

# Supplementary Material to “Estimating Option Pricing Models Using a Characteristic Function-Based Linear State Space Representation”

H. Peter Boswijk<sup>a,b</sup>, Roger J. A. Laeven<sup>a,c,d</sup>, Evgenii Vladimirov<sup>a,b</sup>

<sup>a</sup> Amsterdam School of Economics, University of Amsterdam, 1018 WB Amsterdam, The Netherlands

<sup>b</sup> Tinbergen Institute, 1082 MS Amsterdam, The Netherlands

<sup>c</sup> EURANDOM, 5600 MB Eindhoven, The Netherlands

<sup>d</sup> CentER, Tilburg University, 5000 LE Tilburg, The Netherlands

June 3, 2024

## Abstract

This supplementary material contains four appendices to the paper “Estimating Option Pricing Models Using a Characteristic Function-Based Linear State Space Representation”. For context, notation and definitions, see the paper. This supplement provides details on (i) the proof 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.

## 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

$$\begin{aligned}
 \zeta_t^\phi(u, \tau) &:= \widehat{\phi}_t(u, \tau) - \phi_t(u, \tau) \\
 &= -u_t \sum_{j=2}^n e^{(iu-1)m_j} \cdot \widehat{O}_t(\tau, m_j) \Delta m_j + u_t \int_{-\infty}^{\infty} e^{(iu-1)m} \cdot O_t(\tau, m) dm \\
 &= -u_t \left[ \sum_{j=2}^n e^{(iu-1)m_j} \cdot O_t(\tau, m_j) \Delta m_j + \sum_{j=2}^n e^{(iu-1)m_j} \cdot \zeta_t(\tau, m_j) \Delta m_j \right] \\
 &\quad + u_t \int_{-\infty}^{\infty} e^{(iu-1)m} \cdot O_t(\tau, m) dm,
 \end{aligned}$$

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

$$\begin{aligned}
\zeta_t^\phi(u, \tau) &= -u_t \underbrace{\sum_{j=2}^n e^{(iu-1)m_j} \cdot \zeta_t(\tau, m_j) \Delta m_j}_{=:\zeta_t^{(1)}(u, \tau)} \\
&\quad + \underbrace{u_t \int_{-\infty}^{m_1} e^{(iu-1)m} \cdot O_t(\tau, m) dm + u_t \int_{m_n}^{\infty} e^{(iu-1)m} \cdot O_t(\tau, m) dm}_{=:\zeta_t^{(2)}(u, \tau)} \\
&\quad + \underbrace{u_t \sum_{j=2}^n \int_{m_{j-1}}^{m_j} \left[ e^{(iu-1)m} \cdot O_t(\tau, m) - e^{(iu-1)m_j} \cdot O_t(\tau, m_j) \right] dm}_{=:\zeta_t^{(3)}(u, \tau)} \\
&= \zeta_t^{(1)}(u, \tau) + \zeta_t^{(2)}(u, \tau) + \zeta_t^{(3)}(u, \tau). \tag{A.1}
\end{aligned}$$

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*

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

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

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

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

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

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  $\bar{m} = \max_{1 \leq j \leq n} m_j$ , and the corresponding strike prices by  $\underline{K}$  and  $\bar{K}$ . 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  $\bar{K} \asymp n^{\bar{\alpha}}$  with  $\underline{\alpha} > 0$  and  $\bar{\alpha} > 0$ . Then, as  $n \rightarrow \infty$ , we have

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

and

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

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

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

$$\begin{aligned} \left| \frac{1}{F_t} \int_{\bar{m}}^{\infty} e^{(iu-1)m} \cdot O_t(\tau, m) dm \right| &\leq \int_{\bar{m}}^{\infty} \left| e^{(iu-1)m} \right| \cdot \left| \frac{O_t(\tau, m)}{F_t} \right| dm \leq C_t e^{-(1+\bar{\delta})\bar{m}}, \\ \left| \frac{1}{F_t} \int_{-\infty}^{\underline{m}} e^{(iu-1)m} \cdot O_t(\tau, m) dm \right| &\leq \int_{-\infty}^{\underline{m}} \left| e^{(iu-1)m} \right| \cdot \left| \frac{O_t(\tau, m)}{F_t} \right| dm \leq C_t e^{\underline{\delta}\underline{m}}, \end{aligned}$$

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

$$\begin{aligned} \zeta_t^{(2)}(u, \tau) &= u_t \int_{-\infty}^{\underline{m}} e^{(iu-1)m} \cdot O_t(\tau, m) dm + u_t \int_{\bar{m}}^{\infty} e^{(iu-1)m} \cdot O_t(\tau, m) dm \\ &= \mathcal{O}_p\left(e^{-\underline{\delta}|\underline{m}|\right) + \mathcal{O}_p\left(e^{-(1+\bar{\delta})|\bar{m}|\right) = \mathcal{O}_p\left(e^{-\underline{\delta}|\underline{m}| \wedge (1+\bar{\delta})|\bar{m}|\right) = \mathcal{O}_p\left(n^{-(\underline{\delta}\alpha \wedge (1+\bar{\delta})\bar{\alpha})}\right). \end{aligned}$$

For the discretization errors, we use the following decomposition:

$$\begin{aligned} &\int_{m_{j-1}}^{m_j} \left[ e^{(iu-1)m} \cdot O_t(\tau, m) - e^{(iu-1)m_j} \cdot O_t(\tau, m_j) \right] dm \\ &= \int_{m_{j-1}}^{m_j} \left[ \left( e^{(iu-1)m} - e^{(iu-1)m_j} \right) \cdot O_t(\tau, m_j) + e^{(iu-1)m} (O_t(\tau, m) - O_t(\tau, m_j)) \right] dm. \end{aligned}$$

