Inference for Latent Affine Jump-Diffusions and Its Application to the Option Market

Zur Erlangung des akademischen Grades eines

Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

von der KIT-Fakultät für

Wirtschaftswissenschaften

des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

M.Sc. Lingjie Ni

Referent:	Prof. Dr. Maxim Ulrich
Korreferentin:	Prof. Dr. Marliese Uhrig-Homburg
Tag der mündlichen Prüfung:	24. April 2024

Karlsruhe 2024

Acknowledgements

I would like to express my deep gratitude to my advisor, Professor Dr. Maxim Ulrich, for his unwavering support, enriching discussions, valuable advice, and the academic freedom he provided, which allowed me to explore my ideas fully. Many thanks also to Professor Dr. Marliese Uhrig-Homburg, Professor Dr. Fabian Krüger, and Professor Dr. Andreas Oberweis for serving on my thesis committee.

I am grateful to all my colleagues at the Chair of Financial Economics and Risk Management, as well as the students of C-RAM. Special thanks to Dr. Elmar Jakobs, Dr. Stephan Florig, Dr. Simon Walther, Lukas Zimmer, Ralph Seehuber, and Maximilian Kübler for their insightful discussions and constant encouragement, which have been pivotal in making the past years deeply rewarding.

Additionally, I would like to thank my wife Tian for her boundless patience and continual encouragement. Finally, heartfelt thanks to my parents, whose persistent support has been my foundation throughout my life.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Dynamic asset pricing theory based on the constraints of equilibrium and arbitrage provides a unifying principle to determine the functional relationship between asset prices and fundamental economic characteristics, such as state variables, structural parameters, and market prices of risk. Continuous-time stochastic processes, which are used to model fluctuations in asset prices, play a central role in this approach because of their analytical tractability. As a general class of multivariate stochastic processes, affine jump diffusions (AJDs) have demonstrated unprecedented success over the past three decades due to their superior flexibility and tractability. By selecting proper building blocks for volatility and jumps, the AJD processes can accommodate a broad range of statistical features that are consistent with many empirically observed patterns. The analytical tractability stems from the fact that the conditional characteristic function of the selected processes can be derived by solving a set of ordinary differential equations (ODEs). The use of dynamic asset pricing models with AJDs (AJD models[1]) dates back to the 1970s, as discussed by Piazzesi (2010) and Garcia et al. (2010). There is an extensive literature on term structure models that begins with the work by Vasicek (1977), who proposes a one-factor model for short-rate in order to obtain a closed-form formula for bond yields. In the context of pricing equity options, the seminal work by Black and Scholes (1973) (BS) uses geometric Brownian motion as the dynamic of the underlying asset price to establish the closed-formula for pricing European options. As empirical evidence resoundingly rejects the assumption of constant volatility in the BS model, Heston (1993) proposes a stochastic volatility (SV) model while still providing an analytical option pricing formula. Duffie and Kan (1996) discover that the SV model by Heston is a special case of the AJD model,

[1]For notational simplicity, the AJD models are restricted to dynamic asset pricing models with AJD processes within this thesis.

for which a similar analytical pricing formula can be obtained. Thereafter, the general class of AJD models is subsequently examined by Duffie et al. (2000), who establish a framework for determining the price of equity and fixed income derivatives. By using the general solution from Duffie et al. (2000)'s work, advanced specifications of volatility and jumps can be added to the equity derivative pricing models to further ameliorate their empirical fit. For the volatility part, many studies use multiple volatility factors beyond Heston's SV model (see, e.g., Chernov et al. (2003), Andersen et al. (2015b), among others). More recently Gruber et al. (2021) uses a matrix state process to model volatility. As a result of the work of Merton (1976), jumps in the return dynamic are often defined as a compound Poisson process with a normal distribution of jump sizes and a constant jump intensity. Nevertheless, Kou (2002) proposes a specification of separate positive and negative jumps to fit the tails on both sides of the return dynamics. Thus, there are two jumps with exponentially distributed sizes and constant intensities in both directions. Moreover, stochastic jump intensity can be used to capture the phenomenon of self-existing jump clustering, which has been shown to be critical in modeling the 2008 crash. (see, e.g., Fulop et al. (2015), Andersen et al. (2015b) and Andersen et al. (2020)). It is noteworthy to stress that the aforementioned AJD models for the equity indices are mainly used in determining the price of derivatives written on this index.

Following Pan (2002), Eraker (2004) and Broadie et al. (2007), among others, this thesis uses a well-established approach in the literature to specify market prices of risk for dynamic asset pricing models based on the no-arbitrage framework. In this approach, the pricing kernel does not directly enter the state variables, while it is derived from a pre-defined functional relationship to a group of relevant risk factors which includes the risk-free rate, diffusive return shocks, volatility shocks and jumps. Moreover, risk premiums are defined as a set of parameters that determine compensation for various risk factors. Then, a vector of structural parameters of interest for asset pricing can be assembled by adding the risk premium parameters together with the statistical parameters for the dynamics of the risk factors under physical measure ($\mathbb{P}$). As the other side of the coin, the "fundamental theorem of asset pricing" proposed by Harrison and Kreps (1979), implies the equivalence between the existence of an equivalent martingale measure ($\mathbb{Q}$) and the absence of arbitrage. By assuming that the dynamics of asset pricing models under $\mathbb{P}$ and $\mathbb{Q}$ follow the same structure, the risk premium parameters can be interpreted as the "difference" between the statistical parameters of the dynamics under $\mathbb{P}$ and $\mathbb{Q}$. In the context of derivative pricing, statistical observations of option prices provide insight

into the underlying equivalent martingale measure. Therefore, option surfaces allow researchers to perform some statistical inference about the parameters under $\mathbb{Q}$.

The aim of an empirical analysis of dynamic asset pricing models is usually to extract information about state variables, structural parameters, and market prices of risk from observed prices. In many cases, state variables are intrinsically latent in the context of underlying asset prices, even when the latter are sampled at ultra-high frequency. More specifically, the centrepiece of the empirical analysis conducted with AJD models is to estimate the models with observed prices, which entails two major tasks: the inference of static statistical parameters and the recovery of time-varying dynamic states. One popular statistical solution to estimation problems in the literature is to use filtering methodologies based on a state-space model (SSM) representation. In general, the SSM describes the probabilistic dependence between the latent state variables and the observed measurements. By using standard discretization methods, the dynamics of AJD models can be transformed from continuous-time stochastic processes into discretized transition equations. When coupled with an error assumption for the observed measurements, one can translate the continuous-time models into an SSM interpretation. Therefore, the issue that measurements are only sampled discretely, whereas theoretical models stipulate that prices and state variables fluctuate continuously in time, is resolved. After that, one can use the appropriate filtering method to obtain the latent state estimates and the likelihood of the SSM given the observations from the market. By assuming Gaussian measurement errors and a Gaussian approximation of transition densities, the term structure models with AJD processes can be translated into a linear Gaussian SSM that can be tackled via Kalman filtering. The Kalman filter, on the other hand, cannot be directly applied to an empirical analysis of the option market, as latent states are connected to pricing models nonlinearly or even only in a semi-closed form. Therefore, particle filters, which are simulation-based filtering methods, are required for nonlinear and non-Gaussian SSMs. Furthermore, a number of studies use higher order moments of risk-neutral densities as additional informative measurements in the SSM-based approaches, including Feunou and Okou (2018) and Fulop and Li (2019). As an alternative to the SSM-based solution, Andersen et al. (2015a) leverage the rich information from the option surfaces to address the inference problem via a penalized nonlinear least squares (PNLS) approach that estimates the risk-neutral parameters and recovers the latent states. In this thesis, inspired by the PNLS approach in the work by Andersen et al. (2015a), I present a Kalman filtering-based quasi-maximum likelihood (QML)

approach. Using this approach, it is possible to estimate AJD models in the empirical analysis of the option market rapidly and efficiently by incorporating linear observations of higher order moments from the risk-neutral density.

In the last three decades, AJD models have been used in numerous empirical studies across various asset classes. In particular, AJD models can accurately identify the risk-neutral density implied by derivatives written on equity index (see, e.g., Broadie et al. (2007), Bardgett et al. (2019), among others). In recent empirical studies, Andersen et al. (2015b) and Andersen et al. (2020) investigate the relationship between the left tail factor and the equity risk premium using the AJD model with a stochastic negative jump intensity. Using forward-looking information from option surfaces, they find that the left tail factor is strongly correlated with equity premiums in the U.S. and European markets. In particular, negative jump intensity, as the left tail factor, can predict future equity returns over the medium- and long-term. In addition, they argue that the left tail factor provides economic information that is distinct from the regular volatility factor. This is because the volatility factor is only a potent predictor of actual future return variations, but does not explain future equity returns. Recently, the increased liquidity of derivatives markets for single stocks makes it possible to accurately estimate the cross-sectional risk-neutral densities of underlying asset returns as well as to identify the time-varying factors that represent the key statistical characteristics of complex fluctuations in the conditional risk-neutral density. In this thesis, the empirical analysis of options written on single stocks is conducted using an AJD model with specifications similar to the model of Andersen et al. (2020), who employ a separate process specifying stochastic negative jump intensity as the left tail factor. To the best of my knowledge, Bégin et al. (2020) are the only researchers, who have conducted a parametric study of cross-sectional option surfaces. Nevertheless, in their analysis, the positive and negative jumps are not modeled separately. Hence, the empirical study in this thesis not only demonstrates the robustness and efficiency of the Kalman filter-based QML estimation approach, but also fills a gap in the empirical analysis of the left tail risk embedded in cross-sectional option data using a parametric model.

1.2 Structure of the Thesis

This thesis begins with a review of the mathematical fundamentals of dynamic asset pricing models with AJDs. Then, a discussion of the inference methods used in the

literature for AJDs follows. In addition to the discussion, a Kalman filter-based QML estimation approach that integrates existing techniques for inference of AJD processes in the empirical analysis of dynamic asset pricing models is introduced. Finally, an empirical analysis of dynamic asset pricing models based on AJDs with application to index and stock options is carried out. In more detail, this thesis is structured as follows:

Chapter 2 first introduces the general definition of AJD processes. Using the results of Duffie et al. (2000), a general transformation of the discounted characteristic function of AJD processes can be obtained by solving a set of ordinary differential equations (ODEs). Furthermore, the higher order moments of AJD-implied densities can be determined by solving additional sets of ODEs. Within the arbitrage-free framework, one can retrieve the fair market price of a derivative contract by calculating its expectation of discounted future payoff under the risk-neutral measure. Using the inverse Fourier transformation, one can compute the expectation since the characteristic function of the model dynamics is feasible based on the results from Duffie et al. (2000). In accordance with Andersen et al. (2015b) and Bardgett et al. (2019), the methodology from Fang and Oosterlee (2008) for computing the inverse Fourier transformation via cos-expansions is reviewed. Furthermore, several essential model dynamics are discussed in detail, which can be viewed as building blocks for later empirical analysis. Finally, the non-parametric risk-neutral moments embodied in the option surfaces are discussed.

Chapter 3 reviews the widely used methodologies for inferring AJD processes in the literature. This chapter focuses on the SSM-based inference approach. First, a general definition of SSM is introduced. A discussion of two types of filtering techniques follows, namely the Kalman filter and the particle filter. As the most critical technique for subsequent empirical analysis, the Kalman filter is explained in detail. Furthermore, the PNLS approach by Andersen et al. (2015a) is discussed as it serves as an inspiration for the approach in the empirical study later. Finally, the inference methods from different resources are combined, resulting in a Kalman filter-based QML approach. By using higher order moments, this approach replaces the minimization-based estimates of the latent states in the PNLS approach with the posterior estimates from Kalman filtering.

Chapter 4 validates the soundness of the aforementioned QML approach in extracting latent states and estimating parameters under $\mathbb{Q}$. The validation is based on synthetic data obtained from a Monte Carlo simulation.

Chapter 5 performs an empirical analysis of an AJD model with a left tail factor. The

first application uses the surfaces of options written on the S&P 500 index as input data. Numerous researchers have explored this application in the literature using various estimation techniques. In order to validate the aforementioned QML approach empirically, several diagnostics following the literature are conducted in this application. The extracted trajectories of the negative jump intensity are consistent with the results from Andersen et al. (2015b) and Andersen et al. (2020). In the subsequent application, option surfaces written on single stocks from the S&P 500 index are used. The purpose of this study is to fill a gap in the literature regarding parametric approaches aimed at characterizing the economic implications of the cross-section of left tail factors. Furthermore, it demonstrates that the cross-section of left tail factors contains information about the equity risk premium. The application also entails decomposing both spot variance and left tail factors into systematic and idiosyncratic components. This analysis reveals that the idiosyncratic components of these factors primarily drive the equity risk premium.

Finally, Chapter 6 summarizes the main results of this thesis and briefly discusses future research directions.

2 Mathematical Fundamentals

It is widely recognized that empirical facts do not support the widely used Black-Scholes option pricing model among academicians and investment professionals because it assumes constant volatility and log-normality of future risk-neutral return distributions. Bates (2000), for instance, shows that the post-crash risk-neutral return distributions of 1987 are negatively skewed and leptokurtic. Therefore, the single Brownian motion is inadequate for modeling asset returns in the equity derivatives markets. To overcome this challenge, one needs more advanced tools in econometrics. In particular, stochastic volatility and jumps are integrated into model dynamics as natural extensions of asset return modeling in a multivariate setting. These extensions lead to a general class of stochastic processes known as affine jump-diffusions, which play a key role in dynamic asset pricing models. According to the literature on empirical asset pricing, AJDs are commonly used to build dynamic asset pricing models in reduced form based on the absence of arbitrage constraints on asset prices. The reduced models, as discussed by Garcia et al. (2010), do not directly establish a link between the representative agent's preference and the risk premiums associated with various risks. Researchers usually treat these risk premiums as free parameters in order to capture the empirical facts reflected in the data.

This chapter presents the mathematical fundamentals of AJDs for building the reduced-form dynamic asset pricing model and relevant technicalities for the empirical analysis of option markets. More specifically, this chapter is organized as follows: Section 2.1 discusses the general definition of AJDs and their transformations for derivative pricing. Section 2.2 illustrates the popular AJD models for option pricing in reduced form. Section 2.3 discusses the model-free risk-neutral moments, as derived by Bakshi et al. (2003), and their transformation into cumulants.

2.1 The Affine Jump-Diffusions

In general, AJDs are a class of stochastic processes, consisting of affine stochastic processes with jumps, in a continuous-time setting. Following Duffie et al. (2000), the general definition of AJDs is that the process X_t exists in some state space $D \subseteq \mathbb{R}^n$ and is given by the solutions to the stochastic differential equations (SDEs) as follows:

$$dX_t = \mu\left(X_t\right)\, dt + \sigma\left(X_t\right)\, dW_t + dJ_t\,, \tag{2.1}$$

where the drift term $\mu\left(X_t\right)$, the volatility term $\sigma\left(X_t\right)$, and the jump term J_t preserve a certain affine relation to the state X_t. The detailed definition can be found in Definition A.1

For a concrete example of an AJD model for option pricing, please refer to Illustration A.1, which shows the data generating processes of the model proposed by Bates (1996). This example illustrates how to establish a connection between the theoretical definition of AJD and its application in practice. For more details, see Appendix A.3.1.

2.1.1 Transform Analysis

Following Definition A.1, the (conditional) "discounted" characteristic function ψ with instantaneous riskless short-term rate $r(X_t) : \mathbb{C}^n \times D \times [0, +\infty) \times [0, +\infty) \to \mathbb{C}$, which determines the expectation of a transformation of $X_{t+\tau}$ conditional on $\mathcal{F}_t$ is given with $\tau > 0$, as follows:

$$\psi(u, X_t, t, \tau; \theta^{\mathcal{X}}) := \mathbb{E}\left[\exp\left(-\int_t^{t+\tau} r\left(X_s\right) ds\right) e^{u \cdot X_{t+\tau}} \middle| \mathcal{F}_t\right]. \tag{2.2}$$

Using an AJD model as a basic structure is appealing because the discounted conditional characteristic function is *log-affine* in the state vector.

Following Duffie et al. (2000) and Chen and Joslin (2012), among others, the key to the transform analysis for the "discounted" characteristic function ψ is that

$$\psi(u, X_t, t, \tau; \theta^{\mathcal{X}}) = e^{\alpha(\tau) + \beta(\tau) \cdot X_t}, \tag{2.3}$$

where $\alpha : [0, \tau] \to \mathbb{C}$ and $\beta : [0, \tau] \to \mathbb{C}^n$ are the deterministic functions satisfying the

complex-valued ODEs with[2]:

$$\begin{aligned}\dot{\alpha}(s) &\stackrel{!}{=} -r_0 + k_0 \cdot \beta(s) + \frac{1}{2}\beta(s)^\top h_0 \beta(s) + c_0(\varsigma_\nu(\beta(s)) - 1),\\ \dot{\beta}(s) &\stackrel{!}{=} -r_1 + K_1^\top \beta(s) + \frac{1}{2}\beta(s)^\top H_1 \beta(s) + C_1(\varsigma_\nu(\beta(s)) - 1),\end{aligned} \tag{2.4}$$

with *initial* values: $\alpha(0) = 0$, $\beta(0) = u.$ and $s \in [0, \tau]$. In addition, the functions with a dot above are defined as the derivative with respect to time, i.e., $\dot{\alpha}(s) = \frac{\partial \alpha(s)}{\partial s}$ and $\dot{\beta}(s) = \frac{\partial \beta(s)}{\partial s}$ across this thesis.

As a special case, the definition of the (conditional) characteristic function[3]of $X_{t+\tau}$ in probability theory is given without "discounting" by setting $r_0 = 0$ and $R_1 = 0$ in Definition A.1, as follows:

$$\psi^0(u, X_t, t, \tau; \theta^{\mathcal{X}}) := \mathbb{E}\left[e^{u \cdot X_{t+\tau}} \middle| \mathcal{F}_t\right]. \tag{2.5}$$

Furthermore, without loss of generality, the ODEs can be formulated as *initial* value problems (IVP) with $t = 0$, as researchers often refer to $t = 0$ as the current time point and τ as the time to maturity of certain financial instruments (see examples: zero-coupon bonds and European vanilla options in Section 2.1.3). Therefore, the corresponding "discounted" characteristic function is

$$\begin{aligned}\psi(u, X_0, 0, \tau; \theta^{\mathcal{X}}) &= \mathbb{E}\left[\exp\left(-\int_0^\tau r\left(X_s\right) ds\right) e^{u \cdot X_\tau} | \mathcal{F}_0\right]\\ &= e^{\alpha(\tau) + \beta(\tau) \cdot X_0}.\end{aligned}$$

Nevertheless, if one follows the original notation of Proposition 1 in Duffie et al. (2000) closely, we have the "discounted" characteristic function ψ that

$$\psi(u, X_t, t, \tau; \theta^{\mathcal{X}}) = e^{\alpha(t) + \beta(t) \cdot X_t},$$

where $\alpha : [t, t+\tau] \to \mathbb{C}$ and $\beta : [t, t+\tau] \to \mathbb{C}^n$ are the deterministic functions satisfying

[2]In this set-ups, the ODEs are interpreted as derivative of time to maturity τ instead of time t.
[3]For the general definition of characteristic function, see Definition A.2

the complex-valued ODEs :

$$\begin{aligned}\dot{\alpha}(s) &\overset{!}{=} r_0 - k_0 \cdot \beta(s) - \frac{1}{2}\beta(s)^\top h_0 \beta(s) - c_0(\varsigma_\nu(\beta(s)) - 1),\\ \dot{\beta}(s) &\overset{!}{=} r_1 - K_1^\top \beta(s) - \frac{1}{2}\beta(s)^\top H_1 \beta(s) - C_1(\varsigma_\nu(\beta(s)) - 1),\end{aligned} \tag{2.6}$$

with *terminal* values: $\alpha(t+\tau) = 0$, $\beta(t+\tau) = u$ and $s \in [t, t+\tau]$.[4] It is clear that this ODE-based approach has the advantage that, except for the non-affine case, we can use the whole toolkit of numerical methods for ordinary differential equations (e.g. Runge-Kutta methods) for computing the expectation. Further, this benefit indicates that when jumps are taken into account, the jump transform should have at least a quasi-analytical form. Otherwise, the system appears neat, but it will be computationally expensive to solve. In Duffie et al. (2000) (particularly Proposition 1, page 1351), additional information and technical details are provided in order to enable the *Itô formula* to be applied and to ensure that the *Itô integral* is a martingale with respect to $\mathcal{F}_s$.

Once again, in order to better connect transformation analysis to practice, Illustration A.1 demonstrates its use in formulating Eq. (2.4), i.e., ODEs with IVP. For more details, please refer to Appendix A.3.2.

2.1.2 Extended Transformations and Implied Cumulants

Loosely speaking, the distribution of a random variable X can be fully determined by its moments in probability theory. As alternative quantities, the cumulants of a random variable X are also used. The cumulants generating function $\mathcal{K}_X(w)$ and the moments $M_X(w)$ generating function of a random variable X[5]are defined as

$$\mathcal{K}_X(w) = \ln \mathbb{E}\left[e^{wX}\right], \quad w \in \mathbb{R}, \tag{2.7}$$

and

$$M_X(w) = \mathbb{E}\left[e^{wX}\right], \quad w \in \mathbb{R}, \tag{2.8}$$

respectively.

[4]The key difference between these two formulations lies in the derivative's direction in time, intuitively meaning that the ODEs in Eq. (2.6) characterize $\alpha(s)$ and $\beta(s)$ evolving *backward* from $t+\tau$ to t along the time axis, while in Eq. (2.4), $\alpha(s)$ and $\beta(s)$ evolve *forward* along the time axis. This leads to a sign flip on the right-hand side, if one compares Eq. (2.4) with Eq. (2.6).

[5]The definition can be extended to a multivariate version, where X is a multivariate random vector. $\mathcal{K}_X(u) = \ln \mathbb{E}\left[e^{u \cdot X}\right]$ and $M_X(u) = \mathbb{E}\left[e^{u \cdot X}\right]$, $u \in \mathbb{R}^n$.

The n-th cumulants $\mathcal{K}_{n,X}$ is obtained by differentiating the above function n times and evaluating the result at zero:

$$\mathcal{K}_{n,X} = \mathcal{K}_X^{(n)}(0) = \left.\frac{\partial_X^n \mathcal{K}}{\partial w^n}\right|_{w=0}. \tag{2.9}$$

Similar to n-th cumulants, the n-th *raw* moments $M_n = \mathbb{E}[X^n]$ are defined as follows:

$$M_{n,X} = M^{(n)}(0)_X = \left.\frac{\partial^n M}{\partial w^n}\right|_{w=0}. \tag{2.10}$$

Furthermore, the relationship between $\mathcal{K}_{n,X}$ and $M_{n,X}$ until $n = 4$ can be summarized as follows:

$$\begin{aligned}
\mathcal{K}_{1,X} &= M_{1,X}, \\
\mathcal{K}_{2,X} &= M_{2,X} - M_{1,X}^2, \\
\mathcal{K}_{3,X} &= M_{3,X} - 3M_{2,X}M_{1,X} + 2M_{1,X}^3, \\
\mathcal{K}_{4,X} &= M_{4,X} - 4M_{3,X}M_{1,X} - 3M_{2,X}^2 + 12M_{2,X}M_{1,X}^2 - 6M_{1,X}^4.
\end{aligned} \tag{2.11}$$

Technically, the cumulant $\mathcal{K}_{n,X}$ and moment $M_{n,X}$ differ from each other. Nevertheless, these quantities contain similar information, especially when they are used to tackle the inference problem. Therefore, in later chapters, both terms are treated interchangeably, especially when one refers to the information embedded in these two quantities.

As discussed in Illustration A.1, many econometric option pricing models (reduced models)[6]assume a primary process that directly models the underlying return. Additionally, the other accompanying processes capture additional empirical characteristics of the underlying return distribution, while the underlying return dynamic has no impact on the other dynamic factors. Additionally, the cumulants of the underlying return process also play a crucial role in the estimation procedure in this thesis. Therefore, the following part of this subsection focuses on the derivation of the cumulants for this special case within the AJD class.

Definition 2.1.

Let X_t^Y be a special case of AJDs. As in this case, the first entry does not enter into the dynamics of the rest of the elements in the state vector. Following Definition A.1, X_t^Y is

[6]A detailed discussion of the econometric option pricing models in reduced form is provided in Section 2.2.

defined as follows:

$$X_t^Y := (Y_t, X_t^{(2,n)}) \in D \subseteq \mathbb{R}^n, \tag{2.12}$$

where the element $Y_t \in \mathbb{R}$ locates at the first entry, and the rest elements $X_t^{(2,n)} := (X_t^{(2)}, \ldots, X_t^{(n)}) \in \mathbb{R}^{n-1}$. The $\theta^{\mathcal{X}^Y}$, which specifies the parameters of the X_t^Y's dynamic, follows Definition A.1 with additional fixed specifications: $k_1 = 0, h_1 = 0, c_1 = 0$ and $r_1 = 0$ [7].

Following Eq. (2.5), the characteristic function of $Y_{t+\tau}$ is given via

$$\begin{aligned}\mathbb{E}_t\left[e^{u_1 Y_{t+\tau}}\right] &= \psi^0((u_1, 0, \ldots, 0), X_t^Y, t, \tau; \theta^{\mathcal{X}^Y}) \\ &= e^{\alpha(\tau) + u_1 Y_t + \beta_{2,n}(\tau)\cdot X_t^{(2,n)}},\end{aligned}$$

where the function $\beta_{2,n} : [0, \tau] \to \mathbb{C}^{n-1}$ represents the scale-valued functions corresponding to the second through nth entries in β, as defined in Eq. (2.3).[8] Then, the characteristic function of $Y_{t+\tau} - Y_t$ is

$$\mathbb{E}_t\left[e^{u_1(Y_{t+\tau} - Y_t)}\right] = e^{\alpha(\tau) + \beta_{2,n}(\tau)\cdot X_t^{(2,n)}}. \tag{2.13}$$

Therefore, given the definition in Eq. (2.7), the cumulants generating function of random variable $\Delta Y_{t,\tau} := Y_{t+\tau} - Y_t$ with $\beta_1(0) = u_1$ and $\beta_{2,n}(0) = 0$ is obtained as follows:

$$\begin{aligned}\mathcal{K}_{\Delta Y_{t,\tau}}(u_1) &= \ln \mathbb{E}_t\left[e^{u_1 \Delta Y_{t,\tau}}\right] \\ &\overset{2.13}{=} \alpha(\tau) + \beta_{2,n}(\tau)\cdot X_t^{(2,n)} \\ \iff \qquad \mathcal{K}_{\Delta Y_{t,\tau}}(u_1; x_{2,n}) &= \alpha(u_1, \tau) + \beta_{2,n}(u_1, \tau)\cdot x_{2,n},\end{aligned} \tag{2.14}$$

where $u_1 \in \mathbb{R}$ and $x_{2,n} \in \mathbb{R}^{n-1}$.

Furthermore, the mth-cumulants $\mathcal{K}_{m,\psi^0}$ can be obtained by differentiating the $\psi^0(u, X^Y, t, \tau; \theta^{\mathcal{X}^Y})$ as discussed in Eq. (2.3), as follows,

$$\begin{aligned}\mathcal{K}_{m,\Delta Y_{t,\tau}}(x_{2,n}) &= \left.\frac{\partial^m \mathcal{K}_{\Delta Y_{t,\tau}}(u_1, t, \tau; x_{2,n})}{\partial u_1^m}\right|_{u_1=0} \\ &\overset{2.14}{=} \left.\frac{\partial^m \alpha(u_1, \tau)}{\partial u_1^m}\right|_{u_1=0} + \left.\frac{\partial^m \beta_{2,n}(u_1, \tau)}{\partial u_1^m}\right|_{u_1=0}\cdot x_{2,n} \\ &=: F_{0,\tau}^{(m)} + F_{1,\tau}^{(m)}\cdot x_{2,n},\end{aligned} \tag{2.15}$$

[7] This definition indicates that Y_t only impacts itself and does not have any impact on the dynamics of any other elements in the state vector.

[8] i.e., $\beta(\tau) = (\beta_1(\tau), \beta_{2,n}(\tau)^\top)^\top$, where $\beta_{2,n}(\tau) = (\beta_2(\tau), \ldots, \beta_n(\tau))^\top$.

where $\frac{\partial^m \beta_{2,n}(u_1,\tau)}{\partial u_1^m} = \left(\frac{\partial^m \beta_2(u_1,\tau))}{\partial u_1^m}, \ldots, \frac{\partial^m \beta_n(u_1,\tau))}{\partial u_1^m}\right)^\top$, $F_{0,\tau}^{(m)}$ and $F_{1,\tau}^{(m)}$ are constant coefficients, which depends only on τ and the order m implying the linear relationship between the cumulants and the states at time t.

In Feunou and Okou (2018), the constant coefficients with respect to the model specification from Andersen et al. (2015b) are calculated analytically up to the fourth order. As described in Feunou and Okou (2018), the cumulants can be obtained by solving several new ODEs for different orders *analytically* (see Appendix of Feunou and Okou (2018)).

As an alternative approach, it is possible to obtain the derivatives of the cumulant generating function by solving additional sets of ODEs in a more general way. The next part of this subsection demonstrates how the derivatives can be computed via the approach of solving ODEs. The first-order derivative can be obtained from the extended version of the characteristic function proposed by Duffie et al. (2000).

The definition of the "extended" transformation function $\tilde{\psi} : \mathbb{R}^n \times \mathbb{C}^n \times D \times [0, +\infty) \times [0, +\infty) \to \mathbb{C}$ is

$$\tilde{\psi}(v, u, X_t, t, \tau; \theta^{\mathcal{X}}) := \mathbb{E}\left[\exp\left(-\int_t^{t+\tau} r\left(X_s\right) ds\right)(v \cdot X_{t+\tau}) e^{u \cdot X_{t+\tau}} | \mathcal{F}_t\right]. \tag{2.16}$$

The $\tilde{\psi}$ can be obtained via the extension of discounted characteristic function ψ in Eq. (2.3), as follows,

$$\begin{aligned} \tilde{\psi}(v, u, X_t, t, \tau; \theta^{\mathcal{X}}) &= \psi(u, X_t, t, \tau)\left(\partial_1\alpha(\tau) + \partial_1\beta(\tau) \cdot X_t\right) \\ &= e^{\alpha(\tau)+\beta(\tau)\cdot X_t}\left(\partial_1\alpha(\tau) + \partial_1\beta(\tau) \cdot X_t\right), \end{aligned} \tag{2.17}$$

where $\partial_1\alpha : [0, \tau] \to \mathbb{R}$ and $\partial_1\beta : [0, \tau] \to \mathbb{R}^n$ are the deterministic functions satisfying the ODEs [9]:

$$\begin{aligned} \dot{\partial_1\alpha}(s) &\stackrel{!}{=} k_0 \cdot \partial_1\beta(s) + \partial_1\beta(s)^\top h_0 \beta(s) + c_0 \nabla_{\varsigma_\nu}(\beta(s)) \cdot \partial_1\beta(s), \\ \dot{\partial_1\beta}(s) &\stackrel{!}{=} K_1^\top \partial_1\beta(s) + \partial_1\beta(s)^\top H_1 \beta(s) + C_1 \nabla_{\varsigma_\nu}(\beta(s)) \cdot \partial_1\beta(s), \end{aligned} \tag{2.18}$$

with initial values $\partial_1\alpha(0) = 0$, $\partial_1\beta(0) = v$ and $s \in [0, \tau]$. In addition, $\nabla_{\varsigma_\nu}(\beta(s))$ is the gradient with respect to $\beta(s)$.

Similar to Eq. (2.5), the corresponding extended version of characteristic function without

discount is defined, as follows,

$$\begin{aligned}\tilde{\psi}^0(v,u,X_t,t,\tau;\theta^{\mathcal{X}}) &= \mathbb{E}\left[(v\cdot X_{t+\tau})e^{u\cdot X_{t+\tau}}|\mathcal{F}_t\right], \\ &= \psi^0(u,X_t,t,\tau)\left(\partial_1\alpha(t)+\partial_1\beta(t)\cdot X_t\right).\end{aligned} \tag{2.19}$$

Following the definition in Proposition A.1 and Eq. (2.19) with $v = d_1 := (1, 0 \dots 0)^\top$ and $u = 0$,

$$\begin{aligned}\mathbb{E}_t\left[Y_{t+\tau}\right] &= \mathbb{E}\left[(d_1\cdot X_{t+\tau})e^{0\cdot X_{t+\tau}}\middle|\,\mathcal{F}_t\right] \\ &= \tilde{\psi}^0(d_1,0,X_t^Y,t,\tau;\theta^{\mathcal{X}}) \\ &= e^{\alpha(\tau)+\beta(\tau)\cdot X_t}\left(\partial_1\alpha(\tau)+\partial_1\beta(\tau)\cdot X_t\right).\end{aligned} \tag{2.20}$$

Since $\beta(0) = u = 0$, we have $\alpha(\tau) = 0$, $\beta(\tau) = 0$. In addition, since $\partial_1\beta_1(0) = 1$, we have $\partial_1\beta_1(\tau) = 1$. Therefore, we have

$$\begin{aligned}&\mathbb{E}_t\left[Y_{t+\tau}\right] = \partial_1\alpha(\tau) + Y_t + \partial_1\beta_{2,n}(\tau)\cdot X_t^{(2,n)} \\ \iff\ &\mathbb{E}_t\left[\Delta Y_{t,\tau}\right] = \partial_1\alpha(\tau) + \partial_1\beta_{2,n}(\tau)\cdot X_t^{2,n},\end{aligned}$$

where $\partial_1\alpha(\tau)$ and $\partial_1\beta_{2,n}(\tau)$ satisfy Eq. (2.18) with initial values: $\partial_1\alpha(0) = 0$ and $\partial_1\beta_{2,n}(0) = 0$. Therefore, we have $F_{0,\tau}^{(1)} = \partial_1\alpha(\tau), F_{1,\tau}^{(1)} = \partial_1\beta_{2,n}(\tau)$ following the definition in Eq. (2.15) and first-order of cumulant with the notation in cumulant generating function as follows:

$$\mathcal{K}_{1,\Delta Y_{t,\tau}}(x_{2,n}) = \partial_1\alpha(\tau) + \partial_1\beta_{2,n}(\tau)\cdot x_{2,n} \tag{2.21}$$

Moreover, for further the m-th higher-order derivatives $m \geq 2$ can be computed recursively. Following Theorem 1 in Chen and Joslin (2012), we have ODEs for the m-th higher order deviates as follows:

$$\begin{aligned}\dot{\partial_m\alpha}(s) &\stackrel{!}{=} k_0\cdot\partial_m\beta(s) + \frac{1}{2}\sum_{i=0}^{m}\binom{m}{i}\partial_i\beta(s)^\top h_0\partial_{m-i}\beta(s) \\ &\quad + \partial_{m-1}\left(c_0\nabla\varsigma_\nu(\beta(s))\cdot\partial_1\beta(s)\right), \\ \dot{\partial_m\beta}(s) &\stackrel{!}{=} K_1^\top\partial_m\beta(s) + \frac{1}{2}\sum_{i=0}^{m}\binom{m}{i}\partial_i\beta(s)^\top H_1\partial_{m-i}\beta(s) \\ &\quad + \partial_{m-1}\left(C_1\nabla\varsigma_\nu(\beta(s))\cdot\partial_1\beta(s)\right),\end{aligned} \tag{2.22}$$

with initial values: $\partial_m\alpha(0) = 0$, $\partial_m\beta(0) = 0$[10]and $s \in [0,\tau]$. In addition,

[9]See the heuristic derivation of "extended" characteristic function for the special case of X_t^Y in Appendix A.2

$\partial_{m-1}\left(C_1\nabla\vartheta(\beta(s))\cdot\partial_1\beta(s)\right)$ is the $(m-1)$-th derivative of $C_1\nabla\vartheta(\beta(s))\cdot\partial_1\beta(s)$ with respect to $\beta(s)$. Then, following the same idea of first-order cumulant in Appendix A.2, we have

$$\mathcal{K}_{m,\Delta Y_{t,\tau}}(x_{2,n}) = \partial_m\alpha(\tau) + \partial_m\beta_{2,n}(\tau)\cdot x_{2,n}. \tag{2.23}$$

For illustrative purpose, the example with dynamics of Bates model can be found in Appendix A.3.3.

2.1.3 Risk-Neutral Pricing

According to the "fundamental theorem of asset pricing" proposed by Harrison and Kreps (1979), who establish the link between the existence of an equivalent martingale measure and *the absence of arbitrage*, the fair market value of a "regular" contingent claim - which is a random variable in the form of function $H(X_{t+\tau})$, at time t is determined by its discounted expectation under the equivalent martingale measure under technical conditions.[11] Let $\mathbb{Q}$ be the equivalent martingale associated with the risk-free rate $r(X_t) = r_0 + R_1\cdot X_t$ from Definition A.1 The fair market value of payoff $H(X_{t+\tau})$ at time t with maturity τ is defined, as follows:

$$\mathcal{H}(t,\tau) = \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_t^{t+\tau} r(X_s)ds\right)H(X_{t+\tau})\,|\mathcal{F}_t\right]. \tag{2.24}$$

In the case that zero-coupon bonds pay one at the maturity $t+\tau$, using risk-neutral pricing, the fair market value of the zero-coupon bonds at time t is

$$\begin{aligned}\mathcal{B}(t,\tau) &= \mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(-\int_t^{t+\tau} r(X_s)ds\right)\right]\\ &= \mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(-\int_t^{t+\tau} r(X_s)ds\right)e^{0\cdot X_{t+\tau}}\right]\\ &= e^{\alpha(\tau)+\beta(\tau)\cdot X_t},\end{aligned} \tag{2.25}$$

where $\alpha(\tau)$ and $\beta(\tau)$ are the solution of ODEs in Eq. (2.6) with initial values $\alpha(0)=0$ and $\beta(0)=0$.

[10] For $\beta_1(u_1,0)=u_1$ in the case of X_t^Y, we have the higher order derivatives $\frac{\partial^m\beta_1(u_1,0)}{\partial u_1^m}=0$ with $u_1\in\mathbb{R}$ and $m\geq 2$.

[11] For more illustrative discussions in details about the relationship between the absence of arbitrage and equivalent martingale, see Duffie (2010)

Furthermore, the yield-to-maturity is

$$\begin{aligned} yld(t,\tau) &= -\frac{\ln \mathcal{B}(t,\tau)}{\tau} \\ &= -\frac{\alpha(\tau)+\beta(\tau)\cdot X_t}{\tau} \end{aligned} \tag{2.26}$$

Due to the focus of the empirical application being on the option market, the payoffs of European call: $\mathcal{C}(t,\tau,K)$ and put: $\mathcal{P}(t,\tau,K)$ options written on stocks with moneyness m at maturity $T=t+\tau$ are defined in terms of strike K, as follows:

$$\mathcal{C}(T,0,K) := (S_T-K)^+ = \max(S_T-K,0) \tag{2.27}$$

and

$$\mathcal{P}(T,0,K) := (K-S_T)^+ = \max(K-S_T,0), \tag{2.28}$$

respectively. Following Definition A.1, let the log underlying price be the first element of the process, i.e., $X_t = (\ln S_t, \cdots)^\top$. Therefore, the market fair value of a call option contract is, by definition,

$$\begin{aligned} \mathcal{C}(t,\tau,K) &= \mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(-\int_t^{t+\tau} r(X_s)ds\right)(S_{t+\tau}-K)^+\right] \\ &= \mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(-\int_t^{t+\tau} r(X_s)ds\right)(e^{d_1\cdot X_{t+\tau}}-K)^+\right], \end{aligned} \tag{2.29}$$

where $d_1 := (1,0,\cdots,0)^\top \in \mathbb{R}^n$. In contrast to bond pricing, it is not possible to obtain the fair market value of an option contract directly by solving ODEs. The next subsection explains the essential technicalities of pricing European options.

2.1.4 Inverse Fourier Pricing Techniques

2.1.4.1 Inverse Fourier Method

Chen and Joslin (2012) propose a generalized version of the transform for AJD processes, which is defined as follows:

$$\begin{aligned} G(f,a,b,t,\tau;\theta^{\mathcal{X}}) &= \mathbb{E}\left[\exp\left(-\int_t^{t+\tau} r(X_s)ds\right)e^{a\cdot X_{t+\tau}}f(b\cdot X_{t+\tau})\middle|\mathcal{F}_t\right] \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{f}(\nu)\psi(a+i\nu b,X_t,t,\tau;\theta^{\mathcal{X}})d\nu, \end{aligned} \tag{2.30}$$

where $\psi(a + i\nu b, X_t, t, \tau; \theta^{\mathcal{X}})$ is defined in Eq. (2.3), and $\hat{f}$ is the Fourier transform of f (see Definition A.3).

The general transformation can be further derived as follows:

$$\begin{aligned} G(f, a, b, t, \tau; \theta^{\mathcal{X}}) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\nu) \psi(a + i\nu b, X_t, t, \tau; \theta^{\mathcal{X}}) d\nu \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\nu) e^{\alpha(\tau) + \beta(\tau) \cdot X_t} d\nu, \end{aligned} \tag{2.31}$$

where $\alpha(\tau)$ and $\beta(\tau)$ are the solution from Eq. (2.4) with initial values $\alpha(0) = 0$ and $\beta(0) = a + i\nu b$.

By using the risk-neutral pricing framework and Eq. (2.29), one can calculate the fair value of a European call option via Eq. (2.30), as follows:

$$\begin{aligned} \mathcal{C}(t, \tau, K) &= \mathbb{E}_t^{\mathbb{Q}} \left[\exp\left(- \int_t^{t+\tau} r(X_s) ds \right) (S_{t+\tau} - K)^+ \right] \\ &= \mathbb{E}_t^{\mathbb{Q}} \left[\exp\left(- \int_t^{t+\tau} r(X_s) ds \right) \left(e^{d_1 \cdot X_{t+\tau}} - K \right)^+ \right], \\ &= G(f(b \cdot X_{t+\tau}) = \left(e^{\ln S_T} - K \right)^+, a = 0, b = d_1, 0, t + \tau, \theta^{\mathcal{X}}), \end{aligned} \tag{2.32}$$

where $X_t = (\ln S_t, \ldots)^\top$ and $d_1 = (1, 0, \ldots, 0)^\top$.

It is possible to adjust f to $\tilde{f}$ by selecting a and b differently, which is useful when considering option payouts. The $\tilde{f}$ is still applicable to Eq. (2.30). The adjustment is as follows:

$$e^{a \cdot X_{t+\tau}} f(b \cdot X_{t+\tau}) = e^{(a - cb) \cdot X_{t+\tau}} \tilde{f}(b \cdot X_{t+\tau}), \ c \in \mathbb{R}, \tag{2.33}$$

where $\tilde{f}(b \cdot X_{t+\tau}) := e^{c(b \cdot X_{t+\tau})} f(b \cdot X_{t+\tau})$.

In the case of pricing a European call option, assuming $a = 0, c = -w$ and $b = d_1$, it implies:

$$\begin{aligned} (S_{t+\tau} - K)^+ &= e^{a \cdot X_{t+\tau}} f(b \cdot X_{t+\tau}) \\ &\overset{12}{=} e^{(a - cb) \cdot X_{t+\tau}} \tilde{f}(b \cdot X_{t+\tau}), \end{aligned} \tag{2.34}$$

where $f = e^{wd_1 \cdot X_{t+\tau}} \tilde{f}$ with $\tilde{f} = e^{-wd_1 \cdot X_{t+\tau}} \left(e^{d_1 \cdot X_{t+\tau}} - K\right)^+$.

Therefore, we have

$$\begin{aligned}\mathcal{C}(t,\tau,K) &= \mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(-\int_t^{t+\tau} r(X_s)ds\right) e^{wd_1 \cdot X_{t+\tau}} e^{-wd_1 \cdot X_{t+\tau}} \left(e^{d_1 \cdot X_{t+\tau}} - K\right)^+\right], \\ &= G(\tilde{f}(b \cdot X_{t+\tau}) = e^{-wd_1 \cdot X_{t+\tau}} \left(e^{d_1 \cdot X_{t+\tau}} - K\right)^+, a = wd_1, b = d_1, t, \tau, \theta^{\mathcal{X}}).\end{aligned}$$

Prior to applying Chen and Joslin (2012)'s generalized transform to the price of a European call option, a result from Filipovic (2009) should be taken into account.

Let

$$\tilde{f}(y) := e^{-wy}(e^y - K)^+, \ w > 1, \tag{2.35}$$

where w is a pre-chosen small constant. Filipovic (2009) shows that the Fourier transformation of $\tilde{f}(y)$ is given as follows:

$$\begin{aligned}\hat{\tilde{f}}(\nu) &= \int_{\mathbb{R}} e^{-(w+i\nu)y}(e^y - K)^+ dy, \\ &= \frac{K^{-(w-1+i\nu)}}{(w+i\nu)(w-1+i\nu)}, \ w > 1.\end{aligned} \tag{2.36}$$

Moreover, by setting $y = \ln S_{t+\tau} = d_1 \cdot X_{t+\tau}$ the Fourier transformation of $\tilde{f} = e^{-wd_1 \cdot X_{t+\tau}} \left(e^{d_1 \cdot X_{tG+\tau}} - K\right)^+$ is given in Eq. (2.36). Following Chen and Joslin (2012)'s general transformation (see Eq. (2.30)), we have

$$\begin{aligned}\mathcal{C}(t,\tau,K) &= G(\tilde{f}(b \cdot X_{t+\tau}) = e^{-wd_1 \cdot X_{t+\tau}} \left(e^{d_1 \cdot X_{t+\tau}} - K\right)^+, a = wd_1, b = d_1, t, \tau, \theta^{\mathcal{X}}) \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{\tilde{f}}(\nu)\psi((a+i\nu b), X_t, t, \tau; \theta^{\mathcal{X}})d\nu \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{\tilde{f}}(\nu)\psi((w+i\nu)d_1, X_t, t, \tau; \theta^{\mathcal{X}})d\nu \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{K^{-(w-1+i\nu)}}{(w+i\nu)(w-1+i\nu)}\psi((w+i\nu)d_1, X_t, t, \tau; \theta^{\mathcal{X}})d\nu\end{aligned}$$

.

12

$$\begin{aligned}&e^{(a-cb)\cdot X_{t+\tau}} \tilde{f}(b \cdot X_{t+\tau}) \\ =&e^{(0+wd_1)\cdot X_{t+\tau}} \tilde{f}(d_1 \cdot X_{t+\tau}) \\ =&e^{(0+wd_1)\cdot X_{t+\tau}} e^{(-wd_1)\cdot X_{t+\tau}} (e^{d_1 \cdot X_{t+\tau}} - K)^+ \\ =&e^{w \ln S_{t+\tau}} e^{-w \ln S_{t+\tau}} (S_{t+\tau} - K)^+\end{aligned}$$

Finally, the European call price coincides with

$$\mathcal{C}(t,\tau,K) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{K^{-(w-1+i\nu)}}{(w+i\nu)(w-1+i\nu)} e^{\alpha(\tau;\nu)+\beta(\tau;\nu)\cdot X_t} d\nu \tag{2.37}$$

where $\alpha(\tau;\nu)$ and $\beta(\tau;\nu)$ are the solutions of ODEs in Eq. (2.4) with initial values $\alpha(0) = 0$ and $\beta(0) = (w+i\nu)d_1 = (w+i\nu, 0, \ldots, 0)^\top$. By using standard numerical integration methods, one can evaluate the integration in Eq. (2.37).

2.1.4.2 Cosine Method[13]

The numerical integration in Eq. (2.37) is computationally expensive, especially when the latter estimation approach directly incorporates the observations of options from the market. This requires evaluation of a large amount of option prices at the same time. Therefore, a faster numerical approach is desired. Along with the literature (see e.g., Andersen et al. (2015b) and Bardgett et al. (2019), among others), the cosine method proposed by Fang and Oosterlee (2008) delivers outstanding performance in both speed and accuracy. This part discusses the main idea of the cosine method and how it can be applied to AJD models. Following the general idea of risk-neutral pricing in Section 2.1.3, we have the fair value of contingent claim $v(t+\tau, 0; Y_{t+\tau})$ at t, as follow:

$$\begin{aligned} v(t,\tau;x) &= e^{-r\tau}\mathbb{E}^{\mathbb{Q}}\left[v(Y_{t+\tau}, t+\tau) \middle| Y_t = x\right] \\ &= e^{-r\tau}\int_{\mathbb{R}} v(y, t+\tau) f(y;x)\, dy, \end{aligned} \tag{2.38}$$

where

- Y_t takes the "initial" value x with the information at t;[14]
- y is the realization of $Y_{t+\tau}$ with the information at $t+\tau$;
- $f(y;x)$ is the (conditional) density function under $\mathbb{Q}$;
- for simplicity, the interest rate is assumed constant, i.e., $r_t \equiv r$.

