Parametric Inference and Dynamic State Recovery from Option Panels*

Torben G. Andersen[†] Nicola Fusari[‡] Viktor Todorov[§]

August 2013

Abstract

We develop a new parametric estimation procedure for option panels observed with error. We exploit asymptotic approximations assuming an ever increasing set of option prices in the moneyness (cross-sectional) dimension, but with a fixed time span. We develop consistent estimators for the parameters and the dynamic realization of the state vector governing the option price dynamics. The estimators converge stably to a mixed-Gaussian law and we develop feasible estimators for the limiting variance. We also provide semiparametric tests for the option price dynamics based on the distance between the spot volatility extracted from the options and one constructed nonparametrically from high-frequency data on the underlying asset. Furthermore, we develop new tests for model fit over specific regions of the volatility surface and for the stability of the risk-neutral dynamics over time. A comprehensive Monte Carlo study indicates that the inference procedures work well in empirically realistic settings. In an empirical application to S&P 500 index options, guided by the new diagnostic tests, we extend existing asset pricing models by allowing for a flexible dynamic relation between volatility and priced jump tail risk. Importantly, we document that the priced jump tail risk typically responds in a more pronounced and persistent manner than volatility to large negative market shocks.

Keywords: Option Pricing, Inference, Risk Premia, Jumps, Latent State Vector, Stochastic Volatility, Specification Testing, Stable Convergence.

JEL classification: C51, C52, G12.

---

*We wish to thank Lars Peter Hansen (the co-editor) and anonymous referees for their detailed comments that led to significant improvements of the paper. We also thank Snehal Banerjee, David Bates, Tim Bollerslev, Federico Bugni, Peter Christoffersen, Rama Cont, Drew Creal, Francis Diebold, Dobrislav Dobrev, Kris Jacobs, Jean Jacod, Ravi Jagannathan, Jia Li, Tom McCurdy, Per Mykland, Piotr Orlowski, Dimitris Papanikolau, Jonathan Parker, Markus Reiss, Eric Renault, Andras Sali, Olivier Scaillet, Neil Shephard, Costis Skiadas, George Tauchen, Rasmus T. Varneskov, Mark Watson and Dacheng Xiu as well as many seminar participants. Andersen gratefully acknowledges support from CREATES funded by the Danish National Research Foundation. Todorov's work was partially supported by NSF Grant SES-0957330. We are also grateful for support from the Zell Center for Risk at the Kellogg School.

[†]Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208; NBER, Cambridge, MA; and CREATES, Aarhus, Denmark; e-mail: t-andersen@northwestern.edu.

[‡]Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208; e-mail: n-fusari@northwestern.edu.

[§]Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208; e-mail: v-todorov@northwestern.edu.

1 Introduction

A voluminous literature spanning several decades has, unambiguously, established that time-varying volatility and jumps are intrinsic features of financial prices. Moreover, there has been substantial interest in linking return premiums in the economy to the compensation for such latent risks. In parallel, the trading of derivative contracts has grown explosively, in part reflecting a desire among investors to manage their volatility and jump risk exposures. As a result, ever more comprehensive data for, in particular, exchange-traded options have become available over time. These options span a variety of expiration dates (tenors) and strike prices (moneyness), effectively providing an option or "implied volatility" surface for each trading day, indexed by moneyness and tenor. This sequence of option surfaces – which we label an *option panel* – provides an ideal input for estimation of dynamic asset pricing models. Specifically, in a frictionless parametric setting, the surface allows for perfect recovery of the risk-neutral parameters and the (multivariate) state vector driving the option price dynamics. Within realistic asset pricing models, both reduced form and structural, the state vector typically contains much more information than just the current spot volatility level. For example, it may include short- and long-run volatility factors as well as components governing the jump intensity. Importantly, these state variables are intrinsically latent from the perspective of the underlying asset price, even if a continuous trajectory of the latter is observed.

In practice, however, drawing inference from the option panel is complicated by the presence of non-trivial observation errors in the option prices. The size of those errors vary across strikes and tenor and depend on general market conditions, including possibly the state vector itself. The convolution of the theoretical option price with the observation errors renders the state vector and parameters unobservable directly from a noisy option panel. This situation resembles the recovery of volatility from high-frequency returns, where the volatility is convoluted with Gaussian innovations as well as jumps and microstructure noise. In that context, an elegant solution is to resort to in-fill asymptotics in the time dimension and "average out" the Gaussian return innovations along with the remaining confounding factors. This effectively renders volatility observable subject to quantifiable estimation error. Below, we follow an analogous strategy in the spatial domain.

Consequently, we develop rigorous inference techniques for the implied (latent) state vector and risk-neutral parameters, while avoiding parametric assumptions about the actual measure governing the state vector dynamics. This is feasible as we develop asymptotic distributional approximations assuming only that the number of options underlying each volatility surface is large, so we may treat the time dimension as fixed. We may also allow the observation errors for the option prices to exhibit limited dependence in the spatial (across strikes and tenors) and time series dimension. We accommodate variation in the number of option quotes as well as the strike range and tenor across time – as in the data – and there is no requirement of stationarity in the pattern of maturity and moneyness. Similarly, the observation error may have a non-ergodic and time-varying distribution.

Our estimation method is penalized nonlinear least squares (NLS). The objective function has two parts. The primary component is the mean-square-error in fitting the observed option prices using the parametric option pricing model. The second piece of the objective function penalizes estimates depending on how much the option-implied volatility state deviates from a local nonparametric estimate of spot volatility constructed from high-frequency data on the underlying asset. This constraint stems from the no-arbitrage condition that the current (aggregate) diffusion coefficient must be identical under the actual and risk-neutral measures. Assuming the option price errors "average out" sufficiently when pooled in the objective function, we can consistently estimate both the parameters of the risk-neutral density and the realized trajectory of the state vector.

We further establish the asymptotic properties of our estimator. The convergence is stable, i.e., it holds jointly with any (bounded) random variable defined on the probability space. The limiting distribution is mixed Gaussian with an asymptotic variance that can depend on any random variable adapted to the filtration. The limiting law reflects the flexibility of the estimation approach: we can accommodate option errors that depend in unknown ways on the volatility state as well as option characteristics such as moneyness and tenor. We also provide consistent estimators for the asymptotic variance, thus enabling feasible inference. In analogy to standard NLS, if the option errors are heteroskedastic, we may enhance efficiency by weighting the option fit appropriately for the differing degrees of moneyness and tenor. Consequently, in contrast to much earlier work on option pricing allowing for observation error, e.g., Bates (2000), Jones (2006), and Eraker (2004), we do not impose any parametric assumption on the pricing errors, and we allow them to display significant heteroskedasticity.

As noted previously, the recovery of the volatility state from the option surface has important features in common with the "realized volatility" estimation of stochastic volatility (or time-integrals thereof) based on high-frequency asset returns, see, e.g., Andersen and Bollerslev (1998), Andersen et al. (2003), and Barndorff-Nielsen and Shephard (2002, 2006). In either case, the volatility realization at specific points in time may be recovered pathwise. Moreover, both estimators converge stably with an asymptotic variance that depends on the observed trajectories of asset prices, but do not require stationarity or ergodicity of the volatility process. While the high-frequency (jump-robust) estimator of volatility is based on "averaging out" the noise in the high-frequency return data, the option-based volatility estimator "averages out" the observation errors across the option surface. The major difference is that the option-based estimator exploits a parametric pricing model while the estimator based on high-frequency returns is fully nonparametric. If the option pricing model is valid, the two volatility estimates should not differ in a statistical sense. We formalize and operationalize this observation. Under correct model specification, we establish a joint stable convergence law for the two estimators, enabling us to devise a formal model specification test based on the distance between the two volatility measures.

We propose additional new diagnostic tests for the option price dynamics. The first explores the stability of the risk-neutral parameter estimates over distinct time periods. If the model is misspecified, the period-by-period estimates will, in general, converge to a pseudo-true value, see, e.g., White (1982) and Gourieroux et al. (1984). However, the latter changes over time as the trajectory of the state vector varies across estimation intervals and, for incorrect model specification, this cannot be accommodated by an invariant parameter vector. Hence, we develop a test based on the discrepancy between the parameter estimates over subsequent time periods.

Yet another diagnostic focuses on model performance over specific parts of the implied volatility surface. The empirical option pricing literature typically gauges performance based on the time-averaged fit for a limited set of options. In contrast, we may test for adequacy of the model implied option pricing day-by-day. This diagnostic exploits our feasible limit theory by quantifying the statistical error over the relevant portion of the surface, and then determines if the pricing errors are significant. In essence, the approach disentangles the impact of observation errors (noise) in the option prices from the systematic errors stemming from a misspecified model.

We explore the finite-sample properties of the estimators through an extensive Monte Carlo study using the double-jump stochastic volatility model of Duffie et al. (2000), commonly used in the option pricing literature, as well as a two-factor model. The scale of this simulation study exceeds what has been undertaken previously in the related literature. We find our inference technique to perform well within realistically calibrated settings.

In the empirical application we propose a new three-factor stochastic volatility model and estimate it using an extensive option panel for the S&P 500 index. The model generalizes the existing two-factor specifications by allowing the intensity of the jump tail to depend on an additional factor that is not a component of market volatility (although it can depend on it). Our diagnostic tests reveal that this feature is crucial for explaining the observed dynamic dependencies between short maturity out-of-the-money puts and at-the-money options. The results imply a significant time variation in the risk-neutral jump tail risk. Furthermore, the left and right jump tails exhibit very different dynamics with the latter resembling the dynamics of market volatility more closely. Finally, we document that the response of the priced left tail risk often is substantially more pronounced and persistent than for the volatility process following market crises.

The rest of the paper is organized as follows. Section 2 introduces our formal setup. Section 3 develops our estimators and derives the feasible limit theory. In Section 4, we develop diagnostic tests for the option price dynamics. Section 5 contains a Monte Carlo study of the proposed estimators. In Section 6, we exploit our new inference tools to analyze the option price dynamics of the S&P 500 index. Section 7 concludes. All proofs are deferred to the appendix. In a supplementary appendix we collect additional results pertaining to the Monte Carlo and the empirical application.

2 The Basic Modeling Framework

2.1 Setup and Notation

We first establish some notation. The underlying univariate asset price process is denoted X_t and is defined on a filtered probability space $\left(\Omega^{(0)}, \mathcal{F}^{(0)}, (\mathcal{F}_t^{(0)})_{t\geq 0}, \mathbb{P}^{(0)}\right)$ over the calendar time interval $[0, T]$, for $T > 0$ fixed. It is assumed to be governed by the following general dynamics (under $\mathbb{P}^{(0)}$),

$$\frac{dX_t}{X_{t-}} \;=\; \alpha_t dt \,+\, \sqrt{V_t}\, dW_t \,+\, \int_{x>-1} x\, \widetilde{\mu}(dt, dx)\,, \tag{1}$$

where α_t and V_t are càdlàg; W_t is a $\mathbb{P}^{(0)}$-Brownian motion; μ is an integer-valued random measure counting the jumps in X, with compensator $\widetilde{\nu}^{\mathbb{P}}(dt, dx) = a_t dt \otimes \nu^{\mathbb{P}}(dx)$ for some process a_t and Lévy measure $\nu^{\mathbb{P}}(dx)$, and the associated martingale measure is $\widetilde{\mu} = \mu - \widetilde{\nu}^{\mathbb{P}}$. Furthermore, we denote the expectations operator under $\mathbb{P}^{(0)}$ by $\mathbb{E}[\cdot]$. We assume X satisfies the following condition.

Assumption A0. *The process X in equation (1) satisfies:*

(i) There exists a sequence of stopping times T_r increasing to infinity, and for each integer r there exists a bounded process $V_t^{(r)}$ satisfying $V_t = V_t^{(r)}$ for $t < T_r$, and there are positive constants K_r such that $\mathbb{E}\left\{ |V_t^{(r)} - V_s^{(r)}|^2 \,\middle|\, \mathcal{F}_s^{(0)} \right\} \leq K_r |t-s|$ *for every* $0 \leq s \leq t \leq T$.

(ii) $\int_{x>-1} (|x|^{\beta} \wedge 1)\, \nu^{\mathbb{P}}(dx) < \infty,$ *for some* $\beta \in [0, 2)$.

(iii) $\inf_{t\in[0,T]} V_t > 0$ *and the processes α_t, V_t and a_t are locally bounded.*

Assumption A0 is quite weak and satisfied for almost all standard continuous-time asset pricing models. A0(i) is satisfied if V_t is governed by a (multivariate) stochastic differential equation. Assumption A0(ii) restricts the so-called Blumenthal-Getoor index of the jumps (see, e.g., Section 3.2 in Jacod and Protter (2012)) to be below β. Some of our results, such as Theorem 5 below, depend on the value of this coefficient. Finally, assumption A0(iii) implies that, at each point in time, the price process has a non-vanishing continuous martingale component. We note that assumption A0 does not involve any integrability or stationarity conditions for the model.

The risk-neutral probability measure, $\mathbb{Q}$, is guaranteed to exist by no-arbitrage restrictions on the price process, see, e.g., section 6.K in Duffie (2001), and is locally equivalent to $\mathbb{P}$. It transforms discounted asset prices into (local) martingales. In particular, for X under $\mathbb{Q}$, we have,

$$\frac{dX_t}{X_{t-}} \;=\; (r_t - \delta_t)\, dt \;+\; \sqrt{V_t}\, dW_t \;+\; \int_{x>-1} x\, \widetilde{\mu}(dt, dx)\,, \tag{2}$$

where r_t is the instantaneous risk-free interest rate and δ_t is the instantaneous dividend yield. Moreover, with slight abuse of notation, W_t now denotes a $\mathbb{Q}$-Brownian motion and the jump martingale measure is defined with respect to the risk-neutral compensator $\widetilde{\nu}^{\mathbb{Q}}(dt, dx)$.

We further assume the diffusive volatility and jump processes are governed by a (latent) state vector, so that $V_t = \xi_1(\mathbf{S}_t)$ and $\widetilde{\nu}^{\mathbb{Q}}(dt, dx) = \xi_2(\mathbf{S}_t) \otimes \nu^{\mathbb{Q}}(dx)$, where $\nu^{\mathbb{Q}}(dx)$ is a Lévy measure[1]; ξ_1 and ξ_2 are known functions in C^2, and $\mathbf{S}_t$ denotes the $p \times 1$ state vector. Moreover, r_t and δ_t are smooth functions of $\mathbf{S}_t$, and the latter follows a jump-diffusive Markov process under $\mathbb{Q}$. This specification nests most continuous-time models used in empirical work, including the affine jump-diffusion class of Duffie et al. (2000). The setting allows for volatility processes whose dynamics closely approximate long-memory type dependence, but since $\mathbf{S}_t$ is finite dimensional and follows a Markov process, we do rule out genuine long-memory volatility processes.

We stress that we do not impose any restriction on the dependence between the latent state vector $\mathbf{S}_t$ and either W_t or the jump measure μ. That is, the so-called leverage effect, working through either the diffusive or the jump component of X_t, or both, is allowed for.

