Filtering and likelihood estimation of latent factor jump-diffusions with an application to stochastic volatility models

esposito, francesco paolo and cummins, mark

dublin city university, business school

1 May 2015

Filtering and Likelihood Estimation of Latent Factor Jump-Diffusions with an Application to Stochastic Volatility Models

F. P. Esposito ✉ [*] and M. Cummins ✉ [*]

[*]DCU Business School

01/05/2015

Abstract

In this article we use a partial integral-differential approach to construct and extend a non-linear filter to include jump components in the system state. We employ the enhanced filter to estimate the latent state of multivariate parametric jump-diffusions. The devised procedure is flexible and can be applied to non-affine diffusions as well as to state dependent jump intensities and jump size distributions. The particular design of the system state can also provide an estimate of the jump times and sizes. With the same approch by which the filter has been devised, we implement an approximate likelihood for the parameter estimation of models of the jump-diffusion class. In the development of the estimation function, we take particular care in designing a simplified algorithm for computing. The likelihood function is then characterised in the application to stochastic volatility models with jumps. In the empirical section we validate the proposed approach via Monte Carlo experiments. We deal with the volatility as an intrinsic latent factor, which is partially observable through the integrated variance, a new system state component that is introduced to increase the filtered information content, allowing a closer tracking of the latent volatility factor. Further, we analyse the structure of the measurement error, particularly in relation to the presence of jumps in the system. In connection to this, we detect and address an issue arising in the update equation, improving the system state estimate.

Keywords: latent state-variables, non-linear filtering, finite difference method, multi-variate jump-diffusions, likelihood estimation.

1 Introduction

The estimation of parametric models of stochastic differential equations (SDE) has become a subject of growing interest in recent years, see Sørensen (2004), Aït-Sahalia (2006) for a survey. There are many approaches available, each designed to deal with specific problems connected to the inference exercise. It is difficult to classify the solution methods throughout the problems posed by the estimation of parametric SDEs. A partial categorisation discriminates by

- *Moment-based estimator.* Seminal papers are the GMM of Hansen and Scheinkman (1995), the indirect inference of Gourieroux et al. (1993), the efficient method of moments of Gallant and Tauchen (1996) or the martingale estimating functions of Bibby and Sørensen (1995). Studies based on the characteristic function include Singleton (2001), Chako and Viceira (2003).
- *Likelihood-based estimator.* Seminal papers are the finite difference approach as pioneered by Lo (1988), the simulation of likelihood of Pedersen (1995a) or the Markov Chain Monte Carlo (MCMC) methods as independently derived by Jones (1999), Elerian et al. (2001), Eraker (2001). Another interesting approach is the polynomial expansion as in Aït-Sahalia (1999), Aït-Sahalia (2002), Aït-Sahalia (2008).

Comparison studies have been performed by Jensen and Poulsen (2002), Lindström (2007), Hurn et al. (2007) and in relation to filtering problems, see Lund (1997), Duffee and Stanton (2012) and Christoffersen et al. (2014).

The main issue with the estimation exercise is related to the fact that the likelihood of the stochastic model is generally not known in closed form, making the use of an exact likelihood estimator virtually impossible, except for a few limited special cases. A further problem is represented by the imperfect sample information about the system to be estimated. In the first instance, the system is observed only at discrete times, which poses the problem of how to optimally project the system forward in time, given the current information. Secondly, problems of greater interest in finance involve the system state being only partially or indirectly observable, namely: stochastic volatility (e.g. Heston, 1993, Duffie et al., 2000) and term structure models (e.g. Duffie and Kan, 1996, Chen and Scott, 2003). This lack of information issue can be optimally solved by filtering, which basically consists of finding the mean square best estimate of the system state, given the partial set of historical information available. This can be viewed as a projection problem in the space of mean square integrable martingales, see Øksendal (2003). The whole filtering exercise boils down to the construction of the projection operator, jointly with an update procedure for the projection of the system state, once the observable information has been made available. Several authors develop filtering procedures to tackle the latency of the state components. Examples are Bates (2006), Jiang and Oomen (2007). However, these algorithms are specific to an affine structure of the SDE and have in common the use of the spectral function for affine jump-diffusion models, which is known in semi-analytical form (Duffie et al., 2000).

In this paper, we develop a particular of filter that can treat more general jump-diffusion models and produce estimates of the state vector which include latent components. We then apply the filter within the context of a parametric model estimation. We acknowledge that a filter, similar in spirit to the one used here, has been recently applied to pure diffusion models by Hurn et al. (2013). In that paper the authors apply the same procedure used in this article to derive the generic main filtering equation for non-linear pure diffusions. They solve the nonlinearity problem via the application of a quasi-likelihood approach which is coherent with the estimation strategy they adopt. This paper is different in that we independently devise an extension to the non-linear filter which is able to handle multivariate jump components. The form of the jump is quite general, allowing the possibility to handle synchronous or asynchronous jumps, state-dependent jump size distribution along with affine as well as differently specified state-dependent jump-intensities. The nonlinearity problem is solved with a second order approximation which allows for a quasi-analytical form of the filter that can be implemented in a very flexible fashion. Secondly, along the lines of the original approach found in Maybeck (1982), we complement the filter with an estimation technique that adopts the same methodology used to derive the main filtering equation. The econometric procedure consists of an approximate maximum likelihood (AML) approach whereby the likelihood is obtained via the numerical solution of the partial integral-differential equation (PIDE) describing the transition probability of the multivariate jump-diffusion under analysis, with the application of the finite difference method for the construction of the diffusion operator and the use of a discretisation to deal with the jump component. Within the structuring of the main block of the approximated likelihood, we also discuss the issue of the stabilisation of the PIDE operator approximation and report a criterion which provides a major guideline for this purpose. We also characterise in finer detail the form of the general likelihood for the purpose of a simplified computer implementation. Finally, in the empirical section we analyse a stochastic volatility model with jumps with focus on the system state design. Inspired by previous works such as Bollerslev and Zhou (2002), we introduce the integrated variance variable, which is proved to carry significant auxiliary information when estimating the stochastic volatility factor. Moreover, we test the form of the measurement error variable, providing evidence that augmenting the state to model the error as an auxiliary latent system component is significant. Along the lines of Dempster and Tang (2011) we provide evidence that a martingale form for the error is more desirable. Further, we have discovered that in the presence of jumps a pure diffusion system state estimate might experience shocks that can be accommodated via the extension of the measurement error to jump components. Another interesting conclusion of this paper is that, depending on the system design, in the presence of jumps the measurement error might actually be a redundant system state component, whereby its impact on the system total variability is absorbed by the jump component projection.

The paper is organised as follows. Section 2 presents the non-linear filter and as a key contribution to the literature the extension of the filtering procedure for handling jump components. Section 3 describes the estimation procedure and analyses the problem of the stabilisation of the PIDE operator approximation. Section 4 contains the empirical analysis of a suite of stochastic volatility models. It first depicts the system equations used in the Monte Carlo simulation and further analyses from a statistical perspective the system design, with particular attention to the use of the integrated variance for the sake of the latent state estimation and the form of the measurement error as an auxiliary latent state variable. A further sub-section presents the estimation of the model parameters via AML and discusses some auxiliary measure of the filter performance. Section 5 concludes.

2 The construction of the nonlinear filter

The problem we tackle is the statistical estimation of a parametric model, which describes the dynamics of a vector-valued stochastic process $(S_t)_{t\in[0,T)}$. We call S_t a *system*, essentially because the stochastic differential equations describing its components' dynamics are interconnected. The system S is arranged into two components $S = (X, Y)$, in relation to their observability. We indicate the observable components as Y, whose dynamics are described as a function of X, the state of the system. The system state X is fully or partially latent, that is its path can only be inferred from the information coming from the measurement Y. In solving the estimation problem, we are therefore concerned with the device estimating the latent state of the system and with the construction of the full likelihood for parametric estimation purposes. This section is dedicated to the solution of the former problem, which, as a key contribution to the literature, is extended to include jump components. The construction of the likelihood is pursued in Section 3.

Filtering is the problem of finding the best estimate in a mean square sense of the state of the system, that is the $\mathcal{G}_t$-measurable random variable $\bar{X}_t$ that minimises the path-wise distance from the true state X_t. Let the probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ and let the flow of information as represented by the set $\mathcal{G} \subset \mathcal{F}$, be respectively defined as the algebra of events representing the observable trajectories and the full set of information about the system (X, Y). The solution to the problem defined above, is the projection from the space $\mathcal{L}^2(\mathbb{P})$ onto the space $\mathcal{K} \subset \mathcal{L}^2(\mathbb{P})$ of the $\mathcal{G}_t$-measurable random variables. The projection operator corresponds to the expectation $\mathbb{E}[\cdot|\mathcal{G}_t]$, see Øksendal (2003). The following aims to construct an approximation of the projection operator, when the stochastic process is a jump-diffusion. Actually, because the observables are recorded only at discrete times, we need two projection operators providing the latent state estimates. The approach undertaken here, following the cited seminal literature, consists of the derivation of two equations defining the operators of projection $\mathbb{E}[X_t|\mathcal{G}_{t-\delta_t}]$ and $\mathbb{E}[X_t|\mathcal{G}_t]$. In order to simplify notation, we will indistinctly indicate $\mathbb{E}_{t|s}[X] =: \mathbb{E}[X_t|\mathcal{G}_s] := \bar{X}_{t|s},\ s \leq t$. Corresponding to the previous expectations, the non-linear filter is composed of the following equations. The *time-propagation equation* moves the state estimates between the observation times $t - \delta_t$ and t, the time segments being not necessarily equally spaced, whereas the *update equation* generates the new estimate of the partially latent state vector X_t when a new observation Y_t is available. The update equation is given in a convenient simplified form, as a function of Y and its projection $\bar{Y}$, of the projected state vector $\bar{X}$ and their second order cross-moments. The problem amounts to the construction of the projection and update operators of the first two central moments of the system state. Formally, the framework is given by the parametric system state

$$\mathrm{d}X = b(X^-;\theta)\mathrm{d}t + A(X^-;\theta)\mathrm{d}W + J(z;X^-,\theta)\mathrm{d}N \tag{1}$$

The functions b, A include dependency on the parametric vector $\theta \in \Theta$. The jump size component vector

J depends on the mark point z, whose distribution is parametric and may depend on the state. The random drivers of the system are the Brownian vector W and the Poisson counting process N, with stochastic intensity $\lambda(X^-;\theta)$. The random functions b, A and J are assumed to satisfy conditions that grant a unique solution for Eq. (1) (see e.g. Platen and Bruti-Liberati, 2010), $\forall\theta\in\Theta$. In Eq. (1) we make explicit the dependency on the left limit of X, that is its level immediately before the jump, if any. Subsequently, this notation is dropped, whereby we focus on the construction of the estimation procedure. For a complete treatment of the stochastic integral X and its components, see, e.g., Cont and Tankov (2003), Hanson (2007). For the practical purpose of system estimation, we will assume that the jump size vector of the synchronous jump can be written as $J = G(z)f(X)$, with $G = [g_{ij}(z)]_{ij}$ and $g_{ij} = 0$ when $i \neq j$, where f and g are mapping, respectively, from the domain of X and z, the mark point vector, to $\mathbb{R}^\bullet$. Here, the definition of J is a working tool which makes the jump size dependent at the same time on the mark-point vector z and on the state X, but in a way that allows the factorisation of the jump-component and the state component in the time-propagation equation. The functions f and g increase the flexibility of the statistical model.

The forward equation

Later in the construction of the time-propagation operator, a key role is played by the Kolmogorov forward equation (KFE). In general, considering the SDE (1), the KFE that is the equation describing the transition probabilities of the system, is found as:

Proposition 2.1 (The multi-dimensional jump-diffusion PIDE)**.** *The Kolmogorov forward equation for the Itô process with Poisson jump components* (1) *is*

$$\partial_t\,[\mathbf{p}] = (\mathscr{A}_X + \mathscr{J}_X)\,[\mathbf{p}] \tag{2}$$

where the differential operator $\mathscr{A}_X$ is defined by the position, $C = AA^\star$

$$\mathscr{A}_X[\mathbf{p}] \quad\equiv\quad \tfrac{1}{2}\sum_{ij}\partial^2_{x_i x_j}\,[C_{ij}\mathbf{p}] - \sum_i \partial_{x_i}\,[b_i\mathbf{p}] \tag{3}$$

and the integral operator $\mathscr{J}_X$ is defined as

$$\mathscr{J}_X[\mathbf{p}] \equiv -\,(\lambda\mathbf{p}) + \int_{\mathscr{Z}} \mathrm{d}\mathbb{Q}(z;\,h)\,|\nabla h|\,\,(\lambda\mathbf{p})\circ h \tag{4}$$

Proof. See Hanson (2007) □

In Eq. (4) $\mathbb{Q}$ is the jump size probability measure, $h\colon X^+ \to X^-$ is the post-jump transform, $|\nabla h|$ is the determinant of the Jacobian of h and we indicate by $\circ$ the function composition operator. For ease of presentation, we consider the counting process to be scalar and allow the synchronous jump vector J to be state dependent or not. The jump intensity is the process $\lambda(X)$.

The second component of the system is represented by the observation equation,

$$Y = \mathbf{q}_X(X) + E \tag{5}$$

where E is the *measurement error*, which is left unspecified at the moment. In Eq. (5) we assume a simple linear form for $\mathbf{q}(X) = HX$, through the constant matrix H. This case is relevant for the stochastic volatility model, where H is a pick matrix and for a latent factor term structure model, which targets the estimation of the empirical measure. The extension of Eq. (5) to more general forms requires an approximation to be fully implemented, see e.g. Nielsen et al. (2000), Baadsgaard et al. (2000). See Christoffersen et al. (2014) for a study of non-linearity in the observation equation in the case of an unscented Kalman filter.