By applying the mean value theorem twice, we have

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

and

$$\begin{aligned} \left| e^{(iu-1)m} (O_t(\tau, m) - O_t(\tau, m_j)) \right| &\leq e^{-m_{j-1}} \left| \frac{\partial O_t(\tau, m)}{\partial m} \right|_{m=\tilde{m}} \left| \Delta m_j \right| \\ &\leq e^{-m_{j-1}} \left| \frac{\partial O_t(\tau, m)}{\partial K} \right|_{K=\tilde{K}} \left| e^{\tilde{m}} F_t \Delta m_j \right| \\ &\leq \left| \frac{\partial O_t(\tau, m)}{\partial K} \right|_{K=\tilde{K}} \left| e^{\Delta m_j} F_t \Delta m_j \right|, \end{aligned}$$

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  $m$ ,

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

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  $m > 0$ ,

$$\begin{aligned} \left| \frac{\partial O_t(\tau, m)}{\partial K} \right| &= e^{-r\tau} \mathbb{Q}(F_{t+\tau} > K) = e^{-r\tau} \mathbb{Q}(f_{t+\tau} > e^m) \\ &\leq e^{-r\tau} \mathbb{E}^{\mathbb{Q}}[f_{t+\tau}^{\bar{\delta}+1} | \mathcal{F}_t] e^{-(\bar{\delta}+1)m}, \end{aligned}$$

and, for  $m < 0$ ,

$$\begin{aligned} \left| \frac{\partial O_t(\tau, m)}{\partial K} \right| &= e^{-r\tau} \mathbb{Q}(F_{t+\tau} < K) = e^{-r\tau} \mathbb{Q}(f_{t+\tau}^{-1} > e^{-m}) \\ &\leq e^{-r\tau} \mathbb{E}^{\mathbb{Q}}[f_{t+\tau}^{-\underline{\delta}} | \mathcal{F}_t] e^{\underline{\delta}m}. \end{aligned}$$

Therefore,

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

Combining all these inequalities together, we obtain

$$\begin{aligned} &\left| \frac{1}{F_t} \int_{m_{j-1}}^{m_j} \left[ e^{(iu-1)m} \cdot O_t(\tau, m) - e^{(iu-1)m_j} \cdot O_t(\tau, m_j) \right] dm \right| \\ &= \left| \frac{1}{F_t} \int_{m_{j-1}}^{m_j} \left[ \left( e^{(iu-1)m} - e^{(iu-1)m_j} \right) \cdot O_t(\tau, m_j) + e^{(iu-1)m} (O_t(\tau, m) - O_t(\tau, m_j)) \right] dm \right| \\ &\leq \left( C_t (|u| \vee 1) \Delta m_j e^{-m_{j-1}} e^{-[\bar{\delta}m_{j-1} \vee -(1+\underline{\delta})m_{j-1}]} + C_t e^{\Delta m} e^{-[(\bar{\delta}+1)m_{j-1} \vee -\underline{\delta}m_{j-1}]} \Delta m_j \right) \Delta m_j \\ &\leq C_t e^{-[(\bar{\delta}+1)m_{j-1} \vee -\underline{\delta}m_{j-1}]} (\Delta m_j)^2. \end{aligned}$$

Then, for fixed  $\bar{m}$  and  $\underline{m}$ , and  $\Delta m \rightarrow 0$ , we have

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

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

$$\begin{aligned} &\left| \sum_{j=2}^n \frac{1}{F_t} \int_{m_{j-1}}^{m_j} \left[ e^{(iu-1)m} \cdot O_t(\tau, m) - e^{(iu-1)m_j} \cdot O_t(\tau, m_j) \right] dm \right| \\ &\leq C_t \sum_{j=2}^n e^{-[(\bar{\delta}+1)m_{j-1} \vee -\underline{\delta}m_{j-1}]} (\Delta m_j)^2 \\ &\leq C_t \Delta m \sum_{j=2}^n e^{-[(\bar{\delta}+1)m_{j-1} \vee -\underline{\delta}m_{j-1}]} \Delta m_j. \end{aligned}$$

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

$$\begin{aligned} \sum_{j=2}^n e^{-[(\bar{\delta}+1)m_{j-1} \vee -\underline{\delta}m_{j-1}]} \Delta m_j &\longrightarrow \int_{\underline{m}}^{\bar{m}} e^{-[(\bar{\delta}+1)m \vee -\underline{\delta}m]} dm = \int_{\underline{m}}^0 e^{\underline{\delta}m} dm + \int_0^{\bar{m}} e^{-(\bar{\delta}+1)m} dm \\ &= \frac{1}{\underline{\delta}} \left(1 - e^{-\underline{\delta}\underline{\alpha}}\right) + \frac{1}{\bar{\delta}+1} \left(1 - e^{-(\bar{\delta}+1)\bar{\alpha}}\right) \\ &= \mathcal{O}_p(1) \end{aligned}$$

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

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

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

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

□

## 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

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

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

$$\begin{aligned} \nu_t(\tau, m) &= F_t \sqrt{\tau} \varphi(d_+), \\ d_+ &= -m\varpi^{-1/2}(m) + \frac{1}{2}\varpi^{1/2}(m), \end{aligned}$$

where  $\varphi(x)$  is the standard normal pdf and  $\varpi(m) := \kappa^2(\tau, m)\tau$  is the total implied variance.<sup>1</sup> Hence,

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

---

<sup>1</sup>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.