We denote European-style out-of-the-money option prices for the asset X at time t by $O_{t,k,\tau}$. Assuming frictionless trading in the options market, the option prices are given as,

$$O_{t,k,\tau} \;=\; \begin{cases} \mathbb{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{t+\tau} r_s\, ds}\, (X_{t+\tau} - K)^+ \right], & \text{if} \quad K > F_{t,t+\tau}\ , \\ \mathbb{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{t+\tau} r_s\, ds}\, (K - X_{t+\tau})^+ \right], & \text{if} \quad K \le F_{t,t+\tau}\ , \end{cases} \tag{3}$$

where τ is the tenor, K the strike price, $F_{t,t+\tau}$ the futures price of the underlying asset at time t for the future date $t+\tau$, and $k = \ln(K/F_{t,t+\tau})$ the log-moneyness. The Markovian assumption on the state vector, $\mathbf{S}_t$, implies that $\frac{e^{r_{t,t+\tau}} O_{t,k,\tau}}{F_{t,t+\tau}}$ is a function only of the tenor, state vector, and moneyness (as well as t, if $\mathbf{S}_t$ is not stationary under $\mathbb{Q}$), where $r_{t,t+\tau}$ is the risk-free interest rate for the period

[1]The separability of the Lévy measure in a time-invariant jump measure on the jump size and a stochastic process is a nontrivial restriction. It essentially amounts to restricting the time-variation of jumps of different sizes to be the same. Nevertheless, this assumption is satisfied in most parametric jump models in empirical applications used to date, e.g., it holds for the whole affine jump-diffusion class of models.

$[t, t+\tau]$.[2] We denote the Black-Scholes implied volatility (BSIV) corresponding to $O_{t,k,\tau}$ by $\kappa_{t,k,\tau}$. This merely represents an alternative, and convenient, pricing convention for the options, as the BSIV is a deterministic and strictly monotone transformation of the ratio $e^{r_{t,t+\tau}} O_{t,k,\tau} \,/\, F_{t,t+\tau}$.

2.2 The Parametric Option Pricing Framework

Henceforth, we assume a parametric model for the risk-neutral distribution, characterized by the $q \times 1$ parameter vector θ, with θ_0 signifying the true value, while we do not restrict the objective distribution for the underlying asset beyond what is implied by the equivalence of the two probability measures.[3] For expositional convenience we assume that the functions $\xi_1(\cdot)$ and $\xi_2(\cdot)$ do not depend on the parameter vector.[4] The option panel has a fixed time span, $[0, T]$, and we observe the option surface at given times $t = 1, \ldots, T$. We have a large cross-section of k values, spanning a significant strike range, available each date for several different tenors, τ. This is a natural assumption for active and liquid option markets.

In this section, we focus on the ideal scenario without measurement errors in the option prices. The critical extension to the case involving such errors is provided in Section 3. The theoretical value of the BSIV under the risk-neutral model is denoted $\kappa(k, \tau, \mathbf{S}_t, \theta)$.[5] For each date t, we have a cross-section of option prices $\{O_{t,k_j,\tau_j}\}_{j=1,\ldots,N_t}$ for some integer N_t, where the index j runs across the full set of strikes and tenors available on day t. The number of options for the maturity τ is denoted N_t^τ. The asymptotic theory developed below reflects the distribution of the available options in the sample across the days. Henceforth, we rely on the following notation,

$$N_t = \sum_\tau N_t^\tau, \qquad N = \sum_t N_t, \qquad \underline{N} = \min_{t=1,\ldots,T} N_t. \tag{4}$$

For each pair, (t, τ), $\underline{k}(t, \tau)$ and $\overline{k}(t, \tau)$ denote the minimum and maximum log-moneyness, respectively. The moneyness grid for the options at time t and tenor τ is denoted,

$$\underline{k}(t, \tau) = k_{t,\tau}(0) < k_{t,\tau}(1) ... < k_{t,\tau}(N_t^\tau) = \overline{k}(t, \tau).$$

The asymptotic scheme sequentially adds new strikes to the existing ones within $[\underline{k}(t, \tau), \overline{k}(t, \tau)]$. The following condition formally captures the notion of a large, yet heterogeneous, option panel.

[2] Renault (1997) discusses the homogeneity of the option price with respect to (X_t, K) more generally.

[3] For an alternative approach imposing minimal assumptions on the objective probability measure, while employing a parametric specification for the stochastic discount factor, see Gagliardini et al. (2011). The inference procedures differ very substantially in other important dimensions, as Gagliardini et al. (2011) invoke large time span asymptotics and deal with a small and fixed cross-section of options for each day.

[4] This assumption is also almost universally satisfied for the models used in practical applications.

[5] Recall that $\mathbf{S}_t$ is a Markov process. If the dynamics of $\mathbf{S}_t$ is non-stationary under $\mathbb{Q}$, then κ should also have a subscript t. For notational simplicity, we impose stationarity, but the analysis readily accommodates non-stationarity.

Assumption A1. *For each $t = 1, ..., T$ and each moneyness τ, the number of options $N_t^{\tau} \uparrow \infty$ with $N_t^{\tau}/N_t \to \pi_t^{\tau}$ and $N_t/N \to \varsigma_t$, for some positive numbers π_t^{τ} and ς_t. Moreover, we have $N_t^{\tau}\,\Delta_{t,\tau}(i) \to \psi_{t,\tau}(k)$ uniformly on the interval $(\underline{k}(t,\tau),\ \overline{k}(t,\tau))$, where $\Delta_{t,\tau}(i) = k_{t,\tau}(i) - k_{t,\tau}(i-1)$ and $\psi_{t,\tau}(\cdot)$ takes on finite and strictly positive values.*

Assumption A1 allows for a great deal of intertemporal heterogeneity in the observation scheme. For example, the tenors need not be identical across days and the assumption of a fixed number of maturities at each point in time is imposed only to simplify the exposition. Importantly, we allow for a different number of options in the panel across days, maturities and moneyness. Also, intuitively, the relative number of options on a given date will impact the inference for the state vector on that date compared with other dates. Likewise, the relative number of options across the various maturities and the local "sparseness" of the strikes should influence the quality of inference for parameters and state variables differentially depending on their sensitivity to tenor and moneyness. The quantities ς_t, π_t^{τ} and $\psi_{t,\tau}(k)$ capture these facets of the panel configuration and they do appear explicitly in the asymptotic distribution theory established later.

Of course, although the risk-neutral measure is guaranteed to exist, it is not unique because, in general, financial markets are incomplete. Given our parametric setting, our next assumption is exactly what is required to uniquely identify the *parameterized* $\mathbb{Q}$ measure (as well as the state vector) given the observation scheme in assumption A1.

Assumption A2. *For every $\epsilon > 0$ and $\theta \in \Theta$, for some compact set Θ, we have,*

$$\inf_{\left(\bigcap_{t=1}^{T}\ \{||\mathbf{Z}_t - \mathbf{S}_t|| \le \epsilon\}\ \cap\ \{||\theta - \theta_0|| \le \epsilon\}\right)^c} \ \sum_{t=1}^{T} \sum_{\tau} \int_{\underline{k}(t,\tau)}^{\overline{k}(t,\tau)} \left(\kappa(k, \tau, \mathbf{S}_t, \theta_0) - \kappa(k, \tau, \mathbf{Z}_t, \theta)\right)^2 dk \ > \ 0, \quad \text{a.s.}$$

We emphasize that this identification condition varies across distinct realizations of the state vector. Assumption A1 and A2 imply that, given correct model specification, we can recover the parameter vector as well as the state vector realization without error at any point in time.[6] While the state variables change from period to period, the parameter vector should remain invariant. Similarly, the fit to the option prices provided by the model should be perfect. These restrictions may serve as the basis for specification tests. Moreover, the parametric model has implications for the pathwise behavior of X across all equivalent probability measures. Most notably, the diffusion coefficient of X, $\xi_1(\mathbf{S}_t)$, should be identical for $\mathbb{Q}$ and $\mathbb{P}$. This property is also testable: the diffusion

[6]In a setting with an increasing time span T, the time series of the recovered state vector, $\mathbf{S}_t$, may be further used to estimate, parametrically or nonparametrically, the associated $\mathbb{P}$ law. Hence, an option panel with increasing time span (and wide cross-section) suffices for estimating both the $\mathbb{Q}$ and $\mathbb{P}$ measures, and thus also the risk premiums associated with the state vector dynamics. In principle, there is no need for return data on the underlying asset.

coefficient may be recovered nonparametrically from a continuous record of X and contrasted with the model-implied $\xi_1(\mathbf{S}_t)$. We develop formal tests for such pathwise restrictions of the risk-neutral model in Section 4, covering the relevant case of noisy option and asset price observations.

There are marked differences in the information content of the option panel (with fixed time grid) versus the price path of the underlying asset. This is most readily illustrated in the ideal, and infeasible, setting of frictionless trading and error-free pricing. A continuous record for X allows us to obtain the diffusive volatility, without error, from a local neighborhood of the current time, and to identify the timing and size of any price jump. In contrast, error-free option data enable us to directly observe the state vector, $\mathbf{S}_t$. If the state vector consists of a single volatility factor, V_t, as is often assumed, the two approaches provide equivalent, and perfect, inference about the state of the system. If the model allows for price jumps, the option panel lets us infer the, possibly time-varying, risk-neutral jump intensities and jump distributions, but does not reveal the actual jump realizations. In contrast, the price path for X identifies the jumps, but does not pin down the jump distribution. Finally, if we move to a multi-factor volatility setting, as implied by much recent research, the options data are even more pivotal for inference. For example, if there are two volatility factors, i.e., $V_t = V_{1,t} + V_{2,t}$, the high-frequency data for X directly informs us about the aggregate value, V_t, only, while the option data let us identify $V_{1,t}$ and $V_{2,t}$ separately. While these conclusions only apply for an ideal setting, it clarifies what type of information one may aspire to obtain from either source, even if it will involve estimation and inferential errors in practice.

3 Inference for Option Panels with a Fixed Time Span

We now turn to the empirically relevant case of noisy observations. Figure 1 depicts a nonparametric kernel regression estimate of the relative bid-ask spread in the quotes for S&P 500 index options, in units of BSIV, as a function of moneyness, and normalized by volatility. The spread is non-trivial and increases quite sharply for deep out-of-the-money (OTM) calls. Clearly, the noise in any individual option price is quite significant. This fact motivates our use of an extensive cross-section of option prices to mitigate and diversify the impact of measurement error.[7]

In the remainder of this section, we develop inference procedures for the parameter vector, θ, governing the risk-neutral distribution and the realized trajectory of the state vector $\{\mathbf{S}_t\}_{t=1,\ldots,T}$ based on an option panel, observed with error. We first introduce our assumptions regarding option errors, then define our estimator and, in turn, establish consistency and asymptotic normality.

[7] A similar perspective underlies the Chicago Board Options Exchange (CBOE) computation of the volatility VIX index. It includes all relevant short maturity S&P 500 index options within the prescribed strike range, with the implicit premise that the observation errors largely "wash out" in the integration.

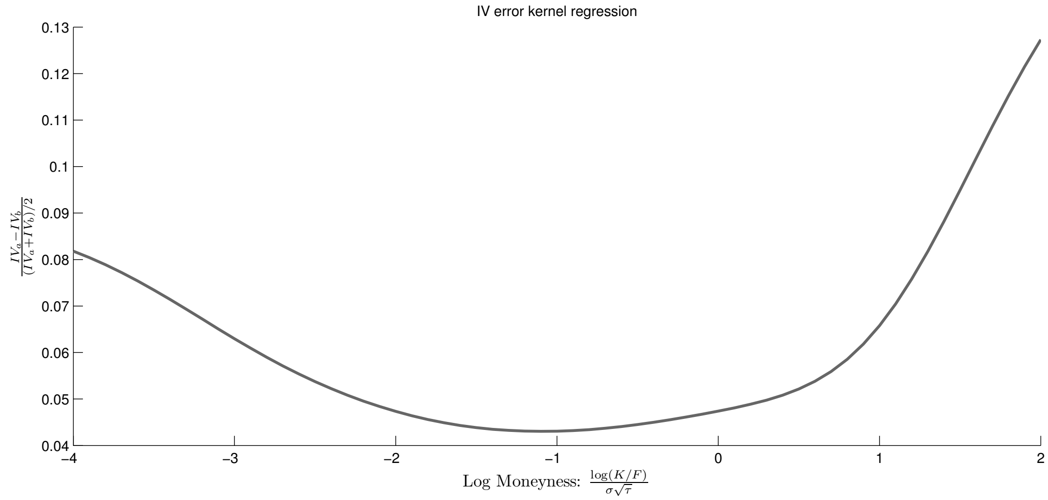

Figure 1: *Kernel regression estimate of the bid-ask spread of option implied volatility as a function of moneyness.* The estimates are based on the best bid and ask quotes for the S&P 500 options on the CBOE at the end-of-trading for each Wednesday during January 1, 1996 – July 21, 2010. We use all available options with maturities up to a year. F and σ denote, respectively, the futures price and the Black-Scholes at-the-money implied volatility at the end of the trading day.

3.1 The Option Error Process

We stipulate that option prices, quoted in terms of BSIV, are observed with error, i.e., we observe

$$\tilde{\kappa}_{t,k,\tau} = \kappa_{t,k,\tau} + \epsilon_{t,k,\tau},$$

where the errors, $\epsilon_{t,k,\tau}$, are defined on a space $\Omega^{(1)} = \bigtimes_{t\in\mathbb{N},k\in\mathbb{R},\tau\in\Gamma} \mathcal{A}_{t,k,\tau}$ for $\mathcal{A}_{t,k,\tau} = \mathbb{R}$, with Γ denoting the set of all possible tenors. $\Omega^{(1)}$ is equipped with the product Borel σ-field $\mathcal{F}^{(1)}$, with transition probability $\mathbb{P}^{(1)}(\omega^{(0)}, d\omega^{(1)})$ from the original probability space $\Omega^{(0)}$ – on which X is defined – to $\Omega^{(1)}$. We define the filtration on $\Omega^{(1)}$ via $\mathcal{F}_t^{(1)} = \sigma\left(\epsilon_{s,k,\tau} : s \le t\right)$. Then the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathbb{P})$ is given as follows,

$$\Omega = \Omega^{(0)} \times \Omega^{(1)},\ \ \mathcal{F} = \mathcal{F}^{(0)} \times \mathcal{F}^{(1)},\ \ \mathcal{F}_t = \cap_{s>t} \mathcal{F}_s^{(0)} \times \mathcal{F}_s^{(1)},\ \ \mathbb{P}(d\omega^{(0)}, d\omega^{(1)}) = \mathbb{P}^{(0)}(d\omega^{(0)})\mathbb{P}^{(1)}(\omega^{(0)}, d\omega^{(1)}).$$

Processes defined on $\Omega^{(0)}$ or $\Omega^{(1)}$ may trivially be viewed as processes on Ω as well, e.g., W_t continues to be a Brownian motion on Ω. We henceforth adopt this perspective without further mention.

