Functional Estimation of Option Pricing Models*

Yannick Dillschneider
Amsterdam School of Economics
University of Amsterdam
and Tinbergen Institute

Evgenii Vladimirov
Econometric Institute
Erasmus University Rotterdam
and Tinbergen Institute

First version: September 5, 2024
This version: February 13, 2025
(Preliminary and incomplete)

Abstract

In this paper, we develop a novel estimation procedure for parametric option pricing models specified under the risk-neutral measure. We set up our estimation strategy to minimize the distance in a functional sense between the option-implied and model-implied logarithm of conditional characteristic functions. Within a broad class of option pricing models, for which the characteristic function is marginally affine, the model's latent state vector can be concentrated out in closed form. As a result, our estimation procedure is computationally fast and easy to implement, while at the same time exploiting all distributional information contained in an option panel about the risk-neutral dynamics of the underlying asset price. We establish the asymptotic properties of the parameter and state estimators and investigate the finite-sample performance of our method in Monte Carlo simulations. In the empirical application, we illustrate the usefulness of our estimation procedure based on by far the largest option panel, both in the time-series and cross-sectional dimensions, considered in the related literature.

JEL classification: C32, C51, C58, G12, G13

Keywords: option pricing; stochastic volatility models; estimation

*We are very grateful to Gustavo Freire, Piotr Orlowski, Olivier Scaillet, Fabio Trojani, and conference participants at the 2024 FinEML Conference at USI Lugano, the 2024 European Winter Meeting of the Econometric Society for helpful comments and suggestions. Email addresses: y.dillschneider@uva.nl and vladimirov@ese.eur.nl.

1 Introduction

For a long time already, researchers and practitioners alike have been using observed market prices of options to uncover the dynamics of the underlying asset price. Early contributions to the option pricing literature already establish that a cross section of observed option prices allows to infer the risk-neutral distribution of the underlying asset price at the traded maturities (e.g., Banz and Miller, 1978, Breeden and Litzenberger, 1978). The granular information content of option prices likewise proves valuable in the estimation of parametric models with rich dynamics, which hardly can be estimated accurately from time series data on the underlying alone. Following the emergence of stochastic volatility models for option pricing, such as the Heston (1993) model, a voluminous amount of research dating back to at least the late 1990s has therefore applied estimation approaches that incorporate option prices, leading to important insights into the fine structure of empirical asset price dynamics, including its diffusive and jump composition.[1]

When estimating parametric option pricing models using an observed option panel,[2] the procedure for a typical model involves the determination of the parameters and latent state variables that govern the model dynamics. The rich information content of the option panel can be exploited especially to estimate the risk-neutral parameters and state vectors. Existing methods for this essentially aim at minimizing the distance between market-observed and model-implied option prices, where the latter have a non-linear dependence on the state vector and generally need to be determined by some numerical procedure.[3] The key drawback of the common estimation approaches is their notoriously complex computations, primarily driven by the high non-linearity and large dimensionality of the (explicit or implicit) state filtering problem, combined with repeated costly evaluations of model option prices. The resulting estimation problem usually can only be realized after significantly downsampling the observed option panel in one or more dimensions. This includes the use of short sample periods, low-frequency time series observations (weekly or even monthly), and the selection of only a small subset of strikes and maturities on each sample date (e.g., short-dated near-the-money options). Even after downsampling and incurring the accompanying economic information loss, these estimation procedures remain computationally demanding and slow.

[1] Examples include Andersen, Fusari, and Todorov (2015b, 2017, 2020), Bates (1996, 2000), Bakshi, Cao, and Chen (1997), Broadie, Chernov, and Johannes (2007), Eraker (2004), Eraker, Johannes, and Polson (2003), Pan (2002), among many others.

[2] Following Andersen, Fusari, and Todorov (2015a), we define an option panel as a time series of option surfaces, each of which consists of option prices indexed by strike and maturity.

[3] Fourier-type numerical integration techniques (e.g., Carr and Madan, 1999) have established themselves as a standard tool for option pricing. An alternative method that is favored in this paper for simulation purposes is the Fourier-cosine expansion suggested by Fang and Oosterlee (2009).

In this paper, we develop a novel estimation approach that overcomes these issues within a broad class of option pricing models, for which we require only the specification of the risk-neutral side of a model. Instead of directly fitting the model to option prices, our method leverages the risk-neutral conditional characteristic function (CCF) of underlying asset returns, which can be approximated by portfolios of observed option prices at any traded maturity using a well-known spanning result originally established in Bick (1982).[4] As such, the CCF captures the entire risk-neutral conditional distribution of the underlying asset return at the given maturity. Within the *marginal-affine* class of models for which the CCF of underlying asset returns has an exponentially affine dependence on the state variables,[5] we set up the estimation problem to minimize the distance in a functional sense between the "observed" and model log CCFs at different maturities, optimally choosing risk-neutral model parameters and state vectors. The affine form of the log CCFs eventually makes it possible to estimate the latent state vectors in closed form as the solutions of linear least-squares problems, involving the cross section of log CCFs on the respective date. Minimizing the distance between log CCFs rather than option prices further allows us to avoid costly option evaluations altogether. As a consequence, our estimation approach is computationally fast and easy to implement while nevertheless exploiting all observed option information through the CCFs, without the need for any prior downsampling of the option panel.

To appropriately capture the information embedded in CCFs, we represent them as functional objects. The formulation of our estimation method invokes Hilbert space techniques akin to generalized method of moments (GMM) estimation with a continuum of moment conditions (C-GMM) (cf. Carrasco and Florens, 2000). By working with CCFs as functional objects, we retain, in principle, all of the contained distributional information about the risk-neutral dynamics of the underlying asset price. Therefore, similar to the established result in Carrasco, Chernov, Florens, and Ghysels (2007) for C-GMM, this estimation procedure can be expected to yield the efficiency of the maximum likelihood estimators, which are generally infeasible to obtain for the state-of-the-art option pricing models. Nevertheless, our estimation procedure also covers and provides theory for the case of the discretized version of such functional objects.

We establish the asymptotic properties of the model parameter and state estimators under a suitably developed infill asymptotic scheme. In this framework, the time span of the panel is fixed, option prices are observed with errors, and the strike grid size approaches zero as the number of options with fixed maturity increases to infinity. The limiting distribution is mixed-Gaussian, with an asymptotic variance depending on the

[4] This spanning result has later been popularized by Bakshi, Kapadia, and Madan (2003), Britten-Jones and Neuberger (2000), Carr and Madan (2001), among others.

[5] This nests the broad class of affine jump-diffusion models in Duffie, Pan, and Singleton (2000), which comprises a vast majority of the state-of-the-art option pricing models.

particular realization of the state vector path. The derived results rely on the established consistency and a stable functional CLT of the option-implied log CCFs in a Hilbert space. For feasible inference, we derive consistent estimators for the asymptotic variance based on the consistent estimation of option price errors.

To assess the finite-sample performance of our estimation procedure, we conduct an extensive Monte Carlo simulations. In particular, we consider widely-used one- and two-factor model specifications and find very good finite-sample performance of parameter and state estimates. In the empirical application, we illustrate the usefulness of our estimation procedure based on what is, to the best of our knowledge, by far the largest option panel, both in time-series and cross-sectional dimensions, considered in the related literature. The estimated volatility states in both of the models closely resemble the market volatility.

The core related literature for our work consists of the extensive body of research that devises option-based estimation approaches for stochastic volatility models. The employed procedures focus either exclusively on the risk-neutral pricing side of the model (*risk-neutral* estimation) or simultaneously account for both the risk-neutral and real-world model dynamics, linked through risk premium specifications (*full* estimation).

The traditional and most widely adopted risk-neutral estimation approach chooses parameters and state variables to minimize the fitting errors between observed and model option prices (or monotonous transformations thereof, such as implied volatilities) in an option panel according to some error metric (e.g., Bakshi et al., 1997, Broadie et al., 2007, Huang and Wu, 2004; see also Jarrow and Kwok, 2015 for a general discussion). Further augmentations of this estimation method include regularization terms in order to ensure economic plausibility or handle identification issues.[6] An important extension in this direction is suggested by Andersen et al. (2015a) (AFT), which for each date incorporates economic regularization of the model-implied spot volatility towards a high-frequency spot volatility estimate. A common feature of these estimation methods is that they work directly with option prices, which leads to a substantial computational burden.[7] For instance, with the common mean-squared error criterion also used by AFT, it is effectively required to explicitly compute state vector estimates for every day in the option panel as the (numerical) solutions to non-linear least-squares problems, thereby necessitating a large number of costly evaluations of model option prices. Despite sharing a closely related risk-neutral estimation idea, our method circuits the computational issues by relying on closed-form state estimates and completely avoiding the calculation of option

[6]The latter are primarily relevant in semi- or non-parametric settings, which typically suffer from ill-posedness of the estimation problem (e.g., Lagnado and Osher, 1997, Cont and Tankov, 2004).

[7]To mitigate the costs of repeated evaluations of the option pricing function, the use of a surrogate model based on machine learning tools has been suggested in the recent literature (cf. H. Chen, Didisheim, and Scheidegger, 2021).

prices. However, due to its reliance on a specific form of the log CCF, our method is restricted to the class of marginal-affine models. Even though we do not incorporate economic regularization as in AFT, it is possible to extend our method in this direction.

Existing full estimation procedures, which explicitly incorporate the real-world dynamics of the model, employ a variety of strategies to obtain estimates for latent state variables. Analogous state filtering techniques as in risk-neutral estimation can be used for this purpose. Boswijk, Laeven, and Lalu (2016), Boswijk, Laeven, Lalu, and Vladimirov (2023) within a CCF-based GMM procedure utilize state filtering that minimizes mean-squared option pricing errors on each day in the option panel. Andersen et al. (2015b) additionally incorporate economic regularization by embedding their method devised in AFT into a Markov chain Monte Carlo procedure. Eventually, all full estimation approaches of this sort inherit a substantial computational complexity. This is occasionally mitigated by imposing strong assumptions, such as the absence of measurement errors for sufficiently many option observations per date, so that the non-linear state filtering problems reduce to cheaper (numerical) inversions of the option pricing formula. Following this route, several "implied-state" estimation methods are devised in the existing literature, including Pan (2002), Santa-Clara and Yan (2010) in a GMM and Aït-Sahalia and Kimmel (2007) in a maximum likelihood setting.[8] Addressing the apparent limitations of these existing estimation approaches, our method may be extended to a full estimation procedure that combines low computational burdens and non-restrictive assumptions on option measurement errors.

Several full estimation approaches in the existing literature moreover make use of dynamic (Bayesian) state filtering techniques. Simulation-based approximations are typically invoked in implementations, since exact dynamic filters are infeasible. Examples include Markov chain Monte Carlo (e.g., Eraker, 2004), particle filtering (e.g., Bardgett, Gourier, and Leippold, 2019, Christoffersen, Jacobs, and Mimouni, 2010, Fulop and Li, 2019, Johannes, Polson, and Stroud, 2009), or the efficient method of moments (e.g., Andersen, Benzoni, and Lund, 2002, Chernov and Ghysels, 2000). By relying on extensive simulations, all of these approaches carry a heavy computational burden.[9] Alternatively, simulations can be circumvented by reverting to (approximative) non-linear Kalman filtering techniques (e.g., Bates, 2000). The construction of certain portfolios of options, whose model prices exhibit an (exponentially) affine relation to the state vector, further enables the use of the linear Kalman filter. In an affine framework, Feunou and Okou (2018) suggest such an estimation approach based on cumulants of underlying asset

[8]Relatedly, for single-factor stochastic volatility models, Aït-Sahalia, Li, and Li (2021) develop a GMM approach that exploits properties of implied volatility surfaces.

[9]This burden may again be reduced by using a surrogate model for option pricing (cf. Dufays, Jacobs, Liu, and Rombouts, 2023). But even in this study, the authors have to significantly downsample the number of options per day to bear the associated computational costs.

returns. Boswijk, Laeven, and Vladimirov (2024) (BLV) instead use the log CCF with quasi-maximum likelihood estimation.[10]

The contribution of BLV is closely related to this paper particularly because they also leverage the form of log CCFs within an affine framework. However, our method differs from and extends BLV in several dimensions. The closed-form state filtering in our method achieves additional computational gains over the linear Kalman filter in BLV, which itself already realizes substantial improvements over competing approaches. As a side effect, our method only relies on risk-neutral model dynamics, while the dynamic filtering in BLV requires the full specification of both risk-neutral and real-world dynamics. Thus, we avoid potential misspecification issues on the real-world model side and also accommodate settings with small time series dimension of the observed option data.[11] Nevertheless, as mentioned before, our method can also be extended towards full estimation. From a technical perspective, our functional formulation further differs from the discrete setup in BLV that uses only a finite number of CCF arguments. Thus, we avoid the ad-hoc choice of arguments and fully exploit the functional nature of CCFs. We also consider a fixed time span asymptotic scheme and impose weaker non-parametric conditions on option errors.

By relying on the CCF, our paper is also conceptually related to the literature that employs the CCF in general parametric estimation procedures. For many popular models (e.g., those in the affine jump-diffusion class), the advantage of such procedures is that the CCF may admit a tractable expression, whereas the likelihood function does not. Without the use of options, one common estimation approach is to formulate a C-GMM estimator with a continuum of moment conditions implied by the CCF of observables (e.g., Carrasco et al., 2007), invoking the econometric theory developed in Carrasco and Florens (2000). From a theoretical perspective, the literature in this direction emphasizes that CCF-based GMM estimation is able to attain the efficiency of maximum likelihood estimation. Several contributions specifically explore this spectral GMM approach in the setting of an exponentially affine CCF (e.g., Chacko and Viceira, 2003, Jiang and Knight, 2002, Singleton, 2001). Incorporating option prices, Boswijk et al. (2016, 2023) utilize CCF-based GMM estimation with state estimates determined from an option panel, which however suffers from the aforementioned computational problems. Finally, beyond parametric settings, CCFs (or related transforms) obtained from option prices are used in the non-parametric estimation of spot volatility (Todorov, 2019, Todorov and Zhang, 2023), volatility of volatility and the leverage effect (Chong and Todorov, 2024),

[10]Related is also the use of (replicated) variance swaps as in, e.g., Aït-Sahalia, Karaman, and Mancini (2020), Bollerslev, Gibson, and Zhou (2011), Egloff, Leippold, and Wu (2010), Jones (2003), Wu (2011).

[11]For instance, similar to the analysis in Andersen et al. (2020), our approach thus enables a more flexible analysis of various risk premia (i.e., differences of risk-neutral and real-world expectations), without the need to a priori impose a parametric structure on the real-world model dynamics.

and jump variation (Todorov, 2022).

The remainder of the paper is organized as follows. Section 2 sets up the theoretical framework. Subsequently, Section 3 develops our estimation approach and the main econometric theory. In Section 4, we support our estimation approach with simulation results. In Section 5, we report empirical results. Finally, Section 6 concludes the paper. The appendix contains additional supplementary material.

2 Theoretical Framework

Let us denote by F_t a forward asset price at time t defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$. We assume that the market is arbitrage-free, which implies the existence of a risk-neutral probability measure $\mathbb{Q}$, locally equivalent to the real-world probability measure $\mathbb{P}$. Since our interest is in extracting the information content from an option panel, we formulate the model dynamics only under the risk-neutral measure $\mathbb{Q}$, while we leave the dynamics under $\mathbb{P}$ unspecified. In particular, we assume that F_t is governed by the following risk-neutral dynamics:

$$\frac{dF_t}{F_t} = \sqrt{V_t} dW_t + \int_{\mathbb{R}} (e^x - 1)\, \widetilde{\mu}(dt, dx), \tag{2.1}$$

where V_t is an adapted, locally bounded variance process; W_t is a standard $\mathbb{Q}$-Brownian motion; and $\widetilde{\mu}$ is a counting jump measure with risk-neutral compensator $\widetilde{\nu}_t(dt, dx)$. We elaborate on the underlying process dynamics and corresponding assumptions later in the paper.

To capture the distributional properties of F_t, in this paper, we work with the risk-neutral conditional characteristic function (CCF) of standardized log returns:

$$\varphi_t(u, \tau) := \mathbb{E}^{\mathbb{Q}}\left[e^{\mathrm{i}u \frac{\log F_{t+\tau} - \log F_t}{\sqrt{\tau}\kappa_{t,\tau}}} \middle| \mathcal{F}_t \right] \text{ with } \tau > 0,\ u \in \mathbb{R}. \tag{2.2}$$

We consider returns scaled by a measure of total volatility $\sqrt{\tau}\kappa_{t,\tau}$ as a form of standardization for different maturities and different levels of volatility. This is similar to the standardized moneyness levels considered, e.g., in Andersen et al. (2015b), who scale the log strike-to-forward ratio by the level of total volatility.[12] Although we adopt this scaling for the purpose of robustness of the estimation procedure, we emphasize that our theoretical results and the estimation procedure do not require or impose this scaling.

Let us further denote by $O_t(\tau, m)$ the time-t forward price of a European-style out-of-the-money (OTM) option written on the forward gross return $F_{t+\tau}/F_t$ with a time-to-

[12] The scaled version can be easily obtained from the unscaled CCF of log returns with arguments of the form $\frac{u}{\sqrt{\tau}\kappa_{t,\tau}}$. Following Andersen et al. (2015b), in the implementation, we choose $\kappa_{t,\tau}$ to be the at-the-money implied volatility.

maturity $\tau > 0$ and log-moneyness strike m.[13] The OTM option is a call option if $m > 0$ and a put option otherwise. In the absence of arbitrage, the OTM forward option prices are given by the conditional risk-neutral expectation of the corresponding option payoffs:

$$O_t(\tau, m) = \begin{cases} \mathbb{E}^{\mathbb{Q}}[(F_{t+\tau}/F_t - e^m)^+ \mid \mathcal{F}_t] & \text{if } m > 0, \\ \mathbb{E}^{\mathbb{Q}}[(e^m - F_{t+\tau}/F_t)^+ \mid \mathcal{F}_t] & \text{if } m \leq 0. \end{cases}$$

Plain-vanilla options can be used to replicate more complex non-trivial payoffs. Using the payoff-spanning result established in Bick (1982) (see also Bakshi et al., 2003, Britten-Jones and Neuberger, 2000, Carr and Madan, 2001), the CCF of log returns can be spanned as a portfolio of OTM option prices:

$$\varphi_t(u, \tau) = 1 - \left(\frac{u^2}{\tau \kappa_{t,\tau}^2} + \mathrm{i} \frac{u}{\sqrt{\tau} \kappa_{t,\tau}} \right) \int_{\mathbb{R}} e^{\left(\mathrm{i} \frac{u}{\sqrt{\tau} \kappa_{t,\tau}} - 1 \right) m} O_t(\tau, m) dm. \tag{2.3}$$

The option-implied CCF is akin to the construction of the VIX index and is also used, e.g., by Todorov (2019) for non-parametric esimation of spot volatility and BLV for the estimation of parametric models and latent state filtering using the Kalman filter.

Importantly, the portfolio representation result (2.3) is model-independent, i.e., it does not rely on the particular model specification employed here. Hence, one may construct an empirical version of (2.3) when using observed option prices. In practice, however, we do not observe option prices for a continuum of strikes and the observed prices are generally not frictionless. Nevertheless, we can approximate the expression in (2.3) based on a finite number $n_{t,\tau}$ of noisy observed option prices $\widehat{O}_t(\tau, m_j)$ at discrete log-moneyness strikes m_j.[14] Similar to Todorov (2019) and BLV, we employ a Riemann sum approximation and define

$$\widehat{\varphi}_t(u, \tau) := 1 - \left(\frac{u^2}{\tau \kappa_{t,\tau}^2} + \mathrm{i} \frac{u}{\sqrt{\tau} \kappa_{t,\tau}} \right) \sum_{j=2}^{n_{t,\tau}} e^{\left(\mathrm{i} \frac{u}{\sqrt{\tau} \kappa_{t,\tau}} - 1 \right) m_j} \widehat{O}_t(\tau, m_j) \Delta m_j, \tag{2.4}$$

where $\Delta m_j = m_j - m_{j-1}$. Thus, given a liquid set of observed option prices, the CCF of log returns becomes essentially an observable quantity. In Section 3, we discuss assumptions imposed on option errors and provide asymptotic results for the computationally feasible CCF $\widehat{\varphi}_t(u, \tau)$ (cf. Proposition 1).

[13]It is convenient to work with options on forward gross returns since they are uniformly bounded by basic no-arbitrage considerations: $0 \leq O_t(\tau, m) \leq 1$. Denote by $\widetilde{O}_t(\tau, m)$ the time-t forward price of a corresponding option written on the forward price $F_{t+\tau}$. By construction, these option prices satisfy the simple scaling relation $O_t(\tau, m) = \widetilde{O}_t(\tau, m)/F_t$.

[14]A minimal requirement for an empirical approximation of the payoff-spanning result in Bick (1982) to be meaningful is that it does not admit arbitrary values in an incomplete market setting. By the theory of Bondarenko, Dillschneider, Schneider, and Trojani (2024), the CCF has this robustness property since the underlying complex-exponential payoff is bounded.

We further assume that the variance and jump processes are driven by a d-dimensional state vector $\mathbf{x}_t$ on a state space $D \subset \mathbb{R}^d$, which follows a jump-diffusive Markov process. In particular, we let $V_t = v(\mathbf{x}_t)$ and $\widetilde{\nu}_t(dt, dx) = \lambda(\mathbf{x}_t)dt \otimes \nu(dx)$ for some deterministic functions v: $D \to \mathbb{R}^+$ and λ: $D \to \mathbb{R}^+$ as well as a state-independent jump size measure ν. The CCF of (standardized) log returns satisfies $\varphi_t(u, \tau) = \varphi_t(u, \tau; \theta_0, \mathbf{x}_t)$ given the (unknown) true model parameters θ_0 that governs the model dynamics, where the specification is such that the CCF is of exponential-affine form:

$$\varphi_t(u, \tau; \theta, \mathbf{z}_t) = \exp\big(\alpha_t(u, \tau; \theta) + \beta_t(u, \tau; \theta)^\intercal \mathbf{z}_t\big). \tag{2.5}$$

Here, $\alpha_t(u, \tau; \theta)$ and $\beta_t(u, \tau; \theta)$ are $\mathbb{C}$- and $\mathbb{C}^d$-valued functions, θ is a parameter vector, and $\mathbf{z}_t \in D$ is a state vector. The time variation in the coefficients is coming from the measure of volatility $\kappa_{t,\tau}$, i.e., we can equivalently write $\alpha_t(u, \tau; \theta) := \alpha(\frac{u}{\sqrt{\tau}\kappa_{t,\tau}}, \tau; \theta)$ and $\beta_t(u, \tau; \theta) := \beta(\frac{u}{\sqrt{\tau}\kappa_{t,\tau}}, \tau; \theta)$. The formulation in equation (2.5) nests the majority of option pricing models considered in the literature, including the affine jump-diffusion (AJD) class of Duffie et al. (2000). We note that while the defining property of the AJD class is an exponentially affine joint CCF of log returns and states (Duffie, Filipović, and Schachermayer, 2003), here we only require the marginal CCF of log returns to be exponentially affine. This allows us to accommodate a broader class of models than the standard AJD specifications. Any model with a CCF of the form in (2.5) will be called *marginal-affine* in this paper.

Under the suitably developed asymptotic scheme, in Section 3, we quantify the approximation errors in the computationally feasible CCF in (2.4), and show that the two CCFs $\widehat{\varphi}_t$ and φ_t are asymptotically the same under reasonable regularity conditions (cf. Lemma B.1). Therefore, under the correct model specification, we can use the option-implied CCF $\widehat{\varphi}_t$ to infer the true parameter vector θ_0 and state vector $\mathbf{x}_t$. The same holds true when considering complex logarithm[15] of the CCFs. Specifically, the exponential-affine form of the parametric CCF in (2.5) implies an affine form for the log CCF $\psi_t(u, \tau; \theta, \mathbf{z}_t) := \log \varphi_t(u, \tau; \theta, \mathbf{z}_t)$:

$$\psi_t(u, \tau; \theta, \mathbf{z}_t) = \alpha_t(u, \tau; \theta) + \beta_t(u, \tau; \theta)^\intercal \mathbf{z}_t. \tag{2.6}$$

Due to the use of finitely many noisy option prices, the logarithm of the option-implied CCF in (2.4), $\widehat{\psi}_t(u, \tau) := \log \widehat{\varphi}_t(u, \tau)$, yields a noisy version of the true log CCF $\psi_t(u, \tau) = \psi_t(u, \tau; \theta_0, \mathbf{x}_t)$:

$$\widehat{\psi}_t(u, \tau) = \alpha_t(u, \tau; \theta_0) + \beta_t(u, \tau; \theta_0)^\intercal \mathbf{x}_t + \xi_t(u, \tau), \tag{2.7}$$

[15] As such, the complex logarithm is multi-valued. To uniquely determine the logarithms of the CCFs $\varphi_t(u, \tau; \theta, \mathbf{z}_t)$ and $\widehat{\varphi}_t(u, \tau)$, we construct them as the *distinguished* logarithms (cf. Section 3.1). In essence, we normalize them to zero at $u = 0$ and further make sure that there are no discontinuities in the resulting functions with respect to u when crossing the branches. In practice, we ensure this by taking the logarithm sequentially on a fine grid starting with arguments closest to the origin.

where $\xi_t(u,\tau)$ is the complex-valued measurement error, stemming from the approximation of the log CCF. We emphasize that the left-hand side variable of equation (2.7) is essentially an observed quantity, which can be computed as a function of a portfolio of observed option prices.

Equation (2.7) is a functional complex-valued linear factor model. By considering a finite number of CCF arguments and taking the real and imaginary parts, BLV transform this functional form into a real-valued linear vector representation, which serves as the measurement equation in their estimation procedure. In this paper, we develop the estimation procedure directly with the complex-valued functional objects. Akin to the GMM with a continuum of moment conditions (cf. Carrasco and Florens, 2000, Carrasco et al., 2007), this allows us to exploit all the information embedded in the CCF.