Therefore, given Assumption 2, we obtain

$$\begin{aligned}
\mathbb{E} \left[ \left| \zeta_t^{(1)}(u, \tau) \right|^2 \middle| \mathcal{F}_t \right] &\leq |u_t|^2 \sum_{j=2}^n e^{-2m_j} \cdot \mathbb{E} \left[ \zeta_t(\tau, m_j)^2 \middle| \mathcal{F}_t \right] (\Delta m_j)^2 \\
&\leq |u_t|^2 \sum_{j=2}^n e^{-2m_j} \cdot \sigma_{\varkappa}^2 \kappa^2(\tau, m_j) F_t^2 \tau \frac{1}{2\pi} e^{-(\varpi_j^{-1} m_j^2 - m_j + \frac{1}{4} \varpi_j)} (\Delta m_j)^2 \\
&\leq C_t \Delta m \sum_{j=2}^n \varpi_j e^{-\varpi_j^{-1} m_j^2 - m_j - \frac{1}{4} \varpi_j} \Delta m_j \\
&\leq C_t \Delta m \sum_{j=2}^n \varpi_j e^{-d_-^2(m_j)} \Delta m_j,
\end{aligned}$$

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

$$\int_{-\infty}^{\infty} \varpi(m) \exp(-d_-^2(m)) dm =: \int_{-\infty}^{\infty} h(m) dm,$$

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):

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

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  $m$ ,

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

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  $\zeta_t^{(1)}(u, \tau) = \mathcal{O}_p(\sqrt{\Delta m}) = \mathcal{O}_p(\sqrt{n^{-1} \log n})$ .

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,

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

and

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

for any  $u_i, u_j \in \mathcal{U}$ , where  $\bar{z}$  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  $\sigma_\varepsilon$  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

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

where

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

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  $\xi_{t,\tau}^{(1)} = \left(\xi_t^{(1)}(u_1, \tau), \dots, \xi_t^{(1)}(u_q, \tau)\right)'$  for a fixed time  $t$  and time-to-maturity  $\tau$ . They are given by

$$\begin{aligned}\Gamma_{t,\tau} &:= \mathbb{E} \left[ \xi_{t,\tau}^{(1)} \overline{\xi_{t,\tau}^{(1)}}' \right] = \left[ \mathbb{E}[\xi_t^{(1)}(u_i, \tau) \xi_t^{(1)}(-u_j, \tau)] \right]_{1 \leq i, j \leq q} \\ &= \sigma_\varepsilon^2 \cdot \left[ \frac{\gamma_t(u_i, u_j, \tau)}{\phi_t(u_i, \tau) \phi_t(-u_j, \tau)} \right]_{1 \leq i, j \leq q} =: \sigma_\varepsilon^2 \cdot \tilde{\Gamma}_{t,\tau}, \\ C_{t,\tau} &:= \mathbb{E} \left[ \xi_{t,\tau}^{(1)} \xi_{t,\tau}^{(1)'} \right] = \left[ \mathbb{E}[\xi_t^{(1)}(u_i, \tau) \xi_t^{(1)}(u_j, \tau)] \right]_{1 \leq i, j \leq q} \\ &= \sigma_\varepsilon^2 \cdot \left[ \frac{c_t(u_i, u_j, \tau)}{\phi_t(u_i, \tau) \phi_t(u_j, \tau)} \right]_{1 \leq i, j \leq q} =: \sigma_\varepsilon^2 \cdot \tilde{C}_{t,\tau}.\end{aligned}$$

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:

$$\begin{aligned}H_{t,\tau} &:= \text{Var} \left[ \begin{pmatrix} \Re(\xi_{t,\tau}^{(1)}) \\ \Im(\xi_{t,\tau}^{(1)}) \end{pmatrix} \right] = \begin{pmatrix} \frac{1}{2} \Re(\Gamma_{t,\tau} + C_{t,\tau}) & \frac{1}{2} \Im(-\Gamma_{t,\tau} + C_{t,\tau}) \\ \frac{1}{2} \Im(\Gamma_{t,\tau} + C_{t,\tau}) & \frac{1}{2} \Re(\Gamma_{t,\tau} - C_{t,\tau}) \end{pmatrix} \\ &= \sigma_\varepsilon^2 \cdot \begin{pmatrix} \frac{1}{2} \Re(\tilde{\Gamma}_{t,\tau} + \tilde{C}_{t,\tau}) & \frac{1}{2} \Im(-\tilde{\Gamma}_{t,\tau} + \tilde{C}_{t,\tau}) \\ \frac{1}{2} \Im(\tilde{\Gamma}_{t,\tau} + \tilde{C}_{t,\tau}) & \frac{1}{2} \Re(\tilde{\Gamma}_{t,\tau} - \tilde{C}_{t,\tau}) \end{pmatrix} \\ &=: \sigma_\varepsilon^2 \cdot \tilde{H}_{t,\tau}.\end{aligned}$$

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:  $H_t = \text{blkdiag}\{H_{t,1}, \dots, H_{t,k}\}$ .

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  $s \neq t = 1, \dots, T$ .

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  $s, t = 1, \dots, T$ . Thus, the proof is established.  $\square$

### A.3 Proof of Proposition 2

#### Setup and notation

Given a parameter value  $\vartheta = (\theta', \sigma_\varepsilon^2)'$ , the true state  $x_t$  can be estimated by the filtered and predicted states as  $\hat{x}_{t|t}(\vartheta)$  and  $\hat{x}_{t|t-1}(\vartheta)$ , 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  $n \rightarrow \infty$ , to the true state when evaluated at the true parameter vector, i.e.,  $\text{plim}_{n \rightarrow \infty} x_t(\theta_0) = x_t$ .

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  $H_t = \mathcal{O}_p(n^{-1} \log n)$ , i.e., it converges to zero as  $n \rightarrow \infty$ . However, the scaled version of  $H_t$  has a non-zero limit. In particular, for the components of the covariance matrix we have for  $k, l = 1, \dots, q$ :

$$\begin{aligned} (\Delta m)^{-1}(\tilde{\Gamma}_{t,i})_{kl} &= \frac{u_{k,t}\bar{u}_{l,t}}{\phi_t(u_k, \tau_i)\phi_t(-u_l, \tau_i)} \sum_{j=2}^n e^{i(u_k - u_l - 2)m_j} \kappa_t^2(\tau_i, m_j) \nu_t^2(\tau_i, m_j) \frac{(\Delta m_j)^2}{\Delta m} \\ &\rightarrow \frac{u_{k,t}\bar{u}_{l,t}}{\phi_t(u_k, \tau_i)\phi_t(-u_l, \tau_i)} \int_{\mathbb{R}} e^{i(u_k - u_l - 2)m} \kappa_t^2(\tau_i, m) \nu_t^2(\tau_i, m) dm =: (\tilde{\mathcal{G}}_{t,i})_{kl} \end{aligned}$$

and

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

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

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

Hence, for the projection matrix we have

$$\begin{aligned} A_t^* &= (Z_t' H_t^- Z_t)^{-1} Z_t' H_t^- = (Z_t' \tilde{H}_t^- Z_t)^{-1} Z_t' \tilde{H}_t^- \\ &= (Z_t' \Delta m \tilde{H}_t^- Z_t)^{-1} Z_t' \Delta m \tilde{H}_t^- \\ &\xrightarrow[n \rightarrow \infty]{p} (Z_t' \tilde{\mathcal{H}}_t^- Z_t)^{-1} Z_t' \tilde{\mathcal{H}}_t^- =: \mathcal{A}_t^*. \end{aligned}$$