A key aspect of preventing the calculation of the expectation analytically is the unknown (conditional) density function of the underlying asset price at maturity $t+\tau$. Once again, the solution lies in the Fourier duality between the density function f and its

[13] This part is mainly based on the results of research projects with C-RAM students. I would like to thank Maximilian Kübler and Julian Meyer for their assistance.

[14] For simplicity, we assume Y_t is a scale value.

Fourier transformation $\hat{f}$, as follow:

$$\hat{f}(t) = \int_{\mathbb{R}} f(y;x) \cdot e^{-ity} dy = \varphi(-t), \ t \in \mathbb{R}.$$

Before discussing the main idea of the cosine method, the following proposition of Fourier cosine expansion is presented:

Furthermore, we will explain several intermediate steps in order to motivate the final result of the cosine method. A key intermediate step is to truncate the domain of the integral in equation Eq. (2.38) at some point. The existence of the integral ensures that this is feasible. According to Proposition Proposition A.2, the density is essentially expanded into an infinite sum of cosines whose coefficients are known up to an integral. While the original support of the function being expanded is on $[0, \pi]$, the result can be applied to any function f whose support is on a finite domain $[a, b]$ given $g : [0, \pi] \to \mathbb{R}, \theta \mapsto f\left(\frac{b-a}{\pi}\theta + a\right)$.

By using Proposition A.2 we have

$$f(y;x) = g\left(\frac{y-a}{b-a}\pi\right) = \sum\nolimits_{k=0}^{\infty}{}' A_k \cos\left(k\frac{y-a}{b-a}\pi\right), \tag{2.39}$$

where

$$A_k = \frac{2}{\pi}\int_0^{\pi} g(\theta;x)\cos(k\theta)\, d\theta \overset{\theta=\frac{y-a}{b-a}\pi}{=} \frac{2}{b-a}\int_a^b f(y;x)\cos\left(k\frac{y-a}{b-a}\pi\right) dy.$$

Since $\cos\left(k\frac{y-a}{b-a}\pi\right) = \mathrm{Re}\left(e^{ik\frac{y-a}{b-a}\pi}\right)$, the coefficient A_k can be further transformed into the following form:

$$\begin{aligned} A_k = \frac{2}{b-a}\int_a^b f(y;x)\mathrm{Re}\left(e^{ik\frac{y-a}{b-a}\pi}\right) dy &= \frac{2}{b-a}\mathrm{Re}\left[e^{-i\frac{ka\pi}{b-a}} \cdot \int_a^b f(y;x)e^{i\frac{k\pi}{b-a}y} dy\right] \\ &\approx \frac{2}{b-a}\mathrm{Re}\left[e^{-i\frac{ka\pi}{b-a}} \cdot \int_{\mathbb{R}} f(y;x)e^{i\frac{k\pi}{b-a}y} dy\right] \\ &= \frac{2}{b-a}\mathrm{Re}\left[e^{-i\frac{ka\pi}{b-a}} \cdot \varphi\left(\frac{k\pi}{b-a}\right)\right] =: F_k, \end{aligned} \tag{2.40}$$

where φ is a function related to the (conditional) characteristic function of $Y_{t+\tau}$.

By its definition the infinite series $\sum_{k=0}^{\infty}{}' A_k \cos\left(k\frac{x-a}{b-a}\pi\right) (= f(\cdot;x))$ is convergent, i.e., the expansion into an orthonormal basis in $L^2(\mathbb{R})$ gives a converging sum. So is the

approximating series

$$\sum\nolimits_{k=0}^{\infty}{}' F_k \cos\left(k\pi\frac{x-a}{b-a}\right),$$

because the approximation error is primarily due to the extension of $[a,b]$ to $\mathbb{R}$. In view of the fact that the decay rate of the density is dominated by the (superlinear) decay rate of the density, we are on the safe side by selecting appropriate a and b.[15] Accordingly, the convergence of the sum suggests that only the first N coefficients need to be computed. As far as we are aware, roughly 10^2 coefficients are sufficient to provide a sufficiently accurate result, depending on maturity (influencing the length of $[a,b]$) and moneyness. Replacing $f(\cdot;x)$ in Eq. (2.38) and limiting the sum to the first $N+1$ coefficients, we have

$$\begin{aligned}
v(x,t,\tau) &= e^{-r\tau}\int_{-\infty}^{\infty} v(y,t+\tau)\sum\nolimits_{k=0}^{\infty}{}' A_k \cos\left(k\frac{x-a}{b-a}\pi\right) dy \\
&\approx e^{-r\tau}\int_{a}^{b} v(y,t+\tau)\sum\nolimits_{k=0}^{\infty}{}' A_k \cos\left(k\frac{x-a}{b-a}\pi\right) dy \\
&\approx e^{-r\tau}\int_{a}^{b} v(y,t+\tau)\sum\nolimits_{k=0}^{N}{}' A_k \cos\left(k\frac{x-a}{b-a}\pi\right) dy \\
&\approx e^{-r\tau}\sum\nolimits_{k=0}^{N}{}' F_k(x)V_k =: \tilde{v}(x,t,\tau),
\end{aligned}$$

where $F_k(x)$ stems from the expansion of $f(\cdot;x)$ into a cosine series (dependent of x) and

$$V_k := \int_a^b v(y,t+\tau)\cos\left(k\pi\frac{y-a}{b-a}\right) dy$$

are the payoff coefficients (independent of x).

The final result of $\tilde{v}(x,t,\tau)$ contains three approximation steps:

1. The final result of expectation under risk-neutral measure is approximated by a truncated integral of $[a,b]$ instead of the integral over $\mathbb{R}$.

2. The infinite series is no longer, not even by computing the infinite sum, the true option's value, but an approximation until N.

3. The approximation of the coefficients A_k in Eq. (2.40) is done by approximating the integral by the full, characteristic function (which initially was true, but no longer is

[15] In fact, the truncation relies on the computation of the first, second and fourth cumulants (see Eq. (2.9)) of the risk neutral distribution at time T implied by $f(\cdot|x)$ by which we then choose

$$[a,b] := \left[\mathcal{K}_{1,Y_{t+\tau}} - L\sqrt{\mathcal{K}_{2,Y_{t+\tau}} + \sqrt{\mathcal{K}_{4,Y_{t+\tau}}}}, \mathcal{K}_{1,Y_{t+\tau}} + L\sqrt{\mathcal{K}_{2,Y_{t+\tau}} + \sqrt{\mathcal{K}_{4,Y_{t+\tau}}}}\right]. \tag{2.41}$$

for the truncated density function $f|_{[a,b]}$).

On the other hand, the advantage of the cosine basis is that analytical results can be derived for several common payoff functions (either European or Digital) without introducing extra noise for their approximation as follows:

For European options defined in the log space of ratio between underlying and strike $Y_{t+\tau} = \ln\left(\frac{S_{t+\tau}}{K}\right)$, the payoff function is given in following form

$$v_{call}(y, t+\tau) = K\left(e^y - 1\right),\ y > 0$$

and

$$v_{put}(y, t+\tau) = K\left(1 - e^y\right),\ y < 0.$$

Thus, we obtain the following payoff coefficients for European call and put

$$V_k^{call} = K\left(\chi_k\left(\bar{a}, b\right) - \Psi_k\left(\bar{a}, b\right)\right)$$

and

$$V_k^{put} = K\left(\Psi_k\left(a, \bar{b}\right) - \chi_k\left(a, \bar{b}\right)\right),$$

for $\bar{a} := \max(a, 0), \bar{b} := \min(b, 0)$ and only if $b > \bar{a}, a < \bar{b}$ respectively. Otherwise, the integration would take place over an area where the call (put) is never in-the-money and the cosine method approximation for $\tilde{v}\left(x, t\right)$ is simply 0.

For the last point in this part, we would like to present another "trick" for the cosine method. It is especially useful for performing multiple option valuations simultaneously. By its definition, the coefficients V_k have to be computed from scratch each time when the strike K changes. Thanks to the transformation of the payoff into the log-space $\ln\left(\frac{S_{t+\tau}}{K}\right)$, the nonlinear relation between the payoff coefficients and the strike is now linear:

$$V_k^{call} = K\underbrace{\left(\chi_k\left(\bar{a}, b\right) - \Psi_k\left(\bar{a}, b\right)\right)}_{=:U_k^{call}} \text{ and } V_k^{put} = K\underbrace{\left(\Psi_k\left(a, \bar{b}\right) - \chi_k\left(a, \bar{b}\right)\right)}_{=:U_k^{put}}.$$

Hence, plugging in the form of F_k from (2.40) we get for European options

$$v\left(x, t\right) \approx \tilde{v}\left(x, t\right) = \frac{2K}{b-a} e^{-r\tau} {\sum_{k=0}^{N}}' \mathrm{Re}\left[e^{-i\frac{ka\pi}{b-a}} \varphi\left(\frac{k\pi}{b-a}; x\right)\right] U_k.$$

Therefore, for different strikes with identical underlying values: S_t, the payoff coefficients

only have to be calculated once. Due to the K dependent definition in $\varphi(\cdot)$, changing K also requires the evaluation of $N+1$ ODEs. Nevertheless, for the AJDs models (especially for the dynamics of the underlying asset according to Proposition A.1), we can show that the strike dependency of the characteristic function can be separated. Consider X_t^Y in Definition 2.1 with $Y_t = \ln\left(\frac{S_t}{K}\right) = x \in \mathbb{R}$, then following Proposition A.1, we have

$$\begin{aligned}
\varphi(\nu_1; Y_t = x) &= \mathbb{E}_t\left[e^{i\nu_1 Y_{t+\tau}}\right] \\
&= \mathbb{E}_t\left[e^{i\nu_1 d_1 \cdot X_{t+\tau}^Y}\right] \\
&= \psi^0(i\nu_1 d_1, X_t^Y, t, \tau; \theta^{\mathcal{X}^Y}) \\
&= e^{\alpha(\tau)+i\nu_1 x+\beta_{2,n}(\tau)\cdot X_t^{(2,n)}} \\
&= e^{i\nu_1 x}\psi^0(0, X_t^Y, t, \tau; \theta^{\mathcal{X}^Y}),
\end{aligned} \tag{2.42}$$

where $\nu_1 \in \mathbb{R}$ and $d_1 := (1, 0, \cdots, 0)^\top \in \mathbb{R}^n$. Finally, the fair price via cosine method is given as follows:

$$v(x,t) \approx \frac{2K}{b-a} e^{-r\tau} \acute{\sum_{k=0}^{N}} \mathrm{Re}\left[e^{ik\pi\frac{(x-a)}{b-a}}\psi^0(0, X_t^Y, t, \tau; \theta^{\mathcal{X}^Y})\right] U_k,\ Y_t = \ln\left(\frac{S_t}{K}\right). \tag{2.43}$$

On the other hand, the dynamics of later concrete models are typically assumed to be $X^{\tilde{Y}} = (\ln S_t, \dots)^\top$ instead of $X_t^Y = X^{\tilde{Y}} - d_1 \ln K = (\ln S_t - \ln K, \dots)^\top$. Their transformations of characteristic functions[16]are

$$\psi^0(\nu_1, X_t^{\tilde{Y}}, t, \tau; \theta^{\mathcal{X}^{\tilde{Y}}}) = e^{\tilde{\alpha}(\tau)+\tilde{\beta}(\tau)\cdot X_t^{\tilde{Y}}},$$

and

$$\psi^0(\nu_1, X_t^Y, t, \tau; \theta^{\mathcal{X}^Y}) = e^{\alpha(\tau)+\beta(\tau)\cdot X_t^Y}$$

respectively. By using *Itô lemma* we immediately obtain

$$dX_t^{\tilde{Y}} = dX_t^Y.$$

As a result, both processes generate the same set of ODEs to solve, and since the solution is unique, we do, in fact, have $\alpha(s) = \tilde{\alpha}(s), \beta(s) = \tilde{\beta}(s),\ \forall s \in [0, \tau]$.

[16]Here we only need the general transformation of AJD.

2.2 Reduced-Form Models

This section reviews several widely used concrete examples within the general class of AJD models (see Definition A.1) for the task of pricing equity derivatives[17]in the empirical asset pricing literature. For options and other derivatives, these models are designed to accommodate various statistical features of asset return dynamics under the risk-neutral measure. In contrast to equilibrium models, investors' preferences are absorbed in risk premiums, and it is difficult to establish a direct link between their preferences and risks. Thus, researchers usually leave the risk premium as free parameters in the models. Following Garcia et al. (2010), this section provides an overview of both the empirical and technical aspects of option pricing models based on AJDs. Furthermore, this section also reviews a number of state-of-the-art models that appear in recent literature.

In general, those reduced-form AJD models consist of one factor capturing the fluctuation of the underlying asset price and some additional aggregate risk factors. These factors include heteroscedastic volatility, stochastic jump intensity, and so on. Additionally, the underlying asset return factor is assumed to be directly observable, while the other factors are assumed to be latent.

The discussion in this section is organised according to building blocks of statistical features for the return dynamic. For reduced-form models, researchers often assume that the model structure remains the same under the physical measure $\mathbb{P}$ and risk-neutral measure $\mathbb{Q}$. As a result, it appears intuitively that risk premiums originate from the "difference" between $\mathbb{P}$ and $\mathbb{Q}$. If model structures are assumed to be the same under $\mathbb{P}$ and $\mathbb{Q}$, the parameterization under $\mathbb{P}$ and parameterization under $\mathbb{Q}$ and their "difference" are determined by the two of them. When either $\mathbb{P}$ or $\mathbb{Q}$ is to be considered, only the corresponding parameterization needs to be specified. The last example in this section illustrates the case of specifying only model dynamics under $\mathbb{Q}$ when focusing on option pricing.

This section is organized as follows: Section 2.2.1 introduces the building block for stochastic volatility. The illustrative example is introduced by Heston (1993), who develops a closed-form for option pricing. Section 2.2.2 reviews the different jump specifications in return. The jump sizes are either normally distributed or exponentially distributed.

[17]Loosely speaking, the methodology of risk-neutral pricing is relatively general, which means the validity of the model is not dependent upon asset class, so long as the AJD model can provide a reasonable fit to the corresponding empirical risk-neutral return density.

Section 2.2.3 discusses models with jumps in volatility and stochastic jump intensity.

2.2.1 Stochastic Volatility

Cox et al. (1985) propose a stochastic process, which contains mean-reversion and self-exciting volatility. This process is often used as the dynamic of interest rates in term-structure models. The definition is given as follows:

Definition 2.2. Cox–Ingersoll–Ross (CIR) model

$$d\alpha_t = \kappa\left(\theta - \alpha_t\right)\, dt + \sigma\sqrt{\alpha_t} dW_t, \tag{2.44}$$

where

κ and θ denote the mean-reversion speed and long-run mean, respectively;

σ corresponds to volatility;

W_t is the standard Wiener process.

As a natural extension of the standard BS model, Heston (1993) proposes a stochastic volatility model, in which a second separate process governs the dynamic of volatility by using the CIR process in Definition 2.2. Thereafter, the CIR process becomes a popular choice for modeling stochastic volatility. The Heston model is defined as follows:

Illustration 2.1. Heston model

To account for volatility clustering and the leverage effect on the return dynamic (see page 3142 Christoffersen et al. (2010)). The price process S_t and volatility process V_t under physical and risk-neutral measures are defined, respectively, as follows:

under $\mathbb{P}$:

$$\begin{aligned} \frac{dS_t}{S_t} &= \mu_t dt + \sqrt{V_t} dW_t \\ dV_t &= \kappa\left(\theta - V_t\right) dt + \sigma\sqrt{V_t} dB_t; \end{aligned} \tag{2.45}$$

under $\mathbb{Q}$:

$$\begin{aligned} \frac{dS_t}{S_t} &= (r_t - q_t)dt + \sqrt{V_t} dW_t^{\mathbb{Q}} \\ dV_t &= \kappa^{\mathbb{Q}}\left(\theta^{\mathbb{Q}} - V_t\right) dt + \sigma\sqrt{V_t} dB_t^{\mathbb{Q}}, \end{aligned} \tag{2.46}$$

where

μ_t denotes the drift in return under physical measure;

r_t and q_t denotes the risk-free rate and dividend rate, respectively;

κ, θ and σ are given as in Definition 2.2 under $\mathbb{P}$;

W_t and B_t are standard Wiener processes, which are correlated with coefficient $\rho \in [-1, 1]$. One can rewrite the Wiener process W_t with another Wiener process $\tilde{W}_t$, which is independent of B_t, as follows:

$$dW_t = \left(\rho d\tilde{W}_t + \sqrt{1-\rho^2} dB_t\right); \tag{2.47}$$

$\left(\kappa^{\mathbb{Q}}, \theta^{\mathbb{Q}}\right)$ are the corresponding values for the process under risk-neutral measure. The σ is the same under both measures.
Furthermore, the variance risk premia is defined as

$$\eta_v = \kappa - \kappa^{\mathbb{Q}}; \tag{2.48}$$

the $\theta^{\mathbb{Q}}$ is restricted as follows:

$$\theta^{\mathbb{Q}} = \frac{\kappa\theta}{\kappa^{\mathbb{Q}}}, \tag{2.49}$$

This restriction is imposed due to $\theta^{\mathbb{Q}}\kappa^{\mathbb{Q}} = \kappa\theta$;
Finally, μ_t can be rewritten as

$$\mu_t = r_t - q_t + \gamma_t, \tag{2.50}$$

where $\gamma_t = \eta_s V_t$.

2.2.2 Jumps in Return

The Black-Scholes model can be further extended by incorporating a jump process, which captures large, infrequent movements caused by rare extreme events. The jump-diffusion model proposed by Merton (1976) includes jumps using the compound Poisson process. For the illustrative purpose, only the definition under physical measure is given, as follows:

Definition 2.3. Merton's jump-diffusion model

$$\frac{dS_t}{S_t} = \mu_t dt + \sigma dW_t + \left(e^{Z_i^s} - 1\right) dN_t - c(t)\bar{\mu}_s dt, \tag{2.51}$$

where

μ_t denotes the drift in return under physical measure;

σ denotes the constant volatility;

W_t is standard Wiener processes with the same setting as in Illustration 2.1;

N_t is Poisson process for counting jump activities with constant instantaneous intensity $c(t) = c_0$;

Z_t^S denote the random variable of jump size, when jump happens, and $Z_t^s \sim \mathcal{N}(\mu_s, \sigma_s)$, and $\bar{\mu}_s = \mathbb{E}\left[e^{Z_t}\right] - 1 = e^{\mu_s + \frac{1}{2}\sigma_s^2} - 1$;

By adding jumps to the return dynamic of the Heston model, Bates (1996) proposes a stochastic volatility jump model. The jumps are modeled as a compounded Poisson process, which is composed of a normally distributed jump size and a constant jump intensity. The dynamics of the Bates model are defined as follows:

Illustration 2.2. Bates model

under $\mathbb{P}$

$$\begin{aligned} \frac{dS_t}{S_t} &= \mu_t dt + \sqrt{V_t} dW_t + \left(e^{Z_t^s} - 1\right) dN_t - c(t)\bar{\mu}_s dt \\ dV_t &= \kappa\left(\theta - V_t\right) dt + \sigma\sqrt{V_t} dB_t; \end{aligned} \tag{2.52}$$

under $\mathbb{Q}$

$$\begin{aligned} \frac{dS_t}{S_t} &= (r_t - q_t)dt + \sqrt{V_t} dW_t^{\mathbb{Q}} + \left(e^{Z_t^s} - 1\right) dN_t^{\mathbb{Q}} - c(t)^{\mathbb{Q}}\bar{\mu}_s^{\mathbb{Q}} dt \\ dV_t &= \kappa^{\mathbb{Q}}\left(\theta^{\mathbb{Q}} - V_t\right) dt + \sigma\sqrt{V_t} dB_t^{\mathbb{Q}}, \end{aligned} \tag{2.53}$$

where

μ_t, r_t and q_t denote the same terms as in Illustration 2.1;

κ, θ and σ are given as in Definition 2.2 under $\mathbb{P}$;

W_t and B_t are standard Wiener processes with the same setting as in Illustration 2.1;

N_t, $c(t)$ and Z_t^s follow the same definition of the Poisson process, jump intensity and random jump size as in the Merton model (see Definition 2.3);

$\left(\kappa^{\mathbb{Q}}, \theta^{\mathbb{Q}}\right)$ are the corresponding values for the volatility process under risk-neutral measure and σ is the same under both measures as in Illustration 2.1;

$\left(\bar{\mu}_s^{\mathbb{Q}}, \sigma_s^{\mathbb{Q}}, c_0^{\mathbb{Q}}\right)$ [18] are the corresponding values for jump part under risk-neutral measure;

μ_t is defined in a similar manner as in Illustration 2.1, whereas γ_t contains a additional jump part, as follows,

$$\gamma_t = \eta_s V_t + c_0\bar{\mu} - c_0^{\mathbb{Q}}\bar{\mu}^{\mathbb{Q}}. \tag{2.54}$$

[18] In Illustration 2.2, these three terms are specified differently under both measures. As in literature, however, the same settings for jump in return impose that $c_0 = c_0^{\mathbb{Q}}$ and $\sigma_s = \sigma_s^{\mathbb{Q}}$.

$W_{i,t}^{\mathbb{Q}}, B_{i,t}^{\mathbb{Q}}, N^{\mathbb{Q}}$ and $Z^{\mathbb{Q}}$ are the corresponding risk-neutral Wiener processes, Poisson processes and jump sizes.

The Bates model (Illustration 2.2) only includes a two-sided Gaussian random jump size, but it is widely observed that large infrequent price movements happen differently in both directions, which causes the tails to differ. In order to accommodate this feature, Kou (2002) proposes a model with double exponential jumps in return. For simplicity only the dynamic under physical measure of Kou's model is discussed. The definition is given by

Definition 2.4. Double exponential jump model
under $\mathbb{P}$

$$\begin{aligned}\frac{dS_t}{S_t} &= \mu_t dt + \sigma dW_t + \left(e^{Z_t^+} - 1\right) dN_t^+ - c^+(t)\bar{\mu}^+ dt \\ &\quad + \left(e^{Z_t^-} - 1\right) dN_t^- - c^-(t)\bar{\mu}^- dt,\end{aligned} \tag{2.55}$$

where

μ_t, σ and W_t are defined in the same manner as in Illustration 2.1;

Z_t^+ and $\tilde{Z_t}^-(Z_t^- := -\tilde{Z_t}^-)$[19]are exponentially distributed jump size with mean $1/\lambda^+$ and $1/\lambda^-$ respectively;

$\bar{\mu}^+ = \mathbb{E}\left[e^{Z_t^+}\right] - 1 = (1 - (\lambda^+)^{-1})^{-1}$ and $\bar{\mu}^- = \mathbb{E}\left[e^{Z_t^-}\right] - 1 = (1 - (-\lambda^-)^{-1})^{-1}$;

N_t^+ and N_t^- are Poisson processes for counting jump activities for positive and negative jumps with instantaneous intensities $c^+(t) = c_0$ and $c^-(t) = 1 - c_0$ respectively.

2.2.3 Jumps in Volatility and Stochastic Jump Intensity

A model that incorporates jumps improves the statistical fit of the tails of the return distribution in. One may, however, wonder whether the independent Poisson process with constant intensity is able to capture the empirical patterns of large price movements. Using constant jump intensity remains limited, especially in the case of time-invariant crash risk, where the likelihood of another crash occurrence remains the same as if there had been one the day before. In order to resolve this potential misfit, Bates (2000) and Pan (2002) propose an approach to model the stochastic jump intensity, as follows:

$$c(t) = c_0 + c_1 V_t,^{20} \tag{2.56}$$

[19]Specifically, the probability density function of Z_t^- is $f(x;\lambda^-) = \begin{cases} 0, & \text{if } x > 0 \\ \lambda^- e^{-\lambda^- |x|}, & \text{if } x \leqslant 0 \end{cases}$.

where V_t denotes the volatility term the same as in Illustration 2.1. Based on this style of link between jump intensity and volatility, high volatility follows high jump probability.[21] Even so, there is still a potential misfit if the probability of the occurrence of a jump is tied to volatility in the linear relationship. In the literature, realized volatility is similar before and after the 1987 market crash, but the time-invariant probability of jumps, which is derived from a linear relationship between jump intensity and volatility, is misspecified. Fulop and Li (2019), among others use the self-exciting jump, which means the large price movement is followed by a second large price movement. As a result, AJD models are extended to include another process that governs the dynamic of jump intensity. In addition, recent empirical studies indicate that an increase in asset prices is related to an increase in asset volatility, indicating a co-jump between price and volatility.

Fulop and Li (2019) uses several models[22] that accommodate the aforementioned empirical patterns for their study on S&P 500 index and its variance swaps. Following the definition of their AJD models, the same definition of dynamic of underlying stock price, volatility and jump intensity are given as follows:

Illustration 2.3. SVUJ

under $\mathbb{P}$

$$\begin{aligned} \frac{dS_t}{S_t} &= \mu_t\, dt + \sqrt{V_t} dW_{1,t} + \left(e^{Z_t^s} - 1\right) dN_t, -c(t)\bar{\mu}_s dt, \\ dV_t &= \kappa_v \left(\theta_v - V_t\right)\, dt + \sigma_v \sqrt{V_t} dB_{1,t} + Z_t^v dN_t, \\ dU_t &= \kappa_u \left(\theta_u - U_t\right)\, dt + \sigma_u \sqrt{U_t} dW_{2,t} + \beta dN_t, \end{aligned} \tag{2.57}$$

under $\mathbb{Q}$

$$\begin{aligned} \frac{dS_t}{S_t} &= \left(r_t - q_t\right) dt + \sqrt{V_t} dW_{1,t}^{\mathbb{Q}} + \left(e^{Z_t^{S,\mathbb{Q}}} - 1\right) dN_t^{\mathbb{Q}} \\ &\quad - c(t)\bar{\mu}_s^{\mathbb{Q}} dt, \\ dV_t &= \kappa_v^{\mathbb{Q}} \left(\theta_v^{\mathbb{Q}} - V_t\right)\, dt + \sigma_v \sqrt{V_t} dB_{1,t}^{\mathbb{Q}} + Z_t^{v,\mathbb{Q}} dN_t^{\mathbb{Q}}, \\ dU_t &= \kappa_u^{\mathbb{Q}} \left(\theta_u^{\mathbb{Q}} - V_t\right)\, dt + \sigma_u \sqrt{U_t} dW_{2,t}^{\mathbb{Q}} + \beta^{\mathbb{Q}} dN_t^{\mathbb{Q}} \end{aligned} \tag{2.58}$$

where

μ_t denotes the drift in return under physical measure;

r_t and q_t denotes the risk-free rate and dividend rate, respectively;

[20] In Bates (2000), the relationship is given as $c(t) = c_0 + c_1 V_{1,t} + c_2 V_{2,t}$ due to the multi-factor setting for volatility in his model.

[21] The relationship is true, when c_1 is assumed to be positive.

[22] Both affine and non-affine models are used in their paper.

κ_v, θ_v and σ_v are given as in Definition 2.2 under $\mathbb{P}$ for the factor V_t;

κ_u, θ_u and σ_u are given as in Definition 2.2 under $\mathbb{P}$ for the factor U_t;

$W_{1,t}$, $W_{2,t}$ and $B_{1,t}$ are standard Wiener processes. The processes $W_{1,t}$, $W_{2,t}$ are independent. The processes $W_{1,t}$ and $B_{1,t}$ are correlated with coefficient $\rho \in [-1, 1]$;

N_t is Poisson process with the jump intensity $c(t) = U_t$;

Z_t^s is the jump size with normal distribution with mean μ_s and variance $\sigma_s (= \sigma_s^{\mathbb{Q}})$ and $\bar{\mu}_s = \mathbb{E}\left[e^{Z_t}\right] - 1 = e^{\mu_s + \frac{1}{2}\sigma_s^2} - 1$;

Z_t^v is the random jump size with exponentially distribution with mean $1/\lambda_v$;

β denotes the constant jump size in jump intensity process;

$\left(\kappa_{v(,u)}^{\mathbb{Q}}, \theta_{v(,u)}^{\mathbb{Q}}, \bar{\mu}_s^{\mathbb{Q}}, \beta^{\mathbb{Q}}\right)$ are the corresponding values for the process risk-neutral measure. $W_{1,t}^{\mathbb{Q}}, W_{1,t}^{\mathbb{Q}}, B_{1,t}^{\mathbb{Q}}, N_t^{\mathbb{Q}}, Z_t^{s,\mathbb{Q}}$ and $Z_t^{v,\mathbb{Q}}$ are the corresponding Wiener processes, Poisson process and random jump sizes under risk-neutral measure.

In addition, for the risk premium of variance and jump intensity one has $\eta_{v(,u)} = \kappa_{v(,u)}^{\mathbb{Q}} - \kappa_{v(,u)}$, $\theta_{v(,u)}^{\mathbb{Q}} = \frac{\kappa_{v(,u)}\theta_{v(,u)}}{\kappa_{v(,u)}^{\mathbb{Q}}}$. This risk-premia specification imposed a restriction $\kappa_{v(,u)}\theta_{v(,u)} = \kappa_{v(,u)}^{\mathbb{Q}}\theta_{v(,u)}^{\mathbb{Q}}$ (see Broadie et al. (2007)). The risk-neutral model, as mentioned above, has exactly the same structure as the one that is under the physical measure. Finally, based on the specification of risk premiums for different risks, $\mu_t = r_t - q_t + \gamma_t$ with

$$\gamma_t = \eta_s V_t + U_t\left(\bar{\mu}_s - \bar{\mu}_s^{\mathbb{Q}}\right).^{23}$$

In addition, there is another series of multi-factor AJD models under $\mathbb{Q}$ proposed by Andersen et al. (2015b) and Andersen et al. (2020) [24], who use information from option surfaces to characterize the dynamics of the underlying return and extract latent factors. In Andersen et al. (2015b), they show that their three-factor model provides a superior out-of-sample fit to implied volatility skew and term structure.[25] Following Kou (2002), their models include two separate exponentially distributed jumps in the return dynamic.

[23] In Fulop and Li (2019), they leave μ_t unspecified, but constant.

[24] With the help of the estimation methodology by Andersen et al. (2015a), they avoid specifying the return dynamic physical measure $\mathbb{P}$, as the task is only to extract latent states and estimate static parameters under $\mathbb{Q}$.

[25] Specifically, the IV skew measures the IV gap between short-dated OTM put and OTM call options, whereas the skew term structure is obtained from the difference between the IVs computed from long- and short-dated options (see Andersen et al. (2015b), page 561).

Moreover, there is a separate process that determines jump intensity: U_t, a pure jump process without diffusion risk in it. Based on the estimation results for S&P 500 index option surfaces, it is evident that only the coefficient of negative jump intensity is significantly different from zero. As a result, they allow only negative jumps associated with stochastic jump intensity. When volatility remains constant, as demonstrated in the aforementioned empirical situation, the intensity of jumps can change dramatically.

Meanwhile, the random jump sizes in volatility and jump intensity are deterministically related to the negative jump size in the return process. Additionally, the model proposed by Andersen et al. (2015b) is equipped with two different volatility factors and a jump intensity factor. Andersen et al. (2020), on the other hand, confine their model to a single volatility factor and a single jump intensity factor. Furthermore, they demonstrate that the restricted version of the model can produce nearly the same results for state recovery as Andersen et al. (2015b)'s model.

In this regard, I present only the lean version of their models. The model dynamics are defined as follows:

Definition 2.5. AFTQ
under $\mathbb{Q}$:

$$\begin{aligned}
\frac{dS_t}{S_t} &= (r_t - q_t)\,dt + \sqrt{V_t}dW_{1,t}^{\mathbb{Q}} + \left(e^{Z_t^{+,\mathbb{Q}}} - 1\right) dN_t^{+,\mathbb{Q}} - c^{+}(t)\,\bar{\mu}^{+,\mathbb{Q}}dt \\
&\quad + \left(e^{Z_t^{-,\mathbb{Q}}} - 1\right) dN_t^{-,\mathbb{Q}} - c^{-}(t)\,\bar{\mu}^{-,\mathbb{Q}}dt \\
dV_t &= \kappa_v^{\mathbb{Q}}\left(\theta_v^{\mathbb{Q}} - V_t\right)dt + \sigma_v\sqrt{V_t}dB_{1,t}^{\mathbb{Q}} + \mu_v\left(Z_t^{-,\mathbb{Q}}\right)^2 dN_t^{-,\mathbb{Q}} \\
dU_t &= -\kappa_u^{\mathbb{Q}}U_t dt + \mu_u\left(Z_t^{-,\mathbb{Q}}\right)^2 dN_t^{-,\mathbb{Q}},
\end{aligned} \tag{2.59}$$

where

$\kappa_v^{\mathbb{Q}}, \theta^{\mathbb{Q}}$ and σ_v are given as in Definition 2.2 under $\mathbb{Q}$ for the factor V_t;

$\kappa_u^{\mathbb{Q}}$ and σ_u are given as in Definition 2.2 under $\mathbb{Q}$ for the factor U_t;

$W_{1,t}^{\mathbb{Q}}$, $W_{2,t}^{\mathbb{Q}}$ and $B_{1,t}^{\mathbb{Q}}$ are standard Wiener processes. The processes $W_{1,t}^{\mathbb{Q}}$, $W_{2,t}^{\mathbb{Q}}$ are independent. The processes $W_{1,t}^{\mathbb{Q}}$ and $B_{1,t}^{\mathbb{Q}}$ are correlated with coefficient $\rho \in [-1, 1]$;

$N_t^{+,\mathbb{Q}}$ and $N_t^{-,\mathbb{Q}}$ are Poisson processes for counting jump activities for positive and negative jumps with instantaneous intensities $c^{+}(t) = c_0^{+}$ and $c^{-}(t) = U_t$ respectively;

Moreover, the identity of $N_t^{-,\mathbb{Q}}$ in all the factors implies that S_t shares the negative

exponential jump, V_t and U_t, i.e., the jump in S_t, V_t and U_t arises at the same time with the same realization of jump size;

$Z_t^{+,\mathbb{Q}}$ and $\tilde{Z}_t^{\,-,\mathbb{Q}}$($Z_t^{-,\mathbb{Q}} := -\tilde{Z}_t^{\,-,\mathbb{Q}}$)[26]are exponentially distributed jump size with mean $1/\lambda^+$ and $1/\lambda^-$ respectively;

$\bar{\mu}^{+,\mathbb{Q}} = \mathbb{E}\left[e^{Z_t^{+,\mathbb{Q}}}\right] - 1 = (1 - (\lambda^{+,\mathbb{Q}})^{-1})^{-1}$ and $\bar{\mu}^{-,\mathbb{Q}} = \mathbb{E}\left[e^{Z_t^{-,\mathbb{Q}}}\right] - 1 = (1 - (-\lambda^{-,\mathbb{Q}})^{-1})^{-1}$;

As a final step, the previously discussed building blocks are blended together. This creates a "new" model that will serve as the primary concrete working model in subsequent chapters. The model accommodates stochastic volatility and stochastic negative jump intensity. Additionally, its specifications of jump sizes in return, volatility, and jump intensity are similar to that of Andersen et al. (2020). Therefore, following Andersen et al. (2020) , Bardgett et al. (2019)[27], Gruber et al. (2021) and the assumption of equivalence of model structure under the physical measure ($\mathbb{P}$) and the risk-neutral measure ($\mathbb{Q}$), the data generating processes for the stock price S_t, the instantaneous diffusive variance V_t, and the negative jump intensity U_t under both measures are defined as follows:

Illustration 2.4. SVNUJ

under $\mathbb{P}$:

$$\begin{aligned}
\frac{dS_t}{S_t} &= \mu_t dt + \sqrt{V_t} dW_{1,t} + \left(e^{Z_t^+} - 1\right) dN_t^+ - c^+(t)\bar{\mu}^+ dt \\
&\quad + \left(e^{Z_t^-} - 1\right) dN_t^- - c^-(t)\bar{\mu}^- dt \\
dV_t &= \kappa_v \left(\theta_v - V_t\right) dt + \sigma_v \sqrt{V_t} dB_{1,t} + \mu_v \left(Z_t^-\right)^2 dN_t^- \\
dU_t &= \kappa_u(\theta_u - U_t)dt + \sigma_u \sqrt{U_t} dW_{2,t} + \mu_u \left(Z_t^-\right)^2 dN_t^-,
\end{aligned} \tag{2.60}$$

The log return process for the stock price $\ln S_t$ has the form via applying Itô.

$$d\ln S_t = \left(\mu_t - \frac{1}{2}V_t\right) dt + \sqrt{V_t} dW_{1,t} + Z_t^+ dN_t^+ - c^+(t)\bar{\mu}^+ dt + Z_t^- dN_t^- - c^-(t)\bar{\mu}^- dt.^{28}$$

---

[26]Specifically, the probability density function of $Z_t^{-,\mathbb{Q}}$ is $f(x;\lambda^{-,\mathbb{Q}}) = \begin{cases} 0, & \text{if } x > 0 \\ \lambda^{-,\mathbb{Q}} e^{-\lambda^{-,\mathbb{Q}}|x|}, & \text{if } x \leqslant 0 \end{cases}$.

[26]Based on Definition 2.5, the dynamic of U_t under $\mathbb{Q}$ has the form:

$$dU_t = -\kappa_u^{\mathbb{Q}} U_t dt + \mu_u \left(Z_t^-\right)^2 dN_t^{-,\mathbb{Q}},$$

which is equivalent to specifying $\theta_u = 0$ and $\sigma_u = 0$ in Eq. (2.60) or Eq. (2.61)

[27]According to Bardgett et al. (2019), the dynamic of U_t under $\mathbb{P}$ and $\mathbb{Q}$ is defined as follows:

$$dU_t = \kappa_u(\theta_u - U_t)dt + \sigma_u \sqrt{V_t} dW_{2,t},$$

where jumps are excluded.

under $\mathbb{Q}$:

$$\begin{aligned}
\frac{dS_t}{S_t} &= (r_t - q_t)\,dt + \sqrt{V_t}dW_{1,t}^{\mathbb{Q}} + \left(e^{Z_t^{+,\mathbb{Q}}} - 1\right) dN_t^{+,\mathbb{Q}} - c^{+}(t)\,\bar{\mu}^{+,\mathbb{Q}}dt \\
&\quad + \left(e^{Z_t^{-,\mathbb{Q}}} - 1\right) dN_t^{-,\mathbb{Q}} - c^{-}(t)\,\bar{\mu}^{-,\mathbb{Q}}dt \\
dV_t &= \kappa_v^{\mathbb{Q}}\left(\theta_v^{\mathbb{Q}} - V_t\right)dt + \sigma_v\sqrt{V_t}dB_{1,t}^{\mathbb{Q}} + \mu_v\left(Z_t^{-,\mathbb{Q}}\right)^2 dN_t^{-,\mathbb{Q}} \\
dU_t &= \kappa_{u,\mathbb{Q}}(\theta_u^{\mathbb{Q}} - U_t)dt + \sigma_u\sqrt{U_t}dW_{2,t}^{\mathbb{Q}} + \mu_u\left(Z_t^{-,\mathbb{Q}}\right)^2 dN_t^{-,\mathbb{Q}},
\end{aligned} \tag{2.61}$$

where

$W_{1,t}$, $W_{2,t}$ and $B_{1,t}$ are standard Wiener processes. The processes $W_{1,t}$, $W_{2,t}$ are independent. The processes $W_{1,t}$ and $B_{1,t}$ are correlated with coefficient $\rho \in [-1, 1]$;

N_t^+ and N_t^- are Poisson processes for counting jump activities for positive and negative jumps with instantaneous intensities $c^+(t) = c_0^+$ and $c^-(t) = U_t$ respectively;

The identity of N_t^- and $N_t^{-,\mathbb{Q}}$ imply the same pattern as in Definition 2.5 for both measures;

Z_t^+ and $\tilde{Z_t}^-(Z_t^- := -\tilde{Z_t}^-)$[29]are exponentially distributed jump size with mean $1/\lambda^+$ and $1/\lambda^-$ respectively;

$\bar{\mu}^+ = \mathbb{E}\left[e^{Z_t^+}\right] - 1 = (1 - (\lambda^+)^{-1})^{-1}$ and $\bar{\mu}^- = \mathbb{E}\left[e^{Z_t^-}\right] - 1 = (1 - (-\lambda^-)^{-1})^{-1}$;

$\left(\kappa_v^{\mathbb{Q}}, \theta_v^{\mathbb{Q}}, \kappa_u^{\mathbb{Q}}, \theta_u^{\mathbb{Q}}\right)$ are the corresponding values for the process risk-neutral measure;

$\left(\lambda^{+,\mathbb{Q}}, \lambda^{-,\mathbb{Q}}, \bar{\mu}^{+,\mathbb{Q}}, \bar{\mu}^{-,\mathbb{Q}}\right)$ are the corresponding values for the process risk-neutral measure, meaning that the the decay parameters of jump sizes can be different by following Bardgett et al. (2019) and Gruber et al. (2021);

$W_{1,t}^{\mathbb{Q}}, W_{2,t}^{\mathbb{Q}}, B_{1,t}^{\mathbb{Q}}, N^{+(,-),\mathbb{Q}}$ and $Z^{+(,-),\mathbb{Q}}$ are the corresponding risk-neutral Wiener processes, Poisson processes and exponential jump sizes. Moreover, one follows the literature (see Broadie et al. (2007), Bardgett et al. (2019) and Gruber et al. (2021)) by assuming that both $\mathbb{P}$ and $\mathbb{Q}$ are associated with the same jump intensities.

[28]In the implementation we have

$$d\ln S_t = \left(r_t - q_t + (\eta_t - 0.5)V_t - c^+(t)\bar{\mu}^{+,\mathbb{Q}} - c^-(t)\bar{\mu}^{-,\mathbb{Q}}\right)dt + \sqrt{V_t}dW_{1,t} + Z_t^+ dN_t^+ + Z_t^- dN_t^-$$

[29]Specifically, the probability density function of Z_t^- is $f(x;\lambda^-) = \begin{cases} 0, & \text{if } x > 0 \\ \lambda^- e^{-\lambda^- |x|}, & \text{if } x \leqslant 0 \end{cases}$.

The drift in return $\mu_t = r_t - q_t + \gamma_t$ under $\mathbb{P}$, with excess return γ_t, which is defined of the form

$$\gamma_t = \eta_s V_t + c^+(t)(\bar{\mu}^+ - \bar{\mu}^{+,\mathbb{Q}}) + c^-(t)(\bar{\mu}^- - \bar{\mu}^{-,\mathbb{Q}}). \tag{2.62}$$

As mentioned among Bates (2000), Broadie et al. (2007), Bardgett et al. (2019), Fulop and Li (2019) and Gruber et al. (2021), the "differences" between parameters under $\mathbb{P}$ and $\mathbb{Q}$ are interpreted as risk premiums in arbitrage-free dynamic asset pricing models. Let $1/\lambda^+ - 1/\lambda^{+,\mathbb{Q}}$ (resp.$\big(-1/\lambda^- - (-1/\lambda^{-,\mathbb{Q}})\big)$) denote the mean of the jump risk premium for a positive (resp. negative) jump. The diffusive volatility risk premium [30] and the negative jump intensity premium [31] are defined as $\eta_i = \kappa_i - \kappa_i^{\mathbb{Q}}, i \in \{v, u\}$ respectively. Furthermore, the market risk premium for volatility is assumed to be linear. Finally, ρ, σ_i and $\kappa_i\theta_i, i \in \{v, u\}$ are restricted to be the same under $\mathbb{P}$ and $\mathbb{Q}$[32].

2.3 Model-free Risk-neutral Moments and Cumulants

As mentioned in Bakshi et al. (2003), any payoff function with bounded expectation can be spanned by a continuum of out-of-the-money European calls and puts based on the results of Bakshi and Madan (2000). Therefore, the entire set of twice-continuously differentiable payoff functions, $H(S)$, can be spanned as follows:

$$\begin{aligned} H(S) &= H(S^*) + (S - S^*)H_S(S^*) + \int_{S^*}^{\infty} H_{SS}(K)(S-K)^+ dK \\ &= \int_0^{S^*} H_{SS}(K)(K-S)^+ dK. \end{aligned} \tag{2.63}$$

Assuming the risk-free rate r_f is constant between t and $t+\tau$, the fair market value of the contingent-claim $H(S)$ is as follows:

$$\begin{aligned} \mathbb{E}_t^{\mathbb{Q}}\left[e^{-r_f\tau}H(S)\right] = e^{-r_f\tau}\left(H(S^*) - S^*H_S(S^*)\right) + H_S(S^*)S_t \\ + \int_{S^*}^{\infty} H_{SS}(K)\mathcal{C}\left(t,\tau,K\right)dK \\ + \int_0^{S^*} H_{SS}(K)\mathcal{P}\left(t,\tau,K\right)dK. \end{aligned} \tag{2.64}$$

[30] In Bates (2000), he restricts $\kappa_v^{\mathbb{Q}} - \kappa_v < 0$. It implies negative variance risk premium.

[31] In Bardgett et al. (2019) they restrict $\kappa_u^{\mathbb{Q}} - \kappa_u = 0$, while in Fulop and Li (2019) $\eta_u = \left(\kappa_u^{\mathbb{Q}} - \kappa\right)/\sigma_u$.