Our proposed estimation procedure is based on minimizing the distance between the logarithms of the option-implied and model-implied CCFs. In particular, for an option panel consisting of T days with a set of maturities $\mathcal{T}_t$ on day t, we minimize:

$$\min_{\theta\in\Theta,\,\{\mathbf{z}_t\}_{t=1,\dots,T}} \sum_{t=1}^{T} w_t \sum_{\tau\in\mathcal{T}_t} \|\widehat{\psi}_t(\cdot,\tau) - \alpha_t(\cdot,\tau;\theta) - \beta_t(\cdot,\tau;\theta)^{\mathsf{T}}\mathbf{z}_t\|^2, \tag{2.8}$$

where w_t are some deterministic weights. The norm $\|\cdot\|$ that measures the magnitude of the functional objects is formally defined in Section 3.

The optimization problem (2.8) is reminiscent of minimizing the difference between the market-observed and model-implied option prices (or their monotonic transformations, such as implied volatilities). In fact, the commonly adopted approach in the literature (see, e.g., Bakshi et al., 1997, Broadie et al., 2007) solves the following problem:

$$\min_{\theta\in\Theta,\,\{\mathbf{z}_t\}_{t=1,\dots,T}} \sum_{t=1}^{T} w_t \sum_{i=1}^{N_t} \left(\widehat{O}_t(\tau_i,m_i) - O_t(\tau_i,m_i;\theta,\mathbf{z}_t)\right)^2, \tag{2.9}$$

where N_t is the number of different strike-maturity pairs observed on day t, $\widehat{O}_t(\tau_i,m_i)$ are the market-observed option prices, and $O_t(\tau_i,m_i;\theta,\mathbf{z}_t)$ are the model-implied option prices for the same option characteristics. AFT add a penalty term to this minimization problem to tie the dynamics of the option panel to the dynamics of the underlying asset. This can also, in principle, be incorporated in our estimation approach.

Although the two optimization problems look similar, working with log CCFs as in problem (2.8) offers several important advantages. First, the objective function in (2.8) does not require evaluating option prices for a given parameter vector θ. This is in contrast to problem (2.9), where calculation of the model-implied option prices, $O_t(\tau_i,m_i;\theta,\mathbf{z}_t)$, for each day and for each parameter value, requires the use of numerical option evaluation techniques (e.g., the FFT approach of Carr and Madan (1999) or the COS method of Fang and Oosterlee (2009)). This makes our estimation procedure computationally appealing since the option evaluation methods are often computationally demanding.

Another advantage is due to the fact that the objective function in problem (2.8) is linear in the state vector, while option prices (or implied volatilities) are highly non-linear functions of the state vector. The latter implies that, if the state vector is latent, we have $d_\theta + d \times T$ parameters to explicitly optimize in problem (2.9), where d_θ is the number of model parameters. This imposes a considerable computational burden if one wants to analyze data with a large time series dimension. As we show in the next section, the linear relation allows us to easily concentrate the state vectors out of the criterion function in closed form as the solutions to linear functional least-squares problems. This effectively reduces the number of parameters to explicitly optimize over to d_θ.

We end this section by emphasizing that working with log CCFs makes our estimation procedure computationally fast and easy to implement, and, what is even more important, it allows us to exploit all distributional information about the risk-neutral dynamics of the underlying asset contained in observed option prices.

3 Estimation Procedure

In this section, we formally introduce our estimation procedure and establish the asymptotic properties of the parameter and state estimators.

3.1 Observation scheme

We start by discussing the observation scheme of option prices and option portfolios. For that, we first state an assumption about the underlying process:

Assumption 1. *Under the risk-neutral measure $\mathbb{Q}$:*

(i) F_t is the unique solution to the SDE (2.1) with some positive and locally bounded variance process V_t and jump components such that $\int_{\mathbb{R}}(x^2 \wedge 1)\nu(dx) < \infty$;

(ii) $\mathbb{E}^{\mathbb{Q}}[(F_{t+\tau}/F_t)^{1+\overline{q}} \mid \mathcal{F}_t] < \infty$ and $\mathbb{E}^{\mathbb{Q}}[(F_{t+\tau}/F_t)^{-\underline{q}} \mid \mathcal{F}_t] < \infty$ for all $t \leq T$ and $\tau \leq \overline{T}$ with some $\overline{T} > 0$ as well as some $\overline{q} > 0$ and $\underline{q} > 0$.

Assumption 1(i) imposes regularity conditions on the underlying stochastic process and is satisfied for standard continuous-time option pricing models considered in the literature. Assumption 1(ii) requires the existence of certain conditional moments of the underlying process and its reciprocal, which regulates the tail behavior of option prices (cf. Lee, 2004). As we show in Lemma B.1, this determines the rate of convergence of approximation errors in option-implied CCFs. We emphasize that Assumption 1 does not restrict the discussion to the marginal-affine class of models and is rather weak.

Our data consist of an option panel observed at integer times $t \in \mathbb{T} := \{1, \ldots, T\}$ with fixed T and a large cross-section consisting of options with different strikes and tenors. In particular, on each date t, we observe a non-empty and finite collection of maturities $\mathcal{T}_t \subset (0, \overline{T}]$, and for each maturity $\tau \in \mathcal{T}_t$, we observe $n_{t,\tau}$ OTM options with log-moneyness strikes in $\mathcal{M}_{t,\tau} \subset \mathbb{R}$. Enumerating the log-moneyness strikes in $\mathcal{M}_{t,\tau}$ by $m_{t,\tau}(j)$, we allow for a non-equidistant log-monyeness grid

$$\underline{m}_{t,\tau} := m_{t,\tau}(1) < \ldots < m_{t,\tau}(n_{t,\tau}) =: \overline{m}_{t,\tau}$$

and denote the distance between adjacent log-moneyness strikes by $\Delta_{t,\tau}(j) := m_{t,\tau}(j) - m_{t,\tau}(j-1)$ for $j = 2, \ldots, n_{t,\tau}$. Our asymptotic scheme is of joint type in which the log-moneyness grid size $\sup_{j=2,\ldots,n_{t,\tau}} \Delta_{t,\tau}(j)$ goes to zero, while the log-moneyness limits $-\underline{m}_{t,\tau}$ and $\overline{m}_{t,\tau}$ increase to infinity as $n := \min_{t,\tau} n_{t,\tau} \to \infty$ for fixed maturity sets $\mathcal{T}_t$ and fixed observation times $t \in \mathbb{T}$. The sequence of strike grids is such that the sets $\mathcal{M}_{t,\tau}$ are monotonously increasing with $n_{t,\tau}$. To accommodate an infill asymptotic scheme, we impose the following assumption on the asymptotic behavior of the log-moneyness grid.

Assumption 2. *For the log-moneyness grid, the following holds:*

(i) There exist deterministic sequences $\Delta_n > 0$ and $0 < \eta_n \leq 1$ depending on n such that $\Delta_n \to 0$ and $\eta_n \to 1$ as $n \to \infty$ and

$$\eta_n \Delta_n \leq \inf_{j=2,\ldots,n_{t,\tau}} \Delta_{t,\tau}(j) \leq \sup_{j=2,\ldots,n_{t,\tau}} \Delta_{t,\tau}(j) \leq \Delta_n.$$

(ii) There exist $\underline{\alpha} > 0$ and $\overline{\alpha} > 0$ such that $e^{\underline{m}_{t,\tau}} \asymp n^{-\underline{\alpha}}$ and $e^{\overline{m}_{t,\tau}} \asymp n^{\overline{\alpha}}$.

(iii) The set $\mathcal{M}_{t,\tau}$ is monotonously increasing as $n_{t,\tau} \to \infty$.

Finally, as common in the related literature, we let option prices be observed with measurement errors due to, e.g., the presence of bid-ask spreads, tick sizes of quotes, and liquidity issues. To distinguish sources of randomness, we define the filtration $\{\mathcal{F}_t^X\}_{t\geq 0}$ generated by the stochastic underlying and state processes and the associated sub-σ-algebra $\mathcal{F}^X := \bigvee_{t\geq 0} \mathcal{F}_t^X \subset \mathcal{F}$. Accordingly, true option prices $O_t(\tau, m)$ are $\mathcal{F}_t^X$-adapted, whereas measurement errors are not. The following assumption details the structure of measurement errors:

Assumption 3. *Option prices are observed with an additive error term:*

$$\widehat{O}_t(\tau, m) = O_t(\tau, m) + \zeta_t(\tau, m), \quad t \in \mathbb{T},\ \tau \in \mathcal{T}_t,\ m \in \mathcal{M}_{t,\tau}, \tag{3.1}$$

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

(i) $\varkappa_t(\tau, m)$ are $\mathcal{F}^X$-conditionally independent along tenors τ, moneyness m and time t;

(ii) $\mathbb{E}[\varkappa_t(\tau, m) \mid \mathcal{F}^X] = 0$, $\mathbb{E}[\varkappa_t^2(\tau, m) \mid \mathcal{F}^X] = 1$, and $\sup_{m\in\mathbb{R}} \mathbb{E}[\varkappa_t^4(\tau, m) \mid \mathcal{F}^X] < \infty$;

(iii) $\sigma_t(\tau, m) = O_t(\tau, m)\widetilde{\sigma}_t(\tau, m)$ is $\mathcal{F}_t^X$-adapted with $\sup_{m\in\mathbb{R}} \widetilde{\sigma}_t^2(\tau, m) < \infty$.

Assumption 3 allows for heteroskedastic option errors, while excluding any dependence structure across the log-moneyness, tenor, and time dimensions conditional on $\mathcal{F}^X$. This assumption avoids a parametric specification of the error terms as in alternative estimation procedures (e.g., Bardgett et al., 2019, BLV, and Dufays et al., 2023) and is similar in sprit to AFT and Todorov (2019).

Given the observation scheme of option prices, we want to establish the observational properties of the log of the option-implied CCF, $\widehat{\psi}_t(\cdot,\tau)$, in a suitably developed infill asymptotic scheme. For our purposes, we focus the attention on a bounded interval of CCF arguments $\mathcal{U} = [-U, U] \subset \mathbb{R}$. The symmetry of $\mathcal{U}$ around the origin is for convenience and not restrictive, accounting for the fact that the true CCF $\varphi_t(\cdot,\tau)$ and its observed counterpart $\widehat{\varphi}_t(\cdot,\tau)$ in equation (2.4) are both Hermitian functions by construction. To rigorously define the associated log CCFs $\psi_t(\cdot,\tau)$ and $\widehat{\psi}_t(\cdot,\tau)$, we set them equal to the *distinguished* logarithms[16] of $\varphi_t(\cdot,\tau)$ and $\widehat{\varphi}_t(\cdot,\tau)$, respectively. The absence of zeros of a function over $\mathcal{U}$ is essential for the existence of its distinguished logarithm (cf. Theorem 7.6.3 in Chung, 2000). Hence, we impose the following assumptions:

Assumption 4. *For every $\tau \in \mathcal{T}_t$ and $t \in \mathbb{T}$, the following holds:*

(i) $\kappa_{t,\tau} \geq \delta_\kappa > 0$;

(ii) $\varphi_t(u,\tau) \neq 0$ *and* $\widehat{\varphi}_t(u,\tau) \neq 0$ *for all* $u \in \mathcal{U}$.

Assumption 4(i) imposes a technical lower bound that controls the range of the arguments $\frac{u}{\sqrt{\tau}\kappa_{t,\tau}}$ of the unscaled CCF. Assumption 4(ii) guarantees the existence of the respective distinguished logarithms. The requirement for $\widehat{\varphi}_t(u,\tau)$ will eventually (for large enough n) be automatically satisfied given the absence of zeros of $\varphi_t(\cdot,\tau)$, noting the uniform convergence of $\widehat{\varphi}_t(u,\tau)$ on the compact $\mathcal{U}$ (cf. Lemma B.1). Moreover, it should be noted that both distinguished logarithms $\psi_t(\cdot,\tau)$ and $\widehat{\psi}_t(\cdot,\tau)$ inherit the Hermitian function property.

To accommodate working with $\widehat{\psi}_t(\cdot,\tau)$ as functional objects across several maturities τ, we introduce a Hilbert space setting for vector-valued functions in which we will work. Concretely, take a finite (Borel) measure π over the compact set $\mathcal{U} = [-U, U]$ such that π is symmetric around the origin.[17] One natural choice to generate the measure π when dealing with a continuum of CCF arguments is a PDF (see, e.g., Carrasco et al., 2007, Carrasco and Kotchoni, 2017). Moreover, our general specification also nests the case with a finite number of CCF arguments, obtained by choosing a discrete π that corresponds

[16]The distinguished logarithm of a continuous function $f\colon \mathcal{U} \to \mathbb{C}$ with $f(0) = 1$ is the unique continuous function $g\colon \mathcal{U} \to \mathbb{C}$ with $g(0) = 0$ such that $f(u) = \exp(g(u))$ for all $u \in \mathcal{U}$.

[17]Formally, we thus require for π that $\pi([-u, 0]) = \pi([0, u])$ for all positive $u \leq U$. E.g., when $\pi(du) = \widetilde{\pi}(u)du$ is generated from a density $\widetilde{\pi}$, we require that the density is symmetric around the origin with $\widetilde{\pi}(-u) = \widetilde{\pi}(u)$ for all $u \in \mathcal{U}$.

to some PMF. Therefore, the asymptotic results developed below apply equally to the case when one works with the entire log CCF in functional form and with its discretized version.

For the given measure π, we consider the associated space $L_p^2(\pi)$ of (equivalence classes of) complex-valued functions $\mathbf{f} = (f_1, \dots, f_p)^\mathsf{T}\colon \mathbb{R} \to \mathbb{C}^p$, formally defined as

$$L_p^2(\pi) := \left\{ \mathbf{f}\colon \mathbb{R} \to \mathbb{C}^p : \int_{\mathcal{U}} |f_i(u)|^2 \pi(du) < \infty \text{ for } i = 1, \dots, p \right\}.$$

Writing $\mathbf{f}^\mathsf{H} = (\bar{\mathbf{f}})^\mathsf{T}$ for the complex conjugate transpose, we equip $L_p^2(\pi)$ with the canonical inner product

$$\langle \mathbf{f}, \mathbf{g} \rangle := \int_{\mathcal{U}} \mathbf{g}(u)^\mathsf{H} \mathbf{f}(u) \pi(du) = \sum_{i=1}^{p} \int_{\mathcal{U}} f_i(u) \overline{g_i(u)} \pi(du),$$

which induces the norm $\|\mathbf{f}\| = \langle \mathbf{f}, \mathbf{f} \rangle^{1/2}$.[18] Evidently, it holds by construction that $\mathbf{f} \in L_p^2(\pi)$ exactly when $f_i \in L_1^2(\pi)$ for all $i = 1, \dots, p$. Moreover, if components of $\mathbf{f}$ and $\mathbf{g}$ are Hermitian functions, we have that $\langle \mathbf{f}, \mathbf{g} \rangle$ is real-valued due to the symmetry of $\mathcal{U}$ and π.

To accommodate the time-dependent maturity sets $\mathcal{T}_t$ in a notationally convenient way, we henceforth set $p = |\mathcal{T}|$ for the collection of maturities $\mathcal{T} := \bigcup_{t \in \mathbb{T}} \mathcal{T}_t = \{\tau_1, \dots, \tau_p\}$. Whenever some τ_i is not in the set of observed maturities $\mathcal{T}_t$, we take the respective element of a $\mathbb{C}^p$-valued function $\mathbf{f}(u) := (f(u, \tau_1), \dots, f(u, \tau_p))^\mathsf{T}$ to be zero. Employing this convention, denote by $\boldsymbol{\varphi}_t$ the vector of true CCFs and by $\hat{\boldsymbol{\varphi}}_t$ the corresponding vector of observed CCFs as in equation (2.4) on date t. Analogously, denote by $\boldsymbol{\psi}_t$ the vector of true log CCFs and by $\widehat{\boldsymbol{\psi}}_t$ the associated vector of observed log CCFs on date t.

The introduced Hilbert space setup is immediately compatible with the properties of the CCF $\boldsymbol{\varphi}_t$ and its distinguished logarithm $\boldsymbol{\psi}_t$ as well as their observed counterparts. It is automatically assured that the true CCF $\boldsymbol{\varphi}_t \in L_p^2(\pi)$, due to its boundedness as a characteristic function. In addition, we obtain that also the observed CCF $\hat{\boldsymbol{\varphi}}_t \in L_p^2(\pi)$, as the CCF measurement errors are uniformly bounded on the compact $\mathcal{U}$ (cf. Lemma B.1). Since a distinguished logarithm is continuous on the compact $\mathcal{U}$ and thus bounded, it likewise holds that $\boldsymbol{\psi}_t \in L_p^2(\pi)$ and $\widehat{\boldsymbol{\psi}}_t \in L_p^2(\pi)$.

As discussed in BLV, the measurement errors of the option-implied log CCF $\widehat{\boldsymbol{\psi}}_t$ arise due to the observation, truncation, and discretization errors in the construction of the corresponding option portfolios in equation (2.4). The next proposition establishes observational properties of this approximation in the Hilbert space $L_p^2(\pi)$. The proof is given in Appendix B.2.

[18]To keep the notation simple, we likewise denote by $\|\mathbf{h}\|$ the Euclidean norm of any complex vector $\mathbf{h}$ and by $\|\mathbf{H}\|$ the induced matrix norm of any complex matrix $\mathbf{H}$. Occasionally, we also use the Frobenius matrix norm $\|\mathbf{H}\|_F$.

Proposition 1. *Suppose Assumptions 1–4 hold. Then, for each $t \in \mathbb{T}$, we have*

$$\widehat{\boldsymbol{\psi}}_t \xrightarrow{\mathbb{P}} \boldsymbol{\psi}_t \text{ in } L^2_p(\pi).$$

If, in addition, $(\underline{\alpha} \wedge \overline{\alpha}) > 1/(2(\underline{q} \wedge (1 + \overline{q})))$, we have, independently across $t \in \mathbb{T}$,

$$\Delta_n^{-1/2}\big(\widehat{\boldsymbol{\psi}}_t - \boldsymbol{\psi}_t\big) \xrightarrow{\mathcal{F}^X\text{-}s} \mathcal{N}\big(0, \mathcal{K}^{(t)}, \mathcal{S}^{(t)}\big) \text{ in } L^2_p(\pi),$$

where the $\mathcal{F}^X$-measurable covariance and relation operators, $\mathcal{K}^{(t)}$ and $\mathcal{S}^{(t)}$, are defined in equations (B.17) and (B.18) in Appendix B.2.

Proposition 1 states that the measurement errors in $\widehat{\boldsymbol{\psi}}_t$ converge to zero as the (minimum) number of options $n \to \infty$. In other words, the errors in option prices are 'averaged out' when constructing the option-implied CCF according to equation (2.4) and taking its logarithm. Moreover, if the log-moneyness limits $-\underline{m}_{t,\tau}$ and $\overline{m}_{t,\tau}$ grow sufficiently fast, the measurement errors in $\widehat{\boldsymbol{\psi}}_t$ satisfy an $\mathcal{F}^X$-stable CLT. Here and throughout, we use $\xrightarrow{\mathcal{F}^X\text{-}s}$ to denote $\mathcal{F}^X$-stable convergence, while in Appendix A we provide further details on this form of convergence in the Hilbert space $L^2_p(\pi)$. For the measurement errors in $\widehat{\boldsymbol{\psi}}_t$, Proposition 1 establishes a limiting mixed-complex Gaussian distribution as $n \to \infty$, whose $\mathcal{F}^X$-measurable covariance and relation operators $\mathcal{K}^{(t)}$ and $\mathcal{S}^{(t)}$ depend on the particular realization of the path of the state vector $\mathbf{x}_t$. The functional form of the established CLT proves to be useful for the estimation procedure that we develop in what follows.

3.2 Estimation

For the main methodological contribution of this paper, we now turn to divising our functional estimation procedure for parametric option pricing models. As discussed in Section 2, we restrict our attention to parametric models in the marginal-affine class. The next assumption summarizes conditions we require from the parametric model:

Assumption 5. *(i) The true parameter vector $\theta_0 \in \text{Int}(\Theta)$, where Θ is a compact parameters space containing admissible parameter values;*

(ii) Under the risk-neutral measure $\mathbb{Q}$, the CCF is of the exponential-affine form (2.5) for θ_0 and $\mathbf{x}_t \in D$ at each $t \in \mathbb{T}$;

(iii) $\alpha_t(u, \tau; \theta)$ and $\beta_t(u, \tau; \theta)$ are continuous in u for all $\theta \in \Theta$, $\tau \in \mathcal{T}_t$, and $t \in \mathbb{T}$.

Assumption 5(ii) strengthens Assumption 1(i) on the underlying process by imposing a particular parametric form. It is satisfied for the majority of the parametric option pricing models considered in the literature since it includes the widely employed AJD class. The admissibility conditions on the parameter space in Assumption 5(i) reflect the joint regularity conditions on D, v and λ that guarantee the existence of a solution to the SDE (2.1). For a detailed discussion of general admissibility conditions in the AJD class, see, e.g., Duffie and Kan (1996). Assumption 5(iii) further assures that $\psi_t(\cdot, \tau; \theta, \mathbf{z}_t)$ in equation (2.6) is the distinguished logarithm of $\varphi_t(\cdot, \tau; \theta, \mathbf{z}_t)$ in equation (2.5).

Consistent with the notation employed so far, define the $\mathbb{C}^p$-vectors $\boldsymbol{\psi}_t(u; \theta, \mathbf{z}_t) = \boldsymbol{\alpha}_t(u; \theta) - \boldsymbol{\beta}_t(u; \theta)\mathbf{z}_t$ of model-implied log CCFs according to equation (2.6), depending on $\mathbf{z}_t \in \mathbb{R}^d$ as well as the $\mathbb{C}^p$-vectors $\boldsymbol{\alpha}_t(u; \theta)$ and $\mathbb{C}^{p\times d}$-matrices $\boldsymbol{\beta}_t(u; \theta)$ of affine coefficient functions. By construction, only those elements of the residuals $\boldsymbol{\xi}_t(u; \theta, \mathbf{z}_t) := \widehat{\boldsymbol{\psi}}_t(u) - \boldsymbol{\psi}_t(u; \theta, \mathbf{z}_t)$ corresponding to maturities in $\mathcal{T}_t$ can be non-zero. Recall that the introduced objects should be viewed as vector- and matrix-valued functions of u.

To meaningfully formulate the estimation problem and asymptotic theory within our Hilbert space setting, we maintain the following regularity conditions, which assure that all relevant functions are contained in $L_p^2(\pi)$:

Assumption 6. *For every $t \in \mathbb{T}$ and $\theta \in \Theta$, the following holds:*

(i) Elements of $\boldsymbol{\alpha}_t(\theta)$ and $\boldsymbol{\beta}_t(\theta)$ are in $L_1^2(\pi)$;

(ii) Elements of $\nabla_\theta \boldsymbol{\alpha}_t(\theta)$, $\nabla_\theta \boldsymbol{\beta}_t(\theta)$ and $\nabla_\theta^2 \boldsymbol{\alpha}_t(\theta)$, $\nabla_\theta^2 \boldsymbol{\beta}_t(\theta)$ are in $L_1^2(\pi)$;

(iii) Elements of $\nabla_\theta^2 \boldsymbol{\alpha}_t(\theta)$, $\nabla_\theta^2 \boldsymbol{\beta}_t(\theta)$ are uniformly bounded in θ.

Assumption 6(i) is consistent with the fact that $\boldsymbol{\psi}_t(\theta, \mathbf{z}_t) \in L_p^2(\pi)$ for any given $\theta \in \Theta$ and $\mathbf{z}_t \in \mathbb{R}^d$, since $\psi_t(\cdot, \tau; \theta, \mathbf{z}_t) \in L_1^2(\pi)$ as the distinguished logarithm of a model CCF. Combined with $\widehat{\boldsymbol{\psi}}_t \in L_p^2(\pi)$, this further yields that residuals satisfy $\boldsymbol{\xi}_t(\theta, \mathbf{z}_t) \in L_p^2(\pi)$. The supposed existence of derivatives in Assumption 6(ii) implies the continuity (elementwise in $L_1^2(\pi)$) of $\boldsymbol{\alpha}_t(\theta)$, $\boldsymbol{\beta}_t(\theta)$ and $\nabla_\theta \boldsymbol{\alpha}_t(\theta)$, $\nabla_\theta \boldsymbol{\beta}_t(\theta)$. Due to the compactness of Θ, these functions are therefore uniformly bounded in θ. Accordingly, Assumption 6(iii) may be replaced by continuity of second-order derivatives, which would imply the uniform boundedness.

Under the imposed regularity conditions, let us define a criterion function Q_T that measures the distance in $L_p^2(\pi)$ between the option-implied log CCF $\widehat{\boldsymbol{\psi}}_t$ and the model log CCF $\boldsymbol{\psi}_t(\theta, \mathbf{z}_t)$, i.e., the magnitude of the residuals $\boldsymbol{\xi}_t(\theta, \mathbf{z}_t)$, in the option panel. Using the $L_p^2(\pi)$ norm $\|\cdot\|$, we specifically take

$$Q_T(\theta, \{\mathbf{z}_t\}_{t\in\mathbb{T}}) := \sum_{t\in\mathbb{T}} w_t \, \|\widehat{\boldsymbol{\psi}}_t - \boldsymbol{\alpha}_t(\theta) - \boldsymbol{\beta}_t(\theta)\mathbf{z}_t\|^2 \tag{3.2}$$

with weights $w_t = 1/|\mathcal{T}_t|$ that account for the time-varying number of observed option maturities. The true parameter vector θ_0 and the associated state vectors $\{\mathbf{x}_t\}_{t\in\mathbb{T}}$ can be estimated by minimizing the distance between the option-implied and model log CCFs as quantified by Q_T:

$$(\widetilde{\theta}, \{\widetilde{\mathbf{x}}_t\}_{t\in\mathbb{T}}) := \underset{\theta\in\Theta,\, \{\mathbf{z}_t\}_{t\in\mathbb{T}}\subset D}{\operatorname{argmin}} \; Q_T(\theta, \{\mathbf{z}_t\}_{t\in\mathbb{T}}) \,. \tag{3.3}$$

Problem (3.3) has a particular structure, which allows to concentrate out the state vector. Concretely, the estimation problem (3.3) may be equivalently formulated as:

$$\widetilde{\theta} := \underset{\theta\in\Theta}{\operatorname{argmin}}\ Q_T(\theta, \{\widetilde{\mathbf{x}}_t(\theta)\}_{t\in\mathbb{T}}) \tag{3.4}$$

with $\widetilde{\mathbf{x}}_t := \widetilde{\mathbf{x}}_t(\widetilde{\theta})$ for $t \in \mathbb{T}$ provided that states are uniquely identified, where for each date t separately, parameter-dependent state estimates $\widetilde{\mathbf{x}}_t(\theta)$ are formed according to

$$\widetilde{\mathbf{x}}_t(\theta) := \underset{\mathbf{z}_t\in D}{\operatorname{argmin}}\ \|\widehat{\boldsymbol{\psi}}_t - \boldsymbol{\alpha}_t(\theta) - \boldsymbol{\beta}_t(\theta)\mathbf{z}_t\|^2. \tag{3.5}$$

The restriction of the domain to the state space $D \subset \mathbb{R}^d$ acts as a possibly binding constraint. Hence, $\widetilde{\mathbf{x}}_t(\theta)$ is the solution of a constrained complex-valued projection problem that generally is not available in closed form. As such, problems (3.3) and (3.4) thereby require joint numerical optimization of parameters and states, which makes them high-dimensional problems that are computationally expensive to solve.

