankov, *Financial Modelling with Jump Processes*, Chapman & Hall/CRC, Boca Raton, FL, 2003.
[14] J. Da Fonseca, M. Grasselli, and C. Tebaldi, *A multifactor volatility Heston model*, Quant. Finance, 8 (2008), pp. 591–604.
[15] I. Drechsler and A. Yaron, *What's vol got to do with it*, Rev. Financial Studies, 24 (2011), pp. 1–45.
[16] J. Driessen, P. Maenhout, and G. Vilkov, *The price of correlation risk: Evidence from equity options*, J. Finance, 64 (2009), pp. 1377–1406.
[17] D. Dufresne, *The Integrated Square-Root Process*, working paper, University of Melbourne, 2001.
[18] R. Engle, *Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models*, J. Bus. Econom. Statist., 20 (2002), pp. 339–350.
[19] E. Errais, K. Giesecke, and L. R. Goldberg, *Affine point processes and portfolio credit risk*, SIAM J. Financial Math., 1 (2010), pp. 642–665.
[20] C. Gouriéroux, *Continuous time Wishart process for stochastic risk*, Econometric Rev., 25 (2006), pp. 177–217.
[21] A. G. Hawkes, *Spectra of some self-exciting and mutually exciting point processes*, Biometrika, 58 (1971), pp. 83–90.
[22] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1991.
[23] J. Jacod and A. N. Shiryaev, *Limit Theorems for Stochastic Processes*, Corundlehren Math. Wiss. 288, Springer-Verlag, Berlin, 1987.
[24] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer, New York, 1991.
[25] J. Muhle-Karbe, O. Pfaffel, and R. Stelzer, *Option pricing in multivariate stochastic volatility models of ou type*, SIAM J. Financial Math., 3 (2012), pp. 66–94.
[26] E. Nicolato and E. Venardos, *Option pricing in stochastic volatility models of the Ornstein-Uhlenbeck type*, Math. Finance, 13 (2003), pp. 445–466.
[27] P. Protter, *Stochastic Integration and Differential Equations: Version* 2.1, Stoch. Model. Appl. Probab. 21, Springer, New York, 2005.
[28] D. Serre, *Matrices: Theory and Applications*, Grad. Texts in Math. 216, Springer, New York, 2010.
[29] A. Sokol and R. N. Hansen, *Exponential martingales and changes of measure for counting processes*, Stoch. Anal. Appl., 33 (2015), pp. 823–843.
[30] V. Todorov, *Variance risk-premium dynamics: The role of jumps*, Rev. Financial Studies, 23 (2010), pp. 345–383.