Intuitively, this formal representation may be motivated as follows. The option errors are defined on the space $\Omega^{(1)}$. We equip this space with the simple product topology as, at any point in time, only a countable number of them appear in our estimation. Since the option errors can be associated with any strike, point in time and maturity, we need a "large" space to support them.

Finally, since we want to allow the option prices and the underlying process X to be dependent, we define the probability measure via a transition probability distribution from $\Omega^{(0)}$ to $\Omega^{(1)}$. For the formal analysis of $\Omega^{(1)}$, see chapter I of Dellacherie and Meyer (1978).

Given the presence of observation error, we cannot identify the parameters and state vector simply by inverting the option pricing formula. We must explicitly accommodate the impact of noise on the inference. In particular, if a limited set of options is included, then inference is only feasible under strict parametric assumptions regarding the error distribution. This is problematic, as we have little evidence pertaining to the nature of these price errors. In contrast, a large cross-section allows us to "average out" the errors and remain fully nonparametric regarding their distribution. However, this error "diversification" only works if we can ensure that the effect of the option price errors vanishes in a suitable manner. The following condition suffices for establishing consistency of our estimator (recall the notation $\underline{N} = \min_{t=1,...,T} N_t$).

Assumption A3. *For every $\epsilon > 0$, $t = 1, ..., T$, and any finite positive-valued $\mathcal{F}_T^{(0)}$-adapted process $\zeta_t(k, \tau)$ on $\mathbb{R} \times \Gamma$ continuous in its first argument, we have, for $\underline{N} \to \infty$ and $\theta \in \Theta$,*

$$\sup_{\{||\mathbf{Z}_t - \mathbf{S}_t|| > \epsilon\} \, \cup \, \{||\theta - \theta_0|| > \epsilon\}} \frac{\sum_{j=1}^{N_t} \zeta_t(k_j, \tau_j) \left[\kappa(k_j, \tau_j, \mathbf{S}_t, \theta_0) - \kappa(k_j, \tau_j, \mathbf{Z}_t, \theta)\right] \epsilon_{t,k_j,\tau_j}}{\sum_{j=1}^{N_t} \left[\kappa(k_j, \tau_j, \mathbf{S}_t, \theta_0) - \kappa(k_j, \tau_j, \mathbf{Z}_t, \theta)\right]^2} \xrightarrow{\mathbb{P}} 0\,.$$

If the state vector $\mathbf{S}_t$ has bounded support, assumption A3 follows from a uniform Law of Large Numbers on compact sets for which primitive conditions are well known, see, e.g., Newey (1991). Of course, for typical asset pricing models the (stochastic) volatility process, and thus $\mathbf{S}_t$, has unbounded support. Assumption A3 may then be shown to follow from uniform convergence on a space of functions vanishing at infinity; see, e.g., Theorem 21 in Ibragimov and Has'minskii (1981). Assumption A3$'$ provides sufficient conditions for Assumption A3 to hold.

Assumption A3$'$. *For every $t = 1, ..., T$, we assume: (i) $\kappa(k, \tau, \mathbf{S}_t, \theta)$ is continuously differentiable in $(\mathbf{S}_t, \theta)$, (ii) $\kappa^{-1}(k, \tau, \mathbf{S}_t, \theta) = O((\log(||\mathbf{S}_t||))^{-1})$ as $||\mathbf{S}_t|| \to \infty$ and $\theta \in \Theta$ for some region in $(\underline{k}(t, \tau), \overline{k}(t, \tau))$ with positive Lebesgue measure, (iii) $\epsilon_{t,k,\tau}$ and $\epsilon_{t,k',\tau'}$ are independent conditional on $\mathcal{F}^{(0)}$, whenever $(k, \tau) \neq (k', \tau')$, $\mathbb{E}\left(\epsilon_{t,k,\tau} | \mathcal{F}^{(0)}\right) = 0$, $\mathbb{E}\left(|\epsilon_{t,k,\tau}|^{\max\{p,q\}+\iota} | \mathcal{F}^{(0)}\right) = \zeta_{t,k,\tau}$, for $\zeta_{t,k,\tau}$ being a continuous function in its second argument and $\iota > 0$ arbitrary small.*

Assumption A3$'$ provides conditions directly on the option error and option price. In particular, condition (ii) requires the option price to diverge in step with the state vector. This is pretty intuitive. Indeed, the BSIV for short maturity ATM options approximately equals spot volatility and hence increases indefinitely as volatility does so. The smoothness of the option prices in (i) as well as the independence and unbiasedness of the measurement errors in (iii) are implied by the conditions for the asymptotic limit theory to hold, and they are discussed further below. This leaves the weak moment restriction on the measurement noise in (iii) as the only separate assumption on the error process. In the Appendix we verify that Assumption A3$'$ implies Assumption A3.

We require additional regularity for our limiting distributional results to hold.

Assumption A4. *For the error process, $\epsilon_{t,k,\tau}$, we have,*

(i) $\mathbb{E}\left(\epsilon_{t,k,\tau}|\mathcal{F}^{(0)}\right) = 0$ *,*

(ii) $\mathbb{E}\left(\epsilon^2_{t,k,\tau}|\mathcal{F}^{(0)}\right) = \phi_{t,k,\tau}$*, for* $\phi_{t,k,\tau}$ *being a continuous function in its second argument,*

(iii) $\epsilon_{t,k,\tau}$ *and* $\epsilon_{t',k',\tau'}$ *are independent conditional on* $\mathcal{F}^{(0)}$ *, whenever* $(t, k, \tau) \neq (t', k', \tau')$*,*

(iv) $\mathbb{E}\left(|\epsilon_{t,k,\tau}|^4\,|\mathcal{F}^{(0)}\right) < \infty$*, almost surely.*

Assumption A4 implies that the observation errors, conditional on the filtration $\mathcal{F}^{(0)}$, are independent. Nonetheless, the error process may display a stochastically evolving volatility which can depend on option moneyness and tenor as well as any other process defined on the original probability space such as the entire history of X_t and $\mathbf{S}_t$. Relative to the earlier literature, we avoid parametric modeling of the error and allow for significant flexibility for its conditional distribution, including the variance and higher order moments. Assumption A4 does, however, rule out correlated option errors, although this requirement may also be weakened.

Assumption A4 is analogous to the conditions imposed on the microstructure noise process for high-frequency asset prices in Jacod et al. (2009) and subsequent papers. We stress that part (i) is critical for our results, although it may be weakened by allowing for a bias that vanishes asymptotically. Part (iii) excludes correlation in the error across strikes, but we can accommodate weak (spatial) dependence, at the cost of more complex notation (and proof). On the other hand, if the option errors contain a common component across all strikes, this error, obviously, cannot be "averaged out" by spatial integration in the moneyness dimension. For example, Bates (2000) assumes that option prices on a given day, for moneyness within certain ranges, may contain a common error component. He interprets this as a model specification error. In our setting, such features must be included in the theoretical value $\kappa(k, \tau, \mathbf{S}_t, \theta_0)$ rather than being treated as errors.

Finally, if it is more appropriate to assume unbiased errors for the option price rather than the BSIV – which constitutes a nonlinear transformation of the price – one should instead minimize the distance between observed and model-implied option prices. In our empirical application, we find the BSIVs to be approximately linear in prices across the relevant strike range, so the distinction between unbiasedness of implied volatilities or prices is not a practical concern; see, e.g., Christoffersen and Jacobs (2004) for a discussion of the impact of the error specification.

3.2 Consistency

In order to formally define our inference procedure, we first introduce an arbitrary consistent non-parametric estimator for the spot variance, V_t , obtained from high-frequency data on the underlying asset. We denote this estimator $\widehat{V}_t^n$, where n signifies the number of high-frequency observations of X that are available within a unit interval of time (an explicit example of $\widehat{V}_t^n$ is provided in Section 3.3). Our estimates for the state vector and the risk-neutral parameters based on the option panel (and the high-frequency data) are then obtained as follows,

$$\left(\{\widehat{\mathbf{S}}_t^n\}_{t=1,\ldots,T},\, \widehat{\theta}^n\right) = \underset{\{\mathbf{Z}_t\}_{t=1,\ldots,T},\, \theta\in\Theta}{\operatorname{argmin}} \sum_{t=1}^{T} \left\{ \frac{1}{N_t} \sum_{j=1}^{N_t} \left(\tilde{\kappa}_{t,k_j,\tau_j} - \kappa(k_j,\tau_j,\mathbf{Z}_t,\theta)\right)^2 + \lambda_n \left(\widehat{V}_t^n - \xi_1(\mathbf{Z}_t)\right)^2 \right\}, \quad (5)$$

for a deterministic sequence of nonnegative numbers $\{\lambda_n\}$. The estimation is based on minimizing the mean squared error in fitting the panel of observed option implied volatilities, with a penalization term that reflects how much the implied spot volatility deviates from a model-free volatility estimate. The presence of $\widehat{V}_t^n$ in the objective function serves to help identify the parameter vector by penalizing values that imply "unreasonable" volatility levels.

The presence of the penalization term in (5) is reminiscent of the inclusion of information regarding the $\mathbb{P}$ dynamics in option-based estimation, e.g., Bates (2000) and Pan (2002). There is, however, a fundamental difference. We do not model the $\mathbb{P}$ dynamics and the penalization in (5) concerns the pathwise behavior of the option surface, not its $\mathbb{P}$ law. This is therefore, a more robust (parameter-free) and stronger (pathwise) restriction on the option dynamics.

The consistency of $\left(\widehat{\mathbf{S}}_t^n, \widehat{\theta}^n\right)$ follows from the next theorem.

Theorem 1 *Suppose assumptions A1-A3 hold for some $T \in \mathbb{N}$ fixed and that $\{\widehat{V}_t^n\}_{t=1,\ldots,T}$ is consistent for $\{V_t\}_{t=1,\ldots,T}$, as $n \to \infty$. Then, if $\underline{N} \to \infty$ and $\lambda_n \to \lambda$ for some finite $\lambda \geq 0$ as $n \to \infty$, we have that $\left(\widehat{\mathbf{S}}_t^n, \widehat{\theta}_t^n\right)$ exists with probability approaching 1 and further that,*

$$||\widehat{\mathbf{S}}_t^n - \mathbf{S}_t|| \ \xrightarrow{\mathbb{P}}\ 0, \quad ||\widehat{\theta}^n - \theta_0|| \ \xrightarrow{\mathbb{P}}\ 0, \quad t = 1, \ldots, T. \quad (6)$$

Thus, in the presence of observation errors satisfying assumption A3, we can still recover the state vector as well as the risk-neutral parameters consistently from the option panel. The key difference between the parameters and the state vector is that the latter changes from day to day, while the former must remain invariant across the sample. The longer the time span covered by the sample, the more restrictive is this invariance condition for the risk-neutral measure. Another major distinction stems from the penalization term constructed from high-frequency data as this term involves only the state vector and not directly the risk-neutral parameters.

3.3 The Limiting Distribution of the Estimator

In analogy to the high-frequency based realized volatility estimators, which also rely on in-fill asymptotics, our limiting distribution results involve stable convergence. We use the symbol $\xrightarrow{\mathcal{L}-s}$ to indicate this form of convergence. It is an extension of the standard notion of convergence in law to the case where the limiting sequence converges jointly with any bounded variable defined on the original probability space. It is particularly useful when the limiting distribution depends on $\mathcal{F}_T$, as in our setting. For formal analysis of this concept, see, e.g., section VIII.5.c. in Jacod and Shiryaev (2003). The stable convergence result in the following theorem is critical for enabling our feasible inference as well as the development of our diagnostic tests in Section 4.

Theorem 2 *Assume A1-A4 hold and $\kappa(t,\tau,\mathbf{Z},\theta)$ is twice continuously-differentiable in its arguments. Then, if $\underline{N}\to\infty$ and $\lambda_n^2\,\underline{N}\to 0$, for $n\to\infty$, we have,*

$$\begin{pmatrix} \sqrt{N_1}\,(\widehat{\mathbf{S}}_1^n-\mathbf{S}_1) \\ \vdots \\ \sqrt{N_T}\,(\widehat{\mathbf{S}}_T^n-\mathbf{S}_T) \\ \sqrt{N/T}\,(\widehat{\theta}^n-\theta_0) \end{pmatrix} \quad \xrightarrow{\mathcal{L}-s} \quad \mathbf{H}_T^{-1}\,(\mathbf{\Omega}_T)^{1/2} \begin{pmatrix} \mathbf{E}_1 \\ \vdots \\ \mathbf{E}_T \\ \mathbf{E}' \end{pmatrix}, \tag{7}$$

where $\mathbf{E}_1,\ldots,\mathbf{E}_T$ are $p\times 1$ vectors and $\mathbf{E}'$ is $q\times 1$ vector, all defined on an extension of the original probability space being i.i.d. with standard normal distribution, and we define

$$\mathbf{\Phi} = \begin{pmatrix} \mathbf{\Phi}_T^{1,1} & \ldots & \mathbf{0}_{p\times p} & \mathbf{\Phi}_T^{1,T+1} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{p\times p} & \ldots & \mathbf{\Phi}_T^{T,T} & \mathbf{\Phi}_T^{T,T+1} \\ \mathbf{\Phi}_T^{T+1,1} & \ldots & \mathbf{\Phi}_T^{T+1,T} & \mathbf{\Phi}_T^{T+1,T+1} \end{pmatrix}, \qquad \mathbf{\Phi}=\mathbf{H},\ \mathbf{\Omega}, \tag{8}$$

with the blocks of $\mathbf{H}$ and $\mathbf{\Omega}$ for $t=1,\ldots,T$ given by

$$\begin{cases} \mathbf{H}_T^{t,t} = \sum_\tau \pi_t^\tau \int_{\underline{k}(t,\tau)}^{\overline{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)}\,\nabla_{\mathbf{S}}\kappa(k,\tau,\mathbf{S}_t,\theta_0)\,\nabla_{\mathbf{S}}\kappa(k,\tau,\mathbf{S}_t,\theta_0)'dk, \\ \mathbf{H}_T^{T+1,T+1} = \sum_{t=1}^T \sum_\tau \pi_t^\tau \int_{\underline{k}(t,\tau)}^{\overline{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)}\,\nabla_\theta\kappa(k,\tau,\mathbf{S}_t,\theta_0)\,\nabla_\theta\kappa(k,\tau,\mathbf{S}_t,\theta_0)'dk, \\ \mathbf{H}_T^{t,T+1} = \left(\mathbf{H}_T^{T+1,t}\right)' = \sum_\tau \pi_t^\tau \int_{\underline{k}(t,\tau)}^{\overline{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)}\,\nabla_{\mathbf{S}}\kappa(k,\tau,\mathbf{S}_t,\theta_0)\,\nabla_\theta\kappa(k,\tau,\mathbf{S}_t,\theta_0)'dk, \end{cases}$$

$$
\begin{cases}
\mathbf{\Omega}_T^{t,t} = \sum_\tau \pi_t^\tau \int_{\underline{k}(t,\tau)}^{\overline{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)}\, \phi_{t,k,\tau}\, \nabla_{\mathbf{S}} \kappa(k,\tau,\mathbf{S}_t,\theta_0)\, \nabla_{\mathbf{S}} \kappa(k,\tau,\mathbf{S}_t,\theta_0)' dk, \\
\mathbf{\Omega}_T^{T+1,T+1} = \sum_{t=1}^T \frac{1}{T\varsigma_t} \sum_\tau \pi_t^\tau \int_{\underline{k}(t,\tau)}^{\overline{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)}\, \phi_{t,k,\tau}\, \nabla_\theta \kappa(k,\tau,\mathbf{S}_t,\theta_0)\, \nabla_\theta \kappa(k,\tau,\mathbf{S}_t,\theta_0)' dk, \\
\mathbf{\Omega}_T^{t,T+1} = \left(\mathbf{\Omega}_T^{T+1,t}\right)' = \frac{1}{\sqrt{T\varsigma_t}} \sum_\tau \pi_t^\tau \int_{\underline{k}(t,\tau)}^{\overline{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)}\, \phi_{t,k,\tau}\, \nabla_{\mathbf{S}} \kappa(k,\tau,\mathbf{S}_t,\theta_0)\, \nabla_\theta \kappa(k,\tau,\mathbf{S}_t,\theta_0)' dk.
\end{cases}
$$

