# Estimating Option Pricing Models Using a Characteristic Function-Based Linear State Space Representation[*]

H. Peter Boswijk

Amsterdam School of Economics

University of Amsterdam

and Tinbergen Institute

Roger J. A. Laeven

Amsterdam School of Economics

University of Amsterdam, EURANDOM

and CentER

Evgenii Vladimirov

Amsterdam School of Economics

University of Amsterdam

and Tinbergen Institute

June 3, 2024

## Abstract

We develop a novel filtering and estimation procedure for parametric option pricing models driven by general affine jump-diffusions. Our procedure is based on the comparison between an option-implied, model-free representation of the conditional log-characteristic function and the model-implied conditional log-characteristic function, which is functionally affine in the model's state vector. We formally derive an associated linear state space representation and the asymptotic properties of the corresponding measurement errors. The state space representation allows us to use a suitably modified Kalman filtering technique to learn about the latent state vector and a quasi-maximum likelihood estimator of the model parameters, for which we establish asymptotic inference results. Accordingly, the filtering and estimation procedure brings important computational advantages. We analyze the finite-sample behavior of our procedure in Monte Carlo simulations. The applicability of our procedure is illustrated in two case studies that analyze S&P 500 option prices and the impact of exogenous state variables capturing Covid-19 reproduction and economic policy uncertainty.

*Keywords:* Options; Characteristic Function; Affine Jump-Diffusion; State Space Representation.

*JEL Classification:* Primary: C13; C58; G13; Secondary: C32; G01.

# 1 Introduction

Over the past decades, explosive growth in the trading of option contracts has attracted the attention of academics and practitioners to the development and estimation of increasingly sophisticated option pricing models. The building blocks of many continuous-time option pricing models are semimartingale stochastic processes that govern the dynamics of the underlying asset. These processes are often latent with stochastic diffusive volatility as the prototypical example, as in the classical Heston (1993) model. The literature also suggests the need to allow for a discontinuous jump component, both in the asset price dynamics and in its volatility process, potentially with a time-varying stochastic jump intensity.

An important econometric challenge lies in estimating the parameters of these continuous-time models and in filtering their unobserved and time-varying components, since option prices are highly nonlinear functions of the state vector. This stands in contrast to, for instance, term structure models, where bond yields can be represented as a linear function of the states, at least within the affine framework (see, e.g., Piazzesi, 2010, for a review of the affine term structure literature). To evaluate option prices as a function of the state vector, one typically needs to apply either Fourier-based methods or simulation-based approaches, in both cases at a substantial computational cost. This is one of the reasons why in much of the empirical research on option pricing, only a subset of the available option price data is used, such as at-the-money contracts or weekly (typically Wednesday) options data.

In this paper, we develop a new latent state filtering and parameter estimation procedure for option pricing models governed by general affine jump-diffusion processes. Our procedure leverages the linear relationship between the logarithm of an option-implied, model-free spanning formula for the conditional characteristic function of the underlying asset return on the one hand, and the state vector induced by parametric model specification on the other hand. From this relationship, we formally derive a linear state space representation, and establish the asymptotic properties of the corresponding measurement errors. Linearity of the measurement and state updating equations that make up the state space representation, with coefficient and variance matrices that are (semi-)closed-form functions of the parameters, allows us to exploit Kalman filtering techniques. The proposed estimation procedure is fast and easy to implement, circumventing the typical computational burden in conducting inference on option pricing models.

Exploiting the option-spanning formula of Carr and Madan (2001) for European-style payoff functions, we replicate the risk-neutral conditional characteristic function (CCF) of the underlying log-asset price at the expiration date in a completely model-independent way. In other words, we imply information about the CCF from the option prices without imposing any parametric assumptions on the underlying asset price dynamics. A similar option-spanning approach for the CCF is used by Todorov (2019) to develop an option-based nonparametric spot volatility estimator. On the other hand, a large stream of literature is devoted to parametric option pricing models belonging to the general affine jump-diffusion (AJD) family; canonical

examples are Heston (1993), Duffie, Pan, and Singleton (2000), Pan (2002), and Bates (2006).[1] The defining property of the AJD class is the exponential-affine joint CCF, which is available in semi-closed form. By comparing the two option pricing representations—model-free and model-implied—we can obtain a linear relation between the logarithm of the option-implied CCF and the model-dependent CCF within the affine framework.

The state vector in AJD option pricing models typically contains both observable processes and latent factors. We address the filtering of such latent factors by developing a linear state space representation for this model class. The development includes an asymptotic analysis of the measurement error components, consisting of observation, truncation and discretization errors, under a double asymptotic scheme in the moneyness dimension. The state space representation allows us to employ suitably modified Kalman filtering techniques to learn about the unobserved intrinsic components of the model and estimate the model parameters using quasi-maximum likelihood (QML). We establish novel asymptotic inference results for the QML parameter estimator. QML approaches based on Kalman filtering are often used in the affine term structure literature, where the yields themselves are linear functions of the state vector (see, e.g., Duffee, 1999, de Jong, 2000, Driessen, 2005). Besides the possibility to exploit Kalman filtering and QML estimation techniques, another advantage of our approach is that, once the model-free CCF has been obtained from the data, no further numerical option pricing methods, such as the FFT approach of Carr and Madan (1999) or simulation-based methods, are needed. Therefore, our method reduces computational costs considerably relative to many existing approaches in the option pricing literature. We note that, whereas the parametric CCF is used to price options in Fourier-based methods, here we use the CCF to directly learn about the latent factors and model parameters.

We analyze the developed estimation procedure in Monte Carlo simulations based on several AJD specifications. We consider a one-factor AJD option pricing model, with the stochastic volatility and jump intensity both being affine functions of a single latent process, and a two-factor AJD model specification with an observable exogenous factor. We find good finite-sample performance in both cases, notwithstanding the challenging nature of the econometric problem.

Finally, we illustrate our new filtering and estimation approach in an empirical application to S&P 500 index options. In particular, we filter and estimate the latent volatility and jump intensity from a stochastic volatility model with co-jumps in returns and volatility. We also investigate the impact of the Covid-19 propagation rate on the stock market within this model, by embedding the associated reproduction number as an exogenous factor into the volatility and jump intensity dynamics. Our results show that while the reproduction number has only a mild effect on total diffusive volatility, it contributes substantially to the likelihood of jumps. By contrast, when we consider an Economic Policy Uncertainty index as exogenous factor, the jump intensity process is not affected, but the exogenous factor contributes significantly to diffusive volatility.

Various estimation and filtering strategies for option pricing models have been developed in

---

[1]See also, e.g., Broadie, Chernov, and Johannes (2007), Aït-Sahalia, Cacho-Diaz, and Laeven (2015), Andersen, Fusari, and Todorov (2017) and the references therein.

the literature. These include the (penalized) nonlinear least squares methods in, for instance, Bakshi, Cao, and Chen (1997), Broadie et al. (2007), Andersen, Fusari, and Todorov (2015a); the efficient method of moments of Gallant and Tauchen (1996) as applied in Chernov and Ghysels (2000) and Andersen, Benzoni, and Lund (2002); the implied-state methods initiated by Pan (2002) and further analyzed by Santa-Clara and Yan (2010); the Markov Chain Monte Carlo method in Eraker (2004) and Eraker, Johannes, and Polson (2003); and the particle filtering method, see Johannes, Polson, and Stroud (2009) and Bardgett, Gourier, and Leippold (2019). Recently, Aït-Sahalia, Li, and Li (2024) proposed an efficient and flexible maximum likelihood estimation procedure for continuous-time models with latent components. Using a closed-form approximation of one-step transition quantities, they can avoid requiring observable proxies such as option prices for the latent factors and imposing assumptions on filtered distributions.

Many of the aforementioned estimation methods use as inputs option prices or a monotonic transformation thereof, such as implied volatilities. By contrast, we propose an estimation procedure based on the prices of spanning option portfolios that by the bijection between CCFs and conditional distributions, in principle, contain *all* probabilistic information about the stochastic process governing the dynamics of the underlying asset. In general, estimation strategies based on the transform space of conditional characteristic functions are, of course, not new to the literature. For instance, Carrasco and Florens (2000) develop a generalized method of moments (GMM) estimator with a continuum of moment conditions based on the CCF; see also Singleton (2001), Carrasco, Chernov, Florens, and Ghysels (2007). In applications to option prices, Boswijk, Laeven, and Lalu (2015) and Boswijk, Laeven, Lalu, and Vladimirov (2021) propose to imply the latent state vector from a panel of options and then estimate the model via GMM with a continuum of moments. Bates (2006) develops maximum likelihood estimation and filtering using CCFs. In particular, he proposes a recursive likelihood evaluation by updating the CCF of a latent variable conditional upon observed data. However, unlike our approach, these methods require numerical integration over the dimension of the state vector, thus suffering from a 'curse of dimensionality'.

Our work is also related to Feunou and Okou (2018), who exploit the linear relation between the first four risk-neutral cumulants of the log-asset price and latent factors. They obtain these cumulants via a portfolio of options and employ the Kalman filter to estimate the latent factors. The main difference with our approach is that we exploit the CCF, and the corresponding state space representation we develop, instead of the first four moments. The CCF contains much richer information, leading to more efficient inference. Another difference is in dimension reduction: Feunou and Okou (2018) use a two-step principal components analysis (PCA) to reduce the dimension of the risk-neutral cumulants observed at different maturities. Instead, we use a modified version of a so-called collapsed Kalman filtering approach, originally developed by Jungbacker and Koopman (2015), which does not suffer from information losses relative to the full-dimensional setting.

The paper is organized as follows. Section 2 provides the theoretical framework for aligning the option-implied and model-implied CCFs. In Section 3, we develop the state space representation, and establish the result about the orders of measurement errors, under a double

asymptotic scheme. This allows us to next develop the filtering approach and corresponding estimation procedure. Section 4 presents the Monte Carlo simulation results. We describe the data in Section 5 and the empirical applications in Section 6. Conclusions are in Section 7. In supplementary material, four appendices provide details on $(i)$ the proofs of Propositions 1 and 2, $(ii)$ the computation of conditional moments, $(iii)$ the inter- and extrapolation scheme for option prices and the measurement errors in the CCF replication, and $(iv)$ additional simulation and empirical results.

## 2 Theoretical Framework

In this section, we provide the theoretical framework for our approach. We start with extracting information about the CCF from option prices allowing for general underlying dynamics. Next, we consider the CCF within the AJD class, which is exponentially affine in the model's state variables. Finally, we discuss how to align the two CCFs—option-implied model-free and AJD model-implied—to conduct inference about the model parameters and the latent state variables.

### 2.1 Option-implied CCF

Throughout the paper, we fix a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$. On this probability space, we consider the dynamics of an arbitrage-free financial market. The no-arbitrage assumption guarantees the existence of a risk-neutral probability measure $\mathbb{Q}$, locally equivalent to $\mathbb{P}$. Since we are interested in exploiting information from options, we formulate the model dynamics under $\mathbb{Q}$.

Let us denote by $F_t$ the futures price at time $t$ for a stock or an index futures contract with some fixed maturity. The absence of arbitrage implies that the futures price process is a semimartingale. In this subsection, we assume the following general dynamics for $F_t$ under $\mathbb{Q}$:

$$\frac{\mathrm{d}F_t}{F_t} = v_t \mathrm{d}W_t + \int_{\mathbb{R}} x\tilde{\mu}(\mathrm{d}t, \mathrm{d}x), \qquad F_0 > 0, \tag{1}$$

where $v_t$ is an adapted, locally bounded, but otherwise unspecified stochastic volatility process; $W_t$ is a standard Brownian motion; $\mu$ is a counting random measure with compensator $\nu_t(\mathrm{d}x)\mathrm{d}t$ such that $\tilde{\mu}(\mathrm{d}t, \mathrm{d}x) := \mu(\mathrm{d}t, \mathrm{d}x) - \nu_t(\mathrm{d}x)\mathrm{d}t$ is the associated martingale measure and $\int (x^2 \wedge 1)\nu_t(\mathrm{d}x) < \infty$.

We further denote out-of-the-money (OTM) European-style option prices at time $t$ with time-to-maturity $\tau > 0$ and strike price $K > 0$ by $O_t(\tau, K)$. Under the no-arbitrage assumption, the option prices equal the risk-neutral conditional expectations of the corresponding discounted payoff functions:

$$O_t(\tau, K) = \begin{cases} \mathbb{E}^{\mathbb{Q}}[e^{-r\tau}(F_{t+\tau} - K)^+ | \mathcal{F}_t], & \text{if } K > F_t, \\ \mathbb{E}^{\mathbb{Q}}[e^{-r\tau}(K - F_{t+\tau})^+ | \mathcal{F}_t], & \text{if } K \leq F_t. \end{cases}$$

The OTM price $O_t(\tau, K)$ is a call option price if $K > F_t$ and a put option price if $K \leq F_t$. For simplicity, we assume a constant interest rate $r$.

5

Following Carr and Madan (2001), any twice continuously differentiable European-style payoff function $g(F_{t+\tau})$, with first and second derivatives $g_F$ and $g_{FF}$, can be spanned via a position in risk-free bonds, futures (or stocks) and options with a continuum of strikes:

$$g(F_{t+\tau}) = g(x) + g_F(x)(F_{t+\tau} - x)$$
$$+ \int_0^x g_{FF}(K)(K - F_{t+\tau})^+ \mathrm{d}K + \int_x^\infty g_{FF}(K)(F_{t+\tau} - K)^+ \mathrm{d}K.$$

Here, $x \in \mathbb{R}^+$, the first and second terms on the right-hand side correspond to risk-free bonds and futures positions, and the third and fourth terms correspond to OTM options. Taking conditional expectations under the risk-neutral measure for $x = F_t$, we find that the price at time $t$ of a contingent claim with payoff function $g(F_{t+\tau})$ can be expressed as a weighted portfolio of a risk-free bond and OTM options:

$$\mathbb{E}^{\mathbb{Q}}[e^{-r\tau} g(F_{t+\tau})|\mathcal{F}_t] = e^{-r\tau} g(F_t) + \int_0^\infty g_{FF}(K) O_t(\tau, K) \mathrm{d}K. \tag{2}$$

This general spanning result lies behind the construction of one of most popular 'fear' indices—the VIX index, when $g(F_{t+\tau}) = \log(F_{t+\tau}/F_t)$. Some other applications of the spanning formula (2) include the calculation of the option-implied skewness and kurtosis (Bakshi, Kapadia, & Madan, 2003) and of the corridor implied volatility (Andersen & Bondarenko, 2007).

Applying this result to the complex-valued payoff function $g(x) = e^{iu \log(x/F_t)}$ yields that the discounted CCF of log returns can be spanned as

$$\phi_t(u, \tau) := e^{-r\tau} \mathbb{E}^{\mathbb{Q}}[e^{iu \log(F_{t+\tau}/F_t)}|\mathcal{F}_t]$$
$$= e^{-r\tau} - (u^2 + iu) \int_0^\infty \frac{1}{K^2} e^{iu(\log K - \log F_t)} \cdot O_t(\tau, K) \mathrm{d}K$$
$$= e^{-r\tau} - (u^2 + iu) \frac{1}{F_t} \int_{\mathbb{R}} e^{(iu-1)m} \cdot O_t(\tau, m) \mathrm{d}m, \tag{3}$$

where $m = \log(K/F_t)$ is the log-moneyness of an option with strike price $K$.[2]

It is important to emphasize that the spanning of the CCF in equation (3) is exact and is furthermore completely model independent akin to the VIX construction. Therefore, the CCF of log returns over a particular horizon $\tau$ can be replicated in a model-free way given a single cross-sectional slice of liquid option prices with all strikes (and the same maturity $\tau$). A similar approach of CCF spanning is taken by Todorov (2019) to nonparametrically estimate spot volatilities from option prices (considering the limit as $\tau \downarrow 0$).

The expression for $\phi_t(u, \tau)$ in (3) cannot be computed in reality as we do not observe option prices for a continuum of strikes. Nevertheless, as we detail in Section 3.1, the expression in (3) is easy to approximate using a limited number of observable option prices. While developing our estimation procedure, we take both the resulting approximation errors as well as the observation errors in option prices, and hence in the CCF approximation, into account. Henceforth, we denote by $\widehat{\phi}_t(u, \tau)$ the computationally feasible counterpart of the option-implied CCF; it is explicitly defined in (14) below. In our simulation experiments and empirical applications, we further use an interpolation-extrapolation scheme to improve reliability of the approximation.

---

[2]With slight abuse of notation, we use the same symbol $O_t$ for the option value as a function of $(\tau, K)$ and as a function of $(\tau, m)$.

## 2.2 Affine jump-diffusion CCF

Whereas the CCF in (3) is model independent, the CCF of log returns of the underlying asset is often considered under some parametric assumptions on the return dynamics. A model-implied CCF depends on the model parameters, which we generally do not know, and potentially on the dynamics of other latent processes, which affect the distribution of returns. Therefore, by suitably aligning the model-free and parameter-dependent CCFs, we may learn about the model parameters and the unobservable state dynamics.

We restrict our attention to the broad class of AJD models defined in Duffie et al. (2000). The main attraction of the AJD class is that the Laplace transform has a semi-closed-form expression and is of the exponential-affine form. Suppose that $X_t$ is a Markov process representing an $d_X$-dimensional state vector in $D \subset \mathbb{R}^{d_X}$ with the first component being the log price of an asset. We assume that under the physical and risk-neutral probability measures, the state vector $X_t$ solves the following stochastic differential equation:

$$\mathrm{d}X_t = \mu(X_t; \theta)\mathrm{d}t + \sigma(X_t; \theta)\mathrm{d}W_t + \sum_{i=1}^{d_J} J_{i,t}\mathrm{d}N_{i,t}, \tag{4}$$

where $W_t$ is a standard Brownian motion in $\mathbb{R}^{d_W}$; $\mu \colon D \to \mathbb{R}^{d_X}$ and $\sigma \colon D \to \mathbb{R}^{d_X \times d_W}$ are the drift and diffusion functions; $N_{i,t}$ is a pure jump process with intensity $\{\lambda^i(X_t; \theta) \colon t \geq 0\}$, $\lambda^i \colon D \to \mathbb{R}^+$; $\{J_{i,t}\}_{t \geq 0}$ constitutes a sequence of jump sizes with generic conditional distribution $\nu^i$ on $\mathbb{R}^{d_X}$ for $i = 1, \ldots, d_J$; and $\theta$ is a vector of unknown parameters that governs the model for $X_t$.[3] We note that we allow for multiple jump types each arriving with their own intensity process as in the generalized AJD class in Appendix B of Duffie et al. (2000). The specification (4) can be extended further, e.g., to include a time-dependent structure and infinite activity jumps; see Duffie et al. (2000), in particular Appendix B, and Duffie, Filipović, and Schachermayer (2003) for more details on the AJD class formulation.

Following Duffie et al. (2000), the drift $\mu(x)$, diffusive variance $\sigma(x)\sigma(x)'$ and jump intensities $\lambda^i(x)$ are assumed to be affine on $D$:

$$
\begin{aligned}
\mu(x) &= K_0 + K_1 x, & K_0 &\in \mathbb{R}^{d_X}, \ K_1 \in \mathbb{R}^{d_X \times d_X}, \\
\sigma(x)\sigma(x)' &= H_0 + \sum_{j=1}^{d_X} x_j H_1^{(j)}, & H_0 &\in \mathbb{R}^{d_X \times d_X}, \ H_1^{(j)} \in \mathbb{R}^{d_X \times d_X}, \ j = 1, \ldots, d_X, \\
\lambda^i(x) &= l_{i,0} + l_{i,1} \cdot x, & l_{i,0} &\in \mathbb{R}, \ l_{i,1} \in \mathbb{R}^{d_X}, \ i = 1, \ldots, d_J,
\end{aligned}
$$

where $x_j$ is the $j$-th element of a vector $x$ and $H_1^{(j)}$ for $j=1, \ldots, d_X$ form a $d_X \times d_X \times d_X$ tensor $H_1$ by stacking matrices along a new dimension. The joint regularity conditions on $(D, \mu, \sigma, \lambda, \nu)$ that guarantee a unique solution to the SDE (4) are discussed in Duffie and Kan (1996) and Dai and Singleton (2000). These joint conditions put constraints on the parameter vector $\theta$. Therefore, we consider a model from the AJD class indexed by $\theta$ in a parameter space $\Theta$ containing such admissible parameter values, on which there is a unique solution to (4) that remains in $D$. For instance, in the case of the stochastic volatility component, the admissible parameter values in $\Theta$ ensure that the volatility process remains nonnegative, by satisfying Feller's condition; see also the discussion of the admissibility problem in Singleton (2006, Chapter 5).

---

[3]We follow the notational conventions in Duffie et al. (2000) and write $t$ also when $t-$ should be understood.

Duffie et al. (2000) show that the affine dependence of the functions $\mu(x)$, $\sigma(x)\sigma(x)'$ and $\lambda(x)$ implies an exponential-affine form of the CCF of the state vector $X_t$. Specifically, the discounted joint CCF of $X_{t+\tau}$ conditional on $\mathcal{F}_t$ with $\tau > 0$ is given by

$$\psi_X(\mathbf{u}, \tau) := e^{-r\tau} \mathbb{E}^{\mathbb{Q}}[e^{\mathbf{i}\mathbf{u} \cdot X_{t+\tau}} | \mathcal{F}_t] = e^{\alpha(\mathbf{u}, \tau; \theta) + \beta(\mathbf{u}, \tau; \theta) \cdot X_t}, \tag{5}$$

where $\mathbf{u} \in \mathbb{R}^{d_X}$ is an argument vector and $\alpha(\mathbf{u}, \tau; \theta)$ and $\beta(\mathbf{u}, \tau; \theta)$ are solutions to the following complex-valued system of ordinary differential equations (ODEs) in time:

$$\begin{cases} \dot{\beta}(\mathbf{u}, s) = K_1' \beta(\mathbf{u}, s) + \frac{1}{2}\beta(\mathbf{u}, s)' H_1 \beta(\mathbf{u}, s) + \sum_{i=1}^{d_J} l_1^i (\chi^i(\beta(\mathbf{u}, s)) - 1), \\ \dot{\alpha}(\mathbf{u}, s) = K_0' \beta(\mathbf{u}, s) + \frac{1}{2}\beta(\mathbf{u}, s)' H_0 \beta(\mathbf{u}, s) + \sum_{i=1}^{d_J} l_0^i (\chi^i(\beta(\mathbf{u}, s)) - 1) - r, \end{cases} \tag{6}$$

with initial conditions $\beta(\mathbf{u}, 0) = \mathbf{i}\mathbf{u}$ and $\alpha(\mathbf{u}, 0) = 0$. Here, $\chi^i(c) = \int_{\mathbb{R}^n} \exp(c \cdot z) \mathrm{d}\nu^i(z)$, $c \in \mathbb{C}^{d_X}$, are jump transforms, which determine the conditional jump-size distributions. The ODE for $\beta$ is known as a generalized Riccati equation, whereas the solution for the second ODE can be obtained by simply integrating the right-hand side expression over time.

The affine dependence of the characteristic exponent $\alpha(\mathbf{u}, \tau; \theta) + \beta(\mathbf{u}, \tau; \theta) \cdot X_t$ on the current state $X_t$ is even the defining property of the AJD class under some regularity conditions (see Duffie et al., 2003). In other words, the AJD class can be defined as a class in which characteristic exponents of $X_{t+\tau}$ given $X_t$ are affine functions of $X_t$. In fact, this is a key property in our estimation procedure. While it is also possible to obtain the CCF for some non-affine models, the exponential-affine form allows us to use linear Kalman filtering techniques in the estimation procedure. This is the main motivation why we restrict our attention to the parametric models of the AJD class.[4]

Unlike the option-implied CCF (3), the CCF in (5) is fully parametric, that is, it requires parametric AJD model dynamics of the state vector $X_t$. Although the AJD class is more restrictive than the general dynamics of $F_t$ in (1), it includes a myriad of popular option pricing models such as those in Heston (1993), Duffie et al. (2000), Pan (2002), Bates (2006), Broadie et al. (2007), Boswijk et al. (2015), and Andersen et al. (2017) among many others.

The state process $X_t$ often includes both observed and unobserved state variables that affect the dynamics of the log futures price $\log F_t$. In our empirical application, we consider the presence of both. Therefore, it is convenient to partition the state vector as $X_t' = (w_t', x_t')$, where $w_t$ represents the observable component and $x_t$ includes $d < d_X$ latent state variables. Then, the dynamics of $X_t$ given by equation (4), can be rewritten as

$$\mathrm{d}w_t = \mu^w(w_t, x_t)\mathrm{d}t + \sigma^w(w_t, x_t)\mathrm{d}W_t + \sum_{i=1}^{d_J} J_{i,t}^w \mathrm{d}N_{i,t}, \tag{7}$$

$$\mathrm{d}x_t = \mu^x(w_t, x_t)\mathrm{d}t + \sigma^x(w_t, x_t)\mathrm{d}W_t + \sum_{i=1}^{d_J} J_{i,t}^x \mathrm{d}N_{i,t}, \tag{8}$$

where $\mu^w \colon D \to \mathbb{R}^{d_X - d}$, $\mu^x \colon D \to \mathbb{R}^d$, $\sigma^w \colon D \to \mathbb{R}^{(d_X - d) \times d_J}$, $\sigma^x \colon D \to \mathbb{R}^{d \times d_J}$ and $J_{i,t}^w$ and $J_{i,t}^x$ are marginal jump sizes of $J_{i,t}$ associated with $w_t$ and $x_t$, respectively. In the simplest case, the

---

[4]The considered AJD class could, in principle, be broadened further to the linear-quadratic jump-diffusion class by augmenting the state vector (see Cheng & Scaillet, 2007, for more details).

observable component includes only the log futures prices, that is, $w_t = \log F_t$. In more general settings, the stochastic volatility is often a main latent driver of the log returns dynamics, as e.g., in Heston (1993).

## 2.3  Marrying the two CCFs

Given the two CCFs (3) and (5), we can now align them to conduct inference about the model parameters and the unobservable state variables. For that purpose, first note that the CCF in (5) is joint for the state vector $X_t$. We assume, without loss of generality, that the first component of the state vector $X_t$ is the log futures price. Therefore, we can easily obtain its marginal CCF by plugging in an argument vector of the form $\mathbf{u_1} := (u, 0, \ldots, 0)' \in \mathbb{R}^{d_X}$ with $u \in \mathbb{R}$. To obtain the marginal CCF of log returns, we further subtract the term $iu \log F_t$ in the exponent. That is, the marginal CCF of log returns under the AJD specification is aligned to that in (3) as follows:

$$\phi_t(u, \tau) = \psi_X(\mathbf{u_1}, \tau)e^{-iu \log F_t} = e^{\alpha(\mathbf{u_1}, \tau; \theta) + \tilde{\beta}(\mathbf{u_1}, \tau; \theta) \cdot X_t}, \tag{9}$$

where $\tilde{\beta}(\mathbf{u_1}, \tau; \theta) := \beta(\mathbf{u_1}, \tau; \theta) - i\mathbf{u_1}$; i.e., the first component of $\tilde{\beta}(\mathbf{u_1}, \tau; \theta)$ differs from that of $\beta(\mathbf{u_1}, \tau; \theta)$, since we are interested in the CCF of log returns rather than that of log prices.

Note that the log of the (joint) CCF (also known as cumulant generating function) is linear in the state vector $X_t$. Therefore, under a correctly specified AJD model we obtain a simple linear relation between the log of the option-spanned CCF[5] of log returns and the model's state vector:

$$\log \phi_t(u, \tau) = \alpha(\mathbf{u_1}, \tau; \theta) + \tilde{\beta}(\mathbf{u_1}, \tau; \theta) \cdot X_t. \tag{10}$$

Replacing the cumulant generating function on the left-hand side with its computationally feasible counterpart $\widehat{\phi}_t(u, \tau)$, which we will explicitly define in Section 3.1, we obtain the following equation, which will play a central role in our estimation procedure:

$$\log \widehat{\phi}_t(u, \tau) = \alpha(\mathbf{u_1}, \tau; \theta) + \tilde{\beta}(\mathbf{u_1}, \tau; \theta) \cdot X_t + \xi_t(u, \tau), \qquad u \in \mathbb{R}. \tag{11}$$

Here, $\xi_t(u, \tau)$ is the measurement error, which is related to the observation, truncation and discretization errors in the CCF-spanning option portfolios. We elaborate in detail on the relation between the computable counterpart of the CCF and the source of the measurement errors in the next section.

Equation (11) is the key relation in our analysis and a few remarks shall be made here regarding it. First, (11) is essentially a *functional* linear model since this equation holds for any argument variable of the CCF, $u \in \mathbb{R}$. Furthermore, the functions $\alpha(\mathbf{u_1}, \tau; \theta)$ and $\tilde{\beta}(\mathbf{u_1}, \tau; \theta)$ are parameter-dependent and solutions to the system of Riccati ODEs (6). Therefore, if the state vector $X_t$ is observable, then the model parameters can be estimated by solving a continuum version of a non-linear least-squares problem.

---

[5]Although the logarithm of a complex number is a multivalued function, here, the ambiguity is resolved given the fact that $\phi(0) = 1$ and the CCF is a continuous function. In fact, in practice we ensure that the logarithm of the CCF does not have 'jumps' by taking the logarithm sequentially with respect to $u$, starting from the origin.

Second, in the case in which the state vector is (partially) unobservable, (11) represents a linear latent factor model with a continuum of linear relations. The factors are given by the state components of the AJD model. Therefore, one could apply, e.g., a (functional) principal component analysis to learn about the unobserved factors. In this paper, we utilize a (suitably modified) Kalman filtering technique to conduct inference about the model parameters and the latent factors.

In other words, (11) reveals that, using the present approach, AJD models become amenable to filtering and estimation using approaches from the rich literature on linear factor and state space models. This is reminiscent of the term structure literature, where in affine term structure models (see Piazzesi, 2010, for a review of this class of models) bond yields themselves are assumed to be linear functions of the state vector. For instance, Duffee (1999), de Jong (2000), Driessen (2005) use the Kalman filter in their estimation of affine term structure models.

Furthermore, another advantage of this approach is that it does not require evaluating option prices given a certain parametric model. Therefore, our estimation procedure is computationally more appealing than many alternative approaches, which often involve the Carr-Madan FFT pricer (Carr & Madan, 1999) or the COS method (Fang & Oosterlee, 2008) to price options. This also implies that the usage of the characteristic function is different: with the FFT or COS methods one needs a model-dependent CCF only to evaluate option prices, while here we use the CCF to directly learn about the latent factors and the model parameters.

Finally, given the partition of the state vector into observable and unobservable components, the linear relation between the option-implied and model-implied CCFs in (11) can be rewritten as

$$
\log \widehat{\phi}_t(u, \tau) = \alpha(\mathbf{u_1}, \tau; \theta) + \beta^w(\mathbf{u_1}, \tau; \theta) \cdot w_t + \beta^x(\mathbf{u_1}, \tau; \theta) \cdot x_t + \xi_t(u, \tau), \qquad u \in \mathbb{R}, \qquad (12)
$$

where $\beta^w(\mathbf{u_1}, \tau; \theta) \in \mathbb{C}^{d_X - d}$ and $\beta^x(\mathbf{u_1}, \tau; \theta) \in \mathbb{C}^d$ are such that $\tilde{\beta}' = (\beta^{w\prime}, \beta^{x\prime})$ is the solution to the ODE system (6). Representation (12) serves as the basis for an observation (or measurement) equation in our estimation procedure.

# 3    Estimation Procedure

In this section, we develop our filtering approach and corresponding estimation procedure for the general class of AJD models under consideration. First, we provide the formal state space representation for the defined class of models. Then, we describe our estimation strategy, which uses the collapsed Kalman filter.

## 3.1    State space representation

As discussed in the previous section, we restrict our attention to the parametric models of the AJD class due to their exponential-affine form of the characteristic function. This form will allow us to exploit a linear Kalman filter in the estimation procedure. In the following, we summarize the assumptions we impose on the parametric model:

**Assumption 1**  (i) *The stochastic process $X_t$ is Markov and affine, with finite second moments under both the physical and risk-neutral probability measures $\mathbb{P}$ and $\mathbb{Q}$. In particular, $X_t$ is the unique solution to the SDE* (4) *and its characteristic function is of the exponential-affine form* (5)*;*

(ii) *The true parameter vector $\theta_0$ lies in the interior of a compact parameter space $\Theta$ containing admissible parameter values.*

Assumption 1 guarantees the existence of a unique solution to the SDE (4) within the AJD class. As discussed in Section 2.2, admissible values $\theta \in \Theta$ reflect the regularity conditions imposed on the model such that there is a unique solution to (4), with, e.g., non-negative volatilities and jump intensities. Such admissibility conditions will need to be checked in a case-by-case model analysis. Assumption 1(*i*) also presumes the technical conditions required to represent the AJD process, defined via the affine dependence of its drift, diffusive variance and jump intensities on the state vector, through the exponential-affine characteristic function. For a detailed analysis of the AJD theory, we refer to Duffie et al. (2000) and Duffie et al. (2003). Note that Assumption 1 does not require the state process to be stationary. Stationarity of the latent state variables $x_t$ is reasonable but not essential for the results to follow; the observed state variables $w_t$ (often including the log-forward price) are typically non-stationary.

In our estimation procedure, we discretize the continuous-time model along two dimensions: with respect to time and with respect to the argument of the CCF. The former naturally follows from the discrete sampling times of financial data, which we denote by the integer indices $t = 1, \ldots, T$. The latter allows us to rely on the existing literature about filtering techniques. For that, let us denote the collection of discretely sampled arguments by a set $\mathcal{U} \subseteq \mathbb{R}$ with cardinality $q \in \mathbb{N}$. We further consider options with $k \in \mathbb{N}$ different maturities $\tau$ and $n \in \mathbb{N}$ different log-moneyness values $m$ on each day.

Since the input of our estimation procedure is a portfolio of option prices, we need to take into account the measurement errors in these option portfolios. For that purpose, we assume an observation error scheme on the option prices that constitute the portfolios. The measurement errors will be defined on the common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, but in what follows, the filtration $\{\mathcal{F}_t\}_{t \geq 0}$ is generated by the state process $\{X_t\}_{t \geq 0}$ only. Note that the theoretical option prices $O_t(\tau, m)$ are $\mathcal{F}_t$-measurable, and hence the same applies to functionals of the option prices such as the (theoretical) Black-Scholes implied volatility (BSIV) and vega.

**Assumption 2** *Option prices are observed with an additive error term:*

$$\widehat{O}_t(\tau_i, m_j) := O_t(\tau_i, m_j) + \zeta_t(\tau_i, m_j), \qquad t = 1, \ldots, T, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n, \quad (13)$$

*where the observation errors $\zeta_t(\tau, m)$ are such that:*

(i) *$\zeta_t(\tau, m)$ are $\mathcal{F}_t$-conditionally independent along tenors $\tau$, moneyness $m$ and time $t$;*

(ii) *$\mathbb{E}[\zeta_t(\tau, m)|\mathcal{F}_t] = 0$;*

(iii) *$\mathbb{E}[\zeta_t(\tau, m)^2|\mathcal{F}_t] = \sigma_t^2(\tau, m) < \infty$ with $\sigma_t(\tau, m) := \sigma_\varkappa \kappa_t(\tau, m) \nu_t(\tau, m)$, where $\sigma_\varkappa \in \mathbb{R}^+$, $\kappa_t(\tau, m)$ is the Black-Scholes implied volatility, and $\nu_t(\tau, m)$ is the Black-Scholes vega.*

The additive error assumption is commonly imposed in the option pricing literature. For instance, Andersen et al. (2015a) and Todorov (2019) use additive error assumptions for option prices quoted in terms of BSIV and dollar amount, respectively. Additive observation errors are also often implicitly assumed when calibrating an option pricing model to market-observed prices, since the calibration is often performed using non-linear least squares as in, e.g., Broadie et al. (2007).