Recall that the components of the quasi-likelihood contributions for the collapsed state space representation, see (27), are given by  $l_t^*(\theta, \sigma_{\varepsilon}^2)$ ,  $l_t^+(\theta_1, \sigma_{\varepsilon}^2)$ , and  $J_t(\theta_1)$ , and the average log-likelihoods by  $\bar{l}_T(\vartheta) = l(Y_t; \vartheta)/T$  and analogously  $\bar{l}_T^*(\theta, \sigma_{\varepsilon}^2)$ ,  $\bar{l}_T^+(\theta_1, \sigma_{\varepsilon}^2)$ ,  $\bar{J}_T(\theta_1)$ ,  $\bar{\ell}_T(\vartheta)$  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  $\mathcal{O}_p(n^{-1} \log n)$  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  $\tilde{H}_t^*(\theta)$ ,  $T_t(\theta)$ ,  $c_t(\theta)$ ,  $Q_t(x_t, \theta)$  and  $Z_t(\theta)$  with respect to the parameter vector  $\theta \in \Theta$ , we will show that as  $n \rightarrow \infty$ , 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  $H_t^*(\vartheta) = \sigma_{\varepsilon}^2 A_t^*(\theta) \tilde{H}_t A_t^{*'}(\theta)$ . Since  $\tilde{H}_t = R_n$  and  $A_t^*(\theta) \rightarrow_p \mathcal{A}_t^*(\theta)$  as  $n \rightarrow \infty$ , 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

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

Furthermore,

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

and

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

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

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

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  $|l_t^*(\vartheta) - \ell_t^*(\theta)| = R_n \rightarrow_p 0$  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  $i^{\text{th}}$  component of the parameter vector:

$$\begin{aligned} \frac{\partial H_t^*(\vartheta)}{\partial \theta_i} &= \frac{\partial A_t^*(\vartheta) H_t A_t^{*'}(\vartheta)}{\partial \theta_i} = 2A_t^*(\vartheta) H_t \frac{\partial A_t^{*'}(\vartheta)}{\partial \theta_i} = R_n, \\ \frac{\partial G_t^*(\vartheta)^{-1}}{\partial \theta_i} &= \frac{\partial P_{t|t-1}(\vartheta)^{-1}}{\partial \theta_i} + R_n = \frac{\partial Q_{t-1}(x_{t-1}(\theta); \theta)^{-1}}{\partial \theta_i} + R_n, \\ \frac{\partial P_{t|t}(\vartheta)}{\partial \theta_i} &= R_n, \\ \frac{\partial \hat{x}_{t|t}(\vartheta)}{\partial \theta_i} &= \frac{\partial x_t(\theta)}{\partial \theta_i} + R_n, \\ \frac{\partial \omega_t^*(\vartheta)}{\partial \theta_i} &= \frac{\partial \eta_t(\theta)}{\partial \theta_i} + R_n, \\ \frac{\partial G_t^*(\vartheta)}{\partial \theta_i} &= \frac{\partial Q_{t-1}(x_{t-1}; \theta)}{\partial \theta_i} + R_n. \end{aligned}$$

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

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

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

$$\begin{aligned} \frac{\partial^2 H_t^*(\vartheta)}{\partial \theta_i \partial \theta_j} &= \frac{\partial^2 A_t^*(\vartheta) H_t A_t^{*'}(\vartheta)}{\partial \theta_i \partial \theta_j} = R_n, \\ \frac{\partial^2 G_t^*(\vartheta)^{-1}}{\partial \theta_i \partial \theta_j} &= \frac{\partial^2 P_{t|t-1}(\vartheta)^{-1}}{\partial \theta_i \partial \theta_j} + R_n = \frac{\partial^2 Q_{t-1}(x_{t-1}(\theta); \theta)^{-1}}{\partial \theta_i \partial \theta_j} + R_n, \\ \frac{\partial^2 P_{t|t}(\vartheta)}{\partial \theta_i \partial \theta_j} &= R_n, \\ \frac{\partial^2 \hat{x}_{t|t}(\vartheta)}{\partial \theta_i \partial \theta_j} &= \frac{\partial^2 x_t(\theta)}{\partial \theta_i \partial \theta_j} + R_n, \end{aligned}$$

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

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  $n \rightarrow \infty$ ,

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

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

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  $\hat{\theta}_1$  and  $\hat{\theta}_2$ , and implies that we cannot prove consistency via uniform convergence (as  $n, T \rightarrow \infty$ ) 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  $\hat{\vartheta}$ .