Several comments are in order. First, we reiterate that the limit result in (7) holds stably conditional on the filtration of the original probability space. The limit is mixed-Gaussian, with a mixing variable, $\mathbf{H}_T^{-1} (\mathbf{\Omega}_T)^{1/2}$, that is adapted to $\mathcal{F}_T$.[8] The random asymptotic variance of the estimator signifies that the precision in recovering the state vector varies from period to period, and that the quality of inference in general depends on the values of the state vector and asset prices as well as the number and characteristics of the options, i.e., tenor and moneyness. This provides important flexibility as the features of the option data change from day to day. It also allows us to formally compare estimates across different time periods and we make frequent use of this fact in the next section. We stress that Theorem 2 does not require any form of stationarity or ergodicity of the state vector, respectively volatility, under the statistical distribution. As noted previously, many aspects of the limiting distributional theory established above resemble the corresponding theory for volatility estimators based on high-frequency data, see, e.g., Barndorff-Nielsen et al. (2006).

Our setup may be contrasted to the cross-sectional regressions with common shocks analyzed by Andrews (2005); see also Kuersteiner and Prucha (2011) for extensions. Andrews (2005) analyzes cross-sectional least squares estimators where both the errors and regressors, conditional on an $\mathcal{F}_0$-adapted random vector, are i.i.d. In our setting, the role of the regressors is taken on by the state vector, $\mathbf{S}_t$, but it is not directly observable and, critically, it exhibits strong temporal dependence. Most importantly, the stable convergence results of Theorem 2 are valid for a much wider σ-field than (a subfield of) $\mathcal{F}_0$ – as in Andrews (2005) – enabling feasible inference from option prices that are highly correlated with the evolving return and volatility innovations. Further, in Theorem 5 below, we show that the stable convergence of Theorem 2 holds jointly with that of a high-frequency estimator for spot volatility, another result for which our general stable convergence result is indispensable.

To implement feasible inference, we need to obtain the requisite consistent estimate of the (conditional) asymptotic variance of $\left(\{\widehat{\mathbf{S}}_t^n\}_{t=1,\ldots,T}, \widehat{\theta}^n\right)$, which in turn, can be done using a consistent estimator of the option error, $\epsilon_{t,k,\tau}$. The formal result is stated in the following theorem.

[8] More formally, the matrices should be denoted $H_T(\omega^{(0)})$ and $\Omega_T(\omega^{(0)})$ to highlight the fact that they depend on the particular realization on the original probability space.

Theorem 3 *Under the conditions of Theorem 2, consistent estimates for $\mathbf{H}_T$ and $\mathbf{\Omega}_T$ are given by $\widehat{\mathbf{H}}_T$ and $\widehat{\mathbf{\Omega}}_T$, where for the same partition of the matrixes as in (8), we set*

$$\begin{cases} \widehat{\mathbf{H}}_T^{t,t} = \frac{1}{N_t}\sum_{j=1}^{N_t} \nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t^n,\widehat{\theta})\, \nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t^n,\widehat{\theta})', \\ \widehat{\mathbf{H}}_T^{T+1,T+1} = \sum_{t=1}^{T}\frac{1}{N_t}\sum_{j=1}^{N_t} \nabla_{\theta}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t^n,\widehat{\theta})\, \nabla_{\theta}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t^n,\widehat{\theta})', \\ \widehat{\mathbf{H}}_T^{t,T+1} = \left(\widehat{\mathbf{H}}_T^{T+1,t}\right)' = \frac{1}{N_t}\sum_{j=1}^{N_t} \nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t^n,\widehat{\theta})\, \nabla_{\theta}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t^n,\widehat{\theta})', \end{cases} \tag{9}$$

$$\begin{cases} \widehat{\mathbf{\Omega}}_T^{t,t} = \frac{1}{N_t}\sum_{j=1}^{N_t}\left(\widehat{\kappa}_j-\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t,\widehat{\theta})\right)^2 \nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t^n,\widehat{\theta})\, \nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t^n,\widehat{\theta})', \\ \widehat{\mathbf{\Omega}}_T^{T+1,T+1} = (\frac{\sum_{t=1}^{T}N_t}{T})\sum_{t=1}^{T}\frac{1}{N_t^2}\sum_{j=1}^{N_t}\left(\widehat{\kappa}_j-\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t^n,\widehat{\theta})\right)^2 \nabla_{\theta}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t^n,\widehat{\theta})\, \nabla_{\theta}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t^n,\widehat{\theta})', \\ \widehat{\mathbf{\Omega}}_T^{t,T+1} = \left(\widehat{\mathbf{\Omega}}_T^{T+1,t}\right)' = \sqrt{\frac{\sum_{t=1}^{T}N_t}{TN_t^3}}\sum_{j=1}^{N_t}\left(\widehat{\kappa}_j-\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t^n,\widehat{\theta})\right)^2 \nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t^n,\widehat{\theta})\, \nabla_{\theta}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t^n,\widehat{\theta})'. \end{cases} \tag{10}$$

Based on equations (9) and (10), as well as the limit result in (7), pivotal tests such as t-tests for the parameters are readily constructed. This is a by-product of the stable convergence in equation (7), which ensures that the result holds jointly with the convergence in probability of $\widehat{H}_T$ and $\widehat{\Omega}_T$ to their (random) asymptotic limits.

We also note that Theorem 2 allows for conditional heteroskedasticity in the option price error. When the latter is present, the estimator in (5) is inefficient in the sense that the observations are not weighted optimally. The next theorem accommodates the use of weighted least squares (WLS).

Theorem 4 *Under the setting of Theorem 2, suppose there exist*

$$\widehat{\phi}_{t,k,\tau} \xrightarrow{\mathbb{P}} \phi_{t,k,\tau} > 0, \quad \text{uniformly on } [\underline{k}(t,\tau),\overline{k}(t,\tau)], \tag{11}$$

for $\tau \in \Gamma$ and $t = 1,...,T$, as $\underline{N} \to \infty$ and $n \to \infty$. Define

$$\left(\{\overline{\mathbf{S}}_t^n\}_{t=1,\dots,T},\,\overline{\theta}^n\right) = \underset{\{\mathbf{Z}_t\}_{t=1,\dots,T},\,\theta\in\Theta}{\operatorname{argmin}} \sum_{t=1}^{T}\left\{\sum_{j=1}^{N_t}\frac{(\tilde{\kappa}_{t,k,\tau}-\kappa(k_j,\tau_j,\mathbf{Z}_t,\theta))^2}{\widehat{\phi}_{t,k,\tau}} + \lambda_n\left(\widehat{V}_t^n-\xi_1(\mathbf{Z}_t)\right)^2\right\}, \tag{12}$$

with $\lambda_n^2\,\underline{N} \to 0$. Then $\left(\{\overline{\mathbf{S}}_t^n\}_{t=1,\dots,T},\,\overline{\theta}^n\right)$ is consistent and further,

$$\sqrt{\underline{N}}\begin{pmatrix} \overline{\mathbf{S}}_1^n-\mathbf{S}_1 \\ \vdots \\ \overline{\mathbf{S}}_T^n-\mathbf{S}_T \\ \overline{\theta}^n-\theta_0 \end{pmatrix} \xrightarrow{\mathcal{L}-s} \mathbf{\Lambda}_T^{-1/2}\begin{pmatrix} \mathbf{E}_1 \\ \vdots \\ \mathbf{E}_T \\ \mathbf{E}' \end{pmatrix}, \tag{13}$$

where $\mathbf{E}_1$,...,$\mathbf{E}_T$ are $p\times 1$ vectors and $\mathbf{E}'$ is $q \times 1$ vector, all defined on an extension of the original probability space and jointly constituting an i.i.d. standard normal vector, and for the same *partitioning of $\mathbf{\Lambda}_T$ as in equation (8), the $\mathbf{\Lambda}_T$ matrix is defined by,*

$$\begin{cases} \mathbf{\Lambda}_T^{t,t} = \varsigma_t \sum_\tau \pi_t^\tau \int_{\underline{k}(t,\tau)}^{\overline{k}(t,\tau)} \frac{\phi_{t,k,\tau}^{-1}}{\psi_{t,\tau}(k)} \nabla_{\mathbf{S}} \kappa(k,\tau,\mathbf{S}_t,\theta_0) \nabla_{\mathbf{S}} \kappa(k,\tau,\mathbf{S}_t,\theta_0)' dk, \\ \mathbf{\Lambda}_T^{T+1,T+1} = \sum_{t=1}^T \varsigma_t \sum_\tau \pi_t^\tau \int_{\underline{k}(t,\tau)}^{\overline{k}(t,\tau)} \frac{\phi_{t,k,\tau}^{-1}}{\psi_{t,\tau}(k)} \nabla_\theta \kappa(k,\tau,\mathbf{S}_t,\theta_0) \nabla_\theta \kappa(k,\tau,\mathbf{S}_t,\theta_0)' dk, \\ \mathbf{\Lambda}_T^{t,T+1} = \left(\mathbf{\Lambda}_T^{T+1,t}\right)' = \varsigma_t \sum_\tau \pi_t^\tau \int_{\underline{k}(t,\tau)}^{\overline{k}(t,\tau)} \frac{\phi_{t,k,\tau}^{-1}}{\psi_{t,\tau}(k)} \nabla_{\mathbf{S}} \kappa(k,\tau,\mathbf{S}_t,\theta_0) \nabla_\theta \kappa(k,\tau,\mathbf{S}_t,\theta_0)' dk. \end{cases}$$

A consistent estimate for $\mathbf{\Lambda}_T$ is given by $\widehat{\mathbf{\Lambda}}_T$ where, for the identical partition of the matrices as used in equation (8), we set,

$$\begin{cases} \widehat{\mathbf{\Lambda}}_T^{t,t} = \frac{1}{\sum_{t=1}^T N_t} \sum_{j=1}^{N_t} \widehat{\phi}_{t,k,\tau}^{-1} \nabla_{\mathbf{S}} \kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t^n,\widehat{\theta}) \nabla_{\mathbf{S}} \kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t^n,\widehat{\theta})', \\ \widehat{\mathbf{\Lambda}}_T^{T+1,T+1} = \frac{1}{\sum_{t=1}^T N_t} \sum_{j=1}^{N_t} \widehat{\phi}_{t,k,\tau}^{-1} \nabla_\theta \kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t^n,\widehat{\theta}) \nabla_\theta \kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t^n,\widehat{\theta})', \\ \widehat{\mathbf{\Lambda}}_T^{t,T+1} = \frac{1}{\sum_{t=1}^T N_t} \sum_{j=1}^{N_t} \widehat{\phi}_{t,k,\tau}^{-1} \nabla_{\mathbf{S}} \kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t^n,\widehat{\theta}) \nabla_\theta \kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t^n,\widehat{\theta})'. \end{cases}$$

The weighting of the observations in (12) renders the (conditional) expected Hessian of the objective function equal to the limiting (conditional) covariance matrix of the gradient or score of the objective function, evaluated at $(\{\mathbf{S}_t\}_{t=1,\ldots,T},\, \theta_0)$. This implies that the asymptotic conditional covariance matrix of $\left(\{\overline{\mathbf{S}}_t^n\}_{t=1,\ldots,T},\, \overline{\theta}^n\right)$ is Λ_T^{-1}. The derivation of the WLS estimator in Theorem 4 depends critically on the stable convergence. As in the classical theory for M-estimators, the proof exploits Slutsky's theorem which implies $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, Y)$ for two sequences $X_n \xrightarrow{\mathcal{L}} X$ and $Y_n \xrightarrow{\mathbb{P}} Y$, but, importantly, only when Y is *non-random*. This result may be extended to the case where the limit Y is *random* only under the stable form of convergence, see, e.g., equation (2.2.5) in Jacod and Protter (2012). For our WLS estimator, the limits of the weights and the elements of the Hessian matrix are random, unlike the classical WLS estimator for which they are non-random constants.

Theorem 4 references a generic consistent estimator of the conditional asymptotic variance of the option error, $\phi_{t,k,\tau}$. A consistent estimate for the conditional variance of $\epsilon_{t,k,\tau}$, as a function of k (parametric or nonparametric), for each pair (t,τ), may be constructed in a manner similar to the standard WLS estimators, see, e.g., Robinson (1987) and Newey and McFadden (1994).

The assumption, $\lambda_n^2 \underline{N} \to 0$, in Theorem 2 (and Theorem 4) ensures that the penalty term in (5) (and (12)) has no first-order asymptotic effect in the estimation. This is convenient from an empirical point of view as only boundedness in probability, and not consistency, is required of $\widehat{V}_t^n$. Hence, we can apply the estimation procedure even in settings where $\widehat{V}_t^n$ may be mildly biased due to microstructure noise without having to perform noise-robust corrections.[9]

[9]We are grateful to a referee for pointing this out. Despite this convenient property, we exploit sufficiently sparse high-frequency return observations in our empirical application in Section 6, that this bias will be negligible.