Assumption $2(i)$ excludes in particular dependence of the observation errors across strikes and is also often imposed in the literature (see, for instance, Christoffersen, Jacobs, & Mimouni, 2010, Andersen et al., 2015a and Todorov, 2019). This assumption can be relaxed by introducing a spatial dependence as in Andersen, Fusari, Todorov, and Varneskov (2021). This would, however, result in more complex expressions for the covariance terms in the measurement errors that we derive below. Furthermore, Andersen et al. (2021) find evidence of limited dependence in the observation errors for S&P 500 index options. They also show that this dependence declined sharply for short-dated options in recent years, due to improved liquidity. Since in our empirical application we consider S&P 500 index options with short tenors focusing on the past three years, the independence assumption will play a secondary role for the estimation procedure.

The conditional mean zero Assumption $2(ii)$ is crucial for our first main result. Assumption $2(iii)$ asserts the standard deviation of the observation errors to be proportional to the product of the option's BSIV and vega. The motivation for this structure is as follows. Let $\widehat{\kappa}(m_j)$ and $\kappa(m_j)$ denote the error-distorted and true BSIV of an option, and assume that the relative volatility errors $\varkappa_j = (\widehat{\kappa}(m_j) - \kappa(m_j))/\kappa(m_j)$ are homoskedastic across the strikes, such that $\mathbb{E}[\varkappa_j^2|\mathcal{F}_t] = \sigma_\varkappa^2$. A Taylor-series expansion of the Black-Scholes pricing function $O^{BS}(\widehat{\kappa}(m_j), m_j)$ around $\kappa(m_j)$ then gives $\widehat{O}(m_j) = O^{BS}(\widehat{\kappa}(m_j), m_j) \approx O(m_j) + \nu(m_j)\kappa(m_j)\varkappa_j$, with $\nu(m_j) = \partial O^{BS}(\kappa(m_j), m_j)/\partial \kappa(m_j)$ the theoretical Black-Scholes vega. Homoskedastic errors in relative implied volatilities are also assumed by Christoffersen, Jacobs, and Ornthanalai (2012) and Du and Luo (2019) in their MLE based on the particle filter and the unscented Kalman filter, respectively.

Finally, to assess the error sizes of the CCF approximation specified below, we impose the following assumption on the existence of moments for the underlying asset and on the log-moneyness grid that allows nonequidistant sampling in the moneyness dimension:

**Assumption 3** *(i) The underlying process and its reciprocal process have finite $1 + \overline{\delta}$ and $\underline{\delta}$ moments under the risk-neutral measure, respectively, i.e., $\mathbb{E}^{\mathbb{Q}}[F_{t+\tau}^{1+\overline{\delta}}|\mathcal{F}_t] < \infty$ and $\mathbb{E}^{\mathbb{Q}}[F_{t+\tau}^{-\underline{\delta}}|\mathcal{F}_t] < \infty$ for some $\overline{\delta} > 0$ and $\underline{\delta} > 0$ and with $\tau > 0$;*

*(ii) For the log-moneyness grid $\underline{m} := m_1 < \ldots < m_n =: \overline{m}$, there exists a deterministic sequence $\Delta m$ depending on $n$ such that $\Delta m \to 0$ as $n \to \infty$ and*

$$\eta \Delta m \leq \inf_{j=2,\ldots,n} \Delta m_j \leq \sup_{j=2,\ldots,n} \Delta m_j \leq \Delta m,$$

*where $\Delta m_j := m_j - m_{j-1}$ and $\eta \in (0, 1]$ is some constant.*

Assumption 3 (*i*) requires the risk-neutral conditional moment generating function of $\log F_{t+\tau}$ to be finite when evaluated in $u = 1+\overline{\delta}$ and $u = -\underline{\delta}$ for some $\overline{\delta}, \underline{\delta} > 0$. A significant literature on the long-term behavior and potential moment explosion of affine stochastic volatility processes shows that additional restrictions may be required to ensure that $\sup\{\tau : \mathbb{E}\left[F_{t+\tau}^u | \mathcal{F}_t\right] < \infty\} = \infty$, for $u > 1$ and $u < 0$, even in the setting of a simple Heston model; see, e.g., Andersen and Piterbarg (2007), Keller-Ressel (2011) and Keller-Ressel and Mayerhofer (2015). We note, however, that we will choose $\tau$ to be 'small'. Specifically, it never exceeds $\tau = 0.25$, both in the Monte Carlo simulations and the empirical applications that follow. When $\tau$ is small and $u$ is moderate, say $u = \pm 2$, typical affine stochastic volatility models are far from moment explosion in realistic settings.[6] Furthermore, more formally, the order of finite conditional moments of $F_{t+\tau}$ determines the convergence rate of the truncation errors in our asymptotic scheme. As we establish explicitly in Lemma 2 in the appendix and Propositions 1 and 2 below, the higher the order of existing moments, the faster the truncation errors converge to zero. That is, even if certain moments of $F_{t+\tau}$ would not exist for some fixed $\tau$, this only affects the convergence rate of the truncation errors, as a function of the number $n$ of option contracts. Hence, as soon as the observation errors dominate the truncation errors, our results below are valid; see also the discussion below Proposition 1.

Using $n > 1$ observable option prices with time-to-maturity $\tau > 0$ and log-moneyness values $\{m_j\}_{j=1}^n$, we may approximate the CCF $\phi_t(u, \tau)$ given in (3) by replacing the theoretical option prices by their observed counterparts, and the integral by a Riemann sum:

$$\widehat{\phi}_t(u, \tau) = e^{-r\tau} - u_t \sum_{j=2}^n e^{(\mathrm{i}u-1)m_j} \cdot \widehat{O}_t(\tau, m_j) \Delta m_j, \qquad (14)$$

where we use the notation $u_t := (u^2 + \mathrm{i}u)/F_t$, and where $\widehat{O}_t(\tau, m_j)$ satisfies Assumption 2.

The deviation of the option-spanned CCF from its theoretical counterpart, $\zeta_t^\phi(u, \tau) := \widehat{\phi}_t(u, \tau) - \phi_t(u, \tau)$, stems from *observation*, *truncation* and *discretization* errors, where truncation refers to the fact that the integration interval $[\underline{m}, \overline{m}]$ does not cover the entire real line. The truncation and discretization errors also arise in VIX calculations and depend on the availability of option prices. They will be shown to be of smaller order than the observation errors, and can further be efficiently reduced by using an interpolation-extrapolation scheme (see, e.g., Jiang & Tian, 2005, 2007, Chang, Christoffersen, Jacobs, & Vainberg, 2012, and Appendix C.1). Appendix C.2 illustrates the impact of the three different types of measurement errors on the CCF approximation, and the effectiveness of the interpolation-extrapolation scheme.[7]

---

[6]For example, as illustrated in Figure 2 of Keller-Ressel (2011), even a ten-fold increase in our $\tau$ would not lead to second-order moment explosion in models of the Heston type considered there. We have also analyzed the sensitivity of the non-explosion region to changes in the parameter values of Keller-Ressel's Figure 2 (which are taken from Gatheral, 2006) and find that even when doubling the vol-of-vol parameter and taking the leverage parameter to be $-1$, one does not get close to explosion of, e.g., $\mathbb{E}\left[F_{t+\tau}^{\pm 2} | \mathcal{F}_t\right]$ for values of $\tau$ smaller than unity.

[7]The interpolation-extrapolation scheme may induce some cross-sectional dependence in the observation errors $\zeta_t(\tau, m)$. This is in deviation from Assumption 2, which is only realistic when referring to the errors before application of the interpolation-extrapolation scheme. We will not consider this effect explicitly in Proposition 1 that follows; it would lead to a more complicated expression for the covariance matrix of the measurement errors, but, importantly, would not affect the first main result otherwise.

From the preceding analysis, the functional measurement equation (12) is then obtained using the following log-linearization:

$$\xi_t(u,\tau) := \log \widehat{\phi}_t(u,\tau) - \log \phi_t(u,\tau) = \log\left(1 + \frac{\zeta_t^\phi(u,\tau)}{\phi_t(u,\tau)}\right) = \xi_t^{(1)}(u,\tau) + r_t(u,\tau), \qquad (15)$$

where the log-linearized observation errors $\xi_t^{(1)}(u,\tau)$ are defined by $\zeta_t^{(1)}(u,\tau)/\phi_t(u,\tau)$, with

$$\zeta_t^{(1)}(u,\tau) := -u_t \sum_{j=2}^{n} e^{(iu-1)m_j} \cdot \zeta_t(\tau,m_j)\Delta m_j,$$

and where $r_t(u,\tau)$ is a remainder term that collects the log-linearized truncation and discretization errors as well as the higher-order terms from the required Taylor-series expansion. (The superscript $^{(1)}$ refers to the first, and prime, source of the measurement errors, the observation errors; see also the detailed decomposition in equation (A.1).)

To formulate the first result, we turn the complex-valued functional measurement equation (12) into a real vector measurement equation, as usual in state space model formulations. First, we stack the log CCF and the corresponding measurement errors along $q$ values $u_1,\dots,u_q$ for the CCF argument $u \in \mathcal{U}$, for a fixed expiration period $\tau_i$:

$$\log \widehat{\phi}_{t,i} := \begin{pmatrix} \log \widehat{\phi}_t(u_1,\tau_i) \\ \log \widehat{\phi}_t(u_2,\tau_i) \\ \vdots \\ \log \widehat{\phi}_t(u_q,\tau_i) \end{pmatrix}, \quad r_{t,i} := \begin{pmatrix} r_t(u_1,\tau_i) \\ r_t(u_2,\tau_i) \\ \vdots \\ r_t(u_q,\tau_i) \end{pmatrix}, \quad \xi_{t,i}^{(1)} := \begin{pmatrix} \xi_t^{(1)}(u_1,\tau_i) \\ \xi_t^{(1)}(u_2,\tau_i) \\ \vdots \\ \xi_t^{(1)}(u_q,\tau_i) \end{pmatrix}.$$

In a similar way, we denote by $a_{t,i}$, $b_{t,i}^w$ and $b_{t,i}^x$ the stacked outputs[8] of the functions $\alpha(u,\tau_i)$, $\beta^w(u,\tau_i)$ and $\beta^x(u,\tau_i)$, respectively. Next, to tackle the complex-valued measurement equation (12), we stack the real and imaginary parts, as well as $k$ maturities:

$$\underbrace{\begin{pmatrix} \Re(\log \widehat{\phi}_{t,1}) \\ \Im(\log \widehat{\phi}_{t,1}) \\ \vdots \\ \Re(\log \widehat{\phi}_{t,k}) \\ \Im(\log \widehat{\phi}_{t,k}) \end{pmatrix}}_{=: \, y_t \in \mathbb{R}^p} = \underbrace{\begin{pmatrix} \Re(a_{t,1}) \\ \Im(a_{t,1}) \\ \vdots \\ \Re(a_{t,k}) \\ \Im(a_{t,k}) \end{pmatrix}}_{=: \, \tilde{d}_t} + \underbrace{\begin{pmatrix} \Re(b_{t,1}^w) \\ \Im(b_{t,1}^w) \\ \vdots \\ \Re(b_{t,k}^w) \\ \Im(b_{t,k}^w) \end{pmatrix}}_{=: \, W_t} w_t + \underbrace{\begin{pmatrix} \Re(b_{t,1}^x) \\ \Im(b_{t,1}^x) \\ \vdots \\ \Re(b_{t,k}^x) \\ \Im(b_{t,k}^x) \end{pmatrix}}_{=: \, Z_t} x_t + \underbrace{\begin{pmatrix} \Re(r_{t,1}) \\ \Im(r_{t,1}) \\ \vdots \\ \Re(r_{t,k}) \\ \Im(r_{t,k}) \end{pmatrix}}_{=: \, r_{t,n}} + \underbrace{\begin{pmatrix} \Re(\xi_{t,1}^{(1)}) \\ \Im(\xi_{t,1}^{(1)}) \\ \vdots \\ \Re(\xi_{t,k}^{(1)}) \\ \Im(\xi_{t,k}^{(1)}) \end{pmatrix}}_{=: \, \varepsilon_t}, \quad (16)$$

where $p = 2qk$. Stacking the real and imaginary parts of the measurements is a natural approach when the state vector is real-valued;[9] a complex-valued state vector would have required a complex Kalman filter based on the so-called widely linear complex estimator, as in Dini and Mandic (2012). The stacked observation equation (16) links all available information from option prices with several tenors at time $t$ to the state vectors $w_t$ and $x_t$ in a linear way.

---

[8]Here, we attribute these elements (and system matrices $\tilde{d}_t, W_t$ and $Z_t$ in equation (16)) with an additional time index although the coefficient functions are assumed to be time-invariant in the exposition. This is because in practice we can have different expiration periods for different days.

[9]See Singleton (2001) and Chacko and Viceira (2003), who use this approach in a GMM estimation setting based on the empirical characteristic function.

To complete the state space model, we need to augment the measurement equation (16) by a transition equation for the unobservable state vector $x_t$. This is a linear, discrete-time dynamic system, to be derived from the continuous-time stochastic differential equation. An Euler discretization of the state process (8) would converge to the true transition dynamics as the discretization step $\Delta t \to 0$. However, the maximum likelihood (ML) estimator based on the Euler discretization is, in general, inconsistent for fixed non-zero $\Delta t$ (Lo, 1988), because the discretization has conditional moments different from those of the true process (Piazzesi, 2010). Fortunately, the AJD assumption under $\mathbb{P}$ implies that the first and second conditional moments of $x_{t+1}$ given $\mathcal{F}_t$ are linear and available in semi-closed form (possibly requiring the solution of a system of ODEs):

$$\mathbb{E}[x_{t+1}|\mathcal{F}_t] = c_t + T_t x_t, \tag{17}$$

$$\mathrm{Var}(x_{t+1}|\mathcal{F}_t) = Q_t(x_t), \tag{18}$$

where $Q_t \colon \mathbb{R}^d \to \mathbb{R}^{d \times d}$ is an affine function in $x_t$. The finiteness of the conditional moments is ensured by Assumption 1(i). Both conditional moments will in general be linear in both the observed state $w_t$ and the latent state $x_t$; but because the former does not need filtering, we absorb its effect in the time-varying intercept $c_t$, and similarly in the intercept of the affine function $Q_t$.[10]

In Appendix B, we show how these transition coefficients can be computed for the AJD model. Using this approach, which will in principle be model-dependent and hence has to be applied case by case, we obtain a discrete-time transition equation with the same conditional mean and variance as the true continuous-time process (but possibly different higher-order moments). Quasi-maximum likelihood (QML) estimation based on conditionally normally distributed measurement and transition errors in the state space representation yields consistent estimation results (Fisher & Gilles, 1996). A similar approach has been adopted in the term structure literature (see, e.g., de Jong, 2000, Duffee, 2002).

We summarize the development of the state space representation, and analyze properties of the errors, in the following proposition. Here, we arbitrarily fix $\mathcal{U}$ and its cardinality $q$. The first main result contains a remainder term in the measurement equation that collects the truncation and discretization errors in the construction of $\log \widehat{\phi}(u, \tau)$ and higher-order terms in the log-linearization. This term vanishes under an asymptotic scheme, where $\overline{m} = \max_{1 \le j \le n} m_j \to \infty$, $\underline{m} = \min_{1 \le j \le n} m_j \to -\infty$ and $\Delta m \to 0$. We also denote the corresponding smallest and largest strike prices by $\underline{K}$ and $\overline{K}$, and express the asymptotic orders with respect to the number of option prices $n$ with fixed maturity.

**Proposition 1** *Suppose Assumptions 1, 2 and 3 hold, and in addition $\underline{K} \asymp n^{-\underline{\alpha}}$ and $\overline{K} \asymp n^{\overline{\alpha}}$ with $\underline{\alpha} > 0$ and $\overline{\alpha} > 0$. Then $\{(y_t, x_t), t = 1, \ldots, T\}$ satisfy the linear state space representation*

$$y_t = d_t + Z_t x_t + r_{t,n} + \varepsilon_t, \qquad \mathbb{E}[\varepsilon_t|\mathcal{F}_t] = 0, \qquad \mathbb{E}[\varepsilon_t \varepsilon_t'|\mathcal{F}_t] = H_t, \tag{19}$$

$$x_{t+1} = c_t + T_t x_t + \eta_{t+1}, \qquad \mathbb{E}[\eta_{t+1}|\mathcal{F}_t] = 0, \qquad \mathbb{E}[\eta_{t+1} \eta_{t+1}'|\mathcal{F}_t] = Q_t(x_t), \tag{20}$$

---

[10]The transition matrix $T_t$ will not be time-varying in stationary AJD processes with equidistant observations, but we do not impose this time-constancy in the notation, also to avoid confusion with the sample size $T$.

*where $r_{t,n} = \mathcal{O}_p\left(n^{-(\underline{\delta}\,\underline{\alpha}\wedge(1+\bar{\delta})\bar{\alpha})} \vee n^{-1}\log n\right)$ and $\varepsilon_t = \mathcal{O}_p\left(\sqrt{n^{-1}\log n}\right)$; $d_t = \tilde{d}_t + W_t w_t$ and $Z_t$ are defined in (16) and $c_t$, $T_t$ and $Q_t$ are as given in (17)–(18); and $H_t = \mathrm{blkdiag}\{H_{t,1},\ldots,H_{t,k}\}$, with $H_{t,i} = \sigma_{\varkappa}^2 \cdot \widetilde{H}_{t,i}$, where*

$$\widetilde{H}_{t,i} = \begin{pmatrix} \frac{1}{2}\Re(\widetilde{\Gamma}_{t,i} + \widetilde{C}_{t,i}) & \frac{1}{2}\Im(-\tilde{\Gamma}_{t,i} + \widetilde{C}_{t,i}) \\ \frac{1}{2}\Im(\widetilde{\Gamma}_{t,i} + \widetilde{C}_{t,i}) & \frac{1}{2}\Re(\widetilde{\Gamma}_{t,i} - \widetilde{C}_{t,i}) \end{pmatrix}, \quad i = 1,\ldots,k, \tag{21}$$

*and $\widetilde{\Gamma}_{t,i}$ and $\widetilde{C}_{t,i}$ are covariance and pseudo-covariance matrices of $\xi_{i,t}/\sigma_{\varkappa}$, with elements*

$$(\widetilde{\Gamma}_{t,i})_{kl} = \frac{u_{k,t}\overline{u_{l,t}}\sum_{j=2}^{n} e^{(\mathrm{i}(u_k-u_l)-2)m_j}\kappa_t^2(\tau_i,m_j)\nu_t^2(\tau_i,m_j)(\Delta m_j)^2}{\phi_t(u_k,\tau_i)\phi_t(-u_l,\tau_i)}, \quad k,l = 1,\ldots,q,$$

$$(\widetilde{C}_{t,i})_{kl} = \frac{u_{k,t}u_{l,t}\sum_{j=2}^{n} e^{(\mathrm{i}(u_k+u_l)-2)m_j}\kappa_t^2(\tau_i,m_j)\nu_t^2(\tau_i,m_j)(\Delta m_j)^2}{\phi_t(u_k,\tau_i)\phi_t(u_l,\tau_i)}, \quad k,l = 1,\ldots,q.$$

*Furthermore,*

*(i) $\mathbb{E}[\varepsilon_t\varepsilon_s'] = 0$ and $\mathbb{E}[\eta_t\eta_s'] = 0$ for $s \neq t = 1,\ldots,T$;*

*(ii) $\mathbb{E}[\varepsilon_t\eta_s'] = 0$ for all $s,t = 1,\ldots,T$;*

*(iii) $\mathbb{E}[\varepsilon_t x_1'] = 0$ and $\mathbb{E}[\eta_{t+1}x_1'] = 0$ for $t = 1,\ldots,T$.*

The proof is given in Appendix A. The orders indicate that the remainder term goes to zero faster than the observation term given some minimum non-zero requirements for $\underline{\alpha}$ and $\bar{\alpha}$. In the sequel, we assume that $(\underline{\alpha} \wedge \bar{\alpha}) > 1/(2(\underline{\delta} \wedge (1+\bar{\delta})))$ and neglect the remainder term in the estimation and filtering procedures. The system matrices $Z_t, T_t, Q_t(x_t)$ and system vectors $d_t$ and $c_t$ are known up to a parameter vector $\theta$, assumed to lie in the interior of a compact parameter space $\Theta$ by Assumption 1(ii). Similarly, the system matrix $H_t$ depends on the data and $\theta$ (via $u_t, \phi_t, \kappa_t$ and $\nu_t$), and an additional unknown parameter $\sigma_{\varkappa}^2$. Note that $(d_t, Z_t, H_t)$ are derived from the $\mathbb{Q}$-dynamics of (4), whereas $(c_t, T_t, Q_t(\cdot))$ correspond to the $\mathbb{P}$-dynamics. Therefore, as we establish in Proposition 2, possible deviations between $\mathbb{P}$ and $\mathbb{Q}$, reflecting the presence of factor risk premia, will lead to different convergence rates of the components of the parameter vector; we discuss this, and corresponding identification challenges in finite samples, further in Section 4 and Appendix D. Estimation of $\theta$ and filtering of the latent state vector via (versions of) the Kalman filter is considered in the next sub-section.

## 3.2 Modified and collapsed Kalman filter

Consider the state space representation (19)–(20), where from now on we will ignore the remainder term $r_{t,n}$, and hence assume that the set of strike prices $\{m_j\}_{j=1}^n$ on each day is rich enough to make this term negligible. Define the dataset $Y_t = \{y_1,\ldots,y_t\}$, and linear projections (denoted by $\widehat{\mathbb{E}}$) of the latent state vector conditional on the data: $\widehat{x}_{t|t} = \widehat{\mathbb{E}}[x_t|Y_t]$ and $\widehat{x}_{t|t-1} = \widehat{\mathbb{E}}[x_t|Y_{t-1}]$, with corresponding mean square error matrices $P_{t|t} = \mathbb{E}[(x_t-\widehat{x}_{t|t})(x_t-\widehat{x}_{t|t})']$ and $P_{t|t-1} = \mathbb{E}[(x_t-\widehat{x}_{t|t-1})(x_t-\widehat{x}_{t|t-1})']$. Then a modified version of the Kalman filter reads

as follows:

$$\omega_t = y_t - (d_t + Z_t \widehat{x}_{t|t-1}), \qquad\qquad G_t = Z_t P_{t|t-1} Z_t' + H_t,$$
$$\widehat{x}_{t|t} = \widehat{x}_{t|t-1} + P_{t|t-1} Z_t' G_t^{-1} \omega_t, \qquad\qquad P_{t|t} = P_{t|t-1} - P_{t|t-1} Z_t' G_t^{-1} Z_t P_{t|t-1},$$
$$\widehat{x}_{t+1|t} = c_t + T_t \widehat{x}_{t|t}, \qquad\qquad P_{t+1|t} = T_t P_{t|t} T_t' + Q_t(\widehat{x}_{t|t}),$$

for $t = 1, \ldots, T$. If the latent state process $x_t$ is stationary, the initial conditions $\widehat{x}_{1|0}$ and $P_{1|0}$ for the filter can be set to the unconditional mean and variance, respectively. Indeed, the unconditional mean and variance are the standard choice for the initialization of the Kalman filter (see, e.g., Durbin & Koopman, 2012). For stationary latent processes, the initial conditions are theoretically irrelevant.

In traditional homoskedastic Gaussian state space models, where the distribution of the vector $(\varepsilon_t', \eta_{t+1}')'$, conditional on $x_t$, is Gaussian with a constant variance matrix, the filtered state $\widehat{x}_{t|t}$ is the conditional expectation of the true process $x_t$ given the observations up to time $t$. When the errors are non-Gaussian homoskedastic, the filtered state represents the linear projection (or minimum mean square error linear predictor) instead of the conditional expectation. This property can be used to prove that QML estimation based on the Gaussian likelihood still yields consistent and asymptotically normal parameter estimates (Hamilton, 1994, Chapter 13). In general AJD models, on the other hand, the distribution of the errors will be non-Gaussian with a conditional variance $Q_t(x_t)$ that is an affine function of the true latent state vector $x_t$. Therefore, the Kalman filter recursions have been modified by using $Q_t(\widehat{x}_{t|t})$ instead of the unobserved $Q_t(x_t)$. A similar modification is used in, e.g., de Jong (2000), Monfort, Pegoraro, Renne, and Roussellet (2017) and Feunou and Okou (2018). Although consistency of QML based on this modification has not been proved, Monte Carlo simulation results in these articles suggest that the method works well in practice. Below, we will provide an asymptotic result for QML estimation under double asymptotics ($n, T \to \infty$), i.e., under vanishing measurement errors.

Given the large dimension of the observation vector $p = 2qk$, the Kalman filter and its QML estimation will be computationally challenging if not infeasible. In fact, an important caveat with this approach is that one needs a non-singular innovation variance matrix $G_t$. Since our CCF approximation is based on common option price data for $q$ different arguments $u$ and fixed time-to-maturity $\tau$, this matrix is likely to be (near-)singular for large $q$. Furthermore, with large cross-sectional dimension, the computation of the inverse matrix for each time $t$ adds a significant computational burden to the estimation procedure. To overcome these issues, we consider the collapsed Kalman filter, originally developed by Jungbacker and Koopman (2015), which we describe below. We modify their method to allow for a (near-)singular variance matrix $H_t$, using generalized inverses.

In particular, let the $p \times p$ matrix $H_t$ have rank $r \leq p$, and let $H_t = S\Lambda S'$ be its spectral decomposition, such that $S'S = I_r$, and $\Lambda$ is a diagonal matrix with the non-zero eigenvalues of $H_t$ on the diagonal. Then the Moore-Penrose pseudo-inverse is $H_t^- = S\Lambda^{-1}S'$, and the pseudo-determinant is $|H_t|_- = |\Lambda|$. We will also use the symmetric square root $H_t^{1/2} = S\Lambda^{1/2}S'$ and its pseudo-inverse $H_t^{-1/2} = S\Lambda^{-1/2}S'$.

The idea of the collapsed Kalman filter is to transform the observation vector $y_t$ into an uncorrelated pair of vectors $y_t^*$ and $y_t^+$ such that $y_t^*$ depends on the state vector $x_t$ and has dimension $d \times 1$, whereas $y_t^+$ does not depend on $x_t$ and has dimension $(p-d) \times 1$. Such a transformation can be done using the matrices $A_t^* = (Z_t' H_t^- Z_t)^{-1} Z_t' H_t^-$ and $A_t^+ = (H_t^{-1/2} Z_t)_\perp' H_t^{-1/2}$, of dimensions $d \times p$ and $(p-d) \times p$, respectively. These definitions assume that $\mathrm{rank}(Z_t' H_t^- Z_t) = d$; this is not very restrictive, given that the dimension $d$ of the state vector will typically be much smaller than the dimension $p$ of the observation vector. Since $A_t^* Z_t = I_d$ and $A_t^+ Z_t = 0$, the observation equation is then transformed into

$$\begin{pmatrix} y_t^* \\ y_t^+ \end{pmatrix} := \begin{bmatrix} A_t^* \\ A_t^+ \end{bmatrix} y_t = \begin{pmatrix} d_t^* \\ d_t^+ \end{pmatrix} + \begin{pmatrix} x_t \\ 0 \end{pmatrix} + \begin{pmatrix} \varepsilon_t^* \\ \varepsilon_t^+ \end{pmatrix}, \tag{22}$$

with $d_t^* = A_t^* d_t$, $d_t^+ = A_t^+ d_t$, $\varepsilon_t^* = A_t^* \varepsilon_t$ and $\varepsilon_t^+ = A_t^+ \varepsilon_t$. Using $H_t^- H_t H_t^- = H_t^-$ and $H_t^- H_t H_t^{-1/2} = H_t^{-1/2}$, we have

$$\mathrm{Var}(\varepsilon_t^*) = A_t^* H_t A_t^{*\prime} = (Z_t' H_t^- Z_t)^{-1} =: H_t^*,$$
$$\mathrm{Var}(\varepsilon_t^+) = A_t^+ H_t A_t^{+\prime} =: H_t^+,$$
$$\mathrm{Cov}(\varepsilon_t^*, \varepsilon_t^+) = A_t^* H_t A_t^{+\prime} = (Z_t' H_t^- Z_t)^{-1} Z_t' H_t^- H_t H_t^{-1/2} (H_t^{-1/2} Z_t)_\perp = 0.$$

The representation (22) shows that information about the state vector $x_t$ is contained in the observation equation for $y_t^*$; thus we may ignore the second equation with $y_t^+$ and focus on the collapsed state space model:

$$y_t^* = d_t^* + x_t + \varepsilon_t^*, \qquad \mathbb{E}[\varepsilon_t^* | \mathcal{F}_t] = 0, \ \mathrm{Var}(\varepsilon_t^* | \mathcal{F}_t) = H_t^*, \tag{23}$$
$$x_{t+1} = c_t + T_t x_t + \eta_{t+1}, \qquad \mathbb{E}[\eta_{t+1} | \mathcal{F}_t] = 0, \ \mathrm{Var}(\eta_{t+1} | \mathcal{F}_t) = Q_t(x_t). \tag{24}$$

It should be noted that the transformation matrix $A_t := [A_t^{*\prime}, A_t^{+\prime}]'$ is singular, but can be written as the product of a non-singular matrix and $H_t^{-1/2}$. Thus, when applied to $\varepsilon_t$, it preserves the degeneracy in the distribution of $\varepsilon_t$ (the singularity of its variance matrix): $\mathrm{Var}(A_t \varepsilon_t) = \mathrm{diag}(H_t^*, H_t^+)$ is of the same rank $r$ as $H_t = \mathrm{Var}(\varepsilon_t)$.

In order to define the Gaussian quasi-log-likelihood contributions of the original data $Y_T = (y_1', \ldots, y_T')'$ and the corresponding transformed data $Y_T^*$ and $Y_T^+$, we use the density of the degenerate multivariate normal distribution (with respect to Lebesgue measure on an appropriate $r$-dimensional linear subspace of $\mathbb{R}^p$):

$$p(\varepsilon_t) = (2\pi)^{-r/2} |H_t|_-^{-1/2} \exp\left(-\tfrac{1}{2} \varepsilon_t' H_t^- \varepsilon_t\right),$$

and similarly the densities of $\varepsilon_t^*$ and $\varepsilon_t^+$. Defining $\vartheta := (\theta', \sigma_\varkappa^2)'$, this leads to the Gaussian quasi-log-likelihood function:

$$l(Y_T; \vartheta) = \sum_{t=1}^{T} \log p_\vartheta(y_t | Y_{t-1}),$$

where $p_\vartheta(y_t | Y_{t-1})$ is the (misspecified) Gaussian distribution of $y_t$ conditional on $Y_{t-1}$ (and $w_1, \ldots, w_{t-1}$), which can be evaluated via the prediction error decomposition based on the original state space representation (19)–(20). The collapsed transformation allows us to decompose

the log-likelihood function $l(Y_T; \vartheta)$ into the following three parts:

$$l(Y_T; \vartheta) = l(Y_T^*; \vartheta) + l(Y_T^+; \vartheta) + \sum_{t=1}^{T} \log |A_t|_-. \tag{25}$$

The first term in (25) is the quasi-log-likelihood evaluated by the Kalman filter applied to the collapsed state space system (23)–(24):

$$l(Y_T^*; \vartheta) = -\frac{dT}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{T} \log |G_t^*| - \frac{1}{2} \sum_{t=1}^{T} \omega_t^{*\prime} G_t^{*-1} \omega_t^*,$$

where $\omega_t^*$ are the prediction errors and $G_t^*$ are their mean square error matrices from the Kalman filter.

Since $y_t^+$ does not depend on the state vector $x_t$, the second term in (25) is given by

$$l(Y_T^+; \vartheta) = -\frac{(r-d)T}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{T} \log |H_t^+|_- - \frac{1}{2} \sum_{t=1}^{T} (y_t^+ - d_t^+)'(H_t^+)^-(y_t^+ - d_t^+).$$

The last term in the expression above can be calculated without an explicit evaluation of the matrix $A_t^+$:

$$\begin{aligned}
(y_t^+ - d_t^+)'(H_t^+)^-(y_t^+ - d_t^+) &= (y_t - d_t)' A_t^{+\prime}(A_t^+ H_t A_t^{+\prime})^- A_t^+(y_t - d_t) \\
&= (y_t - d_t)' J_t^+ H_t^-(y_t - d_t) \\
&= (y_t - d_t)' J_t^+ H_t^- J_t^{+\prime}(y_t - d_t) \\
&= (y_t - d_t)' M_{Z_t}' H_t^- M_{Z_t}(y_t - d_t) \\
&= e_t' H_t^- e_t,
\end{aligned}$$

where $M_{Z_t} = I - Z_t(Z_t' H_t^- Z_t)^{-1} Z_t' H_t^- = I - Z_t A_t^*$, $J_t^+ = A_t^{+\prime}(A_t^+ H_t A_t^{+\prime})^- A_t^+ H_t$ and $e_t = M_{Z_t}(y_t - d_t)$, that is, these are the generalized least squares (GLS) residuals from the observation vector $y_t$ with the covariate matrix $Z_t$ and variance matrix $H_t$. For derivation details,[11] see Jungbacker and Koopman (2015).

Finally, the third term in (25), $|A_t|_-$, can be found from the relation

$$|A_t|_-^2 \cdot |H_t|_- = |A_t H_t A_t'|_- = |H_t^*| \cdot |H_t^+|_-, \tag{26}$$

which follows from the fact that the covariance matrix $A_t H_t A_t'$ is block diagonal given the uncorrelated error terms $\varepsilon_t^*$ and $\varepsilon_t^+$.

Given the measurement error structure as implied by Proposition 1, the single scale parameter $\sigma_\varkappa^2$ of the covariance matrix can be factored out as $H_t = \sigma_\varkappa^2 \widetilde{H}_t$ (and hence $H_t^* = \sigma_\varkappa^2 \widetilde{H}_t^*$). The matrix $\widetilde{H}_t$ has a block-diagonal structure; although its blocks depend on the state vector and parameters via the theoretical BSIV $\kappa_t(\tau, m)$ and vega $\nu_t(\tau, m)$, we estimate these quantities directly from the data, hence they are not updated as we optimize over $\vartheta$. Thus, we have from

---

[11] The derivation in Jungbacker and Koopman (2015) is based on the invertible covariance matrix $H_t$, but the same result and the same derivation are valid when using the pseudo-inverse matrix $H_t^-$.

(26) that

$$
\begin{aligned}
\log |A_t|_- &= \frac{1}{2}\left(\log |H_t^*| + \log |H_t^+|_- - \log |H_t|_-\right) \\
&= \frac{1}{2}\left(\log |\widetilde{H}_t^*| + \log |H_t^+|_- - \log |\widetilde{H}_t|_- - \log \sigma_\varkappa^{2(r-d)}\right) \\
&= \frac{1}{2}\left(\log \frac{|\widetilde{H}_t^*|}{|\widetilde{H}_t|_-} + \log |H_t^+|_- - (r-d)\log \sigma_\varkappa^2\right).
\end{aligned}
$$

Therefore, the log-likelihood (25) is equal to

$$
l(Y_T; \vartheta) = c - \frac{1}{2}\sum_{t=1}^{T}\left(\log |G_t^*| + \omega_t^{*\prime}G_t^{*-1}\omega_t^* - \log \frac{|\widetilde{H}_t^*|}{|\widetilde{H}_t|_-} + \frac{e_t'\widetilde{H}_t^- e_t}{\sigma_\varkappa^2} + (r-d)\log \sigma_\varkappa^2\right), \quad (27)
$$

with $c = -\frac{1}{2}Tr\log 2\pi$. Note that the inverses and determinants of the matrices $G_t^*$ and $\widetilde{H}_t^*$ can be computed efficiently since they have small dimensions $d \times d$. This eases maximization of the log-likelihood function (27) substantially (the matrices $\widetilde{H}_t$ do not vary with the parameters, and therefore their inverses and determinants need to be computed only once).