### Asymptotic normality

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

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

where  $\bar{\theta}$  lies between  $\hat{\theta}$  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

$$\begin{aligned}\nabla_1 l_t(\vartheta) &= \nabla_1 l_t^*(\theta, \sigma_\varkappa^2) + \nabla_1 l_t^+(\theta_1, \sigma_\varkappa^2) + \nabla_1 J_t(\theta_1), \\ \nabla_2 l_t(\vartheta) &= \nabla_2 l_t^*(\theta, \sigma_\varkappa^2).\end{aligned}$$

Note, in particular, that

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

When evaluated at  $\theta_{1,0}$ ,  $e_t(\theta_{1,0}) = M_{Z_t}(\theta_{1,0})\varepsilon_t = R_n^{1/2}$ , while  $\tilde{H}_t = R_n$  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  $\hat{\sigma}_\varepsilon^2$ ) as well as Assumption 4 (iii), the central limit theorem yields for the first component

$$\begin{aligned}\sqrt{\Delta m}\sqrt{T} \nabla_1 \bar{l}_T(\theta_0, \hat{\sigma}_\varepsilon^2) &= \sqrt{\Delta m}\sqrt{T} (\nabla_1 \bar{l}_T^*(\theta_0, \hat{\sigma}_\varepsilon^2) + \nabla_1 \bar{l}_T^+(\theta_{1,0}, \hat{\sigma}_\varepsilon^2) + \nabla_1 \bar{J}_T(\theta_{1,0})) \\ &= \sqrt{\Delta m}\sqrt{T} \nabla_1 \bar{l}_T^+(\theta_{1,0}, \sigma_{\varepsilon,0}^2) + o_p(1) \\ &\xrightarrow{d} N(0, \mathcal{I}_1),\end{aligned}$$

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

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

$$\begin{aligned}\sqrt{T} \nabla_2 \bar{l}_T(\theta_0, \hat{\sigma}_\varepsilon^2) &= \sqrt{T} \nabla_2 \bar{l}_T^*(\theta_0, \hat{\sigma}_\varepsilon^2) \\ &= \sqrt{T} \nabla_2 \bar{\ell}_T^*(\theta_0) + o_p(1) \\ &\xrightarrow{d} N(0, \mathcal{I}_2),\end{aligned}$$

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

As for the second-order derivatives, we have

$$\begin{aligned}\nabla_2^2 l_t(\theta, \sigma_\varepsilon^2) &= \nabla_2^2 l_t^*(\theta, \sigma_\varepsilon^2) = \nabla_2^2 \ell_t^*(\theta) + o_p(1), \\ \nabla_1 \nabla_2 l_t(\theta, \sigma_\varepsilon^2) &= \nabla_1 \nabla_2 l_t^*(\theta, \sigma_\varepsilon^2) = \nabla_1 \nabla_2 \ell_t^*(\theta) + o_p(1), \\ \nabla_1^2 l_t(\theta, \sigma_\varepsilon^2) &= \nabla_1^2 l_t^*(\theta, \sigma_\varepsilon^2) + \nabla_1^2 l_t^+(\theta_1, \sigma_\varepsilon^2) + \nabla_1^2 J_t^+(\theta_1),\end{aligned}$$

with

$$\begin{aligned}\nabla_1^2 l_t^+(\theta_1, \sigma_\varepsilon^2) &= -\frac{1}{2\sigma_\varepsilon^2} \frac{\partial^2 e_t(\theta_1)' \tilde{H}_t^- e_t(\theta_1)}{\partial \theta_1 \partial \theta_1'} \\ &= -\frac{1}{\sigma_\varepsilon^2} \left( \frac{\partial e_t(\theta_1)'}{\partial \theta_1} \tilde{H}_t^- \frac{\partial e_t(\theta_1)}{\partial \theta_1'} + (e_t(\theta_1)' \tilde{H}_t^- \otimes I_p) \frac{\partial \text{vec}(\partial e_t(\theta_1)' / \partial \theta_1)}{\partial \theta_1'} \right).\end{aligned}$$

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

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

Therefore,

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

This proves the proposition.  $\square$

## 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:

$$dx_t = \mu(x_t)dt + \sigma(x_t)dW_t + J_t dN_t, \quad (\text{B.1})$$

with  $\mu(x) = k_0 + k_1x$ ,  $\sigma^2(x) = h_0 + h_1x$ ,  $\lambda(x) = l_0 + l_1x$ , 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, Pan, and Singleton (2000). For notational simplicity, let

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

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

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

Dynkin's formula yields that

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

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

$$\begin{aligned} \mathbb{E}[x_T|\mathcal{F}_t] &= x_t + \mathbb{E}\left[\int_t^T \left(\mu(x_s) + \lambda(x_s) \int_{\mathbb{R}} z d\nu(z)\right) ds|\mathcal{F}_t\right] \\ &= x_t + \mathbb{E}\left[\int_t^T (k_0 + k_1x_s + (l_0 + l_1x_s)\mu_J) ds|\mathcal{F}_t\right] \\ &= x_t + \int_t^T (k_0 + l_0\mu_J + (k_1 + l_1\mu_J)\mathbb{E}[x_s|\mathcal{F}_t]) ds \\ &= x_t + \int_t^T (g_0 + g_1\mathbb{E}[x_s|\mathcal{F}_t]) ds, \end{aligned}$$

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

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

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

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

Next, we are interested in deriving the conditional variance:

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

Note that

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

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

$$\begin{aligned} dm_t(T) &= \left[ -g_1 e^{g_1(T-t)}x_t - g_0 e^{g_1(T-t)} \right] dt + e^{g_1(T-t)}(\mu(x_t)dt + \sigma(x_t)dW_t) + e^{g_1(T-t)}J_t dN_t \\ &= e^{g_1(T-t)} \left[ -(l_0 + l_1 x_t)\mu_J dt + \sigma(x_t)dW_t + J_t dN_t \right]. \end{aligned}$$

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:

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

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:

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

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

$$\begin{aligned} \text{Var}(x_T|\mathcal{F}_t) &= -\frac{1}{2g_1} (h_0 + l_0 \mu_{J2}) \left( 1 - e^{2g_1(T-t)} \right) \\ &\quad - \frac{1}{2g_1^2} (h_1 + l_1 \mu_{J2}) \left[ 2g_1 \left( e^{g_1(T-t)} - e^{2g_1(T-t)} \right) x_t - g_0 \left( 1 - e^{g_1(T-t)} \right)^2 \right]. \quad (\text{B.3}) \end{aligned}$$

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.