In this article, we need to include a further component to the system state. This auxiliary component, intrinsically latent by its nature, is a defining object of the jump component, that is its, possibly state dependent, intensity process. We give it here in its level effect form

$$\lambda = \mathbf{q}_\lambda(X) \tag{6}$$

In the following we extend the time propagation equation as conceived by Maybeck (1982) to handle a marked point Poisson component, which can be state-dependent in the jump intensity function and in the jump size distribution. Our work relies on the intuition of using the jump operator of the forward equation to extend the system state projection dynamics to include a jump component and in deriving workable expressions for the estimation of the latent system-state. Furthermore, we also address from an implementation perspective a feature of the update equation arising when jumps are included in the system state equation and offer robust statistics confirming the effectiveness of the solution.

2.1 The time-propagation equation

In order to construct optimal estimates of the state of the system X, which is observed at discrete times only, we need conditions for the evolution of the system state projections between two observation times. This is called the *time-propagation equation*. The idea in Maybeck (1982) is to derive possibly approximated ordinary differential equations for the first two moments of X, cfr. Nielsen et al. (2000), Baadsgaard et al. (2000)

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}\bar{X} & = \int X \partial_t p \, \mathrm{d}X \\ \frac{\mathrm{d}}{\mathrm{d}t}\bar{V} & = \int XX^\star \partial_t p \, \mathrm{d}X - \frac{\mathrm{d}}{\mathrm{d}t}\bar{X}\bar{X}^\star - \bar{X}\frac{\mathrm{d}}{\mathrm{d}t}\bar{X}^\star \end{cases} \tag{7}$$

In Eq. (7), we substitute the KFE for the jump-diffusion transition probability $\partial_t p$ to obtain an exact or an appropriately proxied ordinary differential equation (ODE) system for $\bar{X}_{t|s}$ and $\bar{V}_{t|s}$. The aim is to calculate the solution of (7) for the jump-diffusion (1). To obtain the solution the following integrals are involved $\int X(\mathscr{A}+\mathscr{J})[p]$ and $\int XX^\star(\mathscr{A}+\mathscr{J})[p]$ that because of linearity can be handled separately with respect to each individual KFE operator. For the same reason, further synchronous jumps can be easily included to the state model. We split Eq. (7) into its diffusion and jump component, using linearity of the operators, that is $\frac{\mathrm{d}}{\mathrm{d}t}(\cdot) = \frac{\mathrm{d}}{\mathrm{d}t}(\cdot)_{\mathscr{A}} + \frac{\mathrm{d}}{\mathrm{d}t}(\cdot)_{\mathscr{J}}$. To simplify notation, we indicate the operator $\mathbb{E}^{\mathbb{P}}[\cdot]$ with $(\!| \cdot |\!)$ and $\mathbb{E}^{\mathbb{Q}}[\cdot]$ with $\langle\!| \cdot |\!\rangle$. We find that the diffusion component of Eq. (7) is

Proposition 2.2 (The Diffusion Component of the Time-Propagation Equation, Maybeck, 1982)**.**

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\left(\bar{X}\right)_{\mathscr{A}} &= (\!| b |\!) \\ \frac{\mathrm{d}}{\mathrm{d}t}\left(\bar{V}\right)_{\mathscr{A}} &= (\!| C |\!) + (\!| bX^\star |\!) + (\!| Xb^\star |\!) - (\!| b |\!)\bar{X}^\star - \bar{X}(\!| b |\!)^\star \end{aligned} \tag{8}$$

Proof. See the Appendix. □

The filter $\left(\bar{X}, \bar{V}\right)$ can be extended with the same approach described above, adapting the integration procedure to handle the jump component. We derive the auxiliary filter component providing the following formal ODE system.

Proposition 2.3 (The Jump Component of the Time-Propagation Equation)**.**

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\left(\bar{X}\right)_{\mathscr{J}} &= \langle\!| G |\!\rangle(\!| \lambda f |\!) =: U \\ \frac{\mathrm{d}}{\mathrm{d}t}\left(\bar{V}\right)_{\mathscr{J}} &= \langle\!| G |\!\rangle(\!| \lambda f X^\star |\!) + (\!| \lambda X f^\star |\!)\langle\!| G |\!\rangle + (\!| \lambda f f^\star |\!) \odot \langle\!| gg^\star |\!\rangle - U\bar{X}^\star - \bar{X}U^\star \end{aligned} \tag{9}$$

Proof. See the Appendix. □

In the above, we have used the sign $\odot$ to indicate component-wise multiplication. The jump component (9) represents to the best of our knowledge a novel contribution to the literature and provides an extension to the nonlinear filter of Maybeck (1982) and the most recent applications in finance of Nielsen et al. (2000), Baadsgaard et al. (2000) and Hurn et al. (2013), which can be used for the estimation of the latent state of jump-diffusions. In order to get a workable expression to use for computations the time-propagation equations require the evaluation of the expectations on the RHS of the previous differential expressions.

2.2 Approximating the expectation operator

With Eqs. (8) and (9), we have obtained an ordinary differential system which describes the projection operators for the first two central moments of the state-equation as a function of time. However, it has to be noticed that Eq. (8) and Eq. (9) are only a formal definition, because the RHS is in general unknown. In order to obtain a workable specification, we need to characterise this formal statement of the time-propagation equations. The approach undertaken in this paper is along the lines of the seminal papers cited above. The expectation of a generic scalar function of the state $q(X)$ is approximated by taking a Taylor series expansion of q around the current state estimate $\bar{X}$ and applying the operator $\mathbb{E}[\cdot]$, to both side of the equation, cfr. Maybeck (1982), Nielsen et al. (2000), to obtain

$$\mathbb{E}[q(X)] = q\left(\bar{X}\right) + \tfrac{1}{2}\operatorname{trace}\left[\nabla^2 q\left(\bar{X}\right)\cdot\bar{V}\right] + R, \tag{10}$$

where we neglect the remainder R, which contains a third order central moment function. The truncated second order expansion introduces bias correction and can be seen as a stochastic equivalent of the extended Kalman filter[1]. It is interesting to notice that if the state function $q(X)$ is at most quadratic, the

[1]The extended Kalman filter corresponds to a first order approximation within the same methodology, cfr., e.g., Lund (1997) and in comparison with other methods in Duffee and Stanton (2012), Christoffersen et al. (2014).

expansion in Eq. (10) is exact. In general, we have obtained an estimate of the time-propagation equation for the jump diffusion (1), with state-dependent jump intensities and amplitudes. This approach differs from that undertaken in Hurn et al. (2013), which uses the quasi-likelihood to approximate the integral with numerical quadrature. We believe this approach offers convenience in allowing for the construction of the time-propagation equation for the estimation of the main projection operator in a quasi-analytical form and further it can be coded in a very flexible fashion.

Example: state-independent affine jump-diffusion

From Eqs. (8), (9) and (10) it is evident that when the b and λ are affine, the jump size is state independent and the diffusion matrix is at most a quadratic function of the state, the time propagation equations are exact and can even be solved explicitly. For instance, in the affine jump-diffusion case, when the jump intensity is $\lambda(X) = \lambda_0 + \lambda_1 \cdot X$ and the synchronised jump vector J is state-independent, we get the exact ODE system

$$
\begin{aligned}
\tfrac{\mathrm{d}}{\mathrm{d}t}\bar{X} &= \tilde{a} + \tilde{B}\bar{X} \\
\tfrac{\mathrm{d}}{\mathrm{d}t}\bar{V} &= \tilde{D} + \tilde{B}\bar{V} + \bar{V}\tilde{B}^{\star}
\end{aligned}
\tag{11}
$$

where

$$
\begin{aligned}
\tilde{a} &= a + \lambda_0 \\
\tilde{B} &= B + \langle\!|J|\!\rangle \lambda_1^{\star} \\
\tilde{D} &= AD_{\bar{X}}^2 A^{\star} + (\lambda_0 + \lambda_1 \cdot \bar{X})\langle\!|JJ^{\star}|\!\rangle
\end{aligned}
$$

which admits a closed form solution. In other situations we have to revert to an approximated ODE.

Example: non-affine volatility

When the stochastic system is not affine, we approximate the time-propagation equation via Eq. (10). In this example, we look at a scalar pure diffusion, with an affine drift $a + bX$ and a squared diffusion function $C = \sigma^2 X^{2\gamma}$, hence the ODE driving the system projection is then

$$
\begin{aligned}
\tfrac{\mathrm{d}}{\mathrm{d}t}\bar{X} &= a + b\bar{X} \\
\tfrac{\mathrm{d}}{\mathrm{d}t}\bar{V} &= \sigma^2\bar{X}^{2\gamma} + \sigma^2(2\gamma^2 - \gamma)\bar{X}^{2(\gamma-1)}\bar{V} + 2b\bar{V}
\end{aligned}
\tag{12}
$$

The expression (12) is used later within the experimental section, in junction with a larger system, when conducting an exercise with a non-affine model.

2.3 The update equation

The non-linear filter we have developed in the previous section has the purpose of projecting the system between two consecutive times, carrying over the whole set of information inferred by the observation vector for the sake of delivering the best estimate of the partially observed system state. Once the system is at the observation time t and new information is collected about Y, we need a means to incorporate such quantities into the system state estimate in an optimal way. The update equation consists of a mechanism to estimate the expectation $\bar{X}_{t|t}$ by refreshing the system state projection with the newly arrived information Y_{t^+}, which are the only observable quantities in the context of a latent system state. The optimal filter $\bar{X}$ represents the best estimate of the state under partial information, which is the natural condition under which data on a phenomenon are presented to the researcher.

The update equation, fundamentally, consists of the application of Bayes' rule, when conditioning the state estimates onto the observed information set at current time. Assuming the update equation form is a linear function of the residuals, it can be found that:

Proposition 2.4 (The update of a linear projection, Maybeck, 1982)**.** *The update equation for the non-linear filter defined by Eqs. (8) and (9) is given by*

$$
\begin{aligned}
\bar{X}_{t|t} &= \bar{X}_{t|s} + \Sigma_{xy}\Sigma_{yy}^{-1}\left(Y_t - \bar{Y}_{t|s}\right) \\
\bar{V}_{t|t} &= \bar{V}_{t|s} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}
\end{aligned}
\tag{13}
$$

with

$$\Sigma_{yy} = \mathbb{E}_{t|s}\left[\left(Y_t - \bar{Y}_{t|s}\right)\left(Y_t - \bar{Y}_{t|s}\right)^\star\right]$$
$$\Sigma_{xy} = \mathbb{E}_{t|s}\left[\left(X_t - \bar{X}_{t|s}\right)\left(Y_t - \bar{Y}_{t|s}\right)^\star\right] = \Sigma^\star_{yx}$$

Proof. See the Appendix. □

Embedding new information into $\bar{X}_{t|s}$ about the observed residuals $Y_t - \bar{Y}_{t|s}$ imports into the states estimates information that would be lost otherwise.

For the application we construct in this study, the observation equation of interest is a linear function of the latent state, $Y = HX$ where in general H is a constant matrix. This assumption implies the exact estimates $\bar{Y} = H\bar{X}$, $\Sigma_{yy} = H\bar{V}H^\star$. Whenever Y is a generic function of the state X, $Y = \mathbf{q}(X) + E$, the system variables $\bar{Y}$, Σ_{yy} and Σ_{xy} are approximated via Eq. (10), cfr. Nielsen et al. (2000). We remark that the update equation (13) in the general case of a nonlinear filtering corresponds to a first order expansion of the projection operator $\mathbb{E}_{t|t}$. Expanding this approximation to higher orders, whereby the expectations are approximated via Eq. (10) is impossible.

2.4 The measurement error

In the observation equation (5), we left the measurement error E unspecified. We discuss the modelling of the process E in this section. In the literature the measurement error is generically indicated as a random process E_t with zero mean and constant covariance matrix Σ_ε, a white noise which is at most cross-section correlated. However, in a recent paper, cfr. Dempster and Tang (2011), it has been statistically proved that the measurement error manifests mean reversion and cross correlation with the state. In Dempster and Tang (2011), the authors plug the measurement error into the state equation, a choice that allows one to design an evolutionary equation for E that could better track the underlying state of the system X. The behaviour of the measurement error is actually the result of the filtering process, therefore it is straightforward to expect a mean reverting or even a martingale behaviour which might have a random impulse which is correlated with the diffusions W. The outcome of the inclusion of the measurement error into the state vector consists of effectively transforming E into another latent component of an augmented state vector. This is an implementation strategy that allows greater flexibility in the system state filtering and the parametric estimation, introducing parameters that might be able to modulate the estimation residuals and improve the quality of the fitting.

In this article we extend the intuition in Dempster and Tang (2011) a step further. We report that in the presence of the jump component the measurement error can experience jumps that if neglected might propagate to the system state estimate and arise unexpectedly in other parts of the system. The case that we will be exploring in the empirical section is that of a latent pure diffusion stochastic volatility whose filtered state, according to a certain measure to be specified, might sometimes experience excessive variations or jumps that can be improved or corrected by a measurement error that contemplates jumps. Thus, building on the same strategy, we will augment the system state by a measurement error vector E that, in general, will include a mean reversion term, a diffusion and a jump component. However, having both a diffusion and a jump component in the error term might result in a useless over-parametrisation. The general equation for the measurement error E is

$$\mathrm{d}E = (c + C_0 E)\,\mathrm{d}t + C_1 \mathrm{d}W_E + J_E \mathrm{d}N_E. \tag{14}$$

We assume that the eigenvalues of the constant matrix C_0 grant that the process E is stationary or at least non-explosive, whereas the constant c is such to compensate the drift generated by the jumps $\Sigma_{i=1}^{N_t} J_{t_i}$, if any, and hence the unconditional mean is zero. As the vector process E belongs to the system state, the diffusion component drivers might possibly be correlated with the diffusive impulse of X. The innovation we introduce is to allow the measurement error to jump via the Poisson point process represented by the stochastic differential $J_E \mathrm{d}N_E$. In general, the measurement error jumps may be synchronised or not with the X component's jumps, whereby in the latter case an intensity function should be specified and parametrised. However, we have found experimentally that at least for the model under test, the former case does not allow the flexibility required to explain the unexpected jumps in the state estimate. When the measurement error jump process N_E is not synchronised with the jump in the state, if any, a further hypothesis on the dynamics of the stochastic intensity process should be specified. This may give rise to auxiliary latent measurement error variables. However, in the empirical section we will adopt the simplification that the jump error intensity is an affine function of the state X. In several case studies, this assumption will produce a statistically significant effect on improving the latent state estimates.

