Implied Calibration and Moments Asymptotics in Stochastic Volatility Jump Diffusion Models

S. Galluccio*, Y. Le Cam[‡]

*BNPParibas, London, [‡] French Treasury, Evry University

February 13, 2007

Abstract

In the context of arbitrage-free modelling of financial derivatives, we introduce a novel calibration technique for models in the affine-quadratic class for the purpose of over-the-counter option pricing and risk-management. In particular, we aim at calibrating a stochastic volatility jump diffusion model to the whole market implied volatility surface at any given time. We study the asymptotic behaviour of the moments of the underlying distribution and use this information to introduce and implement our calibration algorithm. We numerically show that the proposed approach is both statistically stable and accurate.

***Keywords**: Linear-Quadratic models, Option pricing, Model calibration, Moments asymptotics

JEL classification: G12, G13.

Acknowledgments: We are grateful to M.Jeanblanc-Piqué, J.-M. Ly and to the participants at the 2005 Financial Modelling Workshop at the Newton Institute (Cambridge), the 2005 ICBI Global Derivatives Conference (Paris), the 2005 Risk-Management Conference (Rome) and the 2005 Derivatives RISK Conference (London) for helpful comments.

Email address: stefano.galluccio@bnpparibas.com (S. Galluccio), yann.le-cam@polytechnique.org (Y. Le Cam)

1 Introduction

The Black-Scholes (BS) paradigm of lognormality of asset returns distribution is in contrast with empirical observations. For this reason, many different generalizations of the BS model have been proposed in the past. Empirical evidence seems to reject stochastic volatility (SV) models since they are not capable of reproducing the observed conditional kurtosis of returns. The presence of jumps is often advocated as a solution to this problem. In fact, evidence of presence of jumps in the asset, in the volatility or in both is reported in Bates (1996), Bakshi *et al.* (1997), Chernov *et al.* (1999), Andersen *et al.*(2002), Pan (2002), Bates (2000), Eraker *et al.* (2003) and Chernov *et al.* (2003), among others.

In another set of studies, departures from BS model are advocated in relation to the implied volatility smile phenomenon. Smile-consistent deterministic volatility extensions of the BS model were first introduced by Dupire (1994) and Derman and Kani (1994). These are usually referred to as local volatility (LV) models. Although LV models provide a simple mechanism for smile generation, they are pledged by a number of shortcomings (Dumas *et al.* (1997), Rebonato (2000), Andersen and Andreasen (2000), Di Graziano and Galluccio (2005)). In a different line of thought, Hull and White (1987), Stein and Stein (1991) and Heston (1993), account for the smile phenomenon through stochastic volatility (SV) models. Finally, modelling the smile through mixed jump-diffusion (JD) processes is proposed in Andersen et al. (2000) (for LV models with jumps) and Duffie et al. (2000) (for SV models with jumps). In general, forcing a *single* SV or JD model to be consistent with the whole set of smiles at different maturities (the so-called "implied volatility surface") is impossible, unless model coefficients are heavily (and unrealistically) time-dependent. In fact, SV (resp. JD) models tend to underestimate smile convexity at short (resp. long) maturities (see Section 3).

It is now an established fact that SV, JD and LV models should be rejected in favour of stochastic volatility jump-diffusion models (SVJD) thanks to their superior market explicative power (Bates (1996), Bakshi *et al.* (1997), Andersen *et al.*(2002), Pan (2002), Bates (2000), Eraker *et al.* (2003) and Chernov *et al.* (2003)) .

A fundamental problem is the estimation of latent parameters in SVJD models. Statistical estimation from historical data series has been given extensive coverage in the past (see also Chernov *et al.* (2003), Craine, Lochstoer and Syrtveit (2000) and Deelstra et al. (2003)). However, for the purpose of pricing and hedging over-the counter derivatives, a model must be consistent with the available market quotations of liquid vanilla options at any given time to avoid arbitrage opportunities. In this respect statistical estimations must be replaced or, at least, complemented by a reverse engineering process (model calibration) that aims at determining model parameters in order to reproduce the observed vanilla option prices[1].

Despite the importance of having a fast, robust and accurate model calibration, the academic

[1] In a complete market, model calibration identifies the (unique) risk-neutral measure and avoids the problem of determining the market price of risk. When market is incomplete, calibration allows selecting one "market" measure among the infinite set of risk-neutral measures (see Björk (1998) for details).

literature on this subject is scarce. In this respect, calibration based on short-term asymptotics is addressed in Medvedev and Scaillet (2004). Bakshi *et al.* (1997) and Andersen and Andreasen (2000) suggest to calibrate a model by minimizing the sum of the squared errors of all available options across all strikes and maturities. This simple non-linear least squares optimization is usually not enough accurate and not statistically robust, as shown in Cont and Tankov (2004). These authors point out that the information contained in the set of available option prices is not sufficient to remove the coefficients degeneracy that is associated to a SVJD process and suggest that calibration can only be achieved provided one adds exogenous information in addition to the available option prices. For, they introduce a calibration algorithm (in the context of exponential-Lévy processes) where the objective function contains a convex functional that is meant to stabilize the (non-convex) optimization problem. Cont and Tankov (2004) focus on calibrating a single smile at the time. Generalizing their approach to more general processes or to cope with the calibration of the whole volatility surface remains an unsolved issue.

In this paper, inspired by Cont and Tankov (2004) results, we attempt to take the next step in this direction and introduce a novel implied calibration methodology for a wide class of SVJD models with time-dependent coefficients (Duffie *et al.* (2000), Piazzesi (2003), Peng and Scaillet (2004)). Our approach is qualitatively inspired at Cont and Tankov's method, but it differs from that in both the nature of the problem (we aim at calibrating the whole volatility surface as opposed to a single smile curve) and in the type of dynamics (we do not restrict ourselves to Lévy processes). Our approach takes heavily into account the asymptotic behavior of the moments of the underlying distribution. To this aim, we derive asymptotic moments formulae in a simplified case and we discuss how to use them to simplify the calibration of long terms smiles. We apply our method to one of the simplest (yet non-trivial) SVJD model with jumps in the asset and we show that an accurate and financially meaningful calibration to the whole volatility surface is possible and the algorithm is statistically robust. However, our study strongly suggests that the algorithmic complexity is such that generalizing the present approach to more complex models might be difficult to achieve. This is the case, for instance, when jumps in volatility are also present or for more general local volatility forms[2]. These shortcomings expose an intrinsic limitation of SVJD models and clearly indicate that, despite their mathematical and financial appeal, further theoretical developments are needed in this area of research.

The rest of the paper is organized as follows. In Section 2 we introduce model and notations, and we determine closed-form formulae for European options. In Section 3 we analytically study the asymptotic properties of SVJD models and analyze the different role played by jumps and stochastic volatility in explaining the market smile. Section 4 introduces and discusses the calibration problem. Section 5 is devoted to a detailed description of our calibration algorithm. Numerical results are presented in Section 6 while Section 7 contains conclusions and prospects for future research.

[2] One important exception being hybrid (Equity-IR) modelling with no jumps in volatility since in that case the approach we present here can be applied without major modifications, as shown in Galluccio and Le Cam (2005).

2 Mathematical setup and option pricing

2.1 The model

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. We shall denote by S_t the stock price at time t and by r_t the (deterministic) spot interest rate. A probability measure $\mathbb{P}^*$, equivalent to the historical probability measure $\mathbb{P}$, is said to be a risk-neutral measure if the discounted price follows a local martingale process under $\mathbb{P}^*$. We stipulate that the asset dynamics follows a jump-diffusion process[3] under any equivalent risk-neutral measure $\mathbb{P}^*$,

$$\begin{cases} dS_t/S_{t^-} = (r_t - d_t - \mu_t)dt + \eta_t dB_t^1 + dJ_t \\ d\eta_t = \lambda(a_t - \eta_t)dt + \alpha_t dB_t^2, \end{cases} \tag{1}$$

where B^1 and B^2 are two Brownian motions with $d\left\langle B^1, B^2\right\rangle_t = \rho_t dt$. Here λ is a constant and $\rho_t, d_t, \alpha_t, a_t$ are deterministic functions of time. Finally, J represents a compound Poisson Process with stochastic intensity $\xi\left(t, S_t, \eta_t\right)$ so that $J_t = \sum_{n\geq 1}\left(e^{Y_n} - 1\right) 1_{\{T_n \leq t\}}$ (T_n are the jump arrival times of the underlying Poisson's Process of stochastic intensity ξ). The sequence $(Y_n)_{n\geq 0}$ is *iid* (independent of B_1 and B_2), $\mathbb{J}$ will denote the law on $\mathbb{R}$ of the size of the jumps of J (the *iid* sequence $(\exp Y_n - 1)_{n\geq 0}$), and μ is the process such that $J_\cdot - \int_0^\cdot \mu_s ds$ is a martingale (the Radon derivative of the compensator of the compound Poisson Process). More precisely, the two dimensional jump diffusion (S, η) is a Feller process with infinitesimal generator $\mathcal{A}$ defined by (if $x = (x_1, x_2)$)

$$\begin{aligned} \mathcal{A}\varphi(t,x) &= \partial_t\varphi(t,x) + \partial_x\varphi(t,x)\zeta(t,x) + \frac{1}{2}tr\left[\partial_{x,x}\varphi(t,x)\Sigma(t,x)\Sigma^\top(t,x)\right] \\ &\quad + \xi(t,x)\int_{\mathbb{R}}\left[\varphi(t,x_1+u,x_2) - \varphi(t,x_1,x_2)\right]d\mathbb{J}(u), \end{aligned} \tag{2}$$

where ζ is the drift vector and Σ the volatility matrix of the diffusion. Finally, $(\mathcal{F}_t)_{t\geq 0}$ is the natural augmentation of the filtration generated by the two dimensional jump diffusion.

As it is well known, the above model is arbitrage-free but is not complete: model calibration will then be used to select a risk-neutral measure, according to the general theory[4]. We point out that the model defined by Eq. (1) does not belong to the affine class, in the sense of Duffie *et al.* (2000).

Our choice is motivated by several facts. Empirical literature (Jones (2003) and Medvedev and Scaillet (2004), among others) shows that simple affine models must be rejected in favor of more general processes. In fact, our model belongs to the so called "linear-quadratic" class (Piazzesi (2003), Peng and Scaillet (2004)), which includes the affine as a special case. In addition, SVJD linear-quadratic models can be easily generalized to include quanto and cross-currency features as well as the effect of stochastic interest rates with possibly stochastic volatility (for the purpose of hybrid derivatives modelling) while the same does not hold, in general, in affine models (Galluccio

[3] For the sake of simplicity, we will thoroughout assume that the divident process d_t is deterministic and that relative dividends are payed continuously in time. Note also that in our formulation no restriction must be imposed on the r.v. $(Y_n)_{n\geq 0}$ to ensure that S_t stays positive.

[4] We again refer to the textbook by Björk (1998) for a rigorous treatment of the link between model calibration and selection of a single risk-neutral measure in incomplete markets.

and Le Cam (2006)).[5] For avoidance of any doubt, we remark that our calibration algorithm (Sections 4 and 5) equally applies to affine models since model parametrization is essentially the same in both settings.

In the applications, it is useful to recast all equations in a more convenient form by introducing the auxiliary vector diffusion process $Z_t := (X_t = \ln(S_t), Y_t = \eta_t)$ and the jump process $N_t := \sum_{n\geq 1} Y_n 1_{\{T_n \leq t\}}$ (we denote by $\mathbb{G}$ the law of Y_n). In the new setting, the system reads as:

$$\begin{cases} dX_t = \left(r_t - Y_t^2/2 - d_t - \mu_t\right) dt + Y_t dW_t^1 + dN_t \\ dY_t = \lambda\left(a_t - Y_t\right) dt + \alpha_t \left(\rho_t dW_t^1 + \kappa_t dW_t^2\right). \end{cases} \tag{3}$$

where $\left(W^1, W^2\right)$ is a now a two dimensional vector of independent Wiener processes with $W^1 = B^1$ and $\kappa_t = \sqrt{1-\rho_t^2}$. The model can be easily handled analytically. In fact, if we assume that the intensity process takes the quadratic form $\xi(t,z) := \xi_t^0 + \xi_t^1 x + \xi_t^2 y + \xi_t^3 y^2$, the jump diffusion vector-valued process is a semimartingale associated to a triplet of characteristics that are affine-quadratic functions of the state variables, as in Peng and Scaillet (2004).

The presence of jumps in the dynamics is well supported by historical time series analysis, as above mentioned. Bates (1996) and (2000) suggests that jumps are needed in addition to stochastic volatility to allow matching both long and short-maturity smiles within a single model. Strong evidence in support of this claim is given in the next Section.