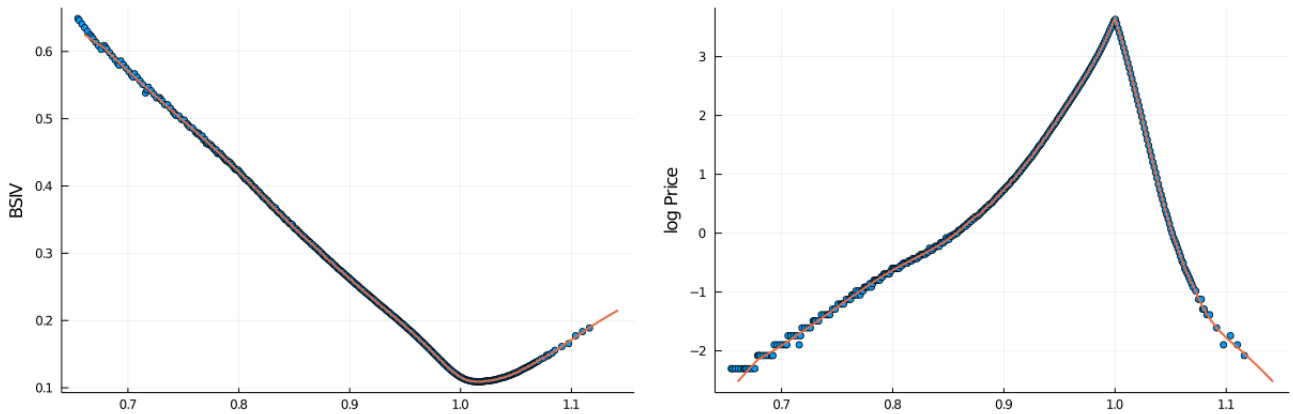
### C.1 Interpolation-extrapolation scheme

#### 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.

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.

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.<sup>2</sup>

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  $i = n^* - 1, \dots, 2$  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 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,

<sup>2</sup>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.

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 Panigirtzoglou (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

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

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

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  $F_\tau$ .<sup>3</sup> 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

---

<sup>3</sup>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.

following no-arbitrage bounds for the right-tail slope  $\bar{\beta}$  (our detailed derivations are available upon request; they are suppressed to save space):

$$0 \leq \bar{\beta} < \min(\beta_{max}, 2),$$

where

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

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

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

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

where

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

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

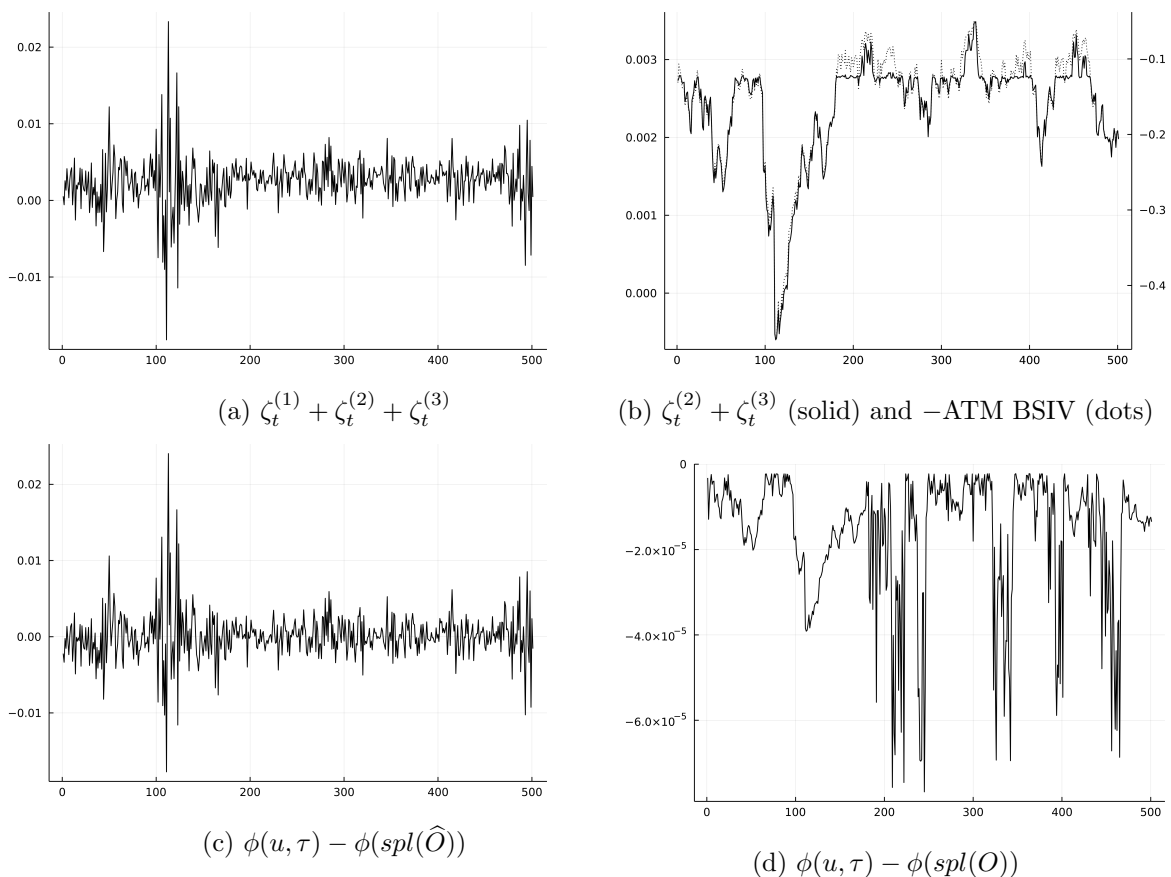
## C.2 CCF replication

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 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.

Panel (a) of Figure C.2 plots the measurement errors when we use a finite set of observed option prices  $\hat{O}_t(\tau, m)$ . 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

Figure C.2: The three types of measurement errors



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.