The QML parameter estimator $\widehat{\vartheta} := (\widehat{\theta}', \widehat{\sigma}_\varkappa^2)'$ is obtained by maximizing (27) with respect to $\vartheta = (\theta', \sigma_\varkappa^2)'$ over the model parameter space $\Theta \times \Sigma_\varkappa$. For fixed $n$, its large-$T$ asymptotic properties are analogous to QML estimation based on the (modified) Kalman filter, as mentioned at the beginning of this sub-section. In cases in which the conditional covariance matrix $Q_t$ does not depend on the latent state vector $x_t$, and the state process is stationary, QML based on the Kalman filter will yield consistent and asymptotically normal estimators. When $Q_t$ is affine in $x_t$, then QML based on the modified Kalman filter appears to have comparable properties in Monte Carlo simulations, but no formal consistency proof is available.

On the other hand, by considering double asymptotics as $n, T \to \infty$ (maintaining a fixed cross-sectional dimension $p$), we are able to obtain an asymptotic inference result. Proposition 1 implies that as $n \to \infty$, the measurement error $\varepsilon_t$ (as well as the remainder term $r_{t,n}$) will tend to zero. This means that in the limit, we may recover the state vector $x_t$ from the observations $y_t$ combined with knowledge of the true parameter vector $\theta_0$ (needed to construct $(d_t, Z_t)$): defining $x_t(\theta) := A_t^*(\theta)(y_t - d_t(\theta)) = y_t^*(\theta) - d_t^*(\theta)$, we have $x_t(\theta_0) = x_t$.[12] That in turn implies that in the limit, the log-likelihood contributions $l_{t+1}^*(\vartheta)$ to $l(Y_T^*; \vartheta)$ can be approximated by

$$
\ell_{t+1}^*(\theta) = c^* - \frac{1}{2}\log |Q_t(x_t(\theta))| - \frac{1}{2}\eta_{t+1}(\theta)'Q_t(x_t(\theta))^{-1}\eta_{t+1}(\theta), \quad (28)
$$

with $c^* = -\frac{1}{2}d\log 2\pi$, $x_t(\theta) = y_t^*(\theta) - d_t^*(\theta)$ and

$$
\begin{aligned}
\eta_{t+1}(\theta) &= x_{t+1}(\theta) - c_t(\theta) - T_t(\theta)x_t(\theta) \\
&= y_{t+1}^*(\theta) - d_{t+1}^*(\theta) - c_t(\theta) - T_t(\theta)(y_t^*(\theta) - d_t^*(\theta)).
\end{aligned}
$$

Up to a Jacobian term (reflecting the transformation from $y_t$ to $y_t^*$), this is the log-likelihood contribution from the heteroskedastic first-order (vector) autoregressive model

$$
y_{t+1}^* = d_{t+1}^* + c_t + T_t(y_t^* - d_t^*) + \eta_{t+1},
$$

---

[12]This idea is related to the term structure model estimation approach proposed by Aït-Sahalia and Kimmel (2010), who assume that a number of yields have zero measurement errors, and hence can be used to back out the state vector without error.

with $\eta_{t+1}|y_t^* \sim N(0, Q_t(y_t^* - d_t^*))$. Observe that $\sigma_\varkappa^2$ does not enter $\ell_{t+1}^*(\theta)$, because $A_t^* = (Z_t'H_t^-Z_t)^{-1}Z_t'H_t^- = (Z_t'\widetilde{H}_t^-Z_t)^{-1}Z_t'\widetilde{H}_t^-$ does not vary with $\sigma_\varkappa^2$.

If $p = d$, then the Jacobian term, to be added to $\ell_{t+1}^*(\theta)$, will simply be $\log|A_{t+1}^*(\theta)|$, and there is no $l(Y_T^+;\vartheta)$ component of the log-likelihood. When $p > d$, the asymptotic analysis requires us to partition $\theta = (\theta_1', \theta_2')'$, where $\theta_1$ is the sub-vector of parameters determining $(d_t(\theta), Z_t(\theta))$, which hence are identified from the "$l_t^+$" contributions to the log-likelihood. The remaining parameters $\theta_2$ only appear in the transition parameter vectors and matrices $c_t(\theta)$, $T_t(\theta)$ and $Q_t(x_t;\theta)$ under the $\mathbb{P}$-dynamics (assuming that the model allows these parameters to be different under $\mathbb{P}$ and $\mathbb{Q}$). The full approximating average log-likelihood can be expressed as

$$\bar{\ell}_T(\vartheta) = \frac{1}{T}\sum_{t=1}^T \left(\ell_t^*(\theta) + l_t^+(\theta_1, \sigma_\varkappa^2) + J_t(\theta_1)\right) = \bar{\ell}_T^*(\theta) + \bar{l}_T^+(\theta_1, \sigma_\varkappa^2) + \bar{J}_T(\theta_1), \qquad (29)$$

where $\ell_t^*(\theta)$ is as defined in (28), $J_t(\theta_1) = \frac{1}{2}\log(|\widetilde{H}_t^*(\theta_1)|/|\widetilde{H}_t|_-)$ and

$$l_t^+(\theta_1, \sigma_\varkappa^2) = c^+ - \frac{1}{2}\frac{e_t(\theta_1)'\widetilde{H}_t^- e_t(\theta_1)}{\sigma_\varkappa^2} - \frac{(r-d)}{2}\log\sigma_\varkappa^2,$$

with $c^+ = -\frac{1}{2}(r-d)\log 2\pi$ and $e_t(\theta_1)$ the GLS residual as defined above. As shown below, under double asymptotics for $n, T \to \infty$, the convergence rate of $\widehat{\theta}_1$ is faster than the usual $\sqrt{T}$ rate that applies to $\widehat{\theta}_2$. This is caused by the fact that the error component in the GLS residuals, $M_{Z_t}(\theta)\varepsilon_t$, vanishes as $n \to \infty$, whereas the signal, which may be expressed as $M_{Z_t}(\theta)[d_t(\theta_0) - d_t(\theta) + Z_t(\theta_0)x_t]$, does not. The different convergence rate is a reflection of the fact that whereas $\theta_2$ is only identified from time-series information in the state updating equation (implying $\sqrt{T}$-consistency), the $\mathbb{Q}$-parameters contained in $\theta_1$ are identified from the cross-sectional information in the full vector $y_t$, with vanishing measurement errors.

The approximation of the modified Kalman-filter average log-likelihood $\bar{l}_T(\vartheta) = T^{-1}l(Y_T;\vartheta)$ $= T^{-1}\sum_{t=1}^T l_t(\vartheta)$ in (27) by the log-likelihood $\bar{\ell}_T(\vartheta)$ in (29), which does not involve filtering unobserved components, leads to consistency and asymptotic normality of $\widehat{\theta}$, under conditions stated next. We use the notation $\nabla$, $\nabla_1$ and $\nabla_2$ for the partial derivatives with respect to $\theta$, $\theta_1$ and $\theta_2$, respectively, and similarly $\nabla^2$, $\nabla_1^2$ and $\nabla_2^2$ for the second partial derivative matrices.

**Assumption 4** *The approximating average log-likelihood and its likelihood contributions satisfy, as $n, T \to \infty$,*

(i) *$\bar{l}_T(\vartheta) - \bar{\ell}_T(\vartheta) = \bar{l}_T^*(\vartheta) - \bar{\ell}_T^*(\vartheta)$ is stochastically equicontinuous;*

(ii) *The true value $\vartheta_0$ lies in the interior of $\Theta \times \Sigma_\varkappa$ and, with $B(\vartheta_0;\delta) = \{\vartheta : |\vartheta - \vartheta_0| \leq \delta\}$,*

$$\lim_{n,T\to\infty} \mathbb{P}\left\{\bar{\ell}_T(\vartheta_0) > \sup_{\vartheta \in (\Theta\times\Sigma_\varkappa)\backslash B(\vartheta_0,\delta)} \bar{\ell}_T(\vartheta)\right\} = 1, \quad \forall\delta > 0;$$

(iii) *The (scaled) score contributions $\sqrt{\Delta m}\nabla_1 l_t^+(\theta_{1,0}, \sigma_{\varkappa,0}^2)$ and $\nabla_2 \ell_t^*(\theta_0)$ satisfy a joint central limit theorem, with variance matrix $\mathcal{I} = \text{blkdiag}(\mathcal{I}_1, \mathcal{I}_2)$;*

*(iv) For any sequence $\vartheta_T \to \vartheta_0$,*

$$\plim_{n,T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \left( \begin{array}{c} \sqrt{\Delta m}\nabla_1 l_t^+(\theta_{1,T},\sigma_{\varkappa,T}^2) \\ \nabla_2 \ell_t^*(\theta_T) \end{array} \right) \left( \begin{array}{c} \sqrt{\Delta m}\nabla_1 l_t^+(\theta_{1,T},\sigma_{\varkappa,T}^2) \\ \nabla_2 \ell_t^*(\theta_T) \end{array} \right)' = \mathcal{I},$$

$$\plim_{n,T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \left( \begin{array}{cc} \Delta m\nabla_1^2 l_t^+(\theta_{1,T},\sigma_{\varkappa,T}^2) & 0 \\ 0 & \nabla_2^2 \ell_t^*(\theta_T) \end{array} \right) = \left( \begin{array}{cc} \mathcal{S}_1 & 0 \\ 0 & \mathcal{S}_1 \end{array} \right) = \mathcal{S}.$$

Assumption 4 $(i)$ guarantees that pointwise convergence to the approximating log-likelihood can be strengthened to uniform convergence; Assumption 4 $(ii)$ is a consistency condition as in Wu (1981), see also Assumption A2 of Andersen et al. (2015a); and Assumptions 4 $(iii)$–$(iv)$ are sufficient for asymptotic normality and consistency of the asymptotic variance matrix of the estimators.

**Proposition 2** *Suppose that the assumptions of Proposition 1, with $(\underline{\alpha}\wedge\overline{\alpha}) > 1/(2(\underline{\delta}\wedge(1+\overline{\delta})))$, and Assumption 4 hold. Also assume that the first- and second-order derivatives, with respect to $\theta$, of the system elements $\widetilde{H}_t^*(\theta)$, $T_t(\theta)$, $c_t(\theta)$, $Q_t(x_t,\theta)$ and $Z_t(\theta)$ are bounded. In addition, suppose that $n,T \to \infty$ with $\sqrt{T}n^{-1}\log n \to 0$, and assume that $r := \min_{1\le t\le T} \text{rank} H_t > d$. Then, the QML estimator $\widehat{\theta}$ is consistent and asymptotically normal, with:*

$$D_{nT}(\widehat{\theta} - \theta_0) = \left( \begin{array}{c} \sqrt{T/(\Delta m)}(\widehat{\theta}_1 - \theta_{1,0}) \\ \sqrt{T}(\widehat{\theta}_2 - \theta_{2,0}) \end{array} \right) \xrightarrow{d} N\left( \left( \begin{array}{c} 0 \\ 0 \end{array} \right), \left( \begin{array}{cc} \mathcal{S}_1^{-1}\mathcal{I}_1\mathcal{S}_1^{-1} & 0 \\ 0 & \mathcal{S}_2^{-1}\mathcal{I}_2\mathcal{S}_2^{-1} \end{array} \right) \right),$$

*where $D_{nT} = \text{blkdiag}\{\sqrt{T/(\Delta m)}I_{\dim\theta_1}, \sqrt{T}I_{\dim\theta_2}\}$, and where*

$$D_{nT}^{-1} \sum_{t=1}^{T} \nabla l_t(\widehat{\vartheta})\nabla l_t(\widehat{\vartheta})' D_{nT}^{-1} \xrightarrow{p} \mathcal{I}, \qquad D_{nT}^{-1} \sum_{t=1}^{T} \nabla^2 l_t(\widehat{\vartheta}) D_{nT}^{-1} \xrightarrow{p} \mathcal{S}.$$

*In addition, $\widehat{\sigma}_\varkappa^2 \xrightarrow{p} \sigma_{\varkappa,0}^2$.*

The proof of Proposition 2 is given in Appendix A. Despite different convergence rates, the proposition implies that standard errors of individual components of $\widehat{\theta}_1$ and $\widehat{\theta}_2$ can be obtained consistently from the conventional QML sandwich formula; the $\sqrt{\Delta m}$ factors, required for convergence, will automatically cancel out from $t$- or Wald test statistics involving $\widehat{\theta}_1$.

# 4 Monte Carlo Study

In this section, we study the finite-sample performance of our estimation procedure. In particular, we consider two AJD specifications: a one-factor model and two versions of an option pricing model with two factors.

## 4.1 SVCDEJ

As a starting point, we illustrate the developed estimation procedure based on a modification of the widely used 'double-jump' stochastic volatility model of Duffie et al. (2000). The modification is due to using double-exponential (rather than Gaussian) jump sizes in returns as in Kou

(2002) and Andersen et al. (2015a), and a stochastic (rather than constant) jump intensity that is a multiple of the stochastic variance as in Pan (2002). We label this specification as 'SVCDEJ' for stochastic volatility model with co-jumps in volatility and double-exponential jumps in returns.

In particular, we assume the following data-generating process for the log forward price under both the $\mathbb{P}$ and $\mathbb{Q}$ probability measures:

$$\mathrm{d}\log F_t = (-\tfrac{1}{2}v_t - \mu\lambda_t)\mathrm{d}t + \sqrt{v_t}\mathrm{d}W_{1,t} + J_t\mathrm{d}N_t, \tag{30}$$

$$\mathrm{d}v_t = \kappa(\bar{v} - v_t)\mathrm{d}t + \sigma\sqrt{v_t}\mathrm{d}W_{2,t} + J_t^v\mathbf{1}_{\{J_t<0\}}\mathrm{d}N_t, \tag{31}$$

where the two standard Brownian motions $W_1$ and $W_2$ are assumed to be correlated with coefficient $\rho \in [-1, 1]$, and $N_t$ is a Poisson jump process with jump intensity proportional to the stochastic variance, $\lambda_t = \delta v_t$, $\delta > 0$. We further assume that $J_t$ is a double-exponentially distributed jump size with generic probability density function

$$f_J(x) = p^+\frac{1}{\eta^+}e^{-x/\eta^+}\mathbf{1}_{\{x\geq 0\}} + p^-\frac{1}{\eta^-}e^{x/\eta^-}\mathbf{1}_{\{x<0\}},$$

where $p^+$ and $p^-$ are probabilities of positive and negative jumps, respectively, and $\eta^+$ and $\eta^-$ are the corresponding conditional means of the jump sizes. We assume that all of these parameters are positive, $p^+ + p^- = 1$ and $\eta^+ < 1$. Given the jump size distribution, the expected relative jump size in returns is

$$\mu := \mathbb{E}[e^J - 1] = \frac{p^+}{1-\eta^+} + \frac{p^-}{1+\eta^-} - 1.$$

We allow the volatility to co-jump only with negative jumps in returns, with exponentially distributed jump sizes $J_t^v$ with mean $\mu_v > 0$. Finally, we assume $\kappa$, $\bar{v}$ and $\sigma$ to be positive and impose Feller's condition $2\kappa\bar{v} > \sigma^2$ and the covariance stationarity condition $\kappa > p^-\delta\mu_v$.

The model in (30)–(31) belongs to the AJD class and exhibits all important ingredients of option pricing models: stochastic volatility, jump components in returns and volatility, time-varying jump intensity and a self-excitation feature (because a negative jump in returns is associated with a positive jump in volatility, which increases the volatility and hence the jump intensity). Furthermore, this specification assumes a double-exponential jump size distribution in returns, which has recently been advocated in the literature (see, e.g., Kou, 2002, Aït-Sahalia et al., 2015, Andersen et al., 2015a and Bardgett et al., 2019).

Our developed estimation and filtering approach uses information from option prices, and is agnostic about equity risk premia. Indeed, the measurements are constructed as portfolios of options rather than the underlying asset. On the other hand, since the transition equation in the state space representation reflects the dynamics of the latent components (under $\mathbb{P}$), it is, in principle, possible to learn about the risk premia associated with the latent processes (for instance, the variance risk premium). However, additional simulation results, reported in Appendix D, suggest that the $\mathbb{Q}$-information in the option prices largely dominates the $\mathbb{P}$-information, making the identification of risk premium parameters weak. Specifically, the simulation results in Table D.2 show that: (i) it is difficult to reliably identify the $\mathbb{P}$-parameters; (ii) notably,

misspecification of the $\mathbb{P}$-dynamics barely changes the identification of the $\mathbb{Q}$-parameters. The parameters characterizing the $\mathbb{P}$-dynamics are identified from the time-series information in the low-dimensional vector $y_t^*$ only — the cross-sectional information in $y_t^+$ identifies the $\mathbb{Q}$-parameters only. Accurate estimation of mean and mean-reversion parameters from time series requires a much longer time span than we consider in our Monte Carlo simulations and empirical illustrations. A similar difficulty of identifying the physical dynamics arises in the term structure literature (see, e.g., Kim & Orphanides, 2012). Therefore, we assume no variance (or state related) risk premia, that is, the latent components have the same dynamics under both probability measures. Importantly, the results in Appendix D suggest that estimation of the $\mathbb{Q}$-parameters is hardly affected by imposing this (possibly invalid) restriction.

The discounted marginal CCF of the log forward prices in the SVCDEJ model can be derived using the results in Duffie et al. (2000) and is given by

$$\psi_X(\mathbf{u}_1, \tau) = e^{-r\tau} \mathbb{E}^{\mathbb{Q}}[e^{iu \log F_{t+\tau}} | \mathcal{F}_t] = e^{\alpha(u,\tau) + \beta_1(u,\tau) \log F_t + \beta_2(u,\tau) v_t}, \tag{32}$$

where $\alpha(u,\tau)$ and $\beta(u,\tau)$ are solutions to the complex-valued ODE system in time:

$$\begin{cases} \dot{\beta}_1(u,s) &= 0, \\ \dot{\beta}_2(u,s) &= -\left(\frac{1}{2} + \mu\delta\right)\beta_1(u,s) - \kappa\beta_2(u,s) + \frac{1}{2}\beta_1^2(u,s) + \rho\sigma\beta_1(u,s)\beta_2(u,s) \\ &\quad + \frac{1}{2}\sigma^2\beta_2^2(u,s) + \delta(\chi(\beta_1(u,s), \beta_2(u,s)) - 1), \\ \dot{\alpha}(u,s) &= \kappa\bar{v}\beta_2(u,s) - r, \end{cases}$$

with initial conditions $\beta_1(u,0) = iu$, $\beta_2(u,0) = 0$ and $\alpha(u,0) = 0$. Here the 'jump transform' takes the form

$$\chi(\beta_1, \beta_2) = \frac{p^+}{1 - \beta_1\eta^+} + \frac{p^-}{(1 + \beta_1\eta^-)(1 - \beta_2\mu_v)}.$$

The CCF of the log price in (32) is used to price options. For the state space representation, we turn it into the CCF of log returns as described in Section 2.3. Using the fact that the solution to the ODE system satisfies $\beta_1(u,\tau) = iu$, the linear relation between the log of the option-implied CCF and the state vector is given by

$$\log \widehat{\phi}_t(u, \tau) = \alpha(u, \tau) + \beta_2(u, \tau)v_t + \xi_t(u, \tau), \qquad u \in \mathbb{R},$$

where $\widehat{\phi}_t(u, \tau)$ is the option-implied CCF, $\tau > 0$ is the time-to-maturity of available options and $\xi_t(u, \tau)$ is the measurement error term due to observation and approximation errors in the option-implied CCF. We use this linear relation to construct the measurement equation as discussed in Section 3.1.

Following Appendix B, the conditional mean and variance of the latent stochastic volatility process are given by

$$\mathbb{E}[v_{t+1}|\mathcal{F}_t] = e^{g_1 \Delta t} v_t + \frac{g_0}{g_1}\left(e^{g_1 \Delta t} - 1\right), \tag{33}$$

$$\text{Var}(v_{t+1}|\mathcal{F}_t) = -\frac{\sigma^2 + 2p^- \delta\mu_v^2}{2g_1^2}\left[2g_1\left(e^{g_1 \Delta t} - e^{2g_1 \Delta t}\right)v_t - g_0\left(1 - e^{g_1 \Delta t}\right)^2\right], \tag{34}$$

with $g_0 = \kappa \bar{v}$ and $g_1 = -\kappa + p^- \delta \mu_v$. Equations (33)–(34) are then used to define the state updating equation:

$$v_{t+1} = c_t + T_t v_t + \eta_{t+1}, \tag{35}$$

where $c_t = \frac{g_0}{g_1} \left( e^{g_1 \Delta t} - 1 \right)$, $T_t = e^{g_1 \Delta t}$ and $\text{Var}(\eta_{t+1}|\mathcal{F}_t) = \text{Var}(v_{t+1}|\mathcal{F}_t) =: Q_t(v_t)$.

The model specification has nine parameters of interest and one additional parameter that characterizes the observation errors. We note that the parameter $\delta$ often enters as a multiple of $p^-$, which can possibly cause identification issues in the estimation procedure. Therefore, to avoid these identification issues, we fix the probability of negative jumps to be $p^- = 0.7$. This is consistent with findings in Aït-Sahalia et al. (2015) and our empirical results for the unrestricted model provided in Appendix D.2, where we also assess the robustness of our empirical results to fixing $p^- = 0.7$.

In the simulation study, we use $T = 500$ time points with $\Delta t = 1/250$, representing two years of data. The time-series of the log prices and true spot volatilities are simulated using an Euler scheme applied to the specification (30)–(31). The initial values are set to $F_0 = 100$ and $v_0 = 0.015$. The options data are generated using the COS method of Fang and Oosterlee (2008) based on the CCF, specified in (32). The true model parameters are displayed in Table 1.

In the simulations, we consider three tenors for options equal to 10, 30 and 60 days. For each tenor, we simulate in an adaptive way a finite number of options with log-moneyness between $\underline{m} = -10 \cdot \sigma_{ATM,\tau} \sqrt{\tau}$ and $\overline{m} = 4 \cdot \sigma_{ATM,\tau} \sqrt{\tau}$, where $\sigma_{ATM,\tau}$ is the BSIV of the ATM option with time-to-maturity $\tau$. Furthermore, the strikes are generated equidistantly with $\Delta K = 0.01 \cdot F_t$. This results in 75 contracts per tenor, on average. For ease of simulation, the tenors are the same for each simulated time point. Finally, we distort the options data by adding the observation errors to the option prices for each tenor $\tau$ and each log-moneyness level $m$ as specified in Assumption 2, i.e.,

$$\widehat{O}_t(\tau, m) = O_t(\tau, m) + \sigma_\varkappa \cdot \kappa_t(\tau, m) \nu_t(\tau, m) \cdot \epsilon,$$

where $\epsilon$ is an i.i.d. standard normal random variable and $\sigma_\varkappa = 0.02$. The distorted option prices, in terms of total implied variance, are then interpolated using a cubic spline and extrapolated linearly in log-moneyness, as described in Appendix C.1.

The covariance matrix of the errors in the measurement equation is calculated according to equation (21). To calculate the pseudo-inverse of the $2q \times 2q$ covariance matrix $\widetilde{H}_{t,i}$ for each of the maturities $i = 1, \ldots, k$, we set the following level of the threshold for the singular values:

$$\text{tol} := \bar{s} \cdot 2q \cdot \max_j s_j,$$

with $\bar{s} = 10^{-5}$ and where the maximum is taken over all singular values $s_j$ of $\widetilde{H}_{t,i}$. We also analyze the robustness of our results to the choice of $\bar{s}$.

With these specifications, we take the number of replications $N$ to be $N = 300$, thus running the estimation procedure of Section 3 300 times. Table 1 provides the Monte Carlo results for the SVCDEJ model, for six different ranges of the argument set $\mathcal{U}$. Some general considerations are: First, negative values of $u$ are not considered because the characteristic function is Hermitian,

## Table 1: Monte Carlo results for the SVCDEJ model

| parameter | $\sigma$ | $\kappa$ | $\bar{v}$ | $\rho$ | $\delta$ | $\eta^+$ | $\eta^-$ | $\mu_v$ | $\sigma_\varkappa$ |
|---|---|---|---|---|---|---|---|---|---|
| | | | | $u = 1,\ldots,5$ | | | | | |
| true value | 0,450 | 8,000 | 0,0150 | -0,950 | 100,000 | 0,0200 | 0,0500 | 0,0500 | 0,020 |
| mean | 0,495 | 7,658 | 0,0162 | -0,900 | 86,900 | 0,0170 | 0,0515 | 0,0523 | 0,023 |
| std | 0,012 | 0,347 | 0,0013 | 0,015 | 10,790 | 0,0066 | 0,0025 | 0,0033 | 0,003 |
| q10 | 0,490 | 7,484 | 0,0155 | -0,908 | 76,066 | 0,0000 | 0,0499 | 0,0499 | 0,021 |
| q50 | 0,496 | 7,705 | 0,0159 | -0,898 | 88,905 | 0,0194 | 0,0510 | 0,0513 | 0,023 |
| q90 | 0,504 | 7,865 | 0,0170 | -0,889 | 94,972 | 0,0214 | 0,0528 | 0,0561 | 0,025 |
| | | | | $u = 1,\ldots,10$ | | | | | |
| true value | 0,450 | 8,000 | 0,0150 | -0,950 | 100,000 | 0,0200 | 0,0500 | 0,0500 | 0,020 |
| mean | 0,454 | 8,122 | 0,0150 | -0,954 | 101,092 | 0,0193 | 0,0493 | 0,0506 | 0,024 |
| std | 0,025 | 0,312 | 0,0008 | 0,037 | 9,089 | 0,0034 | 0,0014 | 0,0011 | 0,002 |
| q10 | 0,429 | 7,656 | 0,0143 | -1,000 | 87,158 | 0,0185 | 0,0479 | 0,0497 | 0,022 |
| q50 | 0,448 | 8,173 | 0,0148 | -0,960 | 103,608 | 0,0200 | 0,0490 | 0,0503 | 0,024 |
| q90 | 0,498 | 8,455 | 0,0162 | -0,893 | 109,667 | 0,0210 | 0,0515 | 0,0516 | 0,027 |
| | | | | $u = 1,\ldots,15$ | | | | | |
| true value | 0,450 | 8,000 | 0,0150 | -0,950 | 100,000 | 0,0200 | 0,0500 | 0,0500 | 0,020 |
| mean | 0,452 | 8,021 | 0,0151 | -0,953 | 99,497 | 0,0201 | 0,0499 | 0,0501 | 0,023 |
| std | 0,015 | 0,176 | 0,0004 | 0,018 | 4,611 | 0,0014 | 0,0008 | 0,0007 | 0,002 |
| q10 | 0,439 | 7,916 | 0,0148 | -0,971 | 97,538 | 0,0193 | 0,0493 | 0,0494 | 0,021 |
| q50 | 0,449 | 8,056 | 0,0150 | -0,956 | 100,593 | 0,0201 | 0,0497 | 0,0501 | 0,023 |
| q90 | 0,463 | 8,177 | 0,0152 | -0,936 | 102,721 | 0,0209 | 0,0502 | 0,0509 | 0,026 |
| | | | | $u = 1,\ldots,20$ | | | | | |
| true value | 0,450 | 8,000 | 0,0150 | -0,950 | 100,000 | 0,0200 | 0,0500 | 0,0500 | 0,020 |
| mean | 0,447 | 7,946 | 0,0150 | -0,962 | 101,097 | 0,0219 | 0,0500 | 0,0487 | 0,024 |
| std | 0,022 | 0,254 | 0,0006 | 0,027 | 6,528 | 0,0013 | 0,0013 | 0,0013 | 0,004 |
| q10 | 0,424 | 7,493 | 0,0145 | -0,997 | 89,601 | 0,0206 | 0,0491 | 0,0469 | 0,022 |
| q50 | 0,440 | 8,005 | 0,0148 | -0,968 | 103,013 | 0,0217 | 0,0497 | 0,0487 | 0,024 |
| q90 | 0,491 | 8,153 | 0,0161 | -0,916 | 106,663 | 0,0235 | 0,0522 | 0,0500 | 0,027 |
| | | | | $u = 1,\ldots,25$ | | | | | |
| true value | 0,450 | 8,000 | 0,0150 | -0,950 | 100,000 | 0,0200 | 0,0500 | 0,0500 | 0,020 |
| mean | 0,453 | 7,889 | 0,0152 | -0,954 | 100,328 | 0,0224 | 0,0504 | 0,0474 | 0,025 |
| std | 0,023 | 0,267 | 0,0007 | 0,027 | 7,204 | 0,0015 | 0,0017 | 0,0027 | 0,006 |
| q10 | 0,430 | 7,504 | 0,0146 | -0,984 | 89,329 | 0,0210 | 0,0492 | 0,0451 | 0,022 |
| q50 | 0,446 | 7,924 | 0,0150 | -0,959 | 102,200 | 0,0224 | 0,0500 | 0,0474 | 0,024 |
| q90 | 0,493 | 8,121 | 0,0162 | -0,909 | 107,179 | 0,0238 | 0,0524 | 0,0495 | 0,033 |
| | | | | $u = 1,\ldots,30$ | | | | | |
| true value | 0,450 | 8,000 | 0,0150 | -0,950 | 100,000 | 0,0200 | 0,0500 | 0,0500 | 0,020 |
| mean | 0,464 | 7,866 | 0,0154 | -0,939 | 97,538 | 0,0217 | 0,0507 | 0,0474 | 0,026 |
| std | 0,021 | 0,239 | 0,0007 | 0,019 | 6,487 | 0,0016 | 0,0019 | 0,0035 | 0,009 |
| q10 | 0,445 | 7,683 | 0,0149 | -0,960 | 91,742 | 0,0207 | 0,0496 | 0,0450 | 0,021 |
| q50 | 0,458 | 7,855 | 0,0152 | -0,941 | 98,945 | 0,0215 | 0,0504 | 0,0473 | 0,022 |
| q90 | 0,497 | 8,079 | 0,0159 | -0,908 | 102,387 | 0,0227 | 0,0518 | 0,0501 | 0,038 |

Note: This table provides Monte Carlo simulation results for the SVCDEJ model, based on 300 replications. Six settings with different ranges of the argument $u$ are considered. Each panel lists, for each parameter, the true value, the Monte Carlo mean and standard deviation, and the 10th, 50th and 90th Monte Carlo percentiles, respectively. We use $T = 500$ time points with $\Delta t = 1/250$. The initial values are set to $F_0 = 100$ and $v_0 = 0.015$. The threshold for singular values is set to $\bar{s} = 10^{-5}$. The probability of negative jumps is fixed to $p^- = 0.7$.

i.e., negative values provide the same information as positive values. Second, the finer the set $\mathcal{U}$, the more the information from the characteristic function is alike. This results in a lower rank of the covariance matrix, i.e., a different required threshold for the singular values controlled by the parameter $\bar{s}$. Therefore, one can choose a different set of arguments $\mathcal{U}$ and adapt the threshold value for the pseudo-inversion to this set. Furthermore, the distance between adjacent arguments in the set $\mathcal{U}$ also affects the choice of its cardinality, i.e., the number of arguments $q$. For that reason, after fixing the distance between the arguments and choosing the threshold parameter $\bar{s}$, we analyze different numbers of arguments $q$. The simulation results in general show a good finite-sample performance. We notice that for smaller ranges of the CCF argument, the estimates exhibit biases for some model parameters. This is expected since the smaller ranges provide coarser information on which we build the filtering and parameter estimation procedures. On the other hand, we also notice that the variance of some parameter estimates increases when using a very large range of arguments (in particular, $u = 1, \ldots, 30$). This is likely due to an increased variance in the CCF approximation for large arguments $u$.

Figure 1: RMSPE and RSSTE for different levels of $\bar{s}$



(a) RMSPE                    (b) RSSTE
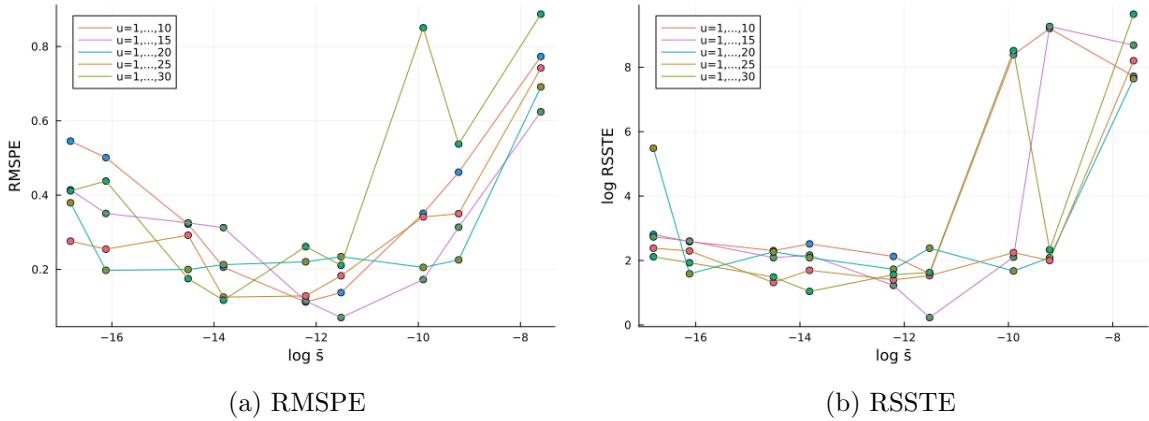
Note: This figure plots the RMSPE and RSSTE for different levels of $\bar{s}$, considering five different ranges of the argument $u$. In particular, we estimate $N = 300$ replications for each $\bar{s} = \{5 \cdot 10^{-8}, 10^{-7}, 5 \cdot 10^{-7}, 10^{-6}, 5 \cdot 10^{-6}, 10^{-5}, 5 \cdot 10^{-5}, 10^{-4}, 5 \cdot 10^{-4}\}$ and plot the resulting RMSPE (Panel (a)) and log RSSTE (Panel (b)) against $\bar{s}$ in log-scale.

To explore the robustness to the choice of the truncation level in the pseudo-inversion of the covariance matrix, we also consider other values of $\bar{s}$. In particular, we run $N = 300$ simulations for each level of $\bar{s}$ using the same parameter values as in Table 1, and construct the root mean square percentage error (RMSPE) metrics, defined as the square root of $N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{d_\theta} \left( (\widehat{\theta}_{i,j} - \theta_{0,j})/\theta_{0,j} \right)^2$, with $d_\theta$ the dimension of $\theta$. Furthermore, we construct the root sum of squared t-statistics error (RSSTE) defined as $\sqrt{\sum_{j=1}^{d_\theta} \left( \text{MCse}(t_{\theta_j}) - 1 \right)^2}$, where $\text{MCse}(t_{\theta_j}) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\widehat{t}_{\theta_{i,j}} - \bar{t}_{\theta_j})^2}$ with $\widehat{t}_{\theta_{i,j}} := (\widehat{\theta}_{i,j} - \theta_{0,j})/\text{se}(\widehat{\theta}_{i,j})$, $\bar{t}_{\theta_j} := \frac{1}{N} \sum_{i=1}^{N} \widehat{t}_{\theta_{i,j}}$, and with $\text{se}(\widehat{\theta}_{i,j})$ being the standard error of the estimated parameter. Figure 1 plots the resulting RMSPEs and RSSTEs for different levels of $\bar{s}$ and five different ranges of the argument $u$. As we can see, the levels of $\bar{s}$ in between $10^{-6}$ and $10^{-5}$ yield the smallest RMSPE and RSSTE. In the following simulations and empirical applications, we therefore set $\bar{s} = 10^{-5}$.