[32] These identity restrictions are imposed in the literature. For further explanation, see Broadie et al. (2007)

The log return of stock price is defined as $R_s(t,\tau) := \ln S_{t+\tau} - \ln S_t$ in the previous section, and $M_{n,R_s(t,\tau)} = \mathbb{E}_t^{\mathbb{Q}}\left[R_s^n(t,\tau)\right]$ follows from the definition of *raw* moments in Eq. (2.10). Furthermore, Bakshi et al. (2003) defines the fair market value of *raw* moments from the second-order to fourth-order: $\mathcal{V}(t,\tau) = \mathbb{E}_t^{\mathbb{Q}}\left[e^{-r_f\tau}R_s^2(t,\tau)\right]$, $\mathcal{W}(t,\tau) = \mathbb{E}_t^{\mathbb{Q}}\left[e^{-r_f\tau}R_s^3(t,\tau)\right]$, and $\mathcal{Z}(t,\tau) = \mathbb{E}_t^{\mathbb{Q}}\left[e^{-r_f\tau}R_s^4(t,\tau)\right]$. Based on equation Eq. (2.64), those fair values are observable weighted portfolios of OTM option contracts:

$$
\begin{aligned}
\mathcal{V}(t,\tau) &= \int_{S_t}^{\infty} \frac{2\left(1-\ln\left(\frac{K}{S_t}\right)\right)}{K^2}\mathcal{C}\left(t,\tau,K\right)dK \\
&\quad + \int_0^{S_t} \frac{2\left(1+\ln\left(\frac{S_t}{K}\right)\right)}{K^2}\mathcal{P}\left(t,\tau,K\right)dK, \\
\mathcal{W}(t,\tau) &= \int_{S_t}^{\infty} \frac{6\ln\left(\frac{K}{S_t}\right) - 3\left(\ln\left(\frac{K}{S_t}\right)\right)^2}{K^2}\mathcal{C}\left(t,\tau,K\right)dK \\
&\quad - \int_0^{S_t} \frac{6\ln\left(\frac{S_t}{K}\right) + 3\left(\ln\left(\frac{S_t}{K}\right)\right)^2}{K^2}\mathcal{P}\left(t,\tau,K\right)dK, \\
\mathcal{Z}(t,\tau) &= \int_{S_t}^{\infty} \frac{12\left(\ln\left(\frac{K}{S_t}\right)\right)^2 - 4\left(\ln\left(\frac{K}{S_t}\right)\right)^3}{K^2}\mathcal{C}\left(t,\tau,K\right)dK \\
&\quad + \int_0^{S_t} \frac{12\left(\ln\left(\frac{S_t}{K}\right)\right)^2 + 4(\ln\left(\frac{S_t}{K}\right))^3}{K^2}\mathcal{P}\left(t,\tau,K\right)dK.
\end{aligned}
\tag{2.65}
$$

Following Bakshi et al. (2003)[33], we have the $\mu(t,\tau) := \mathbb{E}_t^{\mathbb{Q}}\left[R_s(t,\tau)\right]$, which can be derived as follows:

$$
\mu(t,\tau) = e^{r_f\tau} - 1 - \frac{e^{r_f\tau}}{2}\mathcal{V}(t,\tau) - \frac{e^{r_f\tau}}{6}\mathcal{W}(t,\tau) - \frac{e^{r_f\tau}}{24}\mathcal{Z}(t,\tau).
$$

In addition, it is popular to characterize the density using a normalized version of central moments. It is worth stressing that the second-order normalized central moments are the same as the second cumulant. In Bakshi et al. (2003), the second, third, and fourth normalized central moments are defined as variance, skewness, and kurtosis, respectively[34]:

$$
\begin{aligned}
\mathrm{Var}(R_s(t,\tau)) &:= \mathbb{E}_t^{\mathbb{Q}}\left[\left(R_s(t,\tau) - \mathbb{E}_t^{\mathbb{Q}}\left[R_s(t,\tau)\right]\right)^2\right] \\
&= e^{r_f\tau}\mathcal{V}(t,\tau) - \mu(t,\tau)^2,
\end{aligned}
\tag{2.66}
$$

[33]Proof of Theorem 1. in Appendix of Bakshi et al. (2003)

[34]Vilkov (2021) provides implementation details for the various versions of risk-neutral moments.

$$\begin{aligned}\text{skew}(R_s(t,\tau)) &:= \frac{\mathbb{E}_t^{\mathbb{Q}}\left[\left(R_s(t,\tau)-\mathbb{E}_t^{\mathbb{Q}}\left[R_s(t,\tau)\right]\right)^3\right]}{\mathbb{E}_t^{\mathbb{Q}}\left[\left(R_s(t,\tau)-\mathbb{E}_t^{\mathbb{Q}}\left[R_s(t,\tau)\right]\right)^2\right]^{\frac{3}{2}}} \\ &= \frac{e^{r_f\tau}\mathcal{W}(t,\tau)-3\mu(t,\tau)e^{r_f\tau}\mathcal{V}(t,\tau)+2\mu(t,\tau)^3}{(e^{r_f\tau}\mathcal{V}(t,\tau)-\mu(t,\tau)^2)^{\frac{3}{2}}},\end{aligned} \tag{2.67}$$

and

$$\begin{aligned}\text{kurt}(R_s(t,\tau)) &:= \frac{\mathbb{E}_t^{\mathbb{Q}}\left[\left(R_s(t,\tau)-\mathbb{E}_t^{\mathbb{Q}}\left[R_s(t,\tau)\right]\right)^4\right]}{\mathbb{E}_t^{\mathbb{Q}}\left[\left(R_s(t,\tau)-\mathbb{E}_t^{\mathbb{Q}}\left[R_s(t,\tau)\right]\right)^2\right]^2} \\ &= \frac{e^{r_f\tau}\mathcal{Z}(t,\tau)-4\mu(t,\tau)e^{r_f\tau}\mathcal{W}(t,\tau)-6e^{r_f\tau}\mu(t,\tau)^2\mathcal{V}(t,\tau)+3\mu(t,\tau)^4}{\left(e^{r_f\tau}\mathcal{V}(t,\tau)-\mu(t,\tau)^2\right)^2}.\end{aligned} \tag{2.68}$$

The relationships between cumulants and normalized central moments are

$$\begin{aligned}\mathcal{K}_{2,R_s(t,\tau)} &= \mathbb{E}\left[\left(R_s(t,\tau)-\mathbb{E}\left[R_s(t,\tau)\right]\right)^2\right] = \text{Var}(R_s(t,\tau)), \\ \mathcal{K}_{3,R_s(t,\tau)} &= \mathbb{E}\left[\left(R_s(t,\tau)-\mathbb{E}\left[R_s(t,\tau)\right]\right)^3\right], \\ \mathcal{K}_{4,R_s(t,\tau)} &= \mathbb{E}\left[\left(R_s(t,\tau)-\mathbb{E}\left[R_s(t,\tau)\right]\right)^4\right] - 3\text{Var}(R_s(t,\tau))^2.\end{aligned} \tag{2.69}$$

Thus, the model-free cumulants can be obtained as follow:

$$\begin{aligned}\mathcal{K}_{2,R_s(t,\tau)} &= \text{Var}(R_s(t,\tau)), \\ \mathcal{K}_{3,R_s(t,\tau)} &= \text{skew}(R_s(t,\tau))\text{Var}(R_s(t,\tau))^{\frac{3}{2}}, \\ \mathcal{K}_{4,R_s(t,\tau)} &= (\text{kurt}(R_s(t,\tau))-3)\,\text{Var}(R_s(t,\tau))^2.\end{aligned} \tag{2.70}$$

3 Inference Methods

In this chapter, the essential technicalities and methodologies involved in performing inference tasks for latent affine jump-diffusions are discussed. As mentioned in the introduction, one of the primary objectives of this thesis is to provide an overview of how to estimate AJD models in empirical studies with limited computational capabilities. A review of several essential techniques and methodologies for tackling estimation tasks in the literature is presented in this chapter. The methods related to or used in the subsequent empirical application will be explained in more detail. The rest of this chapter is organized as follows: Section 3.1 discusses the primary task of estimating AJD models within the no-arbitrage setting and gives a brief survey of literature about various estimation approaches. Section 3.2 discusses general methods for translating AJD models into a state-space model. Section 3.3 reviews the two widely used filtering techniques used in the literature. Section 3.4 presents the penalized nonlinear least squares method proposed by Andersen et al. (2015a). Finally, Section 3.5 discusses likelihood-based estimation methodologies.

3.1 Review of the Literature

Affine jump-diffusions are widely used to model the dynamics of asset returns and the associated statistical factors, as discussed in Section 2.1. In particular, logarithmic equity price movements are typically modeled using AJDs in the empirical analysis of dynamic asset pricing models. The processes are composed of one primary process which represents the return dynamic, and a number of accompanying processes which describe the persistent time-varying statistical factors that characterize conditional distributions. The accompanying processes often represent hidden factors, such as stochastic volatility terms, jump intensities, and long-run means. For example, the Heston model for option pricing - an affine jump-diffusion version of modeling stochastic volatility of logarithmic

equity returns, is discussed in Illustration 2.1. In addition, the stochastic jump intensity (see, e.g., Fulop and Li (2019) and Andersen et al. (2015b)) and stochastic long-run mean [35] (see, e.g., Bardgett et al. (2019)) are used in the literature. Moreover, in the literature on term structure models (see, e.g., Dai and Singleton (2000), Duan and Simonato (1999), among others), the dynamics of interest rates are also modeled by AJDs. Based on the discussion in Section 2.1 concerning the reduced form dynamic asset pricing models, the actual density of state transitions between sampling points is determined under the physical measure, while the fair prices in the no-arbitrage setting are determined by the state-dependent time-varying conditional density under $\mathbb{Q}$. Thus, observations of forward-looking quantities[36] can be viewed as information about the state-dependent conditional density at each sampling time point. Meanwhile, the trajectories of persistent time-varying latent states can be viewed as a realization of latent factors' dynamics over the entire timeframe. Hence, forward-looking quantities can be used both to determine static parameters under risk-neutral measures and to identify the trajectories of latent states simultaneously. Due to the discrete sampling interval and the state-dependent model implied conditional distribution, the estimation procedure involves two steps, namely, the calibration of static parameters as well as the inference of persistent latent states at every sampling point. Thus, the task of inference for AJDs (more precisely, the latent AJDs) is in general twofold, as follows:

- First, given static model parameters, the recovery of latent states is performed by using information from the observed data for each time point. This filtering step can be seen as a function that returns the trajectories of latent states, given the static parameters of the AJDs.
- Second, parameter calibration is carried out in order to obtain the "best" static parameters for fitting the state-dependent model's implied quantities to the observed data over the entire time span with the filtering function in the first step.

In general, the estimation procedures discussed below follow the twofold approach outlined above. Target models are not necessarily limited by the dynamics modeled by AJD processes. Using rich observations to estimate AJD models, Broadie et al. (2007) are the first to apply a nonlinear least squares approach to estimate static parameters under the risk-neutral measure ($\mathbb{Q}$). They assume that the parameters under the physical measure

[35] To model the term structure of interest rates, Chen (1996) proposes an interest rate model with a stochastic long-run mean.

[36] Quantities include the prices of forward-looking derivatives or their transformations.

($\mathbb{P}$) are known. Andersen et al. (2015a) develop another novel procedure similar to the nonlinear least square approach. They avoid imposing a parametric transition density for the latent states under $\mathbb{P}$. According to Andersen et al. (2015b) and Andersen et al. (2020), rich observations from option markets can be used to identify the static parameters of the models under $\mathbb{Q}$ and the latent states. By incorporating rich observations from the option market, this procedure is capable of estimating static parameters under $\mathbb{Q}$ and inferring latent states.

Suppose the parametric assumption is imposed on the transition of states under $\mathbb{P}$. In that case, the state-space model provides the framework within which standard filtering techniques (e.g., the Kalman filter and particle filter, among others) can be applied to infer latent states. For parametric models in a continuous-time setting, one can apply discretization techniques to obtain the corresponding state-space model. Johannes et al. (2009) initialize a particle filtering-based approach to estimate the trajectories of the latent states for AJD models with asset returns. Christoffersen et al. (2010) use a particle filter and observed return time-series to perform the recovery of latent states, while the static parameters under $\mathbb{Q}$ are calibrated by means of a nonlinear least-square approach. It is also popular in the literature to estimate static parameters and latent states simultaneously by a particle filter. In contrast to the approach of Christoffersen et al. (2010), Bardgett et al. (2019) incorporate index options, VIX options and the underlying's returns in the filtering step inside each particle. The static parameters are calibrated by maximizing the implied likelihood calculated from the particle filter. As an alternative to applying option prices, in Fulop and Li (2019), variance-swaps, which provide information about the second moment of conditional density, are used in each particle.

In contrast to the universality of the particle filter, the Kalman filter is only applicable when the state-space model is linear and Gaussian. Therefore, the Kalman filter is typically used to estimate the term structure model with AJDs in the literature. The linear relationship between bond yield and latent states is concluded in Eq. (2.26). In this case, the Gaussianity of measurement holds if the measurement errors are Gaussian. Transition density with AJDs, however, does not always fulfill the Gaussianity requirement. Duan and Simonato (1999) propose the corresponding approximation for the state equations and make the Kalman filter applicable to the affine term structure model. Furthermore, Feunou and Okou (2018) use the risk-neutral cumulants of return density in the filtering step, since these cumulants are similar to bond yields, which have a linear relationship to the latent states. As discussed in Section 2.1.2, these risk-neutral cumulants can be

obtained from observed option prices in a model-free way. In general, if the primary task is to estimate the risk-neutral density implied by AJDs and infer the trajectories of the latent states for the option market, it is not necessarily necessary to use the index return density under physical measure. With the aid of rich information gathered from option markets, Andersen et al. (2015b) and Andersen et al. (2020) demonstrate that it is possible to estimate the static parameters of their AJD models accurately. As part of the estimation, latent states can be successfully recovered as well.

3.2 State-Space Model and Discretization

3.2.1 State-Space Model

In order to estimate affine term structure models, Duan and Simonato (1999) propose a quasi-maximum likelihood approach based on the Kalman filter. For the estimation task of AJD models using the Kalman filter, these continuous-time models must first be converted into discrete-time models through some discretization methods, and then converted into state-space models.

Let T[37] $> 0, N \in \mathbb{N}, \Delta t = T/N$[38] and $t_n = n\Delta t$. A state-space model contains two discrete time series, which are

$$X_{t_0:t_N} := (X_{t_0}, \ldots, X_{t_N}) \in \mathbb{X} \text{ and } Y_{t_0:t_N} := (Y_{t_0}, \ldots, Y_{t_N}) \in \mathbb{Y}, \tag{3.1}$$

where $\mathbb{X}$ and $\mathbb{Y}$ are multi-dimensional Euclidean spaces.

For notational simplicity, $X_{t_0:t_N}$ and $Y_{t_0:t_N}$ are redefined as

$$X_{0:T} := (X_0, \ldots, X_T) \text{ and } Y_{0:T} := (Y_0, \ldots, Y_T) \tag{3.2}$$

respectively, assuming that the time step between $t-1$ and t is always Δt. In general, the process $Y_{0:T}$ is assumed to be observed, whereas the process $X_{0:T}$ is not. Moreover, for the processes $X_{0:T}$ and $Y_{0:T}$, $y_{0:T} = (y_0, \ldots, y_T)$ and $x_{0:T} = (x_0, \ldots, x_T)$ denote the their realizations, respectively. The static parameters, which specify the concrete dynamics of the models, are summarized in the vector $\theta \in \Theta$.

Definition 3.1. Following Chopin et al. (2020), the definition of a general state-space

[37] In this thesis, all the models are *offline* models, the last time point T is assumed always given.
[38] In this thesis, the time steps are assumed always to be homogeneous.

model is given as follows:

$$\begin{aligned} Y_t &= G\left(X_t, \varepsilon_t; \theta\right), 0 \leq t \leq T \\ X_t &= F\left(X_{t-1}, \epsilon_t; \theta\right), 1 \leq t \leq T \\ X_0 &= \tilde{F}_0\left(\epsilon_0; \theta\right), \end{aligned} \tag{3.3}$$

where G, F denote the deterministic functionals of the measurement equation and state equation, respectively, $\tilde{F}_0$ is the deterministic functional for the initialization of state variables, and ε, ϵ denote the noise in terms of series of i.i.d random variables.

The state-space model can be interpreted in another form with density functions:

$$\mu_0(x_0; \theta) \prod_{t=0}^{T} g(y_t|x_t; \theta) \prod_{t=1}^{T} f(x_t|x_{t-1}; \theta), \tag{3.4}$$

where $g(y_t|x_t; \theta)$ and $f(x_t|x_{t-1}; \theta)$ denote the density function of measurement equation in terms of $Y_t|X_t$, given (x_t, θ) and density function of state equation in terms of $X_t|X_{t-1}$, given (x_{t-1}, θ), respectively. $\mu_0(x_0; \theta)$ denotes the density function of initial distribution.

If there was perfect identification of the observations with their hidden states, it would be possible to evaluate the data distribution. The decomposition

$$p\left(y_{0:T}; \theta\right) = p\left(y_0; \theta\right) \prod_{t=1}^{T} p\left(y_t|y_{0:t-1}; \theta\right) \text{ [39]} \tag{3.5}$$

requires the predictive likelihood

$$p\left(y_t|y_{0:t-1}; \theta\right) = \int g\left(y_t|x_t; \theta\right) f\left(x_t|x_{t-1}; \theta\right) p\left(x_{t-1}|y_{0:t-1}; \theta\right) dx_t dx_{t-1}. \tag{3.6}$$

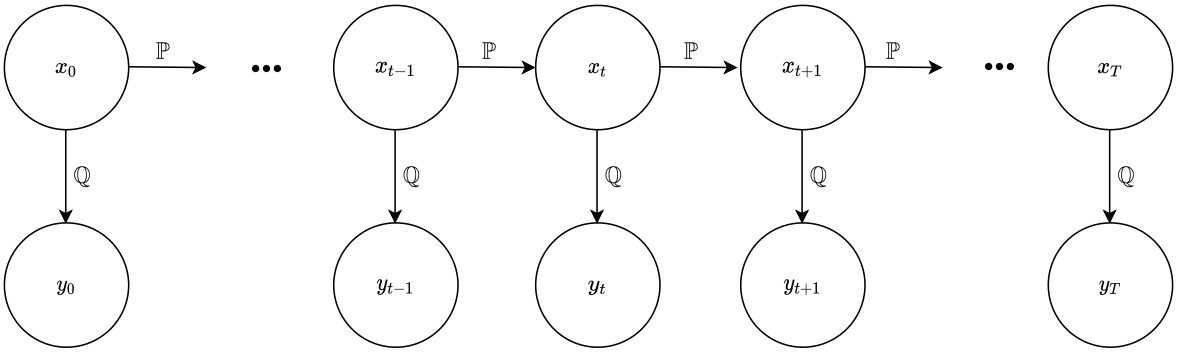
Among others, Duan and Simonato (1999), De Jong (2000) and Duffee and Stanton (2012) suggest a Kalman filter-based approach for estimation of affine term structure models, since bond yields in the context of multi-factor affine models are linear with respect to latent factors by translating them into a linear Gaussian state-space system. In order to fit the AJD models for empirical applications in equity option markets, it is natural to use the option prices, underlying prices, and other observable quantities aggregate from the option panels as observable quantities from the market, whereas the volatility terms, stochastic jump intensities, and other statistical factors can be treated as latent states

[39] $p(\cdot)$ denotes a density function in general.

in the models. As discussed extensively in Section 2.2, for the AJD models, the data generating processes for modeling state equations are assumed under physical measure $\mathbb{P}$ and the measurement equations are modeled under risk-neutral measure $\mathbb{Q}$. Loosely speaking, within the current estimation task, the parameters under $\mathbb{P}$ are tied to state equations, which relate more to state variables as they evolve forward in time, and the parameters under $\mathbb{Q}$ relate more to all observable forward-looking quantities presented by the market. To illustrate the state-space model, its definition is concluded in the Fig. 3.1.

Nevertheless, as demonstrated by Andersen et al. (2015a), if the rich information of

Figure 3.1: The structure of an SSM for a dynamic asset pricing model represented as a graphical model.



forward-looking measurements under risk-neutral measure is taken into account, the trajectories of the latent states over time can also be extracted accurately.

3.2.2 Discretization

Technically the discretization technique shares the essentially same spirit of simulating the trajectories of SDE, i.e., it is equivalent to solving SDE numerically. In accordance with the literature, simulation can be carried out by means of numerical schemes, e.g., the *Euler-Maruyama scheme* (see Fulop and Li (2019) and Christoffersen et al. (2010)) or the *Milstein scheme* (see Bardgett et al. (2019)). For pure diffusive SDE, both schemes exhibit a weak order of convergence[40]of $\gamma = 1$ but our aim is to get consistent estimators for the trajectories necessary for parameter estimations. While the *Milstein* scheme exhibits a strong order of convergence $\gamma = 1$ as opposed to $\gamma = \frac{1}{2}$ for the *Euler-Maruyama scheme* (see Kloeden and Platen (1992)), it involves the numerical approximation of a derivative. This makes it impractical for the AJD models where the runtime increases when options

are incorporated. Throughout this thesis, the *Euler-Maruyama scheme* is used. Based on Platen and Bruti-Liberati (2010)'s general definition of the *Euler-Maruyama scheme* with jumps, the descretization of the Merton model under $\mathbb{P}$ can be derived as

$$\ln S_t = \ln S_{t-1} + (\mu_t - \frac{1}{2}\sigma - c_0\bar{\mu})\Delta t + \sigma\Delta W_t + Z_t^s \Delta N_t, \tag{3.7}$$

where

$$\Delta N_t = N_t - N_{t-1} \sim Bernoulli(\lambda\Delta t) \tag{3.8}$$

and

$$\Delta W_t = W_t - W_{t-1} \sim \mathcal{N}(0,\, \Delta t). \tag{3.9}$$

The static parameters are defined precisely the same as in Definition 2.3.

Moreover, as another illustrative example, the descretization of the Heston model in Illustration 2.1 can be derived as follows:

$$\begin{aligned} \ln S_t &= \ln S_{t-1} + (\mu_t - \frac{1}{2}V_{t-1})\Delta t + \sqrt{V_{t-1}}\left(\rho\Delta W_{1,t} + \sqrt{1-\rho^2}\Delta W_{2,t}\right) \\ V_t &= V_{t-1} + \kappa(\theta - V_{t-1})\Delta t + \sigma\sqrt{V_{t-1}}\Delta W_{1,t}, \end{aligned} \tag{3.10}$$

where

$\{\mu_t, \rho, \kappa, \theta, \sigma\}$ denote the same terms as in the Heston model (As shown in Illustration 2.1, and especially for the definition of μ_t see Eq. (2.50).);

$\Delta W_{1,t}$ and $\Delta W_{2,t}$ are independent random variables, which follows the same definition in Eq. (3.9).

Following the concrete example from Fulop and Li (2019), the discretization of the primary objective model (see Illustration 2.4) can be done by using the *Euler-Maruyama*

[40] Let $T > 0$ and set up the partition $([t_{i-1}, t_i])_{i=1,\ldots,N}$ for $t_0 := 0, t_N := T$ and let $\delta := \max_{i=1,\ldots,N} |t_i - t_{i-1}|$ the maximum time increment. A discrete-time process $X_i \approx X(t_i)$ is said to converge

- *strongly* with order $\gamma > 0$ if there is a constant $C > 0$ independent of δ with

$$\max_{i=0,\ldots,N} \mathbb{E}\big[|X(t_i) - X_i|\big] \leq C\delta^\gamma.$$

- *weakly* with order $\gamma > 0$ w.r.t. a function $F : \mathbb{R} \to \mathbb{R}$ if there is a constant $C > 0$ independent of δ with

$$\max_{i=0,\ldots,N} |\mathbb{E}\big[F(X(t_i))\big] - \mathbb{E}\big[F(X(t_i))\big]| \leq C\delta^\gamma.$$

Strong convergence means pathwise convergence while weak convergence means the convergence of moments ($F(x) = x$) or probabilities.

scheme, as follows:

$$\begin{aligned}
\ln S_t &= \ln S_{t-1} + (\mu_t - \frac{1}{2}V_{t-1} - c_0^+\bar{\mu}^+ - U_{t-1}\bar{\mu}^-)\Delta t \\
&\quad + \sqrt{V_{t-1}}\left(\rho\Delta W_{1,t} + \sqrt{1-\rho^2}\Delta W_{2,t}\right) + Z_t^{s,+}\Delta N_t^+ + Z_t^{s,-}\Delta N_t^- \\
V_t &= V_{t-1} + \kappa_v(\theta_u - V_{t-1})\Delta t + \sigma_v\sqrt{V_{t-1}}\Delta W_{1,t} + \mu_v(Z_t^{s,-})^2\Delta N_t^- \\
U_t &= U_{t-1} + \kappa_u(\theta_u - U_{t-1})\Delta t + \sigma_u\sqrt{U_{t-1}}\Delta W_{3,t} + \mu_u(Z_t^{s,-})^2\Delta N_t^-,
\end{aligned} \tag{3.11}$$

where

$\left\{\mu_t, \bar{\mu}^+, \bar{\mu}^-, c_0^+, \rho, \sigma_v, \sigma_u, Z_t^{s,+}, Z_t^{s,-}\right\}$ are defined in the same as in the definition of the SVNUJ model (see Illustration 2.4, and especially for the definition of μ_t see Eq. (2.62));

$\Delta N_t^+ \sim Bernoulli(c_0^+\Delta t)$ and $\Delta N_t^- \sim Bernoulli(U_{t-1}\Delta t)$;

$\Delta W_{1,t}$, $\Delta W_{2,t}$ and $\Delta W_{3,t}$ are independent normally distributed random variables, which follows the same definition in Eq. (3.9).

3.2.3 Linear Gaussian Approximation

Due to the outstanding tractability and ease of implementation, the Kalman filter (see the discussion in Section 3.3.1) is extensively used in the task of estimating the modern affine term structure model, which is equipped with affine processes defined in Definition A.1. However, the pre-requirement of applying the Kalman filter lies in the fact that the state-space model must fulfill the linear Gaussian structure in both the measurement equation and state equation from Definition 3.1. Naturally, within the affine set-up, the term structure model implied bond yields to be linear w.r.t. latent states (see Eq. (2.26)). Thus, for estimating the term structure model, the state-space model has a linear Gaussian measurement equation if the measurement error is assumed to be Gaussian and addictive. While the actual limitation lies in the state equation, which is linear, but non-Gaussian in affine term structure models that accommodate stochstic volatility. To fit the pre-requirement of Gaussian transition density, Duffee and Stanton (2012) and Monfort et al. (2017) proposed an approximation of state equation by using Gaussian density with the same first two moments of the original conditional transition density. When approximation is applied, the parameter estimates obtained directly from maximum likelihood estimation based on the Kalman filter are not consistent. However, a Monte Carlo study shows that the inconsistency may be of limited significance in practice when the underlying state-space model is linear but with heteroskedastic volatility.

In general, a linear Gaussian state equation can be expressed as follows:

$$X_t = \Phi_0 + \Phi_1 X_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \Sigma_t), \tag{3.12}$$

where $\mathcal{N}$ denotes a multi-variant normal distribution.

Similarly to term structure models, the Heston model (see Illustration 2.1) with discretization under $\mathbb{P}$ is defined as in Eq. (3.10). The state equations are given as follows:

$$V_t = \kappa\theta\Delta t + (1 - \kappa\Delta t) V_{t-1} + \sigma\sqrt{V_{t-1}}\Delta W_t. \tag{3.13}$$

As a result of fitting the first two moments of the conditional density, one can obtain a linear Gaussian approximation that involves the following coefficients:

$$\begin{aligned} &\Phi_0^{sv} = \kappa\theta\Delta t, \quad \Phi_1^{sv} = (1 - \kappa\Delta t), \\ &\Sigma_t^{sv} = \sigma^2 V_{t-1}\Delta t. \end{aligned} \tag{3.14}$$

Furthermore, for Illustration 2.4 with discretization as stated in Eq. (3.11), we have

$$\begin{aligned} V_t &= \kappa_v\theta_v\Delta t + (1 - \kappa_t\Delta t) V_{t-1} + \sigma_v\sqrt{V_{t-1}}\Delta W_{1,t} + \mu_v(Z_t^{s,-})^2\Delta N_t^-, \\ U_t &= \kappa_u\theta_u\Delta t + (1 - \kappa\Delta t) U_{t-1} + \sigma_u\sqrt{U_{t-1}}\Delta W_{2,t} + \mu_u(Z_t^{s,-})^2\Delta N_t^-, \end{aligned} \tag{3.15}$$

Since $\Delta W_{1,t}, \Delta W_{2,t}, Z^{s,-}$ and ΔN_t^- are independent, by applying the same idea the linear Gaussian approximation can be obtained as follows:

$$\Phi_0^{nuj} = \begin{pmatrix} \kappa_v\theta_v\Delta t \\ \kappa_u\theta_u\Delta t \end{pmatrix}, \quad \Phi_1^{nuj} = \begin{pmatrix} 1 - \kappa_v\Delta t & \mu_v\bar{\lambda}^-\Delta t \\ 0 & 1 - \kappa_u\Delta t + \mu_u\bar{\lambda}^-\Delta t \end{pmatrix} \tag{3.16}$$

and

$$\Sigma_t^{nuj} = \begin{pmatrix} \sigma_v^2 V_{t-1}\Delta t + \mu_v^2\sigma_{\lambda^-}^2 U_{t-1}\Delta t & \mu_v\mu_u\sigma_{\lambda^-}^2 U_{t-1}\Delta t \\ \mu_v\mu_u\sigma_{\lambda^-}^2 U_{t-1}\Delta t & \sigma_u^2 U_{t-1}\Delta t + \mu_u^2\sigma_{\lambda^-}^2 U_{t-1}\Delta t \end{pmatrix}. \tag{3.17}$$

The $\bar{\lambda}^-$ and σ_{λ^-} are defined with the help of random variable $\tilde{Z}_t := (Z_t^{s,-})^2$ as follows,

$$\bar{\lambda}^- := \mathbb{E}[\tilde{Z}_t] = \frac{2}{(\lambda^-)^2}, \; \sigma_{\lambda^-} := \sqrt{\mathbb{E}[\tilde{Z}_t^2]} = \sqrt{\frac{24}{(\lambda^-)^4}}.^{41} \tag{3.18}$$

[41] In general, for compounded Poisson process $Y_t = \sum_{i=1}^{N_t} Z_i$, where N_t with $t \geq 0$ is a Poisson

Further, in later simulation studies and empirical analyses, a zero covariance term[42]is used to simplify the variance matrix:

$$\tilde{\Sigma}_t^{nuj} = \begin{pmatrix} \sigma_v^2 V_{t-1}\Delta t + \mu_v^2 \sigma_{\lambda^-}^2 U_{t-1}\Delta t & 0 \\ 0 & \sigma_u^2 U_{t-1}\Delta t + \mu_u^2 \sigma_{\lambda^-}^2 U_{t-1}\Delta t \end{pmatrix}. \tag{3.20}$$

3.3 Filtering Methods [43]

In this section, filtering algorithms for state-space models are discussed. In the literature, various filtering algorithms are applied in the empirical studies of dynamic asset pricing models, including the Kalman filter, the extended Kalman filter, the particle filter, and so forth. In order to perform the empirical analysis of cross-sectional option panels in this thesis, the Kalman filter will be discussed in detail for handling the inference of the corresponding AJD models. To round out the review, a brief introduction is provided to the particle filter, which is a widely used method to estimate models for equity derivatives markets.

process with rate λ and Z_i with $i \geq 1$ are independent and identically distributed random variables. Approximating the mean and variance of a compounded Poisson process can be described as follows:

$$\begin{aligned} \mathbb{E}[Y_t] &= \mathbb{E}[Z_1 + \cdots + Z_{N_t}] = \mathbb{E}[N_t]\,\mathbb{E}[Z_1] = \lambda t \mathbb{E}[Z_1] \\ \mathrm{Var}[Y_t] &= \mathbb{E}[\mathrm{Var}[Y_t|N_t]] + \mathrm{Var}[\mathbb{E}[Y_t|N_t]] \\ &= \mathbb{E}[N_t \mathrm{Var}[Z_1]] + \mathrm{Var}[N_t \mathbb{E}[Z_1]] \\ &= \mathbb{E}[N_t]\,\mathrm{Var}[Z_1] + \mathrm{Var}[N_t]\,(\mathbb{E}[Z_1])^2 \\ &= \lambda t \mathrm{Var}[Z_1] + \lambda t\,(\mathbb{E}[Z_1])^2 \\ &= \lambda t \mathbb{E}\left[Z_i^2\right]. \end{aligned} \tag{3.19}$$

[42]Since we use the Kalman filter as an approximation for our model, we need to choose how we approximate the variance in the state equation. One can treat a covariance matrix with zero covariance terms as another choice of approximation. In general, it is also a reasonable choice. In the variance matrix, only the jump part enters the covariance term and it is almost zero, when λ^- gets larger. Overall, the variance matrix is dominated by the variance coming from diffusion terms for many sets of practical parameters. As a result, the effect of covariance terms with only jump parameters is negligible.

[43]This part is mainly based on the results of a research project with a C-RAM student. I would like to thank Maximilian Kübler for his assistance.

3.3.1 Kalman Filter

Following Durbin and Koopman (2012), the linear Gaussian state-space model (SSM) takes the form

$$y_t = Z_t\alpha_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, H_t),$$
$$\alpha_{t+1} = T_t\alpha_t + R_t\eta_{t+1}, \;\; \eta_{t+1} \sim \mathcal{N}(0, Q_{t+1}).$$

At any time $t+1$, we want to determine which information should be used for updating. As we know from our SSM, the current state α_t directly reflects the distributional properties of α_{t+1}. Observations at time $t+1$ will also reveal further information about α_{t+1}. As a result of the linear Gaussian structure, any observation and any state are normally distributed, potentially in multiple dimensions. Therefore, it is justified to update both the expectation and covariance structure each time a new observation is introduced. Based on the observations, the states are bound by the following laws:

$$\alpha_t \sim \mathcal{N}\left(a_{t|t}, P_{t|t}\right) \text{ and } \alpha_{t+1} \sim \mathcal{N}(a_{t+1}, P_{t+1}), \tag{3.21}$$

where $a_{t|t}$ is used to highlight the additional information y_t (compared to a_t). Assuming information up to $t+1$, the one-step ahead prediction error of y_t helps us find a relationship between at and a_t and a_{t+1} at this particular time. The prediction error is as follows:

$$v_{t+1} = y_{t+1} - \mathbb{E}(y_{t+1}|Y_t) = y_{t+1} - \mathbb{E}(Z_{t+1}\alpha_{t+1} + \epsilon_{t+1}|Y_t) = y_{t+1} - Z_{t+1}a_{t+1},$$

since a_{t+1} is the conditional expected state at time $t+1$ and the error ϵ_{t+1} is independent of previous observations. Considering the normality of the observations, v_{t+1} is also normally distributed. In addition, a_{t+1} and $a_{t|t}$ have a linear relationship, as follows:

$$a_{t+1} = \mathbb{E}[T_t\alpha_t + R_t\eta_{t+1}|Y_t] = T_t\,\mathbb{E}[\alpha_t|Y_t] = T_t\,a_{t|t}.$$

As a result of an affine-linear operation, P_{t+1} can be derived from $P_{t|t}$, as follow:

$$\begin{aligned} P_{t+1} &= \mathrm{Var}[T_t\alpha_t + R_t\eta_{t+1}|Y_t] \\ &= T_t\,\mathrm{Var}[\alpha_t|Y_t]\;T_t^\top + R_t\,Q_{t+1}\,R_t^\top = T_t\,P_{t|t}\,T_t^\top + R_t\,Q_{t+1}\,R_t^\top. \end{aligned}$$

In order to establish a relation between a_{t+1} and a_t, we must find relationships between a_t and $a_{t|t}$, i.e., It is critical that we determine the additional information that y_t provides

in relation to the observations passed. By expressing $a_{t|t}$ in terms of a_t, we have

$$a_{t|t} = \mathbb{E}\left[\alpha_t | Y_t\right] = \mathbb{E}\left[\alpha_t | Y_{t-1}, v_t\right] = \mathbb{E}\left[\left(\alpha_t | v_t\right) | Y_{t-1}\right],$$

since knowledge of the actual observation y_t is equivalent to knowing the explicit value of the one-step ahead prediction error. Using the *regression lemma* (Lemma 1 in subsection 4.2 of Durbin and Koopman (2012)), we are able to reach $a_{t|t}$, and therefore we need to verify that α_t and v_t (given Y_{t-1}) are both normally distributed. It suffices to show that the projection $c^\top \alpha_t + d^\top v_t$ is normally distributed $\forall c \in \mathbb{R}^m$, $\forall d \in \mathbb{R}^p$:

$$c^\top \alpha_t + d^\top v_t = \left(c^\top + d^\top Z_t\right) \alpha_t - Z_t \alpha_t + \epsilon_t \overset{D}{\sim} \left(c^\top + d^\top Z_t\right) \mathcal{N}\left(a_t, P_t\right) - \mathcal{N}\left(Z_t a_t, H_t\right).$$

The two normal distributions are conditionally independent. Finally, using the regression lemma, we can derive

$$a_{t|t} = \mathbb{E}\left[\alpha_t | Y_{t-1}\right] + \operatorname{Cov}\left[\alpha_t, v_t | Y_{t-1}\right] \left(\operatorname{Var}\left[v_t | Y_{t-1}\right]\right)^{-1} v_t. \tag{3.22}$$

The first summand is a_t, according to the notation defined in (3.21). By noting that $\mathbb{E}\left[v_t | Y_{t-1}\right] = 0$, the covariance can be computed directly:

$$\operatorname{Cov}\left[\alpha_t, v_t | Y_{t-1}\right] = \mathbb{E}\left[\alpha_t \left(Z_t \left(\alpha_t - a_t\right) + \epsilon_t\right)^\top | Y_{t-1}\right] = \mathbb{E}\left[\alpha_t \left(\alpha_t - a_t\right) Z_t^\top | Y_{t-1}\right] = P_t \, Z_t^\top.$$

In the case of the conditional variance of v_t, we obtain a similar result:

$$F_t := \operatorname{Var}\left[v_t | Y_{t-1}\right] = \operatorname{Var}\left[Z_t \left(\alpha_t - a_t\right) + \epsilon_t | Y_{t-1}\right] = \operatorname{Var}\left[Z_t \alpha_t + \epsilon_t | Y_{t-1}\right] = Z_t \, P_t \, Z_t^\top + H_t.$$

All in all, we have

$$a_{t|t} = a_t + P_t \, Z_t^\top \, F_t^{-1} \, v_t.$$

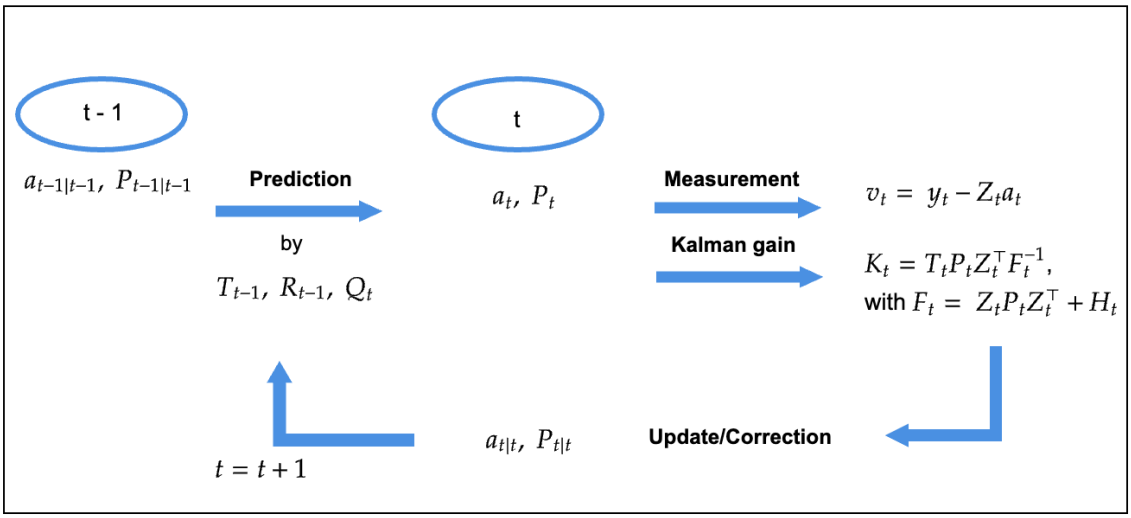
Similarly, we must do the same for the variance-covariance matrix $P_{t|t}$, whose structure is also derived from the regression lemma.

$$\begin{aligned} P_{t|t} &= \operatorname{Var}\left[\alpha_t | Y_{t-1}\right] - \operatorname{Cov}\left[\alpha_t, v_t | Y_{t-1}\right] \left(\operatorname{Var}\left[v_t | Y_{t-1}\right]\right)^{-1} \operatorname{Cov}\left[\alpha_t, v_t | Y_{t-1}\right]^\top \\ &= P_t - P_t \, Z_t^\top \, F_t^{-1} Z_t \, P_t^\top. \end{aligned} \tag{3.23}$$

Based on knowledge of prior observations, the Kalman filter is based on the error we make

in predicting y_t, and takes this error into account for the next prediction. Outliers that deviate from their respective local mean values are therefore potential pitfalls. Fig. 3.2 illustrates the sequential prediction-correction process.

Figure 3.2: Schematic structure of the classical Kalman filter for linear systems.



The matrix F_t is of dimension $p \times p$ and typically observations are low-dimensional (if not real-valued), so the inversion of F_t is cheap in terms of computational effort. Due to the fact that the matrix F_t has a dimension of $p \times p$ and most observations are low-dimensional, the inversion of F_t is relatively simple from a computational perspective. Overall, it would suffice to update the distribution of states at time $t+1$. The Kalman filtering step from t to $t+1$ requires only two initializations

$$a_{t+1} = T_t\, a_{t|t} \text{ and } P_{t+1} = T_t\, P_{t|t}\, T_t^\top + R_t\, Q_{t+1}\, R_t^\top. \tag{3.24}$$

In many cases, the recursion (3.24) is reformulated by using the *Kalman gain* $K_t = T_t P_t\, Z_t^\top\, F_t^{-1}$(see Eq (4.22) of Durbin and Koopman (2012)). We end up with

$$a_{t+1} = T_t\, a_t + K_t\, v_t \text{ and } P_{t+1} = T_t\, P_t\, (T_t - K_t Z_t)^\top + R_t\, Q_{t+1}\, R_t^\top.$$

Furthermore, the log-likelihood of observations $y_0, \ldots, y_T$ is defined as follows:

$$\ell(Y_{0:T}) = \log p(y_0) + \sum_1^\top \log p(y_t|Y_{t-1}). \tag{3.25}$$

Based on the results of the linear Gaussian system in this section, we have $\mathbb{E}(y_t|Y_{t-1}) = Z_t a_t$. Putting $v_t = y_t - Z_t a_t, F_t = \mathrm{Var}(y_t|Y_{t-1})$ and substituting $\mathcal{N}(Z_t a_t, F_t)$ for $p(y_t|Y_{t-1})$ in Eq. (3.25), we obtain

$$\ell(Y_{0:T}) = -\frac{(T+1)p}{2}\log 2\pi - \frac{1}{2}\sum_0^T \left(\log|F_t| + v_t^\top F_t^{-1} v_t\right).^{44} \tag{3.26}$$

[44]For initialization at $t = 0$, $F_0 = \mathrm{Var}(Y_0)$. In addition, it is assumed that a_0 and P_0 are given for $\alpha_0 \sim \mathcal{N}(a_0, P_0)$.

Algorithm 1: Kalman filter for state inference

Input : Observations $y_0, \ldots, y_T$, initial distribution $\mu_0(x_0)$.
Parameters: Coefficients of linear Gaussian SSM $Z_{0:T}, T_{0:T}, R_{0:T}, H_{0:T}$ and $Q_{1:T}$.
Returns : Vector $\{a_{t|t}\}_{t=0,\ldots,T}$ as best guess for the latent states at time $t = 0, \ldots, T$.

while $t \leq T$ **do**

  if $t = 0$ **then**

    Initialization: Set $a_0 \leftarrow \mathbb{E}[\mu_0], P_0 \leftarrow \mathrm{Var}[\mu_0]$.

  end

  Compute the a-priori measurement error: $v_t \leftarrow y_t - Z_t a_t$.

  Compute the Kalman gain at time t: $K_t \leftarrow T_t P_t Z_t^\top F_t^{-1}$, where $F_t := \mathrm{Var}[v_t | Y_{t-1}] = Z_t P_t Z_t^\top + H_t$.

  Update state distribution:

$$\begin{aligned} a_{t|t} &\leftarrow a_t + T_t^{-1} K_t v_t \\ P_{t|t} &\leftarrow P_t - T_t^{-1} K_t Z_t P_t^\top. \end{aligned}$$

  Prediction for next time:

$$\begin{aligned} a_{t+1} &\leftarrow T_t a_{t|t} \\ P_{t+1} &\leftarrow T_t P_{t|t} T_t^\top + R_t Q_{t+1} R_t^\top. \end{aligned}$$

  $t \leftarrow t + 1$

end

3.3.2 Particle Filter

As discussed in the previous section, under certain assumptions and given structural parameters, the latent states embedded in the AJD model can be extracted via a standard filtering algorithm within a SSM framework. The AJD model in continuous-time can be transformed into an SSM using *Euler-Maruyama scheme*. If the SSM from the transformation is linear and with Gaussian noise, one can use the Kalman filter (see Kalman (1960)) to infer the trajectories of the latent states in the AJDs. However, literature shows that if we want to identify and characterize relatively complicated patterns in the financial market, one needs models with stochastic diffusion and jump components. Therefore, in many cases the SSM, which is transformed from the AJD model, is no longer a linear Gaussian SSM. As a result of the nonlinear and non-Gaussian nature of such a system, it is necessary to use a more general filtering algorithm known as particle filtering (see Doucet and Johansen (2009)) in order to infer the trajectories of the latent states. As a type of filtering algorithm, the particle filter casts latent state variables as "particles" into a probability space and re-sample them according to predetermined weights. When the number of particles used approaches infinity, the algorithm provides exact filtering.[45]

In the empirical dynamic asset pricing literature, particle filtering is a widely used method when dealing with nonlinear and non-Gaussian state-space models. In particular, for equity derivatives, there are several existing empirical studies employing particle filtering for the inference task. Christoffersen et al. (2010) apply the particle filter to infer latent states for the Heston model for asset returns. Johannes et al. (2009) use particle filtering to extract latent states from asset returns assuming the static parameters in their AJD model are known. For calibrating the static parameters as well as recovering latent states, Fulop and Li (2019) and Bardgett et al. (2019) use a particle filter with derivatives based on equity indexes along with asset returns.

In general, the particle filter is a more powerful tool to use for tackling inference tasks. However, generality always comes at a price. Since it is a sequential simulation method, the particle filter is typically computationally intensive and has a slow convergence rate, when model dynamics become complex. In particular, when nonlinear relationships appear in the measurement equations, normally only the bootstrap version of the particle filter can be directly used. This is typically not very effective when observations are not

[45]More specifically, particle filtering is an effective Sequential Monte Carlo (SMC) method of estimating state.

informative about the corresponding states. For the subsequent empirical analysis of the cross-sectional option panels, one can use the raw option prices as nonlinear measurements for the corresponding AJD models with stochastic negative jump intensities. This approach, however, leads to an ineffective particle filter. On the other hand, to perform the subsequent empirical analysis on the cross-sectional option panels, they can be transformed into higher order cumulants and then a set of measurements is constructed that contains a linear relationship to the latent states. The particle filter is not an appropriate choice for the inference task, especially when linear measurement equations can be used. The empirical analysis of this thesis does not employ particle filtering, so the details of the algorithm are not discussed here. For readers who are interested in the details, please refer to Appendix A.5.

3.4 Penalized Nonlinear Least Squares

This section explains and presents an inspiring estimation approach, which is considered as the basis of building the Kalman-filtered based quasi-maximum likelihood approach employed in the empirical analysis in this thesis. Andersen et al. (2015a) propose a novel approach to estimate static parameters under the risk-neutral measure and infer the latent states, based on penalized nonlinear least squares (PNLS) to fit the rich data from the option market, as well as an additional penalty for the deviation of estimated volatility from its model-free counterpart. In their approach, BS implied volatility (BSIV) is used instead of the option price. A discussion of the advantages of BSIV over an option contract can be found in Definition A.6. Following the basic idea of the least squares approach, one needs to solve an optimization problem - minimizing the squared errors between the model implied BSIV and the observed BSIV - with respect to the static model parameters under risk-neutral measure and the time-varying latent states L_t. For instance, consider the model dynamic from Illustration 2.4, then we have latent states: $L_t := \{V_t, U_t\}$. In contrast to estimation methods using state-space models, their approach avoids making parametric assumptions regarding the transition density of underlying asset returns. The inference of latent states turns out to be another optimization problem by fitting option surfaces within a fixed time interval given the static model parameters. Therefore, the estimation approach is essentially an optimization task that involves a series of many minor low-dimensional optimizations that can be solved independently. In the rest of this section, this method and its connection to other state-space model-based methods

are discussed in more detail.

The time horizon of the total observations is assumed to be T with a daily frequency of observations. The observations are indexed from 0 to T. Furthermore, N_t denotes the total number of option contracts at each time point. Each individual option at time t is characterised by log-forward moneyness in Eq. (A.8). Since the BSIV is used to represent the option price alternatively, the task of fitting the observed option prices becomes a matter of fitting the model-implied BSIV $\mathrm{IV}^{\mathrm{model}}\left(L_t, \theta; \tau, m^k\right)$ to the corresponding observed BSIV $IV^{\mathrm{data}}_{t,\tau,m}$. In general, in the no-arbitrage dynamic asset pricing models the diffusion "coefficients" stay invariant for $\mathbb{P} \to \mathbb{Q}$, meaning that the realized spot variance should be the same under $\mathbb{P}$ and $\mathbb{Q}$. Thus, loosely speaking an estimate of latent states, which generates larger deviation between model-implied spot variance, is "worse" than the one generating smaller deviation. In the optimization problem, this intuition results in the penalty term. As mentioned at the beginning of this section, the PNLS approach only uses information from the option market, that is, only information from the risk-neutral density.

Following equation (6) of Andersen et al. (2020), the final optimization problem is defined as follows[46]:

$$\begin{aligned}
&\min_{L_{0:T},\theta} \sum_{t=0}^{T} \Big\{ \text{BSIV Fit}_t + \text{Volatility Fit}_t \Big\}, \\
&\text{BSIV Fit}_t = \frac{1}{N_t} \sum_{j=1}^{N_t} \left(\mathrm{IV}^{\mathrm{model}}\left(L_t, \theta; \tau_j, m_j^k\right) - \mathrm{IV}^{\mathrm{data}}_{t,\tau_j,m_j} \right)^2, \\
&\text{Volatility Fit}_t = \frac{\xi_n}{N_t} \frac{\left(\sqrt{\widehat{V}_t^n} - \sqrt{V\left(L_t, \theta\right)} \right)^2}{\widehat{V}_t^n / 2},
\end{aligned} \tag{3.27}$$

where $\widehat{V}_t^n$ [47] denotes a nonparametric estimator of the spot variance at time t, which can be obtained by using high-frequency observations to deduce a consistent estimator. A detailed approach is explained in section A.3 of Andersen et al. (2019). Eq. (3.27) indicates that the final joint estimates of time-varying latent states and static model

[45] The time horizon notation follows the same definition as that in Section 3.2.1.

[46] For notational simplicity, the $\mathbb{Q}$ is omitted.