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:

$$d \log F_t = \left(-\frac{1}{2}v_t - \mu\lambda_t\right)dt + \sqrt{v_t}dW_{1,t} + J_t dN_t, \quad (\text{D.1})$$

$$dv_t = \kappa(\bar{v} - v_t)dt + \sigma\sqrt{v_t}dW_{2,t} + J_t^v dN_t, \quad (\text{D.2})$$

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

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

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

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:

$$v_{t+1} = c_t + T_t v_t + \eta_{t+1}, \quad (\text{D.5})$$

where  $c_t = \frac{g_0}{g_1}(e^{g_1\Delta t} - 1)$ ,  $T_t = e^{g_1\Delta t}$  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  $\kappa > \delta\mu_v$ .

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

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

parameter	$\sigma$	$\kappa$	$\bar{v}$	$\rho$	$\delta$	$\mu_J$	$\sigma_J$	$\mu_v$	$\sigma_x$
$u = 1, \dots, 15$									
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
$u = 1, \dots, 20$									
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
$u = 1, \dots, 25$									
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  $\bar{s} = 10^{-7}$ .

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 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

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  $u = 1, \dots, 20$  and the threshold to  $\bar{s} = 10^{-7}$ . The initial values are set to  $F_0 = 100$  and  $v_0 = 0.02$ .

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 (2015).

### 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, Gourier, and Leippold (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:

$$d \log F_t = \left(-\frac{1}{2}v_t - \mu^- \lambda_t^+ - \mu^+ \lambda_t^-\right)dt + \sqrt{v_t}dW_{1,t} + J_t^- dN_t^- + J_t^+ dN_t^+, \quad (\text{D.6})$$

$$dv_t = \kappa(\bar{v} - v_t)dt + \sigma\sqrt{v_t}dW_{2,t} + J_t^v dN_t^-, \quad (\text{D.7})$$

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  $\delta_0^- = 0$ .

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

parameter	$\sigma$	$\kappa$	$\bar{v}$	$\rho$	$\delta_0^+$	$\delta_1^-$	$\eta^+$	$\eta^-$	$\mu_v$	$\sigma_\varepsilon$
$u = 1, \dots, 15$										
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
$u = 1, \dots, 20$										
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
$u = 1, \dots, 25$										
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  $\bar{s} = 10^{-7}$ .

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 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.

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

parameter	$\sigma$	$\kappa$	$\bar{v}$	$\rho$	$\delta$	$\eta^+$	$\eta^-$	$\mu_v$	$\sigma_\varkappa$
$u = 1, \dots, 15$									
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
$u = 1, \dots, 20$									
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.  $\tilde{\epsilon}_t \sim N(0, 1 - \rho^2)$ .

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

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

parameter	$\sigma$	$\kappa$	$\bar{v}$	$\rho$	$\delta$	$\eta^+$	$\eta^-$	$\mu_v$	$\sigma_x$
$u = 1, \dots, 15$									
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
$u = 1, \dots, 20$									
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  $\rho = 0.5$ .

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

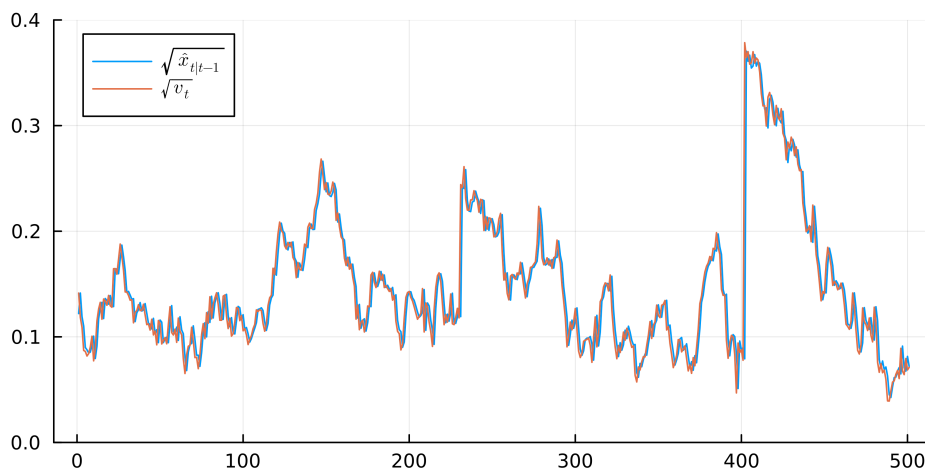
Table D.6: SVCDEJ estimation results

	$\sigma$	$\kappa$	$\bar{v}$	$\rho$	$\delta$	$p^-$	$\eta^+$	$\eta^-$	$\mu_v$	$\sigma_x$
unconstrained										
$\hat{\theta}$	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, $p^- = 0.75$										
$\hat{\theta}$	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, $p^- = 0.65$										
$\hat{\theta}$	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  $u = 1, \dots, 15$  and  $\bar{s} = 10^{-5}$ .

In Table D.7, we also provide additional empirical results for the model specification with

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.

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

	$\sigma$	$\kappa$	$\bar{v}$	$\rho$	$\delta$	$\eta^+$	$\eta^-$	$\mu_v$	$\gamma$	$q$	$\sigma_x$
	short-dated options										
$\hat{\theta}$	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										
$\hat{\theta}$	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  $u = 1, \dots, 15$  and  $\bar{s} = 10^{-5}$ , and with  $p^- = 0.7$ .

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 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.