27

We end this subsection by noting that we have also conducted simulation studies for some related alternative one-factor specifications. In particular, in Appendix D.1, we provide additional simulation results for the 'SVCJ' model with a Gaussian jump size distribution, and the 'SVCEJ' model with two separate counting processes for positive and negative jumps. The former shows a very good finite-sample performance, while the latter, a richer model specification, shows reasonable results, gradually reaching the limits of what can be identified using the present input data and design. In Appendix D.1, we also provide additional robustness checks for the SVCDEJ model, including AR(1) and spatially dependent error structures, and compare its filtered and simulated volatility, which appear to be very close. Finally, additional simulation results (available upon request) indicate improvements with increasing $T$ and $n$, consistent with the asymptotic results established in Proposition 2.

## 4.2 SVCDEJ with external factors

Now we extend the one-factor specification by adding an external factor. This modification can be seen as a two-factor specification, but we will assume that the second factor is observable. The motivation comes from the fact that in some situations we might have an understanding of possible drivers of the risks in the market. Therefore, we would like to embed exogenous variables into the model's risk factors and quantify their impact.

In particular, next to the stochastic volatility component we introduce the exogenous factor $h_t$, which affects the intensity of jumps and the diffusive component. The model reads as follows:

$$\mathrm{d}\log F_t = (-\tfrac{1}{2}V_t - \mu\lambda_t)\mathrm{d}t + \sqrt{v_t}\mathrm{d}W_{1,t} + q\sqrt{h_t}\mathrm{d}W_{3,t} + J_t\mathrm{d}N_t, \tag{36}$$

$$\mathrm{d}v_t = \kappa(\bar{v} - v_t)\mathrm{d}t + \sigma\sqrt{v_t}\mathrm{d}W_{2,t} + J_t^v\mathbf{1}_{\{J_t<0\}}\mathrm{d}N_t, \tag{37}$$

$$\mathrm{d}h_t = \kappa_h(\bar{h} - h_t)\mathrm{d}t + \sigma_h\sqrt{h_t}\mathrm{d}W_{4,t}, \tag{38}$$

where $V_t = v_t + q^2 h_t$ is the total diffusive variance of the process and the jump intensity process $\lambda_t$ is also affected by $h_t$ with $\lambda_t = \delta v_t + \gamma h_t$, $q, \delta, \gamma > 0$. We assume that $W_{3,t}$ and $W_{4,t}$ are independent standard Brownian motions, jointly independent of $(W_{1,t}, W_{2,t})$. The process $h_t$ is exogenous to the SVCDEJ dynamics, meaning that the dynamics of $\log F_t$ and $v_t$ do not affect the dynamics of $h_t$. In turn, the exogenous factor $h_t$ affects the intensity of jumps and the diffusive component of the log return dynamics. This specification is similar to the two-factor model in Andersen et al. (2015a), which includes short- and long-term stochastic volatility components. The difference is that here the exogenous process $h_t$ is observable, although its parameters are unknown.

In the Monte Carlo simulations, we consider two possible estimation approaches. In the first approach, we assume a correct specification of the dynamics of $h_t$ with known true parameters $\kappa_h$, $\bar{h}$ and $\sigma_h$. In practice, these parameters can be pre-estimated given the observed path of the exogenous process. In the second approach, we estimate the misspecified model in which the contribution of $h_t$ is constant throughout the maturity of an option. In other words, under this approach we ignore the dynamics of $h_t$ when pricing options, but let $h_t$ still affect the level of the jump intensity and of the total variance. The motivation is that when the exogenous process

28

Table 2: Monte Carlo results for the SVCDEJ model with external factor

| parameter | $\sigma$ | $\kappa$ | $\bar{v}$ | $\rho$ | $\delta$ | $\eta^+$ | $\eta^-$ | $\mu_v$ | $\gamma$ | $q$ | $\sigma_\varkappa$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| (a) Estimation with known true parameters for $h_t$ | | | | | | | | | | | |
| true value | 0.450 | 8.00 | 0.0150 | -0.950 | 100.00 | 0.0200 | 0.050 | 0.050 | 1.500 | 0.050 | 0.020 |
| mean | 0.459 | 7.973 | 0.016 | -0.944 | 102.74 | 0.022 | 0.050 | 0.048 | 1.449 | 0.048 | 0.022 |
| std | 0.035 | 0.238 | 0.001 | 0.046 | 7.24 | 0.001 | 0.001 | 0.001 | 0.106 | 0.004 | 0.002 |
| q10 | 0.425 | 7.630 | 0.015 | -0.993 | 92.43 | 0.021 | 0.049 | 0.047 | 1.326 | 0.045 | 0.021 |
| q50 | 0.448 | 8.039 | 0.015 | -0.957 | 104.42 | 0.022 | 0.050 | 0.048 | 1.485 | 0.048 | 0.022 |
| q90 | 0.515 | 8.193 | 0.017 | -0.870 | 110.11 | 0.023 | 0.051 | 0.049 | 1.515 | 0.049 | 0.024 |
| (b) Estimation of misspecified model | | | | | | | | | | | |
| true value | 0.450 | 8.00 | 0.0150 | -0.950 | 100.00 | 0.0200 | 0.050 | 0.050 | 1.500 | 0.050 | 0.020 |
| mean | 0.457 | 7.967 | 0.015 | -0.949 | 103.03 | 0.022 | 0.050 | 0.048 | 1.474 | 0.048 | 0.022 |
| std | 0.034 | 0.251 | 0.001 | 0.047 | 7.69 | 0.001 | 0.001 | 0.001 | 0.104 | 0.004 | 0.002 |
| q10 | 0.425 | 7.649 | 0.014 | -1.000 | 92.56 | 0.021 | 0.049 | 0.047 | 1.343 | 0.046 | 0.021 |
| q50 | 0.445 | 8.030 | 0.015 | -0.962 | 104.72 | 0.022 | 0.050 | 0.048 | 1.510 | 0.049 | 0.022 |
| q90 | 0.513 | 8.177 | 0.017 | -0.874 | 110.38 | 0.023 | 0.051 | 0.049 | 1.540 | 0.050 | 0.024 |

Note: This table provides Monte Carlo simulation results for the SVCDEJ model with an exogenous factor, based on 300 replications. Each panel lists, for each parameter, the true value, the Monte Carlo mean and standard deviation, and the 10th, 50th and 90th Monte Carlo percentiles, respectively. We use $T = 500$ time points with $\Delta t = 1/250$. The range for the arguments is set to $u = 1, \ldots, 15$ and the threshold to $\bar{s} = 10^{-5}$. The initial values are set to $F_0 = 100$ and $v_0 = 0.015$. The probability of negative jumps is fixed to $p^- = 0.7$. The parameters of the external factor are set to $\kappa_h = 1$, $\bar{h} = 1$ and $\sigma_h = 0.1$.

is persistent and smooth relative to $v_t$, its dynamics, and correlation, can be neglected when pricing options with short expiration periods. In a similar way, interest rates are often assumed to enter option prices in a deterministic way. Moreover, the true parametric specification for an exogenous variable is likely unknown in practice, but if its dynamics are persistent and smooth, we can find its effect on option prices via this approach. Therefore, in the Monte Carlo experiment, we simulate $h_t$ with a mean-reversion rate that is smaller than that in $v_t$, mimicking the specification we will explore in the empirical application.

For the rest, the Monte Carlo setting for the SVCDEJ model with an external factor is the same as for the SVCDEJ specification in the previous subsection. The parameters of the external factor are set to $\kappa_h = 1$, $\bar{h} = 1$ and $\sigma_h = 0.1$. The simulation results are provided in Table 2. The parameters of the SVCDEJ model exhibit similar good performance under both estimation approaches. Importantly, the parameters related to the external factors, $\gamma$ and $q$, also show similar good performance in the correctly specified model as in the misspecified setting. We emphasize that this is achieved due to simulating a relatively smooth and persistent exogenous process $h_t$ and using short-dated options in the estimation procedure.

# 5 Data

This section describes the data and the data selection process, which we use in our empirical application. Since our estimation procedure utilizes option-implied CCFs, we also pay attention to the construction of these objects in this section. Further details are in Appendix C.

## 5.1 Data description

In this paper, we use options data on the S&P 500 stock market index obtained from the Chicago Board Options Exchange (CBOE). We focus on the period from May 1, 2017, to April 1, 2021, covering in particular the turbulent period in the stock market due to the outbreak of the Covid-19 pandemic. The CBOE provides end-of-day option quotes and a snapshot at 3:45 pm ET, 15 minutes prior to the market closure. We use the latter to calculate mid-quotes since it is considered to be a more accurate representation than the former in view of market liquidity. The data contain both the 'standard' AM-settled SPX options and Weeklys and End-of-Months PM-settled SPXW products. The settlement value for the SPX options is based on the opening level of the S&P 500 index on the settlement day, whereas for the SPXW options it is based on the closing prices of the index.

Table 3: Descriptive statistics for S&P 500 index options

| | $k \leq 0.8$ | $0.8 < k \leq 0.95$ | $0.95 < k \leq 1.03$ | $1.03 < k \leq 1.1$ | $1.1 < k$ | Total |
|---|---|---|---|---|---|---|
| Panel A: Total volume of option contracts (in millions) | | | | | | |
| $\tau \in (2, 9]$ | 4.52 | 52.61 | 167.09 | 13.27 | 1.25 | 238.73 |
| $\tau \in (9, 30]$ | 25.47 | 74.58 | 168.06 | 29.93 | 3.96 | 302.00 |
| $\tau \in (30, 60]$ | 25.85 | 61.97 | 109.53 | 28.80 | 4.22 | 230.38 |
| $\tau \in (60, 90]$ | 11.68 | 21.70 | 40.82 | 11.05 | 3.05 | 88.30 |
| $\tau \in (90, 180]$ | 19.93 | 22.83 | 27.23 | 10.59 | 4.57 | 85.15 |
| $\tau \in (180, 365]$ | 10.74 | 10.91 | 10.82 | 4.54 | 5.12 | 42.12 |
| Total | 98.19 | 244.60 | 523.55 | 98.18 | 22.16 | 986.68 |
| Panel B: Volume of OTM option contracts (in millions) | | | | | | |
| $\tau \in (2, 9]$ | 4.12 | 52.03 | 139.33 | 11.94 | 0.96 | 208.37 |
| $\tau \in (9, 30]$ | 24.04 | 73.54 | 133.21 | 27.87 | 3.36 | 262.03 |
| $\tau \in (30, 60]$ | 23.82 | 61.11 | 83.72 | 27.28 | 3.85 | 199.79 |
| $\tau \in (60, 90]$ | 10.62 | 21.21 | 26.72 | 10.44 | 2.78 | 71.77 |
| $\tau \in (90, 180]$ | 18.77 | 22.15 | 18.23 | 9.92 | 4.30 | 73.35 |
| $\tau \in (180, 365]$ | 10.34 | 10.29 | 6.96 | 4.06 | 4.88 | 36.53 |
| Total | 91.71 | 240.32 | 408.17 | 91.51 | 20.14 | 851.85 |
| Panel C: Average OTM option price ($) | | | | | | |
| $\tau \in (2, 9]$ | 0.43 | 1.38 | 8.69 | 2.69 | 1.73 | 4.32 |
| $\tau \in (9, 30]$ | 0.92 | 5.53 | 23.53 | 5.47 | 2.68 | 10.07 |
| $\tau \in (30, 60]$ | 2.42 | 13.28 | 40.10 | 11.84 | 4.68 | 16.73 |
| $\tau \in (60, 90]$ | 5.05 | 24.89 | 63.56 | 22.00 | 6.17 | 24.93 |
| $\tau \in (90, 180]$ | 10.97 | 50.74 | 104.11 | 47.11 | 11.76 | 43.14 |
| $\tau \in (180, 365]$ | 18.80 | 90.42 | 155.53 | 85.29 | 19.67 | 53.05 |
| Total | 7.97 | 22.29 | 47.73 | 22.35 | 9.64 | 24.02 |

Note: Descriptive statistics for filtered option data on the S&P 500 stock market index. The sample contains daily option data from 1 May 2017 to 1 April 2021. Observations are bucketed into six categories based on the time-to-maturity, $\tau$, and into five categories with respect to the moneyness level, defined as strike-to-forward ratio $k = K/F$.

Given that we need a reliable and wide coverage of option prices for each tenor, we use a fairly generous set of filters. In particular, we retain option observations that satisfy the

Figure 2: Stacked bar chart of time-to-maturity frequency
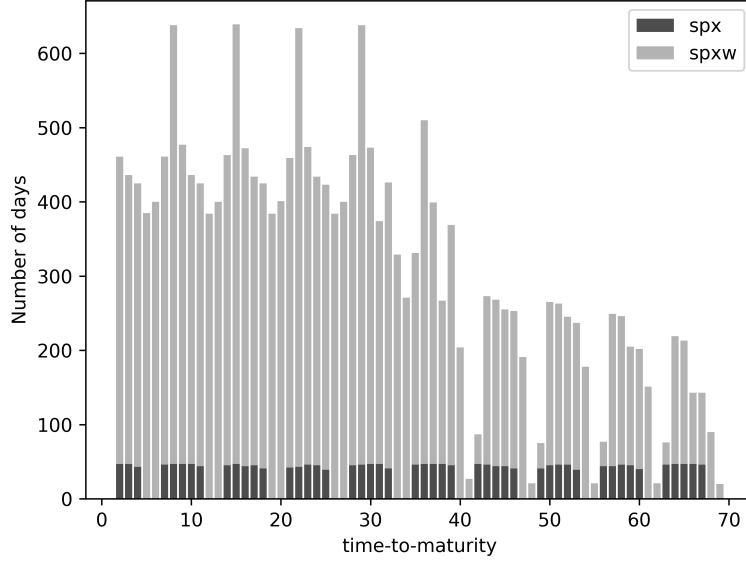


Note: This figure plots a stacked bar chart for the frequency of tenors in S&P 500 index options. The sample contains daily option data from 1 May 2017 to 1 April 2021 (constituting 978 trading days). The indicators 'spx' and 'spxw' correspond to AM-settled 'standard' and PM-settled 'weeklys' and end-of-month contracts, respectively.

following criteria: ($i$) bid price is strictly positive and ask-to-bid ratio is less than a factor 10; ($ii$) the maturity is larger than or equal to 2 calendar days, but less than or equal to 365 calendar days; ($iii$) it is not an early-closure day. The first criterion filters out illiquid observations and the second one limits our consideration in terms of options' maturity. The third criterion rules out shortened trading sessions, which in total constitute 10 days in our sample.

For each tenor, we determine the moneyness based on the forward index level, $F_t(\tau)$. For that, we use the put-call parity to calculate the forward price for close to at-the-money (ATM) options. Specifically, we use up to 5 option pairs with the smallest absolute difference between the call and put prices. The median of their forward-implied prices is taken as the forward index level for the corresponding tenor. The risk-free rates are obtained by interpolating the LIBOR rates to any particular tenor. Finally, given the calculated forward prices and moneyness levels, we retain only out-of-the-money (OTM) options for further exploitation. Descriptive statistics of the S&P 500 index options data sample are provided in Table 3. We observe that the largest portion of the trading volume is due to trades of OTM contracts and options with time-to-maturity less than 60 calendar days. Figure 2 plots the frequency of tenors up to 70 calendar days.

## 5.2 CCF-spanning option portfolios

The construction of the CCF-spanning option portfolios requires reliable option slices with wide coverage of strikes. Given that most of the trading volume is concentrated in option contracts with time-to-maturity of less than 60 days, our empirical application relies on the use of short-dated option slices with expiration period of no more than 2 months. In particular, on each

Table 4: Descriptive statistics for the selected sample

| Number | 1 | 2 | 3 | 4 | 5 | 6 | Total |
|---|---|---|---|---|---|---|---|
| avg. tenor | 6.97 | 12.10 | 18.95 | 26.08 | 33.95 | 53.67 | 25.29 |
| avg. min put | 0.08 | 0.09 | 0.10 | 0.11 | 0.11 | 0.14 | 0.10 |
| avg. min call | 0.09 | 0.09 | 0.10 | 0.11 | 0.12 | 0.14 | 0.11 |
| avg. max price | 22.48 | 31.65 | 40.24 | 47.56 | 54.83 | 71.05 | 44.64 |
| avg. # options | 133.75 | 183.64 | 204.47 | 206.90 | 216.50 | 238.84 | 197.35 |
| avg. min $K/F$ | 0.79 | 0.69 | 0.60 | 0.54 | 0.49 | 0.42 | 0.59 |
| avg. max $K/F$ | 1.06 | 1.09 | 1.12 | 1.14 | 1.16 | 1.21 | 1.13 |
| avg. ATM BSIV | 0.147 | 0.151 | 0.151 | 0.152 | 0.154 | 0.158 | 0.152 |

Note: Descriptive statistics for the selected data sample of options on the S&P 500 stock market index. The sample contains daily option data from 1 May 2017 to 1 April 2021. For each trading day, we select the six option tenors closest to 8, 15, 22, 29, 36 and 61 days from below with the largest trading volume and number of quoted OTM option contracts. The table provides the descriptive statistics for each of the six tenors over the sample.

trading day, we keep the six tenors closest to 8, 15, 22, 29, 36 and 61 days[13] from below with the largest trading volume and number of quoted OTM option contracts. Specifically, starting with the option slices closest to the indicated tenors, we compare them with every next shorter maturity option slice, and prefer the next one if it has a larger trading volume and a larger number of quoted contracts for OTM options. Table 4 provides the descriptive statistics for each of the six selected tenors over the considered time span. We notice the wide coverage of strikes, since the average minimum put and call prices are close to the tick size of $0.05, especially for very short-dated options. We also mention that in the selected option sample, each option slice at each trading day contains at least 55 different quoted contracts. Therefore, no additional filters on the minimum number of contracts are imposed. In total, we have 978 trading days, with six different tenors at each one of them, resulting in a total number of 1,158,059 contracts in the sample.

The inputs of our estimation procedure are option portfolios representing CCFs rather than BSIVs that are commonly used in the literature. Therefore, we pay careful attention to the construction of the option-implied CCF. As discussed in Section 3.1, we use a Riemann sum approximation to obtain a computationally feasible counterpart of the CCF spanning (3). However, in order to reduce the truncation and discretization errors, we further employ an interpolation-extrapolation technique. In particular, we interpolate option prices using cubic splines with carefully selected knot sequences and extrapolate beyond the observable range of strike prices using a parametrization that satisfies the asymptotic results of Lee (2004). The details of the interpolation-extrapolation scheme are provided in Appendix C.1.

The calculation of the option-implied CCF then uses the Riemann sum approximation (14) applied to the result of the interpolation-extrapolation scheme. The construction is conducted for each day and for each maturity separately. In particular, for equation (14), we set $\Delta m = 0.0001$ with a sufficiently wide range of log-moneyness between $\underline{m} = -6$ and $\overline{m} = 2$.

To conclude this section we emphasize again that, contrary to what is common in many

---

[13]The first five of these tenors are the most representative in the sample, see Figure 2.

Table 5: SVCDEJ estimation results

| | $\sigma$ | $\kappa$ | $\bar{v}$ | $\rho$ | $\delta$ | $\eta^+$ | $\eta^-$ | $\mu_v$ | $\sigma_\varkappa$ |
|---|---|---|---|---|---|---|---|---|---|
| $\widehat{\theta}$ | 0.5165 | 8.1700 | 0.0163 | -1.000 | 58.414 | 0.00086 | 0.0620 | 0.1256 | 0.373 |
| s.e. | 0.0257 | 1.8248 | 0.0028 | 0.030 | 15.780 | 0.00014 | 0.0060 | 0.0132 | – |

Note: This table provides the parameter estimates and standard errors for the SVCDEJ model. Descriptive statistics of the options data are in Table 4. The model is estimated based on $u = 1, \ldots, 15$ and $\bar{s} = 10^{-5}$, and with $p^- = 0.7$.

existing approaches, the option prices—or a monotonic transformation thereof—are not used as inputs in our developed estimation procedure. Instead, we use the option portfolios that replicate the CCF of log returns. Furthermore, unlike in many other papers, our option dataset is daily and utilizes the information from short-dated options with maturities between two days and two months.

# 6 Empirical Applications

Having thus constructed the dataset of option-implied CCFs for S&P 500 index options, we now illustrate our estimation procedure in two empirical applications, without and with an external factor.
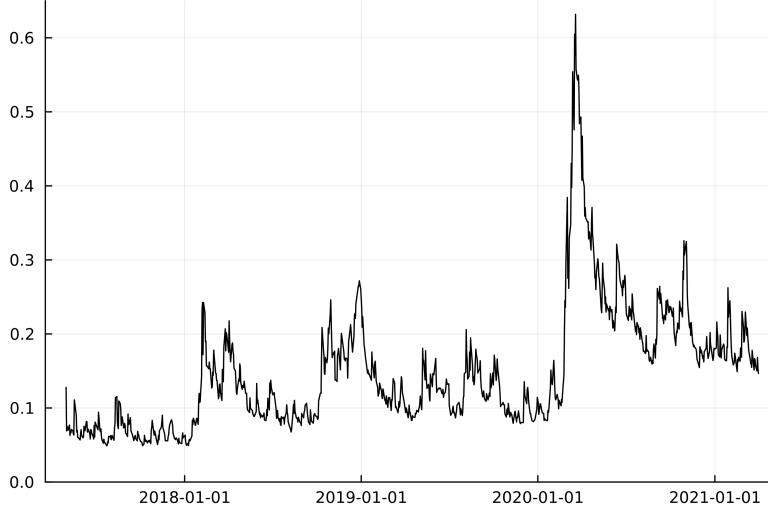
## 6.1 SVCDEJ

We start with estimating the SVCDEJ model specified in Section 4.1, (30)–(31), using the CCF-spanning option portfolios with six short-term tenors described in Section 5.2. Table 5 provides the parameter estimates. Informed by the Monte Carlo results, the estimates are based on the range of CCF arguments $u = 1, \ldots, 15$, a singular value threshold $\bar{s} = 10^{-5}$, and a fixed parameter $p^- = 0.7$. Standard errors are calculated using the familiar sandwich form covariance matrix.

The parameter estimates in Table 5 are meaningful, intuitive and broadly consistent with the literature. We note that the leverage parameter $\rho$ is estimated very close to its boundary value of $-1$, implying almost perfectly correlated diffusive components in returns and volatility. The empirical literature suggests that $\rho$ is negative and large in absolute value. The estimate of $\rho$ being nearly equal to its boundary value might be due to the use of short-dated options that typically exhibit steep implied volatility slopes. Indeed, Andersen et al. (2017) also find this correlation to be close to $-1$ in their dataset dominated by option contracts with maturities of less than 2 months.

Additional empirical results, analyzing the robustness of our empirical findings and model diagnostics, are provided in Appendix D.2. We note that the estimated measurement standard error $\sigma_\varkappa$ corresponds to a standard deviation of about 37% of the implied volatility. This is somewhat larger than what one might expect of measurement errors in option prices only, and might be interpreted to indicate e.g., missing state variables. In agreement with this, some of

Figure 3: SVCDEJ filtered volatility



Note: This figure plots the filtered volatility (i.e., the square root of the filtered state $\widehat{x}_{t+1|t}$) given the parameter estimates of the SVCDEJ model using Kalman filter recursions.

the extensions of the SVCDEJ model considered below and in Appendix D.2 show a slightly lower estimate of $\sigma_{\varkappa}$.

Figure 3 plots the filtered volatility (i.e., the square root of the filtered state $\widehat{x}_{t+1|t}$) given the parameter estimates of the SVCDEJ model. As is clearly visible, the filtered volatility exhibits a relatively stable volatility regime prior to 2020 and jumps up in March 2020 at the outbreak of the Covid-19 pandemic.

## 6.2 SVCDEJ with external factors

Now we turn to model specifications with embedded external factors. In some situations, we might have specific information on possible drivers of the risks in the market, and would like to quantify their effect.

An example is the recent Covid-19 crisis. The Covid-19 pandemic has dramatically affected our lives. It has also had a tremendous impact on the world's economy and financial markets. The beginning of the pandemic, in particular, was associated with a spike in uncertainty. This uncertainty surrounded many aspects: the contagiousness and lethality of the virus, the time required to develop vaccines, the effectiveness of measures, the work-from-home policies, travel bans, and so on. In this application, we explore the impact of the Covid-19 pandemic on the stock market through the lens of option prices. In particular, we consider how the spread of the virus affected the likelihood of jump events and the volatility in the U.S. stock market.

Figure 4, Panel (a), plots the daily cases of Covid-19 infections around the world obtained from the World Health Organization (WHO). The reported number of daily cases, however, does not represent well the contagiousness of the virus. Therefore, Panel (b) of Figure 4 displays the so-called reproduction number $R_t$, according to two measures: the first one is taken from the website 'Our World in Data' and the second one is calculated as the ratio $R_t = I_t/I_{t-7}$, where $I_t$ is the number of infected people in day $t$ and 7 is the reported serial interval for Covid-

Figure 4: Covid-19 daily cases and reproduction numbers



| (a) Daily cases | (b) Reproduction numbers |
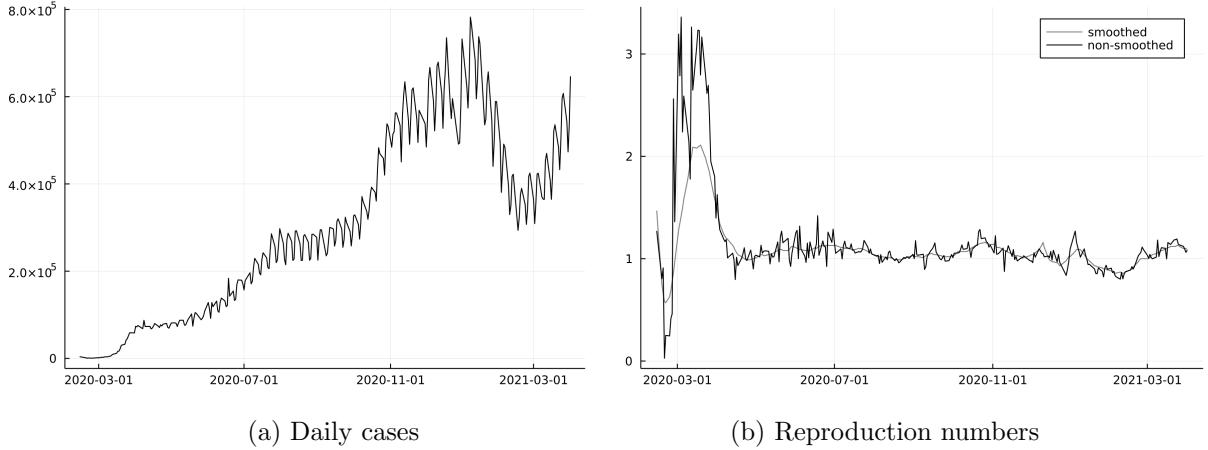
Note: This figure plots the daily Covid-19 cases (worldwide) and two reproduction numbers: the first is taken from the website 'Our World in Data' (using methodology of Arroyo-Marioli, Bullano, Kucinskas, and Rondón-Moreno (2021)); the second is calculated as the ratio $I_t/I_{t-7}$, where $I_t$ is number of infected people in day $t$ and 7 is the reported serial interval for Covid-19. The sample period runs from 14 February 2020 to 1 April 2021.

19. The former is based on the parametric methodology of Arroyo-Marioli et al. (2021) and is smoothed over time.[14] The latter is non-smoothed and based on the assumption that the serial interval is 7 days, which is consistent with the recent epidemiology literature (see, e.g., Maier & Brockmann, 2020, Prem et al., 2020, Flaxman et al., 2020, Arroyo-Marioli et al., 2021). We will use the latter non-parametric and non-smoothed measure as the reproduction number in our application.

To quantify the effect of Covid-19 propagation on the financial market, we embed the reproduction number as an external factor into the (time-varying) levels of the stochastic volatility and jump intensity processes, as described in Section 4.2, equations (36)–(38), with $h_t$ replaced by $R_t$. Given that the reproduction number constitutes a relatively persistent process, we will treat it as a deterministic process when pricing options; in other words, we follow the second estimation approach described in Section 4.2. In a similar way, the risk-free rate and dividend yields are often assumed to be deterministic in the option pricing literature. This allows us to be agnostic about the parametric dynamics of the reproduction number. Furthermore, given the short-dated options under consideration, the errors due to this deterministic treatment are likely to be negligible.[15]

Table 6 provides the parameter estimates for the SVCDEJ model with Covid-19 reproduction numbers as an exogenous factor. With $q$ estimated at 0.0001, the results indicate that the reproduction number dynamics have no substantial effect on the total diffusive volatility. A one unit increase in reproduction numbers, however, leads to an increase in the intensity of jumps by $\gamma$ which is estimated at 1.427. In other words, the reproduction number contributes substantially to the likelihood of jumps. Figure 5 illustrates the dynamics of the jump intensity without and with the added effect of reproduction numbers. Estimation based on either short-

---

[14]In fact, Arroyo-Marioli et al. (2021) use a Kalman smoother.

[15]Similarly, Andersen et al. (2017) and Boswijk et al. (2021) consider an approximation of the return process with 'frozen' spot volatility when estimating their option pricing models with short-dated options.

Table 6: SVCDEJ estimation results with Covid-19 reproduction numbers as external factor

| | $\sigma$ | $\kappa$ | $\bar{v}$ | $\rho$ | $\delta$ | $\eta^+$ | $\eta^-$ | $\mu_v$ | $\gamma$ | $q$ | $\sigma_\varkappa$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\widehat{\theta}$ | 0.568 | 11.416 | 0.0141 | -1.000 | 56.418 | 0.0000 | 0.0573 | 0.141 | 1.427 | 0.0001 | 0.354 |
| s.e. | 0.028 | 2.213 | 0.0021 | 0.009 | 19.138 | 0.0008 | 0.0063 | 0.023 | 0.413 | 0.0001 | – |

Note: This table provides the parameter estimates and standard errors for the SVCDEJ model with Covid-19 reproduction numbers as external factor. The reproduction numbers are set to zero before 14 February 2020, and are taken to be the ratios $R_t = I_t/I_{t-7}$ starting from 14 February 2020. Descriptive statistics of the options data are in Table 4. The model is estimated based on $u = 1, \ldots, 15$ and $\bar{s} = 10^{-5}$, and with $p^- = 0.7$.

Figure 5: SVCDEJ jump intensity with Covid-19 reproduction numbers as external factor
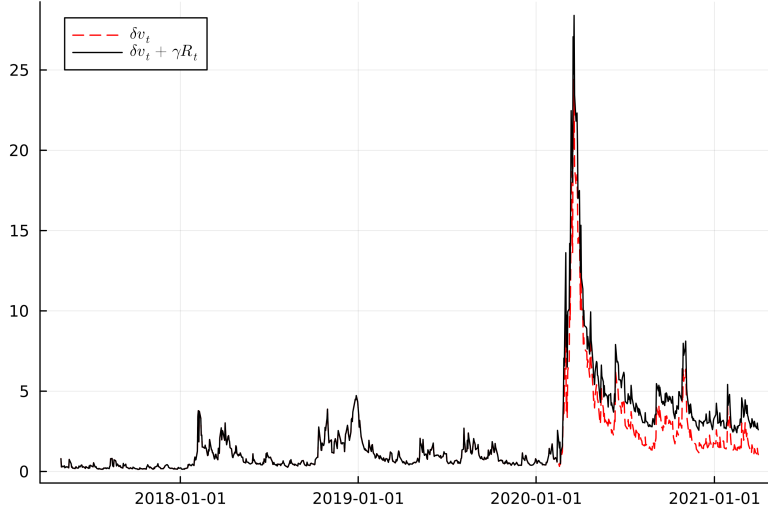


Note: This figure plots the filtered and total jump intensity given the parameter estimates of the SVCDEJ model with Covid-19 reproduction number dynamics.

dated or longer-dated options is considered in Appendix D.2.

It is also possible to investigate the contribution of other external factors to the diffusive volatility and jump intensity. As an example, we provide in Table 7 estimation results for the SVCDEJ model with the Economic Policy Uncertainty (EPU) index embedded as an external factor. The EPU index, developed by Baker, Bloom, and Davis (2016), reflects policy-related economic uncertainty based on newspaper coverage frequency. The estimation results indicate that, unlike the reproduction number, the EPU index has no effect on the jump intensity process, but contributes significantly to the total diffusive volatility of the model, with $q$ estimated at 0.0249; see also Figure 6. Thus, whereas Covid-19 reproduction numbers contribute substantially to the jump intensity dynamics, the policy uncertainty index EPU contributes significantly to the total diffusive volatility.

# 7 Conclusion

We have proposed a novel state filtering and parameter estimation procedure for option pricing models that belong to the affine jump-diffusion class. Our procedure utilizes the log of the option-implied and model-free conditional characteristic function and the model-implied condi-

Table 7: SVCDEJ estimation results with the EPU index as external factor

| | $\sigma$ | $\kappa$ | $\bar{v}$ | $\rho$ | $\delta$ | $\eta^+$ | $\eta^-$ | $\mu_v$ | $\gamma$ | $q$ | $\sigma_\varkappa$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\widehat{\theta}$ | 0.504 | 8.890 | 0.0143 | -1.000 | 79.633 | 0.0001 | 0.0581 | 0.099 | 0.000 | 0.0249 | 0.373 |
| s.e. | 0.016 | 0.379 | 0.0009 | 0.019 | 12.347 | 0.0019 | 0.0030 | 0.009 | 0.029 | 0.0035 | – |

Note: This table provides the parameter estimates and standard errors for the SVCDEJ model with the EPU index as external factor. Descriptive statistics of the options data are in Table 4. The model is estimated based on $u = 1, \ldots, 15$ and $\bar{s} = 10^{-5}$, and with $p^- = 0.7$.

Figure 6: SVCDEJ diffusive volatility with the EPU index as external factor
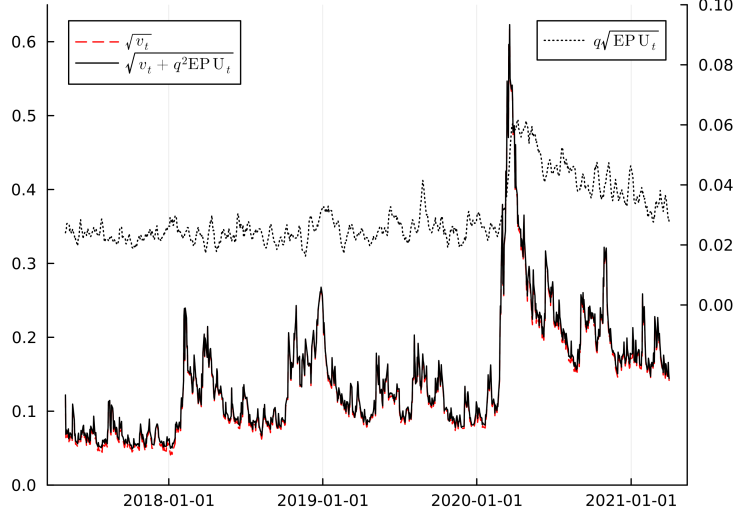


Note: This figure plots the filtered and total diffusive volatility given the parameter estimates of the SVCDEJ model with EPU index dynamics. The total diffusive volatility equals $\sqrt{v_t + q^2 h_t}$, where $h_t$ is the external factor given by the EPU index. (The secondary vertical axis shows the values of $q\sqrt{\text{EPU}_t}$.)

tional log-characteristic function, which is functionally affine in the model's state vector. We have developed a linear state space representation for the considered class of option pricing models, which allows us to exploit suitably modified collapsed Kalman filtering techniques. Our estimation procedure is fast and easy to implement, circumventing the typical computational burden when working with option pricing models. We have demonstrated the applicability of our procedure in two empirical illustrations that analyze S&P 500 index options and the impact of exogenous variables capturing Covid-19 reproduction and economic policy uncertainty data.