2.2 Option Pricing

In our setting, pricing of vanilla European options can be done in quasi-closed form. For $u \in \mathbb{C}$, we introduce the "discounted conditional characteristic function" $\psi(u, z; t, T)$ in the risk-neutral expectation, defined by (process Z is Markov in filtration $\mathbb{F}$, hence the existence of the deterministic function of $z = (x, y)$, ψ)

$$\psi(u, Z_t; t, T) := \mathbb{E}^* \left[\exp\left(-\int_t^T R(s, Z_s) ds \right) e^{uX_T} \middle| \mathcal{F}_t \right], \tag{4}$$

where $R(t, X_t) = r_t$ is the spot interest rate. The Laplace transform of the law of Y_1 is given by $\mathcal{L}(x) = \int e^{ux} d\mathbb{G}(u)$ (under the usual conditions of existence and convergence) and we introduce the auxiliary functions $\Phi_t^i(x) = \xi_t^i \left(\mathcal{L}(x) - 1\right)$. The following result provides the "discounted conditional characteristic function", that is the exponential of a quadratic function of the state variables (time dependency is omitted to simplify the notation):

Proposition 1 *There exist four functions $\gamma(t,T), \beta_1(t,T), \beta_2(t,T)$, and $\delta(t,T)$ such that ψ can be represented as*

$$\psi(u, Z_t; t, T) = \exp\left(\gamma(t,T) + \beta_1(t,T) \cdot X_t + +\beta_2(t,T) \cdot Y_t + \delta(t,T) Y_t^2\right).$$

[5] This is due to the fact that the presence of quanto effects or stochastic interest rates generates non-linear terms in the drift of the process when dynamics is expressed in the domestic risk-neutral measure.

Moreover the four functions satisfy the following system of ODE's:

$$\begin{cases} \frac{\partial \beta_1}{\partial t} & = & -\Phi^1(\beta_1), \\ \frac{\partial \beta_2}{\partial t} & = & -\Phi^2(\beta_1) - 2\lambda a \delta - \left(\rho\alpha\beta_1 + 2\alpha^2\delta\right)\beta_2, \\ \frac{\partial \delta}{\partial t} & = & -\Phi^4(\beta_1) + \frac{1}{2}\left(\beta_1 - \beta_1^2\right) + 2\left(\lambda - \alpha\rho\beta_1\right)\delta - 2\alpha^2\delta^2 \\ \frac{\partial \gamma}{\partial t} & = & -\Phi^0(\beta_1) + (d + \mu - r)\beta_1 - \lambda a \beta_2 - \alpha^2\left(\delta + \frac{1}{2}\beta_2^2\right) \end{cases} \tag{5}$$

with conditions $\beta_1(T,T) = u$, *and* $\beta_2(T,T) = \delta(T,T) = \gamma(T,T) = 0$, *since* $\psi(u, Z_T; T, T) = e^{uX_T}$.

Proof. *See Appendix A.* ■

To simplify the problem we will assume, from now on, that the stochastic intensity of the compound Poisson process is deterministic, i.e. $\xi_t^0 \neq 0, \xi_t^1 = \xi_t^2 = \xi_t^3 = 0$. In this case, the expression of the compensator is $\mu_t = \xi_t^0 \mathbb{E}^*\left[e^Y - 1\right] = \Phi_t^0(1)$. To ease notation, we shall replace ξ_t^0 by ξ_t. The above system contains non-linear, time inhomogeneous second order Riccati equations and cannot be solved in closed form, in general. However, since only a finite set of options at different times to expiry is actually quoted in the market, we can restrict ourselves to handle piecewise constant functions. Let $(T_1, \cdots, T_N)$ be the set of expiry times associated to the quoted vanilla options. Accordingly, if $\theta(t)$ is a generic time-dependent coefficient in the system, we will assume that $\theta(t) = \theta_i$, if $t \in [T_{i-1}, T_i)$, $i = 2, \cdots, N$. With this specification on every interval $[T_{i-1}, T_i)$ all Riccati equations are defined in terms of constant coefficients and then solvable. On every subinterval, terminal conditions are $\beta(T_i) = (u_i^1, u_i^2)$, $\delta(T_i) = u_i^3, \gamma(T_i) = u_i^4$. We then arrive at the following result, where $\Psi_i(t,x) = \left[1 - \exp\left(x(T^i - t)\right)\right]/x$.

Proposition 2 *Assume that* $\alpha_t, a_t, k_t, \xi_t^0$ *are piecewise constant on the intervals* $[T_{i-1}, T_i)$, $i = 2, \cdots, N$. *The solution of the system of ODE's is given, on each* $[T_{i-1}, T_i)$, *by*

$$\begin{aligned} \beta_1(t) &= u_i^1,\ \beta_2(t) = M(t)\left(u_i^2 - K(t)\right), \\ \delta(t) &= \frac{1}{\alpha_i^2}\left(-(B_i + \Gamma_i) - \frac{2\Gamma_i C_i}{e^{4\Gamma_i(T_i - t)} - C_i}\right), \\ \gamma(t) &= u_i^4 - \left[\left(d_i + \Phi_t^0(1)\right)u_i^1 - \Phi_t^0(u_i^1) - (B_i + \Gamma_i)\right](T_i - t) \\ &\qquad -\frac{1}{2}\ln\frac{1 - C_i e^{-4\Gamma_i(T_i - t)}}{1 - C_i} + \int_t^T\left(\lambda a_i + \frac{1}{2}\alpha_i^2\beta_2(s)\right)\beta_2(s)ds, \end{aligned}$$

with the notations and functions :

$$\begin{aligned} A_i &= u_i^1(1 - u_i^1)/4; B_i = \left(\alpha\rho u_i^1 - \lambda\right)/2; \Gamma_i^2 = B_i^2 + \alpha_i^2 A_i; C_i = \frac{\alpha^2 u_4^i + B_i + \Gamma_i}{\alpha^2 u_4^i + B_i - \Gamma_i^i}; \\ p_i &= -2\lambda a_i; z_i = -2\Gamma_i C_i/\alpha_i^2; y_i = -B_i + \Gamma_i/\alpha_i^2; \pi_i = 2(B_i + \Gamma_i) - u_1\rho_1\alpha_i; \\ M(t) &= \frac{(1 - C_i)\,e^{-\pi_i(T_i - t)}}{1 - C_i e^{-4\Gamma_i(T_i - t)}}; K(t) = \frac{(p_i y_i C_i - p_i z_i)\,\Psi_i(t, \pi^i - 4\Gamma_i) - p_i y_i \Psi_i(t, \pi^i)}{1 - C_i}. \end{aligned}$$

Proof. *See Appendix B.* ■

We remark that jumps only appear in the expression of $\gamma(t)$ through the Laplace transform $\mathcal{L}(u_i^1)$. With a proper choice of the distribution of the r.v. Y, the transform can be analytically computed. If $Y \sim \mathcal{N}(q, v^2)$ is a Gaussian random variable then $\mathcal{L}(x) = \exp\left(qx + x^2v^2/2\right)$. This choice provides a simple and intuitive jumps parametrization and, as we show below, it offers great flexibility in the calibration process.

Our goal is the evaluation of a vanilla call option expiring at T and struck at K written on S, whose arbitrage price at time t is $Call_t(S_t, K, t, T) = \mathbb{E}^*\left\{\exp\left(-\int_t^T r_s ds\right)(S_T - K)^+|\mathcal{F}_t\right\}$. To this aim, we use the auxiliary function (we use Markov property again) :

$$G(y, \varsigma, \varphi, Z_t; t, T) = \mathbb{E}^*\left\{e^{-\int_t^T r_s ds + \varsigma \cdot X_T}\chi_{\{\varphi \cdot X_T \leq y\}}\middle|\mathcal{F}_t\right\}, \tag{6}$$

and a number of well-known results on Fourier transforms for option pricing. In fact, as the following Proposition shows, $G(y, \varsigma, \nu, z, t, T)$ can be determined from the knowledge of $\psi(u, z; t, T)$ and the pricing problem is then solved.

Proposition 3 *The price of the call option is given by* $Call_t(S_t, K, t, T) = G_1 - KG_2$, *with*

$$\begin{aligned} G_1 &= G(-\ln K, \zeta_1, \nu_1, Z_t, t, T),\ \zeta_1 = (1,0,0), \nu_1 = (-1,0,0), Z_t = (\ln S_t, \eta_t, r_t), \\ G_2 &= G(-\ln K, \zeta_2, \nu_2, Z_t, t, T),\ \zeta_1 = (0,0,0), \nu_1 = (-1,0,0), Z_t = (\ln S_t, \eta_t, r_t) \end{aligned}$$

and

$$G(y, \varsigma, \nu, Z_t, t, T) = \frac{1}{2}\psi(\varsigma, Z_t, t, T) - \frac{1}{\pi}\int_{(R^d)^+}\frac{1}{k}Im\left\{e^{-iky}\psi\left(\varsigma + ik\nu, Z_t, t, T\right)\right\}dk \tag{7}$$

Proof. *Peng and Scaillet (2004).* ■

3 Model asymptotics and regimes switching

In this section we study the relationship between the dynamics Eq.(1) and the associated shape of the volatility surface. The goal is to provide evidence about the different roles played by jumps and by stochastic volatility in explaining the observed smile in different portions of the time to expiry axis. This result is instrumental in understanding the calibration methodology that will be later introduced.

The analysis of the moments of the asset distribution and their link with the shape of the smile has been already addressed in the literature. In particular, Backus *et al.* (1997) and (in a similar context) Zhang and Xiang (2005) show that if the smile can be parametrized through a quadratic polynomial in the "modified moneyness" $m = \ln(F/K)/\left(\Sigma_{atm}\sqrt{T-t}\right) + \Sigma_{atm}\sqrt{T-t}/2$, where F is the underlying's forward, Σ_{atm} is the at the money (ATM) BS volatility and K is the strike then, approximately, the BS implied volatility at varying m reads as

$$\sigma(m, \tau) \simeq \Sigma_{atm}\sqrt{\tau}\left(1 - \frac{\zeta_1(\tau)}{3!}m - \frac{\zeta_2(\tau)}{4!}\left(1 - m^2\right)\right), \tag{8}$$

where $\zeta_1(t)$ and $\zeta_2(t)$ are the skewness and the kurtosis of the logarithm of the underlying process, and $\tau = T - t$ is the time to expiry. This result descends from a Gram-Charlier expansion of the law of the log-asset price and holds for small values of Σ_{atm}. Formula (8) shows the tight link existing between shape of the smile and moments of the underlying asset process. In particular, when skewness and kurtosis are zero, the smile is flat at Σ_{atm} (as in the BS model). In addition, skewness (through the linear term in m) and kurtosis (through the quadratic term in m) act by respectively tilting and bending the smile.

For the sake of simplicity, we will assume in this section that dividends d_t vanish and that all model coefficients are constant. All conclusions hold (at a qualitative level) in the general setup. Even in this simplified scenario the analytical expression of the characteristic function $\Phi_t(x) := \mathbb{E}_t\left(\exp ix \ln S_T\right)$ of the log-asset price (and *a fortiori*, that of the associated cumulants) is quite involved. For this reason, we analyze separately the impact of jumps and of stochastic volatility.

Pure jump process. In this case, Eq.(1) reduces to the Merton (1973) jump-diffusion model. The first four cumulants of $X_t := \ln(S_t)$ in this setup are well known (appendix C). Skewness $\left(\zeta_1(t) = \Pi_3 \Pi_2^{-3/2}\right)$ and kurtosis $\left(\zeta_2(t) = \Pi_4 \Pi_2^{-2}\right)$ (expression of functions Π_i are recalled in the appendix), are given by :

$$\zeta_1(t) = \frac{1}{\sqrt{t}} \frac{\xi q \left(q^2 + 3v^2\right)}{\left[\eta^2 + \xi\left(q^2 + v^2\right)\right]^{3/2}}, \qquad \zeta_2(t) = \frac{1}{t} \frac{\xi \left(q^4 + 6q^2v^2 + 3v^4\right)}{\left[\eta^2 + \xi\left(q^2 + v^2\right)\right]^2}. \tag{9}$$

These expressions, in conjunction with Eq.(8) , show that the impact of jumps on the volatility smile is restricted at very short times since both skewness and kurtosis are inversely proportional to time and diverge in approaching 0. In other words, high levels of skew and convexity in the smile can be naturally explained by the presence of jumps in the short term, while the presence of (log-normal) jumps is negligible in the medium/long term.

Pure SV model. In this case, Eq.(1) reduces to the Stein-Stein stochastic volatility model and the computation of the moments is a harder task. In this paper, we concentrate on affine-quadratic models but the analytical approach we introduce to evaluate moments asymptotics applies in principle to any SV model.