Table D.8: SVCEJ estimation results

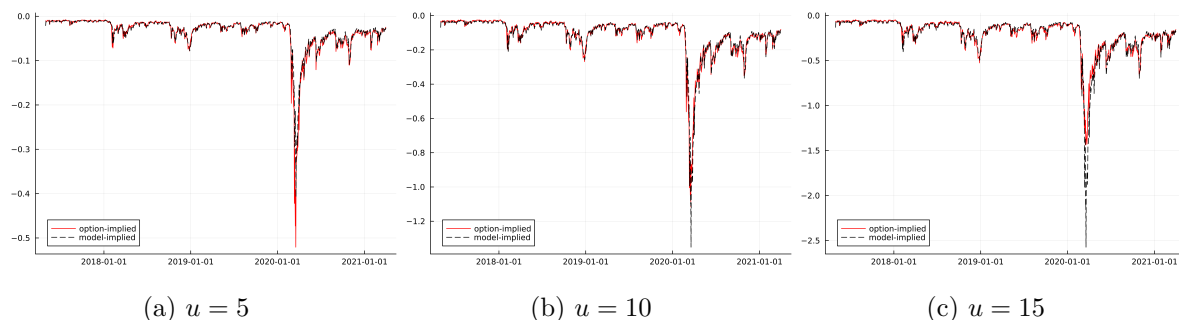
	$\sigma$	$\kappa$	$\bar{v}$	$\rho$	$\delta_0^+$	$\delta_1^-$	$\eta^+$	$\eta^-$	$\mu_v$	$\gamma^+$	$\gamma^-$	$q$	$\sigma_\varepsilon$
	no external factors												
$\hat{\theta}$	0.481	8.31	0.0139	-1.00	3.76	107.3	0.0100	0.0445	0.061	-	-	-	0.221
s.e.	0.007	0.26	0.0002	0.01	0.08	3.99	0.0002	0.0006	0.002				0.004
	$R_0$												
$\hat{\theta}$	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  $u = 1, \dots, 20$  and  $\bar{s} = 10^{-7}$ .

## 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: Real part of the log option-implied and model-implied CCFs

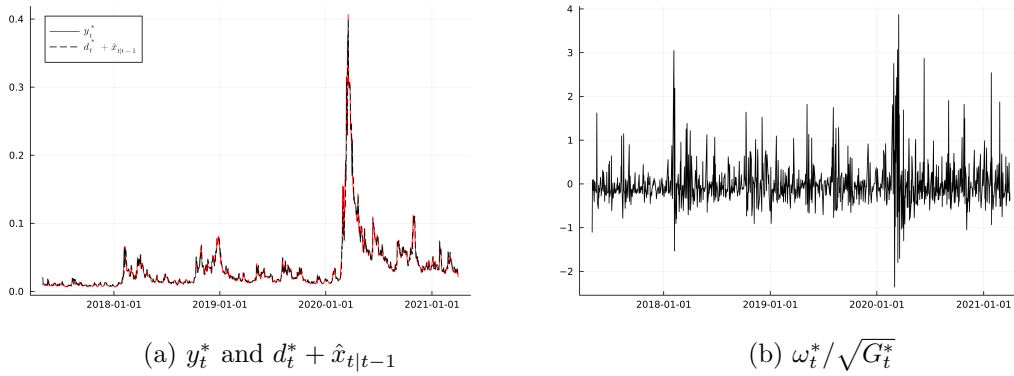


Note: This figure plots the real part of the log option-implied and model-implied CCFs with  $u = \{5, 10, 15\}$  averaged across maturities. The parametric log CCF is based on the SVCDEJ model with parameter estimates in Table 5.

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,  $\log \hat{\psi}_t(u, \tau)$ , 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 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

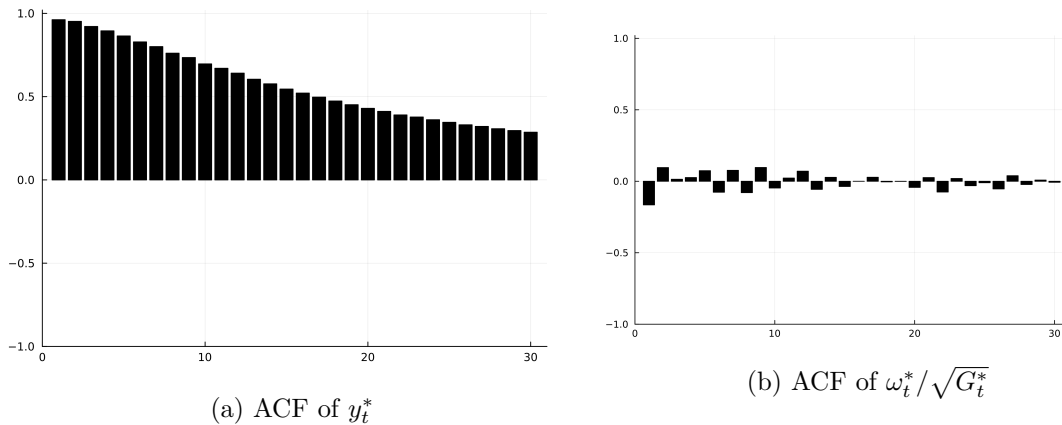
Figure D.3: Fit of the collapsed observation vector  $y_t^*$  and standardized residuals



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.

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



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

- Andersen, T. G., Fusari, N., & Todorov, V. (2015). The risk premia embedded in index options. *Journal of Financial Economics*, 117(3), 558–584.
- 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.
- 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.
- CBOE. (2015). *VIX white paper*. <https://cdn.cboe.com/resources/vix/vixwhite.pdf>.

- Duffie, D., Pan, J., & Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, *68*(6), 1343–1376.
- Fengler, M. R. (2009). Arbitrage-free smoothing of the implied volatility surface. *Quantitative Finance*, *9*(4), 417–428.
- Gatheral, J., & Jacquier, A. (2014). Arbitrage-free SVI volatility surfaces. *Quantitative Finance*, *14*(1), 59–71.
- 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.
- 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.
- Lee, R. W. (2004). The moment formula for implied volatility at extreme strikes. *Mathematical Finance*, *14*(3), 469–480.
- Malz, A. M. (2014). A simple and reliable way to compute option-based risk-neutral distributions. *FEB of New York Staff Report*(677).
- 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.
- Todorov, V. (2019). Nonparametric spot volatility from options. *The Annals of Applied Probability*, *29*(6), 3590–3636.