3 Estimation

In Section 2 we have presented the procedure to estimate the system state X in the presence of partial information via a nonlinear filter, which has been extended to include jump elements. In this paragraph we complement that technique with an estimation method for the parametrisation of multivariate jump-diffusions with latent components. Whereas we have exploited a PIDE approach for the derivation of the symbolic solution of the time-propagation equation (7) for the filtering of the latent state X, here we use an intrinsic filtering approach for the derivation of the likelihood function, whose main component is the solution of the PIDE which describes the transition probability density of the system (1). In treating the latter problem, we follow the route leading to the construction of the approximate likelihood function of the latent state by numerically solving the forward equation (2). We use an AML approach exploiting the inherent optimality properties of the exact likelihood that can be achieved asymptotically by the approximation, cfr. Pedersen (1995a), Poulsen (1999). The same result extends to the PIDE version, provided uniform convergence of the solution, see Lindström (2007). Finally, the approximate likelihood of the observed variable Y is achieved via the numerical integration of the latent state component. The contribution of this article in the estimation exercise involves a particular implementation strategy that allows further simplification of the AML algorithm. In the following we describe the approach undertaken in this paper in the development of the estimation algorithm, which allows a considerable saving of computational time. We also tackle the problem of the stabilisation of the PIDE operator approximation.

3.1 Approximate likelihood function

The objective is to build an approximate likelihood function for the observables, $\mathbb{P}\left(Y_t|Y_{t-1}\right)$, which is able to capture higher order moments implied by the system probability structure. In order to handle the presence of the partially or totally latent system state X, we use the classic Bayesian decomposition of the full likelihood function, which is then concentrated onto the observable variable by marginalising the system state. The contribution of this article in the estimation problem is the characterisation of the likelihood function which hinges on the particular implementation strategy of the likelihood core component $\mathbb{P}\left(X_{t+1}|X_t\right)$. As a result, we achieve a simplification of the estimation algorithm.

We work with the following equation:[2]

$$\mathbb{P}\left(X_{t+1}|Y_t\right)\mathbb{P}\left(Y_t|Y_{t-1}\right) = \int \mathrm{d}X_t\, \mathbb{P}\left(X_{t+1}|X_t\right)\mathbb{P}\left(Y_t|X_t\right)\mathbb{P}\left(X_t|Y_{t-1}\right) \tag{15}$$

In Eq. (15) the likelihood component of the observables Y, that is $\mathbb{P}\left(Y_t|Y_{t-1}\right)$, acts as a normalisation constant, hence it is obtained by integrating the RHS w.r.t. X_{t+1}. Furthermore, the above likelihood function defines an iteration that recursively combines the transition probability of the system state with the density implied by the observation equation. The equation defined in (15) progressively generates the likelihood for the observables and the function $\mathbb{P}\left(X_{t+1}|Y_t\right)$, which acts as a weighting function in the successive step. The initial condition $\mathbb{P}(X_0|Y_{-1})$ is set to $\int \mathrm{d}X_{-1}\mathbb{P}(X_0|X_{-1})$, wherever the initial condition is not observable; otherwise, the initial condition is set to the observables.

In order to portray further the likelihood recursion, we exploit the solution chosen in the article of Dempster and Tang (2011), by augmenting the system state with the measurement error, which renders the observation equation as $Y = q_X(X)$. As a consequence we have that $\mathbb{P}\left(Y_t|X_t\right) \equiv \delta\left[q(X_t) - Y_t\right]$, where δ is the pulse function. Concerning the latter statement, we need to clarify that we implicitly assume that, whenever the observation function is not surjective, the Y belongs to the image of q. Further, we also notice that if any subcomponent X_a of the partition $X = (X_a, X_b)$ does not enter q then we simply refer to $\mathbb{P}\left(Y|X\right) = \mathbb{P}\left(Y|X_b\right)$[3].

To complete the construction of Eq. (15), we build its core component as the solution of the KFE via a combination of a finite difference method (FDM) and an ordinary integral approximation. In recent years, the FDM has received renewed attention in continuous-time financial econometrics, since the

[2]Equation (15) is obtained by plugging into the definition of $\mathbb{P}\left(X_{t+1}|Y_t\right) = \int \mathrm{d}X_t\, \mathbb{P}\left(X_{t+1}|X_t\right)\mathbb{P}\left(X_t|Y_t\right)$ the probabilistic correspondent of the update equation, that is $\mathbb{P}\left(X_t|Y_t\right) = \frac{\mathbb{P}(Y_t|X_t)\mathbb{P}(X_t|Y_{t-1})}{\mathbb{P}(Y_t|Y_{t-1})}$.

[3]However, if the range of Y does not coincide with the image of q, the probability $\mathbb{P}\left(Y|X\right)$ would not be defined. In fact, we have the chain of equalities

$$\mathbb{P}\left(Y|X_a, X_b\right) = \frac{\mathbb{P}(X_a|Y,X_b)\mathbb{P}(Y|X_b)}{\mathbb{P}(X_a|X_b)}$$

and if $Y \notin q(X)$, then $\{Y, X_b\} \equiv \emptyset$ and $\mathbb{P}\left(X_a|Y, X_b\right)$ is not defined. Furthermore, from the RHS of the above chain we notice that because $y = q(x)$ then $\{Y, X_b\} \equiv \{X_b\}$ and therefore $\mathbb{P}\left(Y|X\right) = \mathbb{P}\left(Y|X_b\right)$.

seminal papers of Lo (1988), Pedersen (1995b) and Poulsen (1999). Examples are Jensen and Poulsen (2002), Lindström (2007), Hurn et al. (2010), Lux (2012).

We approximate each component $\mathbb{P}(X_{t+1}|X_t)$ as the solution of the system of ordinary differential equations in the time dimension obtained by applying a finite difference scheme to the diffusion operator and a discretisation of the integral component for the jump operator. Formally, we transform the PIDE (2) into the ODE homogeneous system

$$\tfrac{\mathrm{d}}{\mathrm{d}t}p(t) = (A+J)\cdot p(t) \Rightarrow p(t) = \exp\left[(A+J)t\right]\cdot p(0) \tag{16}$$

that can be formally solved as exhibited. The vector p contains the stack of grid-points, $t = t_i - t_{i-1}$ and $p(0)$ is a representation of the delta function centred at $x_{t_{i-1}}$. Eventually, we focus on the construction of an approximation of the integral-differential operator in the space dimension defining the PIDE for the transition probability distribution. Solving the ODE along the time dimension via the exponentiation of the approximate operator in the space dimension offers a very attractive feature in terms of computational speed and that will be fully exploited in the derivation of the final expression for the likelihood iteration. The construction of the delta function initial condition also involves some caution. A coarse grid might generate abrupt variations in the function approximation that result in oscillatory behaviour of the solution. It might be necessary to force the solution to be positive. Another approach involves the use of the terminal value of a Euler approximation in place of the initial condition, cfr. Poulsen (1999). We choose the first approach, which we justify with the following proposition:

Proposition 3.1. *Assuming the sequence of functions $p^k \to p$ pointwise, with $p \geq 0$, $\forall x$. Thus,*

$$\left|p^k\right| \to p$$

Proof. See Appendix. □

The proposition 3.1 basically says that it is safe forcing the approximated solution to be positive, because its absolute value will still converge to the true solution.

In this paper the solution we adopt for the estimation of the parameters of a multivariate jump-diffusion with latent state components, consists of constructing the likelihood function on a fixed grid space, avoiding dependency of the grid on the system status. Lacking this feature, the AML solution for each time frame would require a localised solution, forcing the computational burden to increase exponentially. In order to achieve this target, we adopt two schemes. First, we implement the solution of the KFE for the system state as described above. The exponentiation of the operator renders the solution dependent on the initial condition only through a matrix operation, thus entailing that the matrix exponential has to be performed only once at each likelihood calculation, instead of as many times as the sample size. However, in order to keep the likelihood calculation on a fixed grid at each recursion, there is a second aspect to deal with. In case the system state contains non-stationary components whose transition densities do not depend on their own levels, their initial condition would have the sole effect of shifting the probability structure along their range of definition. Therefore, to keep the likelihood on a fixed grid, we are more interested in considering their first differences rather than their levels, otherwise forcing the implementation of a specific operator at each likelihood recursion. In the following we derive a finer characterisation of the likelihood iteration in Eq. (15), whose purpose is solely presenting the method although further specifications depend on the partitioning of the state vector and can in general be refined more depending on the system state form. Thus, assuming, for instance, the partitioning of the system state into $X = (X^0, X^1)$ and that $Y = q(X^0)$, with X^0 non stationary[4] and level independent, we can rearrange the likelihood iteration into

$$\mathbb{P}\left(\Delta X^0_{t+1}, X^1_{t+1}|\Delta Y_t\right) = a_t \int \mathrm{d}X^1_t \left[\mathbb{P}\left(\Delta X^0_{t+1}, X^1_{t+1}|0, X^1_t\right)\int_{\Omega_t}\mathrm{d}\Delta X^0_t\, \mathbb{P}\left(\Delta X^0_t, X^1_t|\Delta Y_{t-1}\right)\right] \tag{17}$$

where the vector integration variable $\mathrm{d}X^j = \mathrm{d}x_1 \ldots \mathrm{d}x_{k_j}$ and $a_t^{-1} = \mathbb{P}(Y_t|Y_{t-1})$. The integration domain $\Omega_t := \{\Delta X^0_t \colon q(\Delta X^0_t) - \Delta Y_t = 0\}$. It is important to notice that in order to transform the likelihood iteration into Eq. (17), where the probability mass stays confined onto a user defined multi-interval, instead of drifting alongside the non-stationary system components, the function defining the observation equation q is required to be linear. In the empirical section we will use the proposed approach to produce Monte Carlo experiments with several stochastic volatility model specifications.

[4] In this example we assume that Y is a non stationary process, which is a function of a non stationary state component $q(X_0)$. This is the relevant case for applications. The case where Y is a function of a stationary state can be accommodated with the same approach. The case when Y is at the same time a function of the level of both a stationary and non stationary process cannot be dealt with a fixed grid approach, unless the Y as a function output can be partitioned according to the partitioning of X into stationary and non stationary components. For the non-stationary case we will further need to require that the function q be linear, in order to get $Y + \Delta Y = q(X_0 + \Delta X_0) \Rightarrow \Delta Y = q(\Delta X_0)$.

3.2 Stabilisation of the operator approximation

In this section we discuss an important problem related to the solution of the system state transition density $\mathbb{P}(X_{t+1}|X_t)$. An aspect which is often disregarded when constructing a PDE solution with a FDM is the *operator stabilisation*. In system theory, the concept of stability is in general referred to the sensitivity of the solution to a small perturbation of the system parameters, usually the initial conditions. In the application of finite difference schemes for the solution of PDEs and in particular to evolutionary problems, the study of the stability of the operator approximation is concerned with the problem of suitably choosing the discretisation scheme for the time variable, in order to assure the stability of the solution approximation, see Tavella and Randall (2000), Duffy (2006). In reality, it would be more appropriate to define this problem as how to preserve stability when moving to the time discretisation scheme, assuming the operator in the space dimension is initially stable. However, if we disregard the "special" role that is assigned to time, we might think of this problem for FDM as the problem of stabilising the time-space operator, altogether. In the particular implementation chosen in this paper, we avoid the last issue by taking the analytic solution of the space dimension discretised problem. As a consequence, we are left with the study of the stability of the space-discretised operator. In this section, we take into consideration the stability aspect of the matrix approximation of the original PIDE problem, and provide a criterion which is exploited as a tool for the analysis of this aspect of the likelihood approximation. The design of a stable system is crucial in the construction of the estimation procedure. We offer some guidelines in what follows, without the pretence of exhausting the argument but more to shed some light onto the stabilisation issue related to the implementation of the AML procedure employed in this work. The stabilisation of the operator approximation is in general a task to be pursued on a case-by-case basis.

In general terms, there is not an overall solution to the problem of the stability of a FDM operator and its study is usually confined to the derivative operator approximation ∂. Considering the solution (16), we search for system matrices with negative real part eigenvalues, which would grant stable solutions. However, not every finite difference scheme grants stable operators. See Tavella and Randall (2000) as to how to construct finite difference scheme of any order. In Fig. 1 we show a perfectly legit first order first derivative finite difference scheme which produces a totally unstable matrix. The figure exhibits the real part of the matrix eigenvalues of a simple example whereby the use of a given finite difference scheme produces a problematic approximation of a first order derivative operator. The matrix eigenvalues are all positive entailing that the exponentiation of the matrix will rapidly explode at low variations of t. The situation becomes even more intricate in the case of mixed derivatives with multiple dimensions, where the options to design a finite difference approximation of a given order become numerous. As a general rule, we exploit the following version of the *Gerschgorin's disks* theorem. Let $[a_{ij}] =: A \in \mathbb{C}^{n\times n}$ be a square matrix and $c_i = \sum_{j\neq i} |a_{ij}|$, we have

Theorem 3.1 (Gerschgorin, 1931)**.** *The eigenvalues of the matrix A lie in the union of the disks*

$$\{z \in \mathbb{C}\colon |z - a_{ii}| \leq c_i\}, 1 \leq i \leq n$$

Therefore, the general criterion in constructing a FDM scheme consists of plugging as many negative values as possible on the matrix operator diagonal and keeping the radius of the disks small, in order to obtain stable differential operator approximations. In Fig. 2, we show the ordered real part of the eigenvalue of the matrix approximation of the operators ∂_x, ∂^2_{xx} and ∂^2_{xv}, which have been used in the empirical section. The careful choice of the differentiation scheme allows for the complete stabilisation of the matrix operators.