[47] n denotes the number grids for the intraday time span (see Andersen et al. (2019)).

parameters are given by

$$\{L^{\star}_{0:T}, \theta^{\star}\} = \underset{L_{0:T},\theta}{\arg\min} \sum_{t=0}^{T} \Big\{ \text{BSIV Fit}_t + \text{Volatility Fit}_t \Big\}, \tag{3.28}$$

where the BSIV Fit and Volatility Fit are defined in Eq. (3.27). In the BSIV Fit part, all the option's observations within one time point are treated equally, which may lead to overfitting in the final result. There is a possibility that the estimation result will shift volatility to an unrealistic level. Therefore, as argued by Andersen et al. (2015a) using the volatility part of the optimization problem as a penalty term can help rule out potential ill-behavior of estimators. Furthermore, the ξ_n is fixed at 0.05 in the applications of Andersen et al. (2020).

Andersen et al. (2014) highlights another point with regarding to solving the ultimate optimization problem efficiently, namely without restriction of the transition density under $\mathbb{P}$, the recovery of latent stats for each time point are independent to each other. Given parameters of model under risk-neutral measure $\mathbb{Q}$, the "filtering" of latent states can be defined as following function for each time point t as follows:

$$f_t\left(L_t;\theta\right) := \frac{1}{N_t}\sum_{j=1}^{N_t}\left(\text{IV}^{\text{model}}\left(L_t,\theta;\tau_j,m_j^k\right) - \text{IV}^{\text{data}}_{t,\tau_j,m_j}\right)^2 + \frac{\xi_n}{N_t}\frac{\left(\sqrt{\widehat{V}_t^n} - \sqrt{V\left(L_t,\theta\right)}\right)^2}{\widehat{V}_t^n/2}. \tag{3.29}$$

Using the exchangeability of the minimum (see Proposition A.6) the final two-level optimization problem can be formulated as follows:

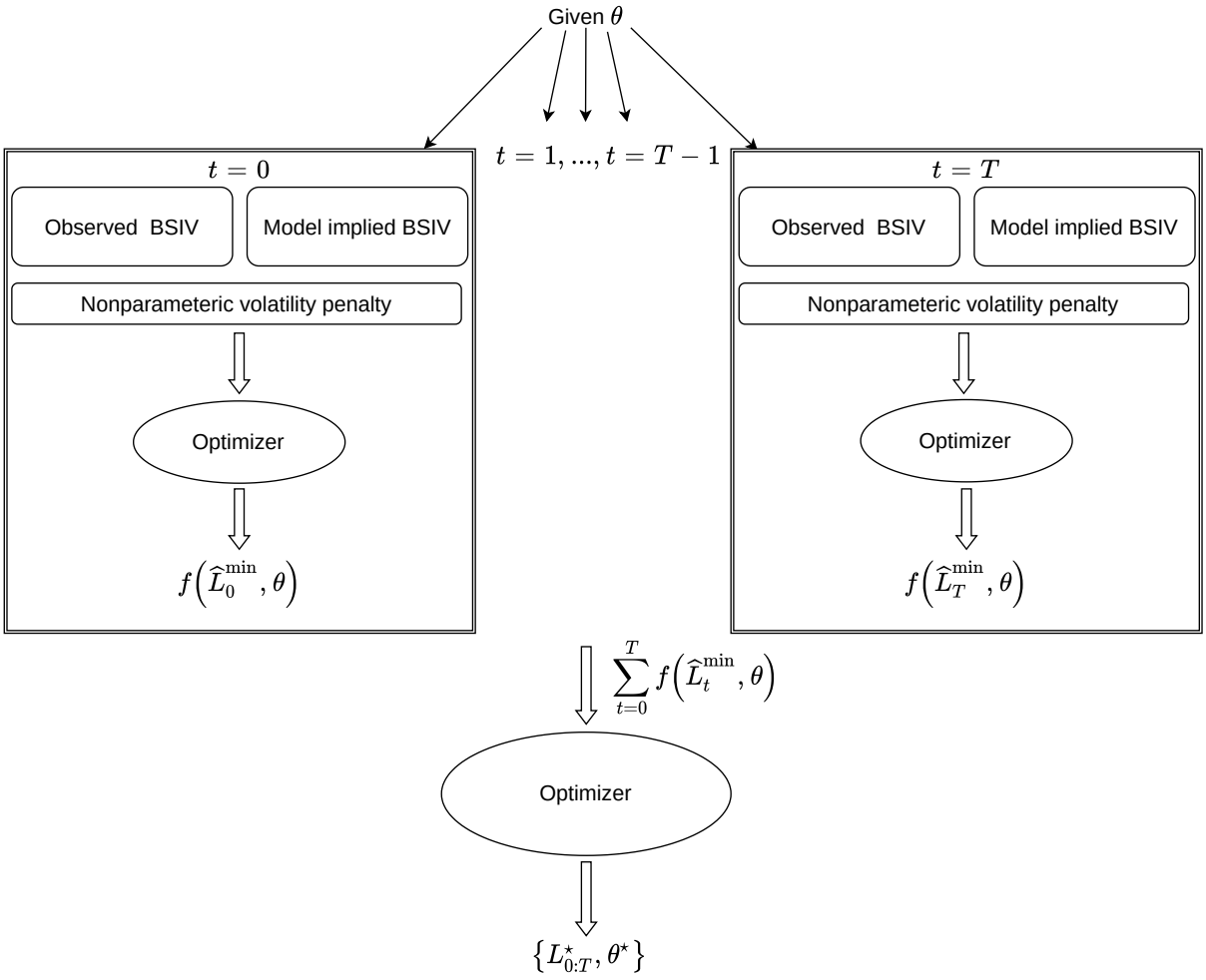
$$\begin{aligned}
\min_{L_{0:T},\theta}\sum_{t=0}^{T} f_t\left(L_t,\theta\right) &= \min_{\theta}\min_{L_{0:T}}\sum_{t=0}^{T} f_t\left(L_t,\theta\right) \\
&= \min_{\theta}\left[\min_{L_0} f_0\left(L_0,\theta\right) + \cdots + \min_{L_T} f_T\left(L_T,\theta\right)\right] \\
&= \min_{\theta}\sum_{t=0}^{T}\hat{f}(\theta)
\end{aligned} \tag{3.30}$$

where

$$\hat{f}(\theta) = f(\hat{L}_t^{\min},\theta) \text{ with } \hat{L}_t^{\min} = \underset{L_t}{\arg\min}\, f_t\left(L_t,\theta\right),\ \forall t \in \{0,\ldots,T\}.$$

Hence, jointly estimating the parameters and the latent trajectory is equivalent to an optimization task of a function whose value itself is determined by optimization. Fig. 3.3

Figure 3.3: Visualization of PNLS.



illustrates the process of solving the two-level optimization problem. The main benefit of collapsed minimization is that the nested inner optimization tasks can be tackled simultaneously. In particular, the original high-dimensional minimization problem is translated into multiple low-dimensional problems. Furthermore, as discussed in Andersen et al. (2015c), the inner minimization problems can be implemented with standard parallelization techniques, e.g., Open MPI (Messages Passing Interface) from an implementation point of view. In general, parallelization can reduce the time required for estimation significantly when a High Performance Computing System is available.

As shown in the Monte Carlo study by Andersen et al. (2015a) and empirical studies by Andersen et al. (2015a) and Andersen et al. (2020), the PNLS estimation procedure can estimate the parameters and recover the trajectories of the latent states accurately, which means that the rich information about the risk-neutral density embedded in the option panels makes the inference of AJD models possible. It appears that this finding generally

supports the soundness of the inference of AJD models under risk-neutral measure and the recovery of latent state trajectories when option panels are considered.

3.5 Likelihood-Based Methods

3.5.1 Particle Metropolis-Hastings [48]

We assume that the structural parameters in AJD models are known in the filtering step discussed above. In many empirical analyses, learning static parameters is also a necessary step to study the relationship between asset price, risk and the market price of risk. Thanks to the filtering step, we can obtain an approximation of the log-likelihood (LLK) function given static parameters and observed measurements within the SSM framework. For estimation of the static parameters, of course, one can directly put the approximated LLK function into some mathematical optimization algorithm (e.g., Bardgett et al. (2019)). On the other hand, we can also follow the idea of Bayesian inference, which incorporates the Metropolis-Hastings algorithm to achieve the estimation task. Due to the fact that this type of Bayesian inference involves a particle filtering step, it is called particle Metropolis-Hastings (PMH) in the literature. As the particle filter is not used in the subsequent empirical analysis. For readers who are interested in the details, please refer to Andrieu et al. (2010) and Dahlin and Schön (2015) for more details.

3.5.2 Quasi-Maximum Likelihood

According to Durbin and Koopman (2012), a linear Gaussian state-space model can be estimated using the Kalman filter with maximum likelihood. The transition density in the state-space model based on the discretized version of the models in the AJD family is not necessarily Gaussian. Thus, the linear Gaussian approximation in Section 3.2.3, i.e., substituting the theoretical transition density with a Gaussian one, is able to match the conditional mean and variance. If there are observations for linear Gaussian measurement equations available, one can obtain an approximation of the likelihood function based on the approximation in the state equations.

Inspired by the PNLS method in Section 3.4, we can add filtering independent measure-

[48] This part is mainly based on the results of research projects with aC-RAM students. I would like to thank Maximilian Kübler for his assistance.

ments to the likelihood evaluation. The filtering independent measurements are denoted as exogenous measurements in the following context. In many cases, due to the nonlinear and non-Gaussian nature of those measurement equations, they only meet the requirements of applying the particle filter, which is a simulation-based filtering technique with a high computation cost. In the empirical studies of AJD models on option markets, the BS implied volatility is a nonlinear mapping of latent states.

Furthermore, the PNLS method is a two-level minimization approach. The inner minimization stage can be viewed as a model-free [49] filtering stage, which is similar to obtaining estimates of latent states based on the posterior distribution in standard filtering methods. When measurement errors are properly assumed, observations of exogenous measurements can contribute to the likelihood function using the estimates of latent states obtained from the filtering step. Furthermore the second stage of the inference procedure shares the same concept as the outer minimization stage in PNLS. Hence, the quasi-maximum likelihood approach with $w = 1$ is equivalent to the PNLS method, since it replaces the minimization problem with Kalman filter-based minimum variance [50] estimations.

As discussed above, we can have two types of measurement equations in the state-space model. Thus, the likelihood function used in the MLE-based estimation approach is twofold. In particular, the first component is the likelihood delivered by the filtering approach, and the second component is based on observations, which are independent of the filtering step. Typically the errors of observations of exogenous measurements are assumed to be normal random variables uncorrelated with errors in the filtering measurement equations. Christoffersen et al. (2010), Ornthanalai (2014), Bégin et al. (2020), Gruber et al. (2021) and others, employ a similar log-likelihood function for the quasi-maximum likelihood approach. The general form of the log-likelihood function is

[49] The model-free refers to the fact that no transition density under $\mathbb{P}$ is assumed.
[50] The minimum variance idea in the Kalman filter refer to the Eq. (A.27)

defined as follows:

$$
\begin{aligned}
\ell(\Theta^{\text{est}}; w) &= (N^f_{\text{total}} + N^e_{\text{total}}) \left((1-w) \frac{\ell_f(\Theta^{\text{est}})}{N^f_{\text{total}}} + w \frac{\ell_e(\Theta^{\text{est}}; \hat{X}_{0:T})}{N^e_{\text{total}}} \right), 0 \leq w < 1 \\
\ell_f &:= \ell(\Theta^{\text{est}}, \{Y^f\}_{0:T}) \text{ in Eq. (3.26)}, \\
\ell_e &:= -\frac{1}{2} \sum_{t=0}^{T} \sum_{i=1}^{N^e_t} \left(\log\left(2\pi(\sigma^{e,rel}_{t,i})^2\right) + \frac{(\varepsilon^e_{t,i})^2}{(\sigma^{e,rel}_{t,i})^2} \right), \\
\varepsilon^e_{t,i} &:= \frac{Y^{e,\text{model}}_{t,i} - Y^{e,\text{data}}_{t,i}}{Y^{e,\text{data}}_{t,i}}, \quad i = 1, \ldots, N^e_t,
\end{aligned}
\tag{3.31}
$$

where

$X_{0:T}$ is the time series of latent states and $\hat{X}_{0:T}$ is its corresponding posterior estimates in Eq. (3.22) from Kalman filter;

N^f_{total} is the total number of observations used in filtering step across the whole time-period;

N^e_{total} is the number of observations of exogenous measurements at each sampling time point, and $N^e_{\text{total}} := \sum_{t=0}^{T} N^e_t$ is the total number of observations of exogenous measurements across the whole time-period;

$Y^{e,\text{model}}_{t,i}$ is the model implied exogenous measurements. It uses the estimates of the latent states $\hat{X}_{0:T}$, which are obtained as side results from evaluation of ℓ_e;

$\sigma^{e,rel}_{t,i}$ is the standard deviation of the measurement error for the observations of exogenous measurements.

The weight w is specified differently in the literature. Christoffersen et al. (2010) have two different specifications with $w = 0$ and $w = 1$. Similarly, Bégin et al. (2020) specifies w = 0.5. Both of them use only asset returns as measurements in the filtering step. Feunou and Okou (2018) set $w = 0$, which means that only the likelihood from the filtering step is considered in the final evaluation. As opposed to the two previous examples, they use cumulants in the filtering step.

[50] In Christoffersen et al. (2010), the maximum likelihood estimation with return in the filtering step is specified with $w = 0$. The specification with $w = 1$ corresponds to their nonlinear least squares estimation, which leaves out σ^2_e.

4 Monte Carlo Study

For the purpose of assessing fairly the selected methods and providing guidance for subsequent empirical applications, this chapter examines the performance of previously proposed methods using synthetic data generated from Monte Carlo simulations. The title of this thesis indicates that estimation approaches should be evaluated in conjunction with AJD model dynamics. In line with the discussion of the pros and cons of various estimation approaches in the previous chapter, I focus on the most promising methodology in this simulation study, that is, the Kalman filter-based quasi-maximum likelihood approach which is also applied in subsequent empirical applications. To evaluate the implemented methods concretely, one must have a representative example from the AJD models. Ideally, the model should be able to capture the statistical characteristics of asset returns, which are widely accepted in the literature on dynamic asset pricing models. Therefore, the SVNUJ model is selected, which nests many popular AJD models in the literature, for the simulation study in this chapter. In order to apply the Kalman filter-based quasi-maximum likelihood approach, it requires a linear Gaussian relationship in the state equations and measurement equations. Therefore, I first examine how the filtering step with approximated Gaussian state equations behaves if informative observations, such as risk-neutral cumulants of the return dynamic, are available. Following this, the performance of the QML approach to estimating static parameters is evaluated. This chapter is structured as follows: Section 4.1 introduces the state-space representation and the parameters considered in the Monte Carlo study. Section 4.2 explains how the various synthetic data used in the simulation study are generated. Section 4.3 introduces the concrete Kalman filter-based QML approach based on observations from the option markets. Section 4.4 presents the results of estimation using synthetic data. In addition, the performance of filtering methods is discussed for the chosen model, as well as the performance of static parameter estimation.

4.1 The Model

4.1.1 Model Dynamic

This simulation study makes use of the dynamics of the SVNUJ model, which is given below:

under $\mathbb{P}$:

$$\begin{aligned}\frac{dS_t}{S_t} &= \mu_t dt + \sqrt{V_t} dW_{1,t} + \left(e^{Z_t^+} - 1\right) dN_t^+ - c^+(t)\bar{\mu}^+ dt \\ &\quad + \left(e^{Z_t^-} - 1\right) dN_t^- - c^-(t)\bar{\mu}^- dt \\ dV_t &= \kappa_v \left(\theta_v - V_t\right) dt + \sigma_v \sqrt{V_t} dB_{1,t} + \mu_v \left(Z_t^-\right)^2 dN_t^- \\ dU_t &= \kappa_u(\theta_u - U_t)dt + \sigma_u \sqrt{U_t} dW_{2,t} + \mu_u \left(Z_t^-\right)^2 dN_t^-,\end{aligned} \tag{4.1}$$

under $\mathbb{Q}$:

$$\begin{aligned}\frac{dS_t}{S_t} &= \left(r_t - q_t\right) dt + \sqrt{V_t} dW_{1,t}^{\mathbb{Q}} + \left(e^{Z_t^{+,\mathbb{Q}}} - 1\right) dN_t^{+,\mathbb{Q}} - c^+(t)\,\bar{\mu}^{+,\mathbb{Q}} dt \\ &\quad + \left(e^{Z_t^{-,\mathbb{Q}}} - 1\right) dN_t^{-,\mathbb{Q}} - c^-(t)\,\bar{\mu}^{-,\mathbb{Q}} dt \\ dV_t &= \kappa_v^{\mathbb{Q}} \left(\theta_v^{\mathbb{Q}} - V_t\right) dt + \sigma_v \sqrt{V_t} dB_{1,t}^{\mathbb{Q}} + \mu_v \left(Z_t^{-,\mathbb{Q}}\right)^2 dN_t^{-,\mathbb{Q}} \\ dU_t &= \kappa_u^{\mathbb{Q}}(\theta_u^{\mathbb{Q}} - U_t)dt + \sigma_u \sqrt{U_t} dW_{2,t}^{\mathbb{Q}} + \mu_u \left(Z_t^{-,\mathbb{Q}}\right)^2 dN_t^{-,\mathbb{Q}}.\end{aligned} \tag{4.2}$$

For simplicity, only the dynamics of the model are presented here. For a detailed explanation of the parameters, see Illustration 2.4.

4.1.2 State-Space Model Representation

The Monte Carlo simulation study carried out in this chapter focuses on the SVNUJ model, which nests many other popular models in the literature. It is necessary to translate continuous-time model dynamics into a state-space representation in order to apply standard filtering techniques. Following the discussion in Section 3.2, the state equation can be obtained from Eq. (3.11) under $\mathbb{P}$, as follows:

$$\begin{aligned}V_t &= \kappa_v \theta_v \Delta t + \left(1 - \kappa_t \Delta t\right) V_{t-1} + \sigma_v \sqrt{V_{t-1}} \Delta W_{1,t} + \mu_v (Z_t^{s,-})^2 \Delta N_t^-, \\ U_t &= \kappa_u \theta_u \Delta t + \left(1 - \kappa \Delta t\right) U_{t-1} + \sigma_u \sqrt{U_{t-1}} \Delta W_{2,t} + \mu_u (Z_t^{s,-})^2 \Delta N_t^-,\end{aligned} \tag{4.3}$$

As discussed in the last chapter, this type of state equation is not linear Gaussian. As a result, a linear Gaussian approximation is required in order to apply the Kalman filter to the above state equations. Based on the results from Section 3.2, we can have the following linear Gaussian representation:

$$X_t = \Phi_0^{nuj} + \Phi_1^{nuj} X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0, \Sigma_t^{nuj}\right), \tag{4.4}$$

where

$$\Phi_0^{nuj} = \begin{pmatrix} \kappa_v \theta_v \Delta t \\ \kappa_u \theta_u \Delta t \end{pmatrix}, \quad \Phi_1^{nuj} = \begin{pmatrix} 1 - \kappa_v \Delta t & \mu_v \bar{\lambda}^- \Delta t \\ 0 & 1 - \kappa_u \Delta t + \mu_u \bar{\lambda}^- \Delta t \end{pmatrix} \tag{4.5}$$

and

$$\Sigma_t^{nuj} = \begin{pmatrix} \sigma_v^2 V_{t-1} \Delta t + \mu_v^2 \sigma_{\lambda^-}^2 U_{t-1} \Delta t & \mu_v \mu_u \sigma_{\lambda^-}^2 U_{t-1} \Delta t \\ \mu_v \mu_u \sigma_{\lambda^-}^2 U_{t-1} \Delta t & \sigma_u^2 U_{t-1} \Delta t + \mu_u^2 \sigma_{\lambda^-}^2 U_{t-1} \Delta t \end{pmatrix}. \tag{4.6}$$

The $\bar{\lambda}^-$ and σ_{λ^-} are defined with the help of random variable $\tilde{Z}_t := (Z_t^{s,-})^2$ as follows,

$$\bar{\lambda}^- := \mathbb{E}[\tilde{Z}_t] = \frac{2}{(\lambda^-)^2},\ \sigma_{\lambda^-} := \sqrt{\mathbb{E}[\tilde{Z}_t^2]} = \sqrt{\frac{24}{(\lambda^-)^4}}.^{51} \tag{4.7}$$

Further, a zero covariance term[52]is used to simplify the variance matrix that is employed in the implementation, as follows:

$$\tilde{\Sigma}_t^{nuj} = \begin{pmatrix} \sigma_v^2 V_{t-1} \Delta t + \mu_v^2 \sigma_{\lambda^-}^2 U_{t-1} \Delta t & 0 \\ 0 & \sigma_u^2 U_{t-1} \Delta t + \mu_u^2 \sigma_{\lambda^-}^2 U_{t-1} \Delta t \end{pmatrix}. \tag{4.8}$$

Feunou and Okou (2018) and Fulop and Li (2019) demonstrate that, for the measurement equations, the moment information of the return dynamics facilitates estimation, especially during the filtering step. The aggregate information collected from the option market, such as variance swaps or model-free cumulants of the return density, can provide valuable insight into the risk-neutral density. However, variance swaps are not

[51]See the derivation in Eq. (3.19).

[52]Since we use the Kalman filter as an approximation for our model, we need to choose how we approximate the variance in the state equation. One can treat a covariance matrix with zero covariance terms as another choice of approximation. In general, it is also a reasonable choice. In the variance matrix, only the jump part enters the covariance term and it is almost zero, when λ^- gets larger. Overall, the variance matrix is dominated by the variance coming from diffusion terms for many sets of practical parameters. As a result, the effect of covariance terms with only jump parameters is negligible. In addition, this further simplification of the variance matrix does not have any impact on the generation of data in simulations.

easily accessible due to the fact that they are OTC contracts. Usually one can only use the cumulants $\mathcal{K}_{n,R_s(t,\tau)}$, which are derived from option surfaces in a model-free way, as informative observations for the measurement equations. Another benefit of employing model-free cumulants as measurement equations is their linear relationship to the latent states. As a result of this feature, the Kalman filter can be applied with the previously introduced linear Gaussian approximation for AJD models.

Cumulants, which form the basis of the likelihood function and the filtering step, are not directly tradeable on the market. It is thus necessary to compute the moments $M_{n,R_s(t,\tau)}$, which can be evaluated using Eq. (2.65) as an intermediate step. Then, model-free estimates of the cumulants can be derived from the moments. For notational simplification, I further use $\mathcal{K}_{t,\tau,n}$ ($M_{t,\tau,n}$) for the n-th cumulant (moment) of $R_s(t,\tau)$ at time t. On the other hand, the model-free estimates of the cumulants can only be obtained with noise, and the bias is caused by two factors: First, the option contracts are observed with noise, and second, model-free estimates of the moments are obtained with errors due to the limited and discrete sampled observations of option surfaces. On the other hand, it should be noted that model-free estimates of cumulants for higher orders can be inaccurate, which may result in less informative observations. Consequently, I limit the cumulants to third order in this simulation study.

As an amelioration of the input information, I use option contracts as the second part of measurement equations that only enter the likelihood evaluation in order to correct any potential bias from model-free estimates of the cumulants. As discussed in Definition A.6, rather than using price values, the BS implied volatility of option contracts is used. Finally, I define the state-space representation for MLE-based estimation with the Kalman filter, as follows:

Definition 4.1. SSM representation of SNVUJ for Kalman filter

- **Measurement equation**:

$$\begin{aligned} \mathcal{K}^{\text{data}}_{t,\tau,n} &= \mathcal{K}^{\text{model}}_{t,\tau,n} + \mathcal{K}^{\text{data}}_{t,\tau,n}\varepsilon^{\mathcal{K}}_{t,\tau,n}, \quad \tau > 0,\ n \in \{2,3\}, \\ \text{IV}^{\text{data}}_{t,i} &= \text{IV}^{\text{model}}_{t,i} + \text{IV}^{\text{data}}_{t,i}\varepsilon^{\text{IV}}_{t,i}, \quad i = 1,\dots,N_{\text{IV},t}, \end{aligned} \tag{4.9}$$

  where τ is the maturity and n is the order of cumulants and $N_{\text{IV},t}$ is the number of contracts available in the input data. Moreover, the error terms in both

measurement equations are assumed to be independent normally distributed:

$$\varepsilon^{\mathcal{K}}_{t,\tau,n} = \frac{\mathcal{K}^{\text{data}}_{t,\tau,n} - \mathcal{K}^{\text{model}}_{t,\tau,n}}{\mathcal{K}^{\text{data}}_{t,\tau,n}} \sim \mathcal{N}(0, (\sigma^{\mathcal{K},\text{rel}}_{t,\tau,n})^2), \tag{4.10}$$

$$\varepsilon^{\text{IV}}_{t,i} = \frac{\text{IV}^{\text{data}}_{t,i} - \text{IV}^{\text{model}}_{t,i}}{\text{IV}^{\text{data}}_{t,i}} \sim \mathcal{N}(0, (\sigma^{\text{IV},\text{rel}}_{t,i})^2), \tag{4.11}$$

where variance of errors in the measurement equation of cumulants is assumed to be maturity and order dependent and the variance of errors in option contracts is contract-wise dependent. Both terms are reserved for heteroscedasticity in general. It is also possible to set both variances to be homogeneous within the same notation at a later stage.

- **State equation**:

$$X_t = \Phi_0^{nuj} + \Phi_1^{nuj} X_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}\left(0, \Sigma_t^{nuj}\right), \tag{4.12}$$

where $X_t = (V_t, U_t)^\top$ and all the coefficients and variance of shocks in state equation are the same as in Eq. (4.5) and Eq. (4.8), respectively.

4.1.3 Parameters

Based on the dynamics of the SVNUJ model (see Illustration 2.4), Θ_{nuj} denotes the vector of static parameters for the data generating processes. The Θ_{nuj} is summarized as follows:

$$\Theta_{nuj} = \{\Theta^{\mathbb{P}}_{nuj}, \Theta^{\mathbb{Q}}_{nuj}, \Theta^{\mathbb{P},\mathbb{Q}}_{nuj}, \Theta^{rp}_{nuj}\}, \tag{4.13}$$

where $\Theta^{\mathbb{P}}_{nuj}$ denotes the collection of all parameters under $\mathbb{P}$:

$$\Theta^{\mathbb{P}}_{nuj} = \{\kappa_v, \theta_v, \kappa_u, \theta_u, \lambda^+, \lambda^-\};$$

$\Theta^{\mathbb{Q}}_{nuj}$ denotes the collection of all parameters under $\mathbb{Q}$:

$$\Theta^{\mathbb{Q}}_{nuj} = \{\kappa^{\mathbb{Q}}_v, \theta^{\mathbb{Q}}_v, \kappa^{\mathbb{Q}}_u, \theta^{\mathbb{Q}}_u, \lambda^{+,\mathbb{Q}}, \lambda^{-,\mathbb{Q}}\};$$

$\Theta^{\mathbb{P},\mathbb{Q}}_{nuj}$ denotes the collection of all parameters, which share the same values under $\mathbb{P}$ and $\mathbb{Q}$:

$$\Theta^{\mathbb{P},\mathbb{Q}}_{nuj} = \{\sigma_v, \sigma_u, \rho, \mu_v, \mu_u, c_0^+\};$$

Θ^{rp}_{nuj} denotes the collection of all parameters related to risk premiums:

$$\Theta^{\text{rp}}_{nuj} = \{\eta_s, \eta_v, \eta_u\}.$$

Due to the restriction of $\kappa_i\theta_i = \kappa_i^{\mathbb{Q}}\theta_i^{\mathbb{Q}}$ with $i \in \{v, u\}$, we have

$$\theta_i^{\mathbb{Q}} = \kappa_i\theta_i/\kappa_i^{\mathbb{Q}} = \kappa_i\theta_i/\left(\kappa_i - \eta_i\right),$$

when κ_i, θ_i and η_i are given. In particular, it means that with this restriction the parameters: $\{\kappa_v^{\mathbb{Q}}, \theta_v^{\mathbb{Q}}, \kappa_u^{\mathbb{Q}}, \theta_u^{\mathbb{Q}}\}$ can be directly inferred, when the parameters: $\{\kappa_v, \theta_v, \eta_v, \kappa_u, \theta_u, \eta_u\}$ are known[53].

The parameters under $\mathbb{Q}$ (including the parameters with the same values under $\mathbb{P}$ and $\mathbb{Q}$) are chosen based on the estimation results of Andersen et al. (2020). For instance, c_0^+ varies from 1.5 to 12.879 across different markets in their estimation results. The value of c_0^+ is selected from their results for the S&P 500 option market. The rest of the parameters under $\mathbb{Q}$ are selected in a similar way. For the values of jump size parameters under $\mathbb{P}$, I follow the numerical relationship from estimation results from Bardgett et al. (2019) and Broadie et al. (2007).

Finally, η_v and η_u are selected based on the estimation results from Fulop and Li (2019). The concrete values are listed in Table 4.1.

Furthermore, subsequent empirical applications focus on the option market and its implied risk-neutral density. The following simulation study is carried out using synthetic option data without including underlying price fluctuations. If $\ln S_t$ is excluded from estimation, it is not possible to identify $\{\eta_s, \lambda^+\}$ with the current model setup. Thus, there are in total 15 free parameters left, as follows:

$$\begin{aligned}
\Theta^{est}_{nuj} &= \{\Theta^{est,\mathbb{P}}_{nuj}, \Theta^{est,\mathbb{Q}}_{nuj}, \Theta^{est,rp}_{nuj}\},\\
\Theta^{est,\mathbb{P}}_{nuj} &= \{\kappa_v, \theta_v, \kappa_u, \theta_u, \lambda^-\},\\
\Theta^{est,\mathbb{Q}}_{nuj} &= \{\lambda^{+,\mathbb{Q}}, \lambda^{-,\mathbb{Q}}, \sigma_v, \sigma_u, \rho, \mu_v, \mu_u, c_0^+\},\\
\Theta^{est,rp}_{nuj} &= \{\eta_v, \eta_u\}.
\end{aligned} \tag{4.14}$$

[53] Alternatively, as in Bardgett et al. (2019), free parameters can be specified as $\kappa_i^{\mathbb{Q}}, \kappa_i, \theta_i^{\mathbb{Q}}$ with $i \in \{v, u\}$. η_i and θ_i with $i \in \{v, u\}$ can be inferred from the risk premium definition and identity restriction.

Table 4.1: Parameters used in the Monte Carlo simulation.

under $\mathbb{P}$		under $\mathbb{Q}$	
Parameter	Value	Parameter	Value
κ_v	10.0000	$\kappa_v^{\mathbb{Q}}$	9.2000
θ_v	0.0200	$\theta_v^{\mathbb{Q}}$	0.0217
κ_u	2.0000	$\kappa_u^{\mathbb{Q}}$	1.5000
θ_u	1.8000	$\theta_u^{\mathbb{Q}}$	2.4000
λ^+	80.0000	$\lambda^{+,\mathbb{Q}}$	60.0000
λ^-	40.0000	$\lambda^{-,\mathbb{Q}}$	20.0000
Risk Premia		under $\mathbb{P}$ & $\mathbb{Q}$	
Parameter	Value	Parameter	Value
η_s	2.5	σ_v	0.4000
η_v	0.8000	σ_u	1.5000
η_u	0.5000	ρ	-0.9000
		μ_v	12.0000
		μ_u	200.0000
		c_0^+	7.5000

This table presents the selected values of parameters, which are used to generate synthetic data for simulation study. Those values of parameters for the data generating process in the simulation study are adapted from different literature.

4.2 Synthetic Data Generation

The primary objective of the simulation is to closely mimic the features of real-world data to be used in subsequent empirical studies. Therefore, the parameters for the data-generating process in the simulation study are adapted from different literature sources. The following discussion will refer to these parameters as true parameters. Table 4.1 summarizes the selected values of the static parameters.

Algorithm 2 provides details for generating synthetic data using the discretization methodology described in Eq. (3.11). In this simulation study, daily observations are generated over a period of eight years (252 sampling points per year). Section 4.2 illustrates the simulated time-series of log-return $\ln S_t$, volatility V_t and negative jump intensity U_t.

[54] $\mu_{t,n} = r_t - q_t + \gamma_{t,n}$, with excess return $\gamma_{t,n}$, which is defined of the form

$$\gamma_{t,n} = \eta_s V_{t,n} + c_0(\bar{\mu}^+ - \bar{\mu}^{+,\mathbb{Q}}) + U_{t,n}(\bar{\mu}^- - \bar{\mu}^{-,\mathbb{Q}}). \tag{4.15}$$

Algorithm 2: Data generation for SVNUJ

The for-loop for the operations with involving index n are ignored due to simplicity. They must be executed $\forall n \in \{1, ..., N\}$.

Input : $t = 1$, $\Delta t \in \mathbb{R}$, $T > 1 \in \mathbb{N}$, $N \in \mathbb{N}$, $\ln S_0 \in \mathbb{R}$, $V_0 \in \mathbb{R}^+$, $U_0 \in \mathbb{R}^+$

Parameters: $\{\Theta^{\mathbb{P}}_{nuj}, \Theta^{\mathbb{P},\mathbb{Q}}_{nuj}, \Theta^{rp}_{nuj}\}$ in Eq. (4.13)

Returns : $\ln S_{1:T,n}$, $V_{1:T,n}$, $U_{1:T,n}$

while $t \leq T$ **do**

$\epsilon^{(1)}_{t,n} \sim \mathcal{N}(0,1)$, $\epsilon^{(2)}_{t,n} \sim \mathcal{N}(0,1)$, $\epsilon^{(3)}_{t,n} \sim \mathcal{N}(0,1)$

$\Delta W^{(1)}_{t,n} \leftarrow \Delta t \epsilon^{(1)}_n$, $\Delta W^{(2)}_{t,n} \leftarrow \Delta t \epsilon^{(2)}_n$, $\Delta W^{(3)}_{t,n} \leftarrow \Delta t \epsilon^{(3)}_n$

$\Delta N^+_{t,n} \sim \text{Bernoulli}(c_0 \Delta t)$, $\Delta N^-_{t,n} \sim \text{Bernoulli}(U_t \Delta t)$

if $\Delta N^+_{t,n} == 1$ **then**

$Z^{s,+}_{t,n} \sim \exp(\lambda^+)$

end

if $\Delta N^-_{t,n} == 1$ **then**

$Z^{s,-}_{t,n} \sim \exp(\lambda^-)$

end

$$\ln S_{t,n} \leftarrow \ln S_{t-1,n} + (\mu_{t,n}{}^{54} - \tfrac{1}{2} V_{t-1,n})\Delta t + \sqrt{V_{t-1}} \left(\rho \Delta W^{(1)}_{t,n} + \sqrt{1-\rho^2} \Delta W^{(2)}_{t,n} \right) + Z^{s,+}_{t,n} \Delta N^+_{t,n} + Z^{s,-}_{t,n} \Delta N^-_{t,n}$$

$$V_t \leftarrow V_{t-1} + \kappa_v(\theta_u - V_{t-1})\Delta t + \sigma_v \sqrt{V_{t-1}} \Delta W^{(1)}_{t,n} + \mu_v (Z^{s,-}_t)^2 \Delta N^-_{t,n}$$

$$U_t \leftarrow U_{t-1} + \kappa_u(\theta_u - U_{t-1})\Delta t + \sigma_u \sqrt{U_{t-1}} \Delta W^{()}_{t,n} + \mu_u (Z^{s,-}_t)^2 \Delta N^-_{t,n}$$

$V_t \leftarrow \max(V_t, 0), \quad U_t \leftarrow \max(U_t, 0)$

$t \leftarrow t + 1$

end

Besides the generation of latent states, for the validation of the whole estimation approach, we need "observed" option prices. Specifically the option surface contains twenty option prices distributed equally over $m \in [-6, 2]$ at every sampling time point[55], where m denotes the standardized log-forward moneyness as in Eq. (A.9) for each maturity. The maturities τ are are 1 week, 1, 2, 3, 6 months(s) and 1 year. To obtain the annualized estimation result, the maturities are transformed into values with the unit of one year,

[55] Andersen et al. (2015a) chose $m \in [-4, 1]$ for their simulation study.

i.e., the set: $\{1/52, 1/12, 0.25, 0.5, 1\}$.

In order to simulate the real-world scenario, shocks must be applied to the simulated true option prices. For the option error, I assume the same as in Eq. (4.11) as follows:

$$\varepsilon_{t,i}^{\mathrm{IV}} = \frac{\mathrm{IV}_{t,i}^{\mathrm{model}} - \mathrm{IV}_{t,i}^{\mathrm{data}}}{\mathrm{IV}_{t,i}^{\mathrm{data}}} \sim \mathcal{N}(0, (\sigma_{t,i}^{\mathrm{IV,rel}})^2), \quad i = 1, \ldots, N_{\mathrm{IV},t}. \tag{4.16}$$

The noise $Z_{t,i}^{\mathrm{IV}}$ typically follows a standard normal distribution that is independent of time, moneyness, and maturity. Due to the set-up of relative errors in the option contracts, we have

$$\varepsilon_{t,i}^{\mathrm{IV}} = \sigma_{t,i}^{\mathrm{IV,rel}} Z_{t,i}^{\mathrm{IV}}. \tag{4.17}$$

Therefore, after generating the synthetic errors, the simulation target $\mathrm{IV}_{t,i}^{\mathrm{data}}$ can be computed as follows:

$$\mathrm{IV}_{t,i}^{\mathrm{data}} = \frac{\mathrm{IV}_{t,i}^{\mathrm{model}}}{1 + \varepsilon_{t,i}^{\mathrm{IV}}}, \tag{4.18}$$

where $\mathrm{IV}_{t,i}^{\mathrm{model}}$ is the "true" $\mathrm{IV}_{t,i}$, which can be computed by using cosine method explained in Section 2.1.4.2 with the simulated latent states $\{V_t, U_t\}$. As this simulation study is designed to investigate the performance of the Kalman filter with the linear Gaussian approximation and its quasi-maximum likelihood estimation approach on the behavior of the SVNUJ model, the impact of measurement errors is not that critical, as long as they are within a reasonable range. For simplicity, I use a single fixed $\sigma_{t,\tau,i}^{\mathrm{rel}}$ of 0.05 that is constant across time, moneynesses and maturities. As a result, this measurement error specification requires only one parameter. In this case, heteroscedasticity remains at the absolute level of measurement errors. With the choice of 0.05, the errors are approximately 5% of the level of simulated "observed" implied volatility. The value of 0.05 is selected based on the magnitude of the estimated error from Andersen et al. (2015a).

After constructing synthetic observations of BS implied volatilities with errors, we can obtain observations of cumulants based on option observations, as discussed in Chapter 2. Due to the noisy inputs of option observations and the limited number of observed contracts, model-free estimates of the cumulants are subject to errors. Unlike observations of options, the true distribution of measurement error is unknown. For simplicity, I assume

the relative errors of cumulant observations are standard normally distributed as follows:

$$\varepsilon^{\mathcal{K}}_{t,\tau,n} = \frac{\mathcal{K}^{\text{data}}_{t,\tau,n} - \mathcal{K}^{\text{model}}_{t,\tau,n}}{\mathcal{K}^{\text{data}}_{t,\tau,n}} \sim \mathcal{N}(0, (\sigma_t^{\mathcal{K}_{\tau,n},\text{rel}})^2), \tag{4.19}$$

where the standard deviation of the relative measurement error is assumed to be homogeneous over time, but maturity- and order-dependent, i.e., $\sigma_t^{\mathcal{K}_{\tau,n},\text{rel}} = \sigma^{\mathcal{K}_{\tau,n},\text{rel}}$. Finally I take the sample standard deviation of the cumulants in each group as estimates of $\sigma^{\mathcal{K}_{\tau,n},\text{rel}}$.

Figure 4.1: Simulated trajectories of latent states for the SVNUJ model.

Note: This figure illustrates the simulated trajectories for dynamics of Eq. (2.60) with the corresponding parameters from Table 4.1.This figure is divided into three panels: the top panel shows the time series of the log return, the middle panel shows the time series of V_t, and the bottom panel shows the time series of U_t.

4.3 Quasi-Maximum Likelihood Estimation

Based on the SSM representation in Definition Definition 4.1, the Kalman filter-based quasi-maximum likelihood approach can be applied. In the literature, MLE-based estimation is extensively used since it is easy to access, if the likelihood or its approximation is available. In this simulation study, I use the quasi-maximum likelihood approach proposed in Section 3.5.2. Following the general definition of the log-likelihood function in Eq. (3.31), we have

$$
\begin{aligned}
\ell(\Theta^{\text{est}_{nuj}}; w) &= (N^{\mathcal{K}}_{\text{total}} + N^{\text{IV}}_{\text{total}}) \left((1-w)\frac{\ell_{\mathcal{K}}(\Theta^{\text{est}})}{N^{\mathcal{K}}_{\text{total}}} + w\frac{\ell_{\text{IV}}(\Theta^{\text{est}}; \hat{X}_{0:T})}{N^{\text{IV}}_{\text{total}}} \right), 0 \le w \le 1, \\
\ell_{\mathcal{K}} &:= \ell(\{y^{\mathcal{K}}\}_{0:T}) \text{ in Eq. (3.26)} \\
&= -\frac{(T+1)p}{2}\log 2\pi - \frac{1}{2}\sum_{0}^{T}\left(\log|F^{\mathcal{K}}_t| + (v^{\mathcal{K}}_t)^\top (F^{\mathcal{K}}_t)^{-1} v^{\mathcal{K}}_t\right), \\
\ell_{\text{IV}} &:= -\frac{1}{2}\sum_{t=0}^{T}\sum_{i=1}^{N^{\text{IV}}_t}\left(\log\left(2\pi\sigma^2_{\text{IV,rel}}\right) + \frac{(\varepsilon^{\text{IV}}_i)^2}{\sigma^2_{\text{IV,rel}}}\right), \\
\varepsilon^{\text{IV}}_{t,i} &:= \frac{\text{IV}^{\text{model}}_{t,i}\left(\Theta^{\text{est}_{nuj}}; \hat{V}_t, \hat{U}_t\right) - \text{IV}^{\text{data}}_{t,i}}{\text{IV}^{\text{data}}_{t,i}}, \quad i = 1, \dots, N^{\text{IV}}_t,
\end{aligned}
\tag{4.20}
$$

where

$X_{0:T} = \{U_{0:T}, V_{0:T}\}$ is the time series of latent states and $\hat{X}_{0:T}$ is its corresponding posterior estimates in Eq. (3.22) from the Kalman filter;

$N^{\mathcal{K}}_{\text{total}}$ is the total number of observations used in the filtering step across the whole time-period;

$v^{\mathcal{K}}_t = y^{\mathcal{K}}_t - \left(Z^{\mathcal{K},nuj}_0 + Z^{\mathcal{K},nuj}_1 a^{nuj}_t\right)$, $F^{\mathcal{K}}_t = \text{Var}(y^{\mathcal{K}}_t | Y^{\mathcal{K}}_{t-1})$;[56]

N^{IV}_t is the number of observations of BS implied volatility at each sampling time point, and $N^{\text{IV}}_{\text{total}} := \sum_{t=0}^{T} N^{\text{IV}}_t$ is the total number of observations of BS implied volatility across the whole time-period;

[56]For example, if we have observations of cumulants following the order of $y^{\mathcal{K}}_t = (\mathcal{K}_{60,2}, \mathcal{K}_{60,3}, \mathcal{K}_{90,2}, \mathcal{K}_{90,3})^\top$ with $m \in \{2,3\}$ and $\tau \in \{60, 90\}$ days, then we define $Z^{\mathcal{K},nuj}_0 := \left(F^2_{0,60}, F^3_{0,60}, F^2_{0,90}, F^3_{0,90},\right)^\top$, $Z^{\mathcal{K},nuj}_1 := \left(F^2_{1,60}, F^3_{1,60}, F^2_{1,90}, F^3_{1,90},\right)^\top$. These are linear coefficients from Eq. (2.15). $a^{nuj}_t = \Phi^{nuj}_0 + \Phi^{nuj}_1 a^{nuj}_{t|t}$, where Φ^{nuj}_0 and Φ^{nuj}_1 are linear coefficients in Eq. (4.5), $a^{nuj}_{t|t} = \{V_{t|t}, U_{t|t}\}$ is the posterior mean of the latent states at each time point of t.

$\text{IV}_{t,i}^{\text{model}}\left(\Theta^{\text{est}_{nuj}}; \hat{V}_t, \hat{U}_t\right)$ is the function of BS implied volatility. It uses the estimates of the latent states $\hat{X}_{0:T}$, which are obtained as side results from evaluation of $\ell_{\mathcal{K}}$;

$\sigma_{\text{IV},rel}$ is the standard deviation of the measurement error for the observations of BS implied volatility. For simplicity, the same constant value is used as in previous data generation.

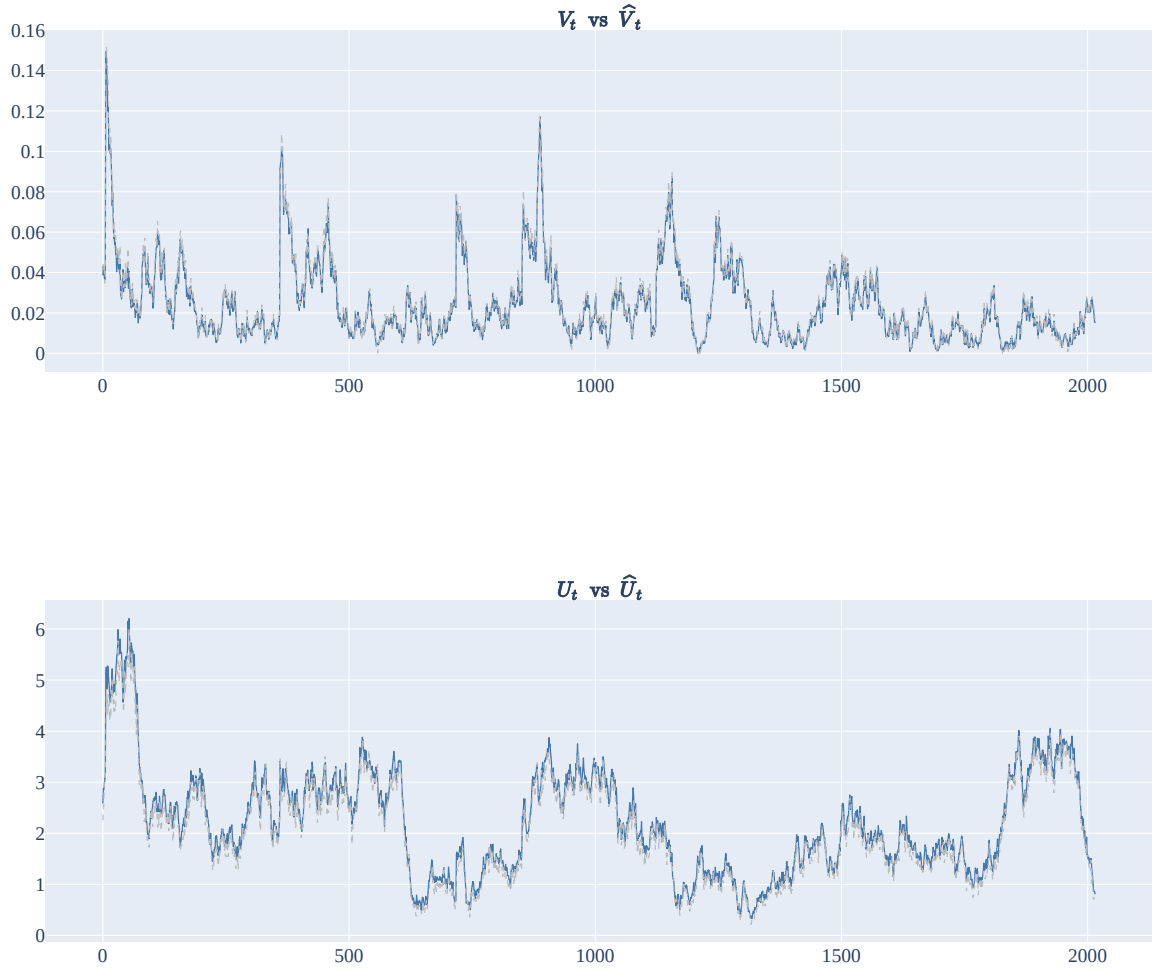
4.4 Estimation Results

4.4.1 Latent States Inference

In this section, validation of the inference of latent states is performed. The estimates of latent states are obtained from the filtering step in the Kalman filter. The state equations are derived from a linear Gaussian approximation on the discretized dynamics of the SVNUJ model, as discussed in Definition 4.1. Additionally, to mimic the real-world scenario the cumulants used in the measurement equations are obtained from model-free estimates. Here, I evaluate exclusively the performance of the Kalman filter with linear Gaussian approximation on the SVNUJ model with cumulants, assuming that the static parameters are known. In essence, the task consists of running a Kalman filter with measurement equations for cumulants. The potential failure to infer latent states can be attributed to two sources of bias. In the first instance, the linear Gaussian approximation of state equations can result in bias. In addition, the cumulants, which are computed in a model-free way, may not provide sufficient information to assist the filter in identifying latent states correctly.