We can extend the above analysis to cover scenarios in which the penalty term is reflected in the limiting distribution. The requirement is that we can establish the joint limiting distribution of the nonparametric estimator $\widehat{V}_t^n$ and the empirical processes arising from the option pricing error – determining the limit in (7). This is feasible for the two nonparametric jump-robust realized volatility estimators defined below,

$$\widehat{V}_t^{\pm,n} = \frac{n}{k_n} \sum_{i \in I^{\pm,n}} (\Delta_i^{t,n} X)^2 \, 1\left(|\Delta_i^{t,n} X| \leq \alpha n^{-\varpi} \right), \quad \Delta_i^{t,n} X = \log\left(X_{t+\frac{i}{n}} \right) - \log\left(X_{t+\frac{i-1}{n}} \right), \tag{14}$$

where $\alpha > 0$, $\varpi \in (0, 1/2)$, k_n denotes a deterministic sequence with $k_n/n \to 0$ and,

$$I^{-,n} = \{-k_n + 1, ..., 0\} \quad \text{and} \quad I^{+,n} = \{1, ..., k_n\}.$$

$V_t^{-,n}$ and $V_t^{+,n}$ are estimators for the spot variance from the left and right, respectively, and may be viewed as "localized" versions of the truncated variation estimator proposed originally by Mancini (2001).[10] If we denote the set of jump times for the variance process by $\mathbf{J} = \{s : \Delta V_s > 0\}$, then, under weaker regularity conditions than in Assumption A0, $V_t^{+,n}$ and $V_t^{-,n}$ are both consistent for V_t, provided $t \notin \mathbf{J}$. We only need to estimate the spot volatility for a finite number of points in time. Since the jump compensator, controlling the discontinuities in V_t, is absolutely continuous in t, the probability of having jumps at any discrete time point is zero, since, almost surely, $t \notin \mathbf{J}$.

The theorem below provides the joint limit distribution of $\widehat{V}_t^{\pm,n}$ and the option-based $\widehat{\mathbf{S}}_t^n$. Defining the coefficient, $\gamma = (2-\beta)\varpi \bigwedge \left[(\frac{1}{\beta} \wedge 1) - 1/2 + (\,[(1-\beta)\varpi] \vee 0\,) \right] \bigwedge 1/4$, we have,

Theorem 5 *Under assumption A0, provided $k_n \to \infty$ with $\sqrt{k_n}/n^\gamma \to 0$, and with β defined as in A0(iii), we have for $T \in \mathbb{N}$,*

$$\sqrt{k_n} \begin{pmatrix} \widehat{V}_1^{+,n} - V_1 \\ \vdots \\ \widehat{V}_T^{+,n} - V_T \end{pmatrix} \xrightarrow{\mathcal{L}-s} \begin{pmatrix} \sqrt{2} V_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sqrt{2} V_T \end{pmatrix} \begin{pmatrix} \widetilde{E}_1 \\ \vdots \\ \widetilde{E}_T \end{pmatrix}, \tag{15}$$

where $\left(\widetilde{E}_1, ..., \widetilde{E}_T \right)'$ is a $T \times 1$ standard normal vector independent of the original filtration $\mathcal{F}$ and defined on an extension of the original probability space.

If the conditions of Theorem 2 hold, the vector $\left(\widetilde{E}_1, ..., \widetilde{E}_T \right)'$ is independent from the vector $(\mathbf{E}_1, ..., \mathbf{E}_T, \mathbf{E}')'$ determining the limit distribution of $\left(\widehat{\mathbf{S}}_1^n, ..., \widehat{\mathbf{S}}_T^n \right)'$ in equation (7).

Moreover, If $(1, ..., T) \cap \mathbf{J} = \emptyset$, the results remain valid for $\widehat{V}_t^{-,n}$ replacing $\widehat{V}_t^{+,n}$, $t = 1, ..., T$.

[10] We stress that, for practical implementation, it is important to let the truncation level reflect the pronounced intraday pattern in volatility – and we do so.

It is optimal to choose ϖ close to 1/2 and, next, k_n close to $n^{1/2}$, provided the jumps are not too active, i.e., their activity index satisfies $\beta < 4/3$ – a fairly mild restriction (recall Assumption A0(ii)).[11] The theorem reveals that the convergence of $\widehat{V}_t^{\pm,n}$ holds jointly with that of $\widehat{\mathbf{S}}_t^n$ and they are asymptotically independent conditional on the filtration of the original probability space.

Using Theorem 5, we can extend Theorem 2 to the case where the penalization has an asymptotic effect on the estimation (an analogous extension holds for the WLS estimator in Theorem 4).

Theorem 6 *Assume A1-A4 are satisfied and $\kappa(t, \tau, \mathbf{Z}, \theta)$ is twice continuously-differentiable in its arguments. Let $\widehat{V}_t^n = \widehat{V}_t^{-,n}$ for $\widehat{V}_t^{-,n}$ defined in (14) and the conditions of Theorem 5 hold. Then, if $\underline{N} \to \infty$ and $n \to \infty$, such that $\lambda_n \to \lambda \geq 0$ and $\lambda_n \sqrt{N/k_n} \to \overline{\lambda} \geq 0$, we have:*

$$\begin{pmatrix} \sqrt{N_1}\,(\widehat{\mathbf{S}}_1^n - \mathbf{S}_1) \\ \vdots \\ \sqrt{N_T}\,(\widehat{\mathbf{S}}_T^n - \mathbf{S}_T) \\ \sqrt{N/T}\,(\widehat{\theta}^n - \theta_0) \end{pmatrix} \xrightarrow{\mathcal{L}-s} (\mathbf{H}_T + \mathbf{D}_T)^{-1}\,(\mathbf{\Omega}_T + \Sigma_T)^{1/2} \begin{pmatrix} \mathbf{E}_1 \\ \vdots \\ \mathbf{E}_T \\ \mathbf{E}' \end{pmatrix}, \tag{16}$$

where $\mathbf{E}_1$,...,$\mathbf{E}_T$ are $p \times 1$ vectors and $\mathbf{E}'$ is a $q \times 1$ vector, all defined on an extension of the original probability space and all mutually independent standard normal vectors. Furthermore, with the same partitioning of $\mathbf{D}_T$ and Σ_T as in equation (8), we define,

$$\mathbf{D}_T^{t,t} = \lambda \nabla_{\mathbf{S}}\xi_1(\mathbf{S}_t)\,\nabla_{\mathbf{S}}\xi_1(\mathbf{S}_t)', \quad \Sigma_T^{t,t} = \overline{\lambda}^2 \varsigma_t 2V_t^2 \nabla_{\mathbf{S}}\xi_1(\mathbf{S}_t)\,\nabla_{\mathbf{S}}\xi_1(\mathbf{S}_t)', \quad t = 1, ..., T,$$

where $\lambda = \lim_n \lambda_n$ and the remainder of the elements in $\mathbf{D}_T$ and Σ_T are zero. Consistent estimates for $\mathbf{D}_T$ and Σ_T are given by $\widehat{\mathbf{D}}_T$ and $\widehat{\Sigma}_T$ where, for the same partition of the matrixes as in (8),

$$\widehat{\mathbf{D}}_T^{t,t} = \lambda_n \nabla_{\mathbf{S}}\xi_1(\widehat{\mathbf{S}}_t)\nabla_{\mathbf{S}}\xi_1(\widehat{\mathbf{S}}_t)', \quad \widehat{\Sigma}_T^{t,t} = \frac{\lambda_n^2 N_t}{k_n} 2(\widehat{V}_t^n)^2 \nabla_{\mathbf{S}}\xi_1(\widehat{\mathbf{S}}_t)\nabla_{\mathbf{S}}\xi_1(\widehat{\mathbf{S}}_t)', \quad t = 1, ..., T,$$

and the rest of the elements of $\widehat{\mathbf{D}}_T$ and $\widehat{\Sigma}_T$ are zero.

The conditions on the sequence λ_n in Theorem 6 are weaker than those in Theorem 2, as $k_n \to \infty$ (recall Theorem 5). The penalization term affects both the (conditional) expected Hessian of the objective function and the (conditional) covariance matrix of the score of the objective function evaluated at $(\{\mathbf{S}_t\}_{t=1,\ldots,T},\, \theta_0)$.

Theorem 6 allows us to study the effect of the penalization in the objective function more formally. We distinguish two cases. First, $N \gg k_n$. Then we have $\lambda = 0$ and thus $\mathbf{D}_T = 0$.

[11] The relative speed condition between k_n and n in Theorem 5 can be weakened slightly, if $\beta \geq 1$, at the cost of more lengthy derivations.

Consequently, the penalty term in the objective function increases the covariance matrix of the estimator. Hence, it is optimal to pick λ_n to ensure $\overline{\lambda} = 0$, i.e., choose λ_n such that the penalization has no first-order asymptotic effect on the estimator. This is intuitive. Since $N \gg k_n$, the recovery of the volatility state is done more efficiently via the options data. Comparatively speaking, the high-frequency data only infuse noise into the estimation procedure.

For $N \ll k_n$ or $N \sim k_n$, it may be preferable to have $\lambda > 0$. The logic is transparent in the (infeasible) case of $k_n = \infty$ (and $N \ll k_n$), i.e., a continuous record of X, from which V_t may be recovered without error. Then option prices (observed with error) are suboptimal for recovering V_t. Of course, the diffusive volatility is only a minor part of the full state and parameter vector, $(\{\mathbf{S}_t^n\}_{t=1,\ldots,T}, \theta_0)$, that we seek to estimate, so options remain critical for the inference.

4 Pathwise Risk-Neutral Model Tests

The hypothesis that our model for the risk-neutral dynamics is well-specified has numerous implications. The previous section develops the limit theory necessary to devise formal tests for this hypothesis. We propose a battery of diagnostics, falling into three categories: the first concerns the fit to the option surface, the second checks for stability of the risk-neutral parameters, and the third assesses the equality between the option-implied volatility and a nonparametric volatility estimate based on high-frequency data. The tests are all pathwise as they involve restrictions on the *observed path* of the option surface and the underlying asset price. Importantly, they do not restrict the statistical law for X, beyond what is implied by the risk-neutral law. As such, they do not rely on a joint hypothesis that the model is correctly specified under both the $\mathbb{P}$ and $\mathbb{Q}$ measures. For simplicity in this section, we derive all tests for the original setup in Theorem 2, i.e., the case of ordinary least squares in which the penalization has no first-order asymptotic effect.

4.1 Option Price Fit

We first develop a test based on the fit afforded by the parametric model. The previous section supplies us with tools to formally separate *observation errors* from *model misspecification errors* in fitting the option prices. The corollary below provides a t-test that captures the quality of the model fit to the option surface at a specific point in time for a given tenor.

Corollary 1 *Let $\mathcal{K} \subset \left(\underline{k}(t,\tau^*)\ \overline{k}(t,\tau^*)\right)$ be a set with positive Lebesgue measure and denote by $N_t^{\mathcal{K}}$ the number of options on day t with tenor τ^* and log-moneyness belonging to the set $\mathcal{K}$. Then, under the assumptions of Theorem 2, we have,*

$$\frac{\sum_{j:k_j\in\mathcal{K}}\left(\tilde{\kappa}_{t,k_j,\tau^*}-\kappa(k_j,\tau^*,\widehat{\mathbf{S}}_t^n,\widehat{\theta}^n)\right)}{\sqrt{\widehat{\Pi}_T'\widehat{\Xi}_T\widehat{\Pi}_T}} \stackrel{\mathcal{L}-s}{\longrightarrow} \mathcal{N}(0,1), \quad \widehat{\Xi}_T=\begin{pmatrix}\widehat{\mathbf{H}}_T^{-1}\widehat{\mathbf{\Omega}}_T(\widehat{\mathbf{H}}_T^{-1})' & \widehat{\mathbf{H}}_T^{-1}\widehat{\mathbf{\Upsilon}}_{1,T}\\ \widehat{\mathbf{\Upsilon}}_{1,T}'(\widehat{\mathbf{H}}_T^{-1})' & \widehat{\mathbf{\Upsilon}}_{2,T}\end{pmatrix}, \quad (17)$$

$$\widehat{\mathbf{\Upsilon}}_{1,T}=\begin{pmatrix}\mathbf{0}_{(t-1)p\times 1}\\ \frac{1}{\sqrt{N_tN_t^{\mathcal{K}}}}\sum_{j:k_j\in\mathcal{K}}\left(\tilde{\kappa}_{t,k_j,\tau^*}-\kappa(k_j,\tau^*,\widehat{\mathbf{S}}_t^n,\widehat{\theta}^n)\right)^2\nabla_{\mathbf{S}}\kappa(k_j,\tau^*,\widehat{\mathbf{S}}_t^n,\widehat{\theta}^n)\\ \mathbf{0}_{(T-t+1)p\times 1}\\ \frac{1}{\sqrt{N_t^{\mathcal{K}}N_t}}\sqrt{\frac{\sum_{t=1}^T N_t}{T}}\sum_{j:k_j\in\mathcal{K}}\left(\tilde{\kappa}_{t,k_j,\tau^*}-\kappa(k_j,\tau^*,\widehat{\mathbf{S}}_t^n,\widehat{\theta}^n)\right)^2\nabla_{\theta}\kappa(k_j,\tau^*,\widehat{\mathbf{S}}_t^n,\widehat{\theta}^n)\end{pmatrix},$$

$$\widehat{\mathbf{\Upsilon}}_{2,T}=\frac{1}{N_t^{\mathcal{K}}}\sum_{j:k_j\in\mathcal{K}}\left(\tilde{\kappa}_{t,k_j,\tau^*}-\kappa(k_j,\tau^*,\widehat{\mathbf{S}}_t^n,\widehat{\theta}^n)\right)^2,$$

$$\widehat{\mathbf{\Pi}}_T=\begin{pmatrix}\mathbf{0}_{1\times(t-1)p} & \frac{-1}{\sqrt{N_t}}\sum_{j:k_j\in\mathcal{K}}\nabla_{\mathbf{S}}\kappa(k_j,\tau^*,\widehat{\mathbf{S}}_t^n,\widehat{\theta}^n)' & \mathbf{0}_{1\times(T-t+1)p} & -\sqrt{\frac{T}{\sum_{t=1}^T N_t}}\sum_{j:k_j\in\mathcal{K}}\nabla_{\theta}\kappa(k_j,\tau^*,\widehat{\mathbf{S}}_t^n,\widehat{\theta}^n)' & \sqrt{N_t^{\mathcal{K}}}\end{pmatrix}.$$

The logic behind the test in Corollary 1 is straightforward. By aggregating the model-implied option fit spatially, we “average out,” and thus alleviate, the effect due to the observation error in the options but we retain the error due to inadequate model fit. Hence, for the result in equation (17) to apply, it is necessary that $\mathcal{K}$ has positive Lebesgue measure and that $\kappa(k,\tau,\mathbf{Z},\theta)$ is a smooth function of log-moneyness. The t-statistics implied by the asymptotic limit result in equation (17) resemble the conditional moment tests proposed by Newey (1985) and Tauchen (1985).

The asymptotic variance of the option fit $\sum_{j:k_j\in\mathcal{K}}\left(\tilde{\kappa}_{t,k_j,\tau^*}-\kappa(k_j,\tau^*,\widehat{\mathbf{S}}_t^n,\widehat{\theta}^n)\right)$, is estimated feasibly by $\widehat{\Pi}_T'\widehat{\Xi}_T\widehat{\Pi}_T$. It accounts for the effect of the estimation error of $\left(\widehat{\mathbf{S}}_t^n,\widehat{\theta}^n\right)$. It is critical for the derivation of Corollary 1 that the convergence in equation (7) holds stably so that the standardization of the model fit in equation (17) yields a variable with a limiting standard normal distribution. The test in equation (17) can, of course, be extended to pool together the estimated errors across options with different tenors as well as for options observed on different days.