To conclude this section, we recall that in the jump-diffusion case with scalar point mark, we approximate the jump operator with a trapezoidal rule that yields an approximation matrix, which is a triangular banded matrix. Usually, this matrix has a minor impact on the overall structure of the system operator. What can be said in general is that this matrix contains positive entries, therefore it will shift the centre of the Gerschgorin's disks to the right, determining a cause for the decrease of the stability. Another problem is related to the fact that although the individual operator blocks might be stable or negligibly destabilising, their sum does not necessarily retain the same spectral structure of the individual factors. Moreover, a few problematic eigenvalues might have been generated by numerical truncation that exhibit a positive real part which tend to shrink when increasing the thickness of the grid. For instance, in the following table[5] we compare the percentage of the matrix trace which can be attributed to problematic eigenvalues for several approximations of the full operator $\mathscr{A}_X + \mathscr{J}_X$:

[5]The table exibits the value of the index in corresponding to a given grid dimension. The correspondent matrix approximation of the PIDE operator is a square matrix whose side is as long as the grid volume expressed in terms of number of steps for each dimension. That corresponds to all the possible combinations of the coverage of the values allowed by the multi-dimensional grid. The index shows the percentage ratio of the sum of the absolute value of positive real part

grid dimension	instability index
10×10	2.27%
20×20	0.16%
30×30	0.04%
50×50	0.00%

We observe that they tend to disappear when the system dimension increases, that is the grid becomes more dense.

In practice, it is important to remark that before using a PIDE solution for the design of the AML procedure, a preliminary study of the system state distribution function output must be run thoroughly and the system grid must be carefully chosen, in order to calibrate the algorithm. Having chosen a fixed grid approach, it is important to make the multi-interval proportionate to the long-run variability of the system components and further rendering a sufficiently fine grid in order to obtain a well enough detailed distribution at low variability levels. The point of strength of this approach is that it is easily adaptable for any type of multivariate jump-diffusion model, whereby the implemented codes require only few modifications to change the system state equations. Moreover, it allows the parameter estimation time to be greatly reduced showcasing the benefit of the sole need of one single matrix exponential at each likelihood cycle, leaving ample possibility for a targeted implementation of the numerical algorithm at the programming level. The procedure also offers the possibility of cutting off the numerical gradient loop, at the cost of a further matrix exponentiation, cfr. Van Loan (1978). We leave this latter step to subsequent implementation and testing.

4 Experimental Section

In this section, we provide empirical evidence in a simulated environment of the efficacy of the described procedures for the estimation of jump-diffusion models. We deal with a stochastic volatility model with jumps whose diffusive component may have a non-affine state function volatility, while the jump component is characterised by a state dependent stochastic intensity which can be non-linear as well. The approach is particularly appealing for the filtering of non-affine processes, where we cannot use, for example, the spectral filtering procedure as in Bates (2006), and even so, in the affine case the jump-diffusion filtering proposed in this paper is particularly convenient. In fact, for the implementation of the cited approach, the estimation of the latent state requires the combined use of numerical integration and differentiation of the characteristic function, constructed via the solution of the ODE system associated with the affine model, cfr. Duffie et al. (2000). On the contrary, the extended filter introduced in this paper and constructed directly in the time domain, produces estimates of the system state trajectories through the less complex recursion described by the Eqs. (8), (9) and (13), an implementation strategy that requires a lower number of approximation layers. Concerning the parameter estimation exercise, the PIDE approach used here is instrumental to the optimisation of the model parameters by which a sample path is most likely to have been generated, and the attention is more for the implications in the filtering context. We refer to articles such as Jensen and Poulsen (2002), Lindström (2007) for a comparative analysis of the AML adopted in this section.

With the Monte Carlo experiment, we check that the estimation of the system state and the model parameters are significant, while focusing the analysis on several aspects of the system development and estimation that we have found to be relevant. Specifically, we augment the system state of the stochastic volatility model with the integrated variance dynamics y_1 and produce statistical evidence that it brings significant information into the system for the sake of the estimation of the latent state voolatility. Further, introducing a measurement error in the observation equation (19), we test the autocorrelation function of the output residuals to investigate the white noise hypothesis, finding that the strong autocorrelation and the presence of a unit root suggest to model the y_1's residual in a martingale form that entails the augmentation of the state equation by the measurement error, as it was an auxiliary latent variable. Moreover, while experimenting several model specifications as a jump-diffusion in x with a pure diffusive v factor, we have detected the presence of unexpected jumps in the projection of the variable v. In the latter situation, we provide evidence that this phenomenon, when present, may be corrected with a pure jump measurement error, which is probably the most suitable form for the measurement error, when the filter is applied to jump-diffusions. Further down the line, we investigate its nature by questioning the need of a measurement error. In fact, we provide evidence that in the majority of the extreme cases, the presence of a jump in the supposedly pure diffusive v factor can be accommodated by

eigenvalues over the sum of the absolute value of the full set of eigenvalues. That number can be thought of as an index of instability of the system matrix: the lower the better.

dropping the variable E and by the likelihood estimation of the parameters. Finally, with another AML exercise, we exhibit the parameter estimates and their likelihood derived standard error, highlighting the very low tracking error of the v estimates with respect to the simulated ones and whereby we also show how to use the filter to derive an estimate of the jumps. The main statistics are presented in Tabs. 2 and 3.

The simulations are produced with a simple Euler scheme at a very tight interval, plus the product of the jump size times a binomial random variable with $p_t = 1 - \exp\left[-\lambda(X_t)\Delta t\right]$.

4.1 System Design

In this section we characterise the form of the system used in the Monte Carlo experiments. We introduce the SDE describing the dynamics of the system and its observables. The baseline state equation is given by

$$\begin{aligned} \mathrm{d}x &= \theta\sqrt{q_x(v)}\,\mathrm{d}W_0 \\ \mathrm{d}v &= \kappa(\omega - v)\,\mathrm{d}t + \sigma\sqrt{v}\left(\rho\,\mathrm{d}W_0 + \sqrt{1-\rho^2}\,\mathrm{d}W_1\right) + J_v\,\mathrm{d}N_1 \\ \mathrm{d}u &= \theta^2 q_x(v)\,\mathrm{d}t \\ \mathrm{d}\pi_x &= -J_0\,\mathrm{d}N_0 \\ \mathrm{d}\pi_u &= J_0^2\,\mathrm{d}N_0 \end{aligned} \tag{18}$$

The dynamic system described by the Eqs. (18) represents the state of the reference multi-variate process we are employing for the Monte Carlo study. The observation equation is given by the linear form

$$\begin{aligned} y_0 &= x + \pi_x \\ y_1 &= u + \pi_u + \varepsilon \end{aligned} \tag{19}$$

The observations are represented by the two dimensional vector $Y = (y_0, y_1)$. The process y_0 is a jump-diffusion with quadratic variation given by the (Lebesgue)-stochastic integral $\int \mathrm{d}t\,\left[\theta^2 q_x(v) + \bar{J}_0^2\lambda_{N_0}\right]$, representing the prototype stochastic volatility model for the experimental exercise. As it can be noticed, the quadratic variation of y_0 depends on the instantaneous variance of the pure diffusion x and the second order (infinitesimal) moment of the Poisson random measure, symbolically described by the stochastic differential $\mathrm{d}\pi_x$. The parameter θ is a scaling factor, whereas the diffusion function is taken as $q_x = v^{2\gamma}$, with $0 < \gamma < 1$. This specification corresponds to a deterministic constant elasticity function[6]. Given the dynamics of v, basically the constant parameter γ regulates the skewness of the unconditional distribution of the unscaled diffusion, determining an asymmetric response in the function v^γ, when v_t is below or above the long run average of one. The case $\gamma = 1/2$ corresponds to an affine volatility function. The state component v is a scalar square-root process, which can be affected by jumps N_1 that are not synchronised with the jumps N_0 in y_0 and that have independent size distributions. The inclusion of a jump in volatility allows the system to resemble the behaviour of the double-jump model introduced by Duffie et al. (2000). Other working hypotheses can be easily implemented. We allow for the presence of correlated diffusive random drivers between the factor x and the stochastic volatility factor v, a fact that in applied financial econometrics is used to reproduce the so called leverage effect, cfr. for instance Glosten et al. (1993). However, we notice that this feature might be reproduced by suitably modelling the jumps and the intensity function. The variable v is also input to the state dependent jump intensities $\lambda_{N_0}(v)$, which can be a linear or a square function of v, allowing state dependency and non linearity also in the stochastic jump-intensity process that have been handled by the jump-extended filter introduced in this paper. For the specific simulation experiment where the jump in variance N_1 is considered, its jump intensity and that of N_0 will be taken as linear in v with the scaling factor of λ_{N_1} given by a fixed proportion of the parameter λ_{N_0}. As they are conceived, the stochastic intensities will produce clustered jumps manifesting their concentration during volatility peaks. The simple non-linear stochastic intensity model with jump frequency described by $\lambda_0 v^2$, will exacerbate the feature just described. The jumps in volatility are assumed to have an affine intensity, with a coefficient which is a fixed fraction of λ_0. The jump size distribution of the Poisson point process component J_0 is specified as exponential, which is also the case for J_1, or extreme value distributions, depending on constant parameters. In the latter case, the jump size distribution is given by the polynomially decaying extreme value distribution $\mathbb{P}\left[J_0 \le z\right] = 1 - (1 + z/\alpha)^{-\eta}$, where we need to require finite order fourth moments in order to grant the well definiteness of the time-propagation equation. In the appendix (cfr. A.5) we calculate the moments

[6] In economics, the elasticity of the utility function $U(c)$ is defined as $-cU''/U'$. This corresponds to an elasticity of the function v^γ of $1-\gamma$.

of the extreme value distribution function used in the experiments. We notice here another interesting aspect of the estimation approach presented in this paper, that is the ability of handling extreme value distributions (EVD) for the jump size, a characteristic that is prevented by the requirement of finite exponential moments in the jump extension of the indirect Hermite expansion approach of Aït-Sahalia (2002) , cfr. Filipović et al. (2013) and Singer (2006). The coefficient $\omega = (\kappa - \lambda_0 \bar{J}_v)/\kappa$, allows the stochastic factor v to oscillate around one with unconditionally unitary mean. For that, in the estimation exercise we will require the mean reversion speed to be higher than the jump drift, in order to obtain stationary variance. As it can be noticed, the jump process N_0 is not compensated, hence y_0 will have a negative drift. Another key point in the design of the system state is the isolation of the diffusion from the jump components, such that the filter state estimates will produce a direct estimate of the the jump process itself. However, this strategy is not directly implementable for the jump component in the v factor, because it does not enter directly in any observable. As it will be seen in the AML exercise, the estimation of this feature is problematic.

The observable y_1 plays a special role. As the state vector (x, v, π_x) is completely unobservable with a dimensionality which is higher than that of the observable y_0, we would expect a high degree of indeterminacy in estimating its projection onto $\mathcal{G}_t$. However, for the case under analysis, we can resort to stochastic calculus to obtain two new variables which increase the information content available by introducing a new observable. We plug into the observation equation the "integrated variance" of the process y_0, which is partially observable. We will prove later that this innovation matters. The integrated variance has been used in other applications in a realised volatility context, see for instance Bollerslev and Zhou (2002), which exploits its moment structure to improve the estimation of a stochastic volatility model. In this paper, the origination point is different. We construct the process $y_1 = u + \pi_u$ considering the SDE which describes the observable $y_0 = x + \pi_x$ and derive the process dynamics for $w = y_0^2$

$$
\begin{aligned}
\mathrm{d}w &= 2y_0 \left(c_0 \,\mathrm{d}t + \theta \sqrt{q_x(v)}\,\mathrm{d}W_0 \right) + \theta^2 q_x(v)\,\mathrm{d}t + \left[(y_0 - J_0)^2 - y_0^2 \right] \mathrm{d}N_0 \\
&= 2y_0 \,\mathrm{d}y_0 + \theta^2 q_x(v)\,\mathrm{d}t + J_0^2 \,\mathrm{d}N_0 \\
&= 2y_0 \,\mathrm{d}y_0 + \mathrm{d}u + \mathrm{d}\pi_u.
\end{aligned}
$$

Hence, let $y_1 = w - 2\int y_0 \,\mathrm{d}y_0$ to obtain the new observable y_1. The augmented state vector for the simulation exercise is therefore $X = (x, v, u, \pi_x, \pi_u)$. In deriving the reference state equation, we highlight the separation of the main random sources into specific state variables, a design choice that allows for the disentangling of the jump variable from the diffusion component. As a consequence, the filter Eqs. (18) and (19) will be able, in particular, to produce the projection of the latent variable π_x, which accumulates the jumps of the observable y_0. We will use the latter filter output to estimate the jump times and sizes; simulation shows that the jumps can be estimated as the tail events of the first difference distribution of the cumulative jump state component, when the tail is cut at the expected unconditional frequency of jumps. The residuals after the distribution cut, can be seen as a projection error that is expected to be negligible.

At this stage of the construction of the stochastic system to be used to experiment the non-linear filter complemented with the chosen AML procedure, we have not yet introduced any measurement error process. We introduce the variable $E = \varepsilon$ into the definition of y_1, as in Eq. (19), postulating a further stochastic factor in explaining the dynamics of the observations for the sequence $y_1(t_n), n \in \mathbb{N}$. However, some questions are crucial in the design of an efficient filter. In the following, we pay particular attention to the form that ε is most likely to exhibit and the implications that entails; further we question the introduction of an additional unobserved process in the system, such as ε. Concerning its form, the classical hypothesis in filtering is that the measurement error is a white noise ε_t. However, when introduced into the system, the residuals that it gathers contradicts this simple hypothesis. We follow Dempster and Tang (2011) in modelling directly the measurement error within the system state and extend this concept further by allowing the measurement error to be affected by jumps. When it is not taken as a simple white noise, the general form of the error is a scalar version of Eq. (14), where its diffusion or pure jump form will be used separately. The empirical evidence shows that in the presence of jumps, a pure diffusion system state component, like v in the experiments carried out in this paper, might in extreme cases exhibit jumps which have not been modelled and that can be accommodated by the introduction of a jump latent measurement error variable. However, the introduction of an auxiliary variable employed as a measurement error for the observation equation might be redundant. In fact, in a simulation exercise presented below, we gather evidence showing that the latter feature is likely to be due to unoptimised parameters. It is further to be noticed that dropping the variable ε can be justified by the presence of a non-zero residual in the cumulative jump projections, once that the jump estimates have been removed. This characteristic signals that in the presence of jumps some background noise in the observation equation might be absorbed by the jump projection.