Although we have focused on Gaussian QML estimation based on Kalman filtering techniques, which delivers good results in our Monte Carlo simulations, the same state space formulation can also be analyzed by more refined methods such as those based on particle filters; see, e.g., Johannes et al. (2009), Christoffersen, Dorion, Jacobs, and Karoui (2014) and Bardgett et al. (2019). Such methods could exploit the non-Gaussianity and heteroskedasticity in the data to obtain more efficient estimates, at the cost of some increased computational complexity. We note that such extensions would still not require option price evaluation by the FFT or COS methods, and thus retain an important advantage of our approach.

Our proposed estimation procedure in principle allows for identification and estimation of factor risk premium parameters, by combining the risk-neutral parameters entering the mea-

surement equation with the objective parameters entering the transition equation. Monte Carlo simulation results suggest, however, that option price data are not very informative about such risk premia, which is why we have concentrated on the case where the objective and risk-neutral measures coincide. Fortunately, the simulation results also suggest that inference on the risk-neutral parameters is quite robust with respect to deviations from this assumption. For more focused inference on (volatility) risk premium parameters, it may be possible to combine the information in daily option prices as considered in this paper with realized measures based on high-frequency returns on the underlying. We intend to explore this in future research.

# Appendix A  Proofs

In this appendix, we provide the proof of Propositions 1 and 2. First, we state and prove some preliminary results.

## A.1  Preliminary results

We start by formally defining the measurement errors in the CCF approximation. Under the observation error structure specified in Assumption 2 and the CCF approximation given by equation (14), the total measurement error in the option-spanned CCF may be written as

$$
\zeta_t^\phi(u,\tau) := \widehat{\phi}_t(u,\tau) - \phi_t(u,\tau)
$$

$$
= -u_t \sum_{j=2}^n e^{(iu-1)m_j} \cdot \widehat{O}_t(\tau,m_j)\Delta m_j + u_t \int_{-\infty}^\infty e^{(iu-1)m} \cdot O_t(\tau,m)\mathrm{d}m
$$

$$
= -u_t \left[ \sum_{j=2}^n e^{(iu-1)m_j} \cdot O_t(\tau,m_j)\Delta m_j + \sum_{j=2}^n e^{(iu-1)m_j} \cdot \zeta_t(\tau,m_j)\Delta m_j \right]
$$

$$
+ u_t \int_{-\infty}^\infty e^{(iu-1)m} \cdot O_t(\tau,m)\mathrm{d}m,
$$

which after reordering of terms can be decomposed into the following three components:

$$
\zeta_t^\phi(u,\tau) = \underbrace{-u_t \sum_{j=2}^n e^{(iu-1)m_j} \cdot \zeta_t(\tau,m_j)\Delta m_j}_{=:\zeta_t^{(1)}(u,\tau)}
$$

$$
+ \underbrace{u_t \int_{-\infty}^{m_1} e^{(iu-1)m} \cdot O_t(\tau,m)\mathrm{d}m + u_t \int_{m_n}^\infty e^{(iu-1)m} \cdot O_t(\tau,m)\mathrm{d}m}_{=:\zeta_t^{(2)}(u,\tau)}
$$

$$
+ \underbrace{u_t \sum_{j=2}^n \int_{m_{j-1}}^{m_j} \left[ e^{(iu-1)m} \cdot O_t(\tau,m) - e^{(iu-1)m_j} \cdot O_t(\tau,m_j) \right] \mathrm{d}m}_{=:\zeta_t^{(3)}(u,\tau)}
$$

$$
= \zeta_t^{(1)}(u,\tau) + \zeta_t^{(2)}(u,\tau) + \zeta_t^{(3)}(u,\tau). \tag{A.1}
$$

The error terms $\zeta_t^{(1)}(u,\tau)$, $\zeta_t^{(2)}(u,\tau)$ and $\zeta_t^{(3)}(u,\tau)$ represent observation, truncation and discretization errors, respectively. In order to characterize the asymptotic orders of these errors, we make use of the following auxiliary result:

**Lemma 1** *Let $f_{t+\tau} = \frac{F_{t+\tau}}{F_t}$ be the futures price normalized to its value at time $t$ for $\tau > 0$. For all $m > 0$, we have the call price bounds*

$$
\frac{O_t(\tau,m)}{F_t} \le \left( \frac{\overline{\delta}}{\overline{\delta}+1} \right)^{\overline{\delta}} \frac{e^{-r\tau}\mathbb{E}^\mathbb{Q}[f_{t+\tau}^{\overline{\delta}+1}|\mathcal{F}_t]}{\overline{\delta}+1} e^{-\overline{\delta}m}, \tag{A.2}
$$

*for each $\overline{\delta} > 0$. Similarly, for all $m < 0$, we have the put price bounds*

$$
\frac{O_t(\tau,m)}{F_t} \le \left( \frac{\underline{\delta}}{\underline{\delta}+1} \right)^{\underline{\delta}} \frac{e^{-r\tau}\mathbb{E}^\mathbb{Q}[f_{t+\tau}^{-\underline{\delta}}|\mathcal{F}_t]}{\underline{\delta}+1} e^{(1+\underline{\delta})m}, \tag{A.3}
$$

*for each $\underline{\delta} > 0$.*

*Proof:* The result is a straightforward adaptation of Theorem 2.1 in Lee (2004). □

Lemma 1 relates moments of the underlying process and of its reciprocal to bounds on option prices. We assume the existence of moments of order $1 + \bar{\delta}$ and $\underline{\delta}$ for the underlying process and its reciprocal, formally stated in Assumption 3($i$). If higher moments exist, as assumed by Qin and Todorov (2019) and Todorov (2019), then we can obtain even tighter bounds for the remainder term in Proposition 1 due to (A.2) and (A.3).

The following lemma establishes the order of magnitude of the truncation and discretization errors under the joint asymptotic scheme, expressed with respect to the number of option prices $n$ with fixed maturity. As in the main text, we denote the smallest and largest log-moneyness by $\underline{m} = \min_{1 \leq j \leq n} m_j$ and $\overline{m} = \max_{1 \leq j \leq n} m_j$, and the corresponding strike prices by $\underline{K}$ and $\overline{K}$. In the proofs, we denote by $C_t$ an $\mathcal{F}_t$-adapted random variable that does not depend on $m$ and that may change from line to line.

**Lemma 2** *Suppose* $\mathbb{E}^{\mathbb{Q}}[F_{t+\tau}^{\bar{\delta}+1}|\mathcal{F}_t] < \infty$ *and* $\mathbb{E}^{\mathbb{Q}}[F_{t+\tau}^{-\underline{\delta}}|\mathcal{F}_t] < \infty$ *for some* $\bar{\delta} > 0$ *and* $\underline{\delta} > 0$, *Assumption 3(ii) holds, and in addition* $\underline{K} \asymp n^{-\underline{\alpha}}$ *and* $\overline{K} \asymp n^{\overline{\alpha}}$ *with* $\underline{\alpha} > 0$ *and* $\overline{\alpha} > 0$. *Then, as* $n \to \infty$, *we have*

$$\zeta_t^{(2)}(u, \tau) = \mathcal{O}_p\left(n^{-(\underline{\delta}\,\underline{\alpha} \wedge (1+\bar{\delta})\overline{\alpha})}\right), \tag{A.4}$$

*and*

$$\zeta_t^{(3)}(u, \tau) = \mathcal{O}_p\left(\frac{\log n}{n}\right), \tag{A.5}$$

*for a fixed* $u \in \mathcal{U}$ *and* $\tau > 0$.

*Proof:* We start with the truncation errors. For $\underline{m} = m_1 < \ldots < m_n = \overline{m}$, with $\underline{m} < 0$ and $\overline{m} > 0$, and using Lemma 1, we can bound the upper and lower truncation parts as follows:

$$\left|\frac{1}{F_t}\int_{\overline{m}}^{\infty} e^{(iu-1)m} \cdot O_t(\tau, m)\mathrm{d}m\right| \leq \int_{\overline{m}}^{\infty} \left|e^{(iu-1)m}\right| \cdot \left|\frac{O_t(\tau, m)}{F_t}\right|\mathrm{d}m \leq C_t e^{-(1+\bar{\delta})\overline{m}},$$

$$\left|\frac{1}{F_t}\int_{-\infty}^{\underline{m}} e^{(iu-1)m} \cdot O_t(\tau, m)\mathrm{d}m\right| \leq \int_{-\infty}^{\underline{m}} \left|e^{(iu-1)m}\right| \cdot \left|\frac{O_t(\tau, m)}{F_t}\right|\mathrm{d}m \leq C_t e^{\underline{\delta}\,\underline{m}},$$

where, as mentioned before, $C_t$ is independent of $m$ and may vary from line to line. Therefore, as $\underline{m} \to -\infty$ and $\overline{m} \to \infty$, we have

$$\zeta_t^{(2)}(u, \tau) = u_t \int_{-\infty}^{\underline{m}} e^{(iu-1)m} \cdot O_t(\tau, m)\mathrm{d}m + u_t \int_{\overline{m}}^{\infty} e^{(iu-1)m} \cdot O_t(\tau, m)\mathrm{d}m$$

$$= \mathcal{O}_p\left(e^{-\underline{\delta}|\underline{m}|}\right) + \mathcal{O}_p\left(e^{-(1+\bar{\delta})|\overline{m}|}\right) = \mathcal{O}_p\left(e^{-(\underline{\delta}|\underline{m}| \wedge (1+\bar{\delta})|\overline{m}|)}\right) = \mathcal{O}_p\left(n^{-(\underline{\delta}\,\underline{\alpha} \wedge (1+\bar{\delta})\overline{\alpha})}\right).$$

For the discretization errors, we use the following decomposition:

$$\int_{m_{j-1}}^{m_j} \left[e^{(iu-1)m} \cdot O_t(\tau, m) - e^{(iu-1)m_j} \cdot O_t(\tau, m_j)\right]\mathrm{d}m$$

$$= \int_{m_{j-1}}^{m_j} \left[\left(e^{(iu-1)m} - e^{(iu-1)m_j}\right) \cdot O_t(\tau, m_j) + e^{(iu-1)m}\left(O_t(\tau, m) - O_t(\tau, m_j)\right)\right]\mathrm{d}m.$$

By applying the mean value theorem twice, we have

$$\left| e^{(iu-1)m} - e^{(iu-1)m_j} \right| \le |iu - 1| \left| e^{(iu-1)\tilde{m}} \right| \Delta m_j \le e^{-m_{j-1}}(|u| \vee 1)\Delta m_j,$$

and

$$\left| e^{(iu-1)m} \left( O_t(\tau, m) - O_t(\tau, m_j) \right) \right| \le e^{-m_{j-1}} \left| \frac{\partial O_t(\tau, m)}{\partial m} \right|_{m=\tilde{m}} \left| \Delta m_j \right.$$

$$\le e^{-m_{j-1}} \left| \frac{\partial O_t(\tau, m)}{\partial K} \right|_{K=\tilde{K}} \left| e^{\tilde{m}} F_t \Delta m_j \right.$$

$$\le \left| \frac{\partial O_t(\tau, m)}{\partial K} \right|_{K=\tilde{K}} \left| e^{\Delta m_j} F_t \Delta m_j, \right.$$

where $\tilde{m} = \log(\tilde{K}/F_t)$ lies between $m$ and $m_j$.

For the first term in the decomposition above, we use that Lemma 1 implies that, for all $m$,

$$\frac{O_t(\tau, m)}{F_t} \le C_t e^{-(\bar{\delta}m \vee -(1+\underline{\delta})m)}.$$

Furthermore, for the second term in the decomposition, we exploit the fact that the derivative with respect to the strike price is the risk neutral distribution or survival function, which can be bounded using the Markov inequality. In particular, for $m > 0$,

$$\left| \frac{\partial O_t(\tau, m)}{\partial K} \right| = e^{-r\tau} \mathbb{Q}(F_{t+\tau} > K) = e^{-r\tau} \mathbb{Q}\left( f_{t+\tau} > e^m \right)$$

$$\le e^{-r\tau} \mathbb{E}^{\mathbb{Q}}[f_{t+\tau}^{\bar{\delta}+1} | \mathcal{F}_t] e^{-(\bar{\delta}+1)m},$$

and, for $m < 0$,

$$\left| \frac{\partial O_t(\tau, m)}{\partial K} \right| = e^{-r\tau} \mathbb{Q}(F_{t+\tau} < K) = e^{-r\tau} \mathbb{Q}\left( f_{t+\tau}^{-1} > e^{-m} \right)$$

$$\le e^{-r\tau} \mathbb{E}^{\mathbb{Q}}[f_{t+\tau}^{-\underline{\delta}} | \mathcal{F}_t] e^{\underline{\delta}m}.$$

Therefore,

$$\left| \frac{\partial O_t(\tau, m)}{\partial K} \right| \le C_t e^{-((\bar{\delta}+1)m \vee -\underline{\delta}m)}.$$

Combining all these inequalities together, we obtain

$$\left| \frac{1}{F_t} \int_{m_{j-1}}^{m_j} \left[ e^{(iu-1)m} \cdot O_t(\tau, m) - e^{(iu-1)m_j} \cdot O_t(\tau, m_j) \right] dm \right|$$

$$= \left| \frac{1}{F_t} \int_{m_{j-1}}^{m_j} \left[ \left( e^{(iu-1)m} - e^{(iu-1)m_j} \right) \cdot O_t(\tau, m_j) + e^{(iu-1)m} \left( O_t(\tau, m) - O_t(\tau, m_j) \right) \right] dm \right|$$

$$\le \left( C_t(|u| \vee 1)\Delta m_j e^{-m_{j-1}} e^{-[\bar{\delta}m_{j-1} \vee -(1+\underline{\delta})m_{j-1}]} + C_t e^{\Delta m} e^{-[(\bar{\delta}+1)m_{j-1} \vee -\underline{\delta}m_{j-1}]} \Delta m_j \right) \Delta m_j$$

$$\le C_t e^{-[(\bar{\delta}+1)m_{j-1} \vee -\underline{\delta}m_{j-1}]} (\Delta m_j)^2.$$

Then, for fixed $\overline{m}$ and $\underline{m}$, and $\Delta m \to 0$, we have

$$\zeta_t^{(3)}(u, \tau) = u_t \sum_{j=2}^{n} \int_{m_{j-1}}^{m_j} \left[ e^{(iu-1)m} \cdot O_t(\tau, m) - e^{(iu-1)m_j} \cdot O_t(\tau, m_j) \right] dm = \mathcal{O}_p(\Delta m). \quad (A.6)$$

The result (A.6), however, can be adapted for the joint asymptotic scheme, where $\overline{m}$, $\underline{m}$ and $\Delta m$ all depend on $n$, with $n \to \infty$. To this end, we first note that

$$\left| \sum_{j=2}^{n} \frac{1}{F_t} \int_{m_{j-1}}^{m_j} \left[ e^{(iu-1)m} \cdot O_t(\tau, m) - e^{(iu-1)m_j} \cdot O_t(\tau, m_j) \right] dm \right|$$

$$\leq C_t \sum_{j=2}^{n} e^{-[(\overline{\delta}+1)m_{j-1} \vee -\underline{\delta} m_{j-1}]} (\Delta m_j)^2$$

$$\leq C_t \Delta m \sum_{j=2}^{n} e^{-[(\overline{\delta}+1)m_{j-1} \vee -\underline{\delta} m_{j-1}]} \Delta m_j.$$

The sum on the far right-hand side of the inequality is a Riemann approximation that converges to the following integral, as $n \to \infty$:

$$\sum_{j=2}^{n} e^{-[(\overline{\delta}+1)m_{j-1} \vee -\underline{\delta} m_{j-1}]} \Delta m_j \longrightarrow \int_{\underline{m}}^{\overline{m}} e^{-[(\overline{\delta}+1)m \vee -\underline{\delta} m]} dm = \int_{\underline{m}}^{0} e^{\underline{\delta} m} dm + \int_{0}^{\overline{m}} e^{-(\overline{\delta}+1)m} dm$$

$$= \frac{1}{\underline{\delta}} \left( 1 - e^{-\underline{\delta} \underline{\alpha}} \right) + \frac{1}{\overline{\delta}+1} \left( 1 - e^{-(\overline{\delta}+1)\overline{\alpha}} \right)$$

$$= \mathcal{O}_p(1)$$

Next, given Assumption 3(ii) on the log-moneyness grid, we can bound $\Delta m$ as

$$\frac{\overline{m} - \underline{m}}{\eta n} \geq \Delta m \geq \frac{\overline{m} - \underline{m}}{n}.$$

Hence, $\Delta m = \mathcal{O}_p \left( n^{-1} \log n \right)$. Thus, the order of magnitude of the discretization errors under the joint asymptotic scheme is given by the order of $\Delta m$:

$$\zeta_t^{(3)}(u, \tau) = \mathcal{O}_p (\Delta m) = \mathcal{O}_p \left( \frac{\log n}{n} \right).$$

$\square$

## A.2 Proof of Proposition 1

Using Lemma 2, Assumption 3 on the moments of the underlying process and observation error Assumption 2, we can decompose the measurement errors in the CCF approximation as

$$\widehat{\phi}_t(u, \tau) - \phi_t(u, \tau) = \zeta_t^{(1)}(u, \tau) + \mathcal{O}_p \left( n^{-(\underline{\delta} \underline{\alpha} \wedge (1+\overline{\delta})\overline{\alpha})} \vee \frac{\log n}{n} \right), \quad (A.7)$$

with

$$\zeta_t^{(1)}(u, \tau) = -u_t \sum_{j=2}^{n} e^{(iu-1)m_j} \cdot \zeta_t(\tau, m_j) \Delta m_j.$$

We now show that $\zeta_t^{(1)}(u, \tau) = \mathcal{O}_p\left(\sqrt{n^{-1} \log n}\right)$. In fact, the standard deviation of the observation errors is proportional to the Black-Scholes vega, which decreases with $|m| \to \infty$. More specifically, the vega is given by

$$\nu_t(\tau, m) = F_t \sqrt{\tau} \varphi(d_+),$$

$$d_+ = -m \varpi^{-1/2}(m) + \frac{1}{2} \varpi^{1/2}(m),$$

where $\varphi(x)$ is the standard normal pdf and $\varpi(m) := \kappa^2(\tau, m)\tau$ is the total implied variance.[16] Hence,

$$\nu_t^2(\tau, m) = F_t^2 \tau \frac{1}{2\pi} e^{-d_+^2} = F_t^2 \tau \frac{1}{2\pi} e^{-(\varpi^{-1}(m)m^2 - m + \frac{1}{4}\varpi(m))}.$$

Therefore, given Assumption 2, we obtain

$$\mathbb{E}\left[\left|\zeta_t^{(1)}(u, \tau)\right|^2 \Big| \mathcal{F}_t\right] \leq |u_t|^2 \sum_{j=2}^{n} e^{-2m_j} \cdot \mathbb{E}\left[\zeta_t(\tau, m_j)^2 | \mathcal{F}_t\right] (\Delta m_j)^2$$

$$\leq |u_t|^2 \sum_{j=2}^{n} e^{-2m_j} \cdot \sigma_\varkappa^2 \kappa^2(\tau, m_j) F_t^2 \tau \frac{1}{2\pi} e^{-(\varpi_j^{-1} m_j^2 - m_j + \frac{1}{4}\varpi_j)} (\Delta m_j)^2$$

$$\leq C_t \Delta m \sum_{j=2}^{n} \varpi_j e^{-\varpi_j^{-1} m_j^2 - m_j - \frac{1}{4}\varpi_j} \Delta m_j$$

$$\leq C_t \Delta m \sum_{j=2}^{n} \varpi_j e^{-d_-^2(m_j)} \Delta m_j,$$

where $d_-(m) := -m \varpi^{-1/2}(m) - \frac{1}{2}\varpi^{1/2}(m)$ and $\varpi_j := \varpi(m_j)$. Then, as $n \to \infty$, the right-hand side summation converges to

$$\int_{-\infty}^{\infty} \varpi(m) \exp\left(-d_-^2(m)\right) \mathrm{d}m =: \int_{-\infty}^{\infty} h(m) \mathrm{d}m,$$

provided that the function $h(m)$ is integrable.

To show the latter, we focus on the tail behavior, since $h$ is continuous and hence bounded on the bounded interval $[\underline{m}, \overline{m}]$. For that, we will make use of the following asymptotic results of Lee (2004):

$$\limsup_{m \to -\infty} \frac{\varpi(m)}{|m|} = \underline{\beta}^* \text{ with } \underline{\beta}^* \in [0, 2], \quad \text{and} \quad \frac{1}{2\underline{\beta}^*} + \frac{\underline{\beta}^*}{8} - \frac{1}{2} = \sup\{q : \mathbb{E}[F_{t+\tau}^{-q} | \mathcal{F}_t] < \infty\},$$

$$\limsup_{m \to \infty} \frac{\varpi(m)}{|m|} = \overline{\beta}^* \text{ with } \overline{\beta}^* \in [0, 2], \quad \text{and} \quad \frac{1}{2\overline{\beta}^*} + \frac{\overline{\beta}^*}{8} - \frac{1}{2} = \sup\{p : \mathbb{E}[F_{t+\tau}^{1+p} | \mathcal{F}_t] < \infty\}.$$

That is, for $m < 0$, the total implied variance $\varpi(m)$ grows at most as fast as $-\underline{\beta}^* m$ for some $\underline{\beta}^* \in [0, 2]$. Given Assumption 3 on the moments of the reciprocal process, we further have $\underline{\beta}^* < 1$, which implies that $h$ is integrable over the negative domain, as for sufficiently small $m$,

$$0 \leq h(m) = \varpi(m) \exp\left(-\varpi^{-1}(m)m^2 - m - \frac{1}{4}\varpi(m)\right) \leq -\underline{\beta}^* m \exp\left(\frac{m}{\underline{\beta}^*} - m\right).$$

---

[16] The total implied variance $\varpi(m)$ is a function of both the moneyness level $m$ and the time-to-maturity $\tau$. For ease of notation, we write it as a function of moneyness only since the time-to-maturity $\tau$ is fixed in our exposition.

Integrability over the positive domain is achieved even without exploiting moment conditions. Therefore, since the right-hand side summation converges under the joint asymptotic scheme to the integral above, we have that $\zeta_t^{(1)}(u, \tau) = \mathcal{O}_p\left(\sqrt{\Delta m}\right) = \mathcal{O}_p\left(\sqrt{n^{-1}\log n}\right)$.

Furthermore, from Assumption 2, it also follows that $\mathbb{E}[\zeta_t^{(1)}(u, \tau)|\mathcal{F}_t] = 0$, while the discretization and truncation errors $\zeta_t^{(2)}(u, \tau)$ and $\zeta_t^{(3)}(u, \tau)$ are $\mathcal{F}_t$-measurable. Hence, the covariance and pseudo-covariance terms of the CCF approximation are given by the second moments of the observation errors $\zeta_t^{(1)}(u, \tau)$, that is,

$$
\begin{aligned}
\operatorname{Cov}(\zeta_t^{\phi}(u_i, \tau), \zeta_t^{\phi}(u_j, \tau)) :&= \mathbb{E}\left[\left(\zeta_t^{\phi}(u_i, \tau) - \mathbb{E}[\zeta_t^{\phi}(u_i, \tau)]\right) \overline{\left(\zeta_t^{\phi}(u_j, \tau) - \mathbb{E}[\zeta_t^{\phi}(u_j, \tau)]\right)}\bigg|\mathcal{F}_t\right] \\
&= \mathbb{E}\left[\zeta_t^{(1)}(u_i, \tau)\zeta_t^{(1)}(-u_j, \tau)\big|\mathcal{F}_t\right] \\
&= u_{i,t}\overline{u_{j,t}} \sum_{j=2}^{n} e^{(\mathrm{i}(u_i - u_j) - 2)m_j} \cdot \sigma_t^2(\tau, m_j)(\Delta m_j)^2 \\
&= \sigma_{\varkappa}^2 \cdot u_{i,t}\overline{u_{j,t}} \underbrace{\sum_{j=2}^{n} e^{(\mathrm{i}(u_i - u_j) - 2)m_j} \cdot \kappa_t^2(\tau, m_j)\nu_t^2(\tau, m_j)(\Delta m_j)^2}_{=:\gamma_t(u_i, u_j, \tau)} \\
&= \sigma_{\varkappa}^2 \cdot \gamma_t(u_i, u_j, \tau),
\end{aligned}
$$

and

$$
\begin{aligned}
\operatorname{PCov}(\zeta_t^{\phi}(u_i, \tau), \zeta_t^{\phi}(u_j, \tau)) :&= \mathbb{E}\left[\left(\zeta_t^{\phi}(u_i, \tau) - \mathbb{E}[\zeta_t^{\phi}(u_i, \tau)]\right) \left(\zeta_t^{\phi}(u_j, \tau) - \mathbb{E}[\zeta_t^{\phi}(u_j, \tau)]\right)\bigg|\mathcal{F}_t\right] \\
&= \mathbb{E}\left[\zeta_t^{(1)}(u_i, \tau)\zeta_t^{(1)}(u_j, \tau)\big|\mathcal{F}_t\right] \\
&= \sigma_{\varkappa}^2 \cdot u_{i,t}u_{j,t} \underbrace{\sum_{j=2}^{n} e^{(\mathrm{i}(u_i + u_j) - 2)m_j} \cdot \kappa_t^2(\tau, m_j)\nu_t^2(\tau, m_j)(\Delta m_j)^2}_{=:c_t(u_i, u_j, \tau)} \\
&= \sigma_{\varkappa}^2 \cdot c_t(u_i, u_j, \tau),
\end{aligned}
$$

for any $u_i, u_j \in \mathcal{U}$, where $\overline{z}$ denotes the complex conjugate of a complex number $z \in \mathbb{C}$. In other words, the covariances of the total measurement errors in the CCF approximation are determined by the properties of the observation errors in option prices only. Note that the terms $\gamma_t(u_i, u_j, \tau)$ and $c_t(u_i, u_j, \tau)$ depend only on option's characteristics such as BSIV, BS vega and moneyness levels. That is, the covariance terms are parametrized using only a single parameter $\sigma_{\varkappa}$ that reflects the variance of the observation errors in option prices.

The measurement equation for the filtering problem is given in terms of the log CCF. Therefore, by applying a Taylor-series expansion to the difference of the logs and using the error decomposition of the CCF approximation (in particular, (A.7)), we have

$$
\begin{aligned}
\xi_t(u, \tau) := \log \widehat{\phi}_t(u, \tau) - \log \phi_t(u, \tau) &= \log\left(1 + \frac{\zeta_t^{(1)}(u, \tau) + \zeta_t^{(2)}(u, \tau) + \zeta_t^{(3)}(u, \tau)}{\phi_t(u, \tau)}\right) \\
&= \xi_t^{(1)}(u, \tau) + r_t(u, \tau),
\end{aligned}
$$

44

where

$$\xi_t^{(1)}(u,\tau) := \frac{\zeta_t^{(1)}(u,\tau)}{\phi_t(u,\tau)} = \mathcal{O}_p\left(\sqrt{\frac{\log n}{n}}\right), \quad \text{and} \quad r_t(u,\tau) = \mathcal{O}_p\left(n^{-(\underline{\delta}\,\underline{\alpha}\wedge(1+\bar{\delta})\bar{\alpha})} \vee \frac{\log n}{n}\right).$$

We note that the remainder term collects the log-linearization of the truncation and discretization errors and higher-order terms from a Taylor-series expansion.

After stacking each component of the measurement equation (12) as well as the observation errors $\xi_t^{(1)}(u,\tau)$ and remainder term $r_t(u,\tau)$ along arguments, real and imaginary parts, and maturity, we obtain the state space measurement equation (19) in Proposition 1.

To derive the covariance matrix of the measurement errors, we first consider the covariance and pseudo-covariance matrices of the stacked vector $\xi_{t,\tau}^{(1)} = \left(\xi_t^{(1)}(u_1,\tau),\ldots,\xi_t^{(1)}(u_q,\tau)\right)'$ for a fixed time $t$ and time-to-maturity $\tau$. They are given by

$$\Gamma_{t,\tau} := \mathbb{E}\left[\xi_{t,\tau}^{(1)}\overline{\xi_{t,\tau}^{(1)}}'\right] = \left[\mathbb{E}[\xi_t^{(1)}(u_i,\tau)\xi_t^{(1)}(-u_j,\tau)]\right]_{1\leq i,j\leq q}$$

$$= \sigma_\varkappa^2 \cdot \left[\frac{\gamma_t(u_i,u_j,\tau)}{\phi_t(u_i,\tau)\phi_t(-u_j,\tau)}\right]_{1\leq i,j\leq q} =: \sigma_\varkappa^2 \cdot \widetilde{\Gamma}_{t,\tau},$$

$$C_{t,\tau} := \mathbb{E}\left[\xi_{t,\tau}^{(1)}\xi_{t,\tau}^{(1)}{}'\right] = \left[\mathbb{E}[\xi_t^{(1)}(u_i,\tau)\xi_t^{(1)}(u_j,\tau)]\right]_{1\leq i,j\leq q}$$

$$= \sigma_\varkappa^2 \cdot \left[\frac{c_t(u_i,u_j,\tau)}{\phi_t(u_i,\tau)\phi_t(u_j,\tau)}\right]_{1\leq i,j\leq q} =: \sigma_\varkappa^2 \cdot \widetilde{C}_{t,\tau}.$$

Next, since $\xi_{t,\tau}^{(1)}$ is a complex-valued random vector, the covariance matrix of the stacked real and imaginary parts of $\xi_{t,\tau}^{(1)}$ is of the following form:

$$H_{t,\tau} := \mathrm{Var}\left[\begin{pmatrix}\Re(\xi_{t,\tau}^{(1)})\\ \Im(\xi_{t,\tau}^{(1)})\end{pmatrix}\right] = \begin{pmatrix}\frac{1}{2}\Re(\Gamma_{t,\tau}+C_{t,\tau}) & \frac{1}{2}\Im(-\Gamma_{t,\tau}+C_{t,\tau})\\ \frac{1}{2}\Im(\Gamma_{t,\tau}+C_{t,\tau}) & \frac{1}{2}\Re(\Gamma_{t,\tau}-C_{t,\tau})\end{pmatrix}$$

$$= \sigma_\varkappa^2 \cdot \begin{pmatrix}\frac{1}{2}\Re(\widetilde{\Gamma}_{t,\tau}+\widetilde{C}_{t,\tau}) & \frac{1}{2}\Im(-\widetilde{\Gamma}_{t,\tau}+\widetilde{C}_{t,\tau})\\ \frac{1}{2}\Im(\widetilde{\Gamma}_{t,\tau}+\widetilde{C}_{t,\tau}) & \frac{1}{2}\Re(\widetilde{\Gamma}_{t,\tau}-\widetilde{C}_{t,\tau})\end{pmatrix}$$

$$=: \sigma_\varkappa^2 \cdot \widetilde{H}_{t,\tau}.$$

This establishes (21).

Given Assumption 2, the error terms $\zeta_{t,\tau}^{(1)}$ and $\xi_{t,\tau}^{(1)}$ are conditionally independent along maturity and time. This implies that the measurement errors $\varepsilon_t$ stacked along maturities are also conditionally independent, thus $\mathbb{E}[\varepsilon_t\varepsilon_s'] = 0$ for $s \neq t$, and their covariance matrix has a block-diagonal form: $H_t = \mathrm{blkdiag}\{H_{t,1},\ldots,H_{t,k}\}$.

The disturbance term in the state updating equation is given by $\eta_{t+1} = x_{t+1} - \mathbb{E}[x_{t+1}|\mathcal{F}_t]$. Therefore, $\eta_t$ constitutes a martingale difference sequence, thus $\mathbb{E}[\eta_t\eta_s'] = 0$ for $s \neq t = 1,\ldots,T$.

Since the measurement errors $\varepsilon_t$ have zero mean conditional on the filtration $\mathcal{F}_t$, we also have that $\mathbb{E}[\varepsilon_t x_t] = 0$. Given that the state process is stationary and the initial condition is the unconditional mean, $\mathbb{E}[\varepsilon_t x_1'] = 0$ and $\mathbb{E}[\eta_{t+1}x_1'] = 0$. This implies that $\mathbb{E}[\varepsilon_t\eta_s'] = 0$ for all $s,t = 1,\ldots,T$. Thus, the proof is established. $\square$

## A.3 Proof of Proposition 2

### Setup and notation

Given a parameter value $\vartheta = (\theta', \sigma_\varkappa^2)'$, the true state $x_t$ can be estimated by the filtered and predicted states as $\widehat{x}_{t|t}(\vartheta)$ and $\widehat{x}_{t|t-1}(\vartheta)$, and by the recovered state $x_t(\theta) = y_t^*(\theta) - d_t^*(\theta) = A_t^*(\theta)(y_t - d_t(\theta))$. As discussed in the main text, $x_t(\theta)$ converges, as $n \to \infty$, to the true state when evaluated at the true parameter vector, i.e., $\text{plim}_{n\to\infty} x_t(\theta_0) = x_t$.

Since the parameter estimates are obtained by maximizing the (quasi) log-likelihood function of the collapsed state space system, we first analyze the limiting behavior of the collapsing transformation. As a straightforward corollary of Proposition 1, the covariance matrix of the observation errors $H_t = \mathcal{O}_p\left(n^{-1}\log n\right)$, i.e., it converges to zero as $n \to \infty$. However, the scaled version of $H_t$ has a non-zero limit. In particular, for the components of the covariance matrix we have for $k, l = 1, \ldots, q$:

$$(\Delta m)^{-1}(\widetilde{\Gamma}_{t,i})_{kl} = \frac{u_{k,t}\overline{u_{l,t}}}{\phi_t(u_k, \tau_i)\phi_t(-u_l, \tau_i)} \sum_{j=2}^{n} e^{(\mathrm{i}(u_k - u_l) - 2)m_j} \kappa_t^2(\tau_i, m_j)\nu_t^2(\tau_i, m_j)\frac{(\Delta m_j)^2}{\Delta m}$$

$$\to \frac{u_{k,t}\overline{u_{l,t}}}{\phi_t(u_k, \tau_i)\phi_t(-u_l, \tau_i)} \int_{\mathbb{R}} e^{(\mathrm{i}(u_k - u_l) - 2)m} \kappa_t^2(\tau_i, m)\nu_t^2(\tau_i, m)\mathrm{d}m =: (\widetilde{\mathcal{G}}_{t,i})_{kl}$$

and

$$(\Delta m)^{-1}(\widetilde{C}_{t,i})_{kl} \to \frac{u_{k,t}u_{l,t}}{\phi_t(u_k, \tau_i)\phi_t(u_l, \tau_i)} \int_{\mathbb{R}} e^{(\mathrm{i}(u_k + u_l) - 2)m} \kappa_t^2(\tau_i, m)\nu_t^2(\tau_i, m)\mathrm{d}m =: (\widetilde{C}_{t,i})_{kl}.$$

Therefore, the limit of the scaled covariance matrix $(\Delta m)^{-1}\widetilde{H}_t$ is $\widetilde{\mathcal{H}}_t = \text{blkdiag}\{\widetilde{\mathcal{H}}_{t,1}, \ldots, \widetilde{\mathcal{H}}_{t,k}\}$ with

$$\widetilde{\mathcal{H}}_{t,i} = \begin{pmatrix} \frac{1}{2}\Re(\widetilde{\mathcal{G}}_{t,i} + \widetilde{\mathcal{C}}_{t,i}) & \frac{1}{2}\Im(-\widetilde{\mathcal{G}}_{t,i} + \widetilde{\mathcal{C}}_{t,i}) \\ \frac{1}{2}\Im(\widetilde{\mathcal{G}}_{t,i} + \widetilde{\mathcal{C}}_{t,i}) & \frac{1}{2}\Re(\widetilde{\mathcal{G}}_{t,i} - \widetilde{\mathcal{C}}_{t,i}) \end{pmatrix}.$$

Hence, for the projection matrix we have

$$A_t^* = (Z_t'H_t^- Z_t)^{-1}Z_t'H_t^- = (Z_t'\widetilde{H}_t^- Z_t)^{-1}Z_t'\widetilde{H}_t^-$$

$$= (Z_t'\Delta m\widetilde{H}_t^- Z_t)^{-1}Z_t'\Delta m\widetilde{H}_t^-$$

$$\xrightarrow[n\to\infty]{p} (Z_t'\widetilde{\mathcal{H}}_t^- Z_t)^{-1}Z_t'\widetilde{\mathcal{H}}_t^- =: \mathcal{A}_t^*.$$

Recall that the components of the quasi-likelihood contributions for the collapsed state space representation, see (27), are given by $l_t^*(\theta, \sigma_\varkappa^2)$, $l_t^+(\theta_1, \sigma_\varkappa^2)$, and $J_t(\theta_1)$, and the average log-likelihoods by $\bar{l}_T(\vartheta) = l(Y_t; \vartheta)/T$ and analogously $\bar{l}_T^*(\theta, \sigma_\varkappa^2)$, $\bar{l}_T^+(\theta_1, \sigma_\varkappa^2)$, $\bar{J}_T(\theta_1)$, $\bar{\ell}_T(\vartheta)$ and $\bar{\ell}_T^*(\theta)$.