Fig. 4.2 displays a visual comparison between the filtered states and the true states by plotting the filtered latent states $\{\hat{V}_{0:T}, \hat{U}_{0:T}\}$ and the true simulated latent states $\{V_{0:T}, U_{0:T}\}$. From visual comparison, it is clear that accurate estimates of the latent states can be obtained from the filtering step of the Kalman filter. In addition, it indicates that the potential bias resulting from the linear Gaussian approximation has a negligible impact on the accuracy of filtering. Therefore, model-free estimates of cumulants, which provide aggregate information on the risk-neutral density of the underlying return, are highly informative regarding persistent latent states in AJD models.

Figure 4.2: Filtered trajectories vs simulated trajectories given the true parameters.



The top panel plots the time series of the filtered volatility term $\{\hat{V}_t\}_{0:T}$ in blue and the time series of the true simulated volatility term $\{V_t\}_{0:T}$ in grey. The bottom panel plots the time series of the filtered negative jump intensity term $\{\hat{U}_t\}_{0:T}$ in blue and the time series of the true simulated negative jump intensity term $\{U_t\}_{0:T}$ in grey.

4.4.2 Static Parameters Estimation

The optimization library Pygmo developed by Biscani and Izzo (2020) is used in my implementation. For the estimation task in the simulation study, the SADE (Self-adaptive Differential Evolution) global optimizer from the library is used. As a result of the high computational burden associated with calculating the implied option prices for the model[57], a large portion of the computation time is spent in the filtering-independent step of the calculation. The following described practice experience is based on the Monte Carlo simulation study in this chapter. In general, the optimization converges well after

[57]OpenMP is used to parallelize the option pricing code.

3×10^3 iterations with a population of 20.[58] For the final runs in the simulation study 5×10^3 iterations with a population of 20 are performed. Each run takes approximately six hours on an Intel Xeon Gold 6230. Table 4.2 presents the estimation results for static parameters based on synthetic data, which is generated from the values of the parameters in the column of true value.

Table 4.2: Estimation results of static parameters for simulation study.

Parameters		True	Mean	Median	Std.	RMSE
Free to estimate	θ_u	1.8000	1.8692	1.8767	0.4383	0.4394
	θ_v	0.0200	0.0176	0.0167	0.0016	0.0029
	ρ	-0.9000	-0.8917	-0.8908	0.0526	0.0528
	η_u	0.5000	0.5558	0.4245	0.5126	0.5105
	η_v	0.8000	2.2454	2.8318	0.9644	1.7322
	κ_u	2.0000	2.0128	1.8717	0.5210	0.5160
	κ_v	10.0000	11.8179	12.3831	1.0180	2.0786
	$\lambda^{+,\mathbb{Q}}$	60.0000	64.0045	63.8071	1.3969	4.2366
	$\lambda^{-,\mathbb{Q}}$	20.0000	20.7002	20.7093	0.1898	0.7250
	λ^{-}	40.0000	41.4503	43.0402	7.7193	7.7782
	μ_u	200.0000	218.7066	217.8206	9.4844	20.9307
	μ_v	12.0000	13.1350	13.1238	0.1762	1.1483
	σ_u	1.5000	1.4097	1.4259	0.1060	0.1384
	σ_v	0.4000	0.4147	0.4134	0.0273	0.0308
	c_0^+	7.5000	8.5627	8.5093	0.5013	1.1728
Inferred	$\theta_u^{\mathbb{Q}}$	2.4000	2.4341	2.4276	0.0501	0.0602
	$\theta_v^{\mathbb{Q}}$	0.0217	0.0215	0.0215	0.0003	0.0004
	$\kappa_u^{\mathbb{Q}}$	1.5000	1.4570	1.4664	0.0612	0.0743
	$\kappa_v^{\mathbb{Q}}$	9.2000	9.5725	9.5448	0.2019	0.4228

Note: This table presents the results of the simulation study designed to demonstrate the quality of quasi-maximum likelihood estimation for the SVNUJ model. Parameters in the "Free to estimate" section are the ones that are actually optimized during the estimation process. Those parameters in the "Inferred" part have been directly derived from the estimation results. Column "True" provides the values of the parameters used in the simulation process. Statistics derived from repeated estimations of 50 sets of simulated data sequences are presented in the remaining columns.

[58] The LLK function is evaluated approximately 6×10^4 times during 3×10^3 iterations of optimization. As a default, the population is set to 20.

5 Empirical Applications

Based on the results of the simulation study reported in Chapter 4, this chapter examines the empirical performance of the aforementioned procedure for parameter estimation and inference of latent states. In particular, by using aggregate information from the option market, it is possible to estimate the risk-neutral parameters of AJD models, which are designed to model equity return dynamics.[59] In this chapter, the empirical validation of the Kalman filter-based QML approach, which is tested in the simulation study, is carried out with options on the S&P 500 index. The estimation results based on index options are further used to validate the future return predictability of the left tail factor modeled by the SVNUJ model. Moreover, with increased liquidity of single stock options, it is possible to estimate the model on a cross-sectional level by analyzing the options written on the constituents of the S&P 500 index. Continuing the empirical study on index options, this chapter examines the economic implications of the cross-sectional left tail factors. The rest of this chapter is organized as follows: In Section 5.1, the model is estimated with S&P 500 index options, in order to validate the Kalman filter-based QML estimation approach empirically. In Section 5.2, an empirical analysis of options written on the constituents of the S&P 500 index is conducted and its economic implications are discussed.

5.1 Application to S&P 500 Index Options

5.1.1 Introduction

In this section, we validate the estimation procedure empirically by applying it to a sample of options written on the S&P 500 index. In general, the performance of the

[59] In addition to the option prices, the future value of the underlying asset must also be considered in order to determine the moneyness of the option contract. If it is not directly feasible, researchers (see Bardgett et al. (2019), among others) often estimate it from the Put-Call-Parity (see Proposition A.5).

Kalman filter-based QML procedure proposed in Eq. (4.20) is examined by replicating the empirical findings about the left tail factor embedded in the risk-neutral S&P 500 return dynamics, which are extensively studied in the literature (see, e.g., Andersen et al. (2015b) and Andersen et al. (2020)). Furthermore, this empirical application compares the performance of estimation approaches with different specifications. Specifically, the question about whether raw options can contribute to improving the identification of static parameters and inference of latent states is addressed. Moreover, it is confirmed that the constant error in the cumulant measurement equations exhibits better performance.

5.1.2 The Model

5.1.2.1 Model Dynamic

This empirical analysis uses the same SVNUJ model as in the simulation study for the S& P 500 index options (see Section 4.1.1). Its dynamics under $\mathbb{P}$ and $\mathbb{Q}$ are defined exactly the same as in Eq. (4.1) and Eq. (4.2), respectively.

5.1.2.2 State-Space Model Representation

For the state-space representation in the filtering step, it is natural to follow the same convention as in Section 4.1.2. The same variance matrix with zero covariance terms is used, as follows:

$$\tilde{\Sigma}_t^{nuj} = \begin{pmatrix} \sigma_v^2 V_{t-1}\Delta t + \mu_v^2\sigma_{\lambda^-}^2 U_{t-1}\Delta t & 0 \\ 0 & \sigma_u^2 U_{t-1}\Delta t + \mu_u^2\sigma_{\lambda^-}^2 U_{t-1}\Delta t \end{pmatrix}. \tag{5.1}$$

5.1.2.3 More on Measurement Errors

In the literature on empirical applications based on the framework of state-space models, measurement errors in each scale-valued measurement equation are typically considered to be Gaussian with zero mean. The errors between different measurement equations are often assumed to be independent. Variances of measurement errors can be assumed to be either homogeneous (see, e.g., Fulop and Li (2019) and Bégin et al. (2020)) or heteroskedastic (see, e.g., Bardgett et al. (2019)) over time. Throughout this section, we assume that the measurement errors are independent Gaussian with zero mean and variances are exogenously given parameters. Let $\varepsilon_{t,\tau,n}^{\mathcal{K}}$ denote measurement error for observa-

tions of cumulants with maturity τ at time point t, and $\sigma_{t,\tau}^{\mathcal{K}_n}$ denote the standard deviation of the measurement errors. Model-free estimates of implied risk-neutral moments exhibit errors due to the integration of the estimators from Bakshi et al. (2003) which can only be calculated numerically based on limited observations. As discussed in Ammann and Feser (2019), linear extrapolation of option surfaces is applied for reaching the truncation point of the numerical integration. Accordingly, different configurations of linear extrapolation result in different point estimates for model-free moments. The CRAM-Database[60] provides two additional versions of model-free moment estimates. Loosely speaking, those two estimates approximate the upper and lower bounds of risk-neutral moments, which are denoted as $M_{t,\tau,n}^{\max}$ and $M_{t,\tau,n}^{\min}$, respectively. [61] Given that the moments and the cumulants of the same distribution are interchangeable, the "boundaries" of cumulants $\mathcal{K}_{t,\tau,n}^{\max}$ and $\mathcal{K}_{t,\tau,n}^{\min}$ are available as well. Thus, the relative spread between $\mathcal{K}_{t,\tau,n}^{\max}$ and $\mathcal{K}_{t,\tau,n}^{\min}$ with respect to $\mathcal{K}_{t,\tau,n}$ is defined as follows:

$$s_{t,\tau,n} := \frac{\mathcal{K}_{t,\tau,n}^{\max} - \mathcal{K}_{t,\tau,n}^{\min}}{\mathcal{K}_{t,\tau,n}}. \tag{5.2}$$

Loosely speaking, one can obtain an approximation of the standard deviation of measurement errors $\hat{\sigma}_t^{\mathcal{K}_{\tau,n}} = \frac{s_{t,\tau,n}}{2}$ based on the relative spread $s_{t,\tau,n}$, for each sampling point t. Therefore, if the measurement errors are assumed to be heteroskedastic, then the variance of measurement errors $(\hat{\sigma}_t^{\mathcal{K}_{\tau,n}})^2$ can directly be used in the estimation approach as a time-varying variance in the error of the corresponding measurement equations. Furthermore, the sample mean $\bar{\sigma}^{\mathcal{K}_{\tau,n}}$ of the whole time series of $\hat{\sigma}_t^{\mathcal{K}_{\tau,n}}$ can be used as a point estimate for the standard deviation of homogeneous measurement errors.

5.1.3 Data

The model is estimated by using cumulants of the risk-neutral S&P 500 index return density and option prices on that index. The end-of-day risk-neutral cumulants are computed from their model-free estimates of risk-neutral moments provided by the CRAM-Database, while the end-of-day prices of S&P 500 index options are collected directly from CBOE with the raw maturities. Furthermore, the risk-neutral moments in the CRAM-Database have been standardized. For the analysis, a sample of data with

[60] A large part of the data used in this thesis is provided by the CRAM research group at the Chair of Financial Economics and Risk Management. I would like to thank them.

[61] For the sake of simplicity, the return density indicator $R_s(t,\tau)$ at t is dropped, and only the time index t and maturity τ are retained for moments and cumulants.

maturities of 30, 60, 91, and 365 days is selected. The corresponding observations of cumulants can be computed directly from the moments. Through the empirical applications presented in this Chapter 4, the second and third cumulants are used as components of the measurement equations.

The sample period is from January 2nd, 2004 to August 21st, 2019, for a total of 3,936 business days. These data cover the global financial crisis of 2008 as well as the European debt crisis of 2010-2012. Moreover, in order to allow for an out-of-sample evaluation of the illustrative model, these 3,936 observations are divided into an in-sample period from January 2004 to December 2017, with 3,524 observations and an out-of-sample period from January 2018 to August 2019, with 412 observations. In Fig. 5.1, the top panel displays the time series for the second cumulants of risk-neutral S&P 500 index return dynamics with maturities of 30, 60, 91, and 365 days. The bottom panel of Fig. 5.1 displays the time series of the third cumulants with the same maturities. In addition, Table 5.1 reports the descriptive statistics of the second cumulants and the third cumulants.

Table 5.1: Summary statistics for the risk-neutral cumulants of the S&P 500 index return.

Second cumulant					
Days to maturity	Mean	Std.	Min.	Max.	ACF
30	0.003570	0.004968	0.000671	0.056632	0.963249
60	0.007446	0.008893	0.001721	0.098159	0.980639
91	0.011807	0.012587	0.003006	0.139756	0.986859
365	0.061382	0.043153	0.018731	0.373113	0.993751

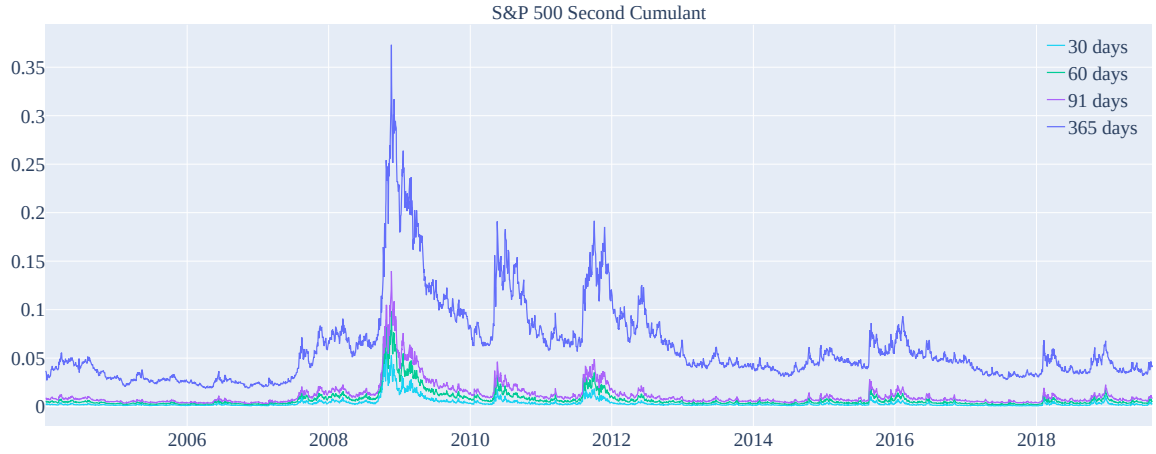
Third cumulant					
Days to maturity	Mean	Std.	Min.	Max.	ACF
30	-0.000620	0.001576	-0.028649	0.000017	0.902393
60	-0.001855	0.003664	-0.047096	-0.000138	0.961815
91	-0.003583	0.006128	-0.074538	-0.000327	0.976384
365	-0.037807	0.036131	-0.290106	-0.005025	0.989830

Note: This table presents summary statistics of the cumulants used for model estimation. These data include the risk-neutral second cumulant and third cumulant of S&P 500 return, which are computed based on risk-neutral moments derived from standardized option observations, with 30-, 60-, 91-, and 365-day maturities. The period covered by this study ranges from January 2004 to August 2019, for a total of 3,936 working days.

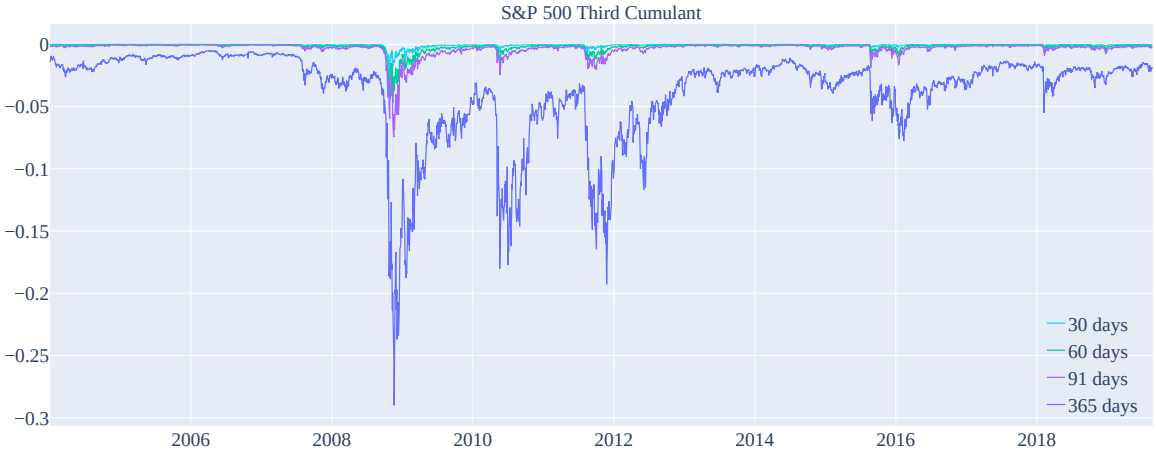
In accordance with Definition A.6, option prices are transformed into BS implied volatilities for use in estimation. Moreover, to ensure the robustness of the estimation, the

Figure 5.1: Time series of the risk-neutral cumulants of the S&P 500 index return.

(a) Second cumulant



(b) Third cumulant



This figure displays the cumulants used for model estimation. The data include the second (in Fig. 5.1a) and third cumulants (in Fig. 5.1b) of the risk-neutral density of the S&P 500 index, computed based on the risk-neutral moments derived from standardized option observations, with 30-, 60-, 91-, and 365-day maturities. The time period ranges from January 2004 to August 2019 for a total of 3,936 business days.

quotes that are used in estimation are selected based on the following criteria:

1. Option contracts are limited to maturities between ten days and one year.[62]

2. Option contracts are limited to OTM contracts (see Definition A.8).[63]

[62] Andersen et al. (2017) demonstrate that to fit options with ultra-short maturity one needs a model with time-varying jump parameters. This is beyond the scope of this application. Due to a lack of liquidity, all options with a maturity of more than one year are dropped.

[63] Normally, in an empirical application there is no exact ATM option (i.e., $m = 0$).

3. Option contracts must have a positive open interest.
4. Option contracts are restricted to standardized log-forward moneyness $m \in [-8, 2]$.
5. Call options and put options are included only if their prices exceed 0.5 and 0.1, respectively.

As a final step, the remaining option contracts are divided into five maturity bins which are $\{9 \text{ days} < \tau \leq 14 \text{ days}, 14 \text{ days} < \tau \leq 30 \text{ days}, 30 \text{ days} < \tau \leq 90 \text{ days}, 90 \text{ days} < \tau \leq 180 \text{ days}, 270 \text{ days} < \tau \leq 365 \text{ days}\}$.[64] To maintain the tractability of the estimation, at most twenty contracts are used, equally spaced within the range of moneynesses for each maturity within the chosen maturity bin. Additionally, if more than one maturity is available within the same maturity bin, the shortest maturity within the bin is retained at every sampling time point (i.e., every business day).

Table 5.2 presents descriptive statistics for option quotes grouped by moneyness and maturity. For the in-sample period, 272,437 observations of option contracts are used as inputs when the option-related part of the likelihood is active. Fig. 5.2 illustrates at-the-money implied volatility for options on the S&P 500 index in three different maturity bins of $\{14 \text{ days} < \tau \leq 30 \text{ days}, 90 \text{ days} < \tau \leq 180 \text{ days}, 270 \text{ days} < \tau \leq 365 \text{ days}\}$, which represent short-term, medium-term and long-term maturities. Furthermore, Fig. 5.3 plots the time series of relative error spreads for cumulants with maturities of 30-, 60-,91- and 365 days.

5.1.4 Quasi-Maximum Likelihood Estimation

As discussed in Section 5.1.2, the state-space model representation of the SVNUJ model and the data introduced in Section 5.1 provide all the ingredients necessary for current empirical application to the options on the S&P 500 index. Nevertheless, there are a number of open issues regarding the concrete specification of the quasi-maximum likelihood approach that still need to be addressed.

In the simulation study, the weight $w = 1$ is used in the log-likelihood function from Eq. (4.20), which means that raw BS implied volatilities are taken into account in the estimation approach. During estimation, the main computation burden is caused by evaluating the model-implied option prices. In the likelihood function, specifying $w = 0$ (i.e., using only cumulants as measurements) allows for a more efficient estimation in

[64] τ denotes days to maturity.

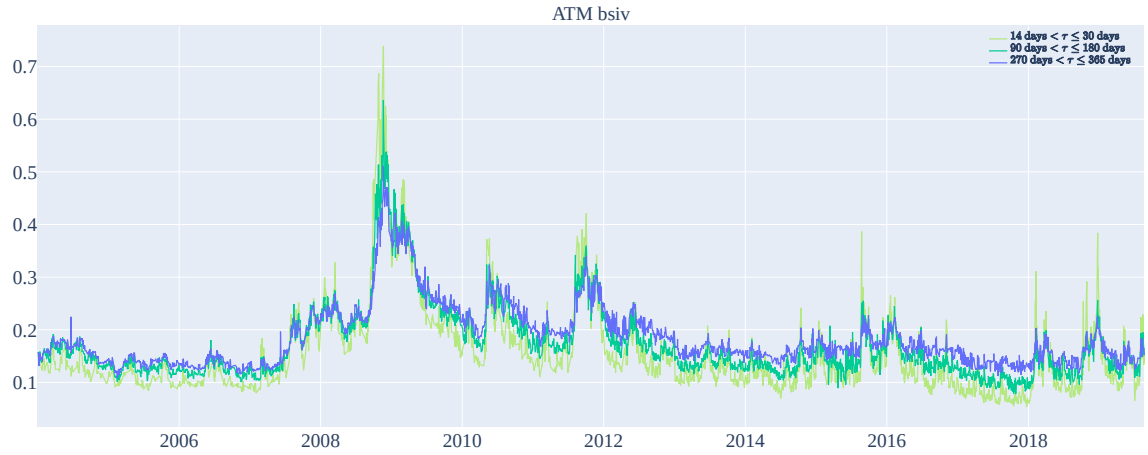
Table 5.2: Number of contracts for the index options and average BS implied volatilities of the index options.

Number of contracts						
	$9 < \tau \leq 14$	$14 < \tau \leq 30$	$30 < \tau \leq 90$	$90 < \tau \leq 180$	$270 < \tau \leq 365$	All
$m \leq -3$	9,533	17,462	23,554	24,927	21,011	96,487
$-3 < m \leq -1$	10,908	18,058	20,892	19,298	18,323	87,479
$-1 < m \leq 1$	11,077	18,277	21,695	21,254	22,846	95,149
$m > 1$	3,893	6,717	8,374	7,664	7,023	33,671
All	35,411	60,514	74,515	73,143	69,203	312,786

Average implied volatility						
	$9 < \tau \leq 14$	$14 < \tau \leq 30$	$30 < \tau \leq 90$	$90 < \tau \leq 180$	$270 < \tau \leq 365$	All
$m \leq -3$	0.270	0.298	0.324	0.346	0.370	0.322
$-3 < m \leq -1$	0.194	0.220	0.239	0.257	0.274	0.237
$-1 < m \leq 1$	0.138	0.154	0.164	0.179	0.189	0.165
$m > 1$	0.121	0.131	0.133	0.132	0.137	0.131
All	0.181	0.201	0.215	0.228	0.242	0.213

Note: This table presents descriptive statistics for option quotes grouped by maturity and moneyness. The data are collected from CBOE. The time period ranges from January 2004 to August 2019 for a total of 3,936 working days.

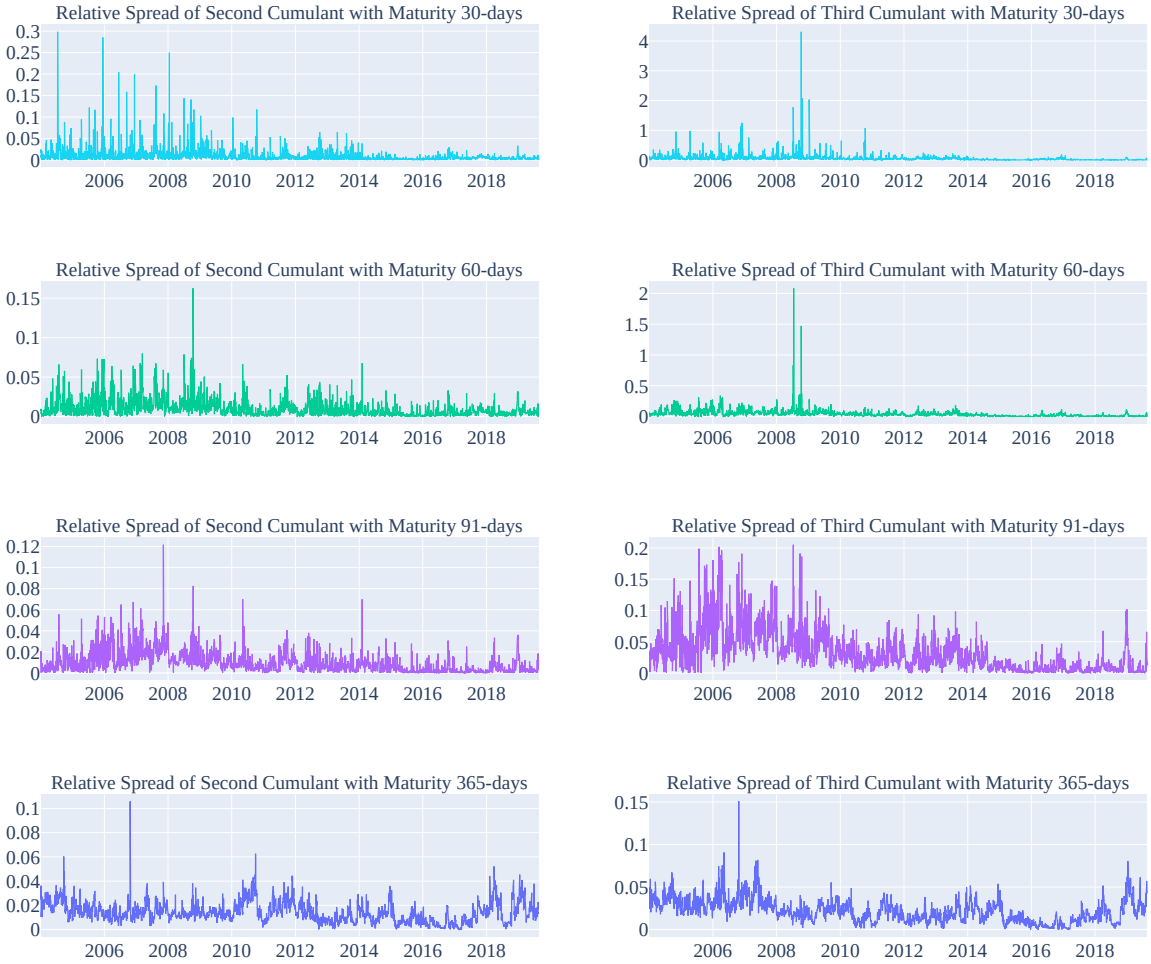
Figure 5.2: BS implied volatilities of ATM options written on the S&P 500 index.



This figure plots at-the-money BS implied volatilities of options written on the S&P index within three different maturity bins of $\{14 \text{ days} < \tau \leq 30 \text{ days}, 90 \text{ days} < \tau \leq 180 \text{ days}, 270 \text{ days} < \tau \leq 365 \text{ days}\}$. The time period ranges from January 2004 to August 2019 for a total of 3,936 working days.

terms of computational time. Arguably the specification of $w = 1$ tends to outperform the specification of $w = 0$ in terms of option pricing performance. Nevertheless, if the

Figure 5.3: Time series of the relative spread of the boundaries on cumulants.



This figure plots the time series of relative spreads between the lower bounds and upper bounds of the cumulants $s_{t,\tau,n}$ defined in Eq. (5.2). The data include the risk-neutral second cumulant and third cumulant of the S&P 500 index returns, which are computed based on risk-neutral moments derived from standardized option observations, with maturity at 30-, 60-, 91-, and 365-days. The time period ranges from January 2004 to August 2019 for a total of 3,936 business days.

primary objective of the estimation task is solely to obtain the filtered trajectories of the latent states, the specification of $w = 0$ may be able to perform similarly to the specification of $w = 1$.[65]

Furthermore, unlike the simulation study, the true values of the cumulants are unknown. Therefore, it is still an open issue about how to specify measurement errors properly. The relative spread between the lower and upper estimates of the cumulants demonstrates the uncertainty of the model-free estimates of the cumulants. Intuitively the wider the

[65]For the hypothesis to be valid, the model-free cumulants must be accurate, i.e., the option panels must be liquid enough.

relative spread, the less informative the cumulants are with regard to the latent states. Based on the time-varying relative spreads displayed in Fig. 5.3, a specification of time-varying standard deviations in measurement errors of cumulants is therefore possible. With this specification, the time-varying standard deviations of measurement errors may provide guidance to the Kalman filter regarding the informativeness of the corresponding cumulants. Alternatively, one can use the mean of the relative spreads to obtain a constant standard deviation for the measurement errors in the cumulants.

Based on the two groups of specification choices, we can have four different specifications which are a combination of the previous discussion, as follows:

I. time varying relative errors in the filtering step and $w = 0$ in the log-likelihood function (i.e., only cumulants are used);

II. time varying relative errors in the filtering step and $w = 1$ in the log-likelihood function (i.e., both cumulants and options are used);

III. constant relative errors in the filtering step and $w = 0$ in the log-likelihood function (i.e., only cumulants are used);

IV. constant relative errors in the filtering step and $w = 1$ in the log-likelihood function (i.e., both cumulants and options are used).

Finally, the actual estimations are carried out on the bwUniCluster 2.0. For each specification, the estimation is performed with 3.5×10^3 iterations in the optimization with the SADE (Self-adaptive Differential Evolution) global optimizer from the same optimization library used in the simulation study. The SADE algorithm employs a population of 20. Thus, there are in total 7×10^4 evaluations of the log-likelihood function in the entire optimization approach.

5.1.5 Estimation Results

5.1.5.1 Estimates of Static Parameters

Table 5.4 presents the estimation results for the static parameters of the SVNUJ model for the four specifications considered for the quasi-maximum likelihood approach discussed in the previous section. Due to the fact that the SVNUJ model used in this empirical study differs from the AJD models used in the literature, it is not possible to establish reference values for many of the parameters estimated. Even so, in the option pricing

literature for options on the S&P 500 index, there is also "common knowledge" about the values of parameters that are shared among the various estimated AJD models.

For example, the ρ indicates a negative correlation between returns and volatilities (see, e.g., Broadie et al. (2007)). In the Spec I, $\rho = 1$ is estimated, which is inconsistent with the empirical features of underlying index returns in general. Additionally, the estimated κ differs significantly from the rest of the specifications. Spec I may not be able to capture the empirical features of the index returns, based on the estimation results of static parameters. Thus, measurement errors with time-varying standard deviation may represent a poor choice of specification in this empirical study, even though the initial notion of this sort of measurement error is appropriate. In contrast, Spec II, which has the same error specification, provides a more reasonable fit of static parameters when compared to Spec I. Therefore, if there is a potential bias in the filtering step, the additional filtering independent measurement of BS implied volatilities may correct for it. Moreover, the closest model to the SVNUJ model is the 2FU model from Andersen et al. (2020), but they only specify the dynamics of index return under $\mathbb{Q}$ (see Definition 2.5). They assume that the stochastic process of U_t only contains jumps, and has a long run mean of zero, i.e., $\sigma_u^{\mathbb{Q}} = 0$ and $\theta_u^{\mathbb{Q}} = 0$. Interestingly, the values of $\sigma_u^{\mathbb{Q}} = 0$ and $\theta_u^{\mathbb{Q}} = 0$ are very close to zero in the results of Spec IV. This is consistent with the model assumption of index return from Andersen et al. (2020). Thus, this provides an indirect evidence that the Kalman filter-based QML approach can be a viable alternative to Andersen et al. (2015a)'s PNLS approach. Furthermore, in both Spec II and Spec III, there are no obvious flaws in the estimation results. Therefore, in order to verify the quality of these two specifications, it is necessary to conduct further analyses.

5.1.5.2 Inference of Latent States

As discussed previously, further assessments are needed to evaluate the accuracy of the estimation results. In the SVNUJ model, all jumps are compound Poisson processes. By definition, jumps modeled by compound Poisson processes are random variables, containing jump intensity and jump size at each sampling point. Considering that jump risks cannot be traded on the market directly, it is difficult to distinguish between the final impact of jump intensity and jump size on model-implied density. Thus, there are always trade-offs between these two components in jumps. This can result in multiple possible combinations of jump parameters yielding similar impacts on the evaluation of the likelihood for the same data input. As a result, U_t as one of the latent states

Table 5.4: Estimation results of static parameters using index options.

Parameters		Spec I	Spec II	Spec III	Spec IV
Free to estimate	η_u	1.871070	3.487719	0.091833	1.968787
	η_v	0.000000	11.451380	9.969258	3.027268
	κ_u	6.627896	5.448218	2.147148	2.860269
	κ_v	100.000000	25.496293	15.324954	8.774694
	$\lambda^{+,\mathbb{Q}}$	139.824486	103.066571	794.396653	162.129886
	$\lambda^{-,\mathbb{Q}}$	9.604448	26.074150	33.315375	19.206782
	λ^{-}	6.924542	137.844590	43.770451	10.995742
	μ_u	128.728216	659.058700	1137.304645	164.280137
	μ_v	36.989211	29.082117	24.344695	17.004407
	ρ	1.000000	-1.000000	-1.000000	-1.000000
	σ_u	0.171105	0.207742	1.795734	0.000100
	σ_v	0.800000	0.524919	0.376289	0.465119
	θ_u	0.331705	0.489239	0.758610	0.000100
	θ_v	0.010556	0.005416	0.004628	0.012330
	c_0^+	39.457550	9.263435	0.004988	26.703292
Inferred	$\kappa_u^{\mathbb{Q}}$	4.756826	1.960499	2.055315	0.891482
	$\kappa_v^{\mathbb{Q}}$	100.000000	14.044912	5.355696	5.747426
	$\theta_u^{\mathbb{Q}}$	0.462179	1.359593	0.792505	0.000321
	$\theta_v^{\mathbb{Q}}$	0.010556	0.009832	0.013243	0.018824

Note: This table presents the results of different specifications for the Kalman filter-based QML estimation approach using cumulants and BS implied volatilities from the S&P 500 index options. Parameters in the "Free to estimate" section are the ones that are actually being optimized during the estimation process. Those parameters in the "Inferred" part are directly derived from the estimation results. There are in total four different specifications for this empirical analysis. Spec I represents the specification of time varying relative errors in the filtering step and $w = 0$ in the log-likelihood function; Spec II represents the specification of time varying relative errors in the filtering step and $w = 1$ in the log-likelihood function; Spec III represents the specification of constant relative errors in the filtering step and $w = 0$ in the log-likelihood function; Spec IV represents the specification of constant relative errors in the filtering step and $w = 1$ in the log-likelihood function.

that drives the negative jump intensity can vary widely in magnitude between different specifications. Following Andersen et al. (2020), to bring the factor representing the negative jumps in the return dynamics to the same magnitude across all specifications, the negative jump variation is used instead of U_t itself as the left tail factor. The negative jump variation is defined as follows:

$$\mathrm{JV}_t^{\mathrm{neg}} := \frac{2}{(\lambda^{\mathbb{Q},-})^2} U_t, \tag{5.3}$$

Fig. 5.4 plots the filtered trajectories of model implied spot volatilities and negative

jump variations which are extracted based on the estimated parameters from the four specifications. Comparing the results from Spec I with the rest of the specifications, it is evident that the results differ significantly in terms of volatility and negative jump variation. We can also see that spot volatilities and negative jump variations are well identified in the other three specifications. In line with the empirical findings of Fulop and Li (2019) and Feunou and Okou (2018), the trajectory of jump intensities can be identified correctly by using aggregate information from the option market, i.e., the risk-neutral cumulants.

Moreover, the trajectories of spot volatilities and negative jump variations from specifications with constant measurement errors in cumulants follow very similar patterns to the trajectories that are inferred by Andersen et al. (2020). To further confirm the informativeness of cumulants in terms of pinning down the trajectory of negative jump variations. The correlation coefficients between the trajectories of spot volatilities and negative jump variations between different specifications are reported in Table 5.5. According to the correlation coefficients for both spot volatility and negative jump variation, Spec III performs almost as well as Spec IV in this empirical analysis with a correlation coefficient higher than 0.99. Therefore, the results of the filtering step with only cumulants suggest that aggregate information from the option market can also deliver reliable performance in pinpointing the evolution of spot volatility and jump intensity for this application.

Table 5.5: Correlations between filtered trajectories of latent states.

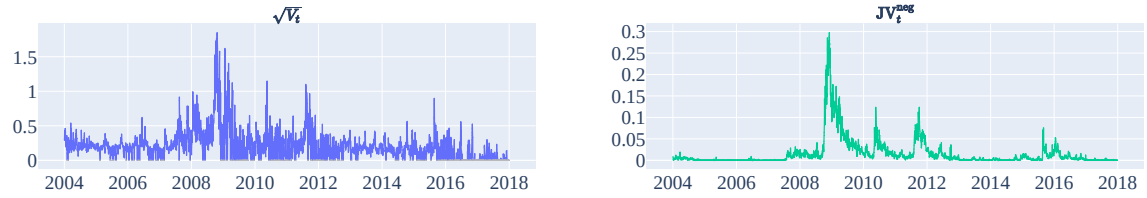
$\sqrt{V_t}$				
	Spec I	Spec II	Spec III	Spec IV
Spec I	1.000000	-	-	-
Spec II	0.821071	1.000000	-	-
Spec III	0.731190	0.948609	1.000000	-
Spec IV	0.744234	0.953622	0.996205	1.000000

$\mathrm{JV}_t^{\mathrm{neg}}$				
	Spec I	Spec II	Spec III	Spec IV
Spec I	1.000000	-	-	-
Spec II	0.931655	1.000000	-	-
Spec III	0.846658	0.944911	1.000000	-
Spec IV	0.882495	0.963885	0.995198	1.000000

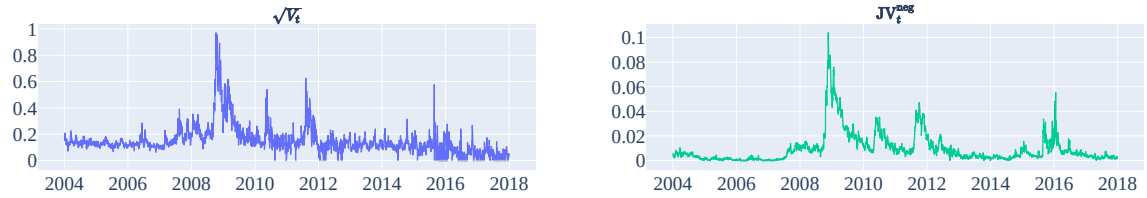
Note: This table presents the correlation coefficients between the trajectories extracted by using different specifications for the quasi-maximum likelihood estimation approach with the cumulants and BS implied volatilities of S&P 500 index options. The top panel shows the correlation between $\sqrt{V_t}$, and the bottom panel shows the correlation between $\mathrm{JV}_t^{\mathrm{neg}}$. There are in total four different specifications for this empirical analysis. Spec I represents for the specification of time varying relative errors in the filtering step and $w = 0$ in the log-likelihood function; Spec II represents the specification of time varying relative errors in the filtering step and $w = 1$ in the log-likelihood function; Spec III represents the specification of constant relative errors in the filtering step and $w = 0$ in the log-likelihood function; Spec IV represents the specification of constant relative errors in the filtering step and $w = 1$ in the log-likelihood function.

Figure 5.4: In-sample filtering results for the different specifications.

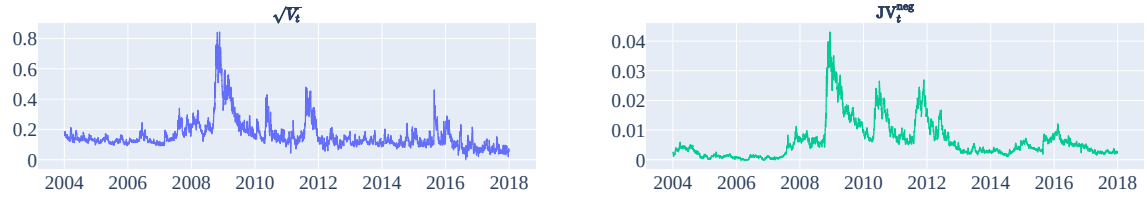
(a) w = 0 with time-varying error



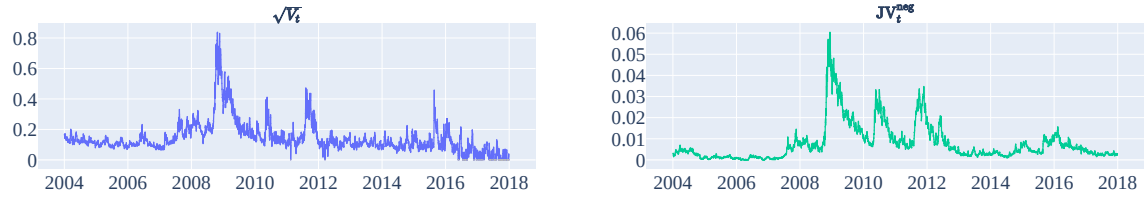
(b) w = 1 with time-varying error



(c) w = 0 with constant error



(d) w = 1 with constant error



This plot shows the results of filtering steps based on the estimated parameters from four different specifications. Panel (a) displays the trajectories of latent states estimated from Spec I, which represents the specification of time varying relative errors in the filtering step and $w = 0$ in the log-likelihood function; Panel (b) displays the trajectories of latent states estimated from Spec II, which represents the specification of time varying relative errors in the filtering step and $w = 1$ in the log-likelihood function; Panel (c) displays the trajectories of latent states estimated from Spec III, which represents the specification of constant relative errors in the filtering step and $w = 0$ in the log-likelihood function; Panel (d) displays the trajectories of latent states estimated from Spec IV, which represents the specification of constant relative errors in the filtering step and $w = 1$ in the log-likelihood function.

5.1.5.3 Option Pricing Performance

A discussion of the model's performance on pricing options is provided in the following section in order to further examine the estimation results. Following Ornthanalai (2014) and Bégin et al. (2020), the pricing errors are reported in the BS implied volatility root-mean-square errors (IVRMSE)), and its relative value (RIVRMSE), based on the same idea of the loss function used in the quasi-maximum likelihood estimation. The IVRMSE and RIVRMSE are defined as

$$\text{IVRMSE} = \sqrt{\frac{1}{N_{\text{total}}^{\text{IV}}} \sum_{k=1}^{N_{\text{total}}^{\text{IV}}} \left(\text{IV}_k^{\text{model}} - \text{IV}_k^{\text{data}}\right)^2} \tag{5.4}$$

and

$$\text{RIVRMSE} = \sqrt{\frac{1}{N_{\text{total}}^{\text{IV}}} \sum_{k=1}^{N_{\text{total}}^{\text{IV}}} \left(\frac{\text{IV}_k^{\text{model}} - \text{IV}_k^{\text{data}}}{\text{IV}_k^{\text{data}}}\right)^2}. \tag{5.5}$$

Table 5.6 reports the in-sample pricing performance of the estimated model with Spec III and Spec IV. Clearly Spec IV outperforms Spec III in this assessment. However, this is not surprising, as Spec III only employs risk-neutral cumulants during estimation. In general, the cumulants as aggregate surface characteristics provide less information in comparison to the original surface. In comparison to option pricing literature, Spec III provides a fairly realistic calibration result for the option pricing task, despite outperforming Spec IV in current assessments. For example, Bégin et al. (2020) report that their in-sample IVRMSE and RIVRMSE are 3.086% and 14.391%, respectively.[66] Therefore, when computational resources are limited, the estimation approach with only cumulants is an appealing alternative, even if the model is intended for option pricing.

5.1.6 Return Predictability

Andersen et al. (2015b) demonstrate a novel insight regarding the left tail factor U_t of the risk-neutral return dynamics for the S&P 500 index, namely that the left tail factor forecasts equity returns over medium- and long-term horizons.[67] Fig. 5.5 presents the returns of the S&P 500 index over different time horizons. Furthermore, Andersen et al. (2020) show that the left tail factor can also be used to predict equity returns in European markets as well. Hence, in this part of the analysis, the return predictability of the left

[66] As a matter of fact, the current model dynamic differs from that of Bégin et al. (2020)'s model assumption.

[67] In their 3F model, there are two volatility factors and a negative jump intensity factor.

tail factor estimated in the previous section is examined. This can be viewed as indirect evidence of the quality of the left tail factor that is extracted from the Klaman filter-based QML approach. To examine predictability, the same set-up of predictive regressions as in Andersen et al. (2020) is used, as follows:

$$r^{\log}_{t,t+h}{}^{68} = \ln S_{t+h} - \ln S_t = \alpha_h + \beta_{v,h} \cdot V_t + \beta_{u,h} \cdot \mathrm{JV}_t^{\mathrm{neg},\perp V} + \epsilon_{t,t+h}, \tag{5.6}$$

where $\mathrm{JV}_t^{\mathrm{neg},\perp V}$ denotes the orthogonal component of the risk-neutral expected negative jump variation, as defined in Eq. (5.8), relative to the spot variance V_t. In the current regression analysis, the t-statistic is computed based on Newey-West SEs with a lag length of $1.3\sqrt{T}$ in accordance with the specification of Andersen et al. (2020). 5.6 presents the results of the predictive regression Eq. (5.6). Over forecasting horizons ranging from 12 weeks to 30 weeks, the results from both specifications indicate substantial return predictability. As compared to Andersen et al. (2020), the quasi-maximum likelihood estimation approach provides very similar evidence of return predictability for the left tail factor. Accordingly, the left tail factor can explain future returns of the S&P 500 index, which is the same conclusion reached by Andersen et al. (2020). Furthermore, the t-statistic and R^2 from Spec III follow the same pattern as the results from Spec IV. Thus, this result implies that cumulants, as aggregate information derived from option surfaces, are sufficient for inferring latent states in the case of options on the S&P 500 index. As an additional illustration, Fig. B.1 presents predicted returns and realized returns at different time horizons. The predicted returns are obtained from predictive regressions based on the estimation results from Spec IV.

[68] To align with Andersen et al. (2015b), log returns is used in this study.

Figure 5.5: S&P 500 index returns.

The top panel of this figure plots cumulative returns of the S&P 500 index between January 2004 and August 2019. The rest panels report the daily, monthly and six-monthly returns of the S&P 500 index with sampling frequency 1-day, 1-month and 6-month.

Table 5.6: Pricing errors on the options used for the estimations.

Spec III					
Overall IVRMSE and RIVRMSE			Sorting by year		
	IVRMSE	RIVRMSE		IVRMSE	RIVRMSE
All	2.485	9.357	2004	1.204	8.475
			2005	1.093	8.917
Sorting by maturity			2006	1.309	9.708
			2007	1.743	9.277
$m \leq -3$	3.273	6.594	2008	5.539	11.819
$-3 < m \leq -1$	1.615	5.152	2009	3.567	9.054
$-1 < m \leq 1$	1.893	9.822	2010	2.516	9.356
$m > 1$	3.469	19.457	2011	3.500	11.395
			2012	1.635	8.959
Sorting by moneyness			2013	1.336	7.454
			2014	1.904	8.429
9 days $< \tau \leq$ 14 days	2.931	11.437	2015	2.800	9.729
14 days $< \tau \leq$ 30 days	3.096	10.875	2016	1.932	9.690
30 days $< \tau \leq$ 90 days	2.811	9.242	2017	1.285	8.626
90 days $< \tau \leq$ 180 days	1.922	7.406			
270 days $< \tau \leq$ 365 days	1.746	8.934			

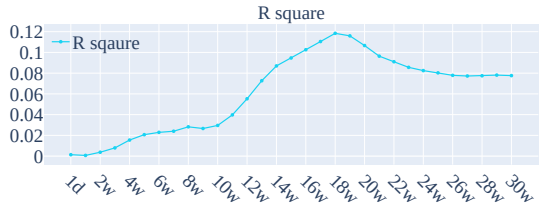
Spec IV					
Overall IVRMSE and RIVRMSE			Sorting by year		
	IVRMSE	RIVRMSE		IVRMSE	RIVRMSE
All	2.066	7.852	2004	1.080	6.674
			2005	0.846	5.822
Sorting by maturity			2006	1.015	6.702
			2007	1.528	7.772
$m \leq -3$	2.633	6.076	2008	4.706	10.804
$-3 < m \leq -1$	1.570	5.482	2009	2.609	7.282
$-1 < m \leq 1$	1.687	8.693	2010	2.067	7.776
$m > 1$	2.495	13.156	2011	2.664	8.602
			2012	1.734	8.044
Sorting by moneyness			2013	1.359	6.272
			2014	1.610	6.891
9 days $< \tau \leq$ 14 days	2.480	10.016	2015	2.274	7.893
14 days $< \tau \leq$ 30 days	2.488	8.897	2016	1.642	8.577
30 days $< \tau \leq$ 90 days	2.193	7.554	2017	1.298	9.097
90 days $< \tau \leq$ 180 days	1.509	6.057			
270 days $< \tau \leq$ 365 days	1.817	7.806			

Note: This table reports the pricing errors on the options used for the estimations. The IVRMSEs and RIVRMSEs are the implied BS volatility root-mean-square errors, and its relative values, respectly. The top panel shows the errors of Spec III. Spec III represents the specification of constant relative errors in the filtering step and $w = 0$ in the log-likelihood function. The bottom panel reports the errors of Spec IV. Spec IV represents the specification of constant relative errors in the filtering step and $w = 1$ in the log-likelihood function. IVRMSEs and RIVRMSEs are given in percentages.