For the illustration of the AML estimation procedure, the system (18) is re-elaborated in order to construct a simplified version of the likelihood function. We deal with the stochastic volatility system in Eq. (18) whereby the integrated variance component has been dropped. In the latter case, the observation error is irrelevant for the parameter estimation exercise. We produce estimates for the two dimensional system (x_0, v). In constructing the likelihood iteration, we apply the approach depicted in the estimation defining Section 3. We remark that the system state and the observables are, respectively, $X = (\Delta x_0, v)$, $Y = \Delta y_0$. Hence, we get the following iterations for the target likelihood:

$$\mathbb{P}(X_{t+1}|Y_t) = a_t \int \mathrm{d}v_t \, \mathbb{P}(\Delta x_{0,t+1}, v_{t+1}|0, v_t)\mathbb{P}(\Delta y_{0,t}, v_t|\Delta y_{0,t-1}) \tag{20}$$

The log-likelihood function for the observables is then

$$\mathscr{L}_N = \tfrac{1}{N}\Sigma_i \log a_{t_i}$$

Ultimately, the AML module constructed in Section 3 provides a moderately time consuming algorithm for the estimation of continuous time parametric models. The integrals involved in the likelihood iteration are discretised to obtain simple matrix multiplications on the defined grid. The full vector of the output parameters is the set $(\theta, \kappa, \sigma, \rho, \gamma, \lambda_0, \alpha_0, \eta)$, which represents, respectively, the diffusion scaling factor for the observed variable x, the speed of mean reversion constant for the stochastic factor v, its volatility coefficient, the correlation factor of the diffusion stochastic factors of x and v, the diffusion exponent factor and the diffusion scaling factor. The remaining factors have different use according to the model they are used for; the α parameter is a scaling parameter for either the exponential or the extreme value jump size distribution, whereas η is employed either as the exponent factor for the EVD or as the scaling factor for the exponential distribution for the jump in the v factor. Whenever unused, certain parameters are dropped.

4.2 Diagnostic check of the system state specification

In the first set of tests, we include the integrated variance into the system and provide statistical evidence of its statistical significance. We postulate a martingale form for the measurement error which is tested successively in a second set of diagnostic checks, where we want to establish the most likely form of the measurement error. Table 2 provides the summary results of the test statistics constructed in this first part of the empirical section; Table 1 contains a legenda of the acronyms used in the latter table. In most cases, the only variables that are left free are the measurement error parameters, while the system state parameters are set at the simulation values. In the case of the testing for the exclusion of measurement error, the model parameters have been estimated with AML. We include several affine models, that also include jumps, which are kept at quite a high frequency, when testing for jumps in the filtered path of a pure diffusion v. We use the quasi-analytical approximation for filtering non-affine model specifications as the CEV and the squared jump intensity. The drift induced by the jump component is the same for the exponential and EVD jumps size, whereas in the case of the squared jump intensity it is slightly smaller, to compensate for the higher variability generated by the squaring of the v factor. The expected jump size is the same for all the models.

The integrated variance

Bollerslev and Zhou (2002) use several sample moments of the integrated variance of equity prices, to estimate parametric stochastic volatility models with the method of moments. The background idea is to extend the information available from observed data to improve the performance of the estimation function. Their main focus, however, is on realised volatility with high frequency data. In this paper, we employ the same idea of extending the observable set, but from a different perspective. We manipulate the state equation to obtain the dynamics of the squared variation, which, up to the integral approximation, is observable. The integrated variance components are then included in the augmented state vector, as in Eq. (18), while a new observable, y_1, is obtained. The question we answer in this section is whether this extension is significant in terms of the estimation ability of the latent state of v. However, in some cases the inclusion of the variable y_1 could even be necessary before being significant. For instance, the filter considered in the estimation exercise with $\rho = 0$ and lacking the variable y_1, for any initial condition, would be able to produce only a curve decaying to the long run mean, because the absence of correlation completely disconnects the projection of the latent state from the observable x. The latter feature is related to the intrinsic martingale structure by which x has been modelled.

We introduce the integrated variance of the process x in order to augment the information used by the filter to produce the latent state estimation. We prove here that the inclusion of the integrated variance within the system steers a significantly informative data flow to the filtering device. We test the latent state v estimation ability of the filter with or without the integrated variance y_1. In case of a system with the latter variable, we include the measurement error ε which we assume to be a martingale diffusion. The model parameters are set to their simulation values, whereas the error parameter is optimised by minimising the mean squared error (MSE) of the y_1 against the filtered one. The estimation ability of the filters is measured with the sample MSE, $\mathbb{E}\left[(v-\bar{v})^2\right]$, of the filtered $\bar{v}$ path. We apply the t-test and the F-test on a large sample of model paths, yielding a set of sample MSE's. The test results are presented in Tab. 2, labelled test T_0. The null hypotheses are based on the assumptions that the plain filter without the integrated variance is yielding a MSE which is on average lower and less volatile than those produced by the augmented state filter. The hypotheses correspond to the irrelevance of the filter with an extended observation equation. We consider the case of a pure diffusion x, whereby it should be easier to estimate the variance path as the variance of the observable which is generated totally by a single risk factor. The models encompass affine and non-affine specification, with a constant elasticity of variance specification, with the parameter γ ranging from 0.2 to 0.9. Nevertheless, the null hypotheses are strongly rejected in all the cases under analysis. It is interesting to notice that in the affine case with $\rho = 0$ the reduced state model produces not only a significantly higher MSE on average, but it also generates a huge variability of this performance measure. The same happens in the case of the highly sensitive response of the x's diffusion to the v factor.

A martingale measurement error

In the previous section, we have assumed that the augmented observation Eq. (19) contains a measurement error ε which is a martingale diffusion. In this section, we include the y_1 variable and test that the latter assumption concerning the measurement error is the most likely. As a consequence, with the martingale assumption, the system state is augmented with the ε dynamics specification.

With the test labelled T_1 we first verify whether assuming a white noise structure for the measurement error is preferable to a martingale equation. With a large sample, that is $2,000$ sample paths generated for a simple affine pure diffusion with quite a high diffusion coefficient for the latent factor v and high negative correlation ρ, we perform a dual tail t-test and F-test for the null that, respectively, the ex-post MSE average and variance of the estimated sample path $\bar{v}$ against the realised trajectories are different within the two samples. We found that these hypotheses are strongly rejected, supporting the conjecture that using a white noise or a martingale form for ε are equivalent. However, this approach is not optimal. In fact, with the sequence of tests labelled T_2 and T_3, we check for the autocorrelation presence in the residuals and even for the presence of a unit root in lag polynomial of, respectively, the sample residuals of the white noise and the martingale measurement error form. Again, we choose a simple model structure, that is that of a pure diffusion affine model. In this case already, the results are clear. We randomly select from a 2,000 sample path for several model parameter specifications, two sample paths upon which we perform a Phillip-Perron test for the presence of unit root and an augmented Dickey-Fuller test for the presence of autocorrelation. The time series which are investigated are the sample measurement error $y_1 - \bar{y}_1$ output of the filtering algorithm. In the case of the martingale error with the test T_3, the autocorrelation of the residuals are tested after the application of the first difference operator, as for the confirmation of the unit root presence. The unit root presence is strongly rejected in the white noise case (test T_1), but the presence of autocorrelated residuals is always confirmed. In the case of the martingale specification of the measurement equation, as in the case of the group of tests T_3, we find very general confirmation of the unit root presence, which is removed by the application of the first difference operator, which is to say that the Dickey-Fuller test generally fails to reject the absence of autocorrelation. This battery of tests on random samples taken from model specification characterised by high and low variance of the latent factor v along with presence and absence of correlated random drivers provides strong evidence for the adoption of the martingale hypothesis, via the inclusion of the measurement equation within the system state. It is interesting to notice the results of the tests T_2-A and T_3-A, whereby the unit root presence is in the first case rejected and in the second case confirmed. Those results are produced by two different measurement error specifications suggesting that the autocorrelated residuals might be self-induced by the smoothing feature of the filtering device. The optimal strategy to deal with this feature is adopting a martingale form which delivers coherent results for both the experiments.

Filtering a pure diffusion variance path

When we bring a jump component into the system equation, a pure diffusion martingale measurement error might reveal itself to be sometimes inadequate. The update Eq. (13) embeds the information streaming from new observables Y_t at time t^+ into the system state equation (1). This mechanism is very convenient because it provides a straightforward method to construct an approximation of the projection operator $\mathbb{E}_{t|t}$. However, when extending the time-propagation of the projection operator $\mathbb{E}_{t|s}$, Eq. (7), to filtering in the presence of jumps, a new feature related to the update equation arises. We acknowledge that in dealing with a system structure like the experimental model (18), when characterised by a pure diffusion latent factor v, that is $J_1 = 0$, the filter might experience jumps in $\bar{v}$, which are not explained by the solution chosen. The source of these shocks reveals to be enclosed within the mechanism of the update equation, which conveys information coming from two processes that manifest synchronous jumps and propagate from the observed variable Y to the system state estimate $\bar{v}$. We observe that this behaviour is not systemic, because it is found only in a few samples of the same simulation and further it is not strictly related to the jump components, because it pairs with a few but not all large jumps manifesting within the very same sample where it happens. It might be connected to the peculiar form of the model or it might depend on the parametric configuration It is, in fact, those two conjectures that we test with the experiments labelled T_4 and T_5 of Tab. 2. Before discussing the results, we briefly introduce the procedure that we implement to infer conclusions about the suggested explanations of this phenomenon. We construct a local statistical measure to detect a jump in the filtered v path. For the several model structures that have been used in these experiments, the parameter vector of the simulated volatility and stochastic intensity factor v is kept constant. Then, via simulation, a non-central Student-t distribution is fit to the standardised first differences of its path, yielding a degree of freedom parameter of 6.49. Thereafter, the latter parameter is used to simulate the distribution of the maximum of the absolute values of a corresponding Student-t random vector whose size is set to be equal to the size of the sample filter. The idea is to estimate the tail of the absolute value of the first differences of v, in order to test the maximum absolute variation of the filtered data for extreme values. The single tail p-value of the test that has just been designed, indicates an estimate of the probability of rejecting the hypothesis that the standardised max variation can be attributed to a diffusive v, signalling the presence of a jump. This test has been applied to the following experiments and referred to as the *jump-test*.

With the experiment labelled T_4 we test a new form for the measurement error, in relation to the presence of possible jumps induced by the update equation in the pure diffusive latent state variable v, for some extreme experimental cases. We provide evidence that allowing for jumps in the measurement error can attenuate fairly large variations or eliminate unexpected jumps in a state variable whose projection is expected to be a diffusion. The augmentation of the system state with a pure jump martingale introduces a new flexibility within the filtered solution that can handle the unexpected shock to the estimate of v. In Figure 3 we exhibit a typical path of $\bar{v}$ obtained by selecting the system state simulation path with the largest jump in x, which exhibits a shock in the filtered v factor. The filter represented by the square markers has been obtained with unoptimised simulation parameters and without any type of measurement error. The remaining dotted lines with different markers showcase the changes in the output filtered v when a pure jump martingale is introduced and then optimised against the target measure which consists in the standardised maximum variation of the output filter. The continuous line path is the realised v path. An attenuation of the jump phenomenon, whenever it happens can also be obtained by the introduction of a diffusion measurement error, whose parameter is suitably optimised. The test labelled T_4 of Tab. 2 shows the results of the jump-test applied to the extreme jump path selected from a large sample of simulated evolutions of several models characterised by a pure diffusion v latent factor, borderline volatility parameters and large jump frequencies and size. The Tab. 2 columns A and B carry the p-value of, respectively, the test statistics when the measurement error is a latent state variable with the dynamics of a pure diffusion or a pure jump martingale. The general result is that the introduction of a pure diffusion measurement error can in general attenuate the jump event likelihood although mostly the jump event cannot be rejected at a 5% confidence level. The pure jump martingale error introduces further flexibility in the model at the cost of an auxiliary parameter, with the capability of completely eliminating the phenomenon. It is interesting to notice the numbers in the case of the EVD models.

However, the unexpected jump which is seldom observed in extreme cases during simulations, can be removed also with another stratagem. In the sequence of testing T_5 in Tab. 2, we use the AML to estimate the model parameters of several sample paths which produce the largest jumps amongst a large simulated sample, which is generated with the same specification and parameters of the experiment described above, but whose system state and observation equation do not contain any measurement error. We use the jump-test to measure the likelihood of a jump in the $\bar{v}$ path, when the filter is applied at the simulation parameters (test T_5-A). The jump is mostly removed when the model parameters are estimated via the likelihood approach. However, in the EVD case, this approach is not sufficient to adjust for the unexpected jump event, in the peak jump sample paths. A careful use of of the measurement error augmentation might be necessary to obtain efficient latent state system estimates. In fact, the use of a measurement error might be redundant, somehow prefiguring an over-parametrisation of the system under analysis. To support this statement, we show in Fig. 4 the value of the likelihood of the system state, when fixing the model parameters to their full likelihood estimates and when considering the filtered state as it had been observed, that is $\mathbb{P}(X_t|X_{t-1})$. We include a pure jump observation error, with a very low fixed size and varying intensity. What is observed is a likelihood which does not change very significantly, making the need of a measurement error questionable. The chart should be viewed as a sequence of likelihood ratio tests. The explanations as to why a measurement error might possibly be an unnecessary component of the system state equation must be sought in the presence of jumps. Setting the system parameters at their maximum likelihood estimates, the use of procedure to extract jump estimates from the filtered path of π_x, leaves a residual that can be interpreted as an induced measurement error. Eventually, the jump projection absorbs a background noise that should otherwise be incorporated by some other variable.

4.3 Estimation of Stochastic volatility models with jumps

To conclude the empirical section, we use the AML procedure in an exercise which puts more emphasis on the parameter estimation exercise that is finally combined with the proposed jump-diffusion extended filter to provide a measure of the ability of this tool in evaluating the path of latent state variables such as the v factor and the x related jump J_0. In this exercise we also deal with a jump-diffusion version of the latent variable v and briefly discuss the implications from an estimation perspective.