Proposition 4 *In a Stein-Stein model, short and long term asymptotics of skewness $\zeta_1(t)$ and excess kurtosis $\zeta_2(t)$ of $\ln S(t)$ are given by*

$$\begin{aligned} \lim_{t\to 0} \frac{\zeta_1(t)}{\sqrt{t}} &= \frac{3\alpha\rho}{\eta_0}, \qquad \lim_{t\to 0} \frac{\zeta_2(t)}{t} = q\eta_0^{-2} + 6\lambda a \eta_0^{-1} - 6\lambda, \\ \lim_{t\to\infty} \zeta_1(t)\sqrt{t} &= \frac{a_1}{n^{3/2}}, \qquad \lim_{t\to\infty} \zeta_2(t) t = \frac{6c_0 - 3(2m+p)n/\lambda}{n^2}, \end{aligned} \tag{10}$$

respectively, with c_0, a_1 defined as in Appendix C and :

$$m = 2a(\eta_0 - a), n = \frac{\alpha^2 + 2\lambda a^2}{2\lambda}, p = \eta_0^2 - \frac{\alpha^2 + 2\lambda a^2}{2\lambda} - 2a(\eta_0 - a), q = 7\alpha^2 + 8\alpha^2\rho^2 + 6\lambda p + 3\lambda m.$$

Proof. *Appendix C.* ■

Two things are worth noticing. First, both skewness and kurtosis now vanish when maturity approaches 0. Second, smile skew and convexity decrease as $1/\sqrt{t}$ and $1/t$ as function of time, respectively, when t tends to infinity. Therefore, while long term smile asymptotics in Stein-Stein and Merton models are the same (apart from multiplicative factors), short term asymptotics are completely different in the two cases. In particular, Eqs. (10) imply that the term structure of both skewness and kurtosis reach a maximum after a finite time $t_{\max}$ in the Stein-Stein model. Interestingly, by proceeding as in Appendix C, it is possible to show that this property is shared by affine models with mean reverting volatility (e.g. Heston model).

Equations (10) indicate the location of the maximum of skewness and kurtosis at $t_{\max}$ is inversely proportional to the volatility mean reversion λ. Since the larger $t_{\max}$ the longer smile skew and convexity are preserved in time, one can use λ to fine tune the degree at which the smile switches from a high convex shape (at short times) to a low convex shape (at long times) as actually observed in the market. This fundamental property is missing in models with pure jumps or in SV models without mean-reverting volatility (such as SABR, Hagan *et al.* (2002)) and so they tend to generate smiles at different maturities that move together too rigidly as a function of the model parameters. More precisely, due to the lognormality of the volatility process, the volvol parameter in the SABR model must rapidly decrease in the time to maturity direction to allow calibrating all smiles. For this reason, these models are unable of fitting the whole volatility surface with time-independent parameters. To gain further insight, we compare the term structure of skewness and kurtosis of the Merton and Stein-Stein model in Fig.1a and Fig.1b for typical values of the parameters (see Eqs. (26)and (30) in Appendix C). No matter how parameters are chosen, in a pure jump model both skewness and kurtosis tend to converge to zero much faster than in a SV model with mean-reverting volatility in general. Figures show that after just one year the jumps-induced skewness and excess kurtosis are totally negligible so that jumps cannot practically generate any smile effect beyond that time. Moreover, in the limit of vanishing time to expiry skewness and kurtosis behave very differently in the two cases.

These theoretical results can be empirically tested thanks to the link between shape of the smile and moments of the underlying process provided by Eq. (8). To this aim, we study the term structure of "butterfly spread" prices observed in the market on a generic trading day. A butterfly spread option with expiry T is a combined position in three call options and, to fix the ideas, can be associated to the quantity $H = \Sigma^{BS}(K^{atm} - \Delta) - 2\Sigma^{BS}(K^{atm}) + \Sigma^{BS}(K^{atm} + \Delta)$ where $\Sigma^{BS}(K)$ is the BS implied volatility at K. Butterfly spreads provide the simplest trading strategy to take a position in the smile's convexity. This is due to the fact that (apart from a multiplicative factor) H is the second derivative of the smile (here thought of as a function of K) taken at K^{atm}. Thus, the higher H the larger the convexity and viceversa. In Fig.2a we show a typical market butterfly as a function of time to expiry. Thanks to Eq. (8) the presence of a butterfly directly translates into that a positive excess kurtosis. We notice two things. First, smile's convexity is positive and quite large in the short term (this is consistent with the presence of jumps but inconsistent with the predictions of a pure SV model, from Eq (10) anf Fig 1b). Second, smile convexity nicely decreases to 0 in the long term but not too quickly since it is still present after 5 years (this is consistent with the presence of SV but is inconsistent with the predictions of a pure jump model, from Eq (10) and Fig 1b). This shows that no matter how parameters are chosen, it is almost impossible to make a pure SV or jumps model consistent with the observed shape of the smile over a time interval of a few years since jumps tend to work well at short maturities and SV at long maturities only. This further reinforces the view that option markets are consistent with the simultaneous presence of both jumps and stochastic volatility in the asset dynamics.

These findings can be also interpreted from a different perspective. From a trading point of view, short-term and long-term smiles have a very different origin. Short term convexity is mainly associated to investors risk aversion to unexpected economic and socio-political events that might induce a jump in the asset price. On the other side, long-term convexity is usually driven by the law of offer/demand induced by large investors, financial institutions and hedge/pension funds buying and selling in and out of the money options for liability management or investment purpose. Traders refer to these two regimes as "Gamma" and "Vega" trading, since the option Gamma (resp. Vega) risk is predominant at short (resp. long) maturities and in presence of large (resp. small) asset variations. These considerations indicate that the market smile is implicitly pricing the risk of large fluctuations (indeed, jumps) in the asset dynamics in the short term and of that of unpredictable (indeed, stochastic) asset volatility in the long end. The threshold between the two regimes will be denoted by T^*. This regime "switching" is therefore an intrinsic market characteristic and plays a fundamental role in our calibration approach.

4 SVJD models: the calibration problem

As above mentioned, model calibration consists of solving a multi-dimensional reverse engineering problem. As discussed by many authors it is impossible, in general, to determine a set of parameters such that market prices are exactly reproduced by a given model.[6] Throughout this paper, by "model calibration" we will then refer to a numerical algorithm such that :

- The difference between market and model option prices is within the bid/ask spread
- The calibrated solution is statistically robust, i.e. weakly sensitive to the input option prices.[7]

Cont and Tankov (2004) show that a *single* smile calibration can be achieved by solving the following non-linear optimization problem (N_S = n. of strikes, $N_E = 1$ = n. of option expires):

$$\{\pi_i\}^* = \arg\min_{\{\pi_i\}} \sum_{j=1}^{N_S} \sum_{k=1}^{N_E} w_{jk} \left| \Sigma\left(T_k, K_j^{(k)}; \{\pi_i\}\right) - \Sigma^{BS}\left(T_k, K_j^{(k)}\right) \right|^2 + \psi F(\{\pi_i\}), \tag{11}$$

where in general $\{\pi_i\}$ is a set of free model parameters, Σ is the model-implied Black-Scholes volatility, Σ^{BS} is the market-implied Black-Scholes volatility, ψ and w_{jk} are weighting constants, $K_j^{(k)}$ is the j-th. strike for options expiring at T_k and $F(\{\pi_i\})$ is a convex regularization functional.

[6] Even from a pure financial point of view this is impossible to achieve. In fact, market imperfections and inefficiencies do not allow to identify option prices exactly (due to the bid/ask spread).

[7] We point out that robustness, in this context, is a fundamental property since otherwise one would be obliged to frequently readjust the replicating portfolio.

The above minimization problem provides in theory a set of "optimal" free parameters $\{\pi_i\}^*$. In our case, however, achieving calibration is much harder since we aim at making *a single model consistent with the whole volatility surface* not just a single smile curve.

In any time interval $[T_{i-1}, T_i)$ between two consecutive option expires our model is specified by a set of 7 independent coefficients: the volatility mean reversion level a_t, the volatility of volatility (volvol) α_t, the constant volatility mean reversion rate λ, the asset-volatility correlation ρ_t, the stochastic jumps intensity ξ_t^0, the jumps average q_t and, finally, the jumps variance v_t^2, with $t \in [T_{i-1}, T_i)$.

As it is well known, any attempt to perform a global calibration on this 7-dimensional manifold is doomed to failure.[8] Understanding the impact of each single parameter on the shape of the smile is instrumental to the problem's solution. We then start by discussing the role of the different parameters when trying to minimize the above functional.

In a pure stochastic volatility framework, the role of coefficients a_t, α_t and ρ_t is indeed well established (Hagan *et al.* (2002)). At the leading order in the volvol α_t, the at-the-money (ATM) volatility is completely specified by a_t. Thus, a_t mainly affects the global level of the smile but has little impact on its overall shape. The instantaneous Equity-volatility correlation ρ_t affects the asymmetry of the smile (or "skew") around the ATM point. Finally, the volatility of volatility α_t rules smile convexity: the higher α_t the more convex the smile and viceversa. In addition, the volvol coefficient has an impact on the process variance (like a_t). Thus, the global level of the smile is affected, too (Fig 2b).[9] Each parameter in the set $\{a, \alpha, \rho\}$ plays a special role in explaining possible smile movements; in other words they are not "degenerate".

A key role is played by the mean reversion parameter λ which we assume constant. A mean reverting Ornstein-Uhlenbeck process converges to its ergodic measure at a speed linked to its "characteristic time" $\tau = 1/\lambda$. Thus, by adjusting the volatility mean reversion λ one can "fine tune" the rate of convergence to the ergodic measure or, equivalently, the rate decrease of the smile convexity at increasing maturities, as previously discussed. In models where $\lambda = 0$, like the one proposed by Hagan *et al.* (2002), it is necessary to artificially impose a decreasing term structure of the volvol to ensure market consistency to option prices. We also remind that λ cannot be statistically inferred from historical time series since it is not a measure change invariant.

When only jumps are present, the picture becomes much more complex. In fact, although jump parameters $\{\xi, q, v\}$ play altogether a role similar to $\{a, \alpha, \rho\}$ in explaining possible smile deformations, their influence on the smile shape cannot be as nicely identified as before; parameters now play "mixed" roles and are thus "degenerate". Merton model provides the simplest example

[8] The causes for this are: *i)* the non-linear optimization problem is not strictly convex and, *ii)* some parameters are degenerate. This implies respectively that: *i)* the objective function has many local minima and, *ii)* it is almost flat in the maximum gradient direction so that both convergence and robustness of the algorithm are not guaranteed (see Cont and Tankov (2004)).

[9] As shown in Fig. 2, by increasing the level of α the ATM volatility increases, as one would intuitively expect. On the opposite, in affine models (Heston) the ATM volatility is inversely proportional to the volvol coefficient. This unrealistic behaviour makes affine models less appealing from a trading perspective than affine-quadratic ones.

in this respect. In this case, the cumulants of $Z_t := \ln(S_t)$ are given by Eq.(9). In a BS setting we have $\xi = 0$ and only the first two cumulants (mean and variance) are different from zero; the implied smile is flat and equal to η. When $\xi \neq 0$ second, third and fourth cumulant play altogether a decisive role in moving the implied volatility away from the BS level. In fact, Eq. (9) and Eq. (8) show that smile deformations around the BS level can be attributed to either the stochastic intensity ξ, the jumps average q or the jumps standard deviation v (for example q^2 and v^2 play an almost identical role in the smile). Therefore, a classification of jump parameters according to their impact on the smile is impossible since different triplets $\{\xi, q, v\}$ can generate almost identical smile shapes. Consequently, the inverse problem (i.e. determining a unique triplet from a given smile or a set of smiles) is in general an ill-defined problem (as already pointed out in Cont and Tankov (2004)). In other words the triplet $\{\xi, q, v\}$ is a "degenerate" set. This identification problem is present in several global minimization algorithm proposed in the literature (Andersen and Andreasen (2000), Bakshi et al. (1997) and Detlefsen (2005) among others).[10]

When both jumps and stochastic volatility are present, the set of (seven) parameters is -roughly speaking- "twice degenerate". To remove these degeneracies from the optimization algorithm on this 7-dimensional manifold, we suggest the following method. We add to problem (11) a number of additional constraints and carry out a regularization of the least-squares optimization[11] :

1. First, (condition C1) the influence of the jumps on the dynamics is strictly restricted at short times while stochastic volatility mainly acts at medium and long expiries to reflect the transition between "Gamma" and "Vega" regimes, as previously discussed.[12] We then split the calibration problem in two consecutive steps. Initially, until a given date T^*, the term structure of diffusion coefficients is kept at a (trial) constant level $\{a_0, \alpha_0, \rho_0\}$, while calibration is performed by only adjusting the jump coefficients. This procedure is motivated by the fact that stochastic volatility has no impact on the smile at $t \leq T^*$ (Fig.1). Once jumps calibration has been achieved, we calibrate the remaining smiles at $t > T^*$ by adjusting $\{a, \alpha, \rho\}$ while the jump parameters are "frozen" to their previously calibrated levels. In this way jumps and stochastic volatility are not "mixed up" in the optimization procedure and several degeneracies

[10] We also remind that calibration methods based on global minimization algorithms might be "trapped" in local minima. Detlefsen (2005) proposes a simulated annealing (Metropolis) algorithm to avoid this problem. Unfortunately, the typical convergence time of Metropolis algorithms are of the order of hours and, as a consequence, they are of no use in practice.

[11] This approach retains (at a qualitative level) some of the interesting features contained in Cont and Tankov's method, namely the regularization of least-squares optimization through addition of a number of constraints to the problem

[12] Central limit theorem implies that the impact of log-normal jumps on the dynamics asymptotically vanishes at increasing times. However, by enforcing jumps to be non-zero only for short periods of time has two main advantages. On one side, it provides a better way to avoid parameters degeneracies when stochastic volatility is also present. On the other hand, it helps implementing PDE's for option pricing, as extensively discussed in Galluccio and Le Cam (2006) since one can avoid numerically solving a complex partial integro-differential equation for the majority of the time axis.

are eliminated.