However, when relaxing the domain to $\mathbb{R}^d$ in problem (3.5), it is possible to provide a closed-form solution. Usual rank conditions assure state identifiability, which are implied by the following assumption:

Assumption 7. *For every $t \in \mathbb{T}$ and $\theta \in \Theta$, the minimal singular value satisfies $\sigma_{\min}(\langle\boldsymbol{\beta}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle) \geq \delta_\sigma > 0$.*

In particular, by relaxing problem (3.5), we obtain parameter-dependent state estimates $\widehat{\mathbf{x}}_t(\theta)$ as follows:

$$\widehat{\mathbf{x}}_t(\theta) := \underset{\mathbf{z}_t\in \mathbb{R}^d}{\operatorname{argmin}}\ \|\widehat{\boldsymbol{\psi}}_t - \boldsymbol{\alpha}_t(\theta) - \boldsymbol{\beta}_t(\theta)\mathbf{z}_t\|^2. \tag{3.6}$$

Problem (3.6) is an unconstrained complex-valued linear projection problem. With Assumptions 6 and 7 in place, it admits the following unique closed-form solution:[19]

$$\widehat{\mathbf{x}}_t(\theta) = \langle\boldsymbol{\beta}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle^{-1}\langle\widehat{\boldsymbol{\psi}}_t - \boldsymbol{\alpha}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle. \tag{3.7}$$

Note that due to the Hermitian properties of the involved functions together with the symmetry of $\mathcal{U}$ and π, it is automatically assured that $\widehat{\mathbf{x}}_t(\theta)$ is real-valued. Obviously, $\widehat{\mathbf{x}}_t(\theta)$ coincides with $\widetilde{\mathbf{x}}_t(\theta)$ whenever the constraint to D is non-binding.

Instead of problem (3.3), the true parameter vector θ_0 and the associated state vectors $\{\mathbf{x}_t\}_{t\in\mathbb{T}}$ can therefore be estimated from the relaxation

$$(\widehat{\theta}, \{\widehat{\mathbf{x}}_t\}_{t\in\mathbb{T}}) := \underset{\theta\in\Theta,\, \{\mathbf{z}_t\}_{t\in\mathbb{T}}\subset\mathbb{R}^d}{\operatorname{argmin}}\ Q_T(\theta, \{\mathbf{z}_t\}_{t\in\mathbb{T}})\,. \tag{3.8}$$

[19]Here and throughout, for a $\mathbb{C}^{p\times d}$-valued function $\mathbf{F}$ and a $\mathbb{C}^{p\times k}$-valued function $\mathbf{G}$, elements of which are in $L_1^2(\pi)$, we extend the inner product notation as $\langle\mathbf{F}, \mathbf{G}\rangle = \int \mathbf{G}^{\mathsf{H}}(u)\mathbf{F}(u)\pi(du)$, yielding a $\mathbb{C}^{k\times d}$ matrix. Writing $\mathbf{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_d)$ and $\mathbf{G} = (\mathbf{g}_1, \ldots, \mathbf{g}_k)$ with $\mathbf{f}_i, \mathbf{g}_i \in L_p^2(\pi)$, the (i, j)-element of $\langle\mathbf{F}, \mathbf{G}\rangle$ is specifically given by $\langle\mathbf{f}_j, \mathbf{g}_i\rangle$.

Concentrating out state estimates $\widehat{\mathbf{x}}_t(\theta)$ according to problem (3.6), parameters and states can be equivalently estimated from

$$\widehat{\theta} := \underset{\theta\in\Theta}{\operatorname{argmin}}\; Q_T(\theta, \{\widehat{\mathbf{x}}_t(\theta)\}_{t\in\mathbb{T}}) \tag{3.9}$$

with $\widehat{\mathbf{x}}_t := \widehat{\mathbf{x}}_t(\widehat{\theta})$ for all $t \in \mathbb{T}$.

Provided correct model specification and the identification of the parameter vector from the panel of log CCFs, the relaxation from problem (3.3) to (3.8) will be asymptotically irrelevant. To ensure identification of the parameter vector, it suffices to impose the following weak assumption on the behavior of a noise-free version of the objective function in problem (3.9) after concentrating out optimal state estimates:

Assumption 8. *For every $\varepsilon > 0$ and $\theta \in \Theta$, we have*

$$\inf_{\|\theta-\theta_0\|>\varepsilon} \sum_{t\in\mathbb{T}} w_t \|\boldsymbol{\psi}_t(\theta_0) - \boldsymbol{\alpha}_t(\theta) - \boldsymbol{\beta}_t(\theta)\mathbf{x}_t(\theta)\|^2 > 0 \;\; a.s.,$$

where $\mathbf{x}_t(\theta) := \langle \boldsymbol{\beta}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle^{-1}\langle \boldsymbol{\psi}_t(\theta_0) - \boldsymbol{\alpha}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle$.

Given the outlined estimation procedure and the asymptotic setup, we can now state the following formal consistency result. Its proof is given in Appendix B.3.

Proposition 2. *Suppose Assumptions 1–8 hold. Then, as $n \to \infty$, we have that $\widehat{\theta} \xrightarrow{\mathbb{P}} \theta_0$ and $\widehat{\mathbf{x}}_t \xrightarrow{\mathbb{P}} \mathbf{x}_t$ for each $t \in \mathbb{T}$.*

Proposition 2 establishes that both the true time-invariant parameter vector θ_0 and the state vectors $\mathbf{x}_t$ can be consistently estimated by solving the non-linear functional least-squares problem (3.8), concentrating out latent states. The result is analogous to solving a non-linear least-squares problem to minimize the distance between observed and model-implied option prices (or implied volatility), as discussed in Section 2.

In addition, we obtain an $\mathcal{F}^X$-stable CLT for the joint asymptotic distribution of the state vector estimators $\widehat{\mathbf{x}}_t$ at each point in time $t \in \mathbb{T}$ and of the time-invariant parameter vector estimator $\widehat{\theta}$. The proof is given in Appendix B.4.

Proposition 3. *Suppose Assumptions 1–8 hold and $(\underline{\alpha} \wedge \overline{\alpha}) > 1/(2(\underline{q} \wedge (1 + \overline{q})))$. Then, as $n \to \infty$, we have*

$$\Delta_n^{-1/2} \begin{pmatrix} \widehat{\mathbf{x}}_1 - \mathbf{x}_1 \\ \vdots \\ \widehat{\mathbf{x}}_T - \mathbf{x}_T \\ \widehat{\theta} - \theta_0 \end{pmatrix} \xrightarrow{\mathcal{F}^X\text{-}s} \mathcal{N}\big(0, A_T^{-1} B_T A_T^{-\top}\big), \tag{3.10}$$

where the $\mathcal{F}^X$-measurable, real-valued matrices A_T and B_T are defined in equations (B.33) and (B.34) in Appendix B.4.

The limiting distribution established in Proposition 3 is mixed-Gaussian with an $\mathcal{F}^X$-measurable covariance matrix $A_T^{-1} B_T A_T^{-\intercal}$ that depends on the realization of the path of the state vector $\mathbf{x}_t$. The mixing result implies that the precision in estimating the state vector is itself random, depending on the information contained in the state vector and option prices. For instance, high volatility days are often associated with increased variance in option errors, leading to noisier state estimates.

The result in Proposition 3 is reminiscent of Theorem 2 in AFT with the main difference in the form of the mixing matrix. As displayed in equations (B.33) and (B.34), in our case, the elements are given by the inner product of functions in the Hilbert space.

As the matrices A_T and B_T that characterize the asymptotic distribution in Proposition 3 depend on the true parameter θ_0 and state vectors $\mathbf{x}_t$ as well as the unspecified measurement error variance $\sigma_t^2(\tau, m)$, the result is infeasible for practical implementation. To obtain a feasible version of Proposition 3 in terms of corresponding matrices $\widehat{A}_T$ and $\widehat{B}_T$, we essentially replace the true parameter and state vectors by their estimated counterparts, $\widehat{\theta}$ and $\widehat{\mathbf{x}}_t$, respectively. Moreover, we use a consistent estimator for $\sigma_t^2(\tau, m)$, constructed by squaring the feasible option errors $\widehat{\zeta}_t(\tau, m) := O(\tau, m; \widehat{\theta}, \widehat{\mathbf{x}}_t) - \widehat{O}_t(\tau, m)$. This construction is valid under sufficiently regular behavior of the model-implied option prices. To formulate the regularity conditions, we denote by $\mathfrak{m}(q, \tau; \theta, \mathbf{z})$ the model-implied moment-generating function (MGF) such that $\mathfrak{m}(q, \tau; \theta_0, \mathbf{x}_t) = \mathbb{E}^{\mathbb{Q}}[(F_{t+\tau}/F_t)^q \mid \mathcal{F}_t]$ holds.

Assumption 9. *(i) $O(m, \tau; \theta, \mathbf{z})$ is continuous in θ and $\mathbf{z}$ for all $m \in \mathbb{R}$ and $\tau \leq \overline{T}$;*

(ii) There exists an $\varepsilon > 0$ such that for all $\theta \in \Theta$ with $\|\theta - \theta_0\| < \varepsilon$ and $\mathbf{z}, \widetilde{\mathbf{z}} \in \mathbb{R}^d$ with $\|\widetilde{\mathbf{z}} - \mathbf{z}\| < \varepsilon$ the following holds: $\mathfrak{m}(-\underline{q}, \tau; \theta, \widetilde{\mathbf{z}}) \leq B(\mathbf{z}) < \infty$ and $\mathfrak{m}(1 + \overline{q}, \tau; \theta, \widetilde{\mathbf{z}}) \leq B(\mathbf{z}) < \infty$ for all $\tau \leq \overline{T}$ as well as some $\overline{q} > 0$ and $\underline{q} > 0$.

Continuity of the model-implied option pricing function as in Assumption 9(i) is a standard assumption, which immediately implies the consistency of the fitted option prices $O(\tau, m; \widehat{\theta}, \widehat{\mathbf{x}}_t)$ pointwise for each τ and m (as well as uniformly on each compact subset of $\Theta \times \mathbb{R}^d$). Employing the usual transform-based option pricing formulas, this form of continuity may be related to sufficiently smooth behavior of the coefficient functions α and β of the (unscaled) CCF. In addition, Assumption 9(ii) imposes locally uniform bounds on the model-implied MGF in a neighborhood of the true parameter vector θ_0 and any possible state vector. As a model-based counterpart to Assumption 1(ii), this assumption regulates the tail behavior of the model-implied option pricing function, such that the consistency also extends to certain integrals involving fitted option prices. Without loss of generality, we may take $\overline{q}$ and $\underline{q}$ identical to Assumption 1(ii).

With the additional regularity conditions in Assumption 9, we obtain that $\widehat{A}_T \xrightarrow{\mathbb{P}} A_T$ and $\widehat{B}_T \xrightarrow{\mathbb{P}} B_T$, leading to the following feasible version of Proposition 3. The proof is given in Appendix B.5.

Proposition 4. *Suppose Assumptions 1–9 hold and* $(\underline{\alpha} \wedge \overline{\alpha}) > 1/(2(\underline{q} \wedge (1+\overline{q})))$. *Then, as* $n \to \infty$, *we have*

$$\Delta_n^{-1/2} \widehat{B}_T^{-1/2} \widehat{A}_T \begin{pmatrix} \widehat{\mathbf{x}}_1 - \mathbf{x}_1 \\ \vdots \\ \widehat{\mathbf{x}}_T - \mathbf{x}_T \\ \widehat{\theta} - \theta_0 \end{pmatrix} \xrightarrow{\mathcal{F}^X\text{-}s} \mathcal{N}(0, I), \tag{3.11}$$

where the real-valued matrices $\widehat{A}_T$ *and* $\widehat{B}_T$ *are consistent feasible estimators of* A_T *and* B_T*, respectively, defined in equations (B.35) and (B.36) in Appendix B.5.*

We conclude this section by noting again that researchers have the flexibility to specify an appropriate measure π. While our exposition focuses on the more general case of continuous functional objects, the proposed estimation procedure and the derived asymptotic results apply equally to the discretized versions of these objects. For instance, in a similar framework, Freire and Vladimirov (2023) compare a fit of parametric models to non-parametric approaches such as PCA and autoencoders based on option-implied log CCFs evaluated at a discrete set of arguments. The results established in this paper provide a theoretical foundation for the estimation technique employed in that study, further highlighting the flexibility and applicability of our framework.

4 Simulations

4.1 One-factor model

As a starting point, we consider a one-factor option pricing model with stochastic volatility, a Gaussian jump size distribution in returns, and exponential co-jumps in volatility. The likelihood of jumps is additionally made stochastic and proportional to the stochastic variance. In particular, we assume the following process, referred to in shorthand as 'SVCJ', for the forward price F_t and state $\mathbf{x}_t = v_t$ under the risk-neutral probability measures:

$$\frac{dF_t}{F_t} = \sqrt{v_t} dW_{1,t} + \int_{\mathbb{R}^2} (e^x - 1)\, \widetilde{\mu}(dt, dx, dy), \tag{4.1}$$

$$dv_t = \kappa(\bar{v} - v_t)dt + \sigma\sqrt{v_t} dW_{2,t} + \int_{\mathbb{R}^2} y\, \mu(dt, dx, dy), \tag{4.2}$$

where two Brownian motions $W_{1,t}$ and $W_{2,t}$ are assumed to be correlated with the coefficient ρ, and the compensator for the jump measure is of the form $\widetilde{\nu}_t(dt, dx, dy) = \lambda(\mathbf{x}_t)dt \otimes \nu(dx, dy)$ with jump size measure

$$\nu(dx, dy) = \left\{ \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(x - \mu_j - \rho_j y)^2}{2\sigma_j^2} \right) \frac{1}{\mu_v} \exp\left(-\frac{y}{\mu_v} \right) \mathbb{1}_{\{y>0\}} \right\} dx \otimes dy, \tag{4.3}$$

and jump intensity $\lambda(\mathbf{x}_t) = \delta v_t$. The model specification has nine parameters that are collected in the parameter vector $\theta = (\kappa, \bar{v}, \sigma, \rho, \delta, \mu_j, \sigma_j, \mu_v, \rho_j)^\intercal$.

Although the model in equations (4.1) and (4.2) is a one-factor option pricing model, it exhibits all main features of option pricing models: stochastic volatility, (co-)jumps in returns and volatility, time-varying stochastic jump intensity, and self-excitation. Furthermore, it embeds many popular one-factor option pricing models as special cases, such as the Heston (1993), Pan (2002), and Bates (1996) models.

Importantly, the SVCJ model in equations (4.1) and (4.2) belongs to the class of AJD models. Hence, it exhibits an affine functional dependence of the log CCF $\psi_t(u, \tau)$ on the state vector $\mathbf{x}_t = v_t$:

$$\psi_t(u, \tau) = \alpha\left(\tfrac{u}{\sqrt{\tau}\kappa_{t,\tau}}, \tau; \theta\right) + \beta\left(\tfrac{u}{\sqrt{\tau}\kappa_{t,\tau}}, \tau; \theta\right) v_t,$$

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

$$\begin{cases} \dot{\alpha}(u,t) &= \kappa\bar{v}\beta(u,t), \\ \dot{\beta}(u,t) &= -\mathrm{i}u\left(\frac{1}{2} + \delta\left(\chi(1,0) - 1\right)\right) - \kappa\beta(u,t) - \frac{u^2}{2} + \mathrm{i}u\rho\sigma\beta(u,t) + \frac{1}{2}\sigma^2\beta^2(u,t) \\ & \quad + \delta\left(\chi(\mathrm{i}u, \beta(u,t)) - 1\right), \end{cases}$$

with initial conditions $\alpha(u, 0) = \beta(u, 0) = 0$ and 'jump transform' of the form

$$\chi(c_1, c_2) = \frac{\exp\left(\mu_j c_1 + \frac{1}{2}\sigma_j^2 c_1^2\right)}{1 - \mu_v c_2 - \rho_j \mu_v c_1}. \tag{4.4}$$

Our interest is in estimating nine risk-neutral parameters of the SVCJ model. Therefore, we ignore a possibly different dynamic under the physical measure, and simulate $T = 504$ time points with $\Delta t = 1/252$ from the same risk-neutral specification (4.1)–(4.3) using an Euler scheme. We fix the model parameters to the estimates reported in Table B.6 of AFT.

Given the simulated paths of log prices and spot volatility, we generate the option data using the COS method of Fang and Oosterlee (2009). In particular, at each point in time t, we simulate options with four maturities of 1, 3, 6, and 12 months, and equidistant strike prices with $\Delta K = 0.01 \times F_t$. We keep the range of strike prices for each maturity sufficiently wide such that the truncation errors in the option-implied CCFs are minimal. The generated option prices are distorted by adding observation errors on the corresponding Black-Scholes implied volatilities (BSIV), i.e.

$$\widehat{\kappa}_t(\tau, m) = \kappa_t(\tau, m) + 0.001 \times \kappa_t(\tau, m) \times \epsilon_t(\tau, m),$$

where $\epsilon_t(\tau, m)$ are i.i.d. standard normal random variables. The noisy BSIVs are then interpolated using a cubic spline to obtain observations with a finer grid of strikes, and then converted to obtain option prices $\widehat{O}_t(\tau, m)$. Finally, the option-implied CCFs are constructed using the Riemann sum approximation as in equation (2.4).

As discussed in Section 3, we work with a class of functions in $L_p^2(\pi)$ that are square-integrable with respect to a finite measure π. We choose $\pi(du) = \phi(u; s)du$ to be generated by a Gaussian PDF $\phi(u; s)$ with zero mean and variance s^2, although any other (symmetric) density function could also be utilized. This choice of the density function allows us to approximate integrals using Gauss-Hermite quadrature, as described in Appendix C. Since in our estimation procedure we would like to control the relative importance of information contained at different arguments of the log CCF, the scale parameter s plays the role of a tuning parameter. In particular, it allows downweighting function values evaluated at larger arguments u with lower signal-to-noise ratio. Therefore, first, we investigate the role of this parameter on the estimation results. Additionally, we explore the robustness to the choice of the quadrature order.

In particular, we run 100 simulations for different levels of s and quadrature orders using the same parameter values for the model as we use in the following Monte Carlo exercise. We assess the optimal level of s via four metrics. First, we look at the fit of option prices by considering the Monte Carlo root mean square error (RMSE) and the root mean square percentage error (RMSPE) of implied volatilities. Although our estimation is conducted using log CCFs and does not require the explicit evaluation of option prices, we can nevertheless assess the pricing accuracy given the parameter and state estimates of the parametric model. Importantly, these metrics can be easily employed in practice to choose the tuning parameters and are commonly used in the literature. Next, to assess the relative accuracy of parameter estimates for a given scale s, we construct the RMSPE for model parameters. Finally, we compute the RMSE between the estimated and the true state variables, which in case of this model is stochastic volatility.

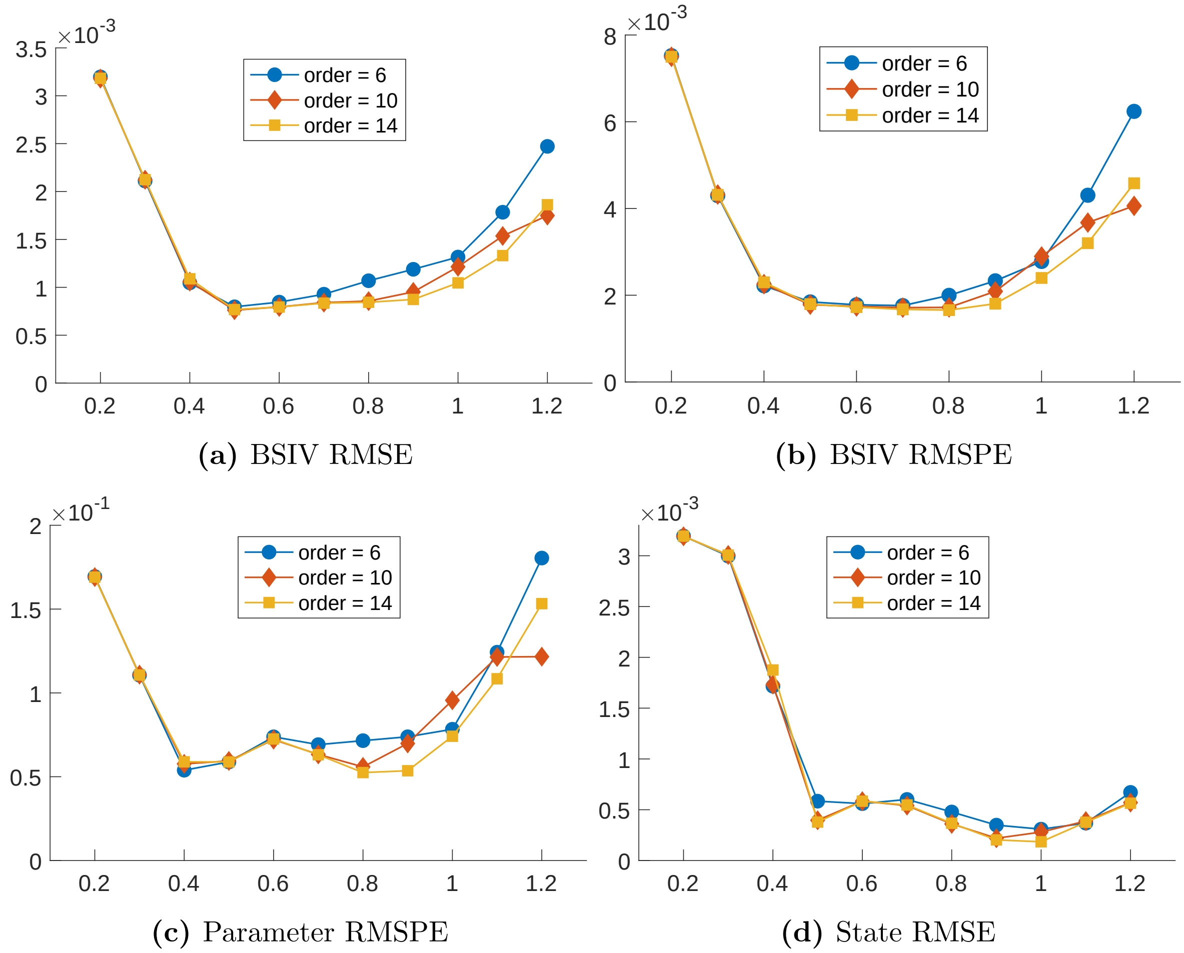
Figure 1 plots the four metrics for different levels of s and three quadrature orders. As we can see, the metrics are generally U-shaped and indicate the optimal range of the scale level between $s = 0.5$ and $s = 0.9$. Importantly, deviations from this range lead only to marginal increase in all four measures. The results are also robust to the choice of the order in the Gauss-Hermite quadrature rule. In the following simulation exercise and empirical application, we consider $s = 0.7$ and the quadrature order of 10.

Table 1 provides the Monte Carlo results for the SVCJ model parameters based on $N = 1000$ replications. The simulation results show a very good finite-sample performance for all parameters of the model as well as the state estimation. We emphasize that we ran the simulation on a regular laptop, with each iteration taking less than a half minute.

4.2 Two-factor model

As an extension of the SVCJ model in Section 4.1, we now consider a two-factor option pricing model in which we have an additional factor for the jump intensity, which follows

Figure 1: Option, parameter, and state vector fit for different levels of s



Note: This figure plots the RMSE and RMSPE for BSIV, RMSPE for parameters, and RMSE for state vector for different values of volatility s and three different quadrature orders. The parameters for the SVCJ model are the same as in Table 1.

a Hawkes process. Specifically, for the model that is henceforth referred to as 'SVHJ', we assume the following dynamics for the forward price F_t and state $\mathbf{x}_t = (v_t, \lambda_t)^\intercal$ under $\mathbb{Q}$:

$$\frac{dF_t}{F_t} = \sqrt{v_t} dW_{1,t} + \int_{\mathbb{R}^3} (e^x - 1)\,\widetilde{\mu}(dt, dx, dy, dz), \tag{4.5}$$

$$dv_t = \kappa(\bar{v} - v_t)dt + \sigma\sqrt{v_t} dW_{2,t} + \int_{\mathbb{R}^3} y\,\mu(dt, dx, dy, dz), \tag{4.6}$$

$$d\lambda_t = \kappa_\lambda(\bar{\lambda} - \lambda_t)dt + \int_{\mathbb{R}^3} z\,\mu(dt, dx, dy, dz). \tag{4.7}$$

The compensator for the jump measure is of the form $\widetilde{\nu}_t(dt, dx, dy, dz) = \lambda(\mathbf{x}_t)dt \otimes \nu(dx, dy, dz)$ with jump intensity $\lambda(\mathbf{x}_t) = \lambda_t$ and jump size measure

$$\nu(dx, dy, dz) = \Bigg\{ \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(x - \mu_j - \rho_j y)^2}{2\sigma_j^2}\right) \times \frac{1}{\mu_v} \exp\left(-\frac{y}{\mu_v}\right) \mathbb{1}_{\{y>0\}} \times \frac{1}{\mu_\lambda} \exp\left(-\frac{z}{\mu_\lambda}\right) \mathbb{1}_{\{z>0\}} \Bigg\} dx \otimes dy \otimes dz.$$

Table 1: Monte Carlo results for the SVCJ model

	κ	$\bar{v}$	σ	ρ	δ	μ_j	σ_j	μ_v	ρ_j	ME_v
true	0.6860	0.0280	0.1540	-0.9000	20.510	-0.0760	0.1380	0.0250	-0.6210	0.00000
mean	0.6933	0.0274	0.1508	-0.9142	20.625	-0.0758	0.1376	0.0253	-0.6169	0.00013
std	0.0082	0.0010	0.0045	0.0200	0.3050	0.0029	0.0010	0.0003	0.0971	0.00043
q10	0.6853	0.0263	0.1451	-0.9374	20.487	-0.0778	0.1367	0.0250	-0.6902	-0.00004
q50	0.6909	0.0278	0.1527	-0.9064	20.516	-0.0764	0.1380	0.0253	-0.6036	0.00000
q90	0.7053	0.0280	0.1541	-0.8994	20.931	-0.0732	0.1381	0.0259	-0.5387	0.00045

Note: This table provides Monte Carlo simulation results for the SVCJ model. For each parameter, we report the true value, the Monte Carlo mean and standard deviation, and the 10th, 50th and 90th Monte Carlo percentiles. The last column reports the same descriptive statistics for the mean errors of the estimated state (ME_v).