In Table 3 we present the parameter estimation for the selected models. As with the case of filtering, we work with affine jump diffusion models with exponential and Pareto type jump size structure, with state dependent affine intensity function. We extend the set of affine models by including a model with an exponential size jump in volatility which has a v proportional stochastic intensity and is not in sync with the jump in the x level. The latter feature keeps the convolution component of the KFE still single dimensional. Exploiting the PIDE solution type approach to construct the core component of the likelihood function, we are able to deal with the parametrisation of two more jump-diffusions which are not affine in their specification. Again, we include one model with a CEV type diffusion function for the x differential, and another model specification with the jump intensity process specified as the square of the variance factor state component. The models are labelled as in Table 1. The model parameters are similar to the ones adopted in the filtering exercise, when studying the measurement error behaviour. However, in this experiment we lower the number of jumps to balance the total amount of variability split between the diffusive and the jump components. In the case of the affine model with the ancillary jump in volatility, although the jumps are supposed to be generated by two distinct random measures, the jump intensity of the volatility component is set to one tenth of the average jump number in the x level. In the latter case, the model generates a rare jump in volatility, which is very large. The parameter κ is such to generate the same pull to the mean, when compensating for the outward drift produced by the jump sequence through the ω parameter. The parameter η represents the size distribution of J_1, when it refers to the affine model with jumps in volatility. The estimation exercise reveals several interesting features. The estimation is very satisfactory although, depending on the specific sample characteristics of the realised jumps, the diffusion correlation and partly the jump intensity and size might be affected. In the case of the jump intensity, we report quite a high parameter variability for the squared intensity and the jump in volatility models, when compared to the rest of the model suite. Concerning the latter model, we record the tendency to squeeze the jump size while increasing the speed of mean reversion and in general the variability of the diffusion component. It seems that the estimator has the tendency to interpret the jumps in volatility as an increased volatility of volatility component.

As noted at the outset, we finally combine the parametrisation component with the filtering, but this time we consider the ability of the filter to estimate the jump time and investigate the relation of the jumps with the measurement error. In Table 3 column A, we report the basis points[7] daily tracking error (TE) of the filtered state component v with respect to its actual sample path, that is the annualised standard deviation of the percentage deviation of the target and the benchmark path. Those numbers are to be considered very low. It is interesting to notice the EVD sample TE value, which is quite an outlier.

[7] In standard financial jargon, one basis point corresponds to 1/10000 in the relative change of a quantity, that is 100 points are equivalent to a 1% variation of the quantity under analysis.

This is due to a large jump in the filtered path, which can be corrected by the inclusion of a measurement error. Another important output we can achieve with the use of the extended jump-diffusion filter, is the projection of the jump component. In fact, if we consider the estimates of the jump components that affect the x path, we can deduce the jump times and sizes by simply differentiating the π_x, while compensating for the induced drift. The jump events result in the left tail when cutting the realised distribution of the first difference of the jump components filter at their expected sample frequency. Figure 5 exhibits the first difference of the π_x state (solid line) components compared to the realised jumps (circles). The chart highlights the typical pattern of the first difference of the projection of the compensated jump component. The empirical distribution exhibits a skewed shape with a large tail corresponding to the jump events. The cutting of the distribution at the unconditional jump frequency leaves a residual that can be interpreted as a general measurement error that can encompass not only an intrinsic projection error, but that might absorb further error components coming from other system parts. The simulation study shows that this error represents a very limited percentage of the variance of x. The Table 3 column B exhibits the percentage of the realised jump times that can be detected with the mentioned procedure. Tab. 3 column C show the average percentage difference of the estimated jump sizes with respect to the corresponding simulation values. What remains after excluding the jumps from the $\Delta\pi_x$ process is a stationary noise whose variance is confined to a small portion of the total expected variance generated by the diffusion and the jump component. The last column D of Tab. 3 shows the yearly standard deviation of residual first difference of the filtered jump component, after the cut of the tail. Similarly, an estimate of the π_u jump process and some residual can be produced by the filter, whereas in the case of jumps in the latent factor v this procedure is not achievable, because there is no way of disentangling the jump factor via a suitable system design. This is due to the fact that the v jump component does not enter the observation equation. However, it must be noticed, see also Aït-Sahalia (2004), that factors like high jump frequency with low size, unbalanced variance attribution in favour of the diffusion component and very low sampling frequency, will in general increase the jump residuals and therefore render the jump estimation more difficult. Or more simply will increase the aliasing of the sampling behaviour with a pure diffusive model.

5 Conclusion

In this article we have extended to partially observable jump-diffusions the nonlinear filter in Maybeck (1982) and more recently applied in Nielsen et al. (2000), Baadsgaard et al. (2000) and Hurn et al. (2013). Specifically, we extend the time-propagation equation component of the filter to jump processes and analyse the implications for the design of the system, with particular attention to the form of the measurement error. Further, with this article we complement the filtering procedure with an econometric method for the estimation of discretely sampled multivariate jump-diffusions with latent state components that is capable of handling the same model class as the filter and exploits the same PIDE philosophy to deliver its core component. The AML procedure for the estimation of the vector parameter defining the model exploits the classic integration of the latent state, while achieving a simple iteration using a proxy of the solution of the KFE, based on the finite difference approximation of the diffusion operator plus a discretisation of the jump part of the KFE. The latter method has recently received renewed attention in the Financial Econometrics literature, cfr. Lindström (2007), Hurn et al. (2010), Lux (2012). In relation to the approximation of the main component of the likelihood solution, we investigate the stabilisation of the PIDE operator approximation and report a general criterion which provides a guideline for the construction of reliable algorithms. The particular implementation choice simplifies the estimation process saving considerable computation time, rendering the estimation technique appealing from this perspective. Furthermore, the numerical component are reduced to a matrix exponentiation and a few numerical integration, which might be possibly refined and optimised at the programming level.

In the experimental section, we deal with an application of the econometric procedure to a stochastic volatility model with jumps, treating the volatility factor as intrinsically latent. We prove that extending the system state to the integrated variance (see, for instance, Bollerslev and Zhou, 2002), imports a wider information content that improves the estimation of the latent variable path. In our analysis, we focus on the dynamics of the measurement error, which can be best dealt with as an auxiliary latent variable via the augmentation of the system state, cfr. Dempster and Tang (2011). In filtering a system with a pure diffusive v factor, we have found that the jump extension of the filter presented in this article might produce, in extreme situations, unexpected jumps in the projection of v, which are detected with an *ad hoc* testing procedure. We have found that a pure jump martingale measurement error can accommodate this inconvenience. However, further down the lines of investigating the nature of this phenomenon, we have found that the AML optimisation of the system parameters might annihilate the likelihood of a jump in the extreme paths, putting forward a possible explanation about the source of this distortion. Moreover, we collect evidence that in most of the cases under analysis, which are already quite extreme, the auxiliary measurement error component might not be necessary, highlighting the risk of possible over-parametrisation. However, in some cases like the EVD, the use of the latter element might be necessary for the design of an efficient filter. In the final Monte Carlo exercise, we use the AML procedure to estimate the model parameters of the full suite of models described in the Eqs. (18) and (19), including also a jump in the latent factor v. The estimation algorithm produces satisfactory output in a relatively short time. The estimates exhibit high sensitivity to the sample path, especially in the cases of the correlation coefficient and the jump frequency. The AML, when applied to the model with a jump in the v factor, seems to over-weight the variance attributed to the diffusion component as opposed to the jump part. This feature is related to the observation equation lacking the sought jump element. Finally, we use the optimised sample parameters to measure the performance of the filter in estimating the latent variable path for the jump-diffusion model set. The filter is able to produce projections that track the realised latent state very closely. Furthermore, the filter is used to produce estimates of the jump times and sizes, using the tail of the first difference of the filtered x jump component. The residuals of the latter sample data shows how the jump variable incorporates a measurement error that can justify the elimination of the ε variable from the observation equation.

With the application of this method, we are able to exploit the high flexibility it delivers with an exercise that produces system states and parameter estimates and statistics not only for affine jump-diffusions with exponential jumps, but that can deal with non linear models, with possibly non-affine state-dependent jump intensity function and with EVD jump size model, a feature which is prevented with other approaches, (see, for instance Aït-Sahalia, 1996, Filipović et al., 2013, Singer, 2006). With respect to the model parametrisation exercise, the likelihood function which is constructed to estimate the jump-diffusion parameters, takes a particularly simple form. The method proposed can be applied to a large family of stochastic models and can work with information sets which are only partially informative, with respect to the evolutionary phenomenon to be described. Moreover the approach is very flexible and once implemented, the codes can be easily adapted to change the model specification. The model produces reliable estimates at a very contained computational cost.

6 Bibliography

Aït-Sahalia, Y., 1996, Testing continuous-time models of the spot interest rate, *Review of Financial Studies* 9, 385–426.

Aït-Sahalia, Y., 1999, Transition densities for interest rate and other nonlinear diffusions, *Review of Financial Studies* 54, 1361–1395.

Aït-Sahalia, Y., 2002, Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach, *Econometrica* 70, 223–262.

Aït-Sahalia, Y., 2004, Disentangling diffusion from jumps, *Journal of Financial Economics* 74, 487–528.

Aït-Sahalia, Y., 2006, Likelihood inference for diffusions: a survey, in *Frontiers in Statistics*, chapter 17, 369–405 (Imperial College Press, London).

Aït-Sahalia, Y., 2008, Closed-form likelihood expansions for multivariate diffusions, *The Annals of Statistics* 36, 906–937.

Baadsgaard, M., J.N. Nielsen, and H. Madsen, 2000, Estimating multivariate exponential-affine term structure models from coupon bond prices using nonlinear filtering, *Econometric Journal* 3, 1–20.

Bates, D., 2006, Maximum likelihood estimation of latent affine processes, *Review of Financial Studies* 19, 909–965.

Bibby, B.M., and M. Sørensen, 1995, Martingale estimation function for discretely sampled diffusions: a closed-form approximation approach, *Bernoulli* 1, 17–39.

Bollerslev, T., and H. Zhou, 2002, Estimating stochastic volatility diffusion using conditional moments of integrated volatility, *Journal of Econometrics* 109, 33–65.

Chako, G., and L.M. Viceira, 2003, Spectral GMM estimation of continuous-time processes, *Journal of Econometrics* 116, 259–292.

Chen, R.R., and L. Scott, 2003, Multi-factor cox-ingersoll-ross models of the term structure: estimates and tests from a kalman filter model, *Journal of Real Estate Finance and Economics* 27, 143–172.

Christoffersen, P., C. Dorion, K. Jacobs, and L. Karoui, 2014, Nonlinear kalman filtering in affine term structure models, Technical Report 14-04, Centre Interuniversitaire sur le Risque, les Politiques Èconomiques et l'Emploi.

Cont, R., and P. Tankov, 2003, *Financial modelling with jump processes* (Chapman & Hall / CRC Press).

Dempster, M.A.H., and K. Tang, 2011, Estimating exponential affine models with correlated measurement errors: applications to fixed income and commodities, *Journal of Banking and Finance* 35, 639–652.

Duffee, G.R., and R.H. Stanton, 2012, Estimation of dynamic term structure models, *Quaterly Journal of Finance* 2, 1–51.

Duffie, D., and R. Kan, 1996, A yield factor model of interest rates, *Mathematical Finance* 6, 379–406.

Duffie, D., J.P. Pan, and K. Singleton, 2000, Trasform analysis and asset pricing for affine jump-diffusions, *Econometrica* 68, 1343–1376.

Duffy, D.J., 2006, *Finite difference methods in financial engineering: a partial differential equation approach* (Wiley).

Elerian, O., S. Chib, and N. Shephard, 2001, Likelihood inference for discretely observed nonlinear diffusions, *Econometrica* 69, 959–993.

Eraker, B., 2001, MCMC analysis of diffusion models with application to finance, *The Journal of Business & Economic Statistics* 19, 177–191.

Filipović, D., E. Mayerhofer, and P. Schneider, 2013, Density approximations for multivariate affine jump-diffusion processes, *Journal of Econometrics* 176, 93–111.

Gallant, A.R., and G.E. Tauchen, 1996, Which moments to match?, *Econometric Theory* 12, 657–681.

Glosten, L.R., R. Jagannathan, and D.E. Runkle, 1993, On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* 42, 27–62.

Gourieroux, C., A. Monfort, and E. Renault, 1993, Indirect inference, *Journal of Applied Econometrics* 8, 85–118.

Hansen, L.P., and J.A. Scheinkman, 1995, Back to the future: generating moment implications for continuous time markov processes, *Econometrica* 63, 767–804.

Hanson, F.B., 2007, *Applied stochastic processes and control for jump-diffusions: modelling, analysis, and computation* (Society for Industrial and Applied Mathematics).

Heston, S.L., 1993, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *The Review of Financial Studies* 6, 327–343.

Hurn, A.S., J.I. Jeisman, and K.A. Lindsay, 2007, Seeing the wood for the trees: a critical evaluation of methods to estimate the parameters of stochastic differential equations, *Journal of Financial Econometrics* 5, 390–455.

Hurn, A.S., K.A. Lindsay, and A.J. McClelland, 2013, A quasi-maximum method for estimating the parameters of multivariate diffusions, *Journal of Econometrics* 172, 106–126.

Hurn, S., J. Jeisman, and K. Lindsay, 2010, Teaching an old dog new tricks: Improved estimation of the parameters of stochastic differential equations by numerical solution of the fokker-planck equation, in *Financial Econometrics Handbook* (Palgrave, London).

Jensen, B., and R. Poulsen, 2002, Transition densities of diffusion processes: numerical comparison of approximation techniques, *Journal of Derivatives* 9, 18–30.

Jiang, G., and R. Oomen, 2007, Estimating latent variables and jump diffusions models using high frequency data, *Journal of Financial Econometrics* 5, 1–30.

Jones, C.S., 1999, Bayesian estimation of continuous-time finance models, University of Rochester.

Lindström, E., 2007, Estimating parameters in diffusion processes using an approximate maximum likelihood approach, *Annals of Operations Research* 151, 269–288.

Lo, A.W., 1988, Maximum likelihood estimation of Itô processes with discretely sampled data, *Econometric Theory* 4, 231–247.