The test will be powerful against alternatives for which the errors in fitting the options in the region $\mathcal{K}$ tend to be highly correlated, as this “blows up” the numerator without affecting the denominator of the ratio in (17). This will typically be the case, as standard models imply smoothness in option prices as a function of moneyness. That is, if the fit is poor for a given strike, due to model misspecification, the model-implied option prices will tend to deviate in the same direction for nearby strikes. Furthermore, the test of Corollary 1 allows us to check the model fit over shorter periods of time. This is more informative about potential sources of model failure than assessing the time-averaged option price fit, as is common practice. For example, we may be able to associate specific types of model failure with broader economic developments that point towards omitted state variables in the model or a fundamental lack of stability in the risk-neutral measure.

4.2 Time-Variation in Parameter Estimates

Our second test is based on the variation of the risk-neutral parameters over time. Under standard regularity conditions, model misspecification will imply that the estimator converges to a pseudo-true parameter vector, see, e.g., White (1982) and Gourieroux et al. (1984). However, in our setting the state variables change from period to period. This should induce a corresponding movement in the pseudo-true parameter vector for the misspecified model. That is, under misspecification, the time-variation in the option prices cannot be "rationalized" by shifts in the state variables, so we should expect "spill-over" in terms of intertemporal variation in the (pseudo-true) risk-neutral parameter estimates over distinct time periods.

Designing a test for parameter variation is straightforward using Theorem 2, as the estimates obtained from option panels spanning disjoint time periods should be asymptotically independent when conditioned on the filtration of the original probability space.[12]

Corollary 2 *In the setting of Theorem 2, denote the risk-neutral parameter estimates from two option panels covering disjoint time periods by* $\widehat{\theta}_1^n$ *and* $\widehat{\theta}_2^n$. *If the risk-neutral model is valid for both of these distinct time periods, we have,*

$$\left(\widehat{\theta}_1^n - \widehat{\theta}_2^n\right)' \left(\widehat{Avar}(\widehat{\theta}_1^n) + \widehat{Avar}(\widehat{\theta}_2^n)\right)^{-1} \left(\widehat{\theta}_1^n - \widehat{\theta}_2^n\right) \xrightarrow{\mathcal{L}-s} \chi^2(q), \tag{18}$$

where $\widehat{Avar}(\widehat{\theta}_1^n)$ *and* $\widehat{Avar}(\widehat{\theta}_2^n)$ *denote consistent estimates of the asymptotic variances of* $\widehat{\theta}_1^n$ *and* $\widehat{\theta}_2^n$ *based on equations (9)-(10) in Theorem 3, and* q *denotes the dimension of the parameter vector. The analogous result applies for a subset of the parameter vector of dimension* $r < q$, *but with* r *replacing* q *in equation (18).*

4.3 Distance between Model-Free and Option-Implied Volatility

Our final diagnostic tests whether the spot volatility estimated nonparametrically from high-frequency data on the underlying asset is equal to the spot volatility, V_t, implied by the option data given the model for the risk-neutral distribution of X. This restriction follows from the fact that the diffusion coefficient of X should be invariant under an equivalent measure change (recall $\mathbb{P}$ and $\mathbb{Q}$ are locally equivalent). Hence, if the option price dynamics is successfully captured by the state

[12] Of course, this can be generalized to the case of overlapping estimation periods by appropriately accounting for the conditional covariance of the two parameter estimates.

vector $\mathbf{S}_t$, the two estimates should not be statistically distinct. This is, of course, the identical constraint that we exploit in our penalization term during estimation. Nonetheless, the condition may be formally tested if we account suitably for the specification of the objective function in (5).

To render the test feasible, we use the two nonparametric jump-robust volatility estimators $\widehat{V}_t^{\pm,n}$ defined in (14) in the previous section. Using the joint asymptotic convergence result for $(\widehat{V}_t^{\pm,n}, \widehat{\mathbf{S}}_t^n)$ in Theorem 5, we can derive the asymptotic behavior of the difference $\widehat{V}_t^{\pm,n} - \xi_1(\widehat{\mathbf{S}}_t^n)$. We state this important result in the following corollary.

Corollary 3 *Under the conditions of Theorems 2 and 5, we have for* $k_n \to \infty$, $\underline{N} \to \infty$ *and* $\lambda_n^2 \underline{N} \to 0$,

$$\left\{ \frac{\xi_1(\widehat{\mathbf{S}}_t^n) - \widehat{V}_t^{-,n}}{\sqrt{\frac{\nabla_{\mathbf{S}}\xi_1(\widehat{\mathbf{S}}_t^n)'\widehat{\chi}_t\nabla_{\mathbf{S}}\xi_1(\widehat{\mathbf{S}}_t^n)}{N_t} + \frac{2(\widehat{V}_t^{-,n})^2}{k_n}}} \right\}_{t=1,\ldots,T} \xrightarrow{\mathcal{L}-s} \begin{pmatrix} \breve{E}_1 \\ \vdots \\ \breve{E}_T \end{pmatrix}, \tag{19}$$

where $\widehat{\chi}_t$ *is the part of* $\widehat{\mathbf{H}}_T^{-1}\widehat{\mathbf{\Omega}}_T(\widehat{\mathbf{H}}_T^{-1})'$ *corresponding to the variance-covariance of* $\widehat{\mathbf{S}}_t^n$ *and where* $(\breve{E}_1, ..., \breve{E}_T)'$ *is a vector of standard normals independent of each other and of* $\mathcal{F}$.

Yet again, we stress that we do not need a parametric model for V_t under the statistical measure, $\mathbb{P}$, to test the equality of the spot volatility implied by the underlying asset dynamics and the model-dependent option-implied dynamics. However, the test does hinge critically on the characterization of the joint stable asymptotic law in Theorem 5. Consequently, this pathwise restriction on the spot volatility cannot be formally tested under the usual approach to option-based parametric inference which precludes the application of this type of limit theory.

The test in Corollary 3 compares two alternative estimators of the spot diffusive volatility: a parametric one based on the option data and a nonparametric one based on the high-frequency record for X. As discussed after Theorem 6, depending on the relative growth of the option data and the high-frequency increments used in estimation of $\widehat{V}_t^{\pm,n}$, we can have either the option-based or the high-frequency based estimator be more efficient for recovery of V_t. In typical applications with a rich set of derivatives data (as is the case in our empirical application later), however, the option-based estimator tends to dominate and hence the second term in the denominator on the left side of (19) is usually substantially larger than the first one. In this case, i.e., for $N \gg k_n$, our test in Corollary 3 is reminiscent of the Hausman (1978) specification test.[13] Under the null of correct risk-neutral model specification, $\xi_1(\widehat{\mathbf{S}}_t^n)$ is more efficient than $\widehat{V}_t^{\pm,n}$ while $\widehat{V}_t^{\pm,n}$ is a robust

[13]We thank a referee for pointing out this connection.

estimator of the spot volatility, which remains valid (i.e., consistent and asymptotically normal) even when the risk-neutral parametric model is misspecified.[14]

Unfortunately, we cannot design a similar test regarding the distance between the option-based estimate of the jump intensity $\xi_2(\widehat{\mathbf{S}}_t^n)$ and a nonparametric one derived from high-frequency data. First, while high-frequency data for X allows us to estimate the "realized" jumps on a given path, it does not produce an estimate of their intensity. The jump intensity depends on the probability measure, and reliable estimation will require applying large time span asymptotics under the $\mathbb{P}$ measure. Secondly, since the jump intensity is tied to the probability distribution, the jump intensity under the risk-neutral and statistical distribution are generally different. In fact, there is strong parametric and nonparametric evidence indicating that they differ significantly.[15]

Finally, to increase power, the tests in Corollaries 1-3 should be applied in parallel. For example, a misspecified model might generate spot volatility estimates that are close to the model-free ones, but in so doing provide a poor fit to option prices or induce parameter instability. Likewise, a faulty model may fit parts of the option panel well – the vector $\{\mathbf{S}_t\}_{t=1,\dots,T}$ provides flexibility in this regard – but this typically produces implausible volatility estimates or unstable parameters.

5 Numerical Experiments

5.1 Model Specification and Parameter Identification

This section provides evidence on the finite sample performance of our inference procedures in the context of a model widely exploited in empirical work, namely the so-called "double-jump" model of Duffie et al. (2000). The model *under the risk-neutral distribution* is specified as,

$$\frac{dX_t}{X_{t-}} = \sqrt{V_t}\, dW_t + dL_{x,t}\,, \qquad dV_t = \kappa_d\,(\overline{v} - V_t)\,dt + \sigma_d\,\sqrt{V_t}\,dB_t + dL_{v,t}\,, \tag{20}$$

where (W_t, B_t) is a two-dimensional Brownian motion with correlation $\operatorname{corr}(B_t, W_t) = \rho_d$; $(L_{x,t}, L_{v,t})$ is a compound Poisson jumps process with intensity λ_j and the distribution of the jump size $(Z_x,\ Z_v)$ is governed by the marginal distribution of Z_v, which is exponential with mean μ_v, while, conditional on Z_v, $\log(Z_x + 1)$ is Gaussian with mean $\mu_x + \rho_j Z_v$ and standard deviation σ_x, and, finally, $L_{x,t}$ is a jump martingale. The model also involves the cross-parameter restriction, $\sigma_d \le \sqrt{2\,\kappa_d\,\overline{v}}$.

[14] Note that in this situation, i.e., for $N \gg k_n$, $\xi_1(\widehat{\mathbf{S}}_t^n)$ is converging at a faster rate than $\widehat{V}_t^{\pm,n}$, so that the asymptotic behavior of $\xi_1(\widehat{\mathbf{S}}_t^n) - \widehat{V}_t^{\pm,n}$ is driven by $\widehat{V}_t^{\pm,n}$.

[15] The aspects of the risk-neutral model for jumps in X we can test from the underlying asset data are those that hold $\mathbb{Q}$-almost surely. This includes the so-called jump activity index, which should be identical under $\mathbb{P}$ and $\mathbb{Q}$. However, to uncover the latter nonparametrically from high-frequency data, we must sample X very finely and this renders the inference highly sensitive to market microstructure effects. Hence, we abstain from testing this restriction.

Finally, for simplicity, we have fixed the risk-free rate and the dividend yield to be zero. The vector of risk-neutral parameters is thus given by $\theta = (\rho_d, \overline{v}, \kappa_d, \sigma_d, \lambda_j, \mu_x, \sigma_x, \mu_v, \rho_j)$.

To ensure that our numerical experiments reflect empirically relevant features of the asset and option price dynamics, we fix the parameters to the consensus values from the literature provided by Broadie et al. (2009). Although our inference procedure only requires a full characterization of the data generating process under the risk-neutral measure, in the simulation experiment, we still need to generate the dynamics of the state variables from the actual probability measure. Hence, we adopt the standard approach of the empirical option pricing literature, see, e.g., Singleton (2006), chapter 15, and assume that X follows the same general model under both the $\mathbb{P}$ and $\mathbb{Q}$ measures, but with differing values for some key parameters, reflecting the presence of risk premiums. The full set of parameter values, adapted from Broadie et al. (2009), is reported in Table 1. We also follow them in fixing $\rho_j = 0$, leaving eight parameters to be estimated for each Monte Carlo sample.

Table 1: Parameter Setting for the Numerical Experiments

Under $\mathbb{P}$				Under $\mathbb{Q}$			
Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
ρ_d	-0.4600	λ_j	1.0080	ρ_d	-0.4600	λ_j	1.0080
$\overline{v}$	0.0144	μ_x	-0.0284	$\overline{v}$	0.0144	μ_x	-0.0501
κ_d	4.0320	σ_x	0.0490	κ_d	4.0320	σ_x	0.0751
σ_d	0.2000	μ_v	0.0315	σ_d	0.2000	μ_v	0.0930

Figure 2 depicts the sensitivity of the option surface with respect to the parameters of the double-jump model across different values of the state variable, i.e., alternative levels of (stochastic) volatility. The figure reveals that the parameters have qualitative different effects on the option surface. This should ensure that the parameters can be identified in practice as long as the option cross-section is sufficiently wide and the option panel spans a time period with a significant degree of variation in the realization of the (volatility) state vector. For example, the long-run mean of the volatility parameter, $\overline{v}$, primarily impacts the longer term options while the short maturity options are determined largely by the current volatility state. The same logic applies to the identification of the mean reversion parameter, κ_d. Not surprisingly, the sensitivity of options with respect to κ_d increases when the (stochastic) volatility state is far from its long-run mean as this enhances the strength of the mean-reversion. Likewise, turning to the jump parameters, it is evident that each of them has a unique effect on the option surface. For example, μ_x has the largest impact on OTM short-maturity put options and its effect decreases for longer maturities. On the other hand, the volatility of jumps parameter, σ_x, has a more symmetric impact on the short maturity puts and calls with a diminishing effect for the longer maturities. Overall, Figure 2 reveals that there are large benefits from using options with a wide range of strikes and levels of moneyness as well as from pooling observations across different days in the estimation.

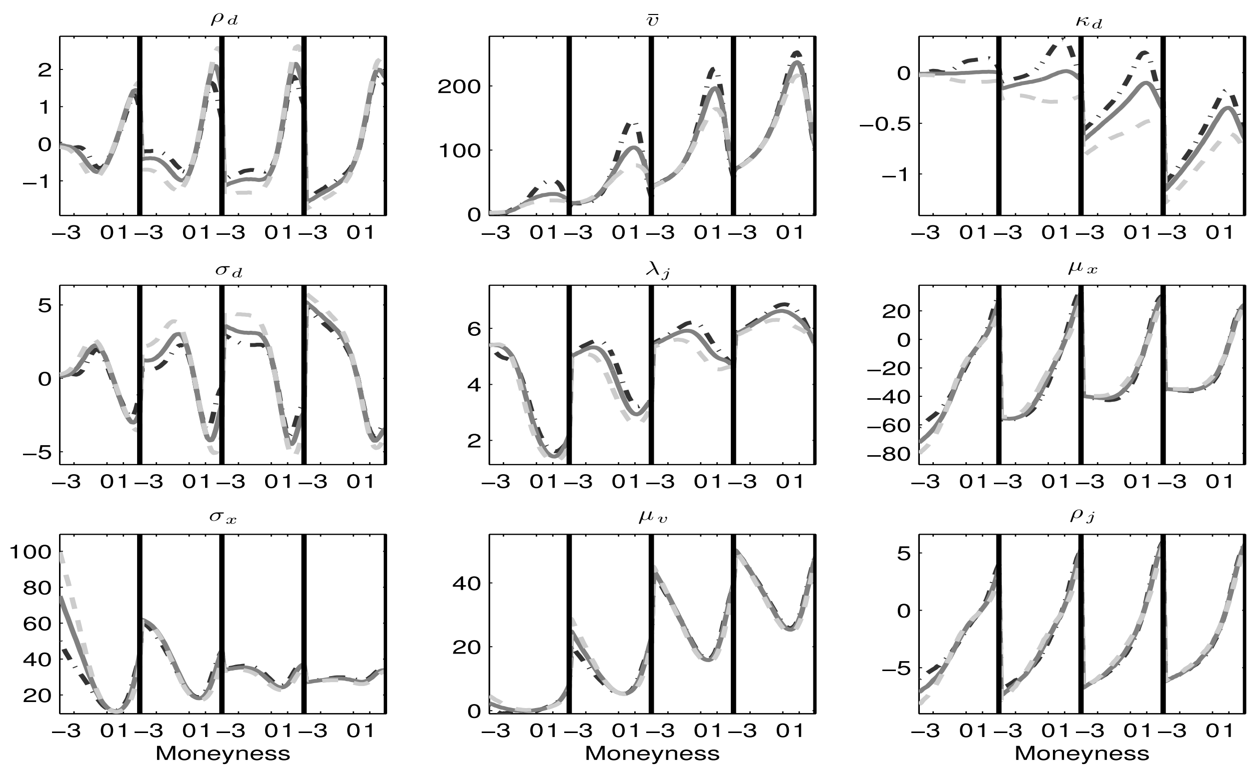

Figure 2: *Option Sensitivity to Parameters in Double-Jump Model.* The figure plots the first derivatives of options (in terms of implied volatility) with respect to the parameters of the double-jump model. The parameters are set at the values reported in Table 1. Moneyness is reported in terms of volatility units, i.e., $\log(K/F)/(\sigma\sqrt{T})$ with notation as in Figure 1. Each of the segments in the plots corresponds to maturities $\tau = 10$, $\tau = 45$, $\tau = 120$ and $\tau = 252$ days (starting from left to right). The solid, dashed and dotted lines correspond to estimates at the 5th, 50th and 95th quantile of volatility respectively.