The first two processes of the SVHJ model are the same as in the one-factor specification. The third process introduces a Hawkes jump intensity based on an exponential kernel. This additional factor decouples the dynamic of the stochastic jump intensity from the variance process, further enriching the model's flexibility to capture the clustering of jumps. Like the SVCJ model, the two-factor specification also belongs to the class of AJD models and can be seen as an extension of the models considered in Boswijk et al. (2016), Du and Luo (2019), and a univariate version of the model in Boswijk et al. (2023). The affine functional dependence of the log CCF $\psi_t(u,\tau)$ on the state vector can be written as

$$\psi_t(u,\tau) = \alpha\left(\frac{u}{\sqrt{\tau}\kappa_{t,\tau}},\tau;\theta\right) + \beta_1\left(\frac{u}{\sqrt{\tau}\kappa_{t,\tau}},\tau;\theta\right)v_t + \beta_2\left(\frac{u}{\sqrt{\tau}\kappa_{t,\tau}},\tau;\theta\right)\lambda_t,$$

where the functional coefficients are given as solutions to the complex-valued ODE system:

$$\begin{cases} \dot{\alpha}(u,t) &= \kappa\bar{v}\beta_1(u,t) + \kappa_\lambda\bar{\lambda}\beta_2(u,t), \\ \dot{\beta}_1(u,t) &= -\frac{\mathrm{i}u}{2} - \kappa\beta_1(u,t) - \frac{u^2}{2} + \mathrm{i}u\rho\sigma\beta_1(u,t) + \frac{1}{2}\sigma^2\beta_1^2(u,t), \\ \dot{\beta}_2(u,t) &= -\mathrm{i}u\left(\chi(1,0,0)-1\right) - \kappa_\lambda\beta_2(u,t) + \left(\chi(\mathrm{i}u,\beta_1(u,t),\beta_2(u,t))-1\right) \end{cases}$$

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

$$\chi(c_1,c_2,c_3) = \frac{\exp\left(\mu_j c_1 + \frac{1}{2}\sigma_j^2 c_1^2\right)}{1-\mu_v c_2 - \rho_j\mu_v c_1} \times \frac{1}{1-\mu_\lambda c_3}. \tag{4.8}$$

The simulation setup mirrors that of the one-factor specification, with the same parameter values for the underlying process and stochastic volatility. The parameters for the jump intensity process are taken from the univariate result of the model in Boswijk et al. (2023).

The Monte Carlo results for the SVHJ model are provided in Table 2. Like the results for the SVCJ model, the parameters of the two-factor model, including those related to the Hawkes jump intensity, exhibit good finite sample performance. We also observe that both state variables – the stochastic variance and the jump intensity – are estimated quite accurately.

Table 2: Monte Carlo results for the two-factor SVHJ model

	κ	$\bar{v}$	σ	ρ	μ_j	σ_j	
true	0.6860	0.0280	0.1540	-0.9000	-0.0760	0.1380	
mean	0.6833	0.0295	0.1646	-0.8714	-0.0773	0.1381	
std	0.0126	0.0062	0.0435	0.1073	0.0041	0.0009	
q10	0.6829	0.0279	0.1531	-0.9064	-0.0833	0.1375	
q50	0.6858	0.0281	0.1540	-0.9005	-0.0767	0.1381	
q90	0.6877	0.0321	0.1607	-0.8751	-0.0742	0.1390	
	μ_v	ρ_j	κ_λ	$\bar{\lambda}$	μ_λ	ME_v	ME_λ
true	0.0250	-0.6210	2.4450	0.3050	2.1760	0.0000	0.0000
mean	0.0247	-0.5841	2.4362	0.2927	2.1796	-0.0003	0.0214
std	0.0014	0.1629	0.0866	0.0412	0.0855	0.0052	0.4341
q10	0.0231	-0.8935	2.4300	0.2779	2.1631	-0.0002	-0.0639
q50	0.0247	-0.5835	2.4472	0.3033	2.1797	0.0000	0.0002
q90	0.0259	-0.3781	2.4698	0.3057	2.2405	0.0007	0.0292

Note: This table provides Monte Carlo simulation results for the two-factor SVHJ model. For each parameter, we report the true value, the Monte Carlo mean and standard deviation, and the 10th, 50th and 90th Monte Carlo percentiles. The last column reports the same descriptive statistics for the mean errors of the estimated volatility and intensity states, ME_v and ME_λ, respectively.

5 Empirical Applications

5.1 Data

We illustrate the developed estimation procedure using the widely considered European-style options on the S&P 500 stock market index available through OptionMetrics. Leveraging the computational efficiency of our method, we analyze a rather large panel with daily observations covering the period from January 1996 to August 2023. To balance the growing amount of available maturities toward the end of the sample, we focus on AM-settled options with maturities between 7 days and 1 year. This limits the number of maturities per day to at most twelve, but the maturities and their number may vary from day to day in the constructed sample. We work with mid-quote option prices after applying standard filters that rule out illiquid observation with (*i*) zero bid quotes, (*ii*) bids exceeding ask quotes, (*iii*) ask-to-bid ratios exceeding a factor of 20, and (*iv*) distance between adjacent strike prices larger than \$300.

Since the inputs of our estimation procedures are log CCFs constructed via portfolio of options, we need a reliable and relatively wide coverage of option strikes at each maturity. This is also in line with the large-n asymptotic scheme of our estimation method. For this reason, we retain option observations that satisfy the following criteria per maturity: (*i*) the number of distinct strike prices is larger than 15, (*ii*) the number of put-call pairs is at least 5, (*iii*) the number of OTM calls and OTM puts is at least 3, (*iv*) the largest standardized moneyness exceeds 1 and the smallest standardized moneyness is below -3, and (v) there is at least one other maturity on the same day for which option observations satisfy the above four criteria. This results in a selection of 4,741,456 contracts covering 6,853 trading days, which is by far the largest option panel, both in the time-series and cross-sectional dimensions, considered in the related estimation literature.

To reduce the impact of the truncation and discretization errors in the option-implied log CCF approximation, we employ an interpolation-extrapolation scheme following BLV. Specifically, we interpolate option prices expressed in terms of their Black-Scholes implied variances using cubic splines with carefully selected knot sequences. The latter are chosen such that the option observations at the knots satisfy standard no-arbitrage conditions and their ask-to-bid ratios are smaller than 5. We also extrapolate the implied variances beyond the observable strike range linearly in log-moneyness m, consistent with the asymptotic behavior established in Lee (2004). For further details on the interpolation-extrapolation scheme, we refer to Appendix C.1 in BLV. The option-implied log CCFs for each day and for each maturity are constructed using the Riemann sum approximation in equation (2.4) applied to the result of the interpolation-extrapolation scheme.

5.2 Empirical results

Using the constructed log CCF panel, we estimate the one- and two-factor parametric option pricing models described in Section 4. As discussed there, we employ a Gauss-Hermite quadrature rule of order 10 to evaluate inner products and choose $\pi(du) = \phi(u;s)du$ using a Gaussian PDF $\phi(u;s)$ with zero mean and scale $s = 0.7$. The standard errors for the parameter estimates are obtained using the feasible asymptotic distribution result established in Proposition 4.

The results for the one-factor SVCJ model are reported in Table 3. The parameter estimates of the SVCJ model are reasonable and broadly in line with those found in the related literature. The jump correlation parameter ρ_j is estimated close to -1, indicating an almost perfect correlation between jump sizes in volatility and returns. Such a feature is often incorporated in the models with exponential jump sizes in returns, where the variance jump sizes are made proportional to the squared (negative) jump sizes in returns (see, e.g., AFT, Andersen et al., 2015b, 2020).

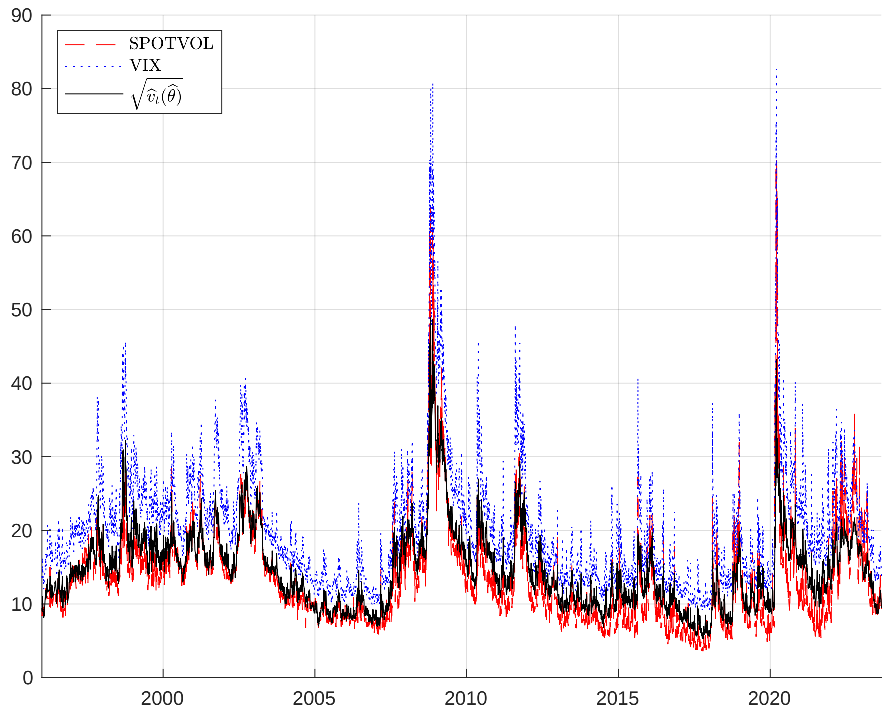
Table 3: Parameter estimates of the SVCJ model

	κ	$\bar{v}$	σ	ρ	δ	μ_j	σ_j	μ_v	ρ_j
$\widehat{\theta}$	2.1704	0.0146	0.2213	-0.5823	46.7497	-0.0827	0.0105	0.0415	-0.9990
s.e.	0.0436	0.0003	0.0039	0.0038	1.7531	0.0023	0.0055	0.0007	0.0222

Note: This table reports the parameter estimates and the standard errors for the one-factor SVCJ model.

Figure 2 plots the estimated volatility from the SVCJ model along with the non-parametric spot volatility measure[20] of Todorov (2019) and the VIX index. We notice that our estimates of the state exhibit a similar time-series pattern as the other two measures of volatility, although they are obtained on each day separately. The estimated volatility also lies below the VIX index and is close to the non-parametric spot volatility measure. This indicates that the volatility estimates obtained from the SVCJ model effectively capture the underlying volatility dynamics in the market.

Figure 2: Estimated states from the SVCJ model



Note: This figure plots the square root of the estimated states from the SVCJ model along with the non-parametric spot volatility measure of Todorov (2019) and the VIX index.

The results for the two-factor model are reported in Table 4. The estimates related to the stochastic volatility factor are generally close to those obtained for the one-factor specification. The difference between some of the parameters, e.g., the long-run volatility $\bar{v}$ and the volatility mean jump size μ_v, can be attributed to the inclusion of a separate stochastic process for the jump intensity. The parameter estimates for the jump intensity are similar to those obtained in a univariate specification for the SPX options in Boswijk et al. (2023). In particular, every new jump arrival increases the jump intensity, on average, by a factor of $\mu_\lambda = 2.085$, which then mean-reverts to the base intensity level $\bar{\lambda} = 0.091$. This translates into an unconditional expected jump intensity of around 0.45, i.e., the market expects jumps to occur on average once in just over two years.

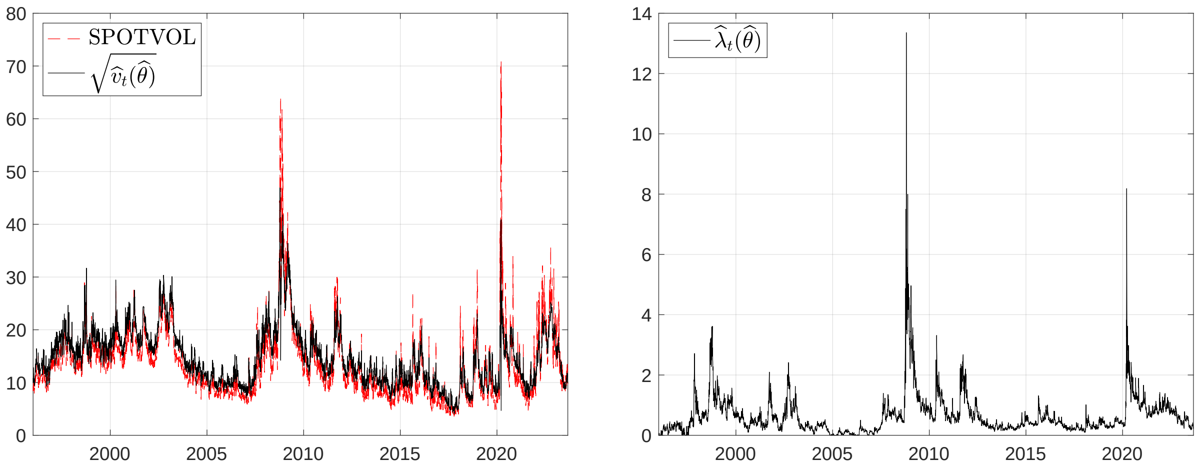
[20]We are thankful to Viktor Todorov for kindly sharing this data.

Table 4: Parameter estimates of the SVHJ model

	κ	$\bar{v}$	σ	ρ	μ_j	σ_j	μ_v	ρ_j	κ_λ	$\bar{\lambda}$	μ_λ
$\widehat{\theta}$	1.322	0.0347	0.394	-0.615	-0.094	0.0001	0.072	-0.999	2.613	0.091	2.085
s.e.	0.016	0.0004	0.002	0.002	0.003	0.5570	0.004	0.062	0.048	0.003	0.066

Note: This table reports the parameter estimates and the standard errors for the two-factor SVHJ model.

Figure 3: Estimated states from the SVHJ model



Note: This figure plots the estimated states from the two-factor SVHJ model. In particular, the left panel displays the square root of the estimated volatility along with the non-parametric spot volatility measure of Todorov (2019) and the right panel plots the estimated jump intensity.

Figure 3 illustrates the estimated states from the SVHJ model. The left panel plots the estimated volatility states alongside the non-parametric spot volatility measure of Todorov (2019), while the right figure displays the estimates of the jump intensity. The volatility estimates from the two-factor specification are even more aligned with the non-parametric spot volatility than the estimates from the one-factor model. Notably, the time-series of the jump intensity estimates resembles a path of a typical self-exciting process: it quickly increases during the turbulent events such as the global financial crisis in 2008 and the covid-19 pandemic, and then it subsequently mean-reverts to its base level.

6 Conclusion

This paper proposes a novel risk-neutral estimation procedure for parametric option pricing models. Using an observed option panel, our procedure minimizes the distance between two functions across different maturities and dates: the logarithm of the option-implied conditional characteristic function (CCF) of (standardized) log underlying returns, which can be approximated from a portfolio of observed option prices, and the model-implied counterpart. Within the marginal-affine class of models, for which the CCF is exponentially affine in the latent state vector, we can concentrate out the state vectors in closed form by solving a linear functional least-squares problem on each date in the panel. This allows us to optimize only over the model's parameter space, circumventing the typical computational costs when estimating parametric option pricing models and avoiding the need to downsample the available option data prior to the estimation.

Although, in this paper, we have focused on the marginal-affine class of models, which admits closed-form estimates for latent state vectors, a similar CCF-based estimation framework can be applied to more general non-affine parametric models. However, the latter still requires tractability of the parametric CCF, but also solving a non-linear least-squares problem numerically. The proposed procedure can be further extended to include economic regularization as in Andersen et al. (2015a), when needed, and can be incorporated into full estimation methods. The latter we intend to explore in future research.

Appendices

A Limit theorems in a Hilbert space

Throughout, consider an $\mathcal{F}$-measurable, $L_p^2(\pi)$-valued random sequence $(\mathbf{f}_n)_{n\in\mathbb{N}}$ with limit $\mathbf{f}$. For our purposes, we are interested exclusively in the case where $\mathbf{f}$ follows a complex Gaussian distribution $\mathcal{N}(0, K, S)$ with covariance operator K and relation operator S.

For convergence in distribution, we say that $\mathbf{f}_n \xrightarrow{d} \mathbf{f}$ in $L_p^2(\pi)$ if $\mathbb{E}[\phi(\mathbf{f}_n)] \to \mathbb{E}[\phi(\mathbf{f})]$ for all $\phi \in C_b(L_p^2(\pi))$, the set of all (complex-valued) bounded and continuous functionals $\phi : L_p^2(\pi) \to \mathbb{C}$. Equivalently, we write $\mathbf{f}_n \xrightarrow{d} \mathcal{N}(0, K, S)$. By Theorem 1.8.4 of van der Vaart and Wellner (2023) (see also the discussion in X. Chen and White, 1998), convergence in distribution in $L_p^2(\pi)$ is characterized by tightness of the sequence and convergence in distribution of the marginals.

Proposition A.1. *The following are equivalent:*

(i) $\mathbf{f}_n \xrightarrow{d} \mathcal{N}(0, K, S)$ *in* $L_p^2(\pi)$*;*

(ii) $(\mathbf{f}_n)_{n\in\mathbb{N}}$ *is tight and* $\langle \mathbf{f}_n, \mathbf{h}\rangle \xrightarrow{d} \mathcal{N}(0, \langle \mathbf{h}, K\mathbf{h}\rangle, \langle \mathbf{h}, S\mathbf{h}\rangle)$ *for all* $\mathbf{h} \in L_p^2(\pi)$.

For the stronger notion of $\mathcal{G}$-stable convergence given some sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, we say that $\mathbf{f}_n \xrightarrow{\mathcal{G}\text{-}s} \mathbf{f}$ in $L_p^2(\pi)$ if $\mathbb{E}[Y\phi(\mathbf{f}_n)] \to \mathbb{E}[Y\phi(\mathbf{f})]$ for all $\phi \in C_b(L_p^2(\pi))$ and every $Y \in L^\infty(\mathcal{G})$, the set of (complex-valued) bounded $\mathcal{G}$-measurable random variables. Equivalently, we write $\mathbf{f}_n \xrightarrow{\mathcal{G}\text{-}s} \mathcal{N}(0, K, S)$. Extending Proposition A.1, stable convergence in $L_p^2(\pi)$ is likewise characterized by tightness of the sequence and stable convergence of the marginals.

Proposition A.2. *The following are equivalent:*

(i) $\mathbf{f}_n \xrightarrow{\mathcal{G}\text{-}s} \mathcal{N}(0, K, S)$ *in* $L_p^2(\pi)$*;*

(ii) $(\mathbf{f}_n)_{n\in\mathbb{N}}$ *is tight and* $\langle \mathbf{f}_n, \mathbf{h}\rangle \xrightarrow{\mathcal{G}\text{-}s} \mathcal{N}(0, \langle \mathbf{h}, K\mathbf{h}\rangle, \langle \mathbf{h}, S\mathbf{h}\rangle)$ *for all* $\mathbf{h} \in L_p^2(\pi)$.

Proof. Since the unit constant function $\mathbb{1} \in L_1^2(\pi)$, we may consider $g = Y\mathbb{1} \in L_1^2(\pi)$. The statement of the proposition will follow if, for $\widetilde{\mathbf{f}}_n := (\mathbf{f}_n; g)$ and $\widetilde{\mathbf{f}} := (\mathbf{f}; g)$, we can show that $\widetilde{\mathbf{f}}_n \xrightarrow{d} \widetilde{\mathbf{f}}$ in $L_{p+1}^2(\pi)$ is equivalent to both (i) and (ii).

For (ii), we use the equivalent characterization in Proposition A.1. Note that $(\mathbf{f}_n)_{n\in\mathbb{N}}$ is tight if and only if $(\widetilde{\mathbf{f}}_n)_{n\in\mathbb{N}}$ is tight, due to the boundedness of Y. Moreover, given any $\mathbf{h} \in L_p^2(\pi)$, from Proposition 3.12 in Häusler and Luschgy (2015), $\langle \mathbf{f}_n, \mathbf{h}\rangle \xrightarrow{\mathcal{G}\text{-}s} \langle \mathbf{f}, \mathbf{h}\rangle$ if and only if $(\langle \mathbf{f}_n, \mathbf{h}\rangle, Y) \xrightarrow{d} (\langle \mathbf{f}, \mathbf{h}\rangle, Y)$ for all $Y \in L^\infty(\mathcal{G})$. Recall that for any (scalar) random sequences (X_n, Y_n) with limit (X, Y), we have equivalence between $(X_n, Y_n) \xrightarrow{d} (X, Y)$ and $w_1 X_n + w_2 Y_n \xrightarrow{d} w_1 X + w_2 Y$ for all $w_1, w_2 \in \mathbb{C}$. Therefore, observing that $w_1\langle \mathbf{f}, \mathbf{h}\rangle + w_2 Y = \langle \widetilde{\mathbf{f}}, \widetilde{\mathbf{h}}\rangle$ with $\widetilde{\mathbf{h}} = (w_1\mathbf{h}; (w_2/\mu(\mathcal{U}))\mathbb{1}) \in L^2_{p+1}(\pi)$, it follows that $\langle \mathbf{f}_n, \mathbf{h}\rangle \xrightarrow{\mathcal{G}\text{-}s} \langle \mathbf{f}, \mathbf{h}\rangle$ for all $\mathbf{h} \in L^2_p(\pi)$ if and only if $\langle \widetilde{\mathbf{f}}_n, \widetilde{\mathbf{h}}\rangle \xrightarrow{d} \langle \widetilde{\mathbf{f}}, \widetilde{\mathbf{h}}\rangle$ for all $\widetilde{\mathbf{h}} \in L^2_{p+1}(\pi)$ and $Y \in L^\infty(\mathcal{G})$. In conclusion, $\widetilde{\mathbf{f}}_n \xrightarrow{d} \widetilde{\mathbf{f}}$ in $L^2_{p+1}(\pi)$ is equivalent to (ii).

For (i), note that for any $\phi \in C_b(L^2_p(\pi))$ and $Y \in L^\infty(\mathcal{G})$, choosing $\widetilde{\phi}(\widetilde{\mathbf{f}}) = \phi(\mathbf{f})\, g$ yields $\widetilde{\phi} \in C_b(L^2_{p+1}(\pi))$ and, due to the convergence in distribution, $\mathbb{E}[Y\,\phi(\mathbf{f}_n)] \to \mathbb{E}[Y\,\phi(\mathbf{f})]$. Hence, $\widetilde{\mathbf{f}}_n \xrightarrow{d} \widetilde{\mathbf{f}}$ in $L^2_{p+1}(\pi)$ immediately implies $\mathbf{f}_n \xrightarrow{\mathcal{G}\text{-}s} \mathbf{f}$ in $L^2_p(\pi)$. For the reverse direction, invoking a functional generalization of the Weierstrass Theorem (e.g., Bruno, 1984), we may find for any $\varepsilon > 0$ and $M > 0$ a uniform ε-approximation of any $\widetilde{\phi} \in C_b(L^2_{p+1}(\pi))$ such that

$$\left| \widetilde{\phi}(\widetilde{\mathbf{f}}) - \sum_{i=1}^{N} \rho_i(g)\,\phi_i(\mathbf{f}) \right| < \varepsilon \quad \text{for all } \|\mathbf{f}\| \le M,$$

where $\rho_i \in C_b(L^2_1(\pi))$ and $\phi_i \in C_b(L^2_p(\pi))$. In light of the tightness of $(\mathbf{f}_n)_{n\in\mathbb{N}}$ (by (ii)), choose $M > 0$ large enough so that $\mathbb{P}[\|f\| > M] < \varepsilon$ and $\mathbb{P}[\|\mathbf{f}_n\| > M] < \varepsilon$ for all n. Then, distinguishing the cases where $\|\mathbf{f}\| \le M$ and $\|\mathbf{f}\| > M$, we have

$$\left| \mathbb{E}\big[\widetilde{\phi}(\widetilde{\mathbf{f}})\big] - \mathbb{E}\left[\sum_{i=1}^{N} \rho_i(g)\,\phi_i(\mathbf{f})\right] \right| < \varepsilon(1 + M')$$

and likewise for each $\mathbf{f}_n$, uniformly for some $M' > 0$ that reflects the bounds of the functionals. Hence, noting that each $\rho_i(g) \in L^\infty(\mathcal{G})$, it follows that $\mathbb{E}[Y\,\phi(\mathbf{f}_n)] \to \mathbb{E}[Y\,\phi(\mathbf{f})]$ for all $\phi \in C_b(L^2_p(\pi))$ and $Y \in L^\infty(\mathcal{G})$ implies $\mathbb{E}[\widetilde{\phi}(\widetilde{\mathbf{f}}_n)] \to \mathbb{E}[\widetilde{\phi}(\widetilde{\mathbf{f}})]$ for all $\widetilde{\phi} \in C_b(L^2_{p+1}(\pi))$. Therefore, $\mathbf{f}_n \xrightarrow{\mathcal{G}\text{-}s} \mathbf{f}$ in $L^2_p(\pi)$ implies $\widetilde{\mathbf{f}}_n \xrightarrow{d} \widetilde{\mathbf{f}}$ in $L^2_{p+1}(\pi)$. In other words, $\widetilde{\mathbf{f}}_n \xrightarrow{d} \widetilde{\mathbf{f}}$ in $L^2_{p+1}(\pi)$ is also equivalent to (i). □

B Proofs of main results

B.1 Preliminary results

Lemma B.1. *Suppose Assumptions 1–3 and 4(i) hold. Then, for fixed $t \in \mathbb{T}$ and $\tau \in \mathcal{T}_t$, we have that*

$$\sup_{u\in\mathcal{U}} \left|\widehat{\varphi}_t(u,\tau) - \varphi_t(u,\tau)\right| = \mathcal{O}_{\mathbb{P}}\left(\sqrt{\frac{\log n}{n}} \vee n^{-(\underline{q}\underline{\alpha}\wedge(1+\overline{q})\overline{\alpha})}\right).$$

Proof. Since t and τ are fixed, for notational convenience let us denote throughout the proof $n := n_{t,\tau}$, $m_j := m_{t,\tau}(j)$, and $\Delta m_j := \Delta_{t,\tau}(j)$. We start by analyzing the errors in the option-implied CCF. Following BLV and Todorov (2019), the total measurement errors in the CCF approximation can be decomposed as $\widehat{\varphi}_t(u,\tau) - \varphi_t(u,\tau) = \sum_{i=1}^{3} \zeta_t^{(i)}(u,\tau)$, where $\zeta_t^{(i)}(u,\tau) := \mathfrak{u}_{t,\tau}(u)\, \widetilde{\zeta}_t^{(i)}(u,\tau)$ with $\mathfrak{u}_{t,\tau}(u) := \frac{u^2}{\tau\kappa_{t,\tau}^2} + \mathrm{i}\frac{u}{\sqrt{\tau}\kappa_{t,\tau}}$ and

$$\widetilde{\zeta}_t^{(1)}(u,\tau) := -\sum_{j=2}^{n} e^{(\mathrm{i}\frac{u}{\sqrt{\tau}\kappa_{t,\tau}}-1)m_j}\ \zeta_t(\tau,m_j)\,\Delta m_j, \tag{B.1}$$

$$\widetilde{\zeta}_t^{(2)}(u,\tau) := \int_{-\infty}^{m_1} e^{(\mathrm{i}\frac{u}{\sqrt{\tau}\kappa_{t,\tau}}-1)m}\ O_t(\tau,m)\,dm + \int_{m_n}^{\infty} e^{(\mathrm{i}\frac{u}{\sqrt{\tau}\kappa_{t,\tau}}-1)m}\ O_t(\tau,m)\,dm, \tag{B.2}$$

$$\widetilde{\zeta}_t^{(3)}(u,\tau) := \sum_{j=2}^{n}\int_{m_{j-1}}^{m_j}\Big(e^{(\mathrm{i}\frac{u}{\sqrt{\tau}\kappa_{t,\tau}}-1)m}\ O_t(\tau,m) - e^{(\mathrm{i}\frac{u}{\sqrt{\tau}\kappa_{t,\tau}}-1)m_j}\ O_t(\tau,m_j)\Big)\,dm. \tag{B.3}$$

The error terms $\zeta_t^{(1)}(u,\tau)$, $\zeta_t^{(2)}(u,\tau)$ and $\zeta_t^{(3)}(u,\tau)$ are referred to as observation, truncation and discretization errors, respectively.