In what follows, we will denote with $R_n$ a (vector or a matrix consisting of) random variable(s), (each elements of) which is $\mathcal{O}_p\left(n^{-1}\log n\right)$ and might change from line to line. Recall that $\Delta m = R_n$.

### Convergence to approximating likelihood

Under the assumptions of Proposition 1, and boundedness of the first- and second-order derivatives of the system elements $\widetilde{H}_t^*(\theta)$, $T_t(\theta)$, $c_t(\theta)$, $Q_t(x_t, \theta)$ and $Z_t(\theta)$ with respect to the parameter vector $\theta \in \Theta$, we will show that as $n \to \infty$, pointwise in $\vartheta$, the difference between $l_t(\vartheta)$ and

$\ell_t(\theta)$, and between their first- and second-order partial derivatives with respect to $\theta$, converges to zero at the rate of $R_n$.

Recall that $H_t^*(\vartheta) = \sigma_\varkappa^2 A_t^*(\theta)\widetilde{H}_t A_t^{*\prime}(\theta)$. Since $\widetilde{H}_t = R_n$ and $A_t^*(\theta) \to_p \mathcal{A}_t^*(\theta)$ as $n \to \infty$, we have that $H_t^*(\vartheta) = R_n$. Since $Z_t^* = I_d$ for the collapsed state space system, $G_t^*(\vartheta) = P_{t|t-1}(\vartheta) + H_t^*(\vartheta)$, and, hence, $G_t^*(\vartheta) - P_{t|t-1}(\vartheta) = R_n$. Using the Woodbury matrix identity, we find:

$$
\begin{aligned}
G_t^*(\vartheta)^{-1} &= (P_{t|t-1}(\vartheta) + H_t^*(\vartheta))^{-1} \\
&= P_{t|t-1}(\vartheta)^{-1} - P_{t|t-1}(\vartheta)^{-1}H_t^*(\vartheta)G_t^*(\vartheta)^{-1} \\
&= P_{t|t-1}(\vartheta)^{-1} + R_n.
\end{aligned}
$$

Using the Kalman filter recursions, it also follows that

$$
\begin{aligned}
P_{t|t}(\vartheta) &= P_{t|t-1}(\vartheta) - P_{t|t-1}(\vartheta)G_t^*(\vartheta)^{-1}P_{t|t-1}(\vartheta) \\
&= H_t^*(\vartheta)G_t^*(\vartheta)^{-1}P_{t|t-1}(\vartheta) \\
&= R_n,
\end{aligned}
$$

and

$$
\begin{aligned}
\widehat{x}_{t|t}(\vartheta) &= \widehat{x}_{t|t-1}(\vartheta) + P_{t|t-1}(\vartheta)G_t^*(\vartheta)^{-1}\omega_t^*(\vartheta) \\
&= \widehat{x}_{t|t-1}(\vartheta) + \omega_t^*(\vartheta) + H_t^*(\vartheta)G_t^*(\vartheta)^{-1}\omega_t^*(\vartheta) \\
&= y_t^*(\theta) - d_t^*(\theta) + H_t^*(\vartheta)G_t^*(\vartheta)^{-1}\omega_t^*(\vartheta) \\
&= x_t(\theta) + H_t^*(\vartheta)G_t^*(\vartheta)^{-1}\omega_t^*(\vartheta) \\
&= x_t(\theta) + R_n.
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
\omega_t^*(\vartheta) &= y_t^*(\theta) - d_t^*(\theta) - \widehat{x}_{t|t-1}(\vartheta) \\
&= x_t(\theta) - \widehat{x}_{t|t-1}(\vartheta) \\
&= T_{t-1}(\theta)(x_{t-1}(\theta) - \widehat{x}_{t-1|t-1}(\vartheta)) + \eta_t(\theta) \\
&= \eta_t(\theta) + R_n,
\end{aligned}
$$

and

$$
\begin{aligned}
G_t^*(\vartheta) &= P_{t|t-1}(\vartheta) + H_t^*(\vartheta) \\
&= T_{t-1}(\theta)P_{t-1|t-1}(\vartheta)T_{t-1}(\theta)' + Q_{t-1}(\widehat{x}_{t-1|t-1}(\vartheta);\theta) + H_t^*(\vartheta) \\
&= Q_{t-1}(\widehat{x}_{t-1|t-1}(\vartheta);\theta) + R_n \\
&= Q_{t-1}(x_{t-1}(\theta);\theta) + R_n,
\end{aligned}
$$

where the last equality follows since $Q_t(x_t)$ is an affine function in $x_t$. From this it also follows that $G_t^*(\vartheta)^{-1} = Q_{t-1}(x_{t-1}(\theta))^{-1} + R_n$.

Combining all of these results, we obtain

$$2|l_t^*(\vartheta) - \ell_t^*(\theta)| = \left|-\log\left(\det(G_t^*(\vartheta))\det(Q_{t-1}(x_{t-1}(\theta);\theta))^{-1}\right)\right.$$
$$\left. -\omega_t^{*\prime}(\vartheta)G_t^*(\vartheta)^{-1}\omega_t^*(\vartheta) + \eta_t(\theta)'Q_{t-1}^{-1}(x_{t-1}(\theta);\theta)\eta_t(\theta)\right|$$
$$\leq \left|\log\det\left(G_t^*(\vartheta)Q_{t-1}^{-1}(x_{t-1}(\theta);\theta)\right)\right| + R_n$$
$$= |\log\det(I_d + R_n)| + R_n.$$

For the order of the first term we can further use that

$$\det(I_d + R_n) = 1 + \det(R_n) + o_p(1).$$

Hence, the order of $\log\det(I + R_n)$ is given by the order of $\det(R_n)$. The latter converges to zero faster than $n^{-1}\log n$ since it is the product of eigenvalues of matrix $R_n$. Therefore, we have that $|l_t^*(\vartheta) - \ell_t^*(\theta)| = R_n \to_p 0$ for all $\vartheta \in \Theta \times \Sigma_\varkappa$.

Consider now the derivatives of the outputs of the Kalman filter applied to the collapsed state space system. Using the expressions derived above, we have the following useful results for the first-order derivatives with respect to the $i^{th}$ component of the parameter vector:

$$\frac{\partial H_t^*(\vartheta)}{\partial\theta_i} = \frac{\partial A_t^*(\vartheta)H_t A_t^{*\prime}(\vartheta)}{\partial\theta_i} = 2A_t^*(\vartheta)H_t\frac{\partial A_t^{*\prime}(\vartheta)}{\partial\theta_i} = R_n,$$
$$\frac{\partial G_t^*(\vartheta)^{-1}}{\partial\theta_i} = \frac{\partial P_{t|t-1}(\vartheta)^{-1}}{\partial\theta_i} + R_n = \frac{\partial Q_{t-1}(x_{t-1}(\theta);\theta)^{-1}}{\partial\theta_i} + R_n,$$
$$\frac{\partial P_{t|t}(\vartheta)}{\partial\theta_i} = R_n,$$
$$\frac{\partial\widehat{x}_{t|t}(\vartheta)}{\partial\theta_i} = \frac{\partial x_t(\theta)}{\partial\theta_i} + R_n,$$
$$\frac{\partial\omega_t^*(\vartheta)}{\partial\theta_i} = \frac{\partial\eta_t(\theta)}{\partial\theta_i} + R_n,$$
$$\frac{\partial G_t^*(\vartheta)}{\partial\theta_i} = \frac{\partial Q_{t-1}(x_{t-1};\theta)}{\partial\theta_i} + R_n.$$

Therefore, the derivatives of the likelihood contributions of the collapsed state space system can be written as

$$\frac{\partial l_t^*(\vartheta)}{\partial\theta_i} = -\frac{1}{2}\text{trace}\left(G_t^*(\vartheta)^{-1}\frac{\partial G_t^*(\vartheta)}{\partial\theta_i}\right) - \omega_t^*(\vartheta)'G_t^{*-1}(\vartheta)\frac{\partial\omega_t^*(\vartheta)}{\partial\theta_i} - \frac{1}{2}\omega_t^*(\vartheta)'\frac{\partial G_t^{*-1}(\vartheta)}{\partial\theta_i}\omega_t^*(\vartheta)$$
$$= -\frac{1}{2}\text{trace}\left((Q_{t-1}(x_{t-1}(\theta);\theta)^{-1} + R_n)\left(\frac{\partial Q_{t-1}(x_{t-1}(\theta);\theta)}{\partial\theta_i} + R_n\right)\right)$$
$$- (\eta_t(\theta) + R_n)'\left(Q_{t-1}(x_{t-1}(\theta);\theta)^{-1} + R_n\right)\left(\frac{\partial\eta_t(\theta)}{\partial\theta_i} + R_n\right)$$
$$- \frac{1}{2}(\eta_t(\theta) + R_n)'\left(\frac{\partial Q_{t-1}(x_{t-1}(\theta);\theta)^{-1}}{\partial\theta_i} + R_n\right)(\eta_t(\theta) + R_n)$$
$$= \frac{\partial\ell_t^*(\theta)}{\partial\theta_i} + R_n.$$

Finally, we have similar results for the second-order derivatives:

$$\frac{\partial^2 H_t^*(\vartheta)}{\partial\theta_i\partial\theta_j} = \frac{\partial^2 A_t^*(\vartheta)H_t A_t^{*\prime}(\vartheta)}{\partial\theta_i\partial\theta_j} = R_n,$$

$$\frac{\partial^2 G_t^*(\vartheta)^{-1}}{\partial\theta_i\partial\theta_j} = \frac{\partial^2 P_{t|t-1}(\vartheta)^{-1}}{\partial\theta_i\partial\theta_j} + R_n = \frac{\partial^2 Q_{t-1}(x_{t-1}(\theta);\theta)^{-1}}{\partial\theta_i\partial\theta_j} + R_n,$$

$$\frac{\partial^2 P_{t|t}(\vartheta)}{\partial\theta_i\partial\theta_j} = R_n,$$

$$\frac{\partial^2 \widehat{x}_{t|t}(\vartheta)}{\partial\theta_i\partial\theta_j} = \frac{\partial^2 x_t(\theta)}{\partial\theta_i\partial\theta_j} + R_n,$$

$$\frac{\partial^2 \omega_t^*(\vartheta)}{\partial\theta_i\partial\theta_j} = \frac{\partial^2 \eta_t(\theta)}{\partial\theta_i\partial\theta_j} + R_n,$$

$$\frac{\partial^2 G_t^*(\vartheta)}{\partial\theta_i\partial\theta_j} = \frac{\partial^2 Q_{t-1}(x_{t-1};\theta)}{\partial\theta_i\partial\theta_j} + R_n,$$

which imply the pointwise convergence of the second-order derivative:

$$\frac{\partial^2 l_t^*(\vartheta)}{\partial\theta\partial\theta'} - \frac{\partial^2 \ell_t^*(\vartheta)}{\partial\theta\partial\theta'} = R_n.$$

## Consistency

Together with the stochastic equicontinuity condition of Assumption 4 $(i)$, the above pointwise convergence results imply, as $n \to \infty$,

$$\sup_{\vartheta\in\Theta\times\Sigma_\varkappa} |\bar{l}_T(\vartheta) - \bar{\ell}_T(\vartheta)| \xrightarrow{p} 0.$$

This implies that for analyzing consistency of $\widehat{\vartheta}$, as $n \to \infty$ followed by $T \to \infty$, we may analyze the approximating average log-likelihood $\bar{\ell}_T(\vartheta) = \bar{\ell}_T^*(\theta) + \bar{l}_T^+(\theta_1, \sigma_\varkappa^2) + \bar{J}_T(\theta_1)$.

Using $e_t(\theta_1) = M_{Z_t}(\theta_1)\varepsilon_t + M_{Z_t}(\theta_1)[d_t(\theta_{1,0}) - d_t(\theta_1) + Z_t(\theta_{1,0})x_t]$, it follows that $\bar{l}_T^+(\theta_1, \sigma_\varkappa^2)$ will diverge to $-\infty$ for $\theta_1 \neq \theta_{1,0}$. This is a reflection of the different convergence rates of $\widehat{\theta}_1$ and $\widehat{\theta}_2$, and implies that we cannot prove consistency via uniform convergence (as $n, T \to \infty$) of $\bar{\ell}_T(\vartheta)$ to a continuous function with unique maximum at $\vartheta_0$. A similar phenomenon occurs in the literature on regressions with unit-root processes, see Saikkonen (1995). On the other hand, based on Assumption 4, $(ii)$ directly implies consistency of $\widehat{\vartheta}$.

## Asymptotic normality

To show the asymptotic normality of $\widehat{\theta}$, we can adapt the usual mean value theorem:

$$0 = \nabla\bar{l}_T(\widehat{\theta}, \widehat{\sigma}_\varkappa^2) = \nabla\bar{l}_T(\theta_0, \widehat{\sigma}_\varkappa^2) + \nabla^2\bar{l}_T(\bar{\theta}, \widehat{\sigma}_\varkappa^2)(\widehat{\theta} - \theta_0),$$

where $\bar{\theta}$ lies between $\widehat{\theta}$ and $\theta_0$. The first-order derivative of the log-likelihood contribution is $\nabla l_t(\vartheta) = (\nabla_1 l_t(\vartheta)', \nabla_2 l_t(\vartheta)')'$ with

$$\nabla_1 l_t(\vartheta) = \nabla_1 l_t^*(\theta, \sigma_\varkappa^2) + \nabla_1 l_t^+(\theta_1, \sigma_\varkappa^2) + \nabla_1 J_t(\theta_1),$$

$$\nabla_2 l_t(\vartheta) = \nabla_2 l_t^*(\theta, \sigma_\varkappa^2).$$

Note, in particular, that

$$\nabla_1 l_t^+(\theta_1, \sigma_\varkappa^2) = \frac{\partial l_t^+(\theta_1, \sigma_\varkappa^2)}{\partial \theta_1} = -\frac{1}{2\sigma_\varkappa^2} \frac{\partial e_t(\theta_1)' \widetilde{H}_t^- e_t(\theta_1)}{\partial \theta_1} = -\frac{1}{\sigma_\varkappa^2} \frac{\partial e_t(\theta_1)'}{\partial \theta_1} \widetilde{H}_t^- e_t(\theta_1).$$

When evaluated at $\theta_{1,0}$, $e_t(\theta_{1,0}) = M_{Z_t}(\theta_{1,0})\varepsilon_t = R_n^{1/2}$, while $\widetilde{H}_t = R_n$ and $\frac{\partial e_t(\theta_1)}{\partial \theta_1'}\big|_{\theta_{1,0}} = \mathcal{O}_p(1)$. Hence, the first-order derivative is unbounded and $R_n^{-1/2}$. Therefore, we need an appropriate standardization for the score vector by the factor $\sqrt{\Delta m} = R_n^{1/2}$ to make its probability limit bounded. However, scaled by the same factor, the first and the third terms of the first-order derivative $\nabla_1 l_t(\vartheta)$ become of a small $R_n^{1/2}$ order since $\nabla_1 l_t^*(\vartheta) = \mathcal{O}_p(1)$ and $\nabla_1 J_t(\theta_1) = \mathcal{O}_p(1)$. Hence, given the consistency result (including consistency of $\widehat{\sigma}_\varkappa^2$) as well as Assumption 4 $(iii)$, the central limit theorem yields for the first component

$$\sqrt{\Delta m}\sqrt{T} \; \nabla_1 \bar{l}_T(\theta_0, \widehat{\sigma}_\varkappa^2) = \sqrt{\Delta m}\sqrt{T} \; \left(\nabla_1 \bar{l}_T^*(\theta_0, \widehat{\sigma}_\varkappa^2) + \nabla_1 \bar{l}_T^+(\theta_{1,0}, \widehat{\sigma}_\varkappa^2) + \nabla_1 \bar{J}_T(\theta_{1,0})\right)$$
$$= \sqrt{\Delta m}\sqrt{T} \; \nabla_1 \bar{l}_T^+(\theta_{1,0}, \sigma_{\varkappa,0}^2) + o_p(1)$$
$$\xrightarrow{d} N(0, \mathcal{I}_1),$$

where $\mathcal{I}_1 := \operatorname{plim}_{n,T\to\infty} T^{-1}\Delta m \sum_{t=1}^T \nabla_1 l_t^+(\vartheta_{1,0})\nabla_1 l_t^+(\vartheta_{1,0})'$. Given the consistency of the parameter estimates and Assumption 4 $(iv)$, the feasible covariance matrix estimator can be obtained as $\widehat{\mathcal{I}}_1 = T^{-1}\Delta m \sum_{t=1}^T \nabla_1 l_t(\widehat{\vartheta})\nabla_1 l_t(\widehat{\vartheta})'$.

On the other hand, we have established $\nabla_2 l_t^*(\vartheta) = \nabla_2 \ell_t^*(\theta) + R_n$. Therefore, under the fairly standard assumptions on the likelihood of the heteroskedastic (V)AR(1) process, as summarized in Assumption 4 $(iii)$–$(iv)$, we find

$$\sqrt{T} \; \nabla_2 \bar{l}_T(\theta_0, \widehat{\sigma}_\varkappa^2) = \sqrt{T} \; \nabla_2 \bar{l}_T^*(\theta_0, \widehat{\sigma}_\varkappa^2)$$
$$= \sqrt{T} \; \nabla_2 \bar{\ell}_T^*(\theta_0) + o_p(1)$$
$$\xrightarrow{d} N(0, \mathcal{I}_2),$$

given that $\Delta m\sqrt{T} \to 0$, such that the approximation errors $R_n$ vanish from the limit distribution. Here $\mathcal{I}_2 = \operatorname{plim}_{n,T\to\infty} T^{-1}\sum_{t=1}^T \nabla_2 \ell_t^*(\theta_0)\nabla_2 \ell_t^*(\theta_0)'$. Given the consistency result and Assumption 4 $(iv)$, the covariance matrix is estimated consistently as $\widehat{\mathcal{I}}_2 := \frac{1}{T}\sum_{t=1}^T \nabla_2 l_t(\widehat{\vartheta})\nabla_2 l_t(\widehat{\vartheta})$. Thus, combining these two results, the standardized score vector $TD_{nT}^{-1}\nabla \bar{l}_T(\theta)$ with $D_{nT} = \operatorname{blkdiag}\{\sqrt{T/(\Delta m)}I_{\dim\theta_1}, \sqrt{T}I_{\dim\theta_2}\}$ converges weakly to a multivariate normal with the block-diagonal covariance matrix $\mathcal{I} = \operatorname{blkdiag}\{\mathcal{I}_1, \mathcal{I}_2\}$.

As for the second-order derivatives, we have

$$\nabla_2^2 l_t(\theta, \sigma_\varkappa^2) = \nabla_2^2 l_t^*(\theta, \sigma_\varkappa^2) = \nabla_2^2 \ell_t^*(\theta) + o_p(1),$$
$$\nabla_1 \nabla_2 l_t(\theta, \sigma_\varkappa^2) = \nabla_1 \nabla_2 l_t^*(\theta, \sigma_\varkappa^2) = \nabla_1 \nabla_2 \ell_t^*(\theta) + o_p(1),$$
$$\nabla_1^2 l_t(\theta, \sigma_\varkappa^2) = \nabla_1^2 l_t^*(\theta, \sigma_\varkappa^2) + \nabla_1^2 l_t^+(\theta_1, \sigma_\varkappa^2) + \nabla_1^2 J_t^+(\theta_1),$$

with

$$\nabla_1^2 l_t^+(\theta_1, \sigma_\varkappa^2) = -\frac{1}{2\sigma_\varkappa^2} \frac{\partial^2 e_t(\theta_1)' \widetilde{H}_t^- e_t(\theta_1)}{\partial \theta_1 \partial \theta_1'}$$
$$= -\frac{1}{\sigma_\varkappa^2} \left(\frac{\partial e_t(\theta_1)'}{\partial \theta_1} \widetilde{H}_t^- \frac{\partial e_t(\theta_1)}{\partial \theta_1'} + (e_t(\theta_1)' \widetilde{H}_t^- \otimes I_p)\frac{\partial \operatorname{vec}(\partial e_t(\theta_1)'/\partial \theta_1)}{\partial \theta_1'}\right).$$

Since $\partial e_t(\theta_1)/\partial\theta_1' = \mathcal{O}_p(1)$, the above term $\nabla_1^2 l_t^+(\theta_1, \sigma_\varkappa^2) = R_n^{-1}$, while, due to the approximation results given above, $\nabla_1^2 l_t^*(\theta, \sigma_\varkappa^2) = \nabla_1^2 \ell_t^*(\theta, \sigma_\varkappa^2) + o_p(1) = \mathcal{O}_p(1)$ and $\nabla_1^2 J_t^+(\theta_1) = \mathcal{O}_p(1)$. Hence, $\Delta m \nabla_1^2 l_t(\theta, \sigma_\varkappa^2) = \Delta m \nabla_1^2 l_t^+(\theta_1, \sigma_\varkappa^2) + o_p(1)$. Combining results, we have shown that the standardized Hessian $TD_{nT}^{-1}\nabla^2\bar{l}_T(\theta, \sigma_\varkappa)D_{nT}^{-1}$ converges to the block diagonal matrix. In fact, using Assumption 4, we have, componentwise,

$$\Delta m \ \nabla_1^2 \bar{l}_T(\bar{\theta}, \hat{\sigma}_\varkappa^2) = \Delta m \ \nabla_1^2 \bar{l}_T^+(\bar{\theta}, \sigma_{\varkappa,0}^2) + o_p(1) \xrightarrow{p} \mathcal{S}_1,$$
$$\nabla_2^2 \bar{l}_T(\bar{\theta}, \hat{\sigma}_\varkappa^2) = \nabla_2^2 \bar{l}_T^*(\bar{\theta}) + o_p(1) \xrightarrow{p} \mathcal{S}_2,$$
$$\sqrt{\Delta m}\nabla_1\nabla_2 \bar{l}_T(\bar{\theta}, \hat{\sigma}_\varkappa^2) = \sqrt{\Delta m}\nabla_1\nabla_2 \ell_T^*(\theta) + o_p(1) \xrightarrow{p} 0.$$

Therefore,

$$D_{nT}(\hat{\theta} - \theta_0) = \left(TD_{nT}^{-1}\nabla^2\bar{l}_T(\bar{\theta}, \hat{\sigma}_\varkappa^2)D_{nT}^{-1}\right)^{-1} TD_{nT}^{-1}\nabla\bar{l}_T(\theta_0, \hat{\sigma}_\varkappa^2)$$
$$\xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathcal{S}_1^{-1}\mathcal{I}_1\mathcal{S}_1^{-1} & 0 \\ 0 & \mathcal{S}_2^{-1}\mathcal{I}_2\mathcal{S}_2^{-1} \end{pmatrix}\right).$$

This proves the proposition. $\qquad\square$

## Appendix B  Conditional Moments

In this appendix, we describe how the conditional mean and variance can be computed for the AJD class of models. In particular, we derive closed-form expressions for the conditional mean and variance in the univariate case, and briefly discuss how these moments can be obtained numerically in the multivariate setting at low computational costs. While semi-closed-form expressions are also available in the multivariate setting, they are more cumbersome to work with in practice since they typically require matrix exponentials and integrations.

We start with considering the univariate version of the AJD process in (8), using shorthand notation as follows:

$$\mathrm{d}x_t = \mu(x_t)\mathrm{d}t + \sigma(x_t)\mathrm{d}W_t + J_t\mathrm{d}N_t, \tag{B.1}$$

with $\mu(x) = k_0 + k_1 x$, $\sigma^2(x) = h_0 + h_1 x$, $\lambda(x) = l_0 + l_1 x$, where all coefficients are real-valued numbers and the standard Brownian motion $W_t$ and the counting process $N_t$ are univariate processes. The jump size distribution $\nu$ on $\mathbb{R}$ is independent of time and of any form of randomness in the model. We further assume that the SDE (B.1) has a unique strong solution and the first two moments are well defined. For more details, see Section 2.2 and Duffie et al. (2000). For notational simplicity, let

$$\mu_J := \mathbb{E}[J], \quad \mu_{J2} := \mathbb{E}[J^2], \quad g_0 := k_0 + l_0\mu_J, \quad g_1 := k_1 + l_1\mu_J.$$

The associated infinitesimal generator $\mathcal{D}$, defined at a bounded $C^2$ function $f\colon D \to \mathbb{R}$, with bounded first and second derivatives $f_x$ and $f_{xx}$, is given by

$$\mathcal{D}f(x) = f_x(x)\mu(x) + \frac{1}{2}f_{xx}(x)\sigma(x)^2 + \lambda(x)\int_{\mathbb{R}}[f(x+z) - f(x)]\mathrm{d}\nu(z).$$

Dynkin's formula yields that

$$\mathbb{E}[f(x_T)|\mathcal{F}_t] = f(x_t) + \mathbb{E}\left[\int_t^T \mathcal{D}f(x_s)\mathrm{d}s\Big|\mathcal{F}_t\right].$$

Therefore, we can find the conditional moments by applying Dynkin's formula for $f(x) = x$:

$$\begin{aligned}
\mathbb{E}[x_T|\mathcal{F}_t] &= x_t + \mathbb{E}\left[\int_t^T \left(\mu(x_s) + \lambda(x_s)\int_{\mathbb{R}} z\mathrm{d}\nu(z)\right)\mathrm{d}s\Big|\mathcal{F}_t\right] \\
&= x_t + \mathbb{E}\left[\int_t^T (k_0 + k_1 x_s + (l_0 + l_1 x_s)\mu_J)\,\mathrm{d}s\Big|\mathcal{F}_t\right] \\
&= x_t + \int_t^T \left(k_0 + l_0\mu_J + (k_1 + l_1\mu_J)\mathbb{E}[x_s|\mathcal{F}_t]\right)\mathrm{d}s \\
&= x_t + \int_t^T \left(g_0 + g_1\mathbb{E}[x_s|\mathcal{F}_t]\right)\mathrm{d}s,
\end{aligned}$$

where Fubini's theorem is used in the third line. Hence, we can obtain the first conditional moment by solving the following ODE:

$$\mathrm{d}\mathbb{E}[x_s|\mathcal{F}_t] = \left(g_0 + g_1\mathbb{E}[x_s|\mathcal{F}_t]\right)\mathrm{d}s,$$

with initial condition $\mathbb{E}[x_t|\mathcal{F}_t] = x_t$. Thus, the conditional expectation is given by

$$m_t(T) := \mathbb{E}[x_T|\mathcal{F}_t] = e^{g_1(T-t)}x_t + \frac{g_0}{g_1}\left(e^{g_1(T-t)} - 1\right). \tag{B.2}$$

Next, we are interested in deriving the conditional variance:

$$\mathrm{Var}(x_T|\mathcal{F}_t) = \mathbb{E}[(x_T - \mathbb{E}[x_T|\mathcal{F}_t])^2|\mathcal{F}_t].$$

Note that

$$x_T - \mathbb{E}[x_T|\mathcal{F}_t] = \mathbb{E}[x_T|\mathcal{F}_T] - \mathbb{E}[x_T|\mathcal{F}_t] = \int_t^T \mathrm{d}\mathbb{E}[x_T|\mathcal{F}_s] = \int_t^T \mathrm{d}m_s(T).$$

The dynamics of the conditional mean for fixed $T > t$ can be obtained by using Itô's lemma:

$$\begin{aligned}
\mathrm{d}m_t(T) &= \left[-g_1 e^{g_1(T-t)}x_t - g_0 e^{g_1(T-t)}\right]\mathrm{d}t + e^{g_1(T-t)}(\mu(x_t)\mathrm{d}t + \sigma(x_t)\mathrm{d}W_t) + e^{g_1(T-t)}J_t\mathrm{d}N_t \\
&= e^{g_1(T-t)}\left[-(l_0 + l_1 x_t)\mu_J\mathrm{d}t + \sigma(x_t)\mathrm{d}W_t + J_t\mathrm{d}N_t\right].
\end{aligned}$$

Note that the process $m_t(T)$ for fixed $T$ is a local martingale. Thus, we can use the Itô isometry to obtain the conditional variance:

$$\begin{aligned}
\mathrm{Var}(x_T|\mathcal{F}_t) &= \mathbb{E}\left[\left(\int_t^T \mathrm{d}m_s(T)\right)^2\Big|\mathcal{F}_t\right] \\
&= \mathbb{E}\left[\int_t^T e^{2g_1(T-s)}\sigma^2(x_s)\mathrm{d}s\Big|\mathcal{F}_t\right] + \mu_{J2}\cdot\mathbb{E}\left[\int_t^T e^{2g_1(T-s)}\lambda(x_s)\mathrm{d}s\Big|\mathcal{F}_t\right] \\
&= \int_t^T e^{2g_1(T-s)}(h_0 + h_1\mathbb{E}[x_s|\mathcal{F}_t])\mathrm{d}s + \mu_{J2}\cdot\int_t^T e^{2g_1(T-s)}(l_0 + l_1\mathbb{E}[x_s|\mathcal{F}_t])\mathrm{d}s \\
&= (h_0 + l_0\mu_{J2})\int_t^T e^{2g_1(T-s)}\mathrm{d}s + (h_1 + l_1\mu_{J2})\cdot\int_t^T e^{2g_1(T-s)}\mathbb{E}[x_s|\mathcal{F}_t]\mathrm{d}s,
\end{aligned}$$

where we have again used Fubini's theorem in the third line. Given the conditional expectation, the second integral on the far right-hand side can be simplified further:

$$
\int_t^T e^{2g_1(T-s)} \mathbb{E}[x_s|\mathcal{F}_t] \mathrm{d}s = \int_t^T e^{2g_1(T-s)} \left[ e^{g_1(s-t)} x_t + \frac{g_0}{g_1} \left( e^{g_1(s-t)} - 1 \right) \right] \mathrm{d}s
$$

$$
= e^{2g_1 T} \left[ \int_t^T e^{-g_1(s+t)} x_t + \frac{g_0}{g_1} \left( e^{-g_1(s+t)} - e^{-2g_1 s} \right) \mathrm{d}s \right]
$$

$$
= e^{2g_1 T} \left[ -\frac{1}{g_1} \left( e^{-g_1(T+t)} - e^{-2g_1 t} \right) x_t - \frac{g_0}{g_1^2} \left( e^{-g_1(T+t)} - e^{-2g_1 t} \right) + \frac{g_0}{2g_1^2} \left( e^{-2g_1 T} - e^{-2g_1 t} \right) \right]
$$

$$
= -\frac{1}{g_1} \left( e^{g_1(T-t)} - e^{2g_1(T-t)} \right) x_t + \frac{g_0}{2g_1^2} \left( 1 - e^{g_1(T-t)} \right)^2 .
$$

Thus, the conditional variance in the univariate case is given by

$$
\mathrm{Var}(x_T|\mathcal{F}_t) = -\frac{1}{2g_1} (h_0 + l_0 \mu_{J2}) \left( 1 - e^{2g_1(T-t)} \right)
$$

$$
-\frac{1}{2g_1^2} (h_1 + l_1 \mu_{J2}) \left[ 2g_1 \left( e^{g_1(T-t)} - e^{2g_1(T-t)} \right) x_t - g_0 \left( 1 - e^{g_1(T-t)} \right)^2 \right]. \quad \text{(B.3)}
$$

Equations (B.2) and (B.3) serve as the basis for the formulation of the transition equation (20) as discussed in Section 3.1. It is crucial for our application to note that the conditional mean (B.2) and conditional variance (B.3) of the univariate AJD process $x_T$ at time $T > t$, conditional on information at time $t$, are affine functions in $x_t$. The affinity of the conditional moments yields the linear state updating equation, which, in turn, allows us to use the linear Kalman filtering technique.

Using the same reasoning, it is in principle also possible to derive the analogues of equations (B.2) and (B.3) for the multivariate AJD process. However, these expressions typically involve matrix exponentials and integrals thereof, which makes them burdensome to work with. Fortunately, the conditional moments can easily be obtained numerically by differentiating the CCF, which, as discussed in Section 2.2 is known in semi-closed form for the AJD class. Indeed, finite difference approximations of the first and second derivatives around the origin yield the moments with high precision and little additional computational costs. One can also easily verify that the affine property of the conditional moments holds in the multivariate case by differentiating the exponentially-affine CCF.

# Appendix C    Interpolation-Extrapolation Scheme and CCF Replication

In this appendix, we discuss in detail the option interpolation-extrapolation scheme we adopt and illustrate the impact of the different measurement errors on the option-implied CCF 'payoff' replication.

## C.1 Interpolation-extrapolation scheme

### C.1.1 Interpolation

For each trading day and for each tenor, we interpolate option prices between moneyness levels using cubic splines. For interpolation, we consider option data expressed in terms of their total implied variance, defined as $\varpi(m, \tau) = \kappa^2(m, \tau) \cdot \tau$, where $\kappa(m, \tau)$ is the Black-Scholes implied volatility for an option with log-moneyness $m$ and tenor $\tau$. This is similar to interpolating on the implied volatility domain, but it will provide us further advantages when we proceed to the extrapolation scheme, discussed in the next subsection.

Cubic splines provide a useful tool for the interpolation of options data and are commonly employed for this purpose in the literature; see, for instance, Jiang and Tian (2007), Malz (2014) among many others. Furthermore, they are also used as an approximation method that allows to smooth out noise in the data; see, for instance, Bliss and Panigirtzoglou (2002), Fengler (2009). For the latter, it is common to penalize the squared second derivative of the spline. This might, however, induce a loss of flexibility of the spline leading to larger approximation errors, especially for short-dated options, which are of pivotal importance in our analysis. In this paper, we therefore use a standard cubic spline, but instead of providing all data as knot points for spline interpolation, we explicitly specify which data points shall be used as knots. This allows us to interpolate in some domains and smooth out in others, taking the 'best' out of the spline interpolation and approximation schemes.