5.2 Monte Carlo Experiments

We now present the findings from an extensive simulation study based on the double-jump model with parameters fixed at the values given in Table 1. We apply our inference procedures on a total of 1,000 Monte Carlo replications.[16]

[16]The computational burden is very significant, but a variety of improvements to the speed in computing option prices within this framework and the use of a network of computers renders the exercise feasible. To the best of our

The Monte Carlo setting aims at broadly mimicking the features of models estimated previously in the literature as well as the data used in our empirical application in Section 6. To this end, the option panel is constructed as follows. We simulate the underlying asset for a year and sample the options every fifth day, corresponding to weekly observations, as is common in empirical work (time is measured in business days). For each such day, we calculate option prices for four maturities: $\tau = 10$, $\tau = 45$, $\tau = 120$ and $\tau = 252$ days, which resemble the available maturities in the actual data. Finally, for each maturity we compute 50 out-of-the-money option prices for an equispaced log-moneyness grid, covering the range $[-4, 1] \cdot \sigma\sqrt{\tau}$, where σ is the ATM BSIV on the given day. This corresponds to using a time-varying coverage of moneyness depending on the level of volatility, again roughly mimicking the available strike ranges in the actual data.[17] For the option error we assume $\epsilon_{t,k,\tau} = \sigma_{t,k,\tau} Z_{t,k,\tau}$, where $Z_{t,k,\tau}$ are standard normal variables, independent across time, moneyness and time-to-maturity, and $\sigma_{t,k,\tau} = 0.5\psi_k/Q_{0.995}$ for ψ_k denoting the relative bid-ask estimate from the kernel regression on the actual data, plotted on Figure 1, and $Q_{0.995}$ denoting the 0.995-quantile of the standard normal distribution. This noise structure allows for significant time-variation of the (conditional) noise variance depending on both the level of volatility and moneyness. Finally, we set $\lambda_n = 0$ in (5) and, for the nonparametric volatility estimator, we use $\widehat{V}_t^{-,n}$ with $n = 400$ – equivalent to sampling every 1-minute over a 6.5 hour trading day – and $k_n = 120$, corresponding to a window of 2 hours. α and ϖ were calculated as in Bollerslev and Todorov (2011) and we refer to that source as well as the supplementary appendix for further details.

In Table 2, we report the results from the Monte Carlo for the parameter vector. The parameter estimates display no significant biases and most are estimated with good precision. The more challenging parameters to estimate, based on the relative size of the inter-quantile range, are ρ_d and σ_d. Given the relatively short one-year samples and the close relation between the option sensitivities for these two parameters, displayed in the first two panels of column one in Figure 2, this finding is not surprising. Of course, longer samples and different underlying parameter values may render it easier to separately identify the two parameters.

Turning next to the diagnostic tests, Table 3 reports on the size of the various tests developed in Section 4. Generally, the small sample behavior is quite satisfactory. The tests for the fit to the

knowledge, it constitutes the first comprehensive simulation study of inference procedures for the double-jump model at this level of complexity. An account of our computational approach is provided in the supplementary appendix.

[17] On average, over 230 bid-ask quotes with positive bid prices for OTM options are reported daily, at the end-of-trading, for the S&P 500 options at the CBOE over $1996 - 2010$, and the number is significantly higher in the second half of the sample. In the simulation, we would garner more precise inference using samples beyond one year, but this horizon provides a sensible compromise, in practice, between the joint objectives of accuracy in estimation and minimization of the period over which we assume invariance of the risk-neutral measure.

Table 2: Monte Carlo Results: Estimation of the Risk-Neutral Parameters

Parameter	True Value	Median	IQR	Parameter	True Value	Median	IQR
ρ_d	-0.4600	-0.4564	0.3911	λ_j	1.0080	1.0080	0.1798
$\overline{v}$	0.0144	0.0144	0.0035	μ_x	-0.0501	-0.0501	0.0134
κ_d	4.0320	4.0321	0.1939	σ_x	0.0751	0.0751	0.0057
σ_d	0.2000	0.2022	0.1240	μ_v	0.0935	0.0935	0.0055

option panel is near perfectly sized, while there is a mild degree of under-rejection for the volatility test as well as the omnibus test for parameter stability. The test in Panel C of Table 3 (recall $\lambda_n = 0$ in the estimation) indicates that, even in the presence of observation error, the option panel alone recovers the path of stochastic volatility with good precision.

Table 3: Monte Carlo Results: Diagnostic Tests

Test	**Nominal size of test**		
	1%	**5%**	**10%**
Panel A: Fit to Option Panel			
Out-of-the-money, short-maturity puts	0.88%	4.80%	10.03%
Out-of-the-money, short-maturity calls	0.89%	4.95%	10.00%
Out-of-the-money, long-maturity puts	1.00%	4.93%	9.96%
Out-of-the-money, long-maturity calls	0.93%	5.15%	10.47%
Panel B: Parameter Stability			
	1.48%	6.33%	10.55%
Panel C: Distance implied-nonparametric volatility			
	1.68%	4.10%	6.45%

Note: Panel A reports rejection frequencies across the full sample for the option fit to specific portions of the option surface on a given Wednesday. This test is based on the result in Corollary 1, using the first two maturities for the first two tests and the last two maturities for the remainder of the tests in this panel. The test in Panel B is given in Corollary 2, and the test in Panel C is provided in Corollary 3.

Corollary 2 can also be used to test for stability of the individual parameters. This is likely a less powerful test, as it fails to exploit the information about the joint fit across the full parameter vector and it does not account for the correlation among the estimates. It is evident from Table 4 that the tests for those parameters, which are estimated relatively imprecisely, e.g., ρ_d and σ_d, tend to be particularly undersized in small samples. In comparison, the omnibus test in Table 3 performs considerably better. We conclude that, overall, the developed method of inference appears to be reasonably sized, even for small samples, given parameters calibrated to commonly observed values.

Table 4: Monte Carlo Results: Tests for Stability of Individual Parameters

Parameter	**Nominal Size**			**Parameter**	**Nominal Size**		
	1%	**5%**	**10%**		**1%**	**5%**	**10%**
ρ_d	0.00%	0.42%	1.90%	λ_j	1.90%	6.96%	10.97%
$\overline{v}$	1.48%	6.96%	10.97%	μ_x	0.63%	6.33%	11.39%
κ_d	1.48%	6.96%	12.66%	σ_x	0.84%	5.06%	8.65%
σ_d	0.63%	4.43%	8.86%	μ_v	0.63%	5.70%	9.49%

Note: The parameter stability test is given in equation (18).

6 Empirical Application

We apply our inference procedure to data on European style S&P 500 equity-index (SPX) options traded at the CBOE. We use the closing bid and ask prices reported by OptionMetrics, applying standard filters and discarding all in-the-money options, options with a tenor of less than 7 days, and options with zero bid prices. We then compute the mid bid-ask BSIV.[18] The data span January 1, 1996 to July 21, 2010. Following earlier empirical work, e.g., Bates (2000) and Broadie et al. (2009), we sample every Wednesday. The full sample includes 760 days, and we use an average of 234 bid-ask quotes per day for estimation. The nonparametric volatility estimate needed for the penalization in the objective function and for the diagnostic tests is constructed from one-minute high-frequency data on the S&P 500 futures covering the extent of the options data. The construction of the high-frequency estimate follows the steps outlined in Section 5. In the estimation, we set $\lambda_n = 0.2$.

We first estimate a two-factor extension of the double-jump volatility model (20) which nests almost all models considered in the empirical option pricing literature (the supplementary appendix reports results for the original one-factor double-jump volatility model). The two-factor model (under the risk-neutral measure) is given by,

$$\begin{aligned}
&\frac{dX_t}{X_{t-}} = (r_t - \delta_t)dt + \sqrt{V_{1,t}}dW_{1,t} + \sqrt{V_{2,t}}dW_{2,t} + dL_{x,t}, \\
&dV_{1,t} = \kappa_{d,1}(\overline{v}_1 - V_{1,t})dt + \sigma_{d,1}\sqrt{V_{1,t}}dB_{1,t} + dL_{v,t}, \quad dV_{2,t} = \kappa_{d,2}(\overline{v}_2 - V_{2,t})dt + \sigma_{d,2}\sqrt{V_{2,t}}dB_{2,t},
\end{aligned} \tag{21}$$

where $(W_{1,t}, W_{2,t}, B_{1,t}, B_{2,t})$ is a four-dimensional Brownian motion with $W_{1,t} \perp W_{2,t}$, $W_{1,t} \perp B_{2,t}$,

[18] The two shortest maturities, beyond seven calendar days, are also used by the CBOE for the VIX computation.

and $W_{2,t} \perp B_{1,t}$ and $\text{corr}\,(W_{1,t}, B_{1,t}) = \rho_{d,1}$ as well as $\text{corr}\,(W_{2,t}, B_{2,t}) = \rho_{d,2}$. Moreover, $(L_{x,t},\ L_{v,t})$ is a bivariate jump process with intensity $\lambda_{j,0} + \lambda_{j,1} V_{1,t-}$ and a joint jump (size) distribution $(Z_1,\ Z_2)$, where the marginal distribution of Z_2 is exponential with mean μ_v and, conditional on Z_2, $\log(Z_1 + 1)$ is normal with mean $\mu_x + \rho_j Z_2$, and variance σ_x^2. For simplicity, we assume that r_t and δ_t are deterministic, where the LIBOR rate for the corresponding maturity represents the interest rate while the dividend yield is obtained from OptionMetrics.[19]

From model (21), we obtain the original one-factor double-jump model (20) if we impose $\lambda_{j,1} = 0$ and $\sigma_{d,2} = \overline{v}_2 = 0$. Importantly, the extended two-factor model (21) allows for a time-varying jump intensity. This is a feature of asset prices implied by many structural models and generated through different channels: time-varying probability of negative jump (rare disaster) in consumption growth, e.g., Wachter (2013), or in its conditional mean and/or variance, e.g., Drechsler and Yaron (2011), or through the time-varying risk aversion of the representative agent, e.g., Du (2010). We further note that jumps now impact $V_{1,t}$ directly, implying that the time-varying jump intensity, $\lambda_{j,1} > 0$, induces self-affectation: a jump today increases the probability of future jumps. Estimation results for the extended model are reported in Table 5.

Table 5: Parameter Estimates of Two-Factor Model

Parameter	Estimate	Standard Error	Parameter	Estimate	Standard Error
$\rho_{d,1}$	−0.9800	0.0253	$\lambda_{j,0}$	0.0217	0.0027
$\overline{v}_1$	0.0331	0.0019	$\lambda_{j,1}$	6.0683	0.8749
$\kappa_{d,1}$	1.2327	0.0638	μ_x	−0.0145	0.0136
$\sigma_{d,1}$	0.2640	0.0113	σ_x	0.0877	0.0082
$\rho_{d,2}$	−0.1824	0.0388	μ_v	0.1501	0.0124
$\overline{v}_2$	0.0066	0.0001	ρ_j	−0.7756	0.0718
$\kappa_{d,2}$	29.8797	0.5951			
$\sigma_{d,2}$	0.2341	0.0569			

Note: Parameter estimates for the two-factor model (21), using S&P 500 equity-index options sampled every Wednesday from January 1996 through July 2010.

The estimates imply that the first volatility factor is very persistent while the second volatility factor has a much lower mean and shorter half-life. Interestingly, the main determinant of the jump intensity is the time-varying component. This is important, as $\lambda_{j,1} > 0$ allows for time-varying jump risk premia, which have been identified by Bates (2000), Pan (2002) and Bollerslev and Todorov (2011). The mean jump intensity is 28.4%, i.e., a jump occurs about every 3-4 years, and the mean

[19]For reviews of the various approaches adopted previously for estimation and inference in the empirical option pricing literature, see, e.g., Bates (2003) and Garcia et al. (2010).

jump size is approximately -13%. These jump estimates are roughly in line with those from studies exploiting both asset returns and options, see, e.g, Singleton (2006), chapter 15, for a review. The correlation between price and volatility jumps is strongly negative. The fact that $\lambda_{j,1}$ and ρ_j both are positive imply intricate interactions between future volatility and jumps: high volatility today increases the likelihood of future jumps while the occurrence of jumps increases the future level of volatility. Finally, the estimated mean volatility is 22.8%, which is consistent with a negative volatility risk premium.