For the observation errors, invoking Lemma 1 in BLV, we obtain

$$\begin{aligned}
\mathbb{E}\Big[|\widetilde{\zeta}_t^{(1)}(u,\tau)|^2 \mid \mathcal{F}^X\Big] &\le \sum_{j=2}^{n} e^{-2m_j}\ \mathbb{E}\big[\zeta_t(\tau,m_j)^2 \mid \mathcal{F}^X\big]\ (\Delta m_j)^2 \\
&\le \mathcal{O}_{\mathbb{P}}(1)\,\Delta_n \sum_{j=2}^{n} e^{-2m_j}\ O_t^2(\tau,m_j)\,\Delta m_j \\
&\le \mathcal{O}_{\mathbb{P}}(1)\,\Delta_n \sum_{j=2}^{n} e^{-2m_j}\ e^{2(-\overline{q}m_j\wedge(1+\underline{q})m_j)}\,\Delta m_j \\
&\le \mathcal{O}_{\mathbb{P}}(1)\,\Delta_n = o_{\mathbb{P}}(1).
\end{aligned}$$

The second inequality results from the measurement error specifications in Assumption 3, in particular Assumption 3(iii); for the third inequality, we use Lemma 1 in BLV together with Assumption 1(ii); the final inequality follows since under the employed asymptotic scheme the summation converges to a finite integral. Hence, $\widetilde{\zeta}_t^{(1)}(u,\tau) = \mathcal{O}_{\mathbb{P}}\Big(\sqrt{\frac{\log n}{n}}\Big) = \mathcal{O}_{\mathbb{P}}\big(\sqrt{\Delta_n}\big)$. Lemma 2 in BLV moreover provides the rates for the truncation and discretization errors: $\widetilde{\zeta}_t^{(2)}(u,\tau) = \mathcal{O}_{\mathbb{P}}\big(n^{-(\underline{q}\underline{\alpha}\wedge(1+\overline{q})\overline{\alpha})}\big)$ and $\widetilde{\zeta}_t^{(3)}(u,\tau) = \mathcal{O}_{\mathbb{P}}\big(\frac{\log n}{n}\big) = \mathcal{O}_{\mathbb{P}}(\Delta_n)$. Importantly, the rates of convergence do not depend on the argument u, yielding that each convergence is uniform on $\mathcal{U}$.

Since Assumption 4(i) implies that $\mathfrak{u}_{t,\tau}$ is bounded on the bounded set $\mathcal{U}$, it further follows from their definition that the same rates of convergence hold for $\zeta_t^{(i)}(u,\tau)$, where each convergence is again uniform on $\mathcal{U}$. Therefore, we obtain that

$$\begin{aligned}
\widehat{\varphi}_t(u,\tau) - \varphi_t(u,\tau) &= \mathcal{O}_{\mathbb{P}}\left(\sqrt{\frac{\log n}{n}}\right) + \mathcal{O}_{\mathbb{P}}\big(n^{-(\underline{q}\underline{\alpha}\wedge(1+\overline{q})\overline{\alpha})}\big) + \mathcal{O}_{\mathbb{P}}\left(\frac{\log n}{n}\right) \\
&= \mathcal{O}_{\mathbb{P}}\left(\sqrt{\frac{\log n}{n}} \vee n^{-(\underline{q}\underline{\alpha}\wedge(1+\overline{q})\overline{\alpha})}\right)
\end{aligned}$$

uniformly on $\mathcal{U}$. □

Lemma B.2. *Suppose Assumptions 1–4 hold with* $(\underline{\alpha} \wedge \overline{\alpha}) > 1/(2(\underline{q} \wedge (1+\overline{q})))$. *Then, for fixed* $t \in \mathbb{T}$ *and* $\tau \in \mathcal{T}_t$, *we have*

$$\Delta_n^{-1/2} \langle \widehat{\varphi}_t(\cdot,\tau) - \varphi_t(\cdot,\tau), h \rangle \xrightarrow{\mathcal{F}^X\text{-}s} \mathcal{N}\big(0, \Sigma^{(t,\tau)}(h), \Gamma^{(t,\tau)}(h)\big)$$

for all $h \in L^2_1(\pi)$, *where* $\Sigma^{(t,\tau)}(h) = \langle h, K^{(t,\tau)} h \rangle$ *and* $\Gamma^{(t,\tau)}(h) = \langle h, S^{(t,\tau)} h \rangle$ *and the* $\mathcal{F}^X$*-measurable covariance and relation operators,* $K^{(t,\tau)}$ *and* $S^{(t,\tau)}$, *are integral operators with kernels given by*

$$k^{(t,\tau)}(u_1,u_2) := \overline{\mathfrak{u}_{t,\tau}(u_1)}\,\mathfrak{u}_{t,\tau}(u_2) \int e^{(\mathrm{i}\frac{u_2-u_1}{\sqrt{\tau}\kappa_{t,\tau}}-2)m} \sigma_t^2(\tau,m)dm, \tag{B.4}$$

$$s^{(t,\tau)}(u_1,u_2) := \mathfrak{u}_{t,\tau}(u_1)\,\mathfrak{u}_{t,\tau}(u_2) \int e^{(\mathrm{i}\frac{u_2+u_1}{\sqrt{\tau}\kappa_{t,\tau}}-2)m} \sigma_t^2(\tau,m)dm, \tag{B.5}$$

respectively, with $\mathfrak{u}_{t,\tau}(u) := \frac{u^2}{\tau\kappa_{t,\tau}^2} + \mathrm{i}\frac{u}{\sqrt{\tau}\kappa_{t,\tau}}$.

Proof. Similar to the proof of Lemma B.1, since t and τ are fixed, we denote $n := n_{t,\tau}$, $m_j := m_{t,\tau}(j)$, and $\Delta m_j := \Delta_{t,\tau}(j)$. We will first show the $\mathcal{F}^X$-stable CLT

$$\Delta_n^{-1/2} \langle \zeta_t^{(1)}(\cdot,\tau), h \rangle \xrightarrow{\mathcal{F}^X\text{-}s} \mathcal{N}\big(0, \Sigma^{(t,\tau)}(h), \Gamma^{(t,\tau)}(h)\big) \tag{B.6}$$

for any $h \in L^2_1(\pi)$.

It is notationally convenient to set $\Delta_n^{-1/2}\zeta_t^{(1)}(u,\tau) = \sum_{j=2}^n f_j^{(t,\tau)}(u)$, defining

$$f_j^{(t,\tau)}(u) := -\Delta_n^{-1/2}\ \mathfrak{u}_{t,\tau}(u)\ e^{(\mathrm{i}\frac{u}{\sqrt{\tau}\kappa_{t,\tau}}-1)m_j} \zeta_t(\tau,m_j)\Delta m_j.$$

For all $h \in L^2_1(\pi)$, we can write

$$\begin{aligned}
\langle f_j^{(t,\tau)}, h \rangle &= \int f_j^{(t,\tau)}(u)\overline{h(u)}\pi(du) \\
&= \Delta_n^{-1/2}\ e^{-m_j}\zeta_t(\tau,m_j)\Delta m_j \underbrace{\int -\mathfrak{u}_{t,\tau}(u)\ e^{\mathrm{i}\frac{u}{\sqrt{\tau}\kappa_{t,\tau}}m_j}\ \overline{h(u)}\pi(du)}_{=:I_j^{(t,\tau)}(h)} \\
&= \Delta_n^{-1/2}\ e^{-m_j}\zeta_t(\tau,m_j) I_j^{(t,\tau)}(h)\Delta m_j
\end{aligned}$$

with $|I_j^{(t,\tau)}(h)| \le M\|h\| < \infty$ uniformly bounded for some $M > 0$, using the Cauchy-Schwarz inequality and Assumption 4(i). Moreover, Assumption 3 implies that $\langle f_j^{(t,\tau)}, h \rangle$ for $j = 1, \dots, n$ are $\mathcal{F}^X$-conditionally independent random variables with zero mean and finite variances of the form

$$\begin{aligned}
\mathbb{E}\Big[\langle f_j^{(t,\tau)}, h \rangle \overline{\langle f_j^{(t,\tau)}, h \rangle} \mid \mathcal{F}^X\Big] &= \mathbb{E}\Big[\iint f_j^{(t,\tau)}(u_1)\overline{h(u_1) f_j^{(t,\tau)}(u_2)} h(u_2)\pi(du_1)\pi(du_2) \mid \mathcal{F}^X\Big] \\
&= \iint \mathbb{E}\Big[f_j^{(t,\tau)}(u_1)\overline{f_j^{(t,\tau)}(u_2)} \mid \mathcal{F}^X\Big]\ \overline{h(u_1)} h(u_2)\pi(du_1)\pi(du_2) \\
&= \int h(u_2)\overline{K_j^{(t,\tau)} h(u_2)}\pi(du_2) \\
&= \langle h, K_j^{(t,\tau)} h \rangle,
\end{aligned}$$

where the integral operator

$$K_j^{(t,\tau)}h(u_2) = \int \mathbb{E}\Big[f_j^{(t,\tau)}(u_2)\overline{f_j^{(t,\tau)}(u_1)} \mid \mathcal{F}^X\Big]\, h(u_1)\pi(du_1)$$
$$= \int k_j^{(t,\tau)}(u_1,u_2)h(u_1)\pi(du_1)$$

with the kernel $k_j^{(t,\tau)}(u_1,u_2) = \mathbb{E}\Big[f_j^{(t,\tau)}(u_2)\overline{f_j^{(t,\tau)}(u_1)} \mid \mathcal{F}^X\Big]$.

To appeal to the stable CLT in Proposition 6.1 of Häusler and Luschgy (2015), the sequence of $\{\langle f_j^{(t,\tau)}, h\rangle\}_{j=1,2,\dots}$ given some $h \in L_1^2(\pi)$ is expressed as a square-integrable martingale difference array, after suitable permutation of strikes. We therefore define the filtration $\{\widetilde{\mathcal{F}}_j\}_{j=0,1,\dots}$ by

$$\widetilde{\mathcal{F}}_j = \mathcal{F}^X \vee \sigma\big(\{\zeta_t(\tau, m_i)\}_{t\in\mathbb{T},\ \tau\in\mathcal{T}_t,\ i=1,\dots,j}\big), \quad j = 0,1,\dots,$$

which form a nested sequence as $n \to \infty$ due to the nested construction of the log-moneyness strike grids. For the limiting σ-algebra, we set $\widetilde{\mathcal{F}} := \bigvee_j \widetilde{\mathcal{F}}_j \supset \mathcal{F}^X$.

In particular, note that $\langle f_j^{(t,\tau)}, h\rangle$ is $\widetilde{\mathcal{F}}_j$-adapted with

$$\mathbb{E}\big[\langle f_j^{(t,\tau)}, h\rangle \mid \widetilde{\mathcal{F}}_{j-1}\big] = \mathbb{E}\big[\langle f_j^{(t,\tau)}, h\rangle \mid \mathcal{F}^X\big] = 0$$

as well as conditional variance and pseudo-variance given by

$$\sum_{j=2}^n \mathbb{E}\Big[\langle f_j^{(t,\tau)}, h\rangle\overline{\langle f_j^{(t,\tau)}, h\rangle} \mid \widetilde{\mathcal{F}}_{j-1}\Big] = \sum_{j=2}^n \mathbb{E}\Big[\langle f_j^{(t,\tau)}, h\rangle\overline{\langle f_j^{(t,\tau)}, h\rangle} \mid \mathcal{F}^X\Big] \xrightarrow{\mathbb{P}} \Sigma^{(t,\tau)}(h),$$
$$\sum_{j=2}^n \mathbb{E}\Big[\langle f_j^{(t,\tau)}, h\rangle\langle f_j^{(t,\tau)}, h\rangle \mid \widetilde{\mathcal{F}}_{j-1}\Big] = \sum_{j=2}^n \mathbb{E}\Big[\langle f_j^{(t,\tau)}, h\rangle\langle f_j^{(t,\tau)}, h\rangle \mid \mathcal{F}^X\Big] \xrightarrow{\mathbb{P}} \Gamma^{(t,\tau)}(h),$$

where $\Sigma^{(t,\tau)}(h)$ and $\Gamma^{(t,\tau)}(h)$ are $\mathcal{F}^X$-measurable with $\mathbb{E}[\Sigma^{(t,\tau)}(h)] < \infty$ and $\mathbb{E}[\Gamma^{(t,\tau)}(h)] < \infty$.

The $\mathcal{F}^X$-measurable asymptotic variance $\Sigma^{(t,\tau)}(h)$ and pseudo-variance $\Gamma^{(t,\tau)}(h)$ can be found via the convergence of the respective covariance and relation operators. In particular, we obtain for the variance that

$$\Delta_n^{-1}\mathbb{E}\Big[\langle \zeta_t^{(1)}(\cdot,\tau), h\rangle\overline{\langle \zeta_t^{(1)}(\cdot,\tau), h\rangle} \mid \mathcal{F}^X\Big] = \sum_{j=2}^n \mathbb{E}\Big[\langle f_j^{(t,\tau)}, h\rangle\overline{\langle f_j^{(t,\tau)}, h\rangle} \mid \mathcal{F}^X\Big]$$
$$= \sum_{j=2}^n \langle h, K_j^{(t,\tau)}h\rangle = \left\langle h, \sum_{j=2}^n K_j^{(t,\tau)}h \right\rangle \xrightarrow{\mathbb{P}} \langle h, K^{(t,\tau)}h\rangle =: \Sigma^{(t,\tau)}(h),$$

where the covariance operator is given by

$$K^{(t,\tau)}h(u_2) = \int k^{(t,\tau)}(u_1,u_2)h(u_1)\pi(du_1)$$

with kernel as in equation (B.4) obtained through

$$\begin{aligned}\sum_{j=2}^{n} k_j^{(t,\tau)}(u_1,u_2) &= \sum_{j=2}^{n} \mathbb{E}\Big[f_j^{(t,\tau)}(u_2)\overline{f_j^{(t,\tau)}(u_1)} \mid \mathcal{F}^X\Big] \\ &= \sum_{j=2}^{n} \Delta_n^{-1}\, \overline{\mathfrak{u}_{t,\tau}(u_1)}\, \mathfrak{u}_{t,\tau}(u_2)\; e^{(\mathrm{i}\frac{u_2-u_1}{\sqrt{\tau}\kappa_{t,\tau}}-2)m_j} \mathbb{E}[\zeta_t^2(\tau,m_j) \mid \mathcal{F}^X](\Delta m_j)^2 \\ &= \overline{\mathfrak{u}_{t,\tau}(u_1)}\, \mathfrak{u}_{t,\tau}(u_2) \sum_{j=2}^{n} e^{(\mathrm{i}\frac{u_2-u_1}{\sqrt{\tau}\kappa_{t,\tau}}-2)m_j} \sigma_t^2(\tau,m_j)\frac{(\Delta m_j)^2}{\Delta_n} \\ &\xrightarrow{\mathbb{P}} \overline{\mathfrak{u}_{t,\tau}(u_1)}\, \mathfrak{u}_{t,\tau}(u_2) \int e^{(\mathrm{i}\frac{u_2-u_1}{\sqrt{\tau}\kappa_{t,\tau}}-2)m} \sigma_t^2(\tau,m)dm =: k^{(t,\tau)}(u_1,u_2).\end{aligned}$$

Similarly, we find for the pseudo-variance that

$$\begin{aligned}\Delta_n^{-1}\mathbb{E}\Big[\langle\zeta_t^{(1)}(\cdot,\tau),h\rangle\langle\zeta_t^{(1)}(\cdot,\tau),h\rangle \mid \mathcal{F}^X\Big] &= \sum_{j=2}^{n}\mathbb{E}\Big[\langle f_j^{(t,\tau)},h\rangle\langle f_j^{(t,\tau)},h\rangle \mid \mathcal{F}^X\Big] \\ = \sum_{j=2}^{n}\langle h, S_j^{(t,\tau)}h\rangle = \Big\langle h, \sum_{j=2}^{n} S_j^{(t,\tau)}h\Big\rangle &\xrightarrow{\mathbb{P}} \langle h, S^{(t,\tau)}h\rangle =: \Gamma^{(t,\tau)}(h),\end{aligned}$$

where the relation operator $S_j^{(t,\tau)}$ is an integral operator with kernel $s_j^{(t,\tau)}(u_1,u_2) = \mathbb{E}\Big[f_j^{(t,\tau)}(u_2)f_j^{(t,\tau)}(u_1) \mid \mathcal{F}^X\Big]$, and the kernel of the relation operator $S^{(t,\tau)}$ as in equation (B.5) is found via

$$\begin{aligned}\sum_{j=2}^{n} s_j^{(t,\tau)}(u_1,u_2) &= \sum_{j=2}^{n} \mathbb{E}\Big[f_j^{(t,\tau)}(u_2)f_j^{(t,\tau)}(u_1) \mid \mathcal{F}^X\Big] \\ &= \sum_{j=2}^{n} \Delta_n^{-1}\, \mathfrak{u}_{t,\tau}(u_1)\, \mathfrak{u}_{t,\tau}(u_2)\; e^{(\mathrm{i}\frac{u_2+u_1}{\sqrt{\tau}\kappa_{t,\tau}}-2)m_j} \mathbb{E}[\zeta_t^2(\tau,m_j) \mid \mathcal{F}^X](\Delta m_j)^2 \\ &= \mathfrak{u}_{t,\tau}(u_1)\, \mathfrak{u}_{t,\tau}(u_2) \sum_{j=2}^{n} e^{(\mathrm{i}\frac{u_2+u_1}{\sqrt{\tau}\kappa_{t,\tau}}-2)m_j} \sigma_t^2(\tau,m_j)\frac{(\Delta m_j)^2}{\Delta_n} \\ &\xrightarrow{\mathbb{P}} \mathfrak{u}_{t,\tau}(u_1)\, \mathfrak{u}_{t,\tau}(u_2) \int e^{(\mathrm{i}\frac{u_2+u_1}{\sqrt{\tau}\kappa_{t,\tau}}-2)m} \sigma_t^2(\tau,m)dm =: s^{(t,\tau)}(u_1,u_2).\end{aligned}$$

To invoke Proposition 6.1 of Häusler and Luschgy (2015), it remains to verify a conditional form of Lindeberg's condition, which is implied by the following conditional form of Lyapunov's condition for some $0 < \delta \leq 2$ (see Remark 6.8 in Häusler and Luschgy

(2015)):

$$\sum_{j=2}^{n}\mathbb{E}\Big[\big|\langle f_j^{(t,\tau)},h\rangle\big|^{2+\delta}\mid\widetilde{\mathcal{F}}_{j-1}\Big]=\sum_{j=2}^{n}\mathbb{E}\Big[\big|\langle f_j^{(t,\tau)},h\rangle\big|^{2+\delta}\mid\mathcal{F}^X\Big]$$
$$=\Delta_n^{-(1+\delta/2)}\sum_{j=2}^{n}e^{-(2+\delta)m_j}\,\mathbb{E}\big[\big|\zeta_t(\tau,m_j)\big|^{2+\delta}\mid\mathcal{F}^X\big]\big|I_j^{(t,\tau)}(h)\big|^{2+\delta}(\Delta m_j)^{2+\delta}$$
$$\leq\Delta_n^{\delta/2}\sum_{j=2}^{n}e^{-(2+\delta)m_j}\,\sigma_t^{2+\delta}(\tau,m_j)\,\mathbb{E}\big[|\varkappa_t(\tau,m_j)|^{2+\delta}\mid\mathcal{F}^X\big]\big|I_j^{(t,\tau)}(h)\big|^{2+\delta}\Delta m_j$$
$$\leq\mathcal{O}_{\mathbb{P}}(1)\,\Delta_n^{\delta/2}\sum_{j=2}^{n}e^{-(2+\delta)m_j}\,\widetilde{\sigma}_t^{2+\delta}(\tau,m_j)\,O_t^{2+\delta}(\tau,m_j)\,\Delta m_j$$
$$\leq\mathcal{O}_{\mathbb{P}}(1)\,\Delta_n^{\delta/2}\sum_{j=2}^{n}e^{-(2+\delta)m_j}\cdot e^{(2+\delta)(-\overline{q}m_j\wedge(1+\underline{q})m_j)}\Delta m_j$$
$$\leq\mathcal{O}_{\mathbb{P}}(1)\,\Delta_n^{\delta/2}=o_{\mathbb{P}}(1).$$

Concretely, the second inequality follows using the uniform boundedness of $|I_j^{(t,\tau)}(h)|$ and the measurement error specifications in Assumption 3, in particular Assumption 3(ii); the third inequality follows from Assumption 3(iii) as well as Lemma 1 of BLV in combination with Assumption 1(ii); the last inequality results because the sum converges to a finite integral under the employed asymptotic scheme. As a consequence, we have the $\widetilde{\mathcal{F}}$-stable CLT

$$\Delta_n^{-1/2}\langle\zeta_t^{(1)}(\cdot,\tau),h\rangle\xrightarrow{\widetilde{\mathcal{F}}\text{-}s}\mathcal{N}\big(0,\Sigma^{(t,\tau)}(h),\Gamma^{(t,\tau)}(h)\big).$$

Since $\mathcal{F}^X\subset\widetilde{\mathcal{F}}$, this implies the $\mathcal{F}^X$-stable CLT in equation (B.6) for any $h\in L_1^2(\pi)$.