2. Second, (condition C2) we impose that the switch between the two regimes at $t \leq T^*$ and $t > T^*$ is smooth. A smooth "transition" between the two regimes (Gamma and Vega) must be imposed in order to guarantee a robust risk-management. To this aim, jumps will be gradually "switched off" to avoid unreasonable discontinuities across the two regimes. We then assume that the stochastic intensity $\xi(t)$ is a continuous (possibly differentiable) strictly decreasing function converging to 0 at $t = T^*$, i.e. *i)* $\xi(T) \in \mathcal{C}^0$, *ii)* $\xi(T^*) = 0$, *iii)* $\xi(t) > \xi(t')$ for $t < t'$. Also, the initial set $\{a_0, \alpha_0, \rho_0\}$ is selected to minimize the difference in parameters value between $t \leq T^*$ and $t > T^*$(Section 5).

3. Third, (condition C3) the volatility mean reversion λ is chosen to make the calibrated set $\{a, \alpha, \rho\}$ is as time-homogeneous as possible. From a statistical point of view, such models are more robust and realistic than those models where all parameters are heavily time-dependent. In addition, when parameters are constant the dynamics of the volatility surface is closer to stationarity and then consistent with empirical observations. This is also beneficial on the risk-management side (Rebonato (2000))

4. Fourth (condition C4), in calibrating jumps at $t \leq T^*$ we do not attempt a global minimization over the set $\{\xi, q, v\}$ to avoid the degeneracy issue above mentioned.

As a side remark, we recall that the jumps intensity is defined under the neutral risk probability measure in our setting. Since this measure reflects market prices anticipations, the estimation of the intensity cannot be performed statistically. This choice ensures in fact a large degree of flexibility in selecting the the risk-neutral intensity. To fix the ideas, assume that in a Merton model (associated to the SDE $dS_t/S_{t-} = \mu dt + \sigma dW_t + j dN_t$ with jumps of constant size), the statistically estimated jumps intensity is constant and equal to Λ. The market being incomplete, it is easy to prove that for any function λ_t, there exists a risk neutral probability measure $\mathbb{P}^*$ such that the intensity of the process is λ_t under $\mathbb{P}^*$. In fact, if $M_t = N_t - \Lambda t$ is the compensated martingale associated to N under the historical probability, and if we define the new probability measure by the following Radon-Nykodim derivative

$$\left.\frac{d\mathbb{P}^*}{d\mathbb{P}^{hist}}\right|_{\mathcal{G}_T} = \xi(h.W)_T \xi\left(\lambda/\Lambda.M\right)_T$$

with $\sigma h_t = \mu - r + j\lambda_t/\Lambda$ (ξ denotes the Doléans Dade exponential) then, from Girsanov theorem, $dS_t/S_{t-} = rdt + \sigma dW_t^* + j dM_t^*$ is a martingale under the neutral risk probability $\mathbb{P}^*$. Here M^* is the compensated martingale associated to a Poisson Process of intensity λ[13].

[13]Recall that $dW_t^* = dW_t - h_t dt$ is a Brownian Motion and $dM_t^* = dN_t - \lambda_t dt$ is the compensated martingale associated to the Poisson's Process N whose intensity is $\Lambda\lambda/\Lambda = \lambda$.

5 Method implementation

In our empirical study we consider options up to 5 years time to expiry since the market is in general illiquid at longer maturities. We consider data from the EuroStoxx 50 equity index, whose ATM volatility matrix is given in Fig 8a[14]. The tenor is $\mathcal{T} = \{1M, 2M, 3M, 6M, 1Y, 2Y, 3Y, 5Y\}$. Similar studies conducted on other indices (S&P 500, FTSE 100, DAX and CAC40) provide qualitatively similar results to the ones presented here.

5.1 Jumps calibration

As anticipated, Cont and Tankov (2004) regularization procedure for generic Lévy processes cannot be directly applied to our problem for two reasons. First, we aim at making the model consistent with the whole volatility surface and in doing so we need to calibrate a *term structure* of model coefficients. Second, the process defined by Eq. (1) is not a Lévy process since its increments are independent but not stationary.

Our method aims at calibrating the whole set of smiles up to (and *including*) T^* by adjusting the jumps average q and standard deviation v once a suitable parametric form for $\xi(t)$ has been assigned. Meanwhile, coefficients $\{a_0, \alpha_0, \rho_0\}$ are kept fixed at a trial (initial) level.

Selection of the expiry threshold T^*. It can be done empirically by observing (Fig.1,2) that, typically, the SV component generates no excess kurtosis (hence, no smile convexity) on smiles of less than 3-4 months from expiry. Selection of T^* is thus important to avoid too much overlap between jumps and stochastic volatility. In our case we fix $T^* = 3M = 0.25$ to get the best results.

Choice of $\xi(t)$. Denote by T^{*+} the date correponding to the first smile after T^*, in our example $T^{*+} = 6M$. Although different choices of $\xi(t)$ provide qualitatively similar results, we find convenient to define $\xi(t) = \{\omega\,(T^{*+}/t - 1)\}^{\delta}$ for $t \in (0, T^{*+}]$ (with $\omega, \delta > 0$), and $\xi(t) = 0$ otherwise . This simple choice satisfies condition C2 and has just two free parameters: ω (affecting the intensity level) and δ (affecting the rate of decrease of the intensity) to calibrate all smiles up to T^*. In practice, $\xi(t)$ must be discretized with caglad[15] piecewise functions $\xi(t) = \xi(t_{i+1})$ for $t \in (t_i, t_{i+1}]$, so that

$$\xi(t) = \xi(t, \omega, \delta) = \begin{cases} \xi_0 = \left\{\omega\left(\frac{6M}{1M} - 1\right)\right\}^{\delta} & \text{for} \quad t \in (0, 1M] \\ \xi_1 = \left\{\omega\left(\frac{6M}{2M} - 1\right)\right\}^{\delta} & \text{for} \quad t \in (1M, 2M] \\ \xi_2 = \left\{\omega\left(\frac{6M}{3M} - 1\right)\right\}^{\delta} & \text{for} \quad t \in (2M, 3M = T^*] \\ \xi_3 = \left\{\omega\left(\frac{6M}{6M} - 1\right)\right\}^{\delta} = 0 & \text{for} \quad t \in (T^*, 6M = T^{*+}] \end{cases}$$

and $\xi(t) = 0$ for $t > T^{*+}$. The problem of the optimal choice of ω and δ is addressed next. We first notice that the speed of convergence of $\xi(t)$ to 0 as $t \to T^*$ has a major impact on the accuracy of

[14] The smooth surface has been obtained by using a BNP Paribas propietary arbitrage-free volatility interpolation algorithm that is capable of matching quoted market prices within their bid-ask spread. With no loss in generality, all smiles are cut-off beyond a point that corresponds to many standard deviations for the ATM strike. Alternative parametrizations have been tried, like the one proposed by Fengler (2005), but results are not significantly affected by this choice. Data are observed on Feb 2nd 2004.

[15] Caglad stands for "left-continuous with right limit".

the jumps calibration. Intuitively, when δ is assigned a large (resp. small) value, jumps intensity converges to zero quickly (resp. slowly). In the former case the model does not generate enough convexity in the smile at T^*. If one then decided to increase ω to compensate, the model would typically generate a too convex smile at the shortest expiry T_1. In the latter case the smile at T^* wuold be too convex and a simultaneous decrease of ω would yield a flat smile at the shortest expiry T_1. In both cases, a *simultaneous* calibration of all smiles with expiry less than T^* is impossible to achieve. These observations indicate that δ must be chosen (other parameters being given) by providing the optimal trade-off between the two opposite scenarios.

Before attempting any minimization on the set $\{q, v, \delta, \omega\}$, for a given set of stoch vol parameters $\mathcal{B} = \{a_0, \rho_0, \alpha_0\}$, we need to gain further insight into the role played by the different parameters.[16] We introduce the function

$$G_{\omega,\delta,\mathcal{B}}(q, v) := \sum_{j=1}^{N_S} \sum_{k=1}^{N_E} w_{jk} \left| \Sigma \left(T_k, K_j^{(k)}; q, v, \omega, \delta, \mathcal{B} \right) - \Sigma^{BS} \left(T_k, K_j^{(k)} \right) \right|^2 .$$

and study the dependence on $(\omega, \delta, \mathcal{B})$ of two objects: (i) the couple $(q^*(\omega, \delta, \mathcal{B}), v^*(\omega, \delta, \mathcal{B}))$ solution of the optimization problem

$$(q^*, v^*) = \arg\min_{(q,v)} G_{\omega,\delta,\mathcal{B}}(q(\omega, \delta, \mathcal{B}), v(\omega, \delta, \mathcal{B}))$$

and (ii) the associated minimum $G^*_{\omega,\delta,\mathcal{B}}(q^*(\omega, \delta, \mathcal{B}), v^*(\omega, \delta, \mathcal{B}))$. We plot $G_{\omega,\delta,\mathcal{B}}$ as a function of (q, v) for different values of ω, δ and α_0. Results not reported here show that changing a_0, ρ_0 has a negligible impact on $G_{\omega,\delta,\mathcal{B}}$ for $t \leq T^*$. Results are gathered in Fig.3,4,5,6 and in Tables 1,2,3,4. We notice that

1. For a given $(\omega, \delta, \mathcal{B})$, the function $G_{\omega,\delta,\mathcal{B}}(q, v)$ is strictly convex around a single minimum (q^*, v^*). Furthermore the convex domain extends to a relatively wide region in the (q, v) space where no other local minima exist. It follows that if $(\delta, \omega, \mathcal{B})$ have been previously chosen, the couple (q^*, v^*) can be found by standard convex optimization routines.

2. The minimum $G^*_{\omega,\delta,\mathcal{B}}$ is strongly dependent on δ and α_0 and weakly dependent on all other parameters. In particular, when α_0 takes too large values (typically beyond 50% for $T^* = 0.25$) the impact of the SV becomes comparable to that of jumps even at short maturities and, as a consequence, $G_{\omega,\delta}(q, v)$ tends to rapidly lose its convex shape because SV and jump parameters become mutually degenerate. The optimal choice of α_0 will be addressed in Section 5.2.

3. The convexity of the objective function is not very pronounced, in general. To increase the convexity (at the expense of calibration accuracy) one can use a Tichonov regularization by adding a convex functional to $G_{\omega,\delta,\mathcal{B}}$ (as in Cont and Tankov (2004)).

[16] Remark that with our parametrization the problem has been reduced from a minimization on the set $\{\xi_i, q_i, v_i\}$ containing $3 * n$ parameters (n is the n. of smiles with expiry less than T^*) to one on the set $\{q, v, \delta, \omega\}$ containing 4 parameters.

4. Results not reported here show that the convexity of $G_{\omega,\delta,\mathcal{B}}$ is generally lost if one assumes $\xi(t) = \xi$ (constant) and attempts estimating ξ jointly with q and v as a solution of the global least-squares problem $(q^*, v^*, \xi^*) = \arg\min_{q,v,\xi} G_{\mathcal{B}}(q, v, \xi)$. Addition of a third jump parameter makes the optimization problem fully degenerate.

Computation of the optimal δ^*. We propose to assign an initial value to the free set $\{\mathcal{B}, \omega\}$, and to determine δ^* (the first time one runs the algorithm) by solving the problem

$$\delta^*(\omega, \mathcal{B}) = \arg\min_{\delta} G_{\omega,\delta,\mathcal{B}}(q^*(\omega, \delta, \mathcal{B}), v^*(\omega, \delta, \mathcal{B})) \tag{12}$$

This method is viable if one can prove that *i)* the problem (12) is convex so that δ^* is well defined and, *ii)* δ^* does not depend on $\{\omega, \mathcal{B}\}$ (one needs to ensure that the optimal δ^* remains unaltered *after* calibration of these parameters). To this aim, in Fig 7a we show several plots of the function $\delta \mapsto G^* = G_{\omega,\delta,\mathcal{B}}(q^*(\omega, \delta, \mathcal{B}), v^*(\omega, \delta, \mathcal{B}))$, by varying the set $(a_0, \rho_0, \alpha_0, \omega)$. Each curve is convex with a single minimum. Hence δ^* is well defined.

In addition, fig 7a shows how $\delta^*(\omega, \mathcal{B})$ is affected by changes in the initial set $\{a_0, \rho_0, \alpha_0, \omega\}$. We first consider a typical set of parameters $\{7\%, -0.4, 10\%, -0.3\}$ as our base case scenario. All other curves in Fig.7a are obtained from the base case by applying a large shock in a single parameter among those in the set $\{a_0, \rho_0, \alpha_0, \omega\}$. Results can be summarized as follows, *i)* δ^* and G^* are not sensibly affected by a shock in ω and a_0 (Series 2 and 3), *ii)* a shock in ρ_0 affects slightly calibration accuracy (G^*) but has almost no impact on the optimal δ^* (Series 4), *iii)* a shock in α_0 affects the optimal δ^* but has almost no impact on the calibration accuracy (G^*). In addition, Fig 7b further investigates the dependency of δ^* on α_0. The picture shows that the functional relationship between δ^* and α_0 is linear only for small values of α_0.