Lund, J., 1997, Non-linear kalman filtering techniques for term structure models, Working Paper, Aarhus School of Business.

Lux, T., 2012, Inference for systems of stochastic differential equations from discretely sampled data: a numerical maximum likelihood approach, Technical Report 1781, Kiel Institute for the World Economy.

Maybeck, P.S., 1982, *Stochastic models, estimation and control*, volume 2 (Academic Press, London).

Nielsen, J.N., M. Vestgaard, and H. Madsen, 2000, Estimation in continuous-time stochstic volatility models using nonlinear filters, *International Journal of Theoretical and Applied Finance* 3, 1–30.

Øksendal, B., 2003, *Stochastic Differential Equations: An Introduction with Applications*, 6th edition (Springer).

Pedersen, A.R., 1995a, Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes, *Bernoulli* 1, 257–279.

Pedersen, A.R., 1995b, A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations, *Board of the Foundation of the Scandinavian Journal of Statistics* 22, 57–71.

Platen, E., and N. Bruti-Liberati, 2010, *Numerical solution of stochastic differential equations with jumps in finance* (Springer).

Poulsen, R., 1999, Approximate maximum likelihood estimation of discretely observed diffusion processes, Technical Report 29, Centre for Analytical Finance, University of Aarhus.

Singer, H., 2006, Moment equations and hermite expansion for nonlinear stochastic differential equations with application to stock price models, *Computational Statistics* 385–397.

Singleton, K.J., 2001, Estimation of affine asset pricing models using the empirical characteristic function, *Journal of Econometrics* 102, 111–141.

Sørensen, H., 2004, Parametric inference for diffusion processes observed at discrete points in time: a survey, *International Statistical Review* 72, 337–354.

Tavella, D., and C. Randall, 2000, *Pricing financial instruments: the finite difference method* (Wiley).

Van Loan, C.F., 1978, Computing integrals involving the matrix exponential, *IEEE Transaction on Automatic Control* 23, 395–404.

Legenda	Description	Test type A	Test type B
T_0	information significance of the integrated variance	one-tail t-test	one-tail F-test
T_1	MSE of the white noise filter vs. the martingale's	two-tail t-test	two-tail F-test
T_2	residuals of the white noise measurement error	unit-root test	autocorr. residuals
T_3	residuals of the martingale measurement error	unit-root test	autocorr. residuals
T_4	jumps in the pure diffusion $\bar{v}$	j-test diff. err.	j-test jump err.
T_5	jumps in the pure diffusion $\bar{v}$	j-test mod. par.	j-test AML par.
	Description		
AFF	affine jump-diffusion with exponential jump size		
EVD	affine jump-diffusion with EVD jump size		
CEV	affine volatility factor with CEV type diffusion and exponential jump size		
SQI	affine jump-diffusion with exponential jump size and quadratic intensity		
AFJ	affine jump-diffusion with exponential jump in x and v		

Table 1: Description of the labels used in tab. 2 and tab. 3

	Model: Type	θ	κ	σ	ρ	γ	λ	α	η	Sampling	Test type A		Test type B	
T_0	AFF_0	2	1	0.9	-0.6	0.5	-	-	-	large	12.74	(1)	13.17	(1)
T_0	AFF_1	2	1	0.9	0	0.5	-	-	-	large	9.72	(1)	31.86	(1)
T_0	CEV_0	2	1	0.9	-0.6	0.9	-	-	-	large	11.79	(1)	42.68	(1)
T_0	CEV_1	2	1	0.9	-0.6	0.2	-	-	-	large	6.29	(0.99)	8.15	(1)
T_1	AFF_0	2	1	0.9	-0.6	0.5	-	-	-	large	9.55	(0)	1.80	(0)
T_2	AFF_0	2	1	0.9	-0.6	0.5	-	-	-	rand	-7.06	(0)	7e+3	(1)
T_2	AFF_0	2	1	0.9	-0.6	0.5	-	-	-	rand	-4.21	(0)	1e+4	(1)
T_2	AFF_1	2	1	0.9	0	0.5	-	-	-	rand	-4.26	(0)	1e+4	(1)
T_2	AFF_1	2	1	0.9	0	0.5	-	-	-	rand	-7.46	(0)	5e+3	(1)
T_2	AFF_2	2	1	0.3	0	0.5	-	-	-	rand	-4.09	(0)	1e+4	(1)
T_2	AFF_2	2	1	0.3	0	0.5	-	-	-	rand	-3.57	(0)	1e+4	(1)
T_3	AFF_0	2	1	0.9	-0.6	0.5	-	-	-	rand	-0.60	(0.43)	85.86	(0)
T_3	AFF_0	2	1	0.9	-0.6	0.5	-	-	-	rand	-1.54	(0.12)	49.06	(0)
T_3	AFF_1	2	1	0.9	0	0.5	-	-	-	rand	-0.32	(0.54)	31.36	(0.05)
T_3	AFF_1	2	1	0.9	0	0.5	-	-	-	rand	-2.21	(0.03)	97.32	(0)
T_3	AFF_2	2	1	0.3	0	0.5	-	-	-	rand	-1.06	(0.26)	93.53	(0)
T_3	AFF_2	2	1	0.3	0	0.5	-	-	-	rand	-1.55	(0.11)	71.26	(0)
T_4	AFF_3	2	1	0.9	0	0.5	30	1	-	peak	8.45	(0.12)	8.14	(0.15)
T_4	AFF_3	2	1	0.9	0	0.5	30	1	-	peak	11.86	(0.02)	8.38	(0.13)
T_4	EVD	2	1	0.9	0	0.5	30	3.10	4.1	peak	30.05	(0)	8.41	(0.12)
T_4	EVD	2	1	0.9	0	0.5	30	3.10	4.1	peak	23.00	(0)	9.49	(0.06)
T_4	CEV_2	2	1	0.9	0	0.9	30	1	-	peak	7.44	(0.24)	7.65	(0.21)
T_4	CEV_2	2	1	0.9	0	0.9	30	1	-	peak	8.83	(0.09)	5.60	(0.75)
T_4	SQI	2	1	0.9	0	0.5	14	1	-	peak	12.39	(0.01)	5.95	(0.63)
T_4	SQI	2	1	0.9	0	0.5	14	1	-	peak	12.34	(0.01)	5.99	(0.61)
T_5	AFF_3	2	1	0.9	-0.6	0.5	30	0.5	-	peak	21.65	(0)	6.52	(0.45)
T_5	AFF_3	2	1	0.9	-0.6	0.5	30	0.5	-	peak	15.41	(0.00)	7.01	(0.32)
T_5	EVD	2	1	0.9	-0.6	0.5	30	1.55	4.1	peak	22.37	(0)	13.92	(0.00)
T_5	EVD	2	1	0.9	-0.6	0.5	30	1.55	4.1	peak	34.76	(0)	15.83	(0)
T_5	CEV_2	2	1	0.9	-0.6	0.9	30	0.5	-	peak	25.07	(0)	8.07	(0.15)
T_5	CEV_2	2	1	0.9	-0.6	0.9	30	0.5	-	peak	20.93	(0)	9.75	(0.05)
T_5	SQI	2	1	0.9	-0.6	0.5	14	0.5	-	peak	28.96	(0)	9.39	(0.07)
T_5	SQI	2	1	0.9	-0.6	0.5	14	0.5	-	peak	9.81	(0.05)	6.53	(0.44)

Table 2: The table contains the test statistic results of the diagnostic check for the structuring of the measurement error process and the inclusion of the new variable, the integrated variance, into the system state equation of the stochastic volatility model (18) used in the econometric application. The description of the labels used within the table are reported in tab. 1.

		θ	κ	σ	ρ	λ	α	η	γ	A	B	C	D
	Mod.	2	1	0.9	-0.6	10	0.5						
AFF	Est.	1.914	2.810	0.884	-0.596	13.784	0.561			4.8	92%	-16%	0.7
	Std	(0.191)	(0.820)	(0.236)	(0.204)	(1.747)	(0.048)						
	Mod.	2	1	0.9	-0.6	10	1.55	4.1					
EVD	Est.	1.691	1.275	1.058	-0.230	6.183	2.266	4.918		9.7	76%	-16%	0.9
	Std	(0.495)	(0.852)	(0.167)	(0.642)	(3.084)	(0.965)	(1.081)					
	Mod.	2	1	0.9	-0.6	10	0.5		0.8				
CEV	Est.	1.997	1.006	1.003	-0.208	11.116	0.748		0.646	6.7	124%	-18%	0.8
	Std	(0.347)	(0.603)	(0.183)	(0.171)	(3.393)	(0.282)		(0.103)				
	Mod.	2	1	0.9	-0.6	7.11	0.5						
SQI	Est.	2.106	2.141	0.665	-0.658	13.290	0.463			4.1	101%	12%	0.7
	Std	(0.404)	(1.863)	(0.153)	(0.254)	(9.258)	(0.117)						
	Mod.	2	2	0.9	-0.6	10	0.5	1					
AFJ	Est.	1.851	3.958	1.416	-0.081	13.777	0.615	0.101		6.5	72%	2%	1.1
	Std	(0.172)	(4.059)	(0.251)	(0.175)	(9.619)	(0.279)	(0.178)					

Table 3: The table contains the model parameters, their estimated values and their estimated standard deviations. For all the models the parameter λ indicates the coefficient in the state dependent process, which is always affine, expect in the case of the squared intensity process. For all the models with an exponential jump in the x level, the coefficient α represent the expected jump size, whereas in the EVD jumps affine model, α and η are the coefficients characterising the jump distribution. Finally, in the case of the affine model with an exponential jump in the v factor, the coefficient η represents the average jump size in v, while the jump intensity factor is assumed to be $\lambda_1 = \lambda/10$. The column A represents a measure of the efficiency of the nonlinear filter in estimating the most relevant system state component, that is the v factor. That column carries the daily tracking error in basis points of the filtered against the realised sample path of v. The column B contains the percentage of the realised jump times in the x levels that are filtered via the system state component π_x. The column C presents the average percentage jump size error, at the estimated jump times, with respect to the corresponding first difference of J_0. The column D exibits the yearly standard deviation of the residuals of first difference of the filtered J_0, once the tail has been removed, with the purpose of estimating the system jumps.

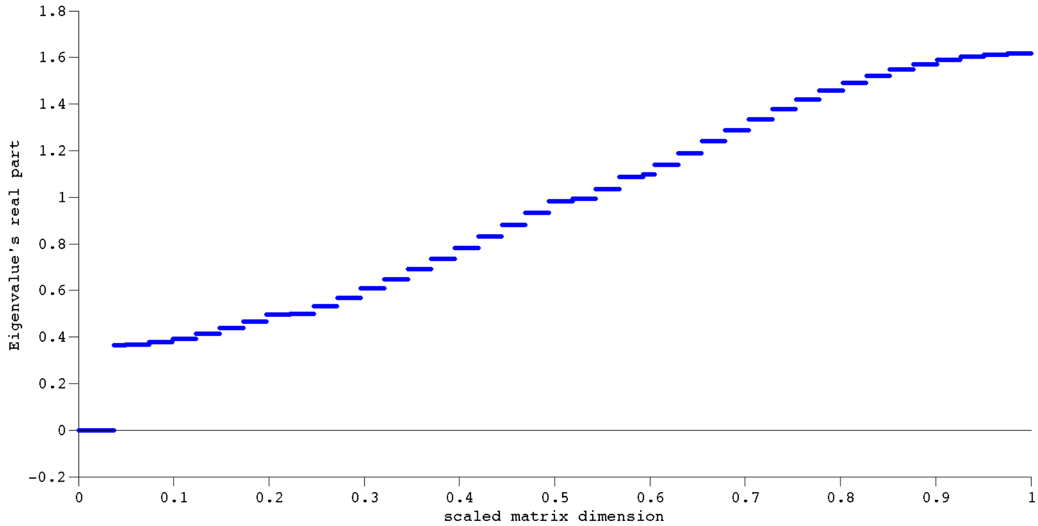

Figure 1: Ordered real part of the FDM matrix for the operator ∂_x, constructed with predominant backward approximation. The order of the approximation is first, which allow the minimum amount of grid points involved. The size of the matrix has been scaled to unity.

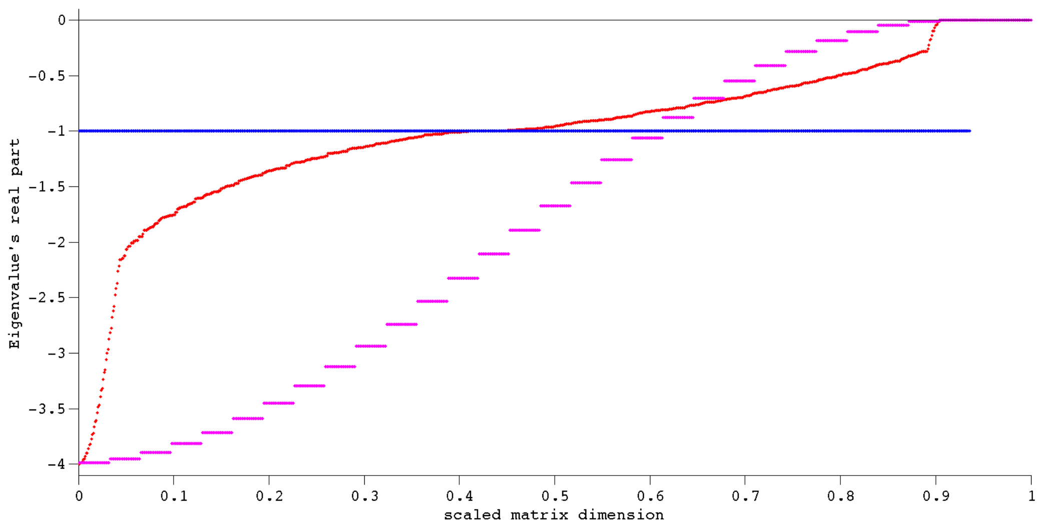

Figure 2: Ordered real part of the FDM matrix for the operator ∂_x, ∂^2_{xx}, ∂^2_{xv} constructed, respectively, with predominant forward approximation, central approximation and mixed upper-right and lower-left corner. The order of the approximation is first and second, which allow the minimum amount of grid points involved. The size of the matrix has been scaled to unity.