Given the derived rates for the errors in the CCF approximation in Lemma B.1, we note that with $(\underline{\alpha}\wedge\overline{\alpha})>1/(2(\underline{q}\wedge(1+\overline{q})))$, the observation errors $\Delta_n^{-1/2}\langle\zeta_t^{(1)}(\cdot,\tau),h\rangle$ will determine the asymptotic distribution, while $\Delta_n^{-1/2}\langle\zeta_t^{(2)}(\cdot,\tau)+\zeta_t^{(3)}(\cdot,\tau),h\rangle=o_{\mathbb{P}}(1)$.[21] In fact, we have for each $h\in L_1^2(\pi)$ that

$$\Delta_n^{-1/2}\langle\widehat{\varphi}_t(\cdot,\tau)-\varphi_t(\cdot,\tau),h\rangle=\Delta_n^{-1/2}\langle\zeta_t^{(1)}(\cdot,\tau),h\rangle+o_{\mathbb{P}}(1)\xrightarrow{\mathcal{F}^X\text{-}s}\mathcal{N}\big(0,\Sigma^{(t,\tau)}(h),\Gamma^{(t,\tau)}(h)\big),$$

as claimed. □

Lemma B.3. *Suppose Assumptions 1–4 hold with* $(\underline{\alpha}\wedge\overline{\alpha})>1/(2(\underline{q}\wedge(1+\overline{q})))$. *Then, we have*

$$\Delta_n^{-1/2}\begin{pmatrix}\langle\widehat{\boldsymbol{\varphi}}_1-\boldsymbol{\varphi}_1,\mathbf{h}_1\rangle\\\vdots\\\langle\widehat{\boldsymbol{\varphi}}_T-\boldsymbol{\varphi}_T,\mathbf{h}_T\rangle\end{pmatrix}\xrightarrow{\mathcal{F}^X\text{-}s}\mathcal{N}\big(0,\boldsymbol{\Sigma}_T(\mathbf{h}),\boldsymbol{\Gamma}_T(\mathbf{h})\big)$$

[21] The truncation and discretization errors are $\mathcal{F}^X$-measurable and might introduce a finite-sample bias, but, as $n\to\infty$, the CCF approximation is asymptotically unbiased.

for all $\mathbf{h} := (\mathbf{h}_1; \dots; \mathbf{h}_T)$ *with* $\mathbf{h}_t := (h_{1,t}, \dots, h_{p,t})^\top \in L^2_p(\pi)$*, where the* $\mathcal{F}^X$*-measurable covariance and relation matrices are given by*

$$\boldsymbol{\Sigma}_T(\mathbf{h}) := \begin{pmatrix} \Sigma^{(1)}(\mathbf{h}_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \Sigma^{(T)}(\mathbf{h}_T) \end{pmatrix}, \tag{B.7}$$

$$\boldsymbol{\Gamma}_T(\mathbf{h}) := \begin{pmatrix} \Gamma^{(1)}(\mathbf{h}_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \Gamma^{(T)}(\mathbf{h}_T) \end{pmatrix}, \tag{B.8}$$

respectively, using

$$\Sigma^{(t)}(\mathbf{h}_t) := \langle \mathbf{h}_t, K^{(t)}\mathbf{h}_t \rangle := \sum_{i=1}^{p} \langle h_{i,t}, K^{(t,\tau_i)} h_{i,t} \rangle, \tag{B.9}$$

$$\Gamma^{(t)}(\mathbf{h}_t) := \langle \mathbf{h}_t, S^{(t)}\mathbf{h}_t \rangle := \sum_{i=1}^{p} \langle h_{i,t}, S^{(t,\tau_i)} h_{i,t} \rangle. \tag{B.10}$$

Here, $K^{(t,\tau)}$ *and* $S^{(t,\tau)}$ *are integral operators with kernels given by equations (B.4) and (B.5).*

Proof. By Lemma B.2, for each fixed t and τ, we have the stable CLT

$$\Delta_n^{-1/2} \langle \widehat{\varphi}_t(\cdot,\tau) - \varphi_t(\cdot,\tau), h \rangle \xrightarrow{\mathcal{F}^X\text{-}s} \mathcal{N}\big(0, \Sigma^{(t,\tau)}(h), \Gamma^{(t,\tau)}(h)\big)$$

for any $h \in L^2_1(\pi)$, where $\Sigma^{(t,\tau)}(h) = \langle h, K^{(t,\tau)} h \rangle$ and $\Gamma^{(t,\tau)}(h) = \langle h, S^{(t,\tau)} h \rangle$ in terms of the covariance and relation operators $K^{(t,\tau)}$ and $S^{(t,\tau)}$. Due to the characterization of $\mathcal{F}^X$-stable convergence, it holds that

$$\mathbb{E}\big[Y\, \mathbb{E}\big[g(\langle \widehat{\varphi}_t(\cdot,\tau) - \varphi_t(\cdot,\tau), h \rangle) \mid \mathcal{F}^X\big]\big] \to \mathbb{E}[Y\, \mathbb{E}[g(Z^{(t,\tau)}(h)) \mid \mathcal{F}^X]] \tag{B.11}$$

for every bounded $\mathcal{F}^X$-measurable $Y \in L^\infty(\mathcal{F}^X)$ and every bounded and continuous function $g \in C_b(\mathbb{C})$, where $Z^{(t,\tau)}(h)$ is an $\mathcal{F}^X$-independent random variable that realizes the complex Gaussian distribution $\mathcal{N}\big(0, \Sigma^{(t,\tau)}(h), \Gamma^{(t,\tau)}(h)\big)$. Equation (B.11) implies that

$$\mathbb{E}\big[g(\langle \widehat{\varphi}_t(\cdot,\tau) - \varphi_t(\cdot,\tau), h \rangle) \mid \mathcal{F}^X\big] \xrightarrow{\mathbb{P}} \mathbb{E}\big[g(Z^{(t,\tau)}(h)) \mid \mathcal{F}^X\big] \tag{B.12}$$

for all $g \in C_b(\mathbb{C})$.

Fix any $t \in \mathbb{T}$. We will next show that

$$\Delta_n^{-1/2} \langle \widehat{\boldsymbol{\varphi}}_t - \boldsymbol{\varphi}_t, \mathbf{h}_t \rangle \xrightarrow{\mathcal{F}^X\text{-}s} \mathcal{N}\big(0, \Sigma^{(t)}(\mathbf{h}_t), \Gamma^{(t)}(\mathbf{h}_t)\big) \tag{B.13}$$

for any $\mathbf{h}_t = (h^{(t,\tau_1)}, \dots, h^{(t,\tau_p)})^\top \in L^2_p(\pi)$, where the covariance and relation, $\Sigma^{(t)}(\mathbf{h}_t) := \langle \mathbf{h}_t, K^{(t)}\mathbf{h}_t \rangle$ and $\Gamma^{(t)}(\mathbf{h}_t) := \langle \mathbf{h}_t, S^{(t)}\mathbf{h}_t \rangle$, are defined in equations (B.9) and (B.10), respectively.

To establish the stable CLT in equation (B.13), exploiting the $\mathcal{F}^X$-conditional independence of measurement errors due to Assumption 3, observe that equation (B.12) and the Bounded Convergence Theorem[22] yield

$$
\begin{aligned}
\mathbb{E}\left[Y \prod_{\tau \in \mathcal{T}_t} g_\tau(\langle \widehat{\varphi}_t(\cdot,\tau) - \varphi_t(\cdot,\tau), h^{(t,\tau)}\rangle)\right] &= \mathbb{E}\left[Y \prod_{\tau \in \mathcal{T}_t} \mathbb{E}\big[g_\tau(\langle \widehat{\varphi}_t(\cdot,\tau) - \varphi_t(\cdot,\tau), h^{(t,\tau)}\rangle) \mid \mathcal{F}^X\big]\right] \\
&\to \mathbb{E}\left[Y \prod_{\tau \in \mathcal{T}_t} \mathbb{E}\big[g_\tau(Z^{(t,\tau)}(h^{(t,\tau)})) \mid \mathcal{F}^X\big]\right] \\
&= \mathbb{E}\left[Y \prod_{\tau \in \mathcal{T}_t} g_\tau(Z^{(t,\tau)}(h^{(t,\tau)}))\right]
\end{aligned}
$$

for every $Y \in L^\infty(\mathcal{F}^X)$ and $g_\tau \in C_b(\mathbb{C})$, where $Z^{(t,\tau)}(h)$ are random variables, independent of $\mathcal{F}^X$ and of each other, that realize the limiting distributions $\mathcal{N}\big(0, \Sigma^{(t,\tau)}(h), \Gamma^{(t,\tau)}(h)\big)$ for $\tau \in \mathcal{T}_t$, while we may set $Z^{(t,\tau)}(h) = 0$ for $\tau \notin \mathcal{T}_t$. By a complex version of the Stone-Weierstrass Theorem, linear combinations of decomposable functions of the form $\prod_{\tau \in \mathcal{T}_t} g_\tau$ uniformly approximate any $\mathbb{C}$-valued continuous function on a compact domain. Hence, using $Z_t(\mathbf{h}_t) := \sum_{\tau \in \mathcal{T}_t} Z^{(t,\tau)}(h^{(t,\tau)})$ and a similar argument as in the proof of Proposition A.2, it follows in particular that

$$
\mathbb{E}[Y\, g(\langle \widehat{\boldsymbol{\varphi}}_t - \boldsymbol{\varphi}_t, \mathbf{h}_t\rangle)] \to \mathbb{E}[Y\, g(Z_t(\mathbf{h}_t))]
$$

for all $Y \in L^\infty(\mathcal{F}^X)$ and $g \in C_b(\mathbb{C})$. Since $Z_t(\mathbf{h}_t)$ realizes the limiting distribution $\mathcal{N}\big(0, \Sigma^{(t)}(\mathbf{h}_t), \Gamma^{(t)}(\mathbf{h}_t)\big)$, the stable CLT in equation (B.13) holds.

Finally, we establish the joint stable CLT

$$
\Delta_n^{-1/2} \begin{pmatrix} \langle \widehat{\boldsymbol{\varphi}}_1 - \boldsymbol{\varphi}_1, \mathbf{h}_1\rangle \\ \vdots \\ \langle \widehat{\boldsymbol{\varphi}}_T - \boldsymbol{\varphi}_T, \mathbf{h}_T\rangle \end{pmatrix} \xrightarrow{\mathcal{F}^X\text{-}s} \mathcal{N}(0, \boldsymbol{\Sigma}_T(\mathbf{h}), \boldsymbol{\Gamma}_T(\mathbf{h})) \tag{B.14}
$$

for $\mathbf{h} := (\mathbf{h}_1; \ldots; \mathbf{h}_T)$ with covariance and relation matrices, $\boldsymbol{\Sigma}_T(\mathbf{h})$ and $\boldsymbol{\Gamma}_T(\mathbf{h})$, defined in equations (B.7) and (B.8), respectively. The stable CLT in equation (B.14) follows from an analogous argument as above. Specifically, again exploiting the $\mathcal{F}^X$-conditional independence of measurement errors, equation (B.12) and the Bounded Convergence

[22] Note that the Bounded Convergence Theorem in its usual form also holds under convergence in probability, rather than almost surely, by the Vitali Convergence Theorem.

Theorem imply that

$$\begin{aligned}\mathbb{E}\Bigg[Y\prod_{t\in\mathbb{T}}g_t(\langle\widehat{\boldsymbol{\varphi}}_t-\boldsymbol{\varphi}_t,\mathbf{h}_t\rangle)\Bigg] &= \mathbb{E}\Bigg[Y\prod_{t\in\mathbb{T}}\mathbb{E}\big[g_t(\langle\widehat{\boldsymbol{\varphi}}_t-\boldsymbol{\varphi}_t,\mathbf{h}_t\rangle)\mid\mathcal{F}^X\big]\Bigg]\\ &\to \mathbb{E}\Bigg[Y\prod_{t\in\mathbb{T}}\mathbb{E}\big[g_t(Z_t(\mathbf{h}_t))\mid\mathcal{F}^X\big]\Bigg]\\ &= \mathbb{E}\Bigg[Y\prod_{t\in\mathbb{T}}g_t(Z_t(\mathbf{h}_t))\Bigg]\end{aligned}$$

for every $Y\in L^\infty(\mathcal{F}^X)$ and $g_t\in C_b(\mathbb{C})$, from which the CLT in equation (B.14) eventually follows. □

B.2 Proof of Proposition 1

We start with the convergence in probability in $L_p^2(\pi)$ of the option-implied log CCF $\widehat{\boldsymbol{\psi}}_t$ for fixed $t\in\mathbb{T}$. By Lemma B.1, the option-implied CCF $\widehat{\varphi}_t(\cdot,\tau)$ converges uniformly on $\mathcal{U}$ such that $\sup_{u\in\mathcal{U}}|\widehat{\varphi}_t(u,\tau)-\varphi_t(u,\tau)|=o_{\mathbb{P}}(1)$. Given continuity of the CCF and the absence of zeros in $\mathcal{U}$ from Assumption 4(ii), Theorem 7.6.3 in Chung (2000) implies the uniform convergence of the log CCFs $\widehat{\psi}_t(\cdot,\tau)$ on $\mathcal{U}$ such that $\sup_{u\in\mathcal{U}}|\widehat{\psi}_t(u,\tau)-\psi_t(u,\tau)|=o_{\mathbb{P}}(1)$. The convergence in $L_1^2(\pi)$ immediately follows, as

$$\|\widehat{\psi}_t(\cdot,\tau)-\psi_t(\cdot,\tau)\|\le\pi(\mathcal{U})^{1/2}\sup_{u\in\mathcal{U}}\big|\widehat{\psi}_t(u,\tau)-\psi_t(u,\tau)\big|=o_{\mathbb{P}}(1).$$

The identity $\|\widehat{\boldsymbol{\psi}}_t-\boldsymbol{\psi}_t\|^2=\sum_{\tau\in\mathcal{T}_t}\|\widehat{\psi}_t(\cdot,\tau)-\psi_t(\cdot,\tau)\|^2$ further implies $\|\widehat{\boldsymbol{\psi}}_t-\boldsymbol{\psi}_t\|=o_{\mathbb{P}}(1)$, which yields the convergence in $L_p^2(\pi)$.

To establish the $\mathcal{F}^X$-stable convergence in $L_p^2(\pi)$ for the option-implied log CCF $\widehat{\boldsymbol{\psi}}_t$, we first show $\mathcal{F}^X$-stable convergence for the option-implied CCF $\widehat{\boldsymbol{\varphi}}_t$, which we subsequently translate to the log CCF using the functional delta method. For the $\mathcal{F}^X$-stable convergence of $\widehat{\boldsymbol{\varphi}}_t$ for $t\in\mathbb{T}$, it suffices by Proposition A.2 to show that (i) the sequence $\Delta_n^{-1/2}(\widehat{\boldsymbol{\varphi}}_t-\boldsymbol{\varphi}_t)$ of random functions is tight and (ii) the marginals $\Delta_n^{-1/2}\langle\widehat{\boldsymbol{\varphi}}_t-\boldsymbol{\varphi}_t,\mathbf{h}\rangle$ converge $\mathcal{F}^X$-stably to a (complex) Gaussian distribution for all $\mathbf{h}\in L_p^2(\pi)$. For (i), we specifically need that for every $\varepsilon>0$, there exists an $M_\varepsilon>0$ such that

$$\sup_n\ \mathbb{P}\big[\Delta_n^{-1/2}\|\widehat{\boldsymbol{\varphi}}_t-\boldsymbol{\varphi}_t\|>M_\varepsilon\big]<\varepsilon. \tag{B.15}$$

From Lemma B.1 and the bound on $\underline{\alpha}\wedge\overline{\alpha}$, we have that $\Delta_n^{-1/2}\|\widehat{\boldsymbol{\varphi}}_t-\boldsymbol{\varphi}_t\|=\mathcal{O}_{\mathbb{P}}(1)$. Together with the fact that $\mathbb{E}[\|\widehat{\boldsymbol{\varphi}}_t-\boldsymbol{\varphi}_t\|]<\infty$ for each n, equation (B.15) holds, which establishes tightness of the sequence. For (ii), Lemmas B.2 and B.3 yield the stable CLT for the marginals given any $\mathbf{h}\in L_p^2(\pi)$,

$$\Delta_n^{-1/2}\langle\widehat{\boldsymbol{\varphi}}_t-\boldsymbol{\varphi}_t,\mathbf{h}\rangle\xrightarrow{\mathcal{F}^X\text{-}s}\mathcal{N}\big(0,\Sigma^{(t)}(\mathbf{h}),\Gamma^{(t)}(\mathbf{h})\big),$$

with $\Sigma^{(t)}(\mathbf{h}) = \langle \mathbf{h}, K^{(t)}\mathbf{h}\rangle$ and $\Gamma^{(t)}(\mathbf{h}) = \langle \mathbf{h}, S^{(t)}\mathbf{h}\rangle$, where the covariance and relation operators $K^{(t)}$ and $S^{(t)}$ in equations (B.9) and (B.10) are given in terms of the integral operators $K^{(t,\tau)}$ and $S^{(t,\tau)}$ with kernels $k^{(t,\tau)}(u_1,u_2)$ and $s^{(t,\tau)}(u_1,u_2)$ as in equations (B.4) and (B.5), respectively. Therefore, by Proposition A.2, we obtain the functional stable CLT

$$\Delta_n^{-1/2}\big(\widehat{\boldsymbol{\varphi}}_t - \boldsymbol{\varphi}_t\big) \xrightarrow{\mathcal{F}^X\text{-}s} \mathcal{N}\big(0, K^{(t)}, S^{(t)}\big) \text{ in } L_p^2(\pi),$$

independently across $t \in \mathbb{T}$.

Given the uniform convergence results of the CCF and the log CCF on $\mathcal{U}$, we can utilize the functional delta method (cf. section 3.10 in van der Vaart and Wellner, 2023) to obtain a functional stable CLT for $\widehat{\boldsymbol{\psi}}_t$. Denote by $\mathcal{C}_1$ the space of $\mathbb{C}$-valued continuous functions f with $f(0) = 1$ and $f(u) \neq 0$ for all $u \in \mathcal{U}$ as well as by $\mathcal{C}_p$ its $\mathbb{C}^p$-valued counterpart. The distinguished log may be viewed as a mapping $\mathrm{Log}\colon \mathcal{C}_1 \subset L_1^2(\pi) \to L_1^2(\pi)$. Its Hadamard derivative at $\varphi \in \mathcal{C}_1$ in direction η such that $\varphi + \eta \in \mathcal{C}_1$ is given by $\mathrm{Log}'(\eta;\varphi) = \eta/\varphi$. Indeed, for any η_ε with $\varphi + \varepsilon\eta_\varepsilon \in \mathcal{C}_1$ and $\|\eta_\varepsilon - \eta\| = o(1)$ as $\varepsilon \downarrow 0$, it holds that

$$\begin{aligned}&\left\| \frac{\mathrm{Log}(\varphi + \varepsilon\eta_\varepsilon) - \mathrm{Log}(\varphi)}{\varepsilon} - \frac{\eta}{\varphi} \right\| \\ &\qquad \le \pi(\mathcal{U})^{1/2} \sup_{u\in\mathcal{U}} \left| \frac{\mathrm{Log}(\varphi + \varepsilon\eta_\varepsilon)(u) - \mathrm{Log}(\varphi)(u)}{\varepsilon} - \frac{\eta(u)}{\varphi(u)} \right| = o(1),\end{aligned}$$

where the convergence of the supremum is established in Proposition 2 of Jongbloed and van der Meulen (2006). Applying each operation componentwise, this result extends analogously to $L_p^2(\pi)$. Theorem 3.10.17 in van der Vaart and Wellner (2023) then allows to conclude that

$$\Delta_n^{-1/2}\big(\widehat{\boldsymbol{\psi}}_t - \boldsymbol{\psi}_t\big) \xrightarrow{\mathcal{F}^X\text{-}s} \mathcal{N}\big(0, \mathcal{K}^{(t)}, \mathcal{S}^{(t)}\big) \text{ in } L_p^2(\pi), \tag{B.16}$$

independently across $t \in \mathbb{T}$. Here, for $\mathbf{h}_t := (h_{1,t}, \ldots, h_{p,t})^\intercal \in L_p^2(\pi)$, the covariance and relation operators, $\mathcal{K}^{(t)}$ and $\mathcal{S}^{(t)}$, are given by

$$\langle \mathbf{h}_t, \mathcal{K}^{(t)}\mathbf{h}_t\rangle := \sum_{i=1}^p \langle h_{i,t}, \mathcal{K}^{(t,\tau_i)} h_{i,t}\rangle, \tag{B.17}$$

$$\langle \mathbf{h}_t, \mathcal{S}^{(t)}\mathbf{h}_t\rangle := \sum_{i=1}^p \langle h_{i,t}, \mathcal{S}^{(t,\tau_i)} h_{i,t}\rangle, \tag{B.18}$$

in terms of the integral operators $\mathcal{K}^{(t,\tau)}$ and $\mathcal{S}^{(t,\tau)}$ with kernels

$$\kappa^{(t,\tau)}(u_1,u_2) = \frac{k^{(t,\tau)}(u_1,u_2)}{\varphi_t(-u_1,\tau)\,\varphi_t(u_2,\tau)}, \tag{B.19}$$

$$\varsigma^{(t,\tau)}(u_1,u_2) = \frac{s^{(t,\tau)}(u_1,u_2)}{\varphi_t(u_1,\tau)\,\varphi_t(u_2,\tau)}, \tag{B.20}$$

respectively.

B.3 Proof of Proposition 2

Denote by $\widehat{q}_T(\theta)$ the objective function $Q_T(\theta, \{\widehat{\mathbf{x}}_t(\theta)\}_{t\in\mathbb{T}})$ in problem (3.9), concentrating out the optimal state estimators $\widehat{\mathbf{x}}_t(\theta)$ as in equation (3.7). Likewise, define $q_T(\theta)$ as an analogous noise-free objective function, using the log CCFs $\boldsymbol{\psi}_t = \mathbf{f}_t(\theta_0)$ and the associated optimal state estimators $\mathbf{x}_t(\theta) := \langle \boldsymbol{\beta}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle^{-1}\langle \boldsymbol{\psi}_t - \boldsymbol{\alpha}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle$. Formally, we thus define

$$\widehat{q}_T(\theta) := \sum_{t\in\mathbb{T}} w_t \,\|\widehat{\boldsymbol{\psi}}_t - \widehat{\mathbf{f}}_t(\theta)\|^2 \quad \text{and} \quad q_T(\theta) := \sum_{t\in\mathbb{T}} w_t \,\|\boldsymbol{\psi}_t - \mathbf{f}_t(\theta)\|^2$$

with $\widehat{\mathbf{f}}_t(\theta) := \boldsymbol{\alpha}_t(\theta) + \boldsymbol{\beta}_t(\theta)\widehat{\mathbf{x}}_t(\theta)$ and $\mathbf{f}_t(\theta) := \boldsymbol{\alpha}_t(\theta) + \boldsymbol{\beta}_t(\theta)\mathbf{x}_t(\theta)$.

We start by verifying that $\widehat{\theta} \xrightarrow{\mathbb{P}} \theta_0$. As a first step, we want to show that $|\widehat{q}_T(\theta) - q_T(\theta)| = o_{\mathbb{P}}(1)$ uniformly on Θ. To establish this, note the bound

$$|\widehat{q}_T(\theta) - q_T(\theta)| \leq \sum_{t\in\mathbb{T}} w_t \left| \|\widehat{\boldsymbol{\psi}}_t - \widehat{\mathbf{f}}_t(\theta)\|^2 - \|\boldsymbol{\psi}_t - \mathbf{f}_t(\theta)\|^2 \right|. \tag{B.21}$$

Each term on the right-hand side of equation (B.21) is in turn bounded by

$$\begin{aligned} \left| \|\widehat{\boldsymbol{\psi}}_t - \widehat{\mathbf{f}}_t(\theta)\|^2 - \|\boldsymbol{\psi}_t - \mathbf{f}_t(\theta)\|^2 \right| &\leq 2\|\boldsymbol{\psi}_t - \mathbf{f}_t(\theta)\| \, \|\boldsymbol{\eta}_t(\theta)\| + \|\boldsymbol{\eta}_t(\theta)\|^2 \\ &\leq 2\big(\|\mathbf{f}_t(\theta_0)\| + \|\mathbf{f}_t(\theta)\|\big) \, \|\boldsymbol{\eta}_t(\theta)\| + \|\boldsymbol{\eta}_t(\theta)\|^2, \end{aligned} \tag{B.22}$$

where $\boldsymbol{\eta}_t(\theta) := \big(\widehat{\boldsymbol{\psi}}_t - \widehat{\mathbf{f}}_t(\theta)\big) - \big(\boldsymbol{\psi}_t - \mathbf{f}_t(\theta)\big)$. Hence, the uniform convergence of $\widehat{q}_T(\theta)$ follows via equations (B.21) and (B.22) if we can verify that $\|\mathbf{f}_t(\theta)\| = \mathcal{O}_{\mathbb{P}}(1)$ and $\|\boldsymbol{\eta}_t(\theta)\| = o_{\mathbb{P}}(1)$ uniformly on Θ for each $t \in \mathbb{T}$.

To show the uniform boundedness of $\|\mathbf{f}_t(\theta)\|$, we use that[23]

$$\begin{aligned} \|\mathbf{f}_t(\theta)\| &\leq \|\boldsymbol{\alpha}_t(\theta)\| + \|\boldsymbol{\beta}_t(\theta)\mathbf{x}_t(\theta)\| \\ &\leq \|\boldsymbol{\alpha}_t(\theta)\| + \|\langle \boldsymbol{\beta}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle\|^{1/2} \, \|\mathbf{x}_t(\theta)\| \\ &\leq \|\boldsymbol{\alpha}_t(\theta)\| + \operatorname{tr}(\langle \boldsymbol{\beta}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle)^{1/2} \, \|\mathbf{x}_t(\theta)\| = \mathcal{O}_{\mathbb{P}}(1) \end{aligned} \tag{B.23}$$

uniformly on Θ and

$$\begin{aligned} \|\mathbf{x}_t(\theta)\| &\leq \|\langle \boldsymbol{\beta}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle^{-1}\| \, \|\langle \boldsymbol{\psi}_t - \boldsymbol{\alpha}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle\| \\ &\leq \|\langle \boldsymbol{\beta}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle^{-1}\| \operatorname{tr}(\langle \boldsymbol{\beta}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle)^{1/2} \, \|\boldsymbol{\psi}_t - \boldsymbol{\alpha}_t(\theta)\| \\ &\leq \|\langle \boldsymbol{\beta}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle^{-1}\| \operatorname{tr}(\langle \boldsymbol{\beta}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle)^{1/2} \big(\|\boldsymbol{\psi}_t\| + \|\boldsymbol{\alpha}_t(\theta)\|\big) = \mathcal{O}_{\mathbb{P}}(1) \end{aligned} \tag{B.24}$$

uniformly on Θ. In fact, the continuity of $\boldsymbol{\alpha}_t(\theta)$ and $\boldsymbol{\beta}_t(\theta)$ implied by Assumption 6 on the compact parameter space Θ yields that $\boldsymbol{\alpha}_t(\theta)$ and $\boldsymbol{\beta}_t(\theta)$ are uniformly bounded on

[23]Consider a matrix-valued function $\mathbf{G}$ with elements in $L_1^2(\pi)$ and a vector v. Since $\langle \mathbf{G}, \mathbf{G}\rangle$ is a positive definite matrix, it has singular value decomposition $\langle \mathbf{G}, \mathbf{G}\rangle = U\Omega U^{\mathsf{H}}$. Setting $C = \Omega^{1/2}U^{\mathsf{H}}$, we have that $\|\mathbf{G}v\| = \|Cv\|$ and $\|C\| = \|\langle \mathbf{G}, \mathbf{G}\rangle\|^{1/2}$. Therefore, $\|\mathbf{G}v\| \leq \|C\|\|v\| = \|\langle \mathbf{G}, \mathbf{G}\rangle\|^{1/2}\|v\|$. In addition, from a componentwise application of the Cauchy-Schwarz inequality under the Frobenius norm, we obtain $\|\langle \mathbf{G}, \mathbf{G}\rangle\| \leq \|\langle \mathbf{G}, \mathbf{G}\rangle\|_F \leq \operatorname{tr}(\langle \mathbf{G}, \mathbf{G}\rangle)$.