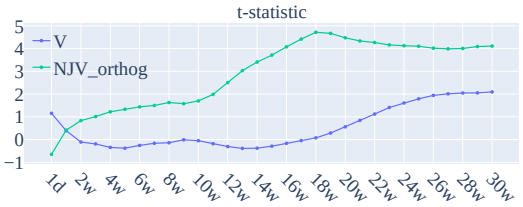
Figure 5.6: Predictive regressions for returns of the S&P 500 index.

(a) w = 0 with constant error

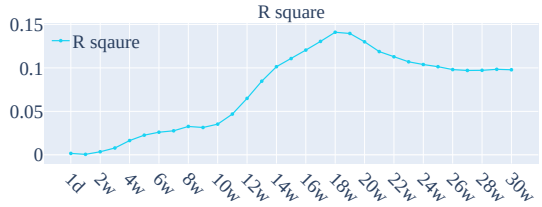
(b) w = 0 with constant error



(c) w = 1 with constant error



(d) w = 1 with constant error



This plot reports the results from prediction regression based on the extracted negative jump variation from Spec III and Spec IV. (a) shows the t-statistics from the result with Spec III and (b) shows the corresponding R^2. (c) shows the t-statistics from the result with Spec IV and (d) shows the corresponding R^2. Spec III represents the specification of constant relative errors in the filtering step and $w = 0$ in the log-likelihood function; Spec IV represents the specification of constant relative errors in the filtering step and $w = 1$ in the log-likelihood function.

5.1.7 Summary

In this subsection, an empirical analysis of the options written on the S&P 500 index is carried out. In terms of estimated parameters, except for Spec I, the Kalman filter-based QML estimation approach with the rest of specifications can yield reasonable realistic results for estimating the risk-neutral parameters in comparison to the results from Andersen et al. (2020). With regard to the results about inference of latent states, specifications with constant relative errors (i.e., Spec III and Spec IV) show better performance. If the goal of estimation is to extract the trajectories of latent states, a constant relative error is a preferred choice. As a result of the investigation of option pricing and return predictability, both the results from Spec III and Spec IV demonstrate positive performance in both of these assessments. The estimated model can capture the key empirical features of the option-implied risk-neutral distribution of the underlying return dynamic. Furthermore, the assessment of return predictability also highlights that the estimation approach solely based on cumulants obtained from a liquid option market delivers decent results. In cases where there are limited computational resources, this approach may be useful for tackling the inference problem.

5.2 Application to Single Stock Options

5.2.1 Introduction

As demonstrated in the empirical application in Section 5.1, the left tail factor, extracted from forward-looking information of equity indices that is not spanned by regular volatility factors, represents tail risk premiums. In more detail, the trajectory of this left tail factor is strongly linked with equity premiums, meaning that the extracted negative jump intensity included in the forward-looking information from the option market can be used to predict medium- and long-term performance. Andersen et al. (2015b) show that the left tail factor differs economically from the regular volatility factor.

Aside from parametric approaches, the relationship between cross-sectional equity returns and forward-looking information in option surfaces has been extensively studied as well. The study by Conrad et al. (2013) examines the relationship between risk-neutral moments and realized returns. Their study finds that there is a negative association between risk-neutral variance and subsequent stock returns, but it is not statistically significant. An et al. (2014) find that changes in Black–Scholes implied volatilities contain the expansionary power of future returns. Furthermore, based on the results of Martin and Wagner (2019), the expected return of a stock is derived from the risk-neutral variance of the market, each stock's risk-neutral variance, and the value-weighted average of their risk-neutral variances.

Other researchers work with the forward-looking information and attempt to demonstrate the relationship between tail risk embedded in the forward-looking information and equity risk-premium. Based on a non-parametric approach, Bollerslev and Todorov (2011) find that jump risk pricing implied by option data accounts for a significant fraction of equity risk premiums. Additionally, these findings are confirmed using the options on equity indices from U.S. and European markets in a parametric manner in Andersen et al. (2015b) and Andersen et al. (2020). In more recent work, Bégin et al. (2020) demonstrate that idiosyncratic risk contributes to the equity premium only when it arises from jump risk. Tail risk thus plays a central role in idiosyncratic risk pricing.

In light of the findings in the literature, a natural follow-up question arises, namely whether the cross-sectional left tail factor in option markets is capable of predicting cross-sectional future equity returns? In general, answering this question will provide insight to a key issue in financial economics, namely which kind of information drives

the equity risk premium. Initially, this question may appear trivial, or one may merely assume that it holds true due to the relationship between the market index and the single stocks. This is not yet even theoretically proven, since index options and options on its constituents are traded separately. This is different from an index that is directly derived from its constituents' price levels. Furthermore, with large amounts of cross-sectional data, a parametric framework is also challenging from a technical point of view for the estimation of an advanced model.

To capture the cross-sectional dynamics of equity option surfaces and to quantify the cross-sectional left tail risk, we again use the SVNUJ model, which is used in the empirical study on the index options. Based on widely documented empirical evidence, this model incorporates stochastic volatility and stochastic negative jump intensity. The model is based on the idea presented in Andersen et al. (2020). The complexity of option pricing and the large number of samples collected from cross-sectional option surfaces present a technical challenge for existing estimation methodologies with limited computing resources.

We use the same Kalman filter-based QML approach as in the previous section to resolve this bottleneck in cross-sectional estimation. As demonstrated previously, the key step of this approach is using the linear relationship between the latent states and estimates of conditional moments from option surfaces to facilitate the filtering step (see, e.g., Feunou and Okou (2018) and Fulop and Li (2019)), since the linear relationship between the non-parametric version of risk-neutral moments is not only informative about the latent states, but also facilitates the filtering step through Kalman filtering under certain assumptions and approximations of transition density.

The key research questions for this empirical application are twofold. First, we would like to examine whether the model can be estimated accurately with cross-sectional data using the Kalman filter-based QML approach. Second, we would like to investigate if future returns can be predicted by the target variable in a cross-sectional manner. With the extracted negative jump intensity, we are able to perform the analysis. In this application, we estimate the model on 254 firms that are or were constituents of the S&P 500 index. To accomplish this, we use the option surfaces of each individual firm between 2004 and 2021. Using a direct comparison to the extensively studied risk-neutral return density of the S&P 500 index, we assess the estimation results from our approach. We confirm that the estimation results deliver a high quality of fit to the cross section of risk-neutral return densities in terms of option pricing performance. Finally, we explore

the implications of the left tail risk factor extracted from the cross-sectional option data. As shown by our empirical results, negative jump intensity is a predictor, confirming the hypothesis regarding the relationship between left tail risk and equity premium. The rest of this section proceeds as follows: Section 5.2.2 introduces the cross-sectional model. Section 5.2.3 reviews the input data. Section 5.2.4 explains the estimation approaches. Section 5.2.5 presents the estimation results. Section 5.2.6 presents the empirical findings, i.e., the cross section of left tail risk factors and its return predictability. Finally, Section 5.2.8 concludes.

5.2.2 The Model

5.2.2.1 Model Dynamic

The same SVNUJ model used in the simulation study and application to index options (see Section 4.1.1) is also employed for the model setups of single stock options. Its dynamics under $\mathbb{P}$ and $\mathbb{Q}$ are defined identically to those in Eq. (2.61).

5.2.2.2 State-Space Model Representation

For the filtering step in the estimation, it is natural to follow the same transformation as in Section 4.1.2. The same variance matrix with non-zero covariance approximation as Eq. (4.6) is used, which takes the form

$$\Sigma_t^{nuj} = \begin{pmatrix} \sigma_v^2 V_{t-1}\Delta t + \mu_v^2\sigma_{\lambda^-}^2 U_{t-1}\Delta t & \mu_v\mu_u\sigma_{\lambda^-}^2 U_{t-1}\Delta t \\ \mu_v\mu_u\sigma_{\lambda^-}^2 U_{t-1}\Delta t & \sigma_u^2 U_{t-1}\Delta t + \mu_u^2\sigma_{\lambda^-}^2 U_{t-1}\Delta t \end{pmatrix}. \tag{5.7}$$

Furthermore, as shown in Section 5.1.5.2, constant relative errors for the measurement errors result in a more robust performance in the filtering step. Therefore, we adopt the same convention for the application of single stock options.

5.2.3 Data

Through out this application, we use daily option surfaces on the constituents of the S&P 500 index from January 2nd, 2004 to December 31st, 2021. We further select the time interval from January 2004 to December 2017 as the in-sample period for the estimation step in the application of cross-sectional data, and the rest as the out-of-sample period. The end-of-day option prices are collected directly from CBOE with raw maturities.

The end-of-day standardized model-free estimates of risk-neutral moments, which are obtained from the option surfaces, are provided directly by the CRAM-Database. The daily returns of individual stocks are obtained from Yahoo Finance for the purpose of further empirical analysis.

On the other hand, due to the fact that individual stock options are American style rather than European style, one needs to account for the early exercise premium. The European option price for single stocks is formed by subtracting the early exercise premium from the American option price. Following the discussion in the literature, the prices of option contracts are transformed into Black-Scholes implied volatilities that are actually used in the estimation. The resulting volatility surfaces are available in the CRAM-Database as well. Moreover, to ensure the robustness of the estimation, we select quotes that are used in estimation based on the following criteria:

1. Option contracts are limited to maturities between ten days and one year.[69]
2. Option contracts are restricted to OTM option contracts.
3. Option contracts must have a positive open interest.
4. Options costs more than 50 cents, irrespective of it being a call or put option.

The remaining option contracts are divided into five maturity bins which are $\{9 \text{ days} < \tau \leq 14 \text{ days}, 14 \text{ days} < \tau \leq 30 \text{ days}, 30 \text{ days} < \tau \leq 90 \text{ days}, 90 \text{ days} < \tau \leq 180 \text{ days}, 270 \text{ days} < \tau \leq 365 \text{ days}\}$.[70] We only keep the shortest maturity within each bin. Furthermore, we keep a maximum of twenty contracts for each remaining maturity, which represent the volatility surfaces well. These contracts are evenly distributed within the range of moneynesses for each maturity.[71] The end-of-day standardized risk-neutral cumulants are computed from model-free estimates of the risk-neutral moments provided by the CRAM-Database. We omit these outliers in the data of risk-neutral moments that violate the no-arbitrary condition for the dynamic asset-pricing model in general, since the cross section of option surfaces contains several ill-behavioral outliers due to the potential liquidity issue.[72] Due to the fact that the risk-neutral moments are standardized, we select data with time-invariant forward-looking horizons across the entire time period

[69] Andersen et al. (2017) demonstrate that to fit the options with ultra-short maturity one needs a model with time-varying jump parameters. This is beyond the scope of this application. All options with a maturity of more than one year are dropped due to a lack of liquidity.

[70] τ denotes days to maturity.

[71] The grids are created equally distributed within the moneyness range of the remaining contracts. We use the nearest neighbourhood approach to select contracts.

[72] The procedure follows the same idea of monotonously property of SVIX from Martin (2017).

of 30, 60, 91, and 365 days. The corresponding observations of cumulants then can be computed directly from the moments.

Since the S&P 500 index constituents change from time to time, the cross-sectional options and cumulants do not always cover enough time periods for the estimation approach to produce a stable estimation result. Thus, we only use stocks with data available for 90% of the time spans within the sample period. Additionally, certain firms may experience challenges during the subsequent empirical analysis. There may be issues during the estimation procedure, during the out-of-sample period latent state filtering step, or due to the lack of return data from Yahoo Finance. As a result, 254 stocks remain for the final empirical analysis.[73] Although we have an estimation approach designed to technically ease estimation, we are still constrained by the issue of speed, when the raw option surfaces are included in the log-likelihood function.[74] Therefore, in addition to the data filtering criteria outlined above, only six raw options across the log-moneyness at each remaining maturity for each Wednesday in the sample are considered for estimation.[75]

5.2.4 Quasi-Maximum Likelihood Estimation

Following Ornthanalai (2014) and Bégin et al. (2020), we use the same quasi-maximum likelihood approach used in estimation with index option data. The likelihood function defined in Eq. (4.20) is written so that different weights can be applied to the likelihood of filtering-dependent part with cumulants versus the likelihood of filtering-dependent part with option surfaces. Clearly, by setting $w = 0$ one can get the fast variant of the likelihood function, which is first introduced in Feunou and Okou (2018). This version of the likelihood function benefits from a pure Kalman filter based approach. It may, however, have the risk that only a limited amount of information about risk-neutral density is utilized. On the other hand, one can specify $w = 1$ using cumulants for extracting latent states and use the rich option surfaces as the filtering independent observations in the estimation procedure. The estimation procedure can be viewed as a light weight version of estimation approach proposed by Andersen et al. (2015a).

[73]For the purpose of clarity, we have classified the firms into different categories in the appendix. See Table C.1 for the firms used, Table C.2 for the firms excluded, and Table C.3 for the firms that fail during the estimation process.

[74]Due to the limited resources on the cluster, we cannot estimate over two hundred individual stocks in a reasonable amount of time.

[75]The moneynesses are selected based on four equally distributed OTM put options and two OTM call options across all the remaining contracts at each maturities for each data sampling time point.For illustrative purposes, a figure demonstrating this for Apple Inc on December 19, 2012, can be found in Fig. B.2.

Furthermore, the following practical expert choices are based on the experience from a simulation study and an application of the estimation procedure to option surfaces on the S&P 500 index. First of all, we choose $w = 1$ in our estimation of cross-sectional option surfaces based on the experience we obtained from our simulation study. Finally, the actual estimations are carried out on the bwUniCluster 2.0. For this empirical application, a certain amount of computational resources is required. In order to perform the maximization of Eq. (4.20), we use the optimization library Pygmo developed by Biscani and Izzo (2020). We select the SADE (Self-adaptive Differential Evolution) global optimizer from the library. As a result of the high computational burden associated with calculating the implied option prices for the model, a large portion of the computation time is spent on the option pricing step of the calculation.[76] In general, the optimization converges well at 3×10^3 iterations with a population of 20. For the empirical application of the cross-sectional data we perform the maximization with 3×10^3 iterations with a population of 20. Thus, there are in total 6×10^4 evaluations of the log-likelihood function in each optimization approach.

5.2.5 Estimation Results

5.2.5.1 Estimates of Static Parameters

In Table 5.7, the results of a QML estimation of the S&P 500 index and its constituents are presented, displaying both optimized and directly inferred parameters. A variety of results are presented, including individual estimates, overall averages, and specific quantiles. Moreover, Table 5.8 reports the average standard errors as well as their quantiles. The similar model 2FU in Andersen et al. (2020) includes only the specification of the dynamic of returns under $\mathbb{Q}$. In their model, the negative jump intensity process contains only jumps, and has a long-run mean of zero, i.e., $\sigma_u^{\mathbb{Q}} = 0$ and $\theta_u^{\mathbb{Q}} = 0$. Interestingly in our estimation results of index option data, the values of $\sigma_u^{\mathbb{Q}}$ and $\theta_u^{\mathbb{Q}}$ in the estimation results are very close to zero in the results. This is consistent with their model assumption of the return process. On the other hand, based on our results, the average values of these two parameters estimated with data from single stocks in the negative jump process differ significantly from zero. Despite the substantial variation across cross-sections, the variance and negative intensity process parameters are comparable to those in the index model. It is interesting to note that the mean reversion speeds of both latent factors

[76] The option pricing code is parallelized with OpenMP.

are greater than those estimated by the index. Additionally, the average of the long-run mean parameters of the negative jump intensity process is higher than the estimate based on market indexes. Therefore, we expect the latter in the inference of negative jump intensity to be higher on average based on cross-sectional data. Hence, we conclude that the cross-sectional risk-neutral dynamics on average show significant differences as compared to the index dynamics, confirming that option trades on single stocks contain different information than option trades on the market index.

Table 5.7: Summary statistics of estimated parameters for S&P 500 index and its constituents.

		SPX	Mean	5%	25%	50%	75%	95%
Free	η_u	1.9942	1.1989	0.0101	0.5438	1.3281	1.9472	2.0000
	η_v	0.2316	5.2983	0.0100	1.0450	5.1392	9.9455	10.0000
	κ_u	2.5074	5.3372	2.6376	4.0374	4.9156	5.8634	8.2744
	κ_v	4.7967	32.5541	8.6329	18.9716	33.7861	49.0509	50.0000
	$\lambda^{+,\mathbb{Q}}$	54.9081	45.2323	1.2686	7.2837	21.8321	70.9701	164.7940
	$\lambda^{-,\mathbb{Q}}$	13.7135	30.1629	20.5597	26.2629	29.3011	33.7043	41.3518
	λ^{-}	6.2378	215.9363	28.7672	104.2367	174.1366	294.1800	500.0000
	μ_u	48.2347	1333.3134	579.5812	1322.2566	1496.4831	1499.9998	1500.0000
	μ_v	10.5777	30.1845	1.0000	3.1544	11.6263	33.9188	111.5239
	ρ	-1.0000	-0.3997	-1.0000	-1.0000	-0.9854	0.3602	1.0000
	σ_u	0.0222	0.6497	0.0001	0.0001	0.2197	0.7739	3.3412
	σ_v	0.4518	0.1377	0.0005	0.0139	0.0572	0.1886	0.5062
	θ_u	0.0001	3.0234	0.7726	1.8498	2.6030	4.1276	6.4950
	θ_v	0.0213	0.0139	0.0001	0.0054	0.0124	0.0196	0.0374
	c_0^+	2.2106	10.2245	0.0026	0.1226	1.7900	19.6653	39.7083
Inferred	$\kappa_u^{\mathbb{Q}}$	0.5132	4.1383	2.0878	2.8437	3.5548	4.2807	7.8742
	$\kappa_v^{\mathbb{Q}}$	4.5650	27.2559	1.2605	12.3221	29.9090	41.9997	49.9897
	$\theta_u^{\mathbb{Q}}$	0.0005	4.1242	1.0341	2.4632	3.3341	5.1530	9.1450
	$\theta_v^{\mathbb{Q}}$	0.0224	0.0182	0.0008	0.0078	0.0153	0.0234	0.0498

Note: This table reports the results of the QML estimation approach using the cumulants and Black-Scholes implied volatilities of S&P 500 index and its constituents. Parameters in the upper panel are QML optimized, while parameters in the lower panel are inferred from the parameters of the upper panel and no-arbitrage restrictions. Parameters for each stock are estimated using daily cumulants and weekly out-of-the-money options. Parameters for each stock are estimated using daily cumulants and available weekly cross-sections of out-of-the-money options, with in-sample period from January 2004 to December 2017. The "Mean" column reports the arithmetic average of each parameter estimated from the cross-sectional samples individually. The "SPX" column reports the corresponding estimated results with data of S&P 500 index. The 5%, 25%, 50%, 75% and 90% columns report the corresponding quantile for the point estimates across the firms.

5.2.5.2 Inference of Latent States

Further assessments of the inference of latent states are presented in this section. As discussed in Section 5.1.5.2, U_t as one of the latent states that drives the negative jump

Table 5.8: Standard errors of point estimates from QML estimation for S&P 500's constituents.

	Mean	5%	25%	50%	75%	95%
η_u	0.0368	0.0154	0.0237	0.0301	0.0410	0.0727
η_v	0.3708	0.0204	0.0919	0.2975	0.5526	0.8894
κ_u	0.0474	0.0171	0.0267	0.0350	0.0489	0.1025
κ_v	0.6887	0.0222	0.1403	0.6762	1.0779	1.7349
$\lambda^{+,\mathbb{Q}}$	3.6204	0.0080	0.0802	0.2642	0.7133	6.3226
$\lambda^{-,\mathbb{Q}}$	0.1129	0.0327	0.0573	0.0895	0.1339	0.2805
λ^{-}	24.2788	1.6344	3.4408	5.1458	10.3406	151.3489
μ_u	15.0897	4.1201	7.8436	11.4792	16.3760	34.2491
μ_v	0.9792	0.0258	0.2332	0.5408	1.2340	3.1516
ρ	6.0961	0.0240	0.1313	0.3266	0.9558	6.4910
σ_u	0.1121	0.0020	0.0121	0.0342	0.1138	0.4957
σ_v	0.0059	0.0002	0.0009	0.0028	0.0081	0.0188
θ_u	0.0885	0.0251	0.0444	0.0678	0.1053	0.2104
θ_v	0.0002	0.0001	0.0001	0.0002	0.0003	0.0006
c_0^+	0.5660	0.0001	0.0039	0.0495	0.5068	3.1134

Note: This table presents the standard errors of point estimates resulting from the QML estimation approach using the cumulants and Black-Scholes implied volatilities of the S&P 500's constituents. The "Mean" column reports the arithmetic average of standard errors of each parameter, as estimated from the cross-sectional samples individually. The 5%, 25%, 50%, 75% and 90% columns report the corresponding quantile for the standard errors.

intensity can vary widely in magnitude between different specifications. By following Andersen et al. (2020), again to bring the factor, which represents the negative jumps in the return dynamics, to the same magnitude across all different specifications, we use the negative jump variation,

$$\mathrm{JV}_t^{\mathrm{neg}} := \frac{2}{(\lambda^{\mathbb{Q},-})^2} U_t, \tag{5.8}$$

instead of U_t itself as the left tail factor.

Fig. 5.7 plots the average of spot volatilities and negative jump variations based on the estimated parameters from individual stock data and the derived factors from market index data. One can clearly observe that the results from the average of the cross section of the factors differ considerably in terms of volatility and negative jump variation compared to values from index data. We observe that the average of cross-sectional volatility is higher than the volatility extracted from the market index in non-crisis time periods. This is consistent with the empirical fact that the implied volatilities of options on single stocks tend to be higher than the implied volatilities of options on the market index.

The average cross-sectional volatility terms in general follow a very similar pattern to the volatility terms of market indexes during the financial crisis. As mentioned above, the jump intensity and size parameters can have certain trade-off effects among themselves, so the negative jump variation will act as a total tail part, which can be used to compare among estimates. The average cross-sectional negative jump variation, however, differs considerably from the jump variation level of the market index. In light of this, we conclude that the tail information embedded in the cross section of the market is not redundant. This empirical pattern is consistent with Bégin et al. (2020), who find that idiosyncratic jump risk matters for the equity risk premium.[77]

5.2.5.3 Option Pricing Performance

The goal of this section is to assess the performance of calibrated models in option pricing. Following Ornthanalai (2014) and Bégin et al. (2020), we report price performance using the BS implied volatility root-mean-square errors (IVRMSE), and its relative value (RIVRMSE), based on the same idea of the loss function used in the quasi-maximum likelihood estimation. The IVRMSE and RIVRMSE are calculated in the same way as in the application to the index data (see Eq. (5.4) and Eq. (5.5)). In principle, the IVRMSE is an absolute measure of implied-volatility pricing error, while the RIVRMSE is a relative measure that is likely to be more effective when comparing pricing errors over time.

Table 5.9 summarizes the in-sample pricing performance of the estimated model with options on the S&P 500 index and options on single stocks across different dimensions of maturity, moneyness and time. The estimated model in the current assessment provides a reasonably realistic calibration result for option pricing compared to results found in the option pricing literature. Using the S&P index option surfaces, we obtain an overall IVRMSE of 2.05%, which is similar to the best performing 3F model in Andersen et al. (2015b) with 1.77% of error. The 3F model is slightly more complicated than our model.[78] Furthermore, from our estimation results, the IVRMSE and RIVRMSE error metrics for the stocks are 6.81% and 10.76%, respectively. Closer to our study, Bégin et al. (2020) is the only study to use a parametric model to characterize risk-neutral density with cross-sectional option data in the literature. In their estimation, they have an IVRMSE of 7.36% and a RIVRMSE of 17.24% for the single stocks.[79] Therefore, our estimated

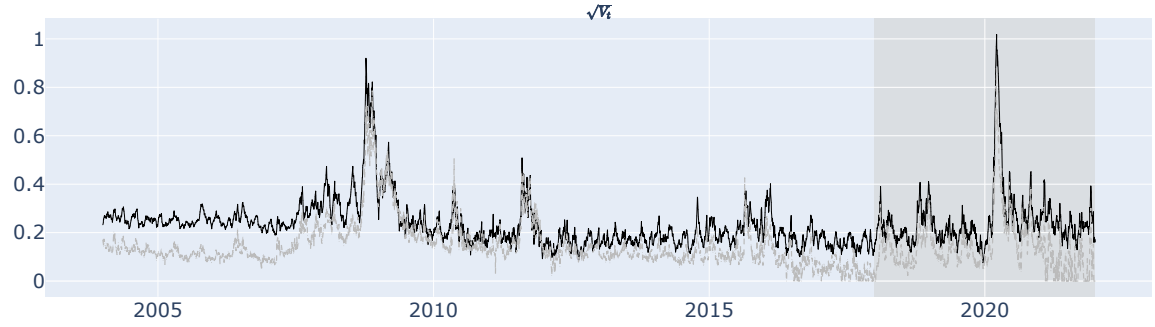
[77]This holds, if we can show that the cross-sectional left tail factors are correlated with cross-sectional equity return premiums.

[78]In their 3F model, there are two volatility factors and a negative jump intensity factor.

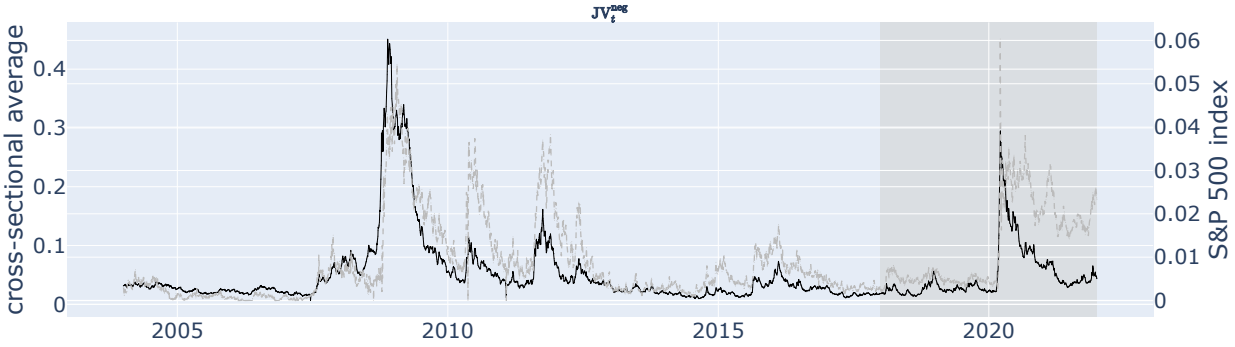
[79]In addition, our results with the S&P 500 index options outperform the corresponding results from

Figure 5.7: Filtering results for index and cross-sectional average.

(a) Volatility



(b) Negative jump intensity



This figure plots the results of filtering steps based on the estimated parameters with data from single stocks and market index. Panel (a) displays the trajectories of volatility extracted with cumulants from risk-neutral density and the optimized parameters from estimation with option data. The black line reports the average volatility level across 254 single stocks. The dashed gray line depicts the volatility level of the market index. Panel (b) displays the trajectories of negative jump variation $\frac{2}{(\lambda^{\mathbb{Q},-})^2}U_t$ extracted with cumulants from risk-neutral density and the optimized parameters from estimation with option data. The black line reports the average negative jump variation level across 254 single stocks. The dashed gray line depicts the negative jump variation level of the market index. The grey areas in all the panels indicate the results of the out-of-sample period. In Panel (b), the left y-axis shows the scale for the cross-sectional average, while the right y-axis shows the scale for the market index.

models perform similarly in the error metric at the absolute level and perform better in the relative error metric in comparison to the results in Bégin et al. (2020). As a result of a comparison with the two closest studies in the literature, we conclude that our estimation procedure provides an empirically credible estimate of the risk-neutral density and the conditional realization of latent states with respect to the option pricing task.

An out-of-sample pricing error is reported in Table 5.10. As can be seen from the relative

Bégin et al. (2020) as well.

pricing errors with respect to maturity and moneyness buckets, they are similar to the in-sample errors, indicating that the model is able to capture salient features of the data.

Table 5.9: In-sample pricing errors for S&P 500 and single stock options.

SPX					
Overall IVRMSE and RIVRMSE			Sorted by year		
	IVRMSE	RIVRMSE		IVRMSE	RIVRMSE
All	2.049	7.425	2004	1.444	6.737
			2005	0.851	5.606
Sorting by maturity			2006	1.085	6.683
			2007	1.578	7.804
14 days $< \tau \leq$ 30 days	2.517	8.873	2008	4.847	10.538
30 days $< \tau \leq$ 90 days	2.215	7.287	2009	2.496	7.420
90 days $< \tau \leq$ 180 days	1.544	6.209	2010	2.140	7.845
270 days $< \tau \leq$ 365 days	1.877	7.401	2011	2.375	8.055
			2012	1.763	7.801
Sorting by moneyness			2013	1.309	6.073
			2014	1.515	6.314
$m \leq -3$	2.596	5.953	2015	1.979	7.113
$-3 < m \leq -1$	1.753	5.262	2016	1.487	7.520
$-1 < m \leq 1$	1.661	8.448	2017	1.226	7.663
$m > 1$	2.125	11.630			

Single stocks					
Overall IVRMSE and RIVRMSE			Sorted by year		
	IVRMSE	RIVRMSE		IVRMSE	RIVRMSE
All	6.812	10.758	2004	6.655	9.631
			2005	4.063	9.143
Sorting by maturity			2006	6.957	9.690
			2007	9.173	11.768
14 days $< \tau \leq$ 30 days	7.779	14.631	2008	9.889	12.674
30 days $< \tau \leq$ 90 days	6.746	9.721	2009	7.474	10.318
90 days $< \tau \leq$ 180 days	6.294	9.965	2010	6.481	10.553
270 days $< \tau \leq$ 365 days	7.724	11.321	2011	5.537	9.729
			2012	5.472	11.138
Sorting by moneyness			2013	6.076	10.104
			2014	6.336	9.451
$m \leq -3$	24.020	19.844	2015	5.894	10.826
$-3 < m \leq -1$	4.816	10.828	2016	8.264	11.498
$-1 < m \leq 1$	5.248	10.051	2017	5.295	11.470
$m > 1$	18.435	17.151			

Note: This table reports aggregated and disaggregated pricing errors of the option pricing model for S&P 500 (top panel) and single stock options (bottom panel). IVRMSE and RIVRMSEs are in percentage terms and stand for implied volatility and relative implied volatility RMSE, respectively. The time horizon is 2004 to 2017.

Table 5.10: Out-of-sample pricing errors for S&P 500 and single stock options.

SPX					
Overall IVRMSE and RIVRMSE			Sorted by year		
	IVRMSE	RIVRMSE		IVRMSE	RIVRMSE
All	3.686	9.413	2018	2.847	8.181
			2019	0.974	5.468
Sorting by maturity			2020	5.707	12.571
			2021	3.592	10.028
14 days $< \tau \leq$ 30 days	4.295	10.875			
30 days $< \tau \leq$ 90 days	3.505	9.453			
90 days $< \tau \leq$ 180 days	3.727	7.847			
270 days $< \tau \leq$ 365 days	3.088	9.224			
Sorting by moneyness					
$m \leq -3$	4.370	6.918			
$-3 < m \leq -1$	2.442	5.239			
$-1 < m \leq 1$	2.854	13.149			
$m > 1$	3.335	15.027			

Single stocks					
Overall IVRMSE and RIVRMSE			Sorted by year		
	IVRMSE	RIVRMSE		IVRMSE	RIVRMSE
All	7.234	14.566	2018	4.436	10.633
			2019	5.356	11.546
Sorting by maturity			2020	10.247	18.623
			2021	6.509	14.611
14 days $< \tau \leq$ 30 days	10.276	22.345			
30 days $< \tau \leq$ 90 days	5.936	11.266			
90 days $< \tau \leq$ 180 days	5.792	10.969			
270 days $< \tau \leq$ 365 days	7.006	12.673			
Sorting by moneyness					
$m \leq -3$	17.919	24.741			
$-3 < m \leq -1$	6.170	11.228			
$-1 < m \leq 1$	6.441	14.752			
$m > 1$	10.674	17.465			

Note: This table reports out-of-sample fitting errors of single stock's option-implied volatility. IVRMSE and RIVRMSE are in percent and stand for Black-Scholes implied volatility and Black-Scholes implied relative implied volatility RMSE, respectively.

5.2.6 Return Predictability: Cross-Sectional Panel Regressions

5.2.6.1 Spot Volatility and Orthogonal Tail Risk

In this section, we examine the most critical implication in this empirical analysis, i.e., the return predictability of negative jump variation extracted from data of single stocks. Andersen et al. (2015b) present a novel insight regarding the left tail factor U_t of risk-

neutral return dynamics for the S&P 500 index, namely that the left tail factor forecasts equity returns over the medium- and long-term.[80] Furthermore, Andersen et al. (2020) show that the left tail factor can be used to predict equity returns on European markets as well. Consequently, in this part of the analysis, we examine the return predictability of the left tail factor estimated in Section 5.2.5.2. To examine the predictability, we follow the set-up of predictive regression for cross-sectional data from the literature (see An et al. (2014), Lewellen et al. (2015), Martin and Wagner (2019), among others). We run panel regressions of realized returns at horizon 1, 3, 6, 9 and 12 months onto the extracted latent factors. Since we intend to examine medium- and long-term predictability, we sample our data on a monthly basis. The panel regression can be summarized, as follows:

$$r_{i,t,t+h}^{\text{simple}\,81} = \frac{S_{t+h} - S_t}{S_t} = \alpha_h + \beta_{v,h} \cdot V_{i,t} + \beta_{u,h} \cdot \text{JV}_{i,t}^{\text{neg},\perp V} + \epsilon_{i,t,t+h}, \quad \forall i \in \{1, ..., 254\}, \tag{5.9}$$

where $\text{JV}_t^{\text{neg},\perp V}$ denotes the orthogonal component of the risk-neutral expected negative jump variation for firm i, as defined in Eq. (5.8), relative to the spot variance V_t. Due to the overlapping observations of returns with horizons longer than one month, the widely used Newey-West standard errors are underestimated in the most extreme cases. Following Martin and Wagner (2019), we employ bootstrap standard errors as a conservative procedure, which accounts for both the time-series and the cross-sectional dependencies in the overlapping cross-sectional observations. Since computation of bootstrap standard errors requires considerable computational power, the procedure is carried out on the bwUniCluster 2.0 and bwForCluster Helix.

Table 5.11 reports the coefficients for the panel regressions of cross-sectional stock returns onto lagged volatilites and orthogonal negative jump variations extracted from option surfaces. All of the coefficients from the regression of stock returns on lagged volatility states are positive and increase with forecasting horizons. A t-statistic of 2.49 at the 9-month forecast horizon suggests relatively weak statistical validity for estimates of $\beta_{V,h}$. It's not until the 12-month horizon that the t-statistic increases to 2.74, surpassing the threshold of 2.5 and thereby emphasizing its predictive ability for returns. Furthermore, the coefficients associated with orthogonal negative jump variation are positive as well across all forecasting zones. Based on the bootstrap procedure, the t-statistics obtained are 3.14 for the 6-month forecasting horizon, 3.98 for the 9-month forecasting horizon, and 4.56 for the 12-month forecasting horizon, which exceed the critical threshold of 2.5.

[80] In their 3F model, there are two volatility factors and a negative jump intensity factor.

[81] To align with Martin and Wagner (2019) and others, simple returns are used in this study.

Both factors show substantial relevance in explaining the cross-sectional variation in U.S. equity premiums, but orthogonal negative jump variation shows greater significance in t-statistics than V_t.

Table 5.11: Cross-sectional panel predictability regressions of spot volatility and orthogonal negative jump variation.

	1-month	3-month	6-month	9-month	12-month
Intercept	0.0087	0.0291	0.0509	0.0777	0.1052
	(2.7355)	(3.8069)	(4.0752)	(4.0689)	(3.9331)
V_i	0.0332	0.0684	0.2005	0.2945	0.4055
	(0.9373)	(0.8214)	(1.7353)	(2.4855)	(2.7441)
$\mathrm{JV}_i^{\mathrm{neg},\perp V}$	0.0582	0.2104	0.5131	0.7269	0.8836
	(1.0127)	(1.7889)	(3.1370)	(3.9788)	(4.5606)
Adjusted R^2 (%)	0.6635	1.6453	5.0340	6.6092	7.7379

Note: This table presents panel regression results for the regressing of future returns on single stocks onto their respective lagged filtered state vector; i.e., $r_{i,t,t+h} = \alpha_h + \beta_{v,h} \cdot V_{i,t} + \beta_{u,h} \cdot \mathrm{JV}_{i,t}^{\mathrm{neg},\perp V} + \epsilon_{i,t,t+h}$. The returns are sampled on a monthly basis from January 2004 to December 2021.The column labels indicate the forecasting horizons, which range from 1 month to 1 year. The coefficients are listed in the table without parentheses, and the parentheses underneath represent the corresponding t-statistics, which are calculated using standard errors obtained through the block bootstrap procedure described in Martin and Wagner (2019). The row labelled "Adjusted R^2(%)" reports adjusted R^2 values, which are presented in percentage terms.

5.2.6.2 Conventional Surface-Related Risk Measures

We investigate whether various conventional surface-related risk measures can cross-sectionally predict future returns. If these measures can indeed forecast future returns, we seek to determine whether the information they carry overlaps with tail risk. The aim of this analysis is to verify whether the cross-sectional return predictability in the U.S. equity market–primarily driven by U_t as per our prior empirical findings–can be readily replicated, or if certain conventional surface-related risk factors can effectively span the factor U_t.

Adopting the methodology of Bakshi et al. (2003), we obtain the risk-neutral variance and skewness, denoted as Bakshi variance ($\mathrm{Var}_t^{\mathrm{Bakshi}}$) and Bakshi skewness ($\mathrm{Skew}_t^{\mathrm{Bakshi}}$) for each single stock. To explore if U_t can be spanned by the first two risk-neutral moments, we perform the following panel regression:

$$\mathrm{JV}_{i,t}^{\mathrm{neg}} = \gamma_0 + \gamma_{bv} \cdot \mathrm{Var}_{i,t}^{\mathrm{Bakshi}} + \gamma_{bs} \cdot \mathrm{Skew}_{i,t}^{\mathrm{Bakshi}} + \epsilon_{i,t}, \quad \forall i \in \{1, ..., 254\}. \tag{5.10}$$

In this regression, the negative jump variation $\mathrm{JV}_{i,t}^{\mathrm{neg}}$ is regressed onto the Bakshi variance and Bakshi skewness. The estimated intercept coefficient is 0.02 with a t-statistic of 2.45. The estimate of γ_{bv} (γ_{bs}) is 2.42 (0.00) with a t-statistic of 4.49 (-2.40). In addition, the adjusted R^2 is 37.99%. We conclude, therefore, that a certain portion of the negative jump variation can be explained by the Bakshi variance.

To investigate conventional surface-related risk measures embedded in implied volatility surfaces, we employ Black-Scholes implied volatilities of raw ATM options (ATM implied volatilities) with the nearest maturity to 30 days, as well as the implied volatility surface skewness (IV skew) measurement to perform an analysis similar to that in the previous subsection. In alignment with Andersen et al. (2015b), the IV skew ($\mathrm{Skew}_t^{\mathrm{IV}}$) is defined as the difference between the Black-Scholes implied volatilities of short-term OTM put and call options. We select the put options with deltas of -0.25 and the call options with deltas of $+0.25$, both with maturities similar to those of the ATM implied volatilities.We regress negative jump variation onto risk measures:

$$\mathrm{JV}_{i,t}^{\mathrm{neg}} = \gamma_0 + \gamma_{iv} \cdot \mathrm{IV}_{i,t}^{\mathrm{atm}} + \gamma_{skew} \cdot \mathrm{Skew}_{i,t}^{\mathrm{IV}} + \epsilon_{i,t}, \quad \forall i \in \{1, ..., 254\}. \tag{5.11}$$

In summary, we obtain the following results: the intercept yields an estimate of -0.07 with a t-statistic of -3.79. The estimated slope coefficient of $\mathrm{IV}_{i,t}^{\mathrm{atm}}$ ($\mathrm{Skew}_t^{\mathrm{IV}}$) is 0.41 (0.10) with a t-statistic of 5.22 (3.54). In this model, the adjusted R^2 is 50.96%, indicating that these two factors account for approximately half of the variation in $\mathrm{JV}_t^{\mathrm{neg}}$.

For further investigation into the explanatory power of future returns related to those conventional risk measures, the study repeats panel regressions similar to the predictive regression with forecasting horizons $h \in \{1, 3, 6, 9, 12\}$, as previously done:

$$r_{i,t,t+h} = \alpha_h + \beta_{bv,h} \cdot \mathrm{Var}_{i,t}^{\mathrm{Bakshi}} + \beta_{bs,h} \cdot \mathrm{Skew}_{i,t}^{\mathrm{Bakshi}} + \epsilon_{i,t,t+h}, \quad \forall i \in \{1, ..., 254\}; \tag{5.12}$$

$$r_{i,t,t+h} = \alpha_h + \beta_{iv,h} \cdot \mathrm{IV}_{i,t}^{\mathrm{atm}} + \beta_{skew,h} \cdot \mathrm{Skew}_{i,t}^{\mathrm{IV}} + \epsilon_{i,t,t+h}, \quad \forall i \in \{1, ..., 254\}. \tag{5.13}$$

Based on the results presented in Table 5.13, the $\mathrm{IV}_t^{\mathrm{atm}}$ demonstrates its ability to predict cross-sectional returns. Specifically, for forecasting horizons of 9 and 12 months, the t-statistics for the $\beta_{iv,h}$ estimates are 2.65 and 2.96, respectively, indicating a significant level of predictability. Conversely, the $\beta_{skew,h}$ estimates, which are nearly zero, imply little to no predictive power for the $\mathrm{Skew}_t^{\mathrm{IV}}$. Additionally, in the predictive regression analysis, the Bakshi moments do not exhibit statistical significance at either horizon.

Table 5.12: Panel regression of alternative measures on negative jump variation.

	$JV_{i,t}^{neg}$	
Intercept	0.0156	-0.0749
	(2.4537)	(-3.7906)
Var_i^{Bakshi}	2.4159	
	(4.4854)	
$Skew_i^{Bakshi}$	-0.0050	
	(-2.4026)	
IV_i^{atm}		0.4142
		(5.2190)
$Skew_i^{IV}$		0.1041
		(3.5399)
Adjusted R^2 (%)	37.9919	50.9589

Note: This table presents the results of panel regression analyses that regress the NJV on two sets of alternative measures. The first set consists of Bakshi moments, which include Bakshi variance (Var_i^{Bakshi}) and Bakshi skewness ($Skew_i^{Bakshi}$). The second set is referred to as the practitioner's measures and contains ATM BSIV (IV_i^{atm}) and IV Skew ($Skew_i^{IV}$). The coefficients are listed in the table without parentheses, and the parentheses underneath represent the corresponding t-statistics, which are calculated using standard errors obtained through the block bootstrap procedure described in Martin and Wagner (2019). The row labelled "Adjusted R^2(%)" reports adjusted R^2 values, which are presented in percentage terms.

Table 5.13: Cross-sectional panel predictability regressions of ATM implied volatility and IV skew.

	1-month	3-month	6-month	9-month	12-month
Intercept	-0.0010	-0.0014	-0.0199	-0.0212	-0.0264
	(-0.0948)	(-0.0584)	(-0.5495)	(-0.5361)	(-0.5552)
IV_i^{atm}	0.0481	0.1412	0.3352	0.4697	0.6219
	(1.1494)	(1.5018)	(2.1340)	(2.6472)	(2.9624)
$Skew_i^{IV}$	-0.0225	-0.0455	-0.0256	0.0087	0.0311
	(-1.0242)	(-0.9965)	(-0.3338)	(0.0911)	(0.2567)
Adjusted R^2 (%)	0.5926	1.7810	4.5487	5.7462	7.0327

Note: This table presents panel regression results for the regressing of future returns on single stocks onto their ATM implied volatility and IV skew; i.e., $r_{i,t,t+h} = \alpha_h + \beta_{iv,h} \cdot IV_{i,t}^{atm} + \beta_{skew,h} \cdot Skew_{i,t}^{IV} + \epsilon_{i,t,t+h}$. The returns are sampled on a monthly basis from January 2004 to December 2021. The column labels indicate the forecasting horizons, which range from 1 month to 1 year. The coefficients are listed in the table without parentheses, and the parentheses underneath represent the corresponding t-statistics, which are calculated using standard errors obtained through the block bootstrap procedure described in Martin and Wagner (2019). The row labelled "Adjusted R^2(%)" reports adjusted R^2 values, which are presented in percentage terms.

The last two analyses reveal the measure: IV_t^{atm} and $\text{Skew}_t^{\text{IV}}$, which exhibit substantial explanatory power for JV_t^{neg}. Notably, among these two dependent variables, IV_t^{atm} have the greatest predictive power for future cross-sectional returns and may overlap with JV_t^{neg}. While the return predicting ability of IV_t^{atm} is not surprising, given its role as part of the input data in our estimation procedure, it is important to understand whether the realization of the tail risk factor can once again play a substantial role in the regression. To address this issue, we perform three separate groups of panel regression analyses, each based on a different set of dependent variables.

First, we incorporate JV_t^{neg} into the panel regressions with forecasting horizons $h \in \{1, 3, 6, 9, 12\}$ months as an additional factor, along with $\text{IV}_{i,t}^{\text{atm}}$ and $\text{Skew}_t^{\text{IV}}$, yielding the following equation

$$\begin{aligned} r_{i,t,t+h} = \alpha_h + \beta_{iv,h} \cdot \text{IV}_{i,t}^{\text{atm}} + \beta_{skew,h} \cdot \text{Skew}_{i,t}^{\text{IV}} + \beta_{U,h} \cdot \text{JV}_{i,t}^{\text{neg}} \\ + \epsilon_{i,t,t+h}, \quad \forall i \in \{1, ..., 254\}. \end{aligned} \tag{5.14}$$

Table 5.14 presents the results, indicating that the estimates of $\beta_{iv,h}$ is no longer statistically significant. Meanwhile, the estimates of $\beta_{U,h}$ begins to show significance starting from a 9-month forecasting horizon. Furthermore, the estimates of $\beta_{skew,h}$ remains consistently insignificant and close to zero. However, due to the high correlation and potential co-linearity between $\text{IV}_{i,t}^{\text{atm}}$ and JV_t^{neg}, we cannot conclusively state that JV_t^{neg} is the only driver of the predictability among these three factors. To further investigate the relationship between $\text{IV}_{i,t}^{\text{atm}}$ and JV_t^{neg}, we continue to perform two additional sets of regression analyses.

In the second regression analysis, we orthogonalize a stock's JV_t^{neg} with regard to its IV_t^{atm} to attach all common movements to the $\text{IV}_{i,t}^{\text{atm}}$. The panel regressions with forecasting horizons $h \in \{1, 3, 6, 9, 12\}$ months take the form

$$r_{i,t,t+h} = \alpha_h + \beta_{iv,h} \cdot \text{IV}_{i,t}^{\text{atm}} + \beta_{U^{\perp iv},h} \cdot \text{JV}_{i,t}^{\text{neg},\perp \text{IV}} + \epsilon_{i,t,t+h}, \quad \forall i \in \{1, ..., 254\}. \tag{5.15}$$

Table 5.15 summarizes the results. Here, $\text{JV}_{i,t}^{\text{neg},\perp \text{IV}}$ represents the orthogonal component with respect to IV_t^{atm}. The results show that the estimates of $\beta_{U^{\perp iv},h}$ become significant starting from a 6-month forecasting period, whereas the estimates of $\beta_{iv,h}$ do not.