Close to ATM options are more liquid than very deep OTM counterparts. Thus, intuitively, information in the former options is more reliable, and we would not like to distort this information by imposing smoothing constraints. Very deep OTM options, on the other hand, may be quite illiquid. Furthermore, the tick size for deep OTM options becomes large relative to their value. This might lead to observing a sequence of the same midpoint quote prices in the data. Figure C.1 provides an example of such 'flat' prices for put options, visible in the right panel for very deep OTM options (i.e., small $k$). These prices clearly violate no-arbitrage assumptions. However, throwing them away would reduce available information, needed to extract the CCF; these prices are not uninformative, but the tick size distorts their information. Therefore, instead of eliminating 'flat' prices, we will just not include them as knot points in our spline interpolation scheme. In other words, we do not require the spline function to go through all data points for deep OTM options, but rather let it approximate the information in them.[17]

More formally, we first include the closest to ATM put option, $m_{n^*}$, in the knot sequence and then iteratively include put options with smaller moneyness level $m_i$ for $i = n^*-1, \ldots, 2$ such that all of the following conditions are satisfied: (i) $P(m_i) < P(m_{i-1})$ and $C(m_i) > C(m_{i-1})$; (ii) $P(m_{i+1}) < P(m_i)$ and $C(m_{i+1}) > C(m_i)$; (iii) daily trading volume for $P(m_i)$ is larger than one. The first two criteria check for no-arbitrage conditions. The third one filters out possible stale prices from being a knot point. Similar mirrored conditions are applied to OTM call prices. The knot sequence thus constructed will likely contain more close to ATM options

---

[17]Recall that we interpolate/approximate data on the total implied variance domain, not in terms of implied volatility, option prices or log prices as considered in Figure C.1.

Figure C.1: Spline interpolation-extrapolation example: April 1, 2021, 15 days to maturity
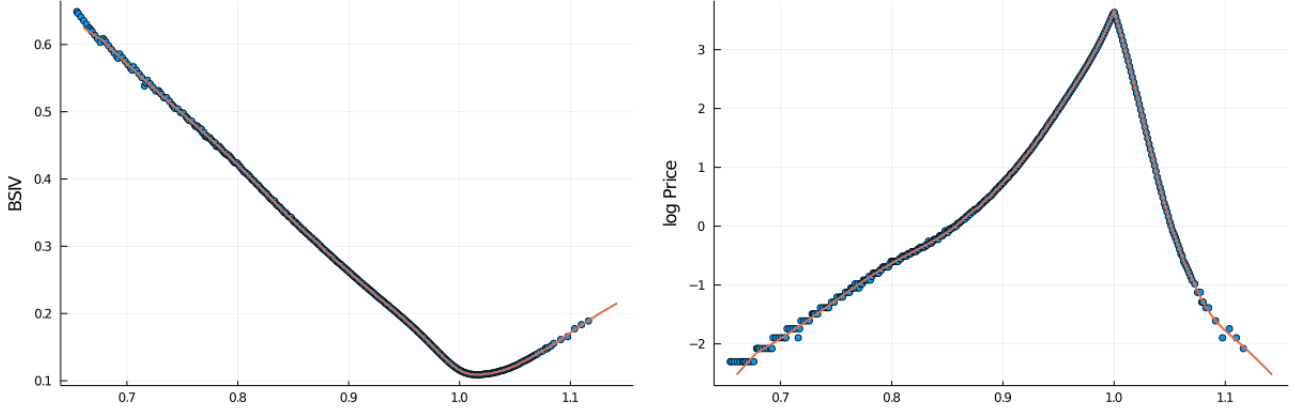


Note: This figure plots an example of the interpolation-extrapolation scheme for options traded on April 1, 2021 with 15 days to maturity. The option data (blue dots) are interpolated using a cubic spline (orange line). Interpolation is conducted on the total implied variance domain. The left panel plots the data in terms of Black-Scholes implied volatility, whereas the right panel plots the data in terms of log prices. Moneyness $K/F_t$ is on the horizontal axis.

and fewer deep OTM options, resulting in more interpolation in the former range and more approximation in the latter one.

We emphasize again that we do not filter out option data that violate no-arbitrage conditions, which would reduce available information for CCF extraction. Instead, we do not include these points into the knot sequence, thus we do not require the spline to go exactly through these points. Furthermore, another reason not to filter out options that violate no-arbitrage conditions, is that we use option-implied CCFs rather than option prices themselves as inputs in our estimation procedure. Similarly, the CBOE does not impose any no-arbitrage filters in the calculation of the VIX index, except for eliminating zero-bid quotes (CBOE, 2015). Figure C.1 provides an example of the interpolation-extrapolation scheme for an option slice traded on April 1, 2020, with 15 days to maturity.

### C.1.2 Extrapolation

Truncation errors are, in a sense, more challenging to address than discretization errors, since they require to make assumptions about the dynamics of option prices (either in dollar or volatility terms) beyond the observable range of strikes. On the other hand, as prices of OTM options decrease with $|m|$, the impact of the truncation errors is expected to be small for highly liquid options that cover a wide range of strike prices (such as index options). Nevertheless, truncation might deteriorate the CCF approximation even for small argument values. This can especially be a relevant issue after a sudden market shock, since options with smaller or larger strikes might not be issued immediately to cover a new range of strikes.

It is common in the literature to use flat extrapolation; see again e.g., Bliss and Panigirt-zoglou (2002), Jiang and Tian (2005) and Malz (2014). Under a flat extrapolation scheme, the implied volatility beyond the observable range of strikes is simply set equal to the volatility of the observable extreme-strike options, i.e., $\kappa(\underline{m}, \tau)$ for the left-hand side of the volatility smile and $\kappa(\overline{m}, \tau)$ for the right-hand side. This approach is very easy to implement. However, the

main caveat of flat extrapolation is that it assumes the Black-Scholes log-normal model to apply in the tails, for the extrapolated range of strikes.

Instead, we extrapolate the total implied variance $\varpi(m, \tau)$ linearly in log-moneyness $m$ beyond the observable range of strikes. This particular linear parametrization is motivated by the asymptotic results of Lee (2004), who analyzed the behavior of the implied volatility smile as strikes tend to infinity. Another example of a parametrization that satisfies Lee's asymptotic results is the SVI model, commonly used among practitioners (Gatheral & Jacquier, 2014). However, it is well known that the SVI approach may not provide a good fit for short-dated options. Thus, different from SVI, we use the more flexible cubic spline for interpolation within the observable range of strikes, as detailed above, and, similar to SVI, extrapolate implied variance linearly in log-moneyness.

The asymptotic results of Lee (2004), exploited also in Appendix A and recalled here for convenience, entail that the implied volatility wings should not grow faster than $|m|^{1/2}$ and, unless the underlying asset has finite moments of all orders, should not grow slower than $|m|^{1/2}$. More specifically, Lee (2004) first shows that

$$\limsup_{m \to -\infty} \frac{\kappa^2(\tau, m)\tau}{|m|} = \underline{\beta}^* \text{ with } \underline{\beta}^* \in [0, 2], \quad \text{and} \tag{C.1}$$

$$\limsup_{m \to \infty} \frac{\kappa^2(\tau, m)\tau}{|m|} = \overline{\beta}^* \text{ with } \overline{\beta}^* \in [0, 2]. \tag{C.2}$$

Furthermore, he establishes that there is a one-to-one correspondence between $\overline{\beta}^*$ and the number of finite moments of the underlying process $F_\tau$, and between $\underline{\beta}^*$ and the number of finite moments of $1/F_\tau$. For instance, for the right tail, the moment formula for implied volatility is given by

$$\frac{1}{2\overline{\beta}^*} + \frac{\overline{\beta}^*}{8} - \frac{1}{2} = \sup\{p : \mathbb{E}[F_\tau^{1+p}] < \infty\}.$$

These results allow us to conjecture the asymptotically valid parametrization to extrapolate implied volatility beyond the observable range of strikes. Hence, we assume that the total variance $\varpi(m) = \kappa^2(\tau, m)\tau$ is an affine function of log-moneyness:

$$\varpi(m) = c + \beta m.$$

An intercept coefficient is introduced to guarantee continuity between the interpolation and extrapolation domains. The intercept coefficients for the left and right tails, denoted by $\underline{c}$ and $\overline{c}$, are exactly determined by the smallest and largest observable strike prices (or corresponding log-moneyness levels) given the slopes $\underline{\beta}$ and $\overline{\beta}$ for the left and right tails, respectively:

$$\underline{c} = \varpi(\underline{m}) - \underline{\beta}\underline{m} \quad \text{and} \quad \overline{c} = \varpi(\overline{m}) - \overline{\beta}\overline{m}.$$

Therefore, what is left to be done is to establish the choice of the slope coefficients $\underline{\beta}$ and $\overline{\beta}$. Note that the formulas (C.1) and (C.2) provide asymptotic bounds for the slope coefficients. Furthermore, finding the number of finite moments of the underlying, and exploiting

the respective moment formulas, would require parametrizing the dynamics of $F_\tau$.[18] The latter is not desirable in our application, since we want to fit another parametric model afterwards. Instead, we simply use the derivatives of the fitted cubic splines at the last observable strikes to determine the slope coefficients:

$$\underline{\beta} = -\left.\frac{\partial \varpi(m)}{\partial m}\right|_{m=\underline{m}} \quad \text{and} \quad \overline{\beta} = \left.\frac{\partial \varpi(m)}{\partial m}\right|_{m=\overline{m}}.$$

Lee's bounds for the slopes constitute an asymptotic result. The chosen slopes $\underline{\beta}$ and $\overline{\beta}$ should satisfy these bounds. (Note that, due to the adopted sign convention in the extrapolation formula, this translates into $\underline{\beta} \in [-2, 0]$ for the left slope.) However, no-arbitrage conditions for our parametrization can be tighter, since we are in a setting with finite log-moneyness levels. To obtain these conditions, we follow the derivation sketched in Jäckel (2014). This yields the following no-arbitrage bounds for the right-tail slope $\overline{\beta}$ (our detailed derivations are available upon request; they are suppressed to save space):

$$0 \le \overline{\beta} < \min(\beta_{max}, 2),$$

where

$$\beta_{max} = \begin{cases} \max\left(\frac{\overline{m}(\overline{\varpi}-2)+\sqrt{\overline{\Delta}}}{\overline{m}^2+1}, \frac{-2\overline{m}+2\sqrt{\overline{m}^2+2\overline{\varpi}^2+4\overline{\varpi}}}{\overline{\varpi}+2}\right), & \text{if } \overline{\Delta} > 0; \\ \frac{-2\overline{m}+2\sqrt{\overline{m}^2+2\overline{\varpi}^2+4\overline{\varpi}}}{\overline{\varpi}+2}, & \text{if } \overline{\Delta} \le 0; \end{cases}$$

with $\overline{\varpi} := \varpi(\overline{m})$ and $\overline{\Delta} := 4\overline{m}^2 - \overline{\varpi}^2 + 4\overline{\varpi}$.

Similarly, for the left-tail slope $\underline{\beta}$,

$$\max(\beta_{min}, -2) < \underline{\beta} \le 0,$$

where

$$\beta_{min} = \begin{cases} \max\left(\frac{\underline{m}(\underline{\varpi}-2)-\sqrt{\underline{\Delta}}}{\underline{m}^2+1}, \frac{-2\underline{m}-2\sqrt{\underline{m}^2+2\underline{\varpi}^2+4\underline{\varpi}}}{\underline{\varpi}+2}\right), & \text{if } \underline{\Delta} > 0; \\ \frac{-2\underline{m}+2\sqrt{\underline{m}^2+2\underline{\varpi}^2+4\underline{\varpi}}}{\underline{\varpi}+2}, & \text{if } \underline{\Delta} \le 0; \end{cases}$$

with $\underline{\varpi} := \varpi(\underline{m})$ and $\underline{\Delta} := 4\underline{m}^2 - \underline{\varpi}^2 + 4\underline{\varpi}$.

## C.2   CCF replication

As discussed in Section 3.1, we replicate the CCF 'payoff' using a Riemann sum approximation, and employ the interpolation-extrapolation scheme detailed in the previous subsection, applied to the set of observable option prices, to reduce the discretization and truncation errors. Figure C.2 illustrates the impact of the different measurement errors on the option-implied CCF. For the illustration, we simulate option prices from the SVCDEJ model using a similar setup as described in Section 4. In particular, at each time point we have a discrete set of strikes

---

[18]Note that flat extrapolation assumes log-normality of the underlying asset in the tails. Since all moments of the log-normal distribution exist, it means that the slope indeed has to be zero in this case.

and additive observation errors in the observed option prices. We fix the time-to-maturity to $\tau = 10$ days, take $u = 20$ and focus only on the real part of the CCF. These values are chosen to emphasize the impact of the measurement errors. The impact of the discretization errors, for instance, is typically smaller for larger maturities and smaller argument values.

Figure C.2: The three types of measurement errors



(a) $\zeta_t^{(1)} + \zeta_t^{(2)} + \zeta_t^{(3)}$

(b) $\zeta_t^{(2)} + \zeta_t^{(3)}$ (solid) and $-$ATM BSIV (dots)

(c) $\phi(u,\tau) - \phi(spl(\widehat{O}))$

(d) $\phi(u,\tau) - \phi(spl(O))$
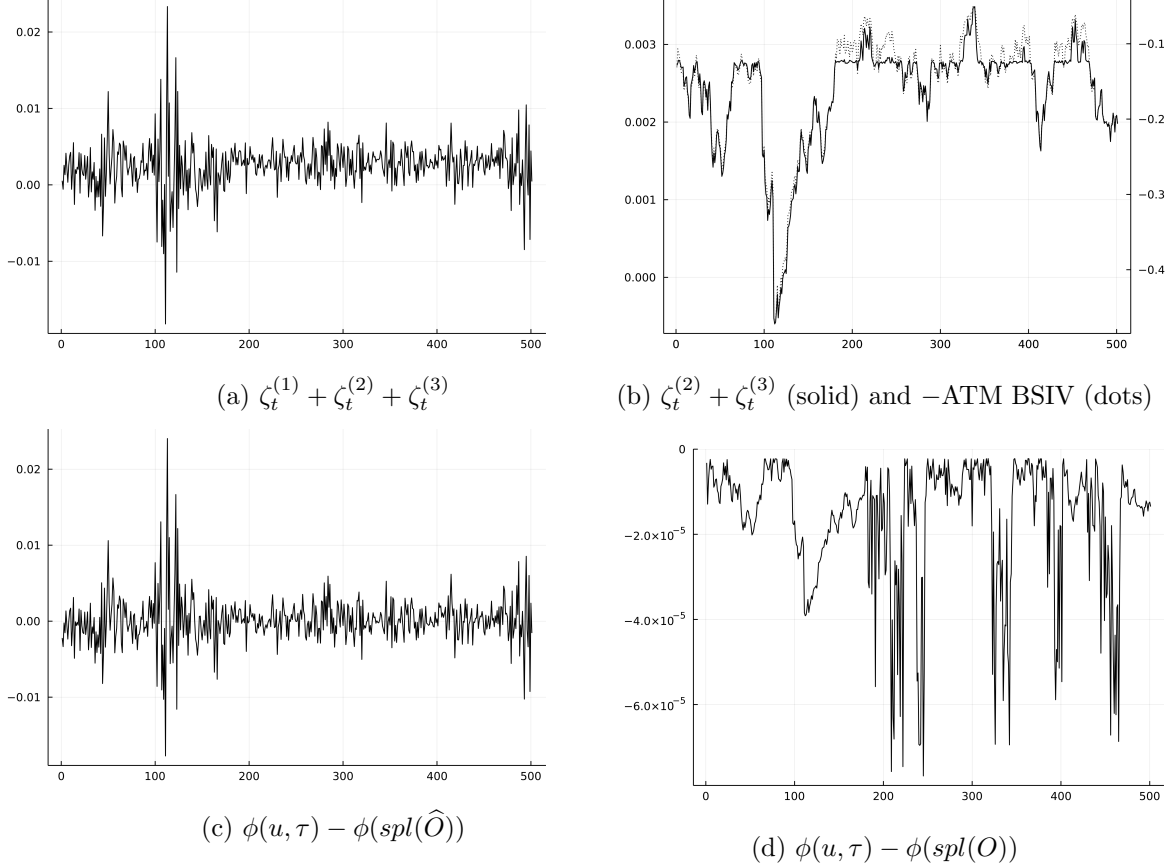
Note: This figure plots the impact of the three types of measurement errors on the option-implied CCF. The figures illustrate the approximation for the real part of the CCF with $\tau = 10$, $u = 20$. The same Monte Carlo simulation setup as described in Section 4 is used here to simulate data from the SVCDEJ model.

Panel (a) of Figure C.2 plots the measurement errors when we use a finite set of observed option prices $\widehat{O}_t(\tau, m)$. That is, it shows the full measurement errors $\zeta_t^{\phi}(u,\tau)$, given by the sum of the observation errors $\zeta_t^{(1)}(u,\tau)$, the truncation errors $\zeta_t^{(2)}(u,\tau)$ and the discretization errors $\zeta_t^{(3)}(u,\tau)$, formally defined in Appendix A. We observe that the errors are not exactly centered at zero, implying a small bias in the CCF approximation. From panel (b), which eliminates the impact of the observation errors $\zeta_t^{(1)}(u,\tau)$ (by using a finite set of *true* option prices $O_t(\tau, m)$), we observe the same small non-zero mean in $\zeta_t^{(2)}(u,\tau) + \zeta_t^{(3)}(u,\tau)$. We overlay this plot with the ATM BSIV to illustrate that the sum of the truncation and discretization errors is strongly negatively correlated with the implied volatility, and hence driven by the volatility dynamics.

For panels (c) and (d), we use a cubic spline interpolation and extrapolate option prices outside of the observed range of strikes, as described in Appendix C.1. Panel (d) plots the errors in the CCF approximation when we use the true finite set of option prices. As we can see,

the discretization and truncation errors are largely reduced by the interpolation-extrapolation scheme (note the scale of the vertical axis). Finally, panel (c) illustrates the errors in the CCF approximation when we apply the same interpolation and extrapolation to option prices observed with error. We observe that the observation errors $\zeta_t^{(1)}(u, \tau)$ (which dominate both panels (a) and (c)) are largely unaffected by the interpolation-extrapolation scheme, but the (already small) bias from the impact of the discretization and truncation errors has been further reduced, leading to errors that are virtually centered around zero.

# Appendix D  Additional Simulation and Empirical Results

In this appendix, we first provide additional simulation results for two related alternative option pricing models, to supplement Section 4. We also consider a model specification that includes a variance risk premium. Furthermore, we analyze in simulations the SVCDEJ model of Section 4 under dependent option pricing errors, and we compare its filtered volatility to the simulated volatility process. Next, we provide some additional empirical results to analyze the robustness of our empirical findings reported in Section 6 and the empirical fit of the SVCDEJ model.

## D.1  Additional simulation results

### D.1.1  SVCJ

We additionally illustrate the developed estimation procedure using the 'double-jump' stochastic volatility model of Duffie et al. (2000) with a Gaussian jump size distribution. In particular, we assume the following process, referred to in shorthand as 'SVCJ', for the log forward price under both the $\mathbb{P}$ and $\mathbb{Q}$ probability measures:

$$\mathrm{d}\log F_t = (-\tfrac{1}{2}v_t - \mu\lambda_t)\mathrm{d}t + \sqrt{v_t}\mathrm{d}W_{1,t} + J_t\mathrm{d}N_t, \tag{D.1}$$

$$\mathrm{d}v_t = \kappa(\bar{v} - v_t)\mathrm{d}t + \sigma\sqrt{v_t}\mathrm{d}W_{2,t} + J_t^v\mathrm{d}N_t, \tag{D.2}$$

where nearly all ingredients are the same as in the SVCDEJ specification in Section 4.1, except for the distribution of the jump sizes. In particular, we assume here that the jump sizes in returns are Gaussian, $J \sim \mathcal{N}(\mu_J, \sigma_J^2)$, and the jump sizes in volatility are independent from the jump sizes in returns with $J^v \sim \exp(1/\mu_v)$.

Similar to the specification in the main text, this model belongs to the AJD class and the log of the option-spanned CCF is linear in the latent state process $v_t$. The conditional mean and variance of the latent stochastic volatility process are given by

$$\mathbb{E}[v_{t+1}|\mathcal{F}_t] = e^{g_1\Delta t}v_t + \frac{g_0}{g_1}\left(e^{g_1\Delta t} - 1\right), \tag{D.3}$$

$$\mathrm{Var}(v_{t+1}|\mathcal{F}_t) = -\frac{\sigma^2 + 2\delta\mu_v^2}{2g_1^2}\left[2g_1\left(e^{g_1\Delta t} - e^{2g_1\Delta t}\right)v_t - g_0\left(1 - e^{g_1\Delta t}\right)^2\right], \tag{D.4}$$

with $g_0 = \kappa\bar{v}$ and $g_1 = -\kappa + \delta\mu_v$. Equations (D.3) and (D.4) are used to define the state updating equation:

$$v_{t+1} = c_t + T_t v_t + \eta_{t+1}, \tag{D.5}$$

where $c_t = \frac{g_0}{g_1}\left(e^{g_1\Delta t} - 1\right)$, $T_t = e^{g_1\Delta t}$ and $\text{Var}(\eta_{t+1}|\mathcal{F}_t) = \text{Var}(v_{t+1}|\mathcal{F}_t)$. We also impose the Feller condition $2\kappa\bar{v} > \sigma^2$ and the covariance stationarity condition $\kappa > \delta\mu_v$.

Table D.1: Monte Carlo results for the SVCJ model

| parameter | $\sigma$ | $\kappa$ | $\bar{v}$ | $\rho$ | $\delta$ | $\mu_J$ | $\sigma_J$ | $\mu_v$ | $\sigma_\varkappa$ |
|---|---|---|---|---|---|---|---|---|---|
| | | | | $u = 1, \ldots, 15$ | | | | | |
| true value | 0.400 | 5.000 | 0.02 | -0.95 | 20.000 | -0.100 | 0.04 | 0.05 | 0.02 |
| mean | 0.410 | 4.869 | 0.0207 | -0.9382 | 17.013 | -0.110 | 0.0343 | 0.0520 | 0.0215 |
| std dev | 0.012 | 0.115 | 0.0007 | 0.0151 | 3.171 | 0.014 | 0.0124 | 0.0026 | 0.0044 |
| q10 | 0.400 | 4.681 | 0.0201 | -0.9516 | 11.209 | -0.136 | 0.0100 | 0.0501 | 0.0166 |
| q50 | 0.405 | 4.905 | 0.0204 | -0.9445 | 18.537 | -0.103 | 0.0404 | 0.0507 | 0.0206 |
| q90 | 0.433 | 4.988 | 0.0219 | -0.9114 | 19.573 | -0.100 | 0.0431 | 0.0570 | 0.0274 |
| | | | | $u = 1, \ldots, 20$ | | | | | |
| true value | 0.400 | 5.000 | 0.02 | -0.95 | 20.000 | -0.100 | 0.04 | 0.05 | 0.02 |
| mean | 0.403 | 4.913 | 0.0202 | -0.9444 | 18.866 | -0.103 | 0.0397 | 0.0502 | 0.0198 |
| std dev | 0.006 | 0.091 | 0.0003 | 0.0085 | 1.420 | 0.006 | 0.0050 | 0.0010 | 0.0062 |
| q10 | 0.396 | 4.830 | 0.0199 | -0.9524 | 17.903 | -0.105 | 0.0375 | 0.0496 | 0.0156 |
| q50 | 0.403 | 4.930 | 0.0202 | -0.9456 | 19.048 | -0.102 | 0.0408 | 0.0501 | 0.0185 |
| q90 | 0.410 | 5.000 | 0.0206 | -0.9373 | 20.088 | -0.100 | 0.0426 | 0.0506 | 0.0234 |
| | | | | $u = 1, \ldots, 25$ | | | | | |
| true value | 0.400 | 5.000 | 0.02 | -0.95 | 20.000 | -0.100 | 0.04 | 0.05 | 0.02 |
| mean | 0.395 | 4.907 | 0.0200 | -0.9555 | 20.424 | -0.097 | 0.0435 | 0.0495 | 0.0207 |
| std dev | 0.009 | 0.129 | 0.0003 | 0.0146 | 1.297 | 0.004 | 0.0029 | 0.0005 | 0.0097 |
| q10 | 0.384 | 4.776 | 0.0196 | -0.9710 | 19.209 | -0.101 | 0.0405 | 0.0489 | 0.0152 |
| q50 | 0.395 | 4.934 | 0.0200 | -0.9564 | 20.392 | -0.097 | 0.0436 | 0.0495 | 0.0178 |
| q90 | 0.404 | 5.010 | 0.0203 | -0.9409 | 22.029 | -0.093 | 0.0473 | 0.0500 | 0.0284 |

Note: This table provides Monte Carlo simulation results for the SVCJ model, based on 500 replications. Three settings with different ranges of the argument $u$ are considered. Each panel lists, for each parameter, the true value, the Monte Carlo mean and standard deviation, and the 10th, 50th and 90th Monte Carlo percentiles, respectively. We use $T = 500$ time points with $\Delta t = 1/250$. The initial values are set to $F_0 = 100$ and $v_0 = 0.02$. The threshold for singular values is set to $\bar{s} = 10^{-7}$.

We use the same simulation setting as in Section 4.1, *mutatis mutandis*. The simulation results are provided in Table D.1. Just like for the SVCDEJ model specification of Section 4.1, the results for the SVCJ model also display high-quality finite-sample properties. We also note that the 'double-jump' specification includes other widely used option pricing models as special cases, such as the stochastic volatility model of Heston (1993).

### D.1.2 SVCJ with a variance risk premium

Since the transition equation in the state space representation reflects the $\mathbb{P}$-dynamics of the latent components, it is, in principle, possible to conduct inference on the risk premia associated with this latent process. In this subsection, we provide Monte Carlo simulation results for the SVCJ model with a variance risk premium (VRP). (To facilitate identification, we focus on the slightly more parsimonious SVCJ model rather than the SVCDEJ model; it will turn out that already in the more parsimonious model, the VRP is weakly identified.) In particular, we model the VRP $\pi_v$ as the difference between the mean-reversion parameters under the $\mathbb{P}$ and $\mathbb{Q}$ measures, that is, in the state transition equation (D.5) we change the mean-reversion

Table D.2: Monte Carlo results for the SVCJ model with variance risk premium

| parameter | $\sigma$ | $\kappa$ | $\bar{v}$ | $\rho$ | $\delta$ | $\mu_J$ | $\sigma_J$ | $\mu_v$ | $\sigma_\varkappa$ | $\pi_v$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | constrained, $\pi_v = 1.0$ | | | | | | |
| true value | 0.400 | 5.000 | 0.0200 | -0.950 | 20.000 | -0.100 | 0.0400 | 0.0500 | 0.0200 | 1.000 |
| mean | 0.402 | 4.924 | 0.0202 | -0.946 | 18.882 | -0.103 | 0.0396 | 0.0503 | 0.0158 | - |
| std dev | 0.007 | 0.073 | 0.0003 | 0.008 | 1.515 | 0.006 | 0.0051 | 0.0010 | 0.0031 | - |
| q10 | 0.396 | 4.841 | 0.0199 | -0.953 | 18.012 | -0.105 | 0.0374 | 0.0496 | 0.0127 | - |
| q50 | 0.401 | 4.931 | 0.0202 | -0.947 | 19.136 | -0.101 | 0.0408 | 0.0501 | 0.0155 | - |
| q90 | 0.409 | 5.003 | 0.0205 | -0.939 | 20.182 | -0.099 | 0.0427 | 0.0508 | 0.0189 | - |
| | | | | constrained, $\pi_v = 0.0$ | | | | | | |
| true value | 0.400 | 5.000 | 0.0200 | -0.950 | 20.000 | -0.100 | 0.0400 | 0.0500 | 0.0200 | 1.000 |
| mean | 0.402 | 4.924 | 0.0202 | -0.946 | 18.882 | -0.103 | 0.0396 | 0.0503 | 0.0158 | - |
| std dev | 0.007 | 0.073 | 0.0003 | 0.008 | 1.515 | 0.006 | 0.0051 | 0.0010 | 0.0031 | - |
| q10 | 0.396 | 4.841 | 0.0199 | -0.953 | 18.012 | -0.105 | 0.0374 | 0.0496 | 0.0127 | - |
| q50 | 0.401 | 4.931 | 0.0202 | -0.947 | 19.136 | -0.101 | 0.0408 | 0.0501 | 0.0155 | - |
| q90 | 0.409 | 5.003 | 0.0205 | -0.939 | 20.182 | -0.099 | 0.0427 | 0.0508 | 0.0189 | - |
| | | | | unconstrained | | | | | | |
| true value | 0.400 | 5.000 | 0.0200 | -0.950 | 20.000 | -0.100 | 0.0400 | 0.0500 | 0.0200 | 1.000 |
| mean | 0.402 | 4.923 | 0.0202 | -0.945 | 18.865 | -0.103 | 0.0395 | 0.0503 | 0.0158 | 4.029 |
| std dev | 0.007 | 0.074 | 0.0003 | 0.008 | 1.546 | 0.006 | 0.0052 | 0.0011 | 0.0031 | 4.942 |
| q10 | 0.396 | 4.835 | 0.0199 | -0.953 | 17.985 | -0.105 | 0.0373 | 0.0496 | 0.0127 | -4.338 |
| q50 | 0.401 | 4.931 | 0.0202 | -0.947 | 19.136 | -0.101 | 0.0408 | 0.0502 | 0.0155 | 4.558 |
| q90 | 0.409 | 5.003 | 0.0206 | -0.939 | 20.182 | -0.099 | 0.0427 | 0.0508 | 0.0189 | 9.883 |

Note: This table provides Monte Carlo simulation results for the SVCJ model with a variance risk premium, based on 500 replications. Each panel lists, for each parameter, the true value, the Monte Carlo mean and standard deviation, and the 10th, 50th and 90th Monte Carlo percentiles, respectively. We use $T = 500$ time points with $\Delta t = 1/250$. The range of arguments is set to $u = 1, \ldots, 20$ and the threshold to $\bar{s} = 10^{-7}$. The initial values are set to $F_0 = 100$ and $v_0 = 0.02$.

parameter to $\kappa^{\mathbb{P}} = \kappa + \pi_v$.

Table D.2 provides Monte Carlo simulation results for the SVCJ model with a VRP. We consider three estimation strategies. First, we fix the VRP parameter to its true value $\pi_v = 1$. Second, we assume no VRP when estimating the model, although the true model is simulated with a non-zero VRP, that is, we fix $\pi_v = 0$ in the estimation procedure. Finally, we estimate the VRP along with all model parameters.

As the results suggest, it is hard to identify the VRP in this setting (see the third panel in Table D.2). It appears that the identification of the VRP is weaker than that of the $\mathbb{Q}$-parameters. This is in line with the asymptotic results established in Proposition 2, which points toward slower convergence rate of the parameters that appear only in the state updating equation. In additional simulation results (not reported here), we see that the identification of the VRP parameter improves for a larger number of time series observations $T$. Accurate estimation, however, requires a much larger time span than we consider in our simulations and empirical application. A similar issue arises in the term structure literature, where calibrated bond prices often imply unrealistic $\mathbb{P}$-dynamics (see, e.g., the discussion in Kim & Orphanides, 2012). However, we also notice that under all three estimation strategies, the identification of the $\mathbb{Q}$-parameters barely changes. That is, even in the misspecified model with the VRP

parameter fixed to zero, the parameter estimates display good finite-sample properties (see the second panel in Table D.2). Consistent with this, we also find (in results not provided here) that if we were to introduce a VRP parameter in the SVCDEJ model in the empirical application of Section 6 (which, supported by the Monte Carlo results, we do not), it would not have a significant effect on the estimates of the model's $\mathbb{Q}$-parameters. For more focused VRP estimation, one can use, e.g., a non-parametric approach based on high-frequency data, as in Bollerslev and Todorov (2011) and Andersen, Fusari, and Todorov (2015b).

### D.1.3 SVCEJ

Instead of the double-exponential jump size distribution considered in Section 4.1, or the Gaussian distribution considered above, one may also consider separate exponential distributions for positive and negative jumps. Following Bardgett et al. (2019), we consider two separate counting processes $N_t^-$ and $N_t^+$ for negative and positive jumps, respectively, and modify the SVCJ specification to obtain the 'SVCEJ' model as follows:

$$\mathrm{d}\log F_t = (-\frac{1}{2}v_t - \mu^-\lambda_t^+ - \mu^-\lambda_t^+)\mathrm{d}t + \sqrt{v_t}\mathrm{d}W_{1,t} + J_t^-\mathrm{d}N_t^- + J_t^+\mathrm{d}N_t^+, \quad (D.6)$$

$$\mathrm{d}v_t = \kappa(\bar{v} - v_t)\mathrm{d}t + \sigma\sqrt{v_t}\mathrm{d}W_{2,t} + J_t^v\mathrm{d}N_t^-, \quad (D.7)$$

where $\lambda_t^-$ and $\lambda_t^+$ are the corresponding jump intensities for negative and positive jumps, and $-J_t^-$ and $J_t^+$ are exponentially distributed negative and positive jump sizes in log returns with means $\eta^-$ and $\eta^+$, respectively. Note that the negative jump sizes have negative support, that is, $J_t^-$ is negative exponential. Given the jump size distributions, the expected relative jump sizes in returns are $\mu^- = \mathbb{E}[e^{J^-} - 1] = -\eta^-/(1 + \eta^-)$ and $\mu^+ = \mathbb{E}[e^{J^+} - 1] = \eta^+/(1 - \eta^+)$. We further let the intensities be affine functions of the stochastic volatility, that is, $\lambda_t^- = \delta_0^- + \delta_1^- v_t$ and $\lambda_t^+ = \delta_0^+ + \delta_1^+ v_t$. However, to keep a moderate number of parameters, we set $\delta_1^+ = 0$ and $\delta_0^- = 0$.

This specification is somewhat richer than the SVCDEJ considered in Section 4.1 since positive jumps are modeled by a separate counting process with its own jump intensity process $\lambda_t^+$. Nevertheless, this specification also belongs to the AJD class and the CCF of log forward prices has a semi-closed form. The state updating equation is defined in a similar way as for the other specifications.

The Monte Carlo simulation results for the SVCEJ model are provided in Table D.3. We notice that most of the parameters exhibit good finite-sample performance. However, the parameters related to the positive jumps are biased and have a large standard deviation.