Θ, i.e., $\|\boldsymbol{\alpha}_t(\theta)\| = \mathcal{O}(1)$ and $\operatorname{tr}(\langle \boldsymbol{\beta}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle) = \mathcal{O}(1)$ uniformly on Θ. Furthermore, we have $\|\boldsymbol{\psi}_t\| = \mathcal{O}_{\mathbb{P}}(1)$ since $\mathbb{E}[\|\boldsymbol{\psi}_t\|] < \infty$. Ultimately, the uniform boundedness in equations (B.23) and (B.24) hinges on the behavior of the matrix inverse $\langle \boldsymbol{\beta}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle^{-1}$ over Θ, for which Assumption 7 yields

$$\|\langle \boldsymbol{\beta}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle^{-1}\| = \sigma_{\max}(\langle \boldsymbol{\beta}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle^{-1}) = \sigma_{\min}(\langle \boldsymbol{\beta}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle)^{-1} = \mathcal{O}(1)$$

uniformly on Θ, where $\sigma_{\max}(M)$ and $\sigma_{\min}(M)$ denote the largest and smallest singular value, respectively, of the $d \times d$ matrix M.

To show the uniform convergence of $\|\boldsymbol{\eta}_t(\theta)\| = o_{\mathbb{P}}(1)$, we use that

$$\|\boldsymbol{\eta}_t(\theta)\| \le \|\widehat{\boldsymbol{\psi}}_t - \boldsymbol{\psi}_t\| + \|\widehat{\mathbf{f}}_t(\theta) - \mathbf{f}_t(\theta)\| = o_{\mathbb{P}}(1) \tag{B.25}$$

uniformly on Θ. As a consequence of the consistency of $\widehat{\psi}_t(\cdot, \tau)$ in $L_1^2(\pi)$ for each fixed t and τ established in Proposition 1, $\|\widehat{\boldsymbol{\psi}}_t - \boldsymbol{\psi}_t\| = \|\boldsymbol{\xi}_t\| = o_{\mathbb{P}}(1)$ holds independently of θ. The uniform convergence in equation (B.25) is thus justified as long as $\|\widehat{\mathbf{f}}_t(\theta) - \mathbf{f}_t(\theta)\| = o_{\mathbb{P}}(1)$ uniformly on Θ. By a similar argument as for the uniform bounds in equations (B.23) and (B.24), we indeed obtain that

$$\begin{aligned}\|\widehat{\mathbf{f}}_t(\theta) - \mathbf{f}_t(\theta)\| &= \|\boldsymbol{\beta}_t(\theta)(\widehat{\mathbf{x}}_t(\theta) - \mathbf{x}_t(\theta))\| \\ &\le \operatorname{tr}(\langle \boldsymbol{\beta}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle)^{1/2} \, \|\widehat{\mathbf{x}}_t(\theta) - \mathbf{x}_t(\theta)\| = o_{\mathbb{P}}(1)\end{aligned} \tag{B.26}$$

uniformly on Θ as well as

$$\begin{aligned}\|\widehat{\mathbf{x}}_t(\theta) - \mathbf{x}_t(\theta)\| &\le \|\langle \boldsymbol{\beta}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle^{-1}\| \, \|\langle \boldsymbol{\xi}_t, \boldsymbol{\beta}_t(\theta)\rangle\| \\ &\le \|\langle \boldsymbol{\beta}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle^{-1}\| \operatorname{tr}(\langle \boldsymbol{\beta}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle)^{1/2} \, \|\boldsymbol{\xi}_t\| = o_{\mathbb{P}}(1)\end{aligned} \tag{B.27}$$

uniformly on Θ.

Denote the infimum in Assumption 8 by $\ell(\varepsilon) > 0$. Exploiting the uniform convergence of $\widehat{q}_T(\theta)$ shown above, we may for each $\delta > 0$ choose some N_δ large enough such that $\mathbb{P}[|\widehat{q}_T(\theta) - q_T(\theta)| \ge \ell(\varepsilon)/2] < \delta$ for all $n \ge N_\delta$ and $\theta \in \Theta$. Restrict the attention to the events where $|\widehat{q}_T(\theta) - q_T(\theta)| < \ell(\varepsilon)/2$ for all $\theta \in \Theta$. By Assumption 8, for any $\varepsilon > 0$, $\|\theta - \theta_0\| > \varepsilon$ implies a.s. that $q_T(\theta) \ge \ell(\varepsilon)$ and further

$$\widehat{q}_T(\theta) > q_T(\theta) - \frac{\ell(\varepsilon)}{2} \ge \frac{\ell(\varepsilon)}{2}.$$

However, due to the optimality of $\widehat{\theta}$ for $\widehat{q}_T(\theta)$ and the identity $q_T(\theta_0) = 0$, observe that

$$\widehat{q}_T(\widehat{\theta}) \le \widehat{q}_T(\theta_0) < q_T(\theta_0) + \frac{\ell(\varepsilon)}{2} = \frac{\ell(\varepsilon)}{2}.$$

Hence, we must have $\|\widehat{\theta} - \theta_0\| \le \varepsilon$ a.s. whenever $|\widehat{q}_T(\theta) - q_T(\theta)| < \ell(\varepsilon)/2$. As a consequence, it follows that $\mathbb{P}[\|\widehat{\theta} - \theta_0\| > \varepsilon] < \delta$ for each $n \ge N_\delta$, which establishes the claimed convergence $\|\widehat{\theta} - \theta_0\| = o_{\mathbb{P}}(1)$.

In order to show that also $\widehat{\mathbf{x}}_t(\widehat{\theta}) \xrightarrow{\mathbb{P}} \mathbf{x}_t = \mathbf{x}_t(\theta_0)$ for each $t = 1, \ldots, T$, we start from

$$\|\widehat{\mathbf{x}}_t(\widehat{\theta}) - \mathbf{x}_t(\theta_0)\| \leq \|\widehat{\mathbf{x}}_t(\widehat{\theta}) - \mathbf{x}_t(\widehat{\theta})\| + \|\mathbf{x}_t(\widehat{\theta}) - \mathbf{x}_t(\theta_0)\|. \tag{B.28}$$

For the first term on the right-hand side of equation (B.28), the uniform convergence of $\|\widehat{\mathbf{x}}_t(\theta) - \mathbf{x}_t(\theta)\|$ obtained from equation (B.27) implies that $\|\widehat{\mathbf{x}}_t(\widehat{\theta}) - \mathbf{x}_t(\widehat{\theta})\| = o_{\mathbb{P}}(1)$. Regarding the second term, it suffices to verify that $\mathbf{x}_t(\theta)$ is continuous at $\theta = \theta_0$, as then $\|\mathbf{x}_t(\widehat{\theta}) - \mathbf{x}_t(\theta_0)\| = o_{\mathbb{P}}(1)$ is implied by $\|\widehat{\theta} - \theta\| = o_{\mathbb{P}}(1)$. From the definition $\mathbf{x}_t(\theta)$, the continuity of $\mathbf{x}_t(\theta)$ follows immediately from the continuity of the elementary vector and matrix operations[24] together with the fact that $\|\theta - \widetilde{\theta}\| = o(1)$ implies

$$\begin{aligned}
&\|\langle \boldsymbol{\psi}_t - \boldsymbol{\alpha}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle - \langle \boldsymbol{\psi}_t - \boldsymbol{\alpha}_t(\widetilde{\theta}), \boldsymbol{\beta}_t(\widetilde{\theta})\rangle\| \\
&\quad \leq \|\langle \boldsymbol{\psi}_t, \boldsymbol{\beta}_t(\theta) - \boldsymbol{\beta}_t(\widetilde{\theta})\rangle\| + \|\langle \boldsymbol{\alpha}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle - \langle \boldsymbol{\alpha}_t(\widetilde{\theta}), \boldsymbol{\beta}_t(\widetilde{\theta})\rangle\| \\
&\quad \leq \operatorname{tr}(\langle \boldsymbol{\beta}_t(\theta) - \boldsymbol{\beta}_t(\widetilde{\theta}), \boldsymbol{\beta}_t(\theta) - \boldsymbol{\beta}_t(\widetilde{\theta})\rangle)^{1/2} \|\boldsymbol{\psi}_t\| + o(1) = o_{\mathbb{P}}(1),
\end{aligned}$$

using again that $\|\boldsymbol{\psi}_t\| = \mathcal{O}_{\mathbb{P}}(1)$. In combination, equation (B.28) thus yields the desired convergence $\|\widehat{\mathbf{x}}_t(\widehat{\theta}) - \mathbf{x}_t(\theta_0)\| = o_{\mathbb{P}}(1)$.

B.4 Proof of Proposition 3

Define $\mathbf{F}_t(\theta, \mathbf{z}) := \boldsymbol{\alpha}_t(\theta) + \boldsymbol{\beta}_t(\theta)\mathbf{z}$. Given differentiability of $\boldsymbol{\alpha}_t(\theta)$ and $\boldsymbol{\beta}_t(\theta)$ with respect to θ by Assumption 6 and the consistency results from Proposition 2, the estimates $\widehat{\theta}$ and $\widehat{\mathbf{x}}_t := \widehat{\mathbf{x}}_t(\widehat{\theta})$ with $t \in \mathbb{T}$ jointly solve

$$\begin{cases} \left\langle \widehat{\boldsymbol{\psi}}_t - \mathbf{F}_t(\widehat{\theta}, \widehat{\mathbf{x}}_t), \boldsymbol{\beta}_t(\widehat{\theta}) \right\rangle = 0, \quad \text{for } t \in \mathbb{T}, \\ \displaystyle\sum_{t \in \mathbb{T}} w_t \left\langle \widehat{\boldsymbol{\psi}}_t - \mathbf{F}_t(\widehat{\theta}, \widehat{\mathbf{x}}_t), \nabla_\theta \mathbf{F}_t(\widehat{\theta}, \widehat{\mathbf{x}}_t) \right\rangle = 0, \end{cases}$$

where $\nabla_\theta \mathbf{F}_t(\widehat{\theta}, \widehat{\mathbf{x}}_t) \in \mathbb{C}^{p \times d_\theta}$ denotes the gradient of $\mathbf{F}_t$ with respect to θ, elements of which are functions in $L_1^2(\pi)$. A first-order exact Taylor expansion yields

$$\begin{aligned}
0 &= \langle \widehat{\boldsymbol{\psi}}_t - \mathbf{F}_t(\theta_0, \mathbf{x}_t), \boldsymbol{\beta}_t(\theta_0)\rangle - \langle \boldsymbol{\beta}_t(\breve{\theta}), \boldsymbol{\beta}_t(\breve{\theta})\rangle \big(\widehat{\mathbf{x}}_t - \mathbf{x}_t\big) \\
&\quad + \Big(-\langle \nabla_\theta \mathbf{F}_t(\breve{\theta}, \breve{\mathbf{x}}_t), \boldsymbol{\beta}_t(\breve{\theta})\rangle + \langle \widehat{\boldsymbol{\psi}}_t - \mathbf{F}_t(\breve{\theta}, \breve{\mathbf{x}}_t), \nabla_\theta \boldsymbol{\beta}_t(\breve{\theta})\rangle\Big) \big(\widehat{\theta} - \theta_0\big)
\end{aligned} \tag{B.29}$$

[24]For vector-valued functions $\mathbf{g}$, $\widetilde{\mathbf{g}}$ and matrix-valued functions $\mathbf{G}$ and $\widetilde{\mathbf{G}}$ with elements in $L_1^2(\pi)$, the inner product $(\mathbf{g}, \mathbf{G}) \mapsto \langle \mathbf{g}, \mathbf{G}\rangle$ is continuous as $\|\mathbf{g} - \widetilde{\mathbf{g}}\| = o(1)$ and $\operatorname{tr}(\langle \mathbf{G} - \widetilde{\mathbf{G}}, \mathbf{G} - \widetilde{\mathbf{G}}\rangle) = o(1)$ imply

$$\|\langle \mathbf{g}, \mathbf{G}\rangle - \langle \widetilde{\mathbf{g}}, \widetilde{\mathbf{G}}\rangle\| \leq \operatorname{tr}(\langle \mathbf{G}, \mathbf{G}\rangle)^{1/2} \|\mathbf{g} - \widetilde{\mathbf{g}}\| + \operatorname{tr}(\langle \mathbf{G} - \widetilde{\mathbf{G}}, \mathbf{G} - \widetilde{\mathbf{G}}\rangle)^{1/2} \|\widetilde{\mathbf{g}}\| = o(1).$$

For matrix-valued functions $\mathbf{G}$ and $\widetilde{\mathbf{G}}$ with elements in $L_1^2(\pi)$, the inner product $\mathbf{G} \mapsto \langle \mathbf{G}, \mathbf{G}\rangle$ is continuous as $\operatorname{tr}(\langle \mathbf{G} - \widetilde{\mathbf{G}}, \mathbf{G} - \widetilde{\mathbf{G}}\rangle) = o(1)$ implies

$$\|\langle \mathbf{G}, \mathbf{G}\rangle - \langle \widetilde{\mathbf{G}}, \widetilde{\mathbf{G}}\rangle\| \leq \big(\operatorname{tr}(\langle \mathbf{G}, \mathbf{G}\rangle)^{1/2} + \operatorname{tr}(\langle \widetilde{\mathbf{G}}, \widetilde{\mathbf{G}}\rangle)^{1/2}\big) \operatorname{tr}(\langle \mathbf{G} - \widetilde{\mathbf{G}}, \mathbf{G} - \widetilde{\mathbf{G}}\rangle)^{1/2} = o(1),$$

using that $\|\langle \mathbf{G}, \widetilde{\mathbf{G}}\rangle\| \leq \|\langle \mathbf{G}, \widetilde{\mathbf{G}}\rangle\|_F \leq \operatorname{tr}(\langle \mathbf{G}, \mathbf{G}\rangle)^{1/2} \operatorname{tr}(\langle \widetilde{\mathbf{G}}, \widetilde{\mathbf{G}}\rangle)^{1/2}$ due to the Cauchy-Schwarz inequality.

for $t \in \mathbb{T}$, and

$$0 = \sum_{t\in\mathbb{T}} w_t \Big\{ \langle \widehat{\boldsymbol{\psi}}_t - \mathbf{F}_t(\theta_0, \mathbf{x}_t), \nabla_\theta \mathbf{F}_t(\theta_0, \mathbf{x}_t)\rangle \\ + \Big(-\langle \boldsymbol{\beta}_t(\breve{\theta}), \nabla_\theta \mathbf{F}_t(\breve{\theta}, \breve{\mathbf{x}}_t)\rangle + \langle \widehat{\boldsymbol{\psi}}_t - \mathbf{F}_t(\breve{\theta}, \breve{\mathbf{x}}_t), \nabla_{\mathbf{x}} \nabla_\theta \mathbf{F}_t(\breve{\theta}, \breve{\mathbf{x}}_t)\rangle\Big) (\widehat{\mathbf{x}}_t - \mathbf{x}_t) \\ + \Big(-\langle \nabla_\theta \mathbf{F}_t(\breve{\theta}, \breve{\mathbf{x}}_t), \nabla_\theta \mathbf{F}_t(\breve{\theta}, \breve{\mathbf{x}}_t)\rangle + \langle \widehat{\boldsymbol{\psi}}_t - \mathbf{F}_t(\breve{\theta}, \breve{\mathbf{x}}_t), \nabla^2_\theta \mathbf{F}_t(\breve{\theta}, \breve{\mathbf{x}}_t)\rangle\Big) (\widehat{\theta} - \theta_0) \Big\}, \tag{B.30}$$

where $\breve{\theta}$ is lying between $\widehat{\theta}$ and θ_0, and $\breve{\mathbf{x}}_t$ is between $\widehat{\mathbf{x}}_t$ and $\mathbf{x}_t$ for each t, respectively. Due to the continuity in θ of $\boldsymbol{\alpha}_t(\theta)$ and $\boldsymbol{\beta}_t(\theta)$ implied by Assumption 6, note that $\mathbf{F}_t(\theta, \mathbf{z})$ is also continuous, since $\|\theta - \widetilde{\theta}\| = o(1)$ and $\|\mathbf{z} - \widetilde{\mathbf{z}}\| = o(1)$ implies that

$$\|\mathbf{F}_t(\theta, \mathbf{z}) - \mathbf{F}_t(\widetilde{\theta}, \widetilde{\mathbf{z}})\| \leq o(1) + \operatorname{tr}(\langle \boldsymbol{\beta}_t(\theta), \boldsymbol{\beta}_t(\theta)\rangle)^{1/2} \|\mathbf{z} - \widetilde{\mathbf{z}}\| = o(1). \tag{B.31}$$

Consider $\mathcal{H} \subset L^2_p(\pi)$ such that $\|h\| \leq M$ for all $h \in \mathcal{H}$. Given the consistency of $\widehat{\theta}$ and $\widehat{\mathbf{x}}_t$ obtained in Proposition 2, we thus have for all $h \in \mathcal{H}$ that $\langle \widehat{\boldsymbol{\psi}}_t - \mathbf{F}_t(\breve{\theta}, \breve{\mathbf{x}}_t), h\rangle = o_{\mathbb{P}}(1)$ uniformly on $\mathcal{H}$ because

$$|\langle \widehat{\boldsymbol{\psi}}_t - \mathbf{F}_t(\breve{\theta}, \breve{\mathbf{x}}_t), h\rangle| \leq |\langle \boldsymbol{\xi}_t, h\rangle| + |\langle \boldsymbol{\psi}_t - \mathbf{F}_t(\breve{\theta}, \breve{\mathbf{x}}_t), h\rangle| \\ \leq \big(\|\boldsymbol{\xi}_t\| + \|\mathbf{F}_t(\theta_0, \mathbf{x}_t) - \mathbf{F}_t(\breve{\theta}, \breve{\mathbf{x}}_t)\|^{1/2}\big) \|h\| = o_{\mathbb{P}}(1),$$

where $\|\boldsymbol{\xi}_t\| = o_{\mathbb{P}}(1)$ follows from Proposition 1 and $\|\mathbf{F}_t(\theta_0, \mathbf{x}_t) - \mathbf{F}_t(\breve{\theta}, \breve{\mathbf{x}}_t)\| = o_{\mathbb{P}}(1)$ from the continuity of $\mathbf{F}_t(\theta, \mathbf{z})$ as $\|\breve{\theta} - \theta_0\| \leq \|\widehat{\theta} - \theta_0\| = o_{\mathbb{P}}(1)$ and $\|\breve{\mathbf{x}}_t - \theta_0\| \leq \|\widehat{\mathbf{x}}_t - \theta_0\| = o_{\mathbb{P}}(1)$. In particular, the result holds when choosing $h = \nabla_\theta \boldsymbol{\beta}_t(\breve{\theta})$ in equation (B.29) as well as $h = \nabla_{\mathbf{x}} \nabla_\theta \mathbf{F}_t(\breve{\theta}, \breve{\mathbf{x}}_t) = \nabla_\theta \boldsymbol{\beta}_t(\breve{\theta})$ and $h = \nabla^2_\theta \mathbf{F}_t(\breve{\theta}, \breve{\mathbf{x}}_t) = \nabla^2_\theta \boldsymbol{\alpha}_t(\breve{\theta}) + \boldsymbol{\beta}_t(\breve{\theta}) \breve{\mathbf{x}}_t$ in equation (B.30), due to the uniform boundedness imposed by Assumption 6.

In matrix form, we can therefore write equations (B.29) and (B.30) as

$$\big(\breve{A}_T + o_{\mathbb{P}}(1)\big) \begin{pmatrix} \widehat{\mathbf{x}}_1 - \mathbf{x}_1 \\ \vdots \\ \widehat{\mathbf{x}}_T - \mathbf{x}_T \\ \widehat{\theta} - \theta_0 \end{pmatrix} = \begin{pmatrix} \langle \boldsymbol{\xi}_1, \boldsymbol{\beta}_1(\theta_0)\rangle \\ \vdots \\ \langle \boldsymbol{\xi}_T, \boldsymbol{\beta}_T(\theta_0)\rangle \\ \sum_{t\in\mathbb{T}} w_t \langle \boldsymbol{\xi}_t, \nabla_\theta \mathbf{F}_t(\theta_0, \mathbf{x}_t)\rangle. \end{pmatrix} \tag{B.32}$$

Given the continuity in θ of $\boldsymbol{\alpha}_t(\theta)$ and $\boldsymbol{\beta}_t(\theta)$ and its derivatives in Assumption 6 as well as the continuity of the relevant matrix operations,[25] the consistency result of Proposition 2

[25]The continuity of $\nabla_\theta \mathbf{F}_t(\theta, \mathbf{z})$ can be established analogously to equation (B.31). Moreover, the continuity of inner products is partially addressed in the proof of Proposition 2. Extending these results, note that for matrix-valued functions $\mathbf{G}$, $\widetilde{\mathbf{G}}$ and $\mathbf{H}$, $\widetilde{\mathbf{H}}$ with elements in $L^2_1(\pi)$, the inner product $(\mathbf{G}, \mathbf{H}) \mapsto \langle \mathbf{G}, \mathbf{H}\rangle$ is continuous as $\operatorname{tr}(\langle \mathbf{G} - \widetilde{\mathbf{G}}, \mathbf{G} - \widetilde{\mathbf{G}}\rangle) = o(1)$ and $\operatorname{tr}(\langle \mathbf{H} - \widetilde{\mathbf{H}}, \mathbf{H} - \widetilde{\mathbf{H}}\rangle) = o(1)$ imply

$$\|\langle \mathbf{G}, \mathbf{H}\rangle - \langle \widetilde{\mathbf{G}}, \widetilde{\mathbf{H}}\rangle\| \leq \operatorname{tr}(\langle \mathbf{G}, \mathbf{G}\rangle)^{1/2} \operatorname{tr}(\langle \mathbf{H} - \widetilde{\mathbf{H}}, \mathbf{H} - \widetilde{\mathbf{H}}\rangle)^{1/2} + \operatorname{tr}(\langle \widetilde{\mathbf{H}}, \widetilde{\mathbf{H}}\rangle)^{1/2} \operatorname{tr}(\langle \mathbf{G} - \widetilde{\mathbf{G}}, \mathbf{G} - \widetilde{\mathbf{G}}\rangle)^{1/2} = o(1).$$

yields $\breve{A}_T \xrightarrow{\mathbb{P}} A_T$ for the $\mathcal{F}^X$-measurable matrix

$$A_T := \begin{pmatrix} \langle \boldsymbol{\beta}_1, \boldsymbol{\beta}_1 \rangle & \dots & 0 & \langle \nabla_\theta \mathbf{F}_1, \boldsymbol{\beta}_1 \rangle \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \langle \boldsymbol{\beta}_T, \boldsymbol{\beta}_T \rangle & \langle \nabla_\theta \mathbf{F}_T, \boldsymbol{\beta}_T \rangle \\ w_1 \langle \boldsymbol{\beta}_1, \nabla_\theta \mathbf{F}_1 \rangle & \dots & w_T \langle \boldsymbol{\beta}_T, \nabla_\theta \mathbf{F}_T \rangle & \sum_{t \in \mathbb{T}} w_t \langle \nabla_\theta \mathbf{F}_t, \nabla_\theta \mathbf{F}_t \rangle \end{pmatrix}, \tag{B.33}$$

where on the right-hand side the dependence of functions on the true parameter vector θ_0 and the true state vectors $\mathbf{x}_t$ is suppressed. The matrix $\breve{A}_T$ is defined similarly, but with the functions depending on $\breve{\theta}$ and $\breve{\mathbf{x}}_t$.

Given the stable CLT in Proposition 1, the scaled vector on the right-hand side of equation (B.32) converges $\mathcal{F}^X$-stably to a complex Gaussian distribution. In fact, due to the Hermitian properties of the involved functions, the imaginary part of the limiting distribution is zero, so that the limit is a real Gaussian distribution. Concretely, by equation (B.16), we thus have that

$$\Delta_n^{-1/2} \begin{pmatrix} \langle \boldsymbol{\xi}_1, \boldsymbol{\beta}_1(\theta_0) \rangle \\ \vdots \\ \langle \boldsymbol{\xi}_T, \boldsymbol{\beta}_T(\theta_0) \rangle \\ \sum_{t \in \mathbb{T}} w_t \langle \boldsymbol{\xi}_t, \nabla_\theta \mathbf{F}_t(\theta_0, \mathbf{x}_t) \rangle \end{pmatrix} \xrightarrow{\mathcal{F}^X\text{-}s} \mathcal{N}(0, B_T)$$

with covariance matrix

$$B_T := \begin{pmatrix} \langle \boldsymbol{\beta}_1, \mathcal{K}^{(1)} \boldsymbol{\beta}_1 \rangle & \dots & 0 & w_1 \langle \nabla_\theta \mathbf{F}_1, \mathcal{K}^{(1)} \boldsymbol{\beta}_1 \rangle \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \langle \boldsymbol{\beta}_T, \mathcal{K}^{(T)} \boldsymbol{\beta}_T \rangle & w_T \langle \nabla_\theta \mathbf{F}_T, \mathcal{K}^{(T)} \boldsymbol{\beta}_T \rangle \\ w_1 \langle \boldsymbol{\beta}_1, \mathcal{K}^{(1)} \nabla_\theta \mathbf{F}_1 \rangle & \dots & w_T \langle \boldsymbol{\beta}_T, \mathcal{K}^{(T)} \nabla_\theta \mathbf{F}_T \rangle & \sum_{t \in \mathbb{T}} w_t^2 \langle \nabla_\theta \mathbf{F}_t, \mathcal{K}^{(t)} \nabla_\theta \mathbf{F}_t \rangle \end{pmatrix}, \tag{B.34}$$

where on the right-hand side the dependence of functions on the true parameter vector θ_0 and the true state vectors $\mathbf{x}_t$ is suppressed. Here, for $\mathbf{g}_t := (g_{1,t}, \dots, g_{p,t})^\intercal$, $\mathbf{h}_t := (h_{1,t}, \dots, h_{p,t})^\intercal \in L_p^2(\pi)$, the covariance operators $\mathcal{K}^{(t)}$ are defined analogous to equation (B.17) as

$$\langle \mathbf{g}_t, \mathcal{K}^{(t)} \mathbf{h}_t \rangle := \sum_{i=1}^{p} \langle g_{i,t}, \mathcal{K}^{(t,\tau_i)} h_{i,t} \rangle.$$

In conclusion, invoking the generalized Slutsky Theorem for stable convergence (e.g., Lemma 1.15.6 in van der Vaart and Wellner, 2023), we obtain the stable CLT in equation (3.10), noting that A_T and B_T are real-valued $\mathcal{F}^X$-measurable matrices.