Figure 3 depicts the weekly Z-scores for the fit to the volatility surface by the two-factor model. Each of the six panels displays one Z-score for each of the 760 weekdays for a total of 4560 separate statistics. Clearly, the confidence bands are violated at a frequency dramatically exceeding the 5% level. Even more strikingly, there are systematic patterns in the violations which may indicate economically important failures of the model.[20] First of all, we notice the relatively poor fit for all long-maturity options as well as a fairly pronounced and statistically significant mispricing of short maturity OTM puts. This occurs despite the presence of two volatility factors in the model which should allow for a reasonable degree of flexibility in fitting the term structure of risk and risk premia. Importantly, there is also a distinct temporal pattern in the directional pricing errors for specific regions of the surface as well as a strong dependence in the mispricing across the different parts of the surface. For example, from the end of 1997 till the end of 1999, the model mostly underprices short-maturity OTM puts and, simultaneously, it overprices short-maturity ATM options and OTM calls (albeit not always in a statistically significant manner). Furthermore, during parts of this period we observe pronounced and persistent underpricing of long-maturity OTM put options. A similar pattern of mispricing across the option surface emerges in early 2009 and lasts through the end of our sample. On the other hand, during the tranquil period of 2004-2007, we observe a good fit for the ATM options but an underpricing of OTM put options. Obviously, from an econometric perspective, the model is strongly rejected. Rejection rates from the formal diagnostic tests developed in Section 4 – including tests for the fit to specific regions of the option surface as well as for parameter stability and coherence between the option-implied and time series estimates of spot volatility – are provided in the supplementary appendix. Clearly, the use of a broad cross-section of options enables us to exploit information regarding the fit across diverse regions of the surface as well as over time, ensuring a more powerful test for empirical option pricing models than applied hitherto. And the verdict is clear. The standard models fail

[20] Under the null hypothesis of correct model specification, there is some correlation between the Z-scores over time because of the error in estimating the parameter vector. This dependence, however, is relatively small because of the high precision in recovering the parameters.

statistically in important dimensions. In terms of the ramifications for economic interpretation, the impact of the rejections is less clear, but we take a first step in addressing this question below.

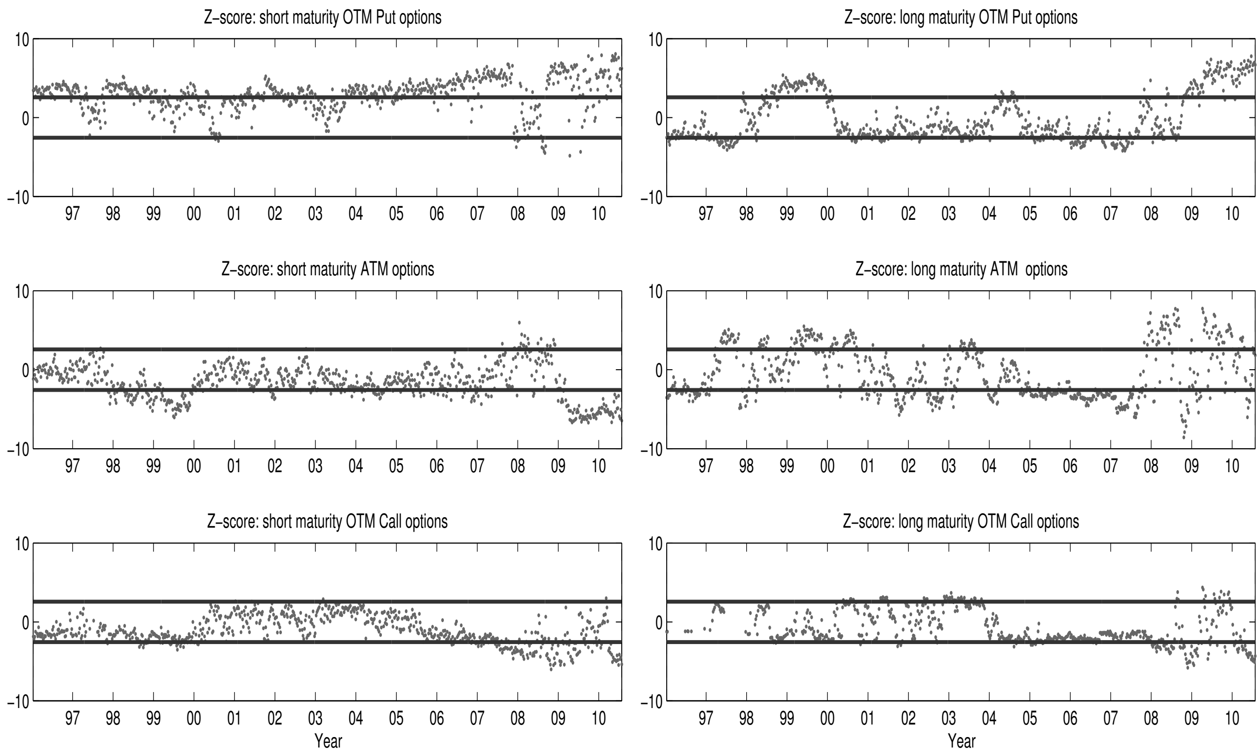

Figure 3: *Option Price Fit for the Two-factor Model.* The short-maturity options are those with the two shortest tenors available on a given day and the long-maturity options are all remaining options with a tenor of less than one year. OTM signifies out-of-the-money and ATM is at-the-money. OTM puts are those with moneyness $[-4, -1] \times \sigma\sqrt{\tau}$, ATM options are those with moneyness $[-1, 1] \times \sigma\sqrt{\tau}$, and OTM calls are those with moneyness $[1, 3] \times \sigma\sqrt{\tau}$, where σ is the ATM BSIV on the given day.

So what is the main source of option mispricing within our general two-factor model? The Z-scores provide guidance. The persistent bias in the fit to specific regions of the option surface suggests that part of the problem stems from the dynamic links between the different sources of risk within the model. This intuition arises from the fact that short-maturity OTM put option prices are determined largely by the left tail of the jump risk, while the short-maturity ATM options (quoted in BSIV) are determined largely by the current level of diffusive volatility. Hence, the tension in fitting, simultaneously, short maturity OTM puts and ATM options, evident in Figure 3, reflects the inability of the model to capture the relation between jump risk and the volatility level implied by the observed option prices. In model (21), this feature is tied to the jump intensity which is a linear function of the persistent component of the diffusive (spot) variance. Given the minor role played by the transient factor in the spot volatility, this defect will not be rectified by allowing the jump intensity to be linear in both volatility factors. Instead, it seems more promising to consider a mechanism that enables the left tail of the volatility surface to display variation that is not linked so tightly to the level of the volatility factors.

These points have broad implications also for popular structural models, like the long-run risk model of Drechsler and Yaron (2011), further extended to allow for ambiguity aversion in Drechsler (2013), as well as the model of time-varying rare disaster risk of Wachter (2013). These models all imply that the risk-neutral jump intensity of the market returns is determined solely by fundamental factors, exactly as the volatilities drive the jump intensity in our two-factor model (21). The above evidence suggests, however, that this might be restrictive as the risk-neutral jump tails, in part, are impacted by different sources of variation than the volatility factors. Thus, motivated by the Z-score displays in Figure 3, we extend the two-factor model (21) by introducing an additional factor in the risk-neutral jump intensity that, importantly, is not a component of spot volatility. We further replace the jump distribution with one that allows for power law jump tail decay to better reflect the nonparametric option-based evidence in Bollerslev and Todorov (2011). The extended model takes the following form,

$$\begin{aligned}
\frac{dX_t}{X_{t-}} &= (r_t - \delta_t)\,dt + \sqrt{V_{1,t}}\,dW_{1,t} + \sqrt{V_{2,t}}\,dW_{2,t} + \int_{\mathbb{R}^2} (e^x - 1)\,\widetilde{\mu}(dt,dx,dy), \\
dV_{1,t} &= \kappa_1\,(\overline{v}_1 - V_{1,t})\,dt + \sigma_1\sqrt{V_{1,t}}\,dB_{1,t} + \mu_{v_1}\int_{\mathbb{R}^2} x^2\,1_{\{x<0\}}\,\mu(dt,dx,dy), \\
dV_{2,t} &= \kappa_2\,(\overline{v}_2 - V_{2,t})\,dt + \sigma_2\sqrt{V_{2,t}}\,dB_{2,t}, \\
dU_t &= -\,\kappa_3\,U_t\,dt + \mu_u\int_{\mathbb{R}^2}\left[(1-\rho_3)\,x^2\,1_{\{x<0\}} + \rho_3\,y^2\right]\mu(dt,dx,dy),
\end{aligned} \tag{22}$$

where $(W_{1,t}, W_{2,t}, B_{1,t}, B_{2,t})$ is defined as for the two-factor model (21). The jump measure μ has a compensator given by $dt \otimes \nu_t^{\mathbb{Q}}(dx,dy)$, where,

$$\nu_t^{\mathbb{Q}}(dx,dy) = \left\{\left(c^- 1_{\{x<0\}}\lambda_- e^{-\lambda_-|x|} + c^+ 1_{\{x>0\}}\lambda_+ e^{-\lambda_+ x}\right) 1_{\{y=0\}} + c^- 1_{\{x=0,\ y<0\}}\lambda_- e^{-\lambda_-|y|}\right\} dx\otimes dy,$$

$$c^- = c_0^- + c_1^- V_{1,t-} + c_2^- V_{2,t-} + c_3^- U_{t-}, \qquad c^+ = c_0^+ + c_1^+ V_{1,t-} + c_2^+ V_{2,t-} + c_3^+ U_{t-}\,.$$

Before commenting on the modifications and extensions relative to model (21), we remark on a few changes in notation motivated by the new specification of the jump processes. The first term in the expression for the jump measure controls price-volatility co-jumps, exactly as in model (21), while the second term governs the independent jumps in the new factor, U. Furthermore, the jumps in U are either directly proportional to the volatility jumps (in V_1) or independent from them. Both jump components of U have the same marginal distribution - they only differ in their interaction with the other jump components of the model. To retain notational symmetry between the jump components of U, we define the second term of the jump measure, μ, over the squared negative realizations. In particular, this implies that all jumps in U are positive.

The distinct features of the extended model may be summarized as follows. First, the distribution of the price jumps is now exponential, so we enforce an empirically realistic decay in the jump intensity as a function of the jump size. Moreover, we allow the left and right jumps to have different tail decay parameters, i.e., we break the dependence between the two tails. Second, the price and volatility co-jumps are now perfectly dependent, with the squared price jumps impacting the volatility dynamics. This is reminiscent of a GARCH specification for the volatility dynamics in discrete time. Given the proliferation of parameters, a parsimonious formulation is desirable. In addition, we allow only the negative price jumps to impact the volatility dynamics, as a more general dependence structure between volatility and price jumps cannot be well identified from the option panel. Third, the negative jumps now have an additional source of time variation, captured by the process U.[21] U is driven, in part, by the negative price jumps which allows for crisis periods to elevate the jump intensity of the system. Importantly, U also contains an independent source of variation, where the degree of dependence between jumps in U and X is captured by ρ_3. Thus, our specification encompasses scenarios of perfect dependence between the jump risks in V_1 and U (for $\rho_3 = 0$) and scenarios of V_1 and U being entirely independent state variables driving the dynamics of the option panel (for $\rho_3 = 1$). In addition, we note that V_1 and U are connected not only through this common jump component, but also via the joint time-variation in the jump intensity. This latter feature generates "cross self-excitation" in which jumps in U excite the occurrence of jumps in V_1 and vice versa. Finally, the extended model (22) allows for separate time-variation in the intensity of positive and negative jumps via discrepancies in the parameters $c_1^{\pm}$ and $c_2^{\pm}$.[22]

The parameter estimates for the three factor model (22) are reported in Table 6. Interestingly, compared with the two-factor model, the role of the factors in generating a persistent volatility process is reversed. The first volatility factor – containing the volatility jumps – is now the more

[21] This feature can also be incorporated into the dynamics for the intensity of the positive price jumps, but since the relevant parameter is not significant, we end up imposing the constraint $c_3^+ = 0$ in our final specification.

[22] From an implementation standpoint, it is critical that the model (22) remains within the general affine class of Duffie et al. (2003). We refer to that paper for the ODE system of equations that must be satisfied by the coefficients of the associated conditional characteristic function. During estimation, we impose the following conditions on the parameters to guarantee covariance stationarity of the three latent factors,

$$\kappa_1 > \frac{2c_1^- \mu_{v_1}}{\lambda_-^2}, \quad \text{and} \quad \kappa_3 > \frac{2c_3^- \kappa_1 \mu_u}{\kappa_1 \lambda_-^2 - 2c_1^- \mu_{v_1}}, \quad \text{and} \quad \kappa_2 < 0, \quad \text{and} \quad \sigma_i^2 \leq 2\kappa_i \bar{v}_i, \quad i = 1, 2.$$

transient component. Both leverage coefficients are significantly negative but none is statistically close to the boundary of -1. Turning to the jump parameters, we note that the jump intensity loadings $c_1^{\pm}$ and $c_2^{\pm}$ are statistically distinct, reflecting differences in the time-variation of the left and right jump tails. This feature is precluded in the original model (21) as it constrains the time variation of the two jump tails to be identical. Moreover, the loading c_3^- on the new jump intensity factor is highly significant, indicating a strong role for this factor in the model dynamics. Furthermore, the tail decay parameters $\lambda^{\pm}$ imply that negative jumps have significantly fatter jump tails than positive jumps, as also established by nonparametric means in Bollerslev and Todorov (2011). The implied negative and positive mean jumps sizes are -4.98% and 1.63%, and the average frequencies are 2.76 and 5.25 jumps per year, respectively. Finally, the new jump intensity factor is estimated to be quite persistent, while the coefficient ρ_3, reflecting the link between the jumps in U and V_1, is estimated imprecisely.

Table 6: Parameter Estimates for the Three-Factor Model

Parameter	Estimate	Std.	Parameter	Estimate	Std.	Parameter	Estimate	Std.
ρ_1	-0.9056	0.0342	σ_2	0.1412	0.0072	c_2^-	0.5113	2.7825
$\overline{v}_1$	0.0092	0.0004	μ_u	0.7636	0.4854	c_2^+	95.8730	16.6387
κ_1	9.6847	0.2337	κ_3	0.8390	0.1487	c_3^-	45.0050	4.8130
σ_1	0.4022	0.0171	ρ_3	0.2946	0.8398	λ_-	20.0751	0.2726
ρ_2	-0.8280	0.0295	c_0^+	2.3751	0.1222	λ_+	61.4484	0.4747
$\overline{v}_2$	0.0385	0.0041	c_1^-	31.9110	2.2798	μ_{v_1}	14.2240	0.4164
κ_2	0.2967	0.0339	c_1^+	92.4620	25.2328			

Note: Parameter estimates of the three-factor model (22) for S&P 500 equity-index options sampled every Wednesday over January 1996-July 2010.

We turn now to the analysis of model performance. Figure 4 displays the weekly model-implied Z-scores for the fit to the different regions of the volatility surface. We observe a substantial improvement compared with the two-factor model for the OTM put and ATM options. In contrast, the gains are moderate for the long maturity OTM calls and non-existent for the short maturity OTM calls. This is not surprising as we decided, for parsimony, not to include a separate jump intensity factor for the right tail in the three-factor model, focusing instead on improving performance in the critical left tail. Overall, the mispricing of the OTM puts and the ATM options in the three factor model is substantially less persistent compared to the two-factor model, indicating that the dynamic interactions among the different risk factors are better accommodated by our extended model. Beyond the general improvement in fit for the short maturity OTM puts and ATM options, we also observe that the periods 1998-1999 and 2009-2010 no longer are associated with overpricing of short-maturity ATM options. Similarly, the short maturity OTM puts are no longer underpriced during the quiet period 2004-2007. Finally, the fit to the long maturity OTM puts and ATM options has improved greatly. Additional formal diagnostics for model performance are collected in the supplementary appendix.[23]