In summary, the optimal exponent δ^* depends only on α_0 to a high degree of accuracy. In other words, once the initial volvol parameter α_0 has been set at inception, one can determine an optimal δ^* for any given set $\{a_0, \rho_0, \omega\}$ as a solution of(12) .

Computation of the optimal ω *and of* (q, v). For a given set $\mathcal{B}$, once δ^* has been determined, we can estimate coefficient ω. Above results indicate that the value of ω does not affect the calibration accuracy. In fact, a change in ω is reflected by a change in the optimal couple $(q^*(\omega, \delta^*, \mathcal{B}), v^*(\omega, \delta^*, \mathcal{B}))$, while the associated value of G^* stays essentially the same. We can then fix ω^* to a value such that the couple $(q^*(\omega^*, \delta^*), v^*(\omega^*, \delta^*))$ is as close as possible to a "prior" couple $\left(q^P, v^P\right)$ arbitrarily chosen. A viable way consists of estimating jumps average and standard deviation from historical data series and to assign $\left(q^P, v^P\right)$ accordingly. This choice has the advantage that the optimal solution (ω^*, q^*, v^*) guarantees that the market-implied model stays "close" (in the probability measure space) to the historically estimated one [17].

In conclusion, jumps calibration can be summarized as follows

[17] The request that couples $\left(q^P, v^P\right)$ and (q^*, v^*) are close is a well defined problem in probabilistic terms. It is indeed equivalent to enforce that the market-implied jumps probability distribution is as close as possible to the historical (objective) one. Girsanov theorem ensures that the jumps distribution is indeed invariant under changes of probability measure (in our case from the objective to the risk-neutral and viceversa).

1. Assign a trial set $\{a_0, \rho_0, \alpha_0, \omega\}$.

2. For a given value of δ, solve the minimization problems

$$\begin{aligned} (q^*(\omega), v^*(\omega)) &= \arg\min_{(q,v)} G_{\omega,\delta,\mathcal{B}}(q(\omega,\delta,\mathcal{B}), v(\omega,\delta,\mathcal{B})) \\ (\omega^*, q^*, v^*) &= \arg\min_{\omega} \left[(q^P - q^*(\omega))^2 + (v^P - v^*(\omega))^2\right]. \end{aligned} \tag{13}$$

3. Finally, the optimal δ^* is obtained by solving the above problems iteratively until the global minimum is reached, that is

$$\delta^* = \arg\min_{\delta} G_{\omega,\delta,\mathcal{B}}(q^*(\omega,\delta,\mathcal{B}), v^*(\omega,\delta,\mathcal{B})). \tag{14}$$

We recall that any value of ω in the interval $[-0.1, -1]$ provides very similar results on tests performed on long time series and on different equity indices.

5.2 Stochastic volatility calibration

In the last section we showed that it is possible to calibrate the jumps once an initial set $\{a_0, \rho_0, \alpha_0\}$ of SV parameters has been assigned. In particular, the choice of the initial volvol parameter α_0 was shown to have a significant impact on jumps calibration. The simplest way to decide how to fix *a priori* the triplet is to run a "pre-calibration". The good news is that once this study has been carried out as shown below, one can safely keep the initial set $\{a_0, \rho_0, \alpha_0\}$ fixed most of the time without readjusting it.

Choice of α_0. The approach we introduce is based on the observation that if $\{a_0, \rho_0, \alpha_0\}$ has been badly selected in the "Gamma" region (so that the process has accumulated too much variance and kurtosis before T^*) calibration of the remaining smiles for options expiring after T^* cannot be achieved. To better illustrate this, we run an empirical test where we consider three sets of parameters $\left\{a_0, \rho_0, \alpha_0^{(i)}\right\}$, with $i = 1, 2, 3$ corresponding to $\alpha_0^{(1)} = 10\%, \alpha_0^{(2)} = 30\%, \alpha_0^{(3)} = 50\%$. For each given set, the model is then calibrated to all smiles up to T^*. Finally, all smiles with expiry beyond T^* are generated. This test is aimed at measuring the terminal variance, skewness and kurtosis generated by the initial set $\{a_0, \rho_0, \alpha_0\}$ and by jumps in the time interval $[0, T^*]$. Table 5 gathers the results. Here we show the difference between market and model implied volatility for smiles at 1Y, 2Y, 3Y and 5Y induced by the calibration at the shorter maturities. As anticipated, when α_0 is assigned a too large value, all model-implied smiles at 6M expiry are inconsistent with the market. In fact, if $\alpha_0 = 50\%$ no matter how $\{a(t), \rho(t), \alpha(t)\}$ are selected hitting the market smile is impossible to achieve because the cumulative variance at T^* is too large. In theory, calibration could still be achievable by allowing $a(t)$ to take large negative values but this solution is not financially sound.

We can formally define, for a given a_0, a "critical" value $\widehat{\alpha}_0$ of the volvol coefficient as follows:

$$\widehat{\alpha}_0 = \sup\{\alpha_0 : \text{all smiles are calibrated within the bid/ask spread; } a(t) > 0, \alpha(t) > 0\}.$$

In other words, $\widehat{\alpha}_0$ is the maximum value of the volvol such that, other parameters being given, all smiles can be matched by means of a sequence $\{a_t, \rho_t, \alpha_t\}$ by simultaneously keeping both $a(t)$ and $\alpha(t)$ positive. In the next section we show that typically $\widehat{\alpha}_0$ is quite large and a full volatility surface calibration can indeed be achieved for any α_0 in the interval $(0, \widehat{\alpha}_0]$. We finally remark that the above picture is not significantly altered by ρ_0 once its sign has been properly assigned (smiles are usually negatively skewed implying ρ_0 should be negative). These two properties are extremely important since they indicate that $\{a_0, \rho_0, \alpha_0\}$ can be assigned with great flexibility without compromising the quality of the calibration.

Calibration of $\{a_t, \rho_t, \alpha_t\}$. We assume that $\{a_0, \rho_0, \alpha_0\}$ has been fixed and that an optimal set $\{\omega^*, q^*, v^*, \delta^*\}$ has been determined accordingly. The next step consists of keeping these parameters fixed and calibrate the remaining part of the volatility surface at $t > T^*$ by adjusting the stochastic volatility coefficients $a(t), \alpha(t)$ and $\rho(t)$. In other words, starting from the first smile after T^*, we proceed recursively and at each interval in between consecutive smiles we attempt solving the following problem

$$
\begin{aligned}
(\alpha^*(t), \rho^*(t), a^*(t)) &= \arg \min_{a(t),\alpha(t),\rho(t)} \sum_{j=1}^{N_S} u_{jk} \left| \Sigma \left(T_k, K_j^{(k)}; \alpha(t), \rho(t), a(t) \right) - \Sigma^{BS} \left(T_k, K_j^{(k)} \right) \right|^2 , \\
\text{for } t &\in [T_k, T_{k+1}), \; k = 1, \cdots, L-1, \quad T_1 = T^*, \qquad (15)
\end{aligned}
$$

where $L - 1$ is the number of smiles with expiry strictly larger than T^*. As above anticipated, this problem is well posed since $\{\alpha^*(t), \rho^*(t), a^*(t)\}$ are not degenerate.

Finally, λ can be fine tuned so that the calibrated term structure of the volvol $\alpha^*(t)$ is as constant as possible. Finding the optimal λ^* can be easily achieved by solving the following least squares optimization,

$$
\lambda^* = \arg \min_{\lambda} \left[\sum_{j=1}^{L-1} \left(\alpha_j^*(\lambda) - \alpha_{j+1}^*(\lambda) \right)^2 \right] , \qquad (16)
$$

where vector $(\alpha_1^*(\lambda), \alpha_2^*(\lambda), \cdots, \alpha_L^*(\lambda))'$ comprises the piecewise constant term structure of α_t^*, for a given value of λ. In short, λ^* is the volatility mean reversion that corresponds to the least oscillating calibrated term structure α_t^*. As before, the good news is that once optimization problem (16) has been solved it is possible to keep λ^* fixed without significantly altering the result in future calibrations. In this way, we empirically established that optimal values for λ are in the interval $[0.4, 0.7]$, independently on the chosen market.

As a final remark we recall that calibrating the stochastic volatility part in the long end can be simplified by using information provided by long term asymptotics, Eqs. (27), (32), in conjuction with Eq. (8). Generalizing the results of Appendix C to account for time-dependent (piece-wise constant) coefficient is indeed straightforward. The approach we advocate here is qualitatively similar to the one suggested by Medvedev and Scaillet (2004) for calibrating a model in the Gamma regime through short term asymptotics.

6 Calibration algorithm and numerical results

6.1 The algorithm

In our empirical test we calibrate the model on a set of increasing time to expiry options, corresponding to $T_1 = 1/12$ (1 month), $T_2 = 1/6$ (2 months), $T_3 = 0.25$ (3 months), $T_4 = 0.5$ (6 months), $T_5 = 1$ (1 year), $T_6 = 2, T_7 = 3$, $T_8 = 5$, while T_0 is the observation date and $T^* = 0.25$. We denote by $\mathcal{T}_<$ the set of option expiries shorter than T^*, that is $\mathcal{T}_< := \{T : T \leq T^*\}$, and $\mathcal{T}_> := \{T : T > T^*\}$. The calibration algorithm is based on a recursive procedure that, starting from the shortest expiry T_1, goes as follows.

1. Be $\lambda^{(0)}$ a "trial" initial value for λ. Run a "pre-calibration" test as described in the previous section to determine, for a given a_0, the critical volvol coefficient $\widehat{\alpha}_0$. Finally, determine an initial set $\{a_0, \rho_0, \alpha_0\}$ by fixing α_0 in $(0, \widehat{\alpha}_0]$.

2. Determine the optimal jump parameters set $\{\omega^*, q^*, v^*, \delta^*\}$ by solving the two problems (13) and (14).

3. Determine the diffusion coefficients by calibrating the smile in the interval $\mathcal{T}_>$. Keep jump parameters frozen at the previously calibrated values, then proceed recursively by sequentially calibrating the remaining smiles starting from the one associated to options with the shortest maturity in $\mathcal{T}_>$. This is done by solving the problem (15) and provides an optimal term structure of SV coefficients $\{\alpha_t^*, \rho_t^*, a_t^*\}$ for $t \geq T^*$.

4. If the prior mean reversion rate $\lambda^{(0)}$ has been badly chosen, step 4) might provide a too rapidly increasing or decreasing term structure $\{\alpha^*(t), \rho^*(t), a^*(t)\}$, as previously discussed. We then proceed (condition C3) by solving the problem (16): choose a new $\lambda^{(1)}$ and restart from step 1). Then proceed recursively until the optimal λ^* has been found or until the desired smoothness of coefficients term structure has been achieved.

It is not necessary to perform all four steps every time. Typically δ^*, λ^* and $\{a_0, \rho_0, \alpha_0\}$ are very stable over time and, once estimated, they need not being readjusted too often.

Extensive empirical studies performed on S&P and EuroStoxx data in the time period spanning the years 2002 - 2005 (not reported here) suggest that the optimal λ^* must lie in the interval $[0.4, 0.7]$, as above mentioned. Interestingly, this is in contrast with the most recent findings of λ based on historical data series (Eraker *et al.* (2000)) that assign to the mean reversion rate much lower values: $\lambda \in [0.013, 0.025]$. This indirectly indicates that the market price of volatility risk is significant in SVJD models.

6.2 Numerical results

We calibrate each smile by selecting three options (i.e., $N_S = 3$) struck at K_i, $i = 1, 2, 3$. They correspond to the at-the-money forward option (K_2), to one in-the-money option (K_1), and to one out of-the-money option (K_3), respectively. This is the minimal number of instruments to calibrate ATM volatility level, smile slope and convexity for a given maturity. To select liquid instruments, for every T_i we fix K_1 (resp. K_3) to a fixed number l of standard deviations from the ATM strike, i.e. $K_1 = K_0 - l\sigma^{ATM}\sqrt{T}$, $K_3 = K_0 + l\sigma^{ATM}\sqrt{T}$. Here, σ^{ATM} is the at-the-money Black implied volatility.[18] Scale parameter l is equal to 1, although larger values can be assigned to calibrate wider portions of the smile. In our tests the spot interest rate is 0.033 and there are no dividends. All weights u_{jk}, w_{jk} are equal to 1.

Table 6 shows the outcome of a typical calibration on the EuroStoxx volatility matrix with $\alpha_0 = 30\%$. We fix $a_0 = 6\%, \rho_0 = -0.6$, and $\lambda = 0.6$. Results can be summarized as follows.

1. Calibration is achieved within the required accuracy (all errors are within the volatility bid-ask spread (about 1%)).

2. The term structures of calibrated coefficients $\{\alpha(t), \rho(t), a(t)\}$ are smooth across the whole time range. In particular, no unreasonable jumps are present in switching between the two regimes.

3. Jumps are gradually switched off since the stochastic intensity $\xi(t)$ nicely converges to 0 in approaching T^*.

4. Calibrated instantaneous correlation $\rho(t)$ converges to -1 at large maturities. This clearly indicates that the market-implied skewness is larger than the one predicted by a SVJD model and is in contrasts with correlation estimations based on historical data.[19] For instance, Eraker *et al.* (2000) report that ρ varies typically in $[-0.4, -0.5]$ for the S&P 500 and in $[-0.3, -0.4]$ for the Nasdaq 100 based on statistical estimations. To take into account these features, dynamics Eq.(1) must be generalized. From a statistical point of view there is strong evidence of presence of jumps in volatility (Eraker *et al.* (2000)). Alternatively, these effects could be accounted for by an extension of the present model to include more complex forms of local volatility (Hagan *et al.* (2002)).