In the third regression analysis, we perform the orthogonalization in the other direction; i.e., we orthogonalize a stock's IV_t^{atm} with regard to its JV_t^{neg} to attach all common movements to the JV_t^{neg}. The panel regressions with forecasting horizons $h \in \{1, 3, 6, 9, 12\}$

Table 5.14: Cross-sectional return predictability: panel regression of ATM implied volatility, IV skew and negative jump variation.

	1-month	3-month	6-month	9-month	12-month
Intercept	0.0025	0.0077	0.0089	0.0215	0.0232
	(0.2664)	(0.4351)	(0.3026)	(0.6379)	(0.5495)
IV_i^{atm}	0.0293	0.0924	0.1779	0.2366	0.3500
	(0.7379)	(1.3332)	(1.3645)	(1.4361)	(1.6581)
$\text{Skew}_i^{\text{IV}}$	-0.0266	-0.0565	-0.0619	-0.0446	-0.0315
	(-1.1609)	(-1.3288)	(-0.7737)	(-0.4318)	(-0.2438)
JV_i^{neg}	0.0421	0.1095	0.3557	0.5247	0.6140
	(0.6779)	(0.8563)	(2.2856)	(3.1750)	(2.9760)
Adjusted R^2 (%)	0.6726	1.9730	5.4882	7.0129	8.2481

Note: This table presents panel regression results for the regressing of future returns on single stocks onto their ATM implied volatility, IV skew and negative jump variation; i.e., $r_{i,t,t+h} = \alpha_h + \beta_{iv,h} \cdot \text{IV}_{i,t}^{\text{atm}} + \beta_{skew,h} \cdot \text{Skew}_{i,t}^{\text{IV}} + \beta_{U,h} \cdot \text{JV}_{i,t}^{\text{neg}} + \epsilon_{i,t,t+h}$. The returns are sampled on a monthly basis from January 2004 to December 2021. The column labels indicate the forecasting horizons, which range from 1 month to 1 year. The coefficients are listed in the table without parentheses, and the parentheses underneath represent the corresponding t-statistics, which are calculated using standard errors obtained through the block bootstrap procedure described in Martin and Wagner (2019). The row labelled "Adjusted R^2(%)" reports adjusted R^2 values, which are presented in percentage terms.

Table 5.15: Cross-sectional return predictability: panel regression of ATM implied volatility and negative jump variation orthogonal to ATM implied volatility.

	1-month	3-month	6-month	9-month	12-month
Intercept	0.0026	0.0081	0.0086	0.0230	0.0261
	(0.2733)	(0.4478)	(0.2855)	(0.6888)	(0.6260)
IV_i^{atm}	0.0367	0.1122	0.2568	0.3517	0.4834
	(0.9214)	(1.3608)	(1.8098)	(2.1342)	(2.3713)
$\text{JV}_i^{\text{neg},\perp\text{IV}}$	0.0530	0.1434	0.4547	0.7078	0.8489
	(0.7318)	(0.9409)	(2.6585)	(3.7840)	(3.6182)
Adjusted R^2 (%)	0.6703	2.0097	5.7601	7.5820	8.8737

Note: This table presents panel regression results for the regressing of future returns on single stocks onto their ATM implied volatility and negative jump variation orthogonal to ATM implied volatility; i.e., $r_{i,t,t+h} = \alpha_h + \beta_{iv,h} \cdot \text{IV}_{i,t}^{\text{atm}} + \beta_{U^{\perp iv},h} \cdot \text{JV}_{i,t}^{\text{neg},\perp\text{IV}} + \epsilon_{i,t,t+h}$. The returns are sampled on a monthly basis from January 2004 to December 2021. The column labels indicate the forecasting horizons, which range from 1 month to 1 year. The coefficients are listed in the table without parentheses, and the parentheses underneath represent the corresponding t-statistics, which are calculated using standard errors obtained through the block bootstrap procedure described in Martin and Wagner (2019). The row labelled "Adjusted R^2(%)" reports adjusted R^2 values, which are presented in percentage terms.

months take the form

$$r_{i,t,t+h} = \alpha_h + \beta_{iv^{\perp U},h} \cdot \text{IV}_{i,t}^{\text{atm},\perp\text{JV}^{\text{neg}}} + \beta_{U,h} \cdot \text{JV}_{i,t}^{\text{neg}} + \epsilon_{i,t,t+h}, \quad \forall i \in \{1, ..., 254\}. \tag{5.16}$$

Table 5.16 presents the summarized results. The results clearly indicate that the estimates of $\beta_{iv^{\perp U},h}$ are close to zero and their t-statistics fall well below the conservative threshold of 2.5. However, the estimates of $\beta_{U,h}$ increase over the forecasting horizons, and their t-statistics surpass this threshold for horizons of 9 months and longer.

Table 5.16: Cross-sectional return predictability: panel regression of ATM implied volatility orthogonal to negative jump variation and negative jump variation.

	1-month	3-month	6-month	9-month	12-month
Intercept	0.0087	0.0260	0.0552	0.0892	0.1143
	(2.0546)	(4.0292)	(5.2508)	(6.4815)	(5.8649)
$\mathrm{IV}_i^{\mathrm{atm},\perp\mathrm{JV}^{\mathrm{neg}}}$	0.0015	0.0134	-0.0368	-0.0824	-0.0650
	(0.0538)	(0.3455)	(-0.6133)	(-0.9373)	(-0.5540)
$\mathrm{JV}_i^{\mathrm{neg}}$	0.0728	0.2179	0.5250	0.7311	0.9683
	(1.0575)	(1.3695)	(2.1000)	(2.7551)	(3.0649)
Adjusted R^2 (%)	0.5384	1.5762	4.9070	6.4497	7.2700

Note: This table presents panel regression results for the regressing of future returns on single stocks onto their ATM implied volatility orthogonal to negative jump variation and negative jump variation; i.e., $r_{i,t,t+h} = \alpha_h + \beta_{iv^{\perp U},h} \cdot \mathrm{IV}_{i,t}^{\mathrm{atm},\perp\mathrm{JV}^{\mathrm{neg}}} + \beta_{U,h} \cdot \mathrm{JV}_{i,t}^{\mathrm{neg}} + \epsilon_{i,t,t+h}$. The returns are sampled on a monthly basis from January 2004 to December 2021. The column labels indicate the forecasting horizons, which range from 1 month to 1 year. The coefficients are listed in the table without parentheses, and the parentheses underneath represent the corresponding t-statistics, which are calculated using standard errors obtained through the block bootstrap procedure described in Martin and Wagner (2019). The row labelled "Adjusted R^2(%)" reports adjusted R^2 values, which are presented in percentage terms.

Overall, the results indicate that ATM implied volatility loses its predictive power for future returns when its co-movement with tail risk is removed. Therefore, the independent information carried by tail risk plays a more significant role than ATM implied volatility. Based on these findings, we can conclude that $\mathrm{JV}_t^{\mathrm{neg}}$ exhibits significant explanatory power over ATM implied volatility, and the information provided by ATM implied volatility is inherently incorporated within $\mathrm{JV}_t^{\mathrm{neg}}$.

5.2.6.3 Systematic vs. Idiosyncratic Components in Risk Factors

In order to understand if either the systematic or idiosyncratic components of V_t and U_t carry the information for the return predictability found in panel regressions in Section 5.2.6.1. Following regressions of the form in Eq. (5.9), we consider the dependent variables in V_t and $\mathrm{JV}_t^{\mathrm{neg},\perp V}$ and decompose them into two components: a systematic component and an idiosyncratic component by projecting each firm's realizations of V_t and $\mathrm{JV}_t^{\mathrm{neg},\perp V}$ onto the corresponding realizations of the market index, respectively. Finally,

we perform predictive panel regressions using the systematic and idiosyncratic components of the single stock's realizations of V_t and $\text{JV}_t^{\text{neg},\perp V}$. The panel regressions with forecasting horizons $h \in \{1, 3, 6, 9, 12\}$ months take the form

$$\begin{aligned} r_{i,t,t+h} = \alpha_h + \beta_{V,h}^{\text{sys}} \cdot V_{i,t}^{\text{sys}} + \beta_{V,h}^{\text{idio}} \cdot V_{i,t}^{\text{idio}} + \beta_{U,h}^{\text{sys}} \cdot \text{JV}_{i,t}^{\text{neg},\perp V,\text{sys}} \\ + \beta_{U,h}^{\text{idio}} \cdot \text{JV}_{i,t}^{\text{neg},\perp V,\text{idio}} + \epsilon_{i,t,t+h}, \quad \forall i \in \{1, ..., 254\}, \end{aligned} \tag{5.17}$$

where the $V_{i,t}^{\text{sys}}$ ($\text{JV}_{i,t}^{\text{neg},\perp V,\text{sys}}$) denotes the systematic component of the corresponding risk factor $V_{i,t}$ ($\text{JV}_{i,t}^{\text{neg},\perp V}$) and the $V_{i,t}^{\text{idio}}$ ($\text{JV}^{\text{neg},\perp V,\text{idio}}$) denotes the idiosyncratic component of $V_{i,t}$ ($\text{JV}_{i,t}^{\text{neg},\perp V}$).

Table 5.17 presents the regression results. We observe that the estimates of $\beta_{V,h}^{\text{idio}}$ for the idiosyncratic component of spot variance consistently increase. The t-statistics of $\beta_{V,h}^{\text{idio}}$ surpass the threshold of 2.5, starting with a 6-month forecast horizon. Therefore, only the information carried by the idiosyncratic components of V_t can predict cross-sectional future returns. In addition, the estimates of $\beta_{U,h}^{\text{idio}}$ for the idiosyncratic component are statistically significant at forecasting horizons of 6 and 9 months.

5.2.7 Return Predictability: Portfolio Sorts

By constructing the portfolio according to An et al. (2014)'s long-short investment strategy, we examine the predictive power of $\text{JV}^{\text{neg},\perp V}$.[82] We report the results in Table 5.18. The average raw and risk-adjusted return differences between High $\text{JV}^{\text{neg},\perp V}$ and Low $\text{JV}^{\text{neg},\perp V}$ portfolios are statistically significant for 6-, 9- and 12-month holding periods. There are statistically significant annualized average returns of 5.43%, 5.74%, and 5.39% for holding periods of 6 months, 9 months, and 12 months, respectively. In summary, $\text{JV}^{\text{neg},\perp V}$ predictability persists for 6, 9 and 12 months, which is in line with the findings in the previous panel regressions.

Furthermore, we examine whether the portfolio's sorting results for systematic and id-

[82]Here we mainly the describe their strategy taken from Jegadeesh and Titman (1993). This involves sorting the extracted orthogonal negative jump variation in the current month t into quintiles. The portfolio is designed with overlapping holding periods h. The strategy holds portfolios selected in the current month and the previous $h-1$ months. In this case, h represents the holding period of 1, 3, 6, 9, and 12 months. At the beginning of every month t, we perform dependent sorts on $\text{JV}^{\text{neg},\perp V}$ over the past month $t-1$. As a result of these rankings, five portfolios are created for $\text{JV}^{\text{neg},\perp V}$. Every month t, the strategy buys stocks in the High $\text{JV}^{\text{neg},\perp V}$ quintile, and sells stocks in the Low $\text{JV}^{\text{neg},\perp V}$ quintile, holding the position for h months. In addition, the strategy closes out the position initiated in month $t-h$. Therefore, under this trading strategy, we revise the weights on $1/h$ of all stocks in our portfolio on a monthly basis. We carry over the rest from the previous month. In this strategy, profits are calculated as a result of a series of portfolios that are rebalanced on a monthly basis to maintain equal weighting.

Table 5.17: Cross-sectional panel predictability regressions of idiosyncratic spot variance, systematic spot variance, idiosyncratic orthogonal jump variation and systematic orthogonal jump variation.

	1-month	3-month	6-month	9-month	12-month
Intercept	0.0064	0.0232	0.0421	0.0692	0.1002
	(1.9305)	(3.1197)	(3.1039)	(3.2998)	(3.4840)
V_i^{sys}	0.0213	0.0606	0.2190	0.2932	0.3902
	(0.4521)	(0.6086)	(1.6020)	(1.7478)	(1.8572)
V_i^{idio}	0.0664	0.1165	0.2231	0.3471	0.4598
	(2.3366)	(2.4150)	(2.6029)	(4.0128)	(4.3062)
$\text{JV}_i^{\text{neg},\perp V,\text{sys}}$	0.2991	0.7729	1.2738	1.5110	1.3817
	(1.8798)	(3.1548)	(3.0492)	(2.7222)	(2.1987)
$\text{JV}_i^{\text{neg},\perp V,\text{idio}}$	0.0418	0.1455	0.3864	0.6236	0.8374
	(0.8174)	(1.4233)	(2.3500)	(3.4117)	(3.8701)
Adjusted R^2 (%)	1.6811	3.2620	6.2345	7.4428	7.9925

Note: This table presents panel regression results for the regressing of future returns on single stocks onto their idiosyncratic spot variance, systematic spot variance, idiosyncratic orthogonal jump variation and systematic orthogonal jump variation; i.e., $r_{i,t,t+h} = \alpha_h + \beta_{V,h}^{\text{idio}} \cdot V_{i,t}^{\text{idio}} + \beta_{V,h}^{\text{sys}} \cdot V_{i,t}^{\text{sys}} + \beta_{U,h}^{\text{idio}} \cdot \text{JV}_{i,t}^{\text{neg},\perp V,\text{idio}} + \beta_{U,h}^{\text{sys}} \cdot \text{JV}_{i,t}^{\text{neg},\perp V,\text{sys}} + \epsilon_{i,t,t+h}$. The returns are sampled on a monthly basis from January 2004 to December 2021. The column labels indicate the forecasting horizons, which range from 1 month to 1 year. The coefficients are listed in the table without parentheses, and the parentheses underneath represent the corresponding t-statistics, which are calculated using standard errors obtained through the block bootstrap procedure described in Martin and Wagner (2019). The row labelled "Adjusted R^2(%)" reports adjusted R^2 values, which are presented in percentage terms.

iosyncratic components of spot volatility and tail risk align with the predictive regressions. Table 5.19 shows that portfolio sorts based on idiosyncratic variance yield statistically significant annualized average returns. These returns are significant, starting from a holding period of one month, and range between 7.76% and 11.18%. Additionally, this strategy consistently yields a Sharpe ratio of around 1.0 across all forecasting periods. Conversely, portfolio sorts based on idiosyncratic left tail risk exhibit annualized average returns of 4.37% (t-statistic: 2.34), 4.76% (t-statistic: 2.51), and 4.88% (t-statistic: 2.54) for holding periods of 6, 9, and 12 months, respectively. The Sharpe ratios for these holding periods are approximately 0.5. However, portfolio sorts focusing on the systematic components of both risk factors do not yield statistically significant average returns.

Table 5.18: Portfolio sorts on spot volatility and orthogonal negative jump variation.

	1-month	3-month	6-month	9-month	12-month
V_i					
Average Return	0.0774	0.0749	0.0838	0.0832	0.0830
	(3.0114)	(2.8147)	(2.7946)	(2.8489)	(3.0594)
FF5 Alpha	0.0922	0.0834	0.0847	0.0820	0.0839
	(3.1286)	(2.8290)	(2.6579)	(2.7148)	(3.0936)
Sharpe Ratio	0.4781	0.5083	0.5136	0.5241	0.5477
$\mathrm{JV}_i^{\mathrm{neg},\perp V}$					
Average Return	0.0439	0.0598	0.0543	0.0574	0.0539
	(1.9114)	(2.4832)	(2.8460)	(3.1139)	(3.0480)
FF5 Alpha	0.0497	0.0601	0.0537	0.0570	0.0541
	(2.1842)	(2.4487)	(2.8361)	(3.2503)	(3.2007)
Sharpe Ratio	0.3195	0.5609	0.5805	0.6939	0.6468

Note: The table showcases the annualized average returns of a long-short strategy, along with the annualized Sharpe ratios for factors V_i and $\mathrm{JV}_i^{\mathrm{neg},\perp V}$. This strategy involves going long on a quintile portfolio with the highest corresponding variable value, while going short on a portfolio with the lowest. The portfolios are maintained over periods of 1, 3, 6, 9, and 12 months, with only monthly rebalancing. The table includes Newey and West (1987) t-statistics, with the number of lags set equal to $1.3\sqrt{T}$, and these statistics are reported in parentheses. Additionally, the table provides the annualized Fama-French 5-factor (see Fama and French (2015)) alpha, denoted as "FF5 Alpha". The Fama-French 5-factors are obtained from Kenneth R.French–Data Library.

Table 5.19: Portfolio sorts systamatic components and idiosyncratic components.

	1-month	3-month	6-month	9-month	12-month
V_i^{sys}					
Average Return	0.0368	0.0385	0.0463	0.0478	0.0464
	(0.9148)	(1.0200)	(1.2113)	(1.2870)	(1.2832)
FF5 Alpha	0.0337	0.0394	0.0445	0.0435	0.0440
	(0.9022)	(1.0242)	(1.1283)	(1.1605)	(1.2366)
Sharpe Ratio	0.1533	0.1964	0.2182	0.2342	0.2241
V_i^{idio}					
Average Return	0.1118	0.0776	0.0829	0.0848	0.0800
	(3.4289)	(4.3864)	(4.9600)	(5.0238)	(5.0885)
FF5 Alpha	0.1304	0.0880	0.0860	0.0876	0.0835
	(4.0717)	(4.7496)	(5.1075)	(5.2741)	(5.5258)
Sharpe Ratio	0.9220	0.8273	1.0353	1.0780	1.0502
$\text{JV}_i^{\text{neg},\perp V,\text{sys}}$					
Average Return	0.0443	0.0683	0.0578	0.0395	0.0285
	(1.1536)	(1.3841)	(1.5868)	(1.4583)	(1.2700)
FF5 Alpha	0.0655	0.0739	0.0546	0.0360	0.0277
	(1.2846)	(1.4252)	(1.5432)	(1.3925)	(1.3055)
Sharpe Ratio	0.2042	0.4154	0.4107	0.3301	0.2566
$\text{JV}_i^{\text{neg},\perp V,\text{idio}}$					
Average Return	0.0268	0.0466	0.0437	0.0476	0.0488
	(0.9161)	(2.0084)	(2.3446)	(2.5140)	(2.5416)
FF5 Alpha	0.0190	0.0477	0.0450	0.0490	0.0497
	(0.6178)	(2.0854)	(2.4351)	(2.6383)	(2.6486)
Sharpe Ratio	0.1565	0.4941	0.5242	0.5764	0.5509

Note: This table presents the annualized average returns of a long-short strategy, along with its corresponding annualized Sharpe ratios for factors of V_i^{sys}, V_i^{idio}, $\text{JV}_i^{\text{neg},\perp V,\text{sys}}$ and $\text{JV}_i^{\text{neg},\perp V,\text{idio}}$. This strategy involves going long on a quintile portfolio with the highest corresponding variable value, while going short on a portfolio with the lowest. Portfolios are held for 1, 3, 6, 9, and 12 month(s), and are rebalanced monthly. The table includes Newey and West (1987) t-statistics, with the number of lags set equal to $1.3\sqrt{T}$, and these statistics are reported in parentheses. Additionally, the table provides the annualized Fama-French 5-factor (see Fama and French (2015)) alpha, denoted as "FF5 Alpha". The Fama-French 5-factors are obtained from Kenneth R.French–Data Library.

5.2.8 Summary

By combining the strengths of several existing estimation methodologies, we apply an affine jump-diffusion model to the cross-section of option surfaces in the U.S. market. According to the literature, the model's estimated static parameters and latent states are in line with the empirical characteristics of option-implied risk-neutral density. Furthermore, the calibration model's option pricing performance assessment confirms not only the appropriateness of the model's implied risk-neutral density, but also the accuracy of the extracted trajectories of spot volatilities and negative jump intensities. When assessing option pricing, the calibrated models yield an overall in-sample root-mean-square error of 6.81% and a relative root-mean-square error of 10.76%.

In addition to demonstrating the model's fit, we also establish that the cross-section of left tail risk factors, represented by the model's negative jump intensity, can predict future stock returns over periods spanning from 6 to 12 months. Panel regression analysis yields statistically significant slope coefficients of 0.51, 0.73, and 0.88 for forecasting horizons of 6, 9, and 12 months, respectively, with corresponding bootstrapped t-statistics of 3.14, 3.98, and 4.56. Furthermore, due to the additional predictive regressions on several conventional surface-related factors, we demonstrate that left tail risk cannot be spanned by those conventional risk measures. Additionally, we decompose both spot variance and tail risk into systematic and idiosyncratic components. Based on the predictive panel regressions, we find that the idiosyncratic components exhibit predictive power across various forecasting horizons.

Finally, we underscore our findings with further evidence based on portfolio sorts involving various factors used in the aforementioned predictive regressions. Specifically, long-short portfolios constructed by sorting the orthogonal left tail risk factor yield statistically significant annualized returns of 5.43%, 5.74%, and 5.39% over holding periods of 6, 9, and 12 months, respectively. The annualized Sharpe ratios of the long-short strategy, based on the left tail risk for these periods, are 0.58, 0.69, and 0.65, respectively. Moreover, long-short portfolios sorted by the idiosyncratic component of spot variance consistently yield significant annualized average returns across all forecasting periods. Meanwhile, portfolios sorted based on the idiosyncratic component of the tail risk factor start to demonstrate significant annualized average returns for holding periods beginning at 6 months. Therefore, these findings indicate that the idiosyncratic components of both variance and left tail risks possess predictive power for equity returns.

6 Conclusion and Outlook

This thesis provides an overview of affine jump-diffusions and their applications to reduced-form dynamic asset pricing models, which are used in the prolific literature on extracting economic information from derivatives markets. Following a review of related estimation methodologies, this thesis aims to provide a robust and efficient approach to estimating static parameters and recovering state variables for AJD models in a no-arbitrage setting by using moments derived from rich observations from the option market for the underlying return density. Empirical studies of the AJD models with an emphasis on a left tail factor are conducted on options written on the S&P 500 index and its constituents. Based on the results of the empirical studies, the conclusion regarding the future return predictability of the left tail factor, which is shown at the index level in the literature, is extended to the cross-sectional level.

In a first step, Chapter 2 reviews the mathematical fundamentals of AJDs and their application to dynamic asset pricing models. In particular, the general definition of AJDs is discussed in light of the seminal work of Duffie et al. (2000). Following their results, a general transformation analysis for AJDs is presented, which is essential for the option pricing methods discussed in this section. The derivation of the AJD-implied moments, which play an instrumental role in the estimation approach, is explained in detail. Additionally, inverse Fourier techniques are discussed in detail for derivative pricing, particularly for the cosine method by Fang and Oosterlee (2008). Moreover, this chapter introduces several essential dynamic asset pricing models based on the arbitrage-free framework, building up in complexity. To conclude the chapter, several issues related to the empirical analysis of the option market are addressed.

Chapter 3 reviews the general settings for the inference procedure for AJD models. State-space models, as a standard tool for estimating discrete-time models with probability relationships between observations and latent states are first presented in this chapter. Based on the *Euler-Maruyama scheme*, the continuous-time AJD models can be trans-

lated into a discrete-time state-space model. With a state-space modeling framework in combination with the filtering methods presented in this chapter, a uniform inference framework is established. The pros and cons of appropriate selection of filtering methods depending on the concrete dynamics of AJD models are also discussed. Moreover, a penalized least squares approach proposed by Andersen et al. (2015a) is explained in addition to the state-space model-based approach. Finally, the aforementioned techniques are combined to form a quasi-maximum likelihood approach based on the Kalman filter.

In Chapter 4, the Kalman filter-based quasi-maximum likelihood approach for the SVNUJ model is validated in two steps via Monte Carlo simulation. The filtering step in the approach is validated via a study assuming the static parameters are known. The validation of the estimation for static parameters is carried out afterwards, since the recovery of latent states is nested in the likelihood function calculation.

The empirical analysis of the SVNUJ model for the U.S. option market is performed in Chapter 5. In order to reproduce the results of Andersen et al. (2015b), an empirical analysis of the options written on the S&P 500 index is conducted. The results from the Kalman filter-based quasi-maximum likelihood approach are consistent with the results of Andersen et al. (2015b), namely that the extracted negative jump intensity as a left tail factor is predictive of future index returns. This chapter additionally extends the results to a cross-sectional level by applying the SVNUJ model to options written on single stocks. Empirical findings from panel regressions affirm the cross-sectional return predictability of the left tail factor. Moreover, by decomposing the spot variance and left tail factors into systematic and idiosyncratic components, and conducting subsequent panel regressions and portfolio sorting analysis on these components, the predictive power of idiosyncratic risk components for cross-sectional returns is established.

Despite the length of the thesis, there are still some issues that remain unattended, particularly those related to the estimation of static parameters under $\mathbb{P}$. As shown in the simulation study in Chapter 4, the parameters under $\mathbb{P}$ cannot be well identified. The underlying reason could be that measurements of cumulants dominates the filtering step, making the dynamics about the evolution of latent states irrelevant. Therefore, one will need some amelioration of the measurements for the dynamics under $\mathbb{P}$. Numerous studies in the literature (see, e.g., Fulop and Li (2019), among others) use return data to pin down the dynamics under $\mathbb{P}$. Mechanically there are two potential issues preventing an accurate and efficient estimation of the parameters under $\mathbb{P}$, when the underlying returns are included. First, in the AJD models, if there are jumps in the return dynamics,

the conditional density of the return is then nonlinear and non-Gaussian. Although technically possible, approximating measurement equations may have a larger bias than approximating state equations. Since measurement equations provide direct insight into state variables, approximation bias could easily be transferred into the recovery of state variables. Second, the return data itself is in general noisy, which means that the higher order risk-neutral moments can still dominate the filtering step. Unfortunately, this potential inefficiency regarding estimating the parameters under $\mathbb{P}$ is often ignored in the literature. Due to the fact that most empirical studies do not include a simulation study before the empirical analysis is conducted. To my current knowledge, only Fulop and Li (2019) conduct a simulation study before proceeding to an empirical study. It appears, however, that the $\mathbb{P}$ parameters are not well identified in both the numerical and simulation studies. One potential amelioration of the measurements would be to use a nonparametric estimate of the continuous return variation[83] derived from the high frequency return data. Since this measure represents the diffusive variation of the return process, it provides direct information about the dynamics of volatility in the AJD models under $\mathbb{P}$. A comprehensive solution to the identification problem under $\mathbb{P}$ would require additional data and the use of particle filtering. Furthermore, there is a potentially statistical pitfall in the assumption of the model, which is sourced from the literature, i.e., a structure-wise misspecification of the model dynamics under $\mathbb{P}$. As outlined by Andersen et al. (2015b), the assumption that model dynamics under $\mathbb{P}$ and $\mathbb{Q}$ exhibit the same structure may not hold at all for real data. As a result, the aforementioned solutions may not resolve the issue either. Therefore, in such a case more advanced econometric tools are required in order to build the wedge between $\mathbb{P}$ and $\mathbb{Q}$ for risk premiums.

Building on the findings presented in this thesis, two data sets emerge as particularly interesting for further extracting useful information embedded in the option market. Firstly, there is scant research on AJD models focusing on ultra short-term options. To my current knowledge, only Andersen et al. (2017) employ an AJD model to characterize the risks implied by weekly options. By using the estimation procedure in this thesis, one can begin with the weekly options on the index and then proceed to the weekly options written on the constituents of the index. Mechanically, this study may reveal whether risk-neutral moment information can improve efficiency in the task of estimating AJD models for short-term options. Empirically, this study can be used to examine whether there exists any additional information that is not captured in models estimated based

[83] See Andersen et al. (2015b) and Andersen et al. (2019) for more details.

on medium- and long-term options. Secondly, as demonstrated by Bates (2019), a set of AJD models can capture the statistical features of intraday fluctuations in S&P 500 future prices. Therefore, given the increased liquidity in the option market, incorporating intraday option surfaces into an empirical study of AJD models presents an interesting and promising research avenue. With this study, it would be possible to examine whether intraday option surfaces can improve the empirical fit of AJD models to the intraday dynamics of underlying returns, since intraday option surfaces should be able to provide information that cannot be obtained from end-of-day surfaces or intraday future prices. Furthermore, this study would offer an opportunity to investigate whether economic news-driven intraday market movements can be explained using state variables extracted from intraday option surfaces.

A Appendix

A.1 Additional Definitions & Propositions

A.1.1 Additional Definitions

Definition A.1. (Affine jump-diffusions (AJD))[84]
Let $(\Omega, \mathcal{F}, P)$ a probability space on some state space $D \subseteq \mathbb{R}^n$. Consider the SDE

$$dX_t = \mu(X_t)\, dt + \sigma(X_t)\, dW_t + dJ_t\,, \tag{A.1}$$

where the building blocks are assumed to satisfy that

- $\mu : D \to \mathbb{R}^n$,
- $\sigma = (\sigma_{i,j})_{i,j=1,\ldots,n} : D \to \mathbb{R}^{n\times n}$,
- W_t is a Brownian motion in $\mathbb{R}^n$, and
- J_t is supposed to be a right-continuous jump process with ς_ν being a jump transform that determines the jump distribution on $\mathbb{R}^n$ with fixed parameters.
  - The jumps arrive with time-varying intensity $\{c(X_t) : t \geq 0\}$ for some mapping $c : D \to [0, \infty)$.
  - Let $\varsigma_\nu = \int_{\mathbb{R}^n} \exp(u \cdot z) d\nu(z)$ be the mapping : $\mathbb{C}^n \to \mathbb{C}$ for $u \in \mathbb{C}^n$, which determines the distribution of jump sizes.

  In this case, the jump can be chosen as a *compound Poisson process* for J_t.

The timing of jump activity happening, which is the counting process of whether a jump occurs or not, is assumed to be independent of jump size in J_t. Furthermore, the sudden

[84]This formulation of definition is based on the results of the research project conducted jointly with Maximilian Kübler, a student at the C-RAM chair. I would like to extend my thanks to him.

shocks shall not be affected by "mild" shocks from Brownian motion W_t. Thus, it is natural to assume the *mutual independence* of J_t and W_t.

Nevertheless, the critical functional relationship for the term *affine* is still missing. To clarify the affinity, consider the dynamics defined in Eq. (2.1), where the stochastic processes are affine if and only if $\forall x \in D$:

- $\mu(x) = k_0 + k_1 \cdot x_1 + \cdots + k_n \cdot x_n = k_0 + K_1 x$ (Drift)
- $\sigma(x)\sigma(x)^\top = h_0 + h_1 \cdot x_1 + \cdots + h_n \cdot x_n = h_0 + H_1 x$ (Variance)
- $c(x) = c_0 + c_1 \cdot x_1 + \cdots + c_n \cdot x_n = c_0 + C_1 \cdot x$ (Poisson counter intensity)
- $r(x) = r_0 + r_1 \cdot x_1 + \cdots + r_n \cdot x_n = r_0 + R_1 \cdot x$ (Instantaneous riskless short-term rate)

for deterministic and time-independent coefficients $k_0, k_1, \ldots, k_n \in \mathbb{R}^n$, $h_0, h_1, \ldots, h_n \in \mathbb{R}^{n\times n}$, $c_0, c_1, \ldots, c_n \in \mathbb{R}$ and $r_0, r_1, \ldots, r_n \in \mathbb{R}$. Finally, all the dynamic related parameters are defined as follows,

$$\theta^{\mathcal{X}} = \{k_0, K_1, h_0, H_1, c_0, C_1, r_0, R_1, \varsigma_\nu\}. \tag{A.2}$$

In most cases, it suggests using a Poisson process for counting jumps. For any time interval $[0, T]$, one can approximate the value of the Poisson process at time T by dividing the interval into $N \in \mathbb{N}$ subintervals of equidistant lengths and performing independent Bernoulli experiments with "a success probability" $\frac{cT}{N}$. This is important when applying a Monte Carlo scheme to approximate the true value of X_t at time T.

Definition A.2. (Characteristic function) Let $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space on some state space $D \subseteq \mathbb{R}^n$ and let $X : \Omega \to D$ a random variable with probability density function $f : D \to \mathbb{R}_{\geq 0}$. The *characteristic function* $\varphi : \mathbb{R}^n \to \mathbb{C}$ of X is defined by

$$\varphi_X(t) := \mathbb{E}\left[e^{it\cdot X}\right] = \int_{\mathbb{R}^n} e^{it\cdot x} f(x)\, dx, \;\; t \in \mathbb{R}^n.$$

For a stochastic process X with $D \subseteq \mathbb{R}^n, n > 2$, the definition is slightly different. We consider for every t of the time set $T = [0, \infty)$ the random variable $X_t := X(t, \cdot)$ and define likewise

$$\varphi_{X_t}(s) := \mathbb{E}\left[e^{is\cdot X_t}\right] = \int_{\mathbb{R}^n} e^{is\cdot x} f_t(x)\, dx, \;\; s \in \mathbb{R}^n.$$

Definition A.3. (Fourier transform)
Let $f \in \mathcal{L}(\mathbb{R}^n; \mathbb{R})$ (i.e. f is integrable under the Lebesgue measure). The *Fourier transform* $\widehat{f} : \mathbb{R}^n \to \mathbb{C}$ is defined as

$$\hat{f}(u) := \int_{\mathbb{R}^n} f(x) \cdot e^{-iu \cdot x} dx, \;\; u \in \mathbb{R}^n. \tag{A.3}$$

The *inverse Fourier transform* is defined as

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}^n} \hat{f}(u) \cdot e^{iu \cdot x} du, \;\; x \in \mathbb{R}^n. \tag{A.4}$$

Definition A.4. (Equivalent Martingale Measure) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with filtration $\{\mathcal{F}_t : t \geq 0\}$. A probability measure $\mathbb{Q}$ is called *equivalent martingale measure* to $\mathbb{P}$ if there is a random variable $Y > 0$ with

- $\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}[Y \cdot \chi_A] \; \forall A \in \mathcal{F}$
- $X_t \exp\left(-\int_0^t r(X_u)\, \mathrm{d}u\right)$ is a martingale under $\mathbb{Q}$ w.r.t. $\{\mathcal{F}_t : t \geq 0\}$.

The first property directly implies $\mathbb{P}(A) > 0 \iff \mathbb{Q}(A) > 0$ for $A \in \mathcal{F}$.

Definition A.5. (Black-Scholes formula)
We make the following assumptions on the market (consisting of at least one *risky asset*, the stock S, and one riskless asset, the *money market*):

1. The risk-free interest rate r is constant over time and the same for borrowing and lending. The standard deviation of the stock's returns, the volatility σ, is constant over time and the stock pays a continuous dividend $q > 0$ per share.
2. The market is *complete*, there are no arbitrage opportunities and borrowing/lending of money at the risk-free interest rate is subject to no restrictions. In particular, it is allowed to buy or sell any quantity of the stock where transactions take place with *no fees attached*.

Let $\mathcal{C}(t, \tau, K)$ ($\mathcal{P}(t, \tau, K)$) denote the price of a European call(put) option on the stock with maturity $\tau > 0$ and strike K. The prices can *analytically* be computed to

$$\begin{aligned} BS_C(\tau, K, \sigma) &:= \mathcal{C}(t, \tau, K) = S(t)e^{-q\tau}\Phi(d_1) - Ke^{-r\tau}\Phi(d_2), \\ BS_P(\tau, K, \sigma) &:= \mathcal{P}(t, \tau, K) = Ke^{-r\tau}\Phi(-d_2) - S(t)e^{-q\tau}\Phi(-d_1), \end{aligned} \tag{A.5}$$

where

$$d_1 := \frac{\log\left(\frac{S(t)}{K}\right) + \left(r - q + \frac{\sigma^2}{2}\right)\tau}{\sigma\sqrt{\tau}} \text{ and } d_2 := d_1 - \sqrt{\tau}.$$

Definition A.6. Black-Scholes implied volatility

Practice generally dictates that it is more convenient to illustrate the BS-implied volatility rather than the option price as a function of moneyness in order to facilitate easy assessment of properly normalized orders of magnitude. In accordance with Garcia et al. (2010), BS-implied volatility can be defined as follows:

According to the Black-Scholes model, implied volatility is defined as the value of the volatility parameter σ, under the risk-neutral measure, which determines the observed European call (or put) option price $\mathcal{C}(t, \tau, K)$ (or $\mathcal{P}(t, \tau, K)$). Therefore, Black-Scholes implied volatility can be formalized as follows:

$$\text{IV}(\tau, K, S_t) := \sigma^{\text{impl}}(\tau, K, S_t), \tag{A.6}$$

where $\sigma^{\text{impl}}(\cdot)$ satisfies the Black-Scholes formula (see Eq. (A.5) for more details) as follows:

$$\mathcal{C}(t, \tau, K) = BS_C(\sigma^{impl}; \tau, K, S_t). \tag{A.7}$$

Moreover, Bégin et al. (2020) argue that Black-Scholes implied volatility is a better choice for quoting the option price by referring to the discussion in Renault (1997) on the benefits of using Black-Scholes as a bijection to work with implied volatility rather than the option price.

Definition A.7. Moneyness of Options There are two major characteristics of options: moneyness, which is a strike-related quantity (such as strike: K, log-moneyness: $\log(K/S_t)$ or log-forward moneyness: $\log(K/F_{t,t+\tau})$), and time to maturity. In light of the same maturity, there are distinct patterns of volatility that depend upon the choice of moneyness. Garcia et al. (2010) argue that the most appropriate choice of moneyness values against which to plot the volatility smiles is log-forward moneyness:

$$m^k = \log(K/F_{t,t+\tau}), \tag{A.8}$$

where the $F_{t,t+\tau}$ is the futures price for the underlying asset at time t, referring to date $t + \tau$.

Moreover, in order to facilitate the comparison of volatility surfaces across different sam-

pling points, Andersen et al. (2015b) defines standardized log-forward moneyness as follows:

$$m = \frac{\log(K/F_{t,t+\tau})}{\mathrm{IV}_{t,\mathrm{ATM}} \cdot \sqrt{\tau}}, \tag{A.9}$$

where the $\mathrm{IV}_{t,\mathrm{ATM}}$ denotes the at-the-money BS implied volatility at time t. Using the standardized log-forward moneyness and risk-neutral pricing, we can define the price of the European-style out-of-the-money option (OTM option) across all the strikes as follows:

Definition A.8. OTM Option

$$O(t,\tau,K) = \begin{cases} \mathcal{P}(t,\tau,K) & m <= 0 \\ \mathcal{C}(t,\tau,K) & m > 0 \end{cases}, \tag{A.10}$$

where $\mathcal{C}(t,\tau,K)$ (respectively, $\mathcal{P}(t,\tau,K)$) denotes the price of a European call (respectively, put) option, τ represents the maturity, K is the strike price, and m is the moneyness defined in Eq. (A.9).

A.1.2 Additional Propositions

Proposition A.1. *(Characteristic function with scale value)*
Let states X_t^Y follows the definition in Definition 2.1. The $\alpha(s)$ and $\beta(s) := (\beta_1(s), (\beta_2(s), \ldots, \beta_n(s))^\top)^\top : [0,\tau] \to \mathbb{C}$ with $\beta_{2,n}(s) := (\beta_2(s), \ldots, \beta_n(s))^\top$, are the functions, which solve ODEs in Eq. (2.4), given initial values $\alpha(0) = 0$ and $\beta(0) = u$ with $u = (u_1, u_2, \ldots, u_n) \in \mathbb{C}^n$ and $u_{2,n} := (u_2, \ldots, u_n) \in \mathbb{C}^{n-1}$. Therefore, the "discounted" characteristic function with scale value is

$$\begin{aligned} \psi(u, X_t^Y, t, \tau; \theta^{\mathcal{X}^Y}) &= e^{\alpha(\tau)+\beta(\tau)\cdot X_t} \\ &= e^{\alpha(\tau)+u_1 Y_t+\beta_{2,n}(\tau)\cdot X_t^{(2,n)}} \end{aligned} \tag{A.11}$$

In addition, assuming $r_0 = 0$ and the rest of elements in R_1: $(r_1, r_2, \ldots, r_n)^\top = 0$[85], the characteristic function for the definition of Eq. (2.5) is

$$\begin{aligned} \psi^0(u, X_t^Y, t, \tau; \theta^{\mathcal{X}^Y}) &= e^{\alpha(\tau)+\beta(\tau)\cdot X_t} \\ &= e^{\alpha(\tau)+u_1 Y_t+\beta_{2,n}(\tau)\cdot X_t^{(2,n)}} \end{aligned} \tag{A.12}$$

[85] r_1 is set to 0 by Definition 2.1, so setting it to 0 is redundant.

Proof.

Since for the state Y_t with $k_1 = 0, h_1 = 0, c_1 = 0$, Y_t implied ODE is $\dot{\beta}_1(s) = 0$, one obtains the solution of ODE with initial value $\beta_1(0) = u_1$, based on its initial value $\beta_1(\tau) = u_1$. $\alpha(\tau)$ and $\beta_{2,n}(\tau)$ are the corresponding values at τ derived from the solution of ODEs in Eq. (2.4).

□

Proposition A.2. *(Fourier cosine expansion)[Theorem 2.1.1, Olson (2017)] Let $f \in \mathcal{L}^2([-\pi, \pi]; \mathbb{R})$. It is then possible to represent $f(t), t \in [-\pi, \pi]$ as an infinite series of compressed cosines and sines as follows*

$$f(t) = \acute{\sum}_{k=0}^{\infty} \left[A_k \cos(kt) + B_k \sin(kt) \right],^{86}$$

where

$$A_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos(kt)\, dt, B_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin(kt)\, dt.$$

If f is odd (even), then $A_k \equiv 0$ ($B_k \equiv 0$) and only sines (cosines) remain.

For any function $g \in \mathcal{L}^2([0, \pi]; \mathbb{R})$ we may define $f : [-\pi, \pi] \to \mathbb{R}, t \mapsto g(|t|)$. The function $f \in \mathcal{L}^2([-\pi, \pi]; \mathbb{R})$ is even and hence we do have for $t \in [0, \pi]$

$$g(t) = f(t) = \acute{\sum}_{k=0}^{\infty} A_k \cos(kt), A_k = \frac{2}{\pi} \int_0^{\pi} f(t) \cos(kt)\, dt.$$

Proposition A.3. *(Cosine series coefficients)*

Let $k \in \mathbb{N}_0$ and $[c, d] \subset [a, b]$ an arbitrary subinterval of $[a, b]$. Both functions

$$\chi_k(c, d) := \int_c^d e^y \cos\left(k\pi \frac{y - a}{b - a} \right) dy,$$
$$\Psi_k(c, d) := \int_c^d \cos\left(k\pi \frac{y - a}{b - a} \right) dy$$

have explicit solutions.

Proof.

The function $\cos\left(k\pi \frac{y-a}{b-a}\right)$ has a known antiderivative for $k \neq 0$, namely

$$\frac{b - a}{k\pi} \sin\left(k\pi \frac{y - a}{b - a} \right).$$

[86] The accent over the summation sign indicates that the first summand will be weighted by $\frac{1}{2}$.

Hence, $\Psi_0(c,d) = d - c$ and $\Psi_k(c,d) = \frac{b-a}{k\pi}\left[\sin\left(k\pi\frac{d-a}{b-a}\right) - \sin\left(k\pi\frac{c-a}{b-a}\right)\right]$ for $k \neq 0$.

For $\chi_k(c,d)$, we apply partial integration twice as follows

$$\begin{aligned}
\chi_k(c,d) &= \int_c^d e^y \cos\left(k\pi\frac{y-a}{b-a}\right) dy \\
&= \left[e^d \cos\left(k\pi\frac{d-a}{b-a}\right) - e^c \cos\left(k\pi\frac{c-a}{b-a}\right)\right] + \frac{k\pi}{b-a}\int_c^d e^y \sin\left(k\pi\frac{y-a}{b-a}\right) dy \\
&= -\left(\frac{k\pi}{b-a}\right)^2 \chi_k(c,d) + \left[e^d \cos\left(k\pi\frac{d-a}{b-a}\right) - e^c \cos\left(k\pi\frac{c-a}{b-a}\right)\right] \\
&+ \frac{k\pi}{b-a}\left[e^d \sin\left(k\pi\frac{d-a}{b-a}\right) - e^c \sin\left(k\pi\frac{c-a}{b-a}\right)\right],
\end{aligned}$$

from which one gets an explicit representation for $\chi_k(c,d)$. □

Proposition A.4. *(Lemma of Itô)[Theorem 4.2.1, Øksendal (2003)]*
Let

$$dX_t = a(t)dt + b(t)dB_t$$

a d-dimensional Itô process with the vectorwise drift function a and the matrix-wise diffusion function b (possibly adapted). Let $f : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^p, (t,x) \mapsto (f_1(t,x), \ldots, f_p(t,x))$ *with* $f \in C^{1,2}([0,\infty) \times \mathbb{R}^n; \mathbb{R}^p)$. *Then, the process*

$$Y(t,\omega) := f(t, X_t)$$

is again an Itô process whose k-th component $Y^{(k)}$ *follows the SDE*

$$dY_t^{(k)} = \frac{\partial}{\partial s} f_k(t, X_t)\, dt + \sum_{i=1}^n \frac{\partial}{\partial x_i} f_k(t, X_t)\, dX_t^{(i)} + \frac{1}{2}\sum_{i,j=1}^n \frac{\partial^2}{\partial x_i \partial x_j} f_k(t, X_t)\, dX_t^{(i)} dX_t^{(j)},$$

with $dB_t^{(i)} dB_t^{(j)} = \delta_{i,j} dt$ *and* $dB_t^{(i)} dt = dt dB_t^{(i)} = 0, i = 1, \ldots, n$.

Proposition A.5. *(Put-Call parity)*
Based on the same assumptions from Definition A.5 (where we do not require completeness of the market), consider the existence of a stock S paying dividend at continuous rate q, a European call $\mathcal{C}(t,\tau,K)$ *and a European put* $\mathcal{P}(t,\tau,K)$ *with equal strike K, equal maturity T and equal underlying* S_t. *We can build a riskless portfolio from stock long,*

put long and call short. If the no-arbitrage principle holds true (e.g. the law of one price is fulfilled), we can deduce the explicit price relation as follows:

$$\mathcal{C}(t,\tau,K) - \mathcal{P}(t,\tau,K) = S_t e^{-q\tau} - Ke^{-r\tau}. \tag{A.13}$$

Proposition A.6. *(Exchangeability of the Minimum)*

Let $X \subset \mathbb{R}^n, Y \subset \mathbb{R}^m$, $M = X \times Y$ and $f : M \to \mathbb{R}$. Then we do have

$$\min_{(x,y)\in M} f(x,y) = \min_{x\in X} \min_{y\in Y} f(x,y).$$

A.2 Higher Orders Derivatives

In this section, I heuristically show that the higher-order derivatives of the characteristic function of X_t^Y defined in Proposition A.1 are again solutions of a set of ODEs. From definition Eq. (2.3), we have

$$\ln \mathbb{E}_t\left[e^{u_1 Y_{t+\tau}}\right] = \alpha(\tau) + \beta(\tau)\cdot X_t^Y \tag{A.14}$$

where $\alpha(s)$ and $\beta(s)$ are defined with initial values $\alpha(0) = 0$ and $\beta(0) = (u_1, 0, \ldots, 0)^\top$ in Eq. (2.4) as following:

$$\begin{aligned}
\dot{\alpha}(s) &\stackrel{!}{=} -r_0 + k_0 \cdot \beta(s) + \frac{1}{2}\beta(s)^\top h_0 \beta(s) + c_0(\varsigma_\nu(\beta(s)) - 1),\ \alpha(0) = 0 \\
\dot{\beta}(s) &\stackrel{!}{=} -r_1 + K_1^\top \beta(s) + \frac{1}{2}\beta(s)^\top H_1 \beta(s) + C_1(\varsigma_\nu(\beta(s)) - 1),\ \beta(0) = (u_1, 0\ldots, 0)^\top,
\end{aligned} \tag{A.15}$$

where $s \in [0,\tau]$ and $\beta(s)$ is a function of $\beta(u_1, s)$ depending on the initial values. Moreover, the function $\partial_1\beta(s) = \frac{\partial\beta(s)}{\partial u_1}$.