B.5 Proof of Proposition 4

We construct feasible versions $\widehat{A}_T$ and $\widehat{B}_T$ of A_T and B_T in equations (B.33) and (B.34) in Section B.4 as follows. Concretely, for $\widehat{A}_T$ and $\widehat{B}_T$, we set

$$\widehat{A}_T := \begin{pmatrix} \langle \boldsymbol{\beta}_1, \boldsymbol{\beta}_1 \rangle & \dots & 0 & \langle \nabla_\theta \mathbf{F}_1, \boldsymbol{\beta}_1 \rangle \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \langle \boldsymbol{\beta}_T, \boldsymbol{\beta}_T \rangle & \langle \nabla_\theta \mathbf{F}_T, \boldsymbol{\beta}_T \rangle \\ w_1 \langle \boldsymbol{\beta}_1, \nabla_\theta \mathbf{F}_1 \rangle & \dots & w_T \langle \boldsymbol{\beta}_T, \nabla_\theta \mathbf{F}_T \rangle & \sum_{t \in \mathbb{T}} w_t \langle \nabla_\theta \mathbf{F}_t, \nabla_\theta \mathbf{F}_t \rangle \end{pmatrix}, \tag{B.35}$$

$$\widehat{B}_T := \begin{pmatrix} \langle \boldsymbol{\beta}_1, \widehat{\mathcal{K}}^{(1)} \boldsymbol{\beta}_1 \rangle & \dots & 0 & w_1 \langle \nabla_\theta \mathbf{F}_1, \widehat{\mathcal{K}}^{(1)} \boldsymbol{\beta}_1 \rangle \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \langle \boldsymbol{\beta}_T, \widehat{\mathcal{K}}^{(T)} \boldsymbol{\beta}_T \rangle & w_T \langle \nabla_\theta \mathbf{F}_T, \widehat{\mathcal{K}}^{(T)} \boldsymbol{\beta}_T \rangle \\ w_1 \langle \boldsymbol{\beta}_1, \widehat{\mathcal{K}}^{(1)} \nabla_\theta \mathbf{F}_1 \rangle & \dots & w_T \langle \boldsymbol{\beta}_T, \widehat{\mathcal{K}}^{(T)} \nabla_\theta \mathbf{F}_T \rangle & \sum_{t \in \mathbb{T}} w_t^2 \langle \nabla_\theta \mathbf{F}_t, \widehat{\mathcal{K}}^{(t)} \nabla_\theta \mathbf{F}_t \rangle \end{pmatrix}, \tag{B.36}$$

where on the right-hand sides the dependence of functions on the estimated parameter vector $\widehat{\theta}$ and the estimated state vectors $\widehat{\mathbf{x}}_t$ is suppressed. Analogous to the definition of $\mathcal{K}^{(t)}$ in equation (B.17), for $\mathbf{g}_t := (g_{1,t}, \dots, g_{p,t})^\intercal$, $\mathbf{h}_t := (h_{1,t}, \dots, h_{p,t})^\intercal \in L_p^2(\pi)$, we define $\widehat{\mathcal{K}}^{(t)}$ as

$$\langle \mathbf{g}_t, \widehat{\mathcal{K}}^{(t)} \mathbf{h}_t \rangle := \sum_{i=1}^p \langle g_{i,t}, \widehat{\mathcal{K}}^{(t,\tau_i)} h_{i,t} \rangle. \tag{B.37}$$

Employing equations (B.4) and (B.19), each $\widehat{\mathcal{K}}^{(t,\tau_i)}$ is an integral operator with kernel given by

$$\widehat{\kappa}^{(t,\tau)}(u_1, u_2) := \frac{\overline{\mathfrak{u}_{t,\tau}(u_1)}\, \mathfrak{u}_{t,\tau}(u_2)}{\widehat{\varphi}_t(-u_1, \tau)\, \widehat{\varphi}_t(u_2, \tau)} \sum_{j=2}^n e^{(\mathrm{i} \frac{u_2 - u_1}{\sqrt{\tau} \kappa_{t,\tau}} - 2) m_j} \widehat{\zeta}_t^2(\tau, m_j) \frac{(\Delta m_j)^2}{\Delta_n}. \tag{B.38}$$

Using the model-implied option prices $O(\tau, m; \theta, \mathbf{z})$ associated to parameter vector θ and state vector $\mathbf{z} \in \mathbb{R}^d$, $\widehat{\zeta}_t(\tau, m_j)$ are feasible estimates of option errors, constructed as

$$\widehat{\zeta}_t(\tau, m) := O(\tau, m; \widehat{\theta}, \widehat{\mathbf{x}}_t) - \widehat{O}_t(\tau, m) = -\zeta_t(\tau, m) + O(\tau, m; \widehat{\theta}, \widehat{\mathbf{x}}_t) - O(\tau, m; \theta_0, \mathbf{x}_t),$$

where the true option prices satisfy $O_t(\tau, m) = O(\tau, m; \theta_0, \mathbf{x}_t)$.

As in the proof of Proposition 3, we may conclude that $\widehat{A}_T \xrightarrow{\mathbb{P}} A_T$ due to the consistency of $\widehat{\theta}$ and $\widehat{\mathbf{x}}_t$ in Proposition 2. In what follows, we will show that also $\widehat{B}_T \xrightarrow{\mathbb{P}} B_T$. Given the convergence of both $\widehat{A}_T$ and $\widehat{B}_T$ to $\mathcal{F}^X$-measurable limiting matrices, the stable CLT in equation (3.10) combined with the generalized Slutsky Theorem for stable convergence (e.g., Lemma 1.15.6 in van der Vaart and Wellner, 2023) yields the feasible stable CLT in equation (3.11).

It now remains to verify that also $\widehat{B}_T \xrightarrow{\mathbb{P}} B_T$. Since by equation (B.37) any non-zero element of $\widehat{B}_T$ is a finite sum of $\langle h, \widehat{\mathcal{K}}^{(t,\tau)} h \rangle$ for some $h \in L_1^2(\pi)$, it suffices to establish that $\langle h, \widehat{\mathcal{K}}^{(t,\tau)} h\rangle \xrightarrow{\mathbb{P}} \langle h, \mathcal{K}^{(t,\tau)} h\rangle$ for any such h, given fixed t and τ. Each term may be written as

$$\langle h, \widehat{\mathcal{K}}^{(t,\tau)} h\rangle = \sum_{j=2}^{n} e^{-2m_j}\, \widehat{J}_j^{(t,\tau)}(h)\, \widehat{\zeta}_t^2(\tau, m_j) \frac{(\Delta m_j)^2}{\Delta_n}$$

when defining

$$\widehat{J}_j^{(t,\tau)}(h) := \iint \frac{\overline{\mathfrak{u}_{t,\tau}(u_1)}\, \mathfrak{u}_{t,\tau}(u_2)}{\widehat{\varphi}_t(-u_1,\tau)\widehat{\varphi}_t(u_2,\tau)} e^{\mathrm{i}\frac{u_2-u_1}{\sqrt{\tau}\kappa_{t,\tau}} m_j} \overline{h(u_1)}\, h(u_2)\, \pi(du_1)\pi(du_2).$$

Due to the absence of zeros in $\widehat{\varphi}_t(\cdot,\tau)$ by Assumption 4(ii) and continuity, $\widehat{\varphi}_t(\cdot,\tau)$ is uniformly bounded away from zero on $\mathcal{U}$. Hence, by the uniform boundedness and applying the Cauchy-Schwarz inequality twice, we have

$$\widehat{J}_j^{(t,\tau)}(h) \xrightarrow{\mathbb{P}} J_j^{(t,\tau)}(h) := \iint \frac{\overline{\mathfrak{u}_{t,\tau}(u_1)}\, \mathfrak{u}_{t,\tau}(u_2)}{\varphi_t(-u_1,\tau)\varphi_t(u_2,\tau)} e^{\mathrm{i}\frac{u_2-u_1}{\sqrt{\tau}\kappa_{t,\tau}} m_j} \overline{h(u_1)}\, h(u_2)\, \pi(du_1)\pi(du_2).$$

In fact, the convergence is uniform in the sense that $\max_j \left|\widehat{J}_j^{(t,\tau)}(h) - J_j^{(t,\tau)}(h)\right| \xrightarrow{\mathbb{P}} 0$, with $|J_j^{(t,\tau)}(h)| \le M\|h\|^2 < \infty$ uniformly bounded for some $M > 0$, using Assumption 4(i).

We proceed in several steps to verify that $\langle h, \widehat{\mathcal{K}}^{(t,\tau)} h\rangle \xrightarrow{\mathbb{P}} \langle h, \mathcal{K}^{(t,\tau)} h\rangle$. First, using the (infeasible) true option errors $\zeta_t(\tau, m)$, we want to show that for any $h \in L_1^2(\pi)$, we have

$$\sum_{j=2}^{n} e^{-2m_j} J_j^{(t,\tau)}(h)\, \zeta_t^2(\tau, m_j) \frac{(\Delta m_j)^2}{\Delta_n} - \sum_{j=2}^{n} e^{-2m_j} J_j^{(t,\tau)}(h)\, \sigma_t^2(\tau, m_j) \frac{(\Delta m_j)^2}{\Delta_n} \xrightarrow{\mathbb{P}} 0. \quad \text{(B.39)}$$

Define $\chi_t(\tau, m_j) := \zeta_t^2(\tau, m_j) - \sigma_t^2(\tau, m_j)$. Due to Assumption 3, $\chi_t(\tau, m_j)$ are $\mathcal{F}^X$-conditionally independent, $\mathbb{E}[\chi_t(\tau, m_j) \mid \mathcal{F}^X] = 0$, and

$$\begin{aligned}\mathbb{E}\big[\chi_t^2(\tau, m_j) \mid \mathcal{F}^X\big] &= \mathbb{E}[\zeta_t^4(\tau, m_j) \mid \mathcal{F}^X] - \sigma_t^4(\tau, m_j) \\ &= O_t^4(\tau, m_j)\, \widetilde{\sigma}_t^4(\tau, m_j) \big(\mathbb{E}[\varkappa_t^4(\tau, m_j) \mid \mathcal{F}^X] - 1\big) < \infty.\end{aligned}$$

Therefore,

$$\begin{aligned}&\mathbb{E}\left[\left|\sum_{j=2}^{n} e^{-2m_j} J_j^{(t,\tau)}(h)\, \chi_t(\tau, m_j) \frac{(\Delta m_j)^2}{\Delta_n}\right|^2 \mid \mathcal{F}^X\right] \\ &\qquad \le \mathcal{O}_{\mathbb{P}}(1) \sum_{j=2}^{n} e^{-4m_j}\, \mathbb{E}\big[\chi_t^2(\tau, m_j) \mid \mathcal{F}^X\big] \frac{(\Delta m_j)^4}{\Delta_n^2} \\ &\qquad \le \mathcal{O}_{\mathbb{P}}(1)\, \Delta_n \sum_{j=2}^{n} e^{-4m_j} \cdot e^{4(-\overline{q}m_j \wedge (1+\underline{q})m_j)} \Delta m_j \\ &\qquad \le \mathcal{O}_{\mathbb{P}}(1)\, \Delta_n = o_{\mathbb{P}}(1),\end{aligned}$$

where the second inequality specifically employs Assumption 3(ii) as well as Assumption 1(ii) with Lemma 1 of BLV; the final inequality follows since the sum converges to a finite integral. Hence, the convergence in equation (B.39) holds.

Next, using the (feasible) option errors $\widehat{\zeta}_t^2(\tau, m)$, we want to show that

$$\sum_{j=2}^{n} e^{-2m_j} J_j^{(t,\tau)}(h)\, \widehat{\zeta}_t^2(\tau, m_j) \frac{(\Delta m_j)^2}{\Delta_n} - \sum_{j=2}^{n} e^{-2m_j} J_j^{(t,\tau)}(h)\, \zeta_t^2(\tau, m_j) \frac{(\Delta m_j)^2}{\Delta_n} \xrightarrow{\mathbb{P}} 0. \quad \text{(B.40)}$$

Define $\widehat{\chi}_t(\tau, m_j) := \widehat{\zeta}_t^2(\tau, m_j) - \zeta_t^2(\tau, m_j)$ and $\widetilde{\zeta}_t(\tau, m) := O(\tau, m; \widehat{\theta}, \widehat{\mathbf{x}}_t) - O(\tau, m; \theta_0, \mathbf{x}_t)$. With these definitions, note that $\widehat{\chi}_t(\tau, m) = \widetilde{\zeta}_t^2(\tau, m) - 2\zeta_t(\tau, m)\, \widetilde{\zeta}_t(\tau, m)$. To establish the convergence in equation (B.40), we thus analyze the two terms separately. In each case, while the original sequence is along $\mathbb{N}$, we use that due to the consistency result in Proposition 2 and Lemma 3.2 in Kallenberg (1997), there exists a subsequence $N' \subset \mathbb{N}$ along which $\widehat{\theta}$ and $\widehat{\mathbf{x}}_t$ converge almost surely.

For the first term, we thereby obtain

$$\begin{aligned} \left| \sum_{j=2}^{n} e^{-2m_j} J_j^{(t,\tau)}(h)\, \widetilde{\zeta}_t^2(\tau, m_j) \frac{(\Delta m_j)^2}{\Delta_n} \right| &\leq \mathcal{O}(1) \sum_{j=2}^{n} e^{-2m_j}\, \widetilde{\zeta}_t^2(\tau, m_j)\, \Delta m_j \\ &\to \mathcal{O}(1) \int e^{-2m}\, \widetilde{\zeta}_t^2(\tau, m)\, dm \\ &\to 0 \end{aligned}$$

almost surely along N'. Specifically, the inequality follows by the uniform boundedness of $J_j^{(t,\tau)}(h)$. Moreover, the sum converges to an integral, which converges to zero as $n \to \infty$ in N' by a (pathwise) application of the Dominated Convergence Theorem, using that $\widetilde{\zeta}_t^2(\tau, m) \to 0$ for all m by the continuity in Assumption 9(i). Under Assumption 9(ii), employing the bound $\widetilde{\zeta}_t^2(\tau, m) \leq O^2(\tau, m; \widehat{\theta}, \widehat{\mathbf{x}}_t) + O^2(\tau, m; \theta_0, \mathbf{x}_t)$, the required integrability condition is satisfied since

$$\begin{aligned} \int e^{-2m}\, \widetilde{\zeta}_t^2(\tau, m)\, dm &\leq \int e^{-2m} \left(O^2(\tau, m; \widehat{\theta}, \widehat{\mathbf{x}}_t) + O^2(\tau, m; \theta_0, \mathbf{x}_t) \right) dm \\ &\leq \mathcal{O}(1) \int e^{-2m} \cdot e^{2(-\overline{q}m \wedge (1+\underline{q})m)}\, dm < \infty \end{aligned}$$

for large enough n, where the second inequality uses Lemma 1 in BLV.

For the second term, we obtain that

$$\begin{aligned} &\left| \sum_{j=2}^{n} e^{-2m_j} J_j^{(t,\tau)}(h)\, \zeta_t(\tau, m_j)\, \widetilde{\zeta}_t(\tau, m_j) \frac{(\Delta m_j)^2}{\Delta_n} \right| \\ &\qquad \leq \mathcal{O}(1) \sum_{j=2}^{n} e^{-2m_j} \left| \zeta_t(\tau, m_j)\, \widetilde{\zeta}_t(\tau, m_j) \right| \Delta m_j \\ &\qquad \leq \mathcal{O}(1) \left(\sum_{j=2}^{n} e^{-2m_j}\, \zeta_t^2(\tau, m_j)\, \Delta m_j \right)^{1/2} \left(\sum_{j=2}^{n} e^{-2m_j}\, \widetilde{\zeta}_t^2(\tau, m_j)\, \Delta m_j \right)^{1/2} \end{aligned}$$

almost surely over N', where the first inequality uses the uniform boundedness of $J_j^{(t,\tau)}(h)$ and the second one is a Cauchy-Schwarz inequality. Analogous to the proof of Lemma B.1, we may show that the sum of $\zeta_t^2(\tau, m_j)$ is $\mathcal{O}_{\mathbb{P}}(1)$. Moreover, as before, the sum of $\widetilde{\zeta}_t^2(\tau, m_j)$ converges to zero almost surely. Hence, we may take another subsequence $N'' \subset N'$ along which the sum of $\zeta_t^2(\tau, m_j)$ is $\mathcal{O}(1)$ and, thereby, the entire second term converges to zero almost surely.

In conclusion, the convergence in equation (B.40) holds almost surely along the subsequence $N'' \subset \mathbb{N}$. Since the construction carries over to any arbitrary starting subsequence $N \subset \mathbb{N}$ without modification, Lemma 3.2 in Kallenberg (1997) yields that convergence in probability along the original sequence $\mathbb{N}$ holds in equation (B.40), as claimed.

Finally, the uniform convergence of $\widehat{J}_j^{(t,\tau)}(h)$ yields also

$$\sum_{j=2}^{n} e^{-2m_j}\, \widehat{J}_j^{(t,\tau)}(h)\, \widehat{\zeta}_t^2(\tau, m_j) \frac{(\Delta m_j)^2}{\Delta_n} - \sum_{j=2}^{n} e^{-2m_j}\, J_j^{(t,\tau)}(h)\, \zeta_t^2(\tau, m_j) \frac{(\Delta m_j)^2}{\Delta_n} \xrightarrow{\mathbb{P}} 0. \quad \text{(B.41)}$$

Therefore, given any t and τ, it holds that $\langle h, \widehat{\mathcal{K}}^{(t,\tau)} h\rangle \xrightarrow{\mathbb{P}} \langle h, \mathcal{K}^{(t,\tau)} h\rangle$ for all $h \in L_1^2(\pi)$, which implies $\widehat{B}_T \xrightarrow{\mathbb{P}} B_T$ and thus concludes the proof.

C Numerical Implementation

For the numerical implementation of our methodology, we specify $L_p^2(\pi)$ with the choice $\pi(du) = \phi(u; s)\, du$, using the centered normal density $\phi(u; s) := \frac{1}{\sqrt{2\pi} s} \exp(-\frac{1}{2}\frac{u^2}{s^2})$ with a given volatility parameter $s > 0$. The density function ϕ in this context acts as a weight function that determines the importance of each CCF argument u in the resulting integral. In our application, where the integration typically involves estimated CCFs or related quantities, we specifically aim at downweighting larger u with lower signal-to-noise ratio. While we choose a Gaussian PDF, other (symmetric) density functions could also be utilized. However, our choice of the PDF is common in the related C-GMM literature (e.g., Carrasco and Kotchoni, 2017) and allows us to conveniently approximate integrals using Gauss-Hermite quadrature.

Formally, for any functions $f, g \in L_1^2(\phi)$ and $\mathcal{U} = [-U, U]$ with large enough U, we use the Gauss-Hermite quadrature to approximate any inner product:

$$\langle f, g\rangle = \int_{\mathcal{U}} f(u)\, \overline{g(u)}\, \phi(u; s)\, du \approx \int_{\mathbb{R}} f(su)\, \overline{g(su)}\, \phi(u; 1)\, du \approx \sum_{i=1}^{N} w_i\, f(s\widetilde{u}_i)\, \overline{g(s\widetilde{u}_i)},$$

where N is the number of sampling points, and $\widetilde{u}_i$ and w_i are the quadrature points and weights, respectively, determined from the Gauss-Hermite quadrature rule. This approximation is exact if the integrand is a polynomial of order $N - 1$.

Whenever the considered functions f and g are both Hermitian, the number of evaluation points can be halved by considering only the positive real line:

$$\langle f, g\rangle = 2\,\mathrm{Re} \int_{[0,U]} f(u)\, \overline{g(u)}\, \phi(u; s)\, du \approx 2\,\mathrm{Re} \sum_{i=1}^{N} w_i\, \mathbb{1}_{\{\widetilde{u}_i \geq 0\}}\, f(s\widetilde{u}_i)\, \overline{g(s\widetilde{u}_i)}.$$

Since the log CCFs and affine coefficients are continuously-differentiable functions in the argument u, in our application, the approximation by the Gauss-Hermite is accurate and robust even for a small number of quadrature points N, as illustrated in Section 4.

References

Aït-Sahalia, Y., Karaman, M., and Mancini, L. (2020). The term structure of equity and variance risk premia. *Journal of Econometrics*, *219*(2), 204–230.

Aït-Sahalia, Y., and Kimmel, R. L. (2007). Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics*, *83*(2), 413–452.

Aït-Sahalia, Y., Li, C., and Li, C. X. (2021). Implied stochastic volatility models. *Review of Financial Studies*, *34*(1), 394–450.

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

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

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

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

Andersen, T. G., Fusari, N., and Todorov, V. (2020). The pricing of tail risk and the equity premium: Evidence from international option markets. *Journal of Business & Economic Statistics*, *38*(3), 662–678.

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

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

Banz, R. W., and Miller, M. H. (1978). Prices for state-contingent claims: Some estimates and applications. *Journal of Business*, *51*(4), 653–672.

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

Bates, D. S. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *The Review of Financial Studies*, *9*(1), 69–107.

Bates, D. S. (2000). Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics*, *94*(1–2), 181–238.

Bick, A. (1982). Comments on the valuation of derivative assets. *Journal of Financial Economics*, *10*(3), 331–345.

Bollerslev, T., Gibson, M., and Zhou, H. (2011). Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *Journal of Econometrics*, *160*(1), 235–245.

Bondarenko, O., Dillschneider, Y., Schneider, P., and Trojani, F. (2024). *What can you really tell from option prices?* (Working paper)

Boswijk, H. P., Laeven, R. J., and Vladimirov, E. (2024). Estimating option pricing models using a characteristic function-based linear state space representation. *Journal of Econometrics*, *244*(1), forthcoming.

Boswijk, H. P., Laeven, R. J. A., and Lalu, A. (2016). *Asset returns with self-exciting jumps: Option pricing and estimation with a continuum of moments.* (Working paper)

Boswijk, H. P., Laeven, R. J. A., Lalu, A., and Vladimirov, E. (2023). *Jump contagion among stock market indices: Evidence from option markets.* (Working paper)

Breeden, D. T., and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business*, *51*(4), 621–651.

Britten-Jones, M., and Neuberger, A. (2000). Option prices, implied price processes, and stochastic volatility. *Journal of Finance*, *55*(2), 839–866.

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

Bruno, V. J. (1984). A Weierstrass approximation theorem for topological vector spaces. *Journal of Approximation Theory*, *42*(1), 1–3.

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

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

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

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

Carrasco, M., and Kotchoni, R. (2017). Efficient estimation using the characteristic function. *Econometric Theory*, *33*(2), 479–526.

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

Chen, H., Didisheim, A., and Scheidegger, S. (2021). *Deep structural estimation: With an application to option pricing.* (Working paper)

Chen, X., and White, H. (1998). Central limit and functional central limit theorems for Hilbert-valued dependent heterogeneous arrays with applications. *Econometric Theory*, *14*(2), 260–284.

Chernov, M., and Ghysels, E. (2000). A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation.

Journal of Financial Economics, *56*(3), 407–458.

Chong, C. H., and Todorov, V. (2024). Volatility of volatility and leverage effect from options. *Journal of Econometrics*, *240*(1), 105669:1–21.

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

Chung, K. L. (2000). *A course in probability theory.* Elsevier.

Cont, R., and Tankov, P. (2004). Non-parametric calibration of jump-diffusion option pricing models. *Journal of Computational Finance*, *7*(3), 1–49.

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

Dufays, A., Jacobs, K., Liu, Y., and Rombouts, J. (2023). Fast filtering with large option panels: Implications for asset pricing. *Journal of Financial and Quantitative Analysis*, 1–32.

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

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

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

Egloff, D., Leippold, M., and Wu, L. (2010). The term structure of variance swap rates and optimal variance swap investments. *Journal of Financial and Quantitative Analysis*, *45*(5), 1279–1310.

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

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

Fang, F., and Oosterlee, C. W. (2009). A novel pricing method for European options based on Fourier-cosine series expansions. *SIAM Journal on Scientific Computing*, *31*(2), 826–848.

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

Freire, G., and Vladimirov, E. (2023). Autoencoder option pricing models. *Available at SSRN*.

Fulop, A., and Li, J. (2019). Bayesian estimation of dynamic asset pricing models with informative observations. *Journal of Econometrics*, *209*(1), 114–138.

Häusler, E., and Luschgy, H. (2015). *Stable convergence and stable limit theorems.* Springer.

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.

Huang, J.-Z., and Wu, L. (2004). Specification analysis of option pricing models based on time-changed Lévy processes. *Journal of Finance*, *59*(3), 1405–1439.

Jarrow, R., and Kwok, S. S. M. (2015). Specification tests of calibrated option pricing models. *Journal of Econometrics*, *189*(2), 397–414.

Jiang, G. J., and Knight, J. L. (2002). Estimation of continuous-time processes via the empirical characteristic function. *Journal of Business & Economic Statistics*, *20*(2), 198–212.

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

Jones, C. S. (2003). The dynamics of stochastic volatility: Evidence from underlying and options markets. *Journal of Econometrics*, *116*(1–2), 181–224.

Jongbloed, G., and van der Meulen, F. H. (2006). Parametric estimation for subordinators and induced OU processes. *Scandinavian Journal of Statistics*, *33*(4), 825–847.

Kallenberg, O. (1997). *Foundations of modern probability*. Springer.

Lagnado, R., and Osher, S. (1997). A technique for calibrating derivative security pricing models: Numerical solution of an inverse problem. *Journal of Computational Finance*, *1*(1), 14–25.

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

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

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

Singleton, K. J. (2001). Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics*, *102*(1), 111–141.

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

Todorov, V. (2022). Nonparametric jump variation measures from options. *Journal of Econometrics*, *230*(2), 255–280.

Todorov, V., and Zhang, Y. (2023). Bias reduction in spot volatility estimation from options. *Journal of Econometrics*, *234*(1), 53–81.

van der Vaart, A. W., and Wellner, J. A. (2023). *Weak Convergence and Empirical Processes* (2nd ed.). Springer.

Wu, L. (2011). Variance dynamics: Joint evidence from options and high-frequency returns. *Journal of Econometrics*, *160*(1), 280–287.