5. The typical levels of $\alpha(t)\rho(t)$ for short maturity options are in line with those reported in Bates (2000), Pan (2002) and Jones (2003) - based on statistical estimations - but differ from the findings of Medvedev and Scaillet (2004) - based on short term asymptotics -, although these authors concentrate on the S&P500 index.[20] In fact, for annualized spot volatility of 0.11,our results indicate that the product $\alpha(t)\rho(t)$ is typically in the interval $[-0.1, -0.15]$ for short dated options. For long maturity options, on the contrary, the product $\alpha(t)\rho(t)$ lies in the interval $[-0.25, -0.3]$.

[18] Alternatively one could select K_1 (resp. K_3) as the strike corresponding to 25% (resp. 75%) of the ATM option's delta.

[19] Although the tests presented here refer to the EuroStoxx 50, the same conclusion applies to other indices, including S&P 500 and FTSE 100.

[20] Results not reported here confirm, however, that calibrated values of $\alpha(t)\rho(t)$ for the S&P500 and the Eurostoxx indeces are usually very close based on our method.

6. Contrary to Medvedev and Scaillet (2004), we find that a simple jumps parametrization allows to compensate the lack of convexity in SV models in the short term and well fit the implied smiles in that region.

The rest of the section addresses the robustness of the proposed algorithm. We perform three different tests that are meant to study the stability of the calibrated solution after the input market volatility surface has been manually shocked. The three most relevant PCA modes are independently analyzed. They consist of a parallel shift, a tilt and a bending of the volatility surface, respectively. In our robustness tests, we keep all parameters fixed at their values before the shock (in particular a_0,ρ_0,α_0, λ and ω are fixed). We then apply the shock and finally re-calibrate the model. If the algorithm were robust, a shock in the input of similar magnitude as those observed in the market should not sensibly alter the location of the previously found minima.

Table 7 shows the results after a shock of 1% has been uniformly applied to the volatility matrix. In the second test (Table 8), a tilt is applied to each smile, that is $K_1 \to K_1 - 0.5\%$, $K_3 \to K_3 + 0.5\%$ and K_2 is unchanged. Finally, (Table 9) we study a market scenario where the smile convexity has increased, that is $K_1 \to K_1 + 0.5\%$, $K_3 \to K_3 + 0.5\%$ and K_2 is unchanged. In all cases, results show that the calibration accuracy is unaffected by the volatility shocks and, more importantly, that the new set of calibrated coefficients is very close to the old one. We can deduce that the algorithm is statistically robust in normal market conditions, i.e. if shocks on the volatility surface are not too large and are in line with the typical market movements from one day to another.

If shocks are much larger in size (a few percentage points) tests not reported here indicate that robustness might be sometimes at risk. In this case one should better determine new optimal values for $\lambda_\eta, \delta, \omega$ and $\{a_0, \rho_0, \alpha_0\}$ before running a new calibration.

7 Conclusions

In this paper we have introduced a market-implied calibration technique that can be used for certain classes of stochastic volatility jump diffusion models. In particular, we focused on a model within the linear-quadratic class since generalizing our framework to include stochastic interest rates with possibly stochastic volatility in single and multi-currency markets is possible. We have numerically implemented our method and shown that it is possible to calibrate the entire volatility surface in normal market conditions. In addition, the algorithm is statistically stable and accurate. We have derived useful asymptotic formulae for the moments that can be used to simplify the calibration at long maturities. Further theoretical and numerical developments in this direction, as the extension of the proposed algorithm to more general processes, are left to future research.

References

[1] Andersen, L. and J. Andreasen (2000), "Jump-Diffusion Processes: Volatility Smile Fitting and Numerical Methods for Option Pricing" , *Review of Derivatives Research,* 4, 231-262.

[2] Andersen, T., Benzoni, L. and J. Lund (2002), "An Empirical Investigation of continuous-time equity return models", *Journal of Finance,* 57, 1239-1284.

[3] Backus, D., Foresi, S., Li, K. and L. Wu (1997), "Accounting for Biases in Black Scholes", Working Paper, Stern School of Business, New York University.

[4] Bakshi, G., Cao, C. and Z. Chen (1997), "Empirical Performance of Alternative Option Pricing models", *Journal of Finance,* 52, 2003 - 2049.

[5] Bates, D. (1996), "Jumps and Stochastic Volatility: Exchange Rates Processes Implicit in Deutsche Mark Options", *Review of Financial Studies,* 9, 69-107.

[6] Bates, D. (2000), "The crash of '87: was it expected ? The evidence from option markets", *Journal of Finance*, 46, 1009-1044.

[7] Björk, T. (1998), "Arbitrage Theory in Continuous Time", Oxford University Press, New York.

[8] Chernov, M., Gallant, R., Ghysels, E. and G. Tauchen (2003), "Alternative Models for Stock Price Dynamics", *Journal of Econometrics*, 116.

[9] Cont, R., and P. Tankov (2004), "Non-parametric calibration of jump-diffusion option pricing models", *Journal of Computational Finance*, 7, 1-43.

[10] Craine, R., Lochstoer, L.A., and K. Syrtveit (2000), "Estimation of a Stochastic-Volatility Jump-Diffusion model", Working Paper, University of California at Berkeley.

[11] Deelstra, G., Ezzine, A. and J. Janssen (2003), "Option Valuation in a Non-Affine Stochastic Volatility Jump-Diffusion Model", Working Paper, University of Bruxelles.

[12] Detlefsen, K. (2005), "Hedging Exotic Options in Stochastic Volatility and Jump Diffusion Models", Working Paper, Humboldt University Berlin.

[13] Derman, E. and I. Kani (1994), "Riding on a Smile", *Risk Magazine,* February, 32-39.

[14] Di Graziano, G. and S. Galluccio (2006), "Evaluating Tracking Hedging Errors for General Processes: Theory and Experiments", Working Paper, BNP Paribas and University of Cambridge.

[15] Duffie, D., Pan, J. and K. Singleton (2000), "Transform Analysis and Asset Pricing for Affine Jump-Diffusions", *Econometrica,* 68, 1343 - 1376.

[16] Dumas, B., Fleming, J. and R.E. Whaley (1997), "Implied Volatility Functions: Empirical Tests", *Journal of Finance,* 53, 2059 - 2106.

[17] Dupire, B. (1994), "Pricing with a Smile", *Risk Magazine,* January, 18-20.

[18] Eraker, B., Johannes, M. and M. Polson (2003), "The Impact of Jumps in Volatility and Returns", *Journal of Finance,* 58, 1269 - 1300.

[19] Fengler, M. (2005), "Arbitrage-free smoothing of the implied volatility surface", Working Paper, Humboldt University Berlin.

[20] Galluccio, S., and Y. Le Cam (2006), "Modelling Hybrids with Jumps and Stochastic Volatility", Working Paper, BNP Paribas and University of Evry.

[21] Hagan, P., Lesniewski, A., Kumar, D. and D. Woodward (2002), "Managing Smile Risk", *Wilmott Magazine*, September, 84-108.

[22] Heston, S. (1993), "Closed-form Solutions for Options with Stochastic Volatility, with Applications to Bond and Currency Options", *Review of Financial Studies*, 6, 327-343.

[23] Hull, J. and A. White (1987), "The Pricing of Options on Assets with Stochastic Volatilities", *Journal of Finance,* 42, 281- 300.

[24] Jones, C.S. (2003), "The Dynamics of Stochastic Volatility: Evidence from Underlying and Option Markets", *Journal of Econometrics* 116, 181-224.

[25] Merton, R. (1973), "Theory of Rational Option Pricing", *Bell Journal of Economics and Management Science,* 4, 141-183.

[26] Medvedev, A. and O. Scaillet (2004), "A Simple Calibration Procedure of Stochastic Volatility Models with Jumps by Short Term Asymptotics", Working Paper, University of Geneva and FAME.

[27] Pan, J. (2002), "The Jump Risk Premia Implicit in Options: Evidence from an Integrated Time Series Study", *Journal of Financial Econometrics,* 63, 3-50.

[28] Peng, C. and O. Scaillet (2004), "Linear-Quadratic Jump-Diffusion Modelling with Application to Stochastic Volatility", Working Paper, University of Geneva and FAME.

[29] Piazzesi, M., (2003) "An Econometric Model of the Yield Curve with Macroeconomic Jump Effects", NBER Working paper 8246.

[30] Rebonato, R. (2000), "Volatility and Correlation", John Wiley and Sons, NewYork.

[31] Stein, E.M. and J.C. Stein (1991), "Stock Price Distributions with Stochastic Volatility: An Analytic Approach", *Review of Financial Studies*, 4, 727-752.

[32] Zhang, J.E. and Y. Xiang (2005), "Implied Volatility Smirk", Working Paper, University of Hong Kong.

A Appendix

Remark that $e^{-\int_0^t r_s ds}\psi(u, X_t, t, T)$ is a $\mathbb{P}^*$-martingale. Equivalently, the process

$$h(t, X_t) = e^{Y_t+\gamma(t,T)+\beta(t,T)\cdot X_t+\delta(t,T)\left(X_t^2\right)^2}$$

is a $\mathbb{P}^*$-martingale where $Y_s = -\int_0^t r_s ds$ is a deterministic function. Since the predictable finite variation process of this semimartingale must be zero, application of Itô formula to $h(t, X_t)$ allows identifying the drift term which results into the following equation

$$\begin{aligned} 0 &= \partial_t f(t,x,y) + \sum_{i=1,2} \partial_{x_i} h(t,x,y)\zeta_i(t,x) - r_t.h(t,x,y) \\ &\quad + \frac{1}{2}\sum_{i,j=1,2} (\Sigma\Sigma^t)_{i,j}(t,x)\partial^2_{i,j} h(t,x,y). + \mathcal{A}h(t,x,y), \end{aligned}$$

where ζ is the drift vector and Σ the volatility matrix of the diffusion. $\mathcal{A}$ is the infinitesimal generator of the jump process, i.e.,

$$\mathcal{A}f(t, X_1, X_2) = \xi(t, X_1, X_2)\int_{\mathbb{R}} \left[f(t, X_1 + u, X_2) - f(t, X_1, X_2)\right] d\mathbb{G}(u),$$

Because

$$\mathcal{A}h(t, X_1, X_2) = \left(\xi^0 + \xi^1 X_1 + \xi^2 X_2 + \xi^3 X_2^2\right)(\mathcal{L}(\beta_1(t,T)) - 1)h(t, X_1, X_2),$$

after some algebra, we finally get

$$\begin{aligned} 0 &= \partial_t\gamma + \partial_t\beta_1(t)X_1 + \partial_t\beta_2(t)X_2 + \partial_t\delta(t)X_2^2 \\ &\quad +\beta_1(t)\left(r_t - X_2^2/2 - d_t - \mu_t\right) + \lambda(\beta_2(t) + 2\delta(t)X_2)(a(t) - X_2) \\ &\quad +\frac{1}{2}\left(\beta_1(t)^2\left(X_2\right)^2 + 2\beta_1(t)(\beta_2(t) + 2\delta(t)X_2)\rho X_2\alpha_t + \left(2\delta(t) + (\beta_2(t) + 2\delta X_2)^2\right)\alpha_t^2\right) \\ &\quad + \left(\xi^0 + \xi^1 X_1 + \xi^2 X_2 + \xi^3 X_2^2\right)(\mathcal{L}(\beta_1(t,T)) - 1) \end{aligned}$$

For a completely generic choice of X_1 and X_2 this expression is a second order polynomial in X and is identically equal to zero if and only if all its coefficients are identically zero, which provides the four ODE's.

B Appendix

We consider a generic time interval $[T^{i-1}, T^i)$ where all equation coefficients are supposed to be constant. To solve the system of Riccati ODE's, a precise order must be followed. In this appendix we will omit specifying the time dependency of some variables to lighten notation.