Then differentiating on both sides with respect u_1 one further can obtain

$$\begin{aligned}
\frac{\partial \ln \mathbb{E}_t\left[e^{u_1 Y_{t+\tau}}\right]}{\partial u_1} &= \frac{\mathbb{E}_t\left[e^{u_1 Y_{t+\tau}} Y_{t+\tau}\right]}{\mathbb{E}_t\left[e^{u_1 Y_{t+\tau}}\right]} \\
&= \frac{\partial\alpha(\tau)}{\partial u_1} + \frac{\partial\beta(\tau)}{\partial u_1}\cdot X_t^Y
\end{aligned} \tag{A.16}$$

Based on the definition of function $\partial_1\beta := \frac{\partial\beta(u_1,s)}{\partial u_1}$,

$$\partial_1\beta(u_1,s) \overset{87}{=} \int \frac{\partial\dot{\beta}(u_1,s)}{\partial u_1} ds \tag{A.17}$$

Furthermore, the first derivative of $\partial_1\beta$ with respect to s is

$$\frac{\partial(\partial_1\beta(u_1,s))}{\partial s} = \frac{\partial\dot{\beta}(u_1,s)}{\partial u_1} \tag{A.18}$$

$$\begin{aligned}
\frac{\partial(\partial_1\beta(u_1,s))}{\partial s} &= \frac{\partial\dot{\beta}(u_1,s)}{\partial u_1} \\
&\overset{88}{=} \frac{\partial\dot{\beta}(u_1,s)}{\partial\beta(u_1,s)}\frac{\partial\beta(u_1,s)}{\partial u_1} \\
&= \frac{\partial\left(-r_1 + K_1^\top\beta(u_1,s) + \frac{1}{2}\beta(u_1,s)^\top H_1\beta(u_1,s) + C_1(\varsigma_\nu(\beta(u_1,s))-1)\right)}{\partial\beta(u_1,s)}\frac{\partial\beta(u_1,s)}{\partial u_1} \\
&= K_1^\top\frac{\partial\beta(u_1,s)}{\partial u_1} + \left(\frac{\partial\beta(u_1,s)}{\partial u_1}\right)^\top H_1\beta(u_1,s) + C_1\nabla\varsigma_\nu(\beta(u_1,s))\cdot\frac{\partial\beta(u_1,s)}{\partial u_1} \\
&= K_1^\top\partial_1\beta(u_1,s) + \partial_1\beta(u_1,s)^\top H_1\beta(s) + C_1\nabla\varsigma_\nu(\beta(u_1,s))\cdot\partial_1\beta(u_1,s),
\end{aligned} \tag{A.20}$$

with initial values of $\beta(u,0) = (u_1,0,\ldots,0)$, we have initial values for $\partial_1\beta(u_1,0)$ are independent of u_1

$$\begin{aligned}
\partial_1\beta(u_1,0) &= \left(\frac{\partial\beta_1(u_1,0)}{\partial u_1},\ldots,\frac{\partial\beta_n(u_1,0)}{\partial u_1}\right)^\top \\
&= (1,0,\ldots,0)^\top = d_1
\end{aligned} \tag{A.21}$$

Therefore, the first derivative of the result of Eq. (2.18) can be derived as above.

87

$$\begin{aligned}
\partial_1\beta &:= \frac{\partial\beta(u_1,s)}{\partial u_1} \\
&= \int_0^\tau \frac{\partial^2\beta(u_1,s)}{\partial u_1\partial s} ds \\
&= \int_0^\tau \frac{\partial^2\beta(u_1,s)}{\partial s\partial u_1} ds \\
&= \int_0^\tau \frac{\partial\frac{\partial\beta(u_1,s)}{\partial s}}{\partial u_1} ds \\
&= \int_0^\tau \frac{\partial\dot{\beta}(u_1,s)}{\partial u_1} ds
\end{aligned} \tag{A.19}$$

88 chain rule for multi-variant function

A.3 The Bates Model

Illustration A.1. The data generating processes X_t contains two factors: log stock price and volatility $\{\ln S_t, V_t\}$, whose dynamics are given as follows

$$\begin{aligned} d\ln S_t &= \left(\mu - \frac{1}{2}V_t\right)dt + \sqrt{V_t}\left(\rho dW_{1,t} + \sqrt{1-\rho^2}dW_{2,t}\right) + Z_t^s dN_t, \\ dV_t &= \kappa\left(\theta - V_t\right)dt + \sigma\sqrt{V_t}dW_{1,t}, \end{aligned} \tag{A.22}$$

where

$\{\mu, \kappa, \theta, \rho\}$ are constant parameters;

$W_{1,t}$ and $W_{2,t}$ are standard independent Wiener processes;

N_t is Poisson process for counting jump activities with constant instantaneous intensity c;

Z_t^s denote the random variable of jump size with $\frac{d\nu(z)}{dz} := \frac{1}{\sigma_s\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{z-\mu_s}{\sigma_s}\right)}$

$$\varsigma_\nu(w) = \int_{\mathbb{R}} e^{wz}\nu(z)dz = e^{\mu_s w + \frac{1}{2}\sigma_s w^2};$$

A.3.1 Parameters of the Bates Model in AJD Matrix Interpretation

$\theta^{\mathcal{X}}$ for the Bates model in Illustration A.1 is given as follows:

$$k_0 = \begin{pmatrix} \mu \\ \kappa\theta \end{pmatrix}, \quad K_1 = \begin{pmatrix} 0 & -\frac{1}{2} \\ 0 & -\kappa \end{pmatrix},$$

$$h_0 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad H_1 = \left(\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma\rho \\ \sigma\rho & \sigma^2 \end{pmatrix}\right),$$

$$r_0 = 0, \quad R_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad c_0 = c, \quad C_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

[88] The further economic implication of the model and its parameters are discussed with Illustration 2.2 with more details

A.3.2 AJD Interpretation of the ODEs for the Bates Model

For simplicity, the time argument is omitted. The ODEs for the "dicounted" characteristic function is concretely given, as follows,

$$\begin{aligned}
\dot{\alpha} &= \begin{pmatrix} \mu & \kappa\theta \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + c\left(\exp\left(\mu_s\beta_1 + \frac{1}{2}\sigma_s^2\beta_1^2\right) - 1\right) \\
&= \mu\beta_1 + \kappa\theta\beta_2 + c\left(\exp\left(\mu_s\beta_1 + \frac{1}{2}\sigma_s^2\beta_1^2\right) - 1\right), \\
\begin{pmatrix} \dot{\beta}_1 \\ \dot{\beta}_2 \end{pmatrix} &= \begin{pmatrix} 0 & 0 \\ -\frac{1}{2} & -\kappa \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0, & \begin{pmatrix} \beta_1 & \beta_2 \end{pmatrix} \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \end{pmatrix}^\top \\
&= \begin{pmatrix} 0 \\ -\frac{1}{2}\beta_1 - \kappa\beta_2 + \frac{1}{2}\left(\beta_1^2 + 2\beta_1\beta_2\rho\sigma + (\beta_2\sigma)^2\right) \end{pmatrix}.
\end{aligned}$$

In addition, the following concrete ODEs are obtained:

$$\begin{aligned}
\dot{\alpha} &= \mu\beta_1 + \kappa\theta\beta_2 - c\left(\exp\left(\mu_s\beta_1 - \frac{1}{2}\sigma_s^2\beta_1^2\right) - 1\right), \\
\dot{\beta}_1 &= 0, \\
\dot{\beta}_2 &= -\frac{1}{2}\beta_1 - \kappa\beta_2 + \frac{1}{2}\left(\beta_1^2 + 2\beta_1\beta_2\rho\sigma + (\beta_2\sigma)^2\right).
\end{aligned} \tag{A.23}$$

To be more specific, in some applications $\alpha(0) = 0$ and $\beta(0) = (u_1, u_2)^\top$, the "discounted" characteristic function is

$$\psi(u, X_t, t, \tau; \theta^{\mathcal{X}^c}) = e^{\alpha(\tau) + u_1 \ln S_t + \beta_2(\tau)V_t}. \tag{A.24}$$

A.3.3 Cumulants for Bates Model

By using the dynamic of Illustration A.1 and specifying $\Delta Y_{t,\tau} = \ln S_{t+\tau} - \ln S_t := R_s(t,\tau)$ one have

$$\mathcal{K}_{1,R_s(t,\tau)}(V_t) = \mathbb{E}_t\left[R_s(t,\tau)\right] = \partial_1\alpha(\tau) + \partial_1\beta_2(\tau)V_t, \tag{A.25}$$

and further with $m \geq 2$

$$\mathcal{K}_{m,R_s(t,\tau)}(V_t) = \partial_m\alpha(\tau) + \partial_m\beta_2(\tau)V_t. \tag{A.26}$$

A.4 (Linear) Minimum Mean Square Estimation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space and $X, Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ two random variables on the Hilbert space $\mathcal{L}^2$.
Define $\mathcal{F}_Y := \sigma\left(\bigcup_{A\in\mathcal{B}(\mathbb{R})} Y^{-1}(A)\right)$ as the σ-algebra generated by Y and the subspace $\mathcal{L}^2(\Omega, \mathcal{F}_Y, \mathbb{P}) \subseteq \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. Optimally "predicting" X given Y means to minimize

$$\min_{Z\in\mathcal{L}^2(\Omega,\mathcal{F}_Y,\mathbb{P})} \mathbb{E}\left[(X-Z)^2|Y\right].$$

The solution makes use of the basic concepts of the conditional expectation $\mathbb{E}[X|Y]$:

$$\begin{aligned}\mathbb{E}\left[(X-Z)^2|Y\right] &= \mathbb{E}\left[(X-\mathbb{E}[X|Y]+\mathbb{E}[X|Y]-Z)^2|Y\right]\\ &= \mathbb{E}\left[(X-\mathbb{E}[X|Y])^2|Y\right] + 2\underbrace{\mathbb{E}\left[(X-\mathbb{E}[X|Y])(\mathbb{E}[X|Y]-Z)|Y\right]}_{(*)}\\ &\quad + \mathbb{E}\left[(\mathbb{E}[X|Y]-Z)^2|Y\right].\end{aligned}$$

Now, since Z and, by construction, $\mathbb{E}[X|Y]$ are $\mathcal{F}_Y$-measurable, we may simplify $(*)$ to

$$\begin{aligned}\mathbb{E}\left[(X-\mathbb{E}[X|Y])(\mathbb{E}[X|Y]-Z)|Y\right] &= \mathbb{E}\left[X-\mathbb{E}[X|Y]|Y\right](\mathbb{E}[X|Y]-Z)\\ &= (\mathbb{E}[X|Y]-\mathbb{E}[X|Y])(\mathbb{E}[X|Y]-Z)\\ &= 0 \ \mathbb{P}\text{-a.s.}\end{aligned}$$

Hence, what we have in fact is

$$\mathbb{E}\left[(X-Z)^2|Y\right] = \mathbb{E}\left[(X-\mathbb{E}[X|Y])^2|Y\right] + \mathbb{E}\left[(\mathbb{E}[X|Y]-Z)^2|Y\right],$$

from which the pointwise optimizer is deduced to $Z = \mathbb{E}[X|Y]$.

Assuming that X, Y are together *jointly normally distributed*, we can show a particular easy structure of the conditional expectation. Integrating over the respective joint density is tedious and there is yet another, more systematic way. The r.v.

$$X - \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}Y$$

is jointly normal with Y and uncorrelated (and by normality: independent) to Y:

$$\operatorname{Cov}\left(X - \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(Y)}Y, Y\right) = \operatorname{Cov}(X,Y) - \operatorname{Cov}(X,Y)\frac{\operatorname{Cov}(Y,Y)}{\operatorname{Var}(Y)} = 0.$$

Hence, we do have by $\mathbb{E}\left[X - \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(Y)}Y|Y\right] = \mathbb{E}\left[X - \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(Y)}Y\right]$:

$$\begin{aligned}\mathbb{E}[X|Y] &= \mathbb{E}\left[X - \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(Y)}Y + \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(Y)}Y|Y\right] \\ &= \mathbb{E}[X] + \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(Y)}(Y - \mathbb{E}[Y]).\end{aligned} \tag{A.27}$$

A.5 Particle Filtering[89]

In general, the assumption of a linear structure and normally distributed noise represents a too severe constraint for most time series given, i.e. fitting a model is better done by allowing a more general structure. Following Dahlin and Schön (2015), we may represent the nonlinear state-space model as follows

$$x_0 \sim \mu(x_0), \;\; x_t|x_{t-1} \sim f(x_t|x_{t-1}), \;\; y_t|x_t \sim g(y_t|x_t). \tag{A.28}$$

The initial distribution μ_0, the transition and observation functionals f, g might contain a vector θ of parameters which we assume *known*[90]for the moment. In most applications, however, they need to be estimated concurrently, too. We want to outline the general idea of a particle filter, orientating ourselves at Doucet and Johansen (2009), Johannes et al. (2009), and Creal (2012).

In the following, let $y_{0:t} = (y_0, \ldots, y_t)$ and $x_{0:t} = (x_0, \ldots, x_t)$ denote the vector of the first $t+1$ accumulated observations and latent states, respectively. In this Bayesian context, we are interested in estimating the trajectory of latent states, conditional on the observations, i.e. the full *posterior distribution*

$$p(x_{0:t}|y_{0:t}) = \frac{p(x_{0:t}, y_{0:t})}{p(y_{0:t})}. \tag{A.29}$$

[89]This part is mainly based on the results of research projects with a C-RAM students. I would like to thank Maximilian Kübler for his assistance.

[90]In an empirical study, the parameters θ are typically estimated from observed data $(y_0, \ldots, y_T)$.

We do have $p(x_{0:t}, y_{0:t}) = p(x_{0:t})\, p(y_{0:t}|x_{0:t})$. Following a markovian setup, we have

$$p(x_{0:t}) = \mu(x_0) \prod_{i=1}^{t} f(x_i|x_{i-1}) \quad \text{and} \quad p(y_{0:t}, x_{0:t}) = \prod_{i=0}^{t} g(y_i|x_i).$$

However, $p(y_{0:t}) = \int p(x_{0:t}, y_{0:t})\, dx_{0:t}$ needs to be calculated, too. In most models, the distributions are not known in closed form and hence, numerical methods are required.

In Bayesian Inference, we aim to compute the expectation as trustable approximation of the true latent state

$$\mathbb{E}[x_t|y_{0:t}] = \int x_t p(x_t|y_{0:t})\, dx_t.$$

From a Monte Carlo point of view, we would sample from $x_t \sim p(\cdot|y_{0:t})$ and approximate the integral by computing the empirical mean

$$\mathbb{E}[x_t|y_{0:t}] \approx \frac{1}{m} \sum_{i=1}^{m} x_t^i,$$

that converges $\mathbb{P}$-a.s. against $\mathbb{E}[x_t|y_{0:t}]$ by the *strong law of large numbers*. However, sampling from the unknown posterior is unfeasible. Instead, we sample from an *importance distribution* $x_t \sim \pi(\cdot|y_{0:t})$ and approximate with suitable weights $\hat{w}_t^i > 0$

$$\mathbb{E}[x_t|y_{0:t}] \approx \sum_{i=1}^{m} \hat{w}_t^i x_t^i.$$

Since

$$\mathbb{E}[x_t|y_{0:t}] = \int x_t p(x_t|y_{0:t})\, dx_t = \int \left[x_t \frac{p(x_t|y_{0:t})}{\pi(x_t|y_{0:t})} \right] \pi(x_t|y_{0:t})\, dx_t,$$

the choice of the *importance weights* $\hat{w}_t^i := \frac{1}{m} \frac{p\left(x_t^i|y_{0:t}\right)}{\pi\left(x_t^i|y_{0:t}\right)} > 0$ is obvious. By

$$p\left(x_t^i|y_{0:t}\right) \propto p\left(y_{0:t}|x_t^i\right) p\left(x_t^i\right),$$

we can construct *unnormalized* weights through the definition

$$\tilde{w}_t^i := \frac{p(y_{0:t}|x_t^i)\, p(x_t^i)}{\pi(x_t^i|y_{0:t})}.$$

To account for the missing m, we *normalize* them by letting

$$w_t^i := \frac{\tilde{w}_t^i}{\sum_{i=1}^m \tilde{w}_t^i}.$$

General particle filters use sequentially a set of weighted particles $\{w_t^i, x_t^i\}, i = 1, \ldots, m$ to approximate the posterior with the discrete *Dirac measure* $\delta(\cdot)$ through

$$p(x_t|y_{0:t}) \approx \sum_{i=1}^m w_t^i \delta\left(x_t - x_t^i\right).$$

We can further derive a recursion for our weights by

$$\begin{aligned} p(x_{0:t}|y_{0:t}) &\propto p(y_t|x_{0:t}, y_{0:t-1})\, p(x_{0:t}|y_{0:t-1}) \\ &= g(y_t|x_t)\, p(x_t|x_{0:t-1}, y_{0:t-1})\, p(x_{0:t-1}|y_{0:t-1}) \\ &= g(y_t|x_t)\, f(x_t|x_{t-1})\, p(x_{0:t-1}|y_{0:t-1}) \end{aligned}$$

The importance weights w_t^i could thus be chosen *recursively* as

$$w_t^i = \frac{p(x_{0:t}^i|y_{0:t})}{\pi(x_{0:t}^i|y_{0:t})} \propto \frac{g(y_t|x_t^i)\, f\left(x_t^i|x_{t-1}^i\right) p\left(x_{0:t-1}^i|y_{0:t-1}\right)}{\pi(x_{0:t}^i|y_{0:t})}.$$

The update of our importance distribution has to take into account the new observation coming in and the suggested form of our full posterior distribution

$$\pi(x_{0:t}|y_{0:t}) = \pi(x_t|x_{0:t-1}, y_{0:t})\, \pi(x_{0:t-1}|y_{0:t-1})$$

is appropriate (prediction of next state based on the next observation). This translates into the *weight recursion*

$$w_t^i \propto \frac{g(y_t|x_t^i)\, f\left(x_t^i|x_{t-1}^i\right)}{\pi\left(x_t^i|x_{0:t-1}^i, y_{0:t}\right)} w_{t-1}^i, \quad t = 1, \ldots, T.$$

It is non-trivial in a nonlinear setting to compute $\pi\left(x_t^i|x_{0:t-1}^i, y_{0:t}\right)$. There is a simple choice, namely $\pi(x_t|x_{0:t-1}, y_{0:t}) = f\left(x_t^i|x_{t-1}^i\right)$. Having said this, the problem is that one gives away the informational content of y_t which might lead to *sample impoverishment* with degenerate particles. There is no simple way to overcome this deficiency of the

so-called *Bootstrap filter* (BPF). Moreover, we can decompose the denominator to

$$\begin{aligned}\pi\left(x_t^i|x_{0:t-1}^i, y_{0:t}\right) &\propto p\left(y_t|x_t^i\right) f\left(x_t^i|x_{0:t-1}^i\right) \\ &= \pi\left(y_t|x_{0:t-1}^i\right) \pi\left(x_t|x_{0:t-1}^i, y_t\right) p\left(x_{0:t-1}^i|y_{0:t-1}\right) \\ &\propto w_{t-1}\pi_1\left(y_t|x_{0:t-1}^i\right) \cdot \pi_2\left(x_t|x_{0:t-1}^i, y_t\right)\end{aligned} \tag{A.30}$$

for approximating functionals π_1, π_2 with suitable support.

Algorithm 3: (Auxiliary) Particle filter for state inference

Input : Observations $y_0, \ldots, y_T$, $m \in \mathbb{N}$ particles, initial distribution $\mu_0(x_0)$.

Returns : Vector $\hat{x}$ as best guess for the latent state vector x to drive the system. $\hat{x}_T$ is estimated solely based on a normalized version of $\{w_T^i\}_{i=1,\ldots,m}$.

while $t \leq T$ **do**

  if $t = 0$ **then**

    Initialization: Set $\pi_0 = \mu_0$ as initial sampling distribution, $\hat{x}$ denoting the vector of state estimates over time. Sample from the proposed distribution μ_0, get a sequence of m particles $x_0^i, i = 1, \ldots, m$ with weight $w_0^i = \frac{1}{m}$ each.

  end

  1. Weighting: For each particle x_{t-1}^i compute the *look-ahead* weight $\tau_t^i = \pi_1\left(y_t|x_{0:t-1}^i\right) w_{t-1}^i$. Normalize the weights to

$$\tau_t^i = \frac{\tau_t^i}{\sum_{l=1}^m \tau_t^l}.$$

  2. Estimation: Compute the mean $\hat{x}_{t-1} = \sum_{i=1}^m x_{t-1}^i \cdot \tau_t^i$ as state estimation.

  3. Resampling: Generate new particles - prior propagation - $\{x_{t-1}^i\}$ (multinomial distribution on $\{\tau_t^i\}_{i=1,\ldots,m}$) and replace the old particles.

  4. Propagation: For each particle x_{t-1}^i sample a transitioned particle by letting

$$x_t^i \sim \pi_2\left(\cdot|x_{0:t-1}^i, y_{0:t}\right).$$

  5. Importance weighting: For each particle x_t^i, update its weight by

$$w_t^i = \frac{g\left(y_t|x_t^i\right) f\left(x_t^i|x_{t-1}^i\right)}{\tau_t^i \cdot \pi_2\left(x_t^i|x_{0:t-1}^i, y_{0:t}\right)} w_{t-1}^i. \tag{A.31}$$

  $t = t + 1$

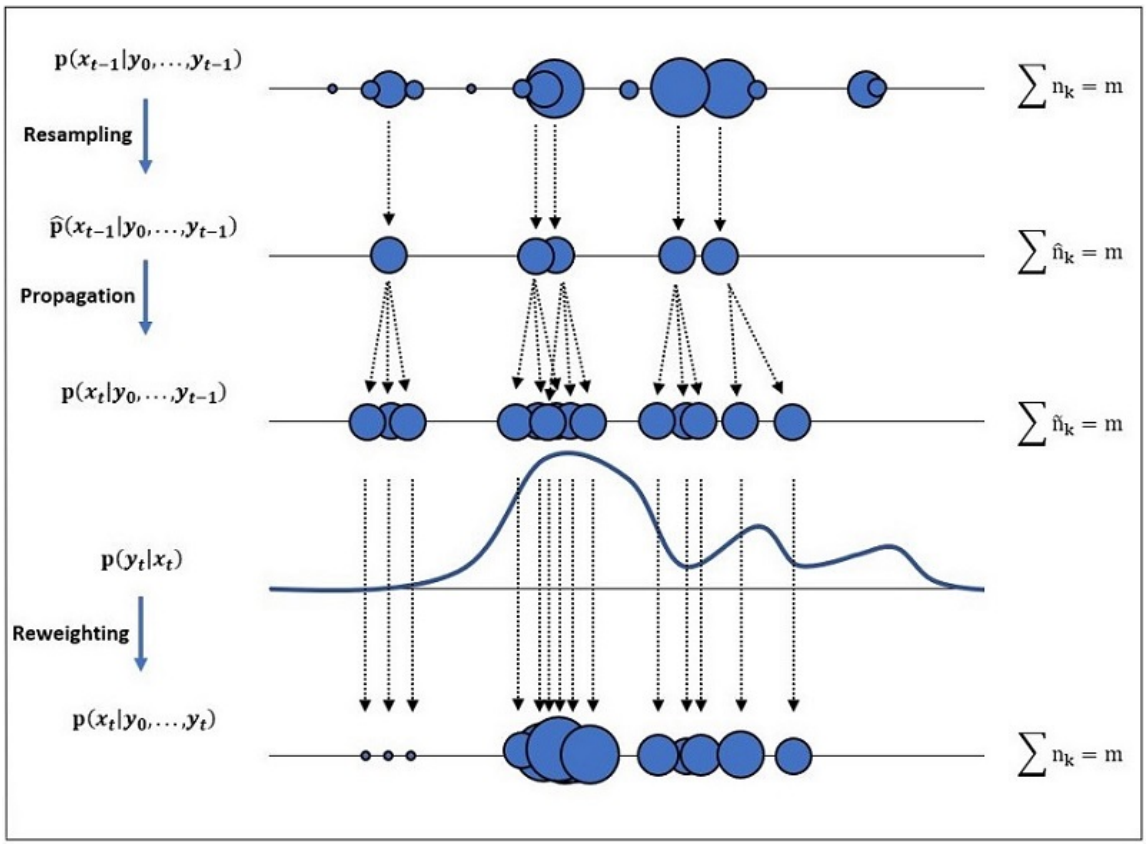
end

In principal, we could repeat this method for each and every observation introduced into the system. It is clear however that if we *don't* resample at the beginning of every iteration, we lose focus of the important particles, i.e. those with high weight. Scenarios that are unlikely, i.e. particles x_t^i $(i = 1, \ldots, m)$ with low weight at time t_t, still have a more or less severe impact on the propagated probability distribution at time t_t.

We want to get rid of *unfit* particles and make use of what is called *importance sampling* in literature. After sampling from the discrete distribution obtained, we again have m (not necessarily different) particles at the beginning of the propagation step of iteration t. It looks like we just deteriorated our local model by reducing the number of atoms our distribution could carry. Fortunately, the *propagation* will ensure that we truly have m different atoms which can lie close to each other but hardly ever coincide (assuming a *continuous* transition density).

We visualize the development of the particle set by taking the example of the Bootstrap filter. The number of particles $m \in \mathbb{N}$ is constant over time but one or more particles may occur multiply, i.e. the *effective sample size*[91] (ESS) may be low at times.

Figure A.1: Scheme of the Bootstrap filter over one iteration.



[91] The ESS is computed as $1 \leq \frac{1}{\sum_{i=1}^{m}(w_t^i)^2} \leq m$ and is low when a few particles dominate $p(\cdot|y_{0:t})$.

The *Auxiliary particle filter* (APF) does essentially the same and differs only to the extent that the weighting steps 1 & 5 take into account the new observation in order to increase the adaptiveness to the SSM specifics.

If $\pi_1\left(y_t|x^i_{0:t-1}\right) \equiv 1$, then the resampling procedure ignores new information that y_t brings into the system and local identification of the true hidden state is in theory much harder. Second, if $\pi_2\left(x^i_t|x^i_{0:t-1}, y_{0:t}\right) = f\left(x^i_t|x^i_{t-1}\right)$, then we have the same problematic. If these two simplifications are used, we again end up with the Bootstrap filter. To be more precise, one could incorporate new information for instance by

$$\pi_1\left(y_t|x^i_{0:t-1}\right) := g\left(y_t|\mathbb{E}\left[x^i_t|x^i_{t-1}\right]\right),$$

since f and g are mostly known analytically. For affine jump-diffusions, we may use the Auxiliary particle filter because it prevents quick sample impoverishment and detects jumps more efficiently.[92]

Nevertheless, there is some degree of freedom. Following the SSM set up for the state evolution, using the expected next particle $\mathbb{E}\left[x^i_t|x^i_{t-1}\right]$ implies zero diffusion in the state equation. Depending on the correlation between state and measurement innovation, the expected state diffusion significantly reduces the standard deviation of the measurement. For near perfect correlation ($|\rho| \approx 1$), the measurement is peaked and resembles very much a dirac distribution. It is then to be feared that most of the particles (if not all) get assigned a very negative log-weight $\log\left(\pi_1\left(y_t|x^i_{0:t-1}\right)\right)$. This could hamper identification of good particles and we need to account for this problem when correlation is reaching *critical extreme values*. Entirely neglecting the SSM connection is no choice, so we always limit the influence of the state innovation to the measurement.

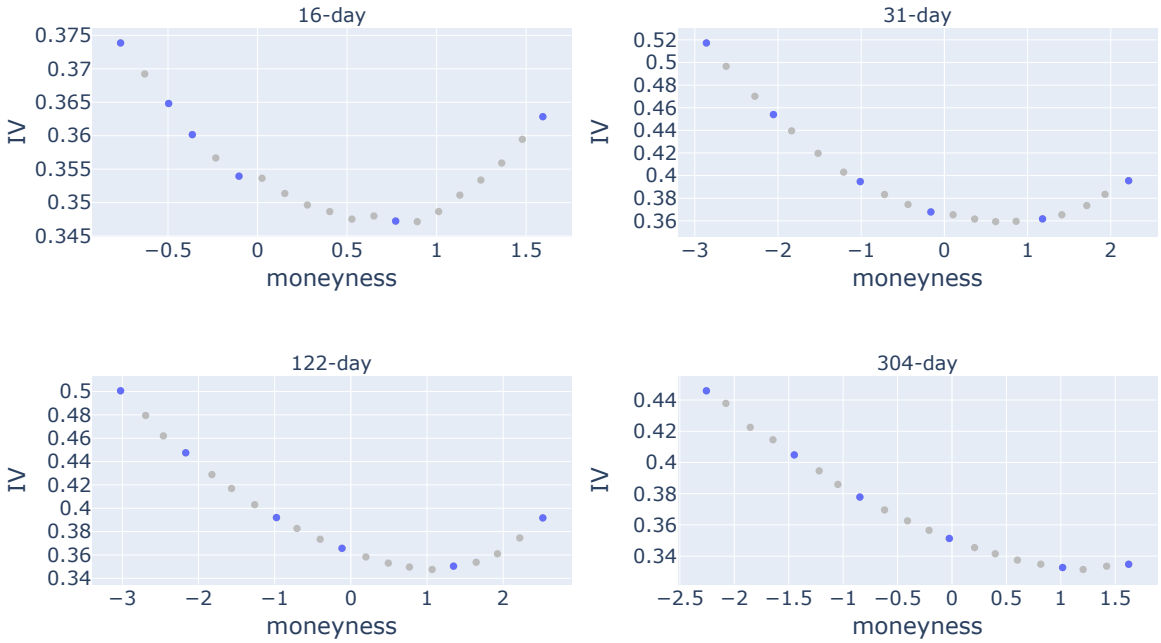
[92] A classical example is given in Johannes et al. (2009), pages 2767, 2768: Imagine the case of Merton's model where at time $t+1$ a positive jump happens. If the jump probability is constant 1 %, from 1000 particles, about 10 will be propagated with jump. Typically, the observation distribution $p\left(y_{t+1}|x_{t+1}\right)$ is shifted to the right but the particles where no jump took place will be assigned vanishing weight since their propagation move does not explain the harsh change at all. The next resampling step will concentrate the entire particles' distribution to the few particles that have at least some explanatory force, thus leading to degeneration.

B Supplementary Figures

Figure B.1: Predicted returns and realized returns.

This figure illustrates the return predictability of the left tail factor. It plots the time series of predicted returns (solid blue lines) and realized returns (dashed grey lines) for the S&P 500 index at 6-, 12-, 16-, 20-, 24- and 30-week horizons. The predicted returns are obtained from the predictive regressions based on the estimation results from Spec IV. Spec IV represents the specification of constant relative errors in the filtering step and $w = 1$ in the log-likelihood function.

Figure B.2: Implied volatilities used in estimation for Apple Inc. as of December 19, 2012.



This plot illustrates the Black-Scholes implied volatilities (IVs) used in the estimation for Apple Inc. as of December 19, 2012. The grey circles denote all available observed IVs subsequent to the data cleaning step, whereas the blue circles represent the IVs effectively employed in our estimation approach.

C Supplementary Tables

Table C.1: Firms included in the return predictability analysis.

	Symbol		Symbol		Symbol		Symbol
1	AMGN	2	MSFT	3	NKE	4	NOC
5	OMC	6	OXY	7	PNC	8	PPL
9	STT	10	TROW	11	YUM	12	CAT
13	CMA	14	FDX	15	LEN	16	MTB
17	NTRS	18	PENN	19	PFE	20	PRU
21	SEE	22	SLB	23	T	24	TGT
25	VRSN	26	XOM	27	ADBE	28	ADI
29	CCI	30	CL	31	CMI	32	EOG
33	MDT	34	QCOM	35	SHW	36	TXN
37	UNH	38	WAT	39	WY	40	AES
41	AIG	42	ATVI	43	CHRW	44	FDS
45	LH	46	MMM	47	MSI	48	NFLX
49	NSC	50	NVDA	51	NWL	52	PG
53	SWK	54	SYK	55	TAP	56	UPS
57	WFC	58	WYNN	59	CAG	60	FITB
61	ORCL	62	PAYX	63	PHM	64	TRV
65	TXT	66	URI	67	VLO	68	WMT
69	CLX	70	KMB	71	SO	72	SYY
73	TRMB	74	TSN	75	TTWO	76	UHS
77	VFC	78	WDC	79	DUK	80	LMT
81	MRO	82	MS	83	NEE	84	NEM
85	PCAR	86	PPG	87	STZ	88	VMC
89	VRTX	90	CMCSA	91	DRI	92	FFIV

Continued on next page

	Symbol		Symbol		Symbol		Symbol
93	FISV	94	LLY	95	LNC	96	PFG
97	PH	98	TJX	99	TSCO	100	UNP
101	VZ	102	WHR	103	FE	104	MO
105	MRK	106	NTAP	107	NUE	108	PEP
109	PGR	110	STX	111	USB	112	VNO
113	WM	114	ZION	115	ABC	116	ADM
117	AFL	118	AMT	119	BDX	120	CI
121	CNC	122	DHI	123	ETR	124	MCD
125	A	126	AAPL	127	ADP	128	ALL
129	EXPD	130	JPM	131	LRCX	132	RL
133	BWA	134	D	135	LUV	136	AZO
137	CME	138	DHR	139	MCHP	140	RCL
141	ALK	142	AVB	143	EW	144	KLAC
145	MCK	146	APD	147	FAST	148	K
149	LVS	150	CSX	151	CTSH	152	CVX
153	MMC	154	MOS	155	MU	156	NRG
157	AMAT	158	AXP	159	DLTR	160	EXC
161	KMX	162	MCO	163	REGN	164	BAX
165	COO	166	DE	167	FCX	168	KEY
169	ROP	170	ADSK	171	MHK	172	ROK
173	AON	174	BBY	175	CCL	176	CSCO
177	CVS	178	ETN	179	PXD	180	APA
181	BIIB	182	CAH	183	DXC	184	EIX
185	EMR	186	GIS	187	GS	188	HAL
189	HSIC	190	HSY	191	INTC	192	IRM
193	RE	194	AEP	195	AKAM	196	BA
197	BK	198	BMY	199	DGX	200	GD
201	GOOGL	202	HAS	203	HIG	204	HUM
205	INTU	206	ISRG	207	LOW	208	ROST
209	BAC	210	DVN	211	HD	212	HES
213	IBM	214	ITW	215	JCI	216	APH
217	EA	218	EBAY	219	EL	220	GE
221	GPN	222	GRMN	223	GWW	224	HOLX
225	HON	226	JNJ	227	MAR	228	ED

Continued on next page

	Symbol		Symbol		Symbol		Symbol
229	GILD	230	GLW	231	HPQ	232	HST
233	JBHT	234	MAS	235	AAP	236	ACN
237	COP	238	DOV	239	KO	240	MGM
241	DISH	242	KR	243	AMZN	244	CPB
245	ABT	246	BEN	247	COF	248	DVA
249	JNPR	250	MET	251	SBUX	252	COST
253	CTAS	254	EMN				

Note: This table lists the symbols of U.S. stocks that have been included in our empirical analysis, accompanied by completed estimation processes and accessible return series.

Table C.2: Firms excluded due to insufficient observations.

No.	Symbol	No.	Symbol	No.	Symbol	No.	Symbol
1	C	2	FIS	3	MPC	4	MRNA
5	PSA	6	PSX	7	SRE	8	TDG
9	TECH	10	TSLA	11	TYL	12	ANTM
13	DOW	14	DPZ	15	NDSN	16	NI
17	PLD	18	PNW	19	SJM	20	UAA
21	UAL	22	ULTA	23	WST	24	XRAY
25	BXP	26	CFG	27	DFS	28	PYPL
29	SEDG	30	SPG	31	STE	32	SYF
33	TDY	34	TFC	35	TMUS	36	UA
37	WBA	38	WRB	39	XEL	40	XYL
41	ABMD	42	AJG	43	ALLE	44	CEG
45	NOW	46	OKE	47	SIVB	48	SNPS
49	SWKS	50	VRSK	51	WELL	52	ZBH
53	DTE	54	LKQ	55	MOH	56	MTCH
57	NDAQ	58	NWS	59	PKG	60	PTC
61	UDR	62	V	63	VIAC	64	AEE
65	ARE	66	AVY	67	BKNG	68	BRO
69	CRM	70	CTLT	71	EQIX	72	EQR
73	MKC	74	SYMC	75	TT	76	TWTR
77	BLK	78	CDW	79	CF	80	DRE
81	LNT	82	MTD	83	NCLH	84	NLSN
85	O	86	ORLY	87	PEG	88	PNR
89	POOL	90	SPGI	91	TPR	92	WAB
93	WBD	94	WRK	95	ZBRA	96	ZTS
97	MSCI	98	NVR	99	NXPI	100	PAYC
101	PKI	102	PM	103	TEL	104	VTR
105	WTW	106	ALGN	107	BSX	108	LHX
109	LIN	110	MNST	111	NWSA	112	ODFL
113	PEAK	114	PVH	115	PWR	116	SNA
117	TFX	118	TMO	119	WEC	120	ANET
121	BIO	122	CB	123	CMS	124	KEYS
125	KHC	126	QRVO	127	APTV	128	CPRT

Continued on next page

No.	Symbol	No.	Symbol	No.	Symbol	No.	Symbol
129	CZR	130	RJF	131	AMCR	132	ANSS
133	AWK	134	BR	135	CBRE	136	CHD
137	EXPE	138	LUMN	139	RMD	140	CPT
141	AME	142	AMP	143	DIS	144	ETSY
145	EVRG	146	KIM	147	REG	148	AMD
149	AVGO	150	DAL	151	DD	152	CTVA
153	KMI	154	CBOE	155	CMG	156	CNP
157	FBHS	158	LW	159	LYB	160	ROL
161	LYV	162	MPWR	163	ALB	164	DG
165	EPAM	166	LB	167	LDOS	168	CDAY
169	CDNS	170	CTRA	171	BKR	172	CINF
173	DXCM	174	GM	175	GNRC	176	GOOG
177	GPC	178	HCA	179	ICE	180	ILMN
181	IPGP	182	MA	183	BF.B	184	ES
185	ESS	186	FLT	187	FMC	188	FRC
189	FRT	190	FTNT	191	HBAN	192	HRL
193	IT	194	ATO	195	ECL	196	FOX
197	FOXA	198	FTV	199	HLT	200	HPE
201	IDXX	202	MAA	203	RF	204	AOS
205	CHTR	206	HII	207	IEX	208	IFF
209	IP	210	IVZ	211	JKHY	212	RHI
213	ABBV	214	CE	215	CRL	216	ENPH
217	GL	218	INCY	219	IQV	220	FANG
221	RSG	222	AIZ	223	BRK.B	224	EXR
225	MKTX	226	SBAC	227	DLR	228	FB
229	L	230	MLM	231	SBNY	232	AAL
233	EFX	234	MDLZ				

Note: This table provides a list of stocks excluded from the empirical analysis due to a deficiency of more than 10% of data points within the specified timeframe of 2017 to 2021.

Table C.3: Firms excluded from empirical analysis for additional reasons.

	Symbol	Excluded reason
1	WMB	estimation failed
2	SCHW	estimation failed
3	F	estimation failed
4	IPG	estimation failed
5	TER	filtering failed
6	JEC	filtering failed
7	RTN	filtering failed
8	BLL	no return data
9	CERN	no return data
10	CTXS	no return data
11	IR	no return data

Note: This table enumerates the stocks that are excluded due to failure in the estimation process, the filtering step, or due to the unavailability of return data from Yahoo Finance. The stock symbols are listed alongside the reasons for their exclusion.

Bibliography

Ammann, M., and A. Feser. 2019. Robust estimation of risk-neutral moments. *Journal of Futures Markets*, 39:1137–1166.

An, B.-J., A. Ang, T. G. Bali, and N. Cakici. 2014. The joint cross section of stocks and options. *The Journal of Finance*, 69:2279–2337.

Andersen, T. G., N. Fusari, and V. Todorov. 2014. Supplementary Appendix to "The Risk Premia Embedded in Index Options". `http://www.kellogg.northwestern.edu/faculty/todorov/htm/papers/opa_app.pdf`.

Andersen, T. G., N. Fusari, and V. Todorov. 2015a. Parametric inference and dynamic state recovery from option panels. *Econometrica*, 83:1081–1145.

Andersen, T. G., N. Fusari, and V. Todorov. 2015b. The risk premia embedded in index options. *Journal of Financial Economics*, 117:558–584.

Andersen, T. G., N. Fusari, and V. Todorov. 2015c. SUPPLEMENT TO "PARAMETRIC INFERENCE AND DYNAMIC STATE RECOVERY FROM OPTION PANELS".

Andersen, T. G., N. Fusari, and V. Todorov. 2017. Short-Term Market Risks Implied by Weekly Options. *The Journal of Finance*, 72:1335–1386.

Andersen, T. G., N. Fusari, and V. Todorov. 2019. Supplementary Appendix to "The Pricing of Tail Risk and the Equity Premium: Evidence from International Option Markets".

Andersen, T. G., N. Fusari, and V. Todorov. 2020. The pricing of tail risk and the equity premium: evidence from international option markets. *Journal of Business & Economic Statistics*, 38:662–678.

Andrieu, C., A. Doucet, and R. Holenstein. 2010. Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72:269–342.

Bakshi, G., N. Kapadia, and D. Madan. 2003. Stock return characteristics, skew laws, and the differential pricing of individual equity options. *The Review of Financial Studies*, 16:101–143.

Bakshi, G., and D. Madan. 2000. Spanning and derivative-security valuation. *Journal of financial economics*, 55:205–238.

Bardgett, C., E. Gourier, and M. Leippold. 2019. Inferring volatility dynamics and risk premia from the S&P 500 and VIX markets. *Journal of Financial Economics*, 131:593–618.

Bates, D. S. 1996. Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *The Review of Financial Studies*, 9:69–107.

Bates, D. S. 2000. Post-'87 crash fears in the S&P 500 futures option market. *Journal of econometrics*, 94:181–238.

Bates, D. S. 2019. How Crashes Develop: Intradaily Volatility and Crash Evolution. *Journal of Finance*, 74:193–238.

Bégin, J.-F., C. Dorion, and G. Gauthier. 2020. Idiosyncratic jump risk matters: Evidence from equity returns and options. *The Review of Financial Studies*, 33:155–211.

Biscani, F., and D. Izzo. 2020. A parallel global multiobjective framework for optimization: pagmo. *Journal of Open Source Software*, 5:2338. URL `https://doi.org/10.21105/joss.02338`.

Black, F., and M. Scholes. 1973. The pricing of options and corporate liabilities. *Journal of political economy*, 81:637–654.

Bollerslev, T., and V. Todorov. 2011. Tails, fears, and risk premia. *The Journal of Finance*, 66:2165–2211.

Broadie, M., M. Chernov, and M. Johannes. 2007. Model specification and risk premia: Evidence from futures options. *The Journal of Finance*, 62:1453–1490.

Chen, H., and S. Joslin. 2012. Generalized transform analysis of affine processes and applications in finance. *The Review of Financial Studies*, 25:2225–2256.

Chen, L. 1996. *Stochastic mean and stochastic volatility: a three-factor model of the term structure of interest rates and its applications in derivatives pricing and risk management*. Blackwell publishers.

Chernov, M., A. R. Gallant, E. Ghysels, and G. Tauchen. 2003. Alternative models for stock price dynamics. *Journal of Econometrics*, 116:225–257.

Chopin, N., O. Papaspiliopoulos, et al. 2020. *An introduction to sequential Monte Carlo*, vol. 4. Springer.

Christoffersen, P., K. Jacobs, and K. Mimouni. 2010. Volatility dynamics for the S&P500: Evidence from realized volatility, daily returns, and option prices. *The Review of Financial Studies*, 23:3141–3189.

Conrad, J., R. F. Dittmar, and E. Ghysels. 2013. Ex ante skewness and expected stock returns. *The Journal of Finance*, 68:85–124.

Cox, J. C., J. E. Ingersoll, and S. A. Ross. 1985. A Theory of the Term Structure of Interest Rates. *Econometrica*, 53:385. URL `https://doi.org/10.2307/1911242`.

Creal, D. 2012. A survey of sequential Monte Carlo methods for economics and finance. *Econometric reviews*, 31:245–296.

Dahlin, J., and T. B. Schön. 2015. Getting started with particle Metropolis-Hastings for inference in nonlinear dynamical models. *arXiv preprint arXiv:1511.01707*.

Dai, Q., and K. J. Singleton. 2000. Specification analysis of affine term structure models. *The Journal of Finance*, 55:1943–1978.

De Jong, F. 2000. Time series and cross-section information in affine term-structure models. *Journal of Business & Economic Statistics*, 18:300–314.

Doucet, A., and A. M. Johansen. 2009. A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of nonlinear filtering*, 12:3.

Duan, J.-C., and J.-G. Simonato. 1999. Estimating and testing exponential-affine term structure models by Kalman filter. *Review of quantitative finance and accounting*, 13:111–135.

Duffee, G. R., and R. H. Stanton. 2012. Estimation of dynamic term structure models. *The Quarterly Journal of Finance*, 2:1250008.

Duffie, D. 2010. *Dynamic asset pricing theory*. Princeton University Press.

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

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

Durbin, J., and S. J. Koopman. 2012. *Time Series Analysis by State Space Methods*. 38. Oxford University Press.

Eraker, B. 2004. Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *The Journal of finance*, 59:1367–1403.

Fama, E. F., and K. R. French. 2015. A five-factor asset pricing model. *Journal of financial economics*, 116:1–22.

Fang, F., and C. W. Oosterlee. 2008. A novel pricing method for European options based on Fourier-cosine series expansions. *SIAM Journal on Scientific Computing*, 31:826–848.

Feunou, B., and C. Okou. 2018. Risk-neutral moment-based estimation of affine option pricing models. *Journal of Applied Econometrics*, 33:1007–1025.

Filipovic, D. 2009. *Term-Structure Models. A Graduate Course.* Springer.

Fulop, A., and J. Li. 2019. Bayesian estimation of dynamic asset pricing models with informative observations. *Journal of Econometrics*, 209:114–138.

Fulop, A., J. Li, and J. Yu. 2015. Self-exciting jumps, learning, and asset pricing implications. *The Review of Financial Studies*, 28:876–912.

Garcia, R., E. Ghysels, and E. Renault. 2010. The econometrics of option pricing. In *Handbook of Financial Econometrics: Tools and Techniques*, pp. 479–552. Elsevier.

Gruber, P. H., C. Tebaldi, and F. Trojani. 2021. The price of the smile and variance risk premia. *Management Science*, 67:4056–4074.

Harrison, J. M., and D. M. Kreps. 1979. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic theory*, 20:381–408.

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.

Jegadeesh, N., and S. Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, 48:65–91.

Johannes, M. S., N. G. Polson, and J. R. Stroud. 2009. Optimal filtering of jump diffusions: Extracting latent states from asset prices. *The Review of Financial Studies*, 22:2759–2799.

Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82:35–45.

Kloeden, P. E., and E. Platen. 1992. *Numerical Solution of Stochastic Differential Equations*. Springer Berlin Heidelberg. URL https://doi.org/10.1007/978-3-662-12616-5.

Kou, S. G. 2002. A jump-diffusion model for option pricing. *Management science*, 48:1086–1101.

Lewellen, J., et al. 2015. The Cross-section of Expected Stock Returns. *Critical Finance Review*, 4:1–44.

Martin, I. 2017. What is the Expected Return on the Market? *The Quarterly Journal of Economics*, 132:367–433.

Martin, I. W., and C. Wagner. 2019. What is the Expected Return on a Stock? *The Journal of Finance*, 74:1887–1929.

Merton, R. C. 1976. Option pricing when underlying stock returns are discontinuous. *Journal of financial economics*, 3:125–144.

Monfort, A., F. Pegoraro, J.-P. Renne, and G. Roussellet. 2017. Staying at zero with affine processes: An application to term structure modelling. *Journal of Econometrics*, 201:348–366.

Newey, W. K., and K. D. West. 1987. Hypothesis testing with efficient method of moments estimation. *International Economic Review*, pp. 777–787.

Øksendal, B. 2003. *Stochastic Differential Equations. An introduction with applications.* 6th ed. Springer.

Olson, T. 2017. *Applied Fourier Analysis: From Signal Processing to Medical Imaging.* Birkhäuser.

Ornthanalai, C. 2014. Levy jump risk: Evidence from options and returns. *Journal of Financial Economics*, 112:69–90.

Pan, J. 2002. The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of financial economics*, 63:3–50.

Piazzesi, M. 2010. Affine term structure models. In *Handbook of financial econometrics: Tools and Techniques*, pp. 691–766. Elsevier.

Platen, E., and N. Bruti-Liberati. 2010. *Numerical solution of stochastic differential equations with jumps in finance*, vol. 64. Springer Science & Business Media.

Renault, E. 1997. Econometric models of option pricing errors. *Econometric Society Monographs*, 28:223–278.

Vasicek, O. 1977. An equilibrium characterization of the term structure. *Journal of financial economics*, 5:177–188.

Vilkov, G. 2021. Option-implied Data and Analysis. URL `https://osf.io/z2486/`.