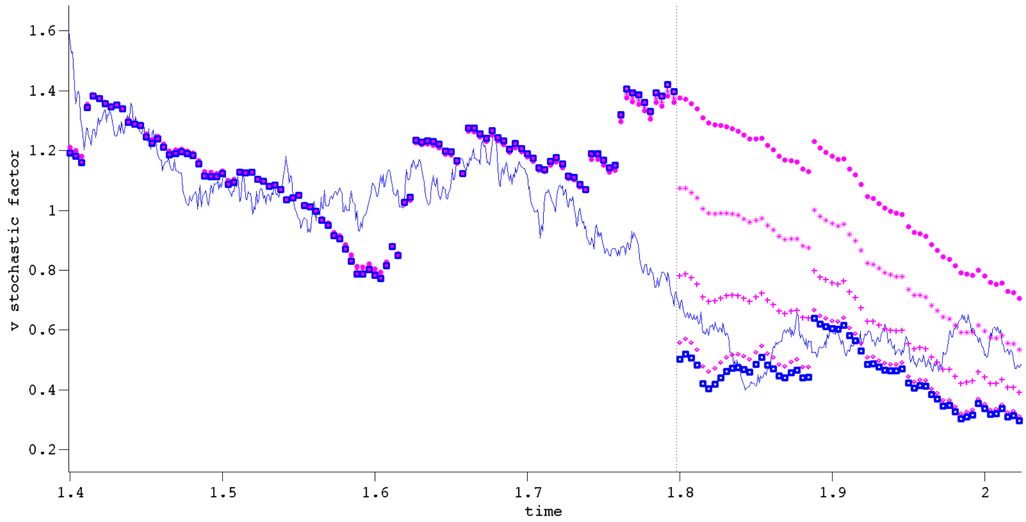

Figure 3: The figure shows the typical pattern of a shock in the filtered path of the volatility factor. The realised v factor (solid line) is compared to a non-linear filter with a pure jump measurement error (dotted lines) when triggering the jump intensity up to its optimised level (circle marks). Introducing jumps in the measurement error provide more flexibility to the system and allow the absorption of possible shocks in the latent state estimates.

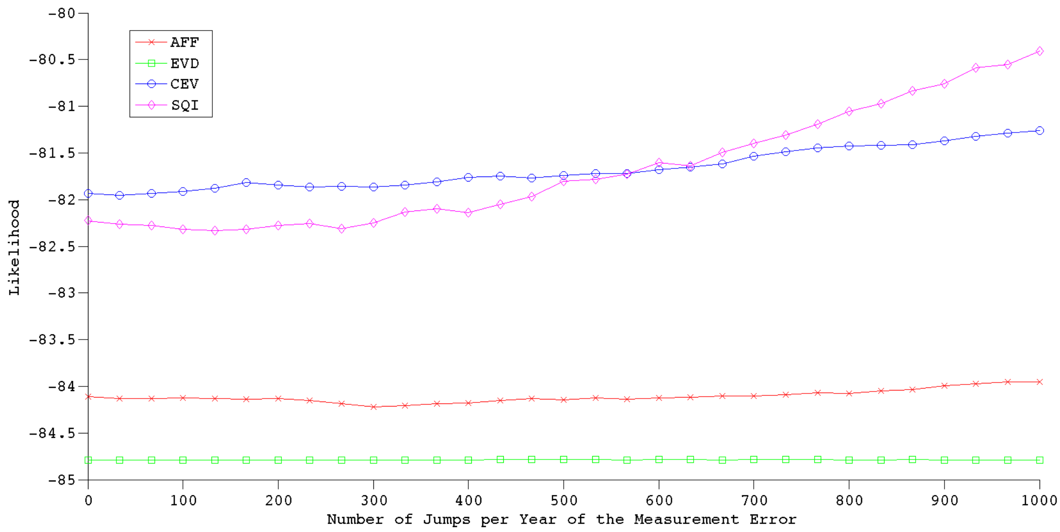

Figure 4: This figure shows the value of the log-likelihood of the filtered system state X as if it was observed. The model parameters are optimised with full AML estimation and then kept constant while the system state is estimated thorough the nonlinear filter. The system state includes a pure jump measurement error, with a jump size set to 0.1, while the jump intensity runs from 0 to 100. The various curves corresponds to the experimental section's model. What is interesting here is the complete absence of a trend and the very limited variability of the target function values.

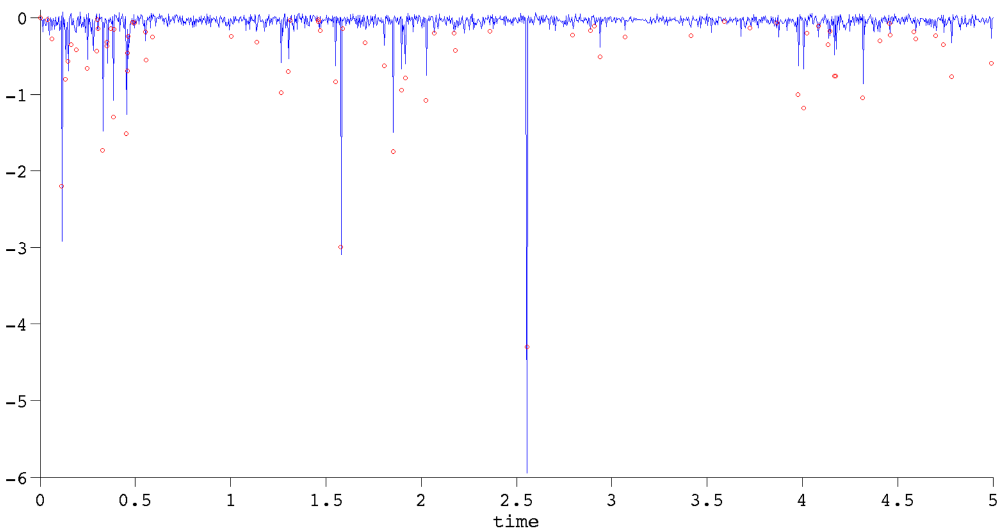

Figure 5: This figure shows the first difference of the filtered π_x path for the AFF model. The red circles indicates the realised jumps in the x level.

A Proof of several Propositions

A.1 Proposition 2.2

Proof. Recall the *Forward Kolmogorov Equation*

$$\partial_t p = \tfrac{1}{2}\sum_{ij}\partial^2_{ij}\left[C_{ij}p\right] - \sum_i \partial_i\left[b_i p\right] \tag{21}$$

Where p represents the state transition density. Therefore, we can take the time derivative of the expectation and combine with (21)

$$\begin{aligned}\bar{X}_{t|s} &= \int \mathrm{d}x\, p_{t|s}X \Rightarrow \tfrac{\mathrm{d}}{\mathrm{d}t}\bar{X} = \int \mathrm{d}x\, \partial_t p X = \\ &\tfrac{1}{2}\int \mathrm{d}x \sum_{ij}\partial^2_{ij}\left[C_{ij}p\right]X - \int \mathrm{d}x \sum_i \partial_i\left[b_i p\right]X \end{aligned}\tag{22}$$

and simplify the expression. In fact, considering the generic component of the last term and integrating by parts, we obtain

$$\begin{aligned} -\int \mathrm{d}x\, \partial_i\left[b_i p\right]X &= -\int\dots\int \mathrm{d}x_1\dots \mathrm{d}x_{i-1}\mathrm{d}x_{i+1}\dots\mathrm{d}x_{\mathrm{n}}\int \mathrm{d}x_i\,\partial_i\left[b_i p\right]X = \\ &\int \mathrm{d}x\,\left[b_i p\right]\partial_i X = \int \mathrm{d}x\,\left[b_i p\right]e_i = e_i\mathbb{E}b_i. \end{aligned}$$

Similarly, integrating by parts the first term we obtain

$$\int \mathrm{d}x\, p\sum_{ij}C_{ij}\partial^2_{ij}X = 0$$

because $\partial^2_{ij}X = \partial_i e_j = 0$.

In order to obtain the the evolutionary equation for the covariance of the $\bar{X}$, we consider the expression

$$V_{t|s} = \mathbb{E}_{t|s}\left[XX^\star\right] - \bar{X}_{t|s}\bar{X}^\star_{t|s}$$

and obtain

$$\tfrac{\mathrm{d}}{\mathrm{d}t}V_{t|s} = \tfrac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[XX^\star\right] - \tfrac{\mathrm{d}}{\mathrm{d}t}\left[\bar{X}\right]\bar{X}^\star - \bar{X}\tfrac{\mathrm{d}}{\mathrm{d}t}\left[\bar{X}\right]^\star$$

Now, combining (21) with the $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[XX^\star\right]$ and with the same argument as above we obtain

$$-\int \mathrm{d}x \sum_i \partial_i\left[b_i p\right]XX^\star = -\int \mathrm{d}x\, p\sum_i b_i\partial_i XX^\star$$

but $\partial_i XX^\star = e_i X^\star + Xe_i^\star$, hence

$$\begin{aligned} -\int \mathrm{d}x\sum_i \partial_i\left[b_i p\right]XX^\star &= -\int \mathrm{d}x\, p\left(bX^\star + Xb^\star\right) = \\ &\mathbb{E}\left[bX^\star\right] + \mathbb{E}\left[Xb^\star\right] \end{aligned}$$

The last component of the expression for $\frac{\mathrm{d}}{\mathrm{d}t}V_{t|s}$ is

$$\tfrac{1}{2}\int \mathrm{d}x\sum_{ij}\partial^2_{ij}\left[C_{ij}p\right]XX^\star = \tfrac{1}{2}\int \mathrm{d}x\, p\sum_{ij}C_{ij}\partial^2_{ij}\left[XX^\star\right]$$

but

$$\sum_{ij}\partial_{ij}XX^\star = \sum_{ij}\partial_i\left(e_jX^\star + Xe_j^\star\right) = \begin{bmatrix} 2 & 1 & \dots & 1 \\ 1 & 2 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 2 \end{bmatrix}$$

hence

$$\tfrac{1}{2}\int \mathrm{d}x\sum_{ij}\partial^2_{ij}\left[C_{ij}p\right]XX^\star = \mathbb{E}C.$$

$\square$

A.2 Proposition 2.3

Proof. We notice that in (4) it is more convenient to revert back to the pre-jump transform $H(X^+) = X^- \Longleftrightarrow X^+ = (I+J)(X^-)$ when calculating the first two moments of the state; using the distributional equality $(\mathcal{J}^\star[u], v) = (u, \mathcal{J}[v])$, we obtain the integrals $\int (X^- + J^-)\mathcal{J}^\star[p]\mathrm{d}X$ and $\int (X^- + J^-)(X^- + J^-)^\star \mathcal{J}^\star[p]\mathrm{d}X$. After some calculus, using the moment integrals, we get the sought jump component of the time propagation equation. □

A.3 Proposition 2.4

Proof. Following Maybeck (1982), we define two functions of the state vector X and the observed vector Y, $\psi(X)$ and $\theta(Y)$ and, applying a version of the iterated expectations

$$\mathbb{E}_{t|s}\left[\psi(X)\theta(Y)\right] = \mathbb{E}_{t|s}\left[\mathbb{E}_{t|t}\left[\psi(X)\right]\theta(Y)\right] \tag{23}$$

To obtain the 13 we assume the form

$$\begin{aligned}\bar{X}_{t|t} &= a_t + A_t\left(Y_t - \bar{Y}_{t|s}\right)\\ \bar{V}_{t|t} &= \Sigma_t\end{aligned}$$

The update equation can therefore be obtained by defining appropriately the functions ψ and θ and then plugging the definitions into the 23. The term a_t is obtained by letting $\psi = X_t - \bar{X}_{t|t}$ and $\theta = 1$, whereas $\psi = X_t - \bar{X}_{t|t}$ and $\theta = \left(Y_t - \bar{Y}_{t|s}\right)^\star$ entails that the 23 can be solved for A_t. The matrix Σ_t can be obtained with $\psi = \left(X_t - \bar{X}_{t|t}\right)\left(X_t - \bar{X}_{t|t}\right)^\star$ and $\theta = 1$ and substituting the definition of $\bar{X}_{t|t}$ in the RHS. □

A.4 Proposition 3.1

Proof. Assume that $f^h \to f$ pointwise and $f \geq 0$, $\forall x$. Recalling the definition of a limit, with $\varepsilon > 0$,

$$\lim_{h\to\infty} f^h(x) = f(x) \Leftrightarrow \exists k\colon \left|f - f^h\right| < \varepsilon \;\forall h > k$$

The variable k depends in general on x. Now, because $-f \geq -\left|f^h\right|$ and $f \geq 0$, we have

$$\varepsilon > \left|f - f^h\right| \geq \left|f - \left|f^h\right|\right| \geq 0$$

and the proposition is proved. □

A.5 Calculus: non-central moments of the EVD

Consider the distribution of the $\mathbb{P}\left[J \leq x\right] = 1 - \left(1 + x/\alpha\right)^{-\eta}$, that is

$$\frac{\eta}{\alpha\left(1 + \frac{x}{\alpha}\right)^{1+\eta}} \tag{24}$$

after the change of variable $x/\alpha = z$ and disregarding the constant α^n, we need to calculate the integral

$$\eta \int_0^{+\infty} \mathrm{d}z \frac{z^n}{(1+z)^{1+\eta}}$$

for $\eta > n$ it can be proved by induction that this integral is equal to

$$\frac{n!}{(\eta - n)\ldots(\eta - 1)}$$

This is true for $n = 1$, assuming it true for n, we see that

$$\int_0^{+\infty} \mathrm{d}z \frac{z^{n+1}}{(1+z)^{1+\eta}} = \frac{n+1}{\eta}\int_0^{+\infty} \mathrm{d}z \frac{z^n}{(1+z)^{1+\eta}}$$

and letting $\bar{\eta} = \eta - 1$ we get

$$\frac{(n+1)n!}{(\bar{\eta}+1)(\bar{\eta} - n)\ldots(\bar{\eta} - 1)\bar{\eta}}$$

substituting back $\bar{\eta}$ and multiplying by η we get the sought result.