### D.1.4 SVCDEJ

As a robustness check for the simulation results reported in Section 4.1, we additionally consider the SVCDEJ model under misspecified error terms. In particular, Table D.4 provides simulation results under a simulated AR(1) error structure, while keeping the estimation procedure the same as in Section 4.1. Furthermore, Table D.5 contains simulation results for errors that are generated with a spatial correlation between two adjacent strike prices of $\rho = 0.5$. As the results

Table D.3: Monte Carlo results for the SVCEJ model

| parameter | $\sigma$ | $\kappa$ | $\bar{v}$ | $\rho$ | $\delta_0^+$ | $\delta_1^-$ | $\eta^+$ | $\eta^-$ | $\mu_v$ | $\sigma_\varkappa$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | $u = 1, \ldots, 15$ | | | | | |
| true value | 0.450 | 8.000 | 0.015 | -0.95 | 2.000 | 100.000 | 0.01 | 0.05 | 0.05 | 0.02 |
| mean | 0.483 | 8.063 | 0.0159 | -0.9242 | 0.512 | 99.475 | 0.0355 | 0.0527 | 0.0529 | 0.0626 |
| std dev | 0.069 | 1.506 | 0.0026 | 0.0796 | 1.690 | 21.104 | 0.0211 | 0.0196 | 0.0351 | 0.1820 |
| q10 | 0.459 | 7.642 | 0.0150 | -0.9577 | 0.028 | 90.701 | 0.0191 | 0.0485 | 0.0473 | 0.0170 |
| q50 | 0.486 | 7.914 | 0.0157 | -0.9261 | 0.125 | 99.718 | 0.0320 | 0.0495 | 0.0482 | 0.0229 |
| q90 | 0.511 | 8.306 | 0.0168 | -0.9049 | 0.578 | 106.121 | 0.0524 | 0.0505 | 0.0504 | 0.0302 |
| | | | | | $u = 1, \ldots, 20$ | | | | | |
| true value | 0.450 | 8.000 | 0.015 | -0.95 | 2.000 | 100.000 | 0.01 | 0.05 | 0.05 | 0.02 |
| mean | 0.462 | 8.178 | 0.0149 | -0.9551 | 1.309 | 110.631 | 0.0166 | 0.0510 | 0.0464 | 0.0482 |
| std dev | 0.042 | 0.804 | 0.0020 | 0.0266 | 1.566 | 18.690 | 0.0130 | 0.0156 | 0.0056 | 0.1425 |
| q10 | 0.443 | 7.849 | 0.0144 | -0.9851 | 0.531 | 101.950 | 0.0121 | 0.0478 | 0.0452 | 0.0143 |
| q50 | 0.454 | 8.099 | 0.0148 | -0.9588 | 0.964 | 107.556 | 0.0155 | 0.0489 | 0.0468 | 0.0184 |
| q90 | 0.488 | 8.410 | 0.0154 | -0.9165 | 1.742 | 115.366 | 0.0190 | 0.0498 | 0.0484 | 0.0368 |
| | | | | | $u = 1, \ldots, 25$ | | | | | |
| true value | 0.450 | 8.000 | 0.015 | -0.95 | 2.000 | 100.000 | 0.01 | 0.05 | 0.05 | 0.02 |
| mean | 0.466 | 7.966 | 0.0154 | -0.9417 | 1.526 | 104.783 | 0.0157 | 0.0515 | 0.0479 | 0.0381 |
| std dev | 0.048 | 0.615 | 0.0022 | 0.0321 | 1.322 | 17.732 | 0.0203 | 0.0140 | 0.0114 | 0.1101 |
| q10 | 0.442 | 7.671 | 0.0146 | -0.9726 | 0.602 | 97.221 | 0.0106 | 0.0488 | 0.0462 | 0.0129 |
| q50 | 0.454 | 7.963 | 0.0150 | -0.9484 | 1.394 | 103.412 | 0.0126 | 0.0496 | 0.0476 | 0.0152 |
| q90 | 0.494 | 8.171 | 0.0159 | -0.8935 | 2.130 | 109.523 | 0.0167 | 0.0505 | 0.0492 | 0.0417 |

Note: This table provides Monte Carlo simulation results for the SVCEJ model, based on 300 replications. Three settings with different ranges of the argument $u$ are considered. Each panel lists, for each parameter, the true value, the Monte Carlo mean and standard deviation, and the 10th, 50th and 90th Monte Carlo percentiles, respectively. We use $T = 500$ time points with $\Delta t = 1/250$. The initial values are set to $F_0 = 100$ and $v_0 = 0.015$. The threshold for singular values is set to $\bar{s} = 10^{-7}$.

indicate, a misspecified error structure may increase the variance of the estimated parameters, but the finite-sample performance is broadly robust to dependent option pricing errors.

Additionally, Figure D.1 displays an example of the filtered volatility based on the estimated parameters of the considered SVCDEJ model next to the true stochastic volatility process. We observe from the figure that the filtered volatility closely follows the true spot volatility, even in the presence of jumps in volatility. A few more examples are provided in our GitHub repository.

## D.2 Additional empirical results

### D.2.1 Robustness checks

Table D.6 provides additional empirical results for the model specification of Section 6. Next to the empirical results with fixed $p^- = 0.7$ reported in Section 6, we provide the estimates for an unrestricted probability of negative jumps and for different fixed values $p^- = 0.65$ and $p^- = 0.75$. Overall, the results indicate similar parameter estimates as in Table 5, which is reassuring for the robustness of our empirical results. We also note larger standard errors of the parameter estimates in the unrestricted model, specifically for the parameter $\delta$, which enters the model as a multiple of $p^-$. This is in line with our simulation results for the unrestricted model (not provided here), which show the limits to identification in the considered unrestricted model.

Table D.4: Monte Carlo results for the SVCDEJ model with AR(1) errors

| parameter | $\sigma$ | $\kappa$ | $\bar{v}$ | $\rho$ | $\delta$ | $\eta^+$ | $\eta^-$ | $\mu_v$ | $\sigma_\varkappa$ |
|---|---|---|---|---|---|---|---|---|---|
| | | | | $u = 1, \ldots, 15$ | | | | | |
| true value | 0,450 | 8,000 | 0,0150 | -0,950 | 100,000 | 0,0200 | 0,0500 | 0,0500 | 0,020 |
| mean | 0,453 | 7,989 | 0,0152 | -0,952 | 98,937 | 0,0202 | 0,0500 | 0,0500 | 0,022 |
| std | 0,019 | 0,351 | 0,0011 | 0,023 | 6,386 | 0,0020 | 0,0016 | 0,0020 | 0,004 |
| q10 | 0,436 | 7,601 | 0,0147 | -0,975 | 89,054 | 0,0190 | 0,0491 | 0,0490 | 0,020 |
| q50 | 0,448 | 8,047 | 0,0150 | -0,957 | 100,667 | 0,0201 | 0,0497 | 0,0501 | 0,022 |
| q90 | 0,494 | 8,196 | 0,0161 | -0,906 | 103,761 | 0,0213 | 0,0516 | 0,0513 | 0,025 |
| | | | | $u = 1, \ldots, 20$ | | | | | |
| true value | 0,450 | 8,000 | 0,0150 | -0,950 | 100,000 | 0,0200 | 0,0500 | 0,0500 | 0,020 |
| mean | 0,445 | 7,963 | 0,0150 | -0,964 | 101,624 | 0,0217 | 0,0499 | 0,0488 | 0,023 |
| std | 0,022 | 0,245 | 0,0006 | 0,028 | 6,419 | 0,0015 | 0,0013 | 0,0015 | 0,003 |
| q10 | 0,423 | 7,572 | 0,0145 | -1,000 | 91,088 | 0,0206 | 0,0489 | 0,0467 | 0,021 |
| q50 | 0,439 | 8,010 | 0,0148 | -0,970 | 103,136 | 0,0216 | 0,0497 | 0,0489 | 0,023 |
| q90 | 0,490 | 8,192 | 0,0160 | -0,916 | 107,386 | 0,0230 | 0,0518 | 0,0505 | 0,026 |

Note: This table provides Monte Carlo simulation results for the SVCDEJ model, based on 300 replications. Two settings with different ranges of the argument $u$ are considered. Each panel lists, for each parameter, the true value, the Monte Carlo mean and standard deviation, and the 10th, 50th and 90th Monte Carlo percentiles, respectively. We use $T = 500$ time points with $\Delta t = 1/250$. The initial values are set to $F_0 = 100$ and $v_0 = 0.015$. The threshold for singular values is set to $\bar{s} = 10^{-5}$. The probability of negative jumps is fixed to $p^- = 0.7$. The error terms have an AR(1) structure, i.e., $\epsilon_t$ is an AR(1) process, $\epsilon_t = \rho\epsilon_{t-1} + \tilde{\epsilon}_t$, with $\rho = 0.8$ and i.i.d. $\tilde{\epsilon}_t \sim N(0, 1 - \rho^2)$.

Table D.5: Monte Carlo results for the SVCDEJ model with spatially correlated errors

| parameter | $\sigma$ | $\kappa$ | $\bar{v}$ | $\rho$ | $\delta$ | $\eta^+$ | $\eta^-$ | $\mu_v$ | $\sigma_\varkappa$ |
|---|---|---|---|---|---|---|---|---|---|
| | | | | $u = 1, \ldots, 15$ | | | | | |
| true value | 0,450 | 8,000 | 0,0150 | -0,950 | 100,000 | 0,0200 | 0,0500 | 0,0500 | 0,020 |
| mean | 0,459 | 7,936 | 0,0153 | -0,945 | 97,303 | 0,0203 | 0,0502 | 0,0501 | 0,031 |
| std | 0,020 | 0,277 | 0,0007 | 0,022 | 6,985 | 0,0018 | 0,0013 | 0,0008 | 0,003 |
| q10 | 0,441 | 7,564 | 0,0148 | -0,967 | 86,923 | 0,0193 | 0,0493 | 0,0491 | 0,029 |
| q50 | 0,453 | 8,004 | 0,0151 | -0,950 | 99,528 | 0,0201 | 0,0498 | 0,0500 | 0,030 |
| q90 | 0,496 | 8,153 | 0,0162 | -0,906 | 102,213 | 0,0218 | 0,0519 | 0,0510 | 0,032 |
| | | | | $u = 1, \ldots, 20$ | | | | | |
| true value | 0,450 | 8,000 | 0,0150 | -0,950 | 100,000 | 0,0200 | 0,0500 | 0,0500 | 0,020 |
| mean | 0,452 | 7,897 | 0,0152 | -0,956 | 99,720 | 0,0222 | 0,0503 | 0,0485 | 0,032 |
| std | 0,023 | 0,332 | 0,0009 | 0,028 | 8,378 | 0,0020 | 0,0020 | 0,0027 | 0,006 |
| q10 | 0,425 | 7,454 | 0,0145 | -0,995 | 89,318 | 0,0207 | 0,0490 | 0,0468 | 0,030 |
| q50 | 0,447 | 7,961 | 0,0149 | -0,959 | 102,023 | 0,0219 | 0,0498 | 0,0484 | 0,031 |
| q90 | 0,491 | 8,174 | 0,0163 | -0,913 | 106,194 | 0,0239 | 0,0523 | 0,0503 | 0,035 |

Note: This table provides Monte Carlo simulation results for the SVCDEJ model, based on 300 replications. Two settings with different ranges of the argument $u$ are considered. Each panel lists, for each parameter, the true value, the Monte Carlo mean and standard deviation, and the 10th, 50th and 90th Monte Carlo percentiles, respectively. We use $T = 500$ time points with $\Delta t = 1/250$. The initial values are set to $F_0 = 100$ and $v_0 = 0.015$. The threshold for singular values is set to $\bar{s} = 10^{-5}$. The probability of negative jumps is fixed to $p^- = 0.7$. The error terms are simulated with a spatial correlation between two adjacent strike prices of $\rho = 0.5$.

Figure D.1: An example of the filtered volatility for the SVCDEJ model
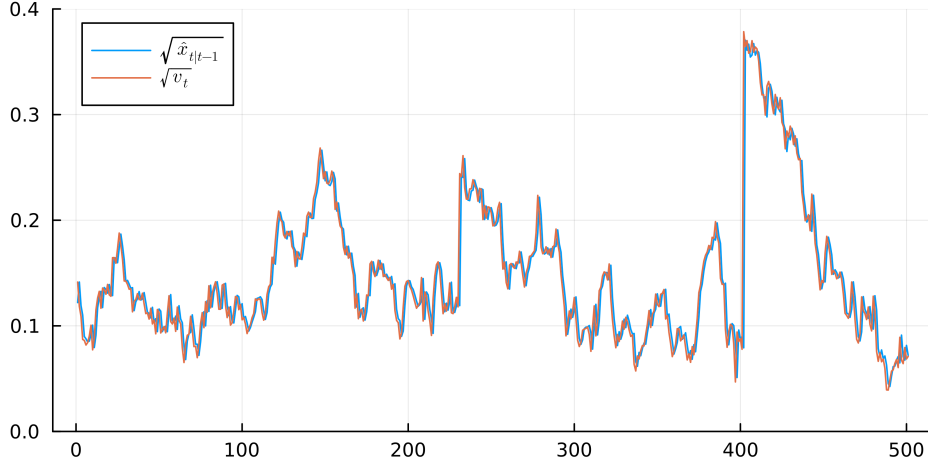


Note: This figure plots the filtered volatility based on the estimated parameters of the SVCDEJ model next to the true stochastic volatility process.

Therefore, in the empirical application in the main text, we focus on the restricted model.

Table D.6: SVCDEJ estimation results

| | $\sigma$ | $\kappa$ | $\bar{v}$ | $\rho$ | $\delta$ | $p^-$ | $\eta^+$ | $\eta^-$ | $\mu_v$ | $\sigma_\varkappa$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | unconstrained | | | | | |
| $\widehat{\theta}$ | 0.5096 | 8.001 | 0.0162 | -1.000 | 157.04 | 0.2651 | 0.0001 | 0.0619 | 0.1230 | 0.3731 |
| s.e. | 0.0192 | 1.461 | 0.0023 | 0.024 | 15.40 | 0.0298 | 0.0000 | 0.0046 | 0.0106 | - |
| | | | | | constrained, $p^- = 0.75$ | | | | | |
| $\widehat{\theta}$ | 0.5083 | 8.703 | 0.0148 | -1.000 | 73.35 | 0.75 | 0.0000 | 0.0586 | 0.1175 | 0.3729 |
| s.e. | 0.2605 | 1.155 | 0.0019 | 0.441 | 1.66 | | 0.0000 | 0.0013 | 0.0233 | - |
| | | | | | constrained, $p^- = 0.65$ | | | | | |
| $\widehat{\theta}$ | 0.5090 | 8.639 | 0.0150 | -1.000 | 72.45 | 0.65 | 0.0002 | 0.0589 | 0.1181 | 0.3729 |
| s.e. | 0.0121 | 0.971 | 0.0012 | 0.014 | 9.79 | | 0.0000 | 0.0027 | 0.0081 | - |

Note: This table provides the parameter estimates and standard errors for the SVCDEJ model. The model is estimated based on $u = 1, \ldots, 15$ and $\bar{s} = 10^{-5}$.

In Table D.7, we also provide additional empirical results for the model specification with an external factor now using either short- or longer-dated options. For short-dated options, we consider the three shortest maturities from our dataset, whereas the longer-dated options consist of the three largest maturities. We find that, although some of the estimated model parameters change considerably, the parameter representing the impact of the external factor on the jump intensity process is of a similar magnitude when estimated from short- or from longer-dated options.

Table D.8 provides empirical results for the alternative model specification SVCEJ detailed in Subsection D.1.3, without and with external state variables. Positive and negative jumps are modeled by separate counting processes with their own jump intensities $\lambda_t^+$ and $\lambda_t^-$, possibly depending on the external state variable with coefficients $\gamma^+$ and $\gamma^-$, respectively. We

Table D.7: SVCDEJ estimation results with Covid-19 reproduction numbers as external factor

| | $\sigma$ | $\kappa$ | $\bar{v}$ | $\rho$ | $\delta$ | $\eta^+$ | $\eta^-$ | $\mu_v$ | $\gamma$ | $q$ | $\sigma_\varkappa$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | short-dated options | | | | | | |
| $\widehat{\theta}$ | 0.776 | 19.132 | 0.0157 | -1.000 | 40.960 | 0.0179 | 0.0593 | 0.200 | 0.895 | 0.0357 | 0.384 |
| s.e. | 0.127 | 7.714 | 0.0038 | 0.121 | 14.735 | 0.0243 | 0.0095 | 0.027 | 0.472 | 0.0332 | - |
| | | | | | longer-dated options | | | | | | |
| $\widehat{\theta}$ | 0.494 | 6.266 | 0.0195 | -0.948 | 28.412 | 0.0304 | 0.0975 | 0.098 | 0.985 | 0.0000 | 0.258 |
| s.e. | 0.116 | 0.864 | 0.0006 | 0.113 | 5.381 | 0.0024 | 0.0035 | 0.016 | 0.098 | 0.0000 | - |

Note: This table provides the parameter estimates and standard errors for the SVCDEJ model with Covid-19 reproduction numbers as external factor using either short-dated ($\tau \in (2, 22)$) or longer-dated ($\tau \in (22, 61)$) options. The model is estimated based on $u = 1, \ldots, 15$ and $\bar{s} = 10^{-5}$, and with $p^- = 0.7$.

Table D.8: SVCEJ estimation results

| | $\sigma$ | $\kappa$ | $\bar{v}$ | $\rho$ | $\delta_0^+$ | $\delta_1^-$ | $\eta^+$ | $\eta^-$ | $\mu_v$ | $\gamma^+$ | $\gamma^-$ | $q$ | $\sigma_\varkappa$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | no external factors | | | | | | | |
| $\widehat{\theta}$ | 0.481 | 8.31 | 0.0139 | -1.00 | 3.76 | 107.3 | 0.0100 | 0.0445 | 0.061 | - | - | - | 0.221 |
| s.e. | 0.007 | 0.26 | 0.0002 | 0.01 | 0.08 | 3.99 | 0.0002 | 0.0006 | 0.002 | | | | 0.004 |
| | | | | | | $R_0$ | | | | | | | |
| $\widehat{\theta}$ | 0.547 | 11.28 | 0.0133 | -1.00 | 0.999 | 83.11 | 0.0150 | 0.0439 | 0.079 | 0.036 | 1.587 | 0.016 | 0.213 |
| s.e. | 0.005 | 0.25 | 0.0003 | 0.01 | 0.045 | 2.97 | 0.0004 | 0.0006 | 0.002 | 0.300 | 0.142 | 0.015 | 0.004 |

Note: This table provides the parameter estimates and standard errors for the SVCEJ model. The model is estimated based on $u = 1, \ldots, 20$ and $\bar{s} = 10^{-7}$.

observe a similar magnitude as in Table 5 (and Table D.6) for most of the parameter estimates, corroborating again the robustness of our empirical results.

## D.2.2 Model diagnostics

The proposed estimation procedure for parametric option pricing models coupled with the derived asymptotic results allow us to obtain several model diagnostics. Although we do not argue that the model considered in Section 6 statistically outperforms competitor models in the literature, the model diagnostics indicate whether the model is, or is far from, a realistic specification.

Figure D.2 illustrates the fit of the model for three CCF arguments, $u = 5$, $10$, and $15$. In particular, it plots the time series of the log option-implied CCF, $\log \widehat{\psi}_t(u, \tau)$, along with its model-implied counterpart, given the parameter estimates in Table 5. We observe that the model-implied CCFs nicely pick up the dynamics of the observed counterparts, yielding a good fit.

Since analyzing the prediction errors individually for each CCF argument is rather cumbersome and impractical, we can consider the fit of the aggregated information, obtained by collapsing the high-dimensional vector $y_t$ into $y_t^*$. Even though this transformation is parameter-dependent, analyzing the fit of $y_t^*$ can be considered to be quite a relevant model diagnostic. Figure D.3 displays the collapsed observation vector $y_t^*$ along with its predicted values obtained

Figure D.2: Real part of the log option-implied and model-implied CCFs



(a) $u = 5$          (b) $u = 10$          (c) $u = 15$
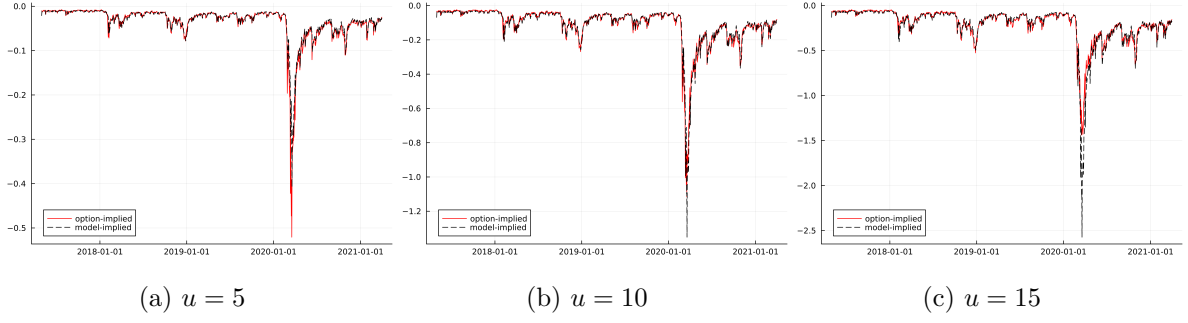
Note: This figure plots the real part of the log option-implied and model-implied CCFs with $u = \{5, 10, 15\}$ averaged accross maturities. The parametric log CCF is based on the SVCDEJ model with parameter estimates in Table 5.

Figure D.3: Fit of the collapsed observation vector $y_t^*$ and standardized residuals



(a) $y_t^*$ and $d_t^* + \hat{x}_{t|t-1}$          (b) $\omega_t^*/\sqrt{G_t^*}$
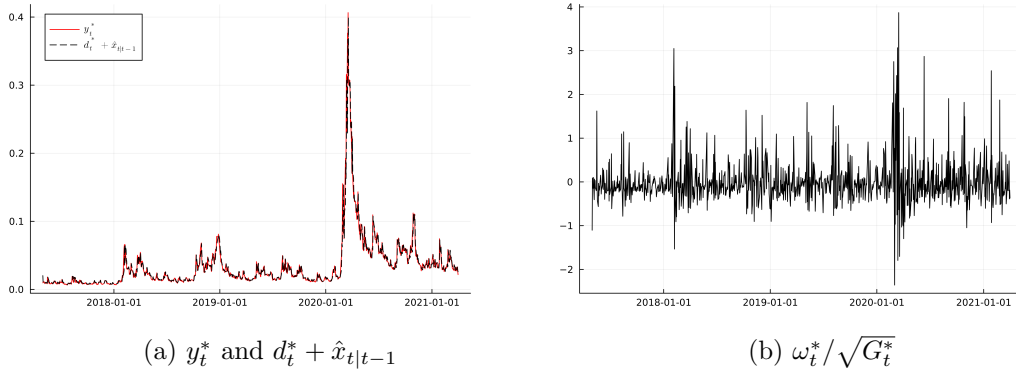
Note: This figure plots the fit of the collapsed observation vector $y_t^*$ and the standardized residuals defined as $\omega_t^*/\sqrt{G_t^*}$ for the SVCDEJ model with parameter estimates in Table 5.

from the Kalman filter and the standardized prediction errors. The latter are commonly used for diagnostic tests in time-series analysis. The figure indicates a good fit of the model, with the standardized residuals being close to white noise. Finally, Figure D.4 plots the ACF of the collapsed vector and the standardized residuals. Notably, the model captures well most of the autocorrelation in the data.

Figure D.4: ACF of the collapsed observation vector $y_t^*$ and the standardized residuals



(a) ACF of $y_t^*$

(b) ACF of $\omega_t^*/\sqrt{G_t^*}$
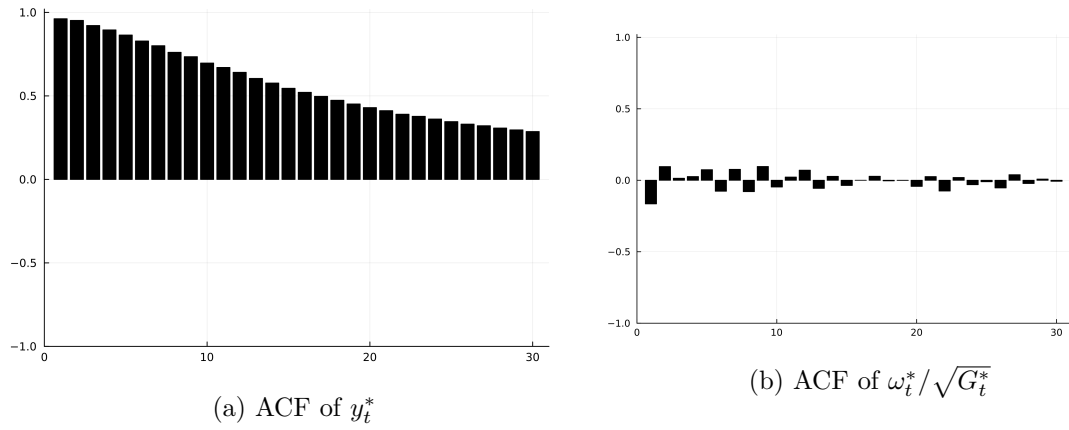
Note: This figure plots the ACF of the collapsed vector and the standardized residuals over 30 lags for the SVCDEJ model with parameter estimates in Table 5.

# References

Aït-Sahalia, Y., Cacho-Diaz, J., & Laeven, R. J. (2015). Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, *117*(3), 585–606.

Aït-Sahalia, Y., & Kimmel, R. L. (2010). Estimating affine multifactor termstructure models using closed-form likelihood expansions. *Journal of Financial Economics*, *98*, 113–144.

Aït-Sahalia, Y., Li, C., & Li, C. X. (2024). Maximum likelihood estimation of latent markov models using closed-form approximations. *Journal of Econometrics*, *240*(2), 105008.

Andersen, L., & Piterbarg, V. (2007). Moment explosions in stochastic volatility models. *Finance and Stochastics*, *11*(1), 29–50.

Andersen, T. G., Benzoni, L., & Lund, J. (2002). An empirical investigation of continuous-time equity return models. *The Journal of Finance*, *57*(3), 1239–1284.

Andersen, T. G., & Bondarenko, O. (2007). Construction and interpretation of model-free implied volatility. *NBER Working Paper No. 13449*.

Andersen, T. G., Fusari, N., & Todorov, V. (2015a). Parametric inference and dynamic state recovery from option panels. *Econometrica*, *83*(3), 1081–1145.

Andersen, T. G., Fusari, N., & Todorov, V. (2015b). The risk premia embedded in index options. *Journal of Financial Economics*, *117*(3), 558–584.

Andersen, T. G., Fusari, N., & Todorov, V. (2017). Short-term market risks implied by weekly options. *The Journal of Finance*, *72*(3), 1335–1386.

Andersen, T. G., Fusari, N., Todorov, V., & Varneskov, R. T. (2021). Spatial dependence in option observation errors. *Econometric Theory*, *37*(2), 205–247.

Arroyo-Marioli, F., Bullano, F., Kucinskas, S., & Rondón-Moreno, C. (2021). Tracking R of COVID-19: A new real-time estimation using the Kalman filter. *PLoS One*, *16*(1), e0244474.

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, *131*(4), 1593–1636.

Bakshi, G., Cao, C., & Chen, Z. (1997). Empirical performance of alternative option pricing models. *The Journal of Finance*, *52*(5), 2003–2049.

Bakshi, G., Kapadia, N., & Madan, D. (2003). Stock return characteristics, skew laws, and the differential pricing of individual equity options. *The Review of Financial Studies*, *16*(1), 101–143.

Bardgett, C., Gourier, E., & Leippold, M. (2019). Inferring volatility dynamics and risk premia from the S&P 500 and VIX markets. *Journal of Financial Economics*, *131*(3), 593–618.

Bates, D. S. (2006). Maximum likelihood estimation of latent affine processes. *The Review of Financial Studies*, *19*(3), 909–965.

Bliss, R. R., & Panigirtzoglou, N. (2002). Testing the stability of implied probability density functions. *Journal of Banking & Finance*, *26*(2-3), 381–422.

Bollerslev, T., & Todorov, V. (2011). Tails, fears, and risk premia. *The Journal of Finance*, *66*(6), 2165–2211.

Boswijk, H. P., Laeven, R. J., & Lalu, A. (2015). Asset returns with self-exciting jumps:

Option pricing and estimation with a continuum of moments. Working paper, University of Amsterdam and Tinbergen Institute.

Boswijk, H. P., Laeven, R. J., Lalu, A., & Vladimirov, E. (2021). Jump contagion among stock market indices: Evidence from option markets. *Tinbergen Institute Discussion Paper 2021-086/III*.

Broadie, M., Chernov, M., & Johannes, M. (2007). Model specification and risk premia: Evidence from futures options. *The Journal of Finance*, *62*(3), 1453–1490.

Carr, P., & Madan, D. (1999). Option valuation using the fast Fourier transform. *Journal of Computational Finance*, *2*(4), 61–73.

Carr, P., & Madan, D. (2001). Optimal positioning in derivative securities. *Quantitative Finance*, *1*(1), 19–37.

Carrasco, M., Chernov, M., Florens, J.-P., & Ghysels, E. (2007). Efficient estimation of general dynamic models with a continuum of moment conditions. *Journal of Econometrics*, *140*(2), 529–573.

Carrasco, M., & Florens, J.-P. (2000). Generalization of GMM to a continuum of moment conditions. *Econometric Theory*, *16*(6), 797–834.

CBOE. (2015). *VIX white paper*. https://cdn.cboe.com/resources/vix/vixwhite.pdf.

Chacko, G., & Viceira, L. M. (2003). Spectral GMM estimation of continuous-time processes. *Journal of Econometrics*, *116*(1-2), 259–292.

Chang, B.-Y., Christoffersen, P., Jacobs, K., & Vainberg, G. (2012). Option-implied measures of equity risk. *Review of Finance*, *16*(2), 385–428.

Cheng, P., & Scaillet, O. (2007). Linear-quadratic jump-diffusion modeling. *Mathematical Finance*, *17*(4), 575–598.

Chernov, M., & Ghysels, E. (2000). A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. *Journal of Financial Economics*, *56*(3), 407–458.

Christoffersen, P., Dorion, C., Jacobs, K., & Karoui, L. (2014). Nonlinear Kalman filtering in affine term structure models. *Management Science*, *60*(9), 2248–2268.

Christoffersen, P., Jacobs, K., & Mimouni, K. (2010). Volatility dynamics for the S&P500: Evidence from realized volatility, daily returns, and option prices. *The Review of Financial Studies*, *23*(8), 3141–3189.

Christoffersen, P., Jacobs, K., & Ornthanalai, C. (2012). Dynamic jump intensities and risk premiums: Evidence from S&P500 returns and options. *Journal of Financial Economics*, *106*(3), 447–472.

Dai, Q., & Singleton, K. J. (2000). Specification analysis of affine term structure models. *The Journal of Finance*, *55*(5), 1943–1978.

de Jong, F. (2000). Time series and cross-section information in affine term-structure models. *Journal of Business & Economic Statistics*, *18*(3), 300–314.

Dini, D. H., & Mandic, D. P. (2012). Class of widely linear complex Kalman filters. *IEEE Transactions on Neural Networks and Learning Systems*, *23*(5), 775–786.

Driessen, J. (2005). Is default event risk priced in corporate bonds? *The Review of Financial*

*Studies*, *18*(1), 165–195.

Du, D., & Luo, D. (2019). The pricing of jump propagation: Evidence from spot and options markets. *Management Science*, *65*(5), 2360–2387.

Duffee, G. R. (1999). Estimating the price of default risk. *The Review of Financial Studies*, *12*(1), 197–226.

Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *The Journal of Finance*, *57*(1), 405–443.

Duffie, D., Filipović, D., & Schachermayer, W. (2003). Affine processes and applications in finance. *The Annals of Applied Probability*, *13*(3), 984–1053.

Duffie, D., & Kan, R. (1996). A yield-factor model of interest rates. *Mathematical Finance*, *6*(4), 379–406.

Duffie, D., Pan, J., & Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, *68*(6), 1343–1376.

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford University Press.

Eraker, B. (2004). Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *The Journal of Finance*, *59*(3), 1367–1403.

Eraker, B., Johannes, M., & Polson, N. (2003). The impact of jumps in volatility and returns. *The Journal of Finance*, *58*(3), 1269–1300.

Fang, F., & Oosterlee, C. W. (2008). A novel pricing method for European options based on Fourier-cosine series expansions. *SIAM Journal on Scientific Computing*, *31*(2), 826–848.

Fengler, M. R. (2009). Arbitrage-free smoothing of the implied volatility surface. *Quantitative Finance*, *9*(4), 417–428.

Feunou, B., & Okou, C. (2018). Risk-neutral moment-based estimation of affine option pricing models. *Journal of Applied Econometrics*, *33*(7), 1007–1025.

Fisher, M., & Gilles, C. (1996). Estimating exponential-affine models of the term structure. *Unpublished working paper. Federal Reserve Bank of Atlanta*.

Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., ... others (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, *584*(7820), 257–261.

Gallant, A. R., & Tauchen, G. (1996). Which moments to match? *Econometric Theory*, 657–681.

Gatheral, J. (2006). *The Volatility Surface*. Wiley Finance.

Gatheral, J., & Jacquier, A. (2014). Arbitrage-free SVI volatility surfaces. *Quantitative Finance*, *14*(1), 59–71.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, *6*(2), 327–343.

Jäckel, P. (2014). Clamping down on arbitrage. *Wilmott*, *2014*(71), 54–69.

Jiang, G. J., & Tian, Y. S. (2005). The model-free implied volatility and its information content. *The Review of Financial Studies*, *18*(4), 1305–1342.

Jiang, G. J., & Tian, Y. S. (2007). Extracting model-free volatility from option prices: An examination of the VIX index. *The Journal of Derivatives*, *14*(3), 35–60.

Johannes, M. S., Polson, N. G., & Stroud, J. R. (2009). Optimal filtering of jump diffusions: Extracting latent states from asset prices. *The Review of Financial Studies*, *22*(7), 2759–2799.

Jungbacker, B., & Koopman, S. J. (2015). Likelihood-based dynamic factor analysis for measurement and forecasting. *The Econometrics Journal*, *18*(2), 1–21.

Keller-Ressel, M. (2011). Moment explosions and long-term behavior of affine stochastic volatility models. *Mathematical Finance*, *21*(1), 73–98.

Keller-Ressel, M., & Mayerhofer, E. (2015). Exponential moments of affine processes. *The Annals of Applied Probability*, *25*(2), 714–752.

Kim, D. H., & Orphanides, A. (2012). Term structure estimation with survey data on interest rate forecasts. *Journal of Financial and Quantitative Analysis*, *47*(1), 241–272.

Kou, S. G. (2002). A jump-diffusion model for option pricing. *Management Science*, *48*(8), 1086–1101.

Lee, R. W. (2004). The moment formula for implied volatility at extreme strikes. *Mathematical Finance*, *14*(3), 469–480.

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

Maier, B. F., & Brockmann, D. (2020). Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science*, *368*(6492), 742–746.

Malz, A. M. (2014). A simple and reliable way to compute option-based risk-neutral distributions. *FRB of New York Staff Report*(677).

Monfort, A., Pegoraro, F., Renne, J.-P., & Roussellet, G. (2017). Staying at zero with affine processes: An application to term structure modelling. *Journal of Econometrics*, *201*(2), 348–366.

Pan, J. (2002). The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Financial Economics*, *63*(1), 3–50.

Piazzesi, M. (2010). Affine term structure models. In *Handbook of Financial Econometrics: Tools and Techniques* (pp. 691–766). Elsevier.

Prem, K., Liu, Y., Russell, T. W., Kucharski, A. J., Eggo, R. M., Davies, N., . . . others (2020). The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: A modelling study. *The Lancet Public Health*, *5*(5), e261–e270.

Qin, L., & Todorov, V. (2019). Nonparametric implied Lévy densities. *The Annals of Statistics*, *47*(2), 1025–1060.

Saikkonen, P. (1995). Problems with the asymptotic theory of maximum likelihood estimation in integrated and cointegrated systems. *Econometric Theory*, *11*(5), 888–911.

Santa-Clara, P., & Yan, S. (2010). Crashes, volatility, and the equity premium: Lessons from S&P 500 options. *The Review of Economics and Statistics*, *92*(2), 435–451.

Singleton, K. J. (2001). Estimation of affine asset pricing models using the empirical charac-

teristic function. *Journal of Econometrics*, *102*(1), 111–141.

Singleton, K. J. (2006). *Empirical Dynamic Asset Pricing: Model Specification and Econometric Assessment*. Princeton University Press.

Todorov, V. (2019). Nonparametric spot volatility from options. *The Annals of Applied Probability*, *29*(6), 3590–3636.

Wu, C.-F. (1981). Asymptotic theory of nonlinear least squares estimation. *The Annals of Statistics*, *9*(3), 501–513.