- *First equation.* Solution subject to the final condition $\beta_1(T^i) = u_i^{(1)}$ is immediate, and reads

$$\beta_1(t) = u_i^{(1)}. \tag{17}$$

- *Third equation.* The equation satisfied by $\delta(t)$ is a second-order Riccati equation with terminal condition $\delta(T^i) = u_i^{(3)}$

$$\frac{\partial\delta(t)}{\partial t} = \frac{1}{2}\left(u_i^{(1)} - \left(u_i^{(1)}\right)^2\right) + 2\left(\lambda - \alpha_i\rho u_i^{(1)}\right)]\delta(t) - 2\alpha^2\delta(t)^2,$$

and we have used $\beta_1(t)$ given Eq.(17). After a little algebra, we can rewrite the equations as

$$\begin{aligned}\frac{\partial\delta(t)}{\partial t} &= 2\left(\frac{1}{\alpha_i^2}\left(B_i^2 + \alpha_i^2 A_i\right) - \left(\alpha_i\delta(t) + \frac{B_i}{\alpha_i}\right)^2\right) \\ &= -2\left(\alpha_i\delta(t) + \frac{B_i}{\alpha_i} + \frac{\Gamma_i}{\alpha_i}\right)\left(\alpha_i\delta(t) + \frac{B_i}{\alpha_i} - \frac{\Gamma^i}{\alpha_i}\right),\end{aligned}$$

or, equivalently, by separating the variables

$$4\Gamma^i dt = \alpha_i d\delta(t)\left[\frac{1}{\left(\alpha_i\delta(t) + \frac{B_i}{\alpha_i} + \frac{\Gamma_i}{\alpha_i}\right)} - \frac{1}{\left(\alpha_i\delta(t) + \frac{B_i}{\alpha_i} - \frac{\Gamma_i}{\alpha_i}\right)}\right]$$

and the solution, given the above final condition $u_i^{(3)}$ is therefore

$$\delta(t) = -\frac{1}{\alpha_i^2}\left((B_i + \Gamma_i) + \frac{2\Gamma_i C_i}{e^{4\Gamma_i(T_i - t)} - C_i}\right). \tag{18}$$

- *Second equation.* This equation is linear and its solution (with terminal condition $\beta_2(T^i) = u_i^{(2)}$is lengthy but straightforward. We have

$$\frac{\partial\beta_2(t)}{\partial t} = -2\lambda a_i\delta - \left(\rho\alpha_i\beta_1 + 2\alpha_i^2\delta\right)\beta_2$$

Introducing a new set of functions

$$U(t) = -2\lambda a_i\delta, \quad V(t) = -\left(\rho\alpha_i\beta_1 + 2\alpha_i^2\delta\right)$$

This equation becomes $\partial\beta_2(t)/\partial t = U(t) + V(t)\beta_2(t)$ so that, formally

$$\beta_2(t) = u_i^{(2)} e^{-\int_t^{T^i} V(s)ds} - e^{-\int_t^{T^i} V(s)ds}\int_t^{T^i} U(x) e^{\int_x^{T^i} V(s)ds} dx.$$

After some algebra it is possible to solve all integrals explicitly, and we finally obtain

$$\beta_2(t) = M(t)\left(u_i^{(2)} - K(t)\right). \tag{19}$$

where $M(t)$ and $K(t)$ have been defined in the text.

-*Fifth equation.* The equation to solve reads as

$$\frac{\partial\gamma}{\partial t} = -\Phi^0(\beta_1) + (d + \mu - r)\beta_1 - \lambda a_i\beta_2 - \alpha_i^2\left(\delta + \beta_2^2/2\right)$$

with terminal condition $\gamma(T^i) = u_i^{(4)}$. Once again, the solution is lengthy but straightforward. Notice that the integrals defining $\beta_2(t)$ and $\beta_2(t)^2$ can be alternatively expressed in terms of hypergeometric functions but the expressions are rather involved. In the applications, both integrals can be easily evaluated through a simple Gaussian quadrature algorithm.

C Appendix

1. *Pure jump model.* In Merton's model, the log price satisfies the SDE (with N and μ defined as in the main text) :

$$dX_t = \left(r_t - \eta^2/2 - \mu\right) dt + Y_t dW_t^1 + dN_t,$$

and Levy Khintchine formula provides the expression of the characteristic function

$$\Phi_t^{Mer}(\theta) = \exp\left(t\varphi(\theta)\right), \text{ with } \varphi(\theta) = i\left(r - \eta^2/2 - \mu\right)\theta - \eta^2\theta^2/2 + \xi\left[\exp\left(i\theta q - \theta^2 v^2/2\right) - 1\right].$$

Differentiating this function leads to the expressions of the first four cumulants Π_i $(i = 1, \cdots, 4)$:

$$\begin{aligned} \mathbb{E}\left(X_t\right) &= \Pi_1 = \left(r - \frac{\eta^2}{2}\right)t, \\ \mathbb{E}\left[\left(X_t - \mathbb{E}\left(X_t\right)\right)^2\right] &= \Pi_2 = Var\left(X_t\right) = \eta^2 t + \xi t\left(q^2 + v^2\right), \\ \mathbb{E}\left[\left(X_t - \mathbb{E}\left(X_t\right)\right)^3\right] &= \Pi_3 = \xi t q\left(q^2 + 3v^2\right), \\ \mathbb{E}\left[\left(X_t - \mathbb{E}\left(X_t\right)\right)^4\right] - 3\Pi_2^2 &= \Pi_4 = \xi t\left(q^4 + 6q^2v^2 + 3v^4\right). \end{aligned}$$

Thus skewness and kurtosis read as, respectively

$$\zeta_1(t) = \frac{1}{\sqrt{t}}\frac{\xi q\left(q^2 + 3v^2\right)}{\left[\eta^2 + \xi\left(q^2 + v^2\right)\right]^{3/2}}, \text{ and } \zeta_2(t) = \frac{1}{t}\frac{\xi\left(q^4 + 6q^2v^2 + 3v^4\right)}{\left[\eta^2 + \xi\left(q^2 + v^2\right)\right]^2}.$$

2. *Stochastic volatility model with no jumps.* Despite we here concentrate on the Stein-Stein model, the arguments used in this section can be easily applied to any SV model with simple modifications to determine the moments asymptotics. In the Stein-Stein model, the log price satisfies the equation $dX_t = \left(r_t - \eta_t^2/2\right) dt + \eta_t dW_t^1$. For future purpose, we introduce the process $Y = X - \mathbb{E}X$ and study the dynamics of the powers of Y (by Ito's lemma):

$$dY_t^2 = \eta_t^2 dt + 2Y_t\eta_t dW_t^1, \tag{20}$$

$$dY_t^3 = 3\eta_t^2 Y_t dt + 3\eta_t Y_t^2 dW_t^1, \tag{21}$$

$$dY_t^4 = 6\eta_t^2 Y_t^2 dt + 4\eta_t Y_t^3 dW_t^1. \tag{22}$$

Similarly, given the volatility process $d\eta_t = \lambda(a - \eta_t)dt + \alpha dW_t^2$ we have

$$d\eta_t^2 = \left(\alpha^2 + 2\lambda(a\eta_t - \eta_t^2)\right) dt + 2\alpha\eta_t dW_t^2, \tag{23}$$

$$d\eta_t^3 = \left(3\eta_t\alpha^2 + 3\eta_t^2\lambda(a - \eta_t)\right) dt + 3\alpha\eta_t^2 dW_t^2, \tag{24}$$

$$d\eta_t^4 = \left(6\eta_t^2\alpha^2 + 4\eta_t^3\lambda(a - \eta_t)\right) dt + 4\eta_t^3\alpha dW_t^2. \tag{25}$$

We denote $g(s) = \mathbb{E}\left(\eta_s\right)$, $f(s) = \mathbb{E}\left(\eta_s^2\right)$, $\varphi(t) = \mathbb{E}\left(\eta_t^3\right)$, and $\psi(t) = \mathbb{E}\left(\eta_t^4\right)$. By taking the expectation of both terms in the volatility SDE leads to a ODE in $g(s)$,

$$\begin{cases} g'(s) + \lambda g(s) = \lambda a \\ g(0) = \eta_0 \end{cases},$$

whose solution is given by

$$g(s) = a + (\eta_0 - a)e^{-\lambda t}.$$

Similarly, by taking the expectation of Eq.(23)leads to the ODE

$$\begin{cases} f'(s) + 2\lambda f(s) = \alpha^2 + 2\lambda a g(s) \\ f(0) = \eta_0^2 \end{cases}.$$

whose solution reads as

$$\begin{aligned} f(t) &= \frac{\alpha^2 + 2\lambda a^2}{2\lambda} + 2a(\eta_0 - a)e^{-\lambda t} + \left(\eta_0^2 - \frac{\alpha^2 + 2\lambda a^2}{2\lambda} - 2a(\eta_0 - a)\right) e^{-2\lambda t} \\ &\equiv n + m e^{-\lambda t} + p e^{-2\lambda t}, \end{aligned}$$

By taking the expectation of Eq.(24) leads to the ODE

$$\begin{cases} \varphi'(t) + 3\lambda\varphi(t) = 3\alpha^2 g(t) + 3\lambda a f(t) \\ \varphi(0) = \eta_0^3 \end{cases},$$

whose solution is

$$\begin{aligned} \varphi(t) &= \left(\frac{a\alpha^2}{\lambda} + an\right) + \left(\frac{3\alpha^2(\eta_0 - a)}{2\lambda} + \frac{3am}{2}\right) e^{-\lambda t} + 3ape^{-2\lambda t} \\ &+ \left(\eta_0^3 - \frac{a\alpha^2}{\lambda} - an - \frac{3\alpha^2(\eta_0 - a)}{2\lambda} - \frac{3am}{2} - 3ap\right) e^{-3\lambda t}. \end{aligned}$$

Finally, by taking the expectation of Eq.(25) leads to the ODE

$$\begin{cases} \psi'(t) + 4\lambda\psi(t) = 6\alpha^2 f(t) + 4\lambda a\varphi(t) \\ \psi(0) = \eta_0^4 \end{cases}$$

whose solution reads as

$$\begin{aligned} \psi(t) &= \left(\frac{a^2\alpha^2}{\lambda} + a^2 n + \frac{3\alpha^2 n}{2\lambda}\right) + \left(\frac{2a\alpha^2(\eta_0 - a)}{\lambda} + 2a^2 m + \frac{2m\alpha^2}{\lambda}\right) e^{-\lambda t} \\ &+ \left(6a^2 p + \frac{3p\alpha^2}{\lambda}\right) e^{-2\lambda t} + 4a\left(\eta_0^3 - \frac{a\alpha^2}{\lambda} - an - \frac{3\alpha^2(\eta_0 - a)}{2\lambda} - \frac{3am}{2} - 3ap\right) e^{-3\lambda t} \\ &+ \left(\eta_0^4 - 4a\eta_0^3 + a^2(3n + 4m + 6p) + \frac{\alpha^2}{\lambda}\left(4a(\eta_0 - a) + 3a^2 - 3p - 2m - \frac{3}{2}n\right)\right) e^{-4\lambda t} \end{aligned}$$

We are now ready to derive the expressions of the four cumulants Π_i $(i = 1, \cdots, 4)$.

Derivation of Π_2. From Eq.(20), we have

$$\begin{aligned} \Pi_2(t) &= \mathbb{E}\left(Y_t^2\right) = \mathbb{E}\left(\langle Y \rangle_t\right) = \int_0^t \mathbb{E}\left(\eta_s^2\right) ds \\ &= \int_0^t f(s) ds = nt + \frac{m}{\lambda}\left(1 - e^{-\lambda t}\right) + \frac{p}{2\lambda}\left(1 - e^{-2\lambda t}\right). \end{aligned}$$

Derivation of Π_3. From Eq.(21) we obtain

$$\Pi_3(t) = \mathbb{E}\left(Y_t^3\right) = 3\mathbb{E}\left(\int_0^t Y_s d\langle Y \rangle_s\right) = 3\int_0^t \mathbb{E}\left(Y_s \eta_s^2\right) ds = 3\int_0^t k(s) ds.$$

with $k(t) = \mathbb{E}\left(Y_t \eta_t^2\right)$. To evaluate this last expectation we introduce a new function $h(t) = \mathbb{E}\left(Y_t \eta_t\right)$ and two new processes

$$\begin{aligned} U_t &= Y_t \eta_t = \int_0^t \left(Y_s \lambda (a - \eta_s) + \alpha\rho\eta_s\right) ds + Y_s \alpha dW_s^2 + \eta_s^2 dW_s^1, \\ V_t &= Y_t \eta_t^2 = \int_0^t \left(Y_s \left(\alpha^2 + 2\lambda(a\eta_s - \eta_s^2)\right) + 2\alpha\rho\eta_s^2\right) ds + Y_s 2\alpha\eta_s dW_s^2 + \eta_s^3 dW_s^1. \end{aligned}$$

By taking the expectation of the last two equations (with $\mathbb{E}\left(Y_t\right) = 0$) we arrive to an ODE for $h(t)$

$$\begin{cases} h'(t) + \lambda h(t) = \alpha\rho g(s) \\ h(0) = 0 \end{cases},$$

whose solution is

$$h(t) = \frac{\alpha\rho a}{\lambda}\left(1 - e^{-\lambda t}\right) + \alpha\rho(\eta_0 - a) t e^{-\lambda t}.$$

As a consequence, the equation for $k(t)$ is

$$\begin{cases} k'(t) + 2\lambda k(t) = 2\lambda a h(s) + 2\alpha\rho f(s) \\ k(0) = 0 \end{cases},$$

and its solution, after some algebra, reads as

$$\begin{aligned} k(t) \quad = \quad & \frac{\alpha\rho\left(a^2 + n\right)}{\lambda}\left(1 - e^{-2\lambda t}\right) + \frac{2\alpha\rho\left(m - a^2\right)}{\lambda}\left(e^{-\lambda t} - e^{-2\lambda t}\right) + 2\alpha\rho p t e^{-2\lambda t} \\ & + 2\lambda a \frac{\alpha\rho(\eta_0 - a)}{\lambda^2} e^{-2\lambda t} + 2\lambda a \frac{\alpha\rho(\eta_0 - a)}{\lambda^2}\left(\lambda t - 1\right) e^{-\lambda t}. \end{aligned}$$

Summing up, we arrive at following the expression of the third cumulant

$$\Pi_3(t) = a_0 + a_1 t + a_2 e^{-\lambda t} + a_3 t e^{-\lambda t} + a_4 e^{-2\lambda t} + a_5 t e^{-2\lambda t},$$

with

$$\begin{aligned} a_0 &= 3\frac{\alpha\rho\left(m - a^2\right)}{\lambda^2} + 3a\frac{\alpha\rho(\eta_0 - a)}{\lambda^2} - 3\frac{\alpha\rho\left(a^2 + n\right)}{2\lambda^2} + \frac{3\alpha\rho p}{2\lambda^2}, \\ a_1 &= 3\frac{\alpha\rho\left(a^2 + n\right)}{\lambda}, a_2 = -6\frac{\alpha\rho\left(m - a^2\right)}{\lambda^2}, a_3 = -6a\frac{\alpha\rho(\eta_0 - a)}{\lambda}, \\ a_4 &= -3a\frac{\alpha\rho(\eta_0 - a)}{\lambda^2} + 3\frac{\alpha\rho\left(a^2 + n\right)}{2\lambda^2} + 3\frac{\alpha\rho\left(m - a^2\right)}{\lambda^2} - \frac{3\alpha\rho p}{2\lambda^2}, a_5 = -\frac{3\alpha\rho p}{\lambda}. \end{aligned}$$

Derivation of ζ_1 and asymptotics. From above results, we get

$$\zeta_1(t) = \frac{\Pi_3(t)}{\Pi_2^{3/2}(t)} = \frac{a_0 + a_1 t + a_2 e^{-\lambda t} + a_3 t e^{-\lambda t} + a_4 e^{-2\lambda t} + a_5 t e^{-2\lambda t}}{\left(m/\lambda + p/2\lambda + nt - m/\lambda e^{-\lambda t} - p/2\lambda e^{-2\lambda t}\right)^{3/2}}. \tag{26}$$

From this general expression, it is possible to deduce the asymptotic behavior of the skewness ζ_1 at large times:

$$\zeta_1(t) \sim_{t\to\infty} \frac{a_1}{n^{3/2}} \frac{1}{\sqrt{t}}, \tag{27}$$

since

$$a_0 + a_1 t + a_2 e^{-\lambda t} + a_3 t e^{-\lambda t} + a_4 e^{-2\lambda t} + a_5 t e^{-2\lambda t} \sim_{t\to\infty} a_1 t \text{ and}$$
$$\left(m/\lambda + p/2\lambda + nt - m/\lambda e^{-\lambda t} - p/2\lambda e^{-2\lambda t}\right)^{3/2} \sim_{t\to\infty} (nt)^{3/2}$$

On the other hand, the asymptotic behavior for t small is :

$$\zeta_1(t) \sim_{t\to 0} \frac{3\alpha\rho}{\eta_0}\sqrt{t}, \tag{28}$$

since (from $a_0 + a_2 + a_4 = 0$ and $a_1 - \lambda a_2 + a_3 - a_4 2\lambda + a_5 = 0$),

$$a_0 + a_1 t + a_2 e^{-\lambda t} + a_3 t e^{-\lambda t} + a_4 e^{-2\lambda t} + a_5 t e^{-2\lambda t} = 3\alpha\rho\eta_0^2 t^2 + O(t^3) \text{ and}$$
$$\left(m/\lambda + p/2\lambda + nt - m/\lambda e^{-\lambda t} - p/2\lambda e^{-2\lambda t}\right)^{3/2} = \eta_0^3 t^{3/2} + \mathcal{O}\left(t^{5/2}\right).$$

Derivation of Π_4. From Eq.(22) the fourth cumulant reads as

$$\mathbb{E}(Y_t^4) = 6\int_0^t \mathbb{E}(\eta_s^2 Y_s^2)ds = 6\int_0^t l(s)ds,$$

with $l(t) = \mathbb{E}(\eta_s^2 Y_s^2)$. The expectation of the SDE

$$d\eta_s^2 Y_s^2 = \left(\eta_t^4 + 4\alpha\rho\eta_t^2 Y_t + \alpha^2 Y_s^2 + 2\lambda Y_s^2(a\eta_t - \eta_t^2)\right)dt + 2Y_t\eta_t^3 dW_t^1 + 2\alpha\eta_t Y_s^2 dW_t^2,$$

directly leads to the ODE satisfied by $l(t)$

$$\begin{cases} l'(t) + 2\lambda l(t) = \psi(t) + 4\alpha\rho k(t) + \alpha^2\Pi_2(t) + 2\lambda a v(t) \\ l(0) = 0 \end{cases} \tag{29}$$

where we have introduced the auxiliary function $v(t) = \mathbb{E}(\eta_s Y_s^2)$. At the same time, from

$$d\eta_s Y_s^2 = \left(\eta_t^3 + 2Y_t\eta_t\alpha\rho + \lambda Y_s^2(a - \eta_t)\right)dt + 2Y_t\eta_t^3 dW_t^1 + \alpha Y_s^2 dW_t^2,$$

we obtain the ODE satisfied by $v(t)$:

$$\begin{cases} v'(t) + \lambda v(t) = \varphi(t) + 2\alpha\rho h(t) + a\lambda\Pi_2(t) \\ v(0) = 0 \end{cases},$$

whose solution is (coefficients b_i are given below) :

$$v(t) = \frac{b_0}{\lambda} - \frac{b_5}{\lambda^2} - \left(\frac{b_0}{\lambda} - \frac{b_2}{\lambda} - \frac{b_3}{2\lambda} - \frac{b_5}{\lambda^2}\right)e^{-\lambda t} - \frac{b_2}{\lambda}e^{-2\lambda t} - \frac{b_3}{2\lambda}e^{-3\lambda t} + b_1 t e^{-\lambda t} + \frac{b_4}{2}t^2 e^{-\lambda t} + \frac{b_5}{\lambda}t.$$

By putting all expressions together, we finally get

$$l(t) = c_0 + c_1 e^{-\lambda t} + c_2 e^{-2\lambda t} + c_3 e^{-3\lambda t} + c_4 e^{-4\lambda t} + c_5 t e^{-\lambda t} + c_6 t^2 e^{-\lambda t} + c_7 t e^{-2\lambda t} + c_8 t^2 e^{-2\lambda t} + c_9 t$$

and, by simple integration,

$$\begin{aligned}\mathbb{E}(Y_t^4) &= 6\int_0^t l(s)ds = \left(\frac{6c_1}{\lambda}+\frac{3c_2}{\lambda}+\frac{2c_3}{\lambda}+\frac{3c_4}{2\lambda}+\frac{6c_5}{\lambda^2}+\frac{12c_6}{\lambda^3}+\frac{3c_7}{2\lambda^2}+\frac{3c_8}{2\lambda^3}\right)+6c_0t\\ &-\left(\frac{6c_1}{\lambda}+\frac{6c_5}{\lambda^2}+\frac{12c_6}{\lambda^3}+\frac{3c_8}{2\lambda^3}\right)e^{-\lambda t}-\left(3\frac{c_2}{\lambda}+\frac{3c_7}{2\lambda^2}\right)e^{-2\lambda t}-2\frac{c_3}{\lambda}e^{-3\lambda t}-\frac{3c_4}{2\lambda}e^{-4\lambda t}\\ &-\left(\frac{3c_7}{\lambda}+\frac{3c_8}{\lambda^2}\right)te^{-2\lambda t}-\left(\frac{6c_5}{\lambda}+\frac{12c_6}{\lambda^2}\right)te^{-\lambda t}-\frac{6c_6}{\lambda}t^2e^{-\lambda t}-\frac{3c_8}{\lambda}t^2e^{-2\lambda t}+3c_9t^2\end{aligned}$$

with :

$$\begin{aligned}&b_0=\frac{a\alpha^2}{\lambda}+an+\frac{2\alpha^2\rho^2a}{\lambda}+am+\frac{ap}{2}, b_4=2\alpha^2\rho^2(\eta_0-a), b_1=\frac{3\alpha^2\left(\eta_0-a\right)}{2\lambda}+\frac{3am}{2}\\ &-\frac{2\alpha^2\rho^2a}{\lambda}-am, b_3=\eta_0^3-\frac{a\alpha^2}{\lambda}-an-\frac{3\alpha^2\left(\eta_0-a\right)}{2\lambda}-\frac{3am}{2}-3ap, b_5=a\lambda n, b_2=3ap-\frac{ap}{2},\\ &c_0=\frac{a^2\alpha^2}{2\lambda^2}+\frac{a^2n}{2\lambda}+\frac{\alpha^2n}{2\lambda^2}+2\frac{\alpha^2\rho^2\left(a^2+n\right)}{\lambda^2}+\frac{\alpha^2m}{2\lambda^2}+\frac{\alpha^2p}{4\lambda^2}+\frac{ab_0}{\lambda}-\frac{3ab_5}{2\lambda^2}, c_1=+\frac{2a^2m}{\lambda}+\frac{m\alpha^2}{\lambda^2}\\ &+\left(2-16\rho^2\right)\frac{a\alpha^2\left(\eta_0-a\right)}{\lambda^2}+8\frac{\alpha^2\rho^2\left(m-a^2\right)}{\lambda^2}-\frac{2ab_0}{\lambda}+\frac{2ab_2}{\lambda}+\frac{ab_3}{\lambda}-\frac{2ab_1}{\lambda}+\frac{2ab_4}{\lambda^2}+\frac{2ab_5}{\lambda^2},\\ &c_2=-c_0-c_1-c_3-c_4, c_3=-\frac{4a\eta_0^3}{\lambda}+\frac{4a^2\alpha^2}{\lambda^2}+\frac{4a^2n}{\lambda}+\frac{6a\alpha^2\left(\eta_0-a\right)}{\lambda^2}+\frac{6a^2m}{\lambda}+\frac{12a^2p}{\lambda}+\frac{ab_3}{\lambda},\\ &c_4=-\frac{\eta_0^4-4a\eta_0^3+3a^2n+4a^2m+6a^2p}{2\lambda}-\frac{2\alpha^2a\left(\eta_0-a\right)}{\lambda^2}+\frac{6p\alpha^2+4m\alpha^2+3\alpha^2n-6a^2\alpha^2}{4\lambda^2},\\ &c_5=2ab_1+\frac{8a\alpha^2\rho^2(\eta_0-a)}{\lambda}-\frac{2ab_4}{\lambda}, c_6=ab_4, c_8=4\alpha^2\rho^2p, c_9=\frac{\alpha^2n}{2\lambda}+\frac{ab_5}{\lambda},\\ &c_7=8a\frac{\alpha^2\rho^2(\eta_0-a)}{\lambda}-4\frac{\alpha^2\rho^2\left(a^2+n\right)}{\lambda}-8\frac{\alpha^2\rho^2\left(m-a^2\right)}{\lambda}+6a^2p+\frac{5p\alpha^2}{2\lambda}-2ab_2.\end{aligned}$$

Derivation of ζ_2 and asymptotics. By definition we have

$$\zeta_2(t)=\frac{\Pi_4(t)}{\Pi_2^2(t)}=\frac{\mathbb{E}(Y_t^4)-3\Pi_2^2(t)}{\Pi_2^2(t)}=\frac{\mathbb{E}(Y_t^4)-3\Pi_2^2(t)}{\left(m/\lambda+p/2\lambda+nt-m/\lambda e^{-\lambda t}-p/2\lambda e^{-2\lambda t}\right)^2} \tag{30}$$

Thanks to previous results, a Taylor expansion around $t=0$ leads to

$$\begin{aligned}\mathbb{E}(Y_t^4) &= 3\eta_0^4t^2+\left(\left(7\alpha^2+8\alpha^2\rho^2\right)\eta_0^2+6\lambda a\eta_0^3-6\lambda\eta_0^4\right)t^3+\mathcal{O}(t^4)\text{ and}\\ \Pi_2^2(t) &= \eta_0^4t^2-2\eta_0^2\left(p\lambda+\frac{\lambda m}{2}\right)t^3+\mathcal{O}(t^4),\end{aligned}$$

which implies that when t approaches 0, the kurtosis behaves as

$$\zeta_2(t)\sim_{t\to0}\left(\left(7\alpha^2+8\alpha^2\rho^2+6\lambda p+3\lambda m\right)\eta_0^{-2}+6\lambda a\eta_0^{-1}-6\lambda\right)t. \tag{31}$$

Finally, by studying the behavior at large times we obtain

$$\begin{aligned}\Pi_4(t) &\sim_{t\to\infty}\left(6c_0-3(2m+p)\frac{n}{\lambda}\right)t\\ \Pi_2^2(t) &\sim_{t\to\infty}n^2t^2+(2m+p)\frac{n}{\lambda}t,\end{aligned}$$

and thus

$$\zeta_2(t)\sim_{t\to\infty}\frac{\left(6c_0-3(2m+p)n/\lambda\right)}{n^2}\frac{1}{t}. \tag{32}$$

Expressions 28, 27, 31, 32, are equivalent to those given in the main text. This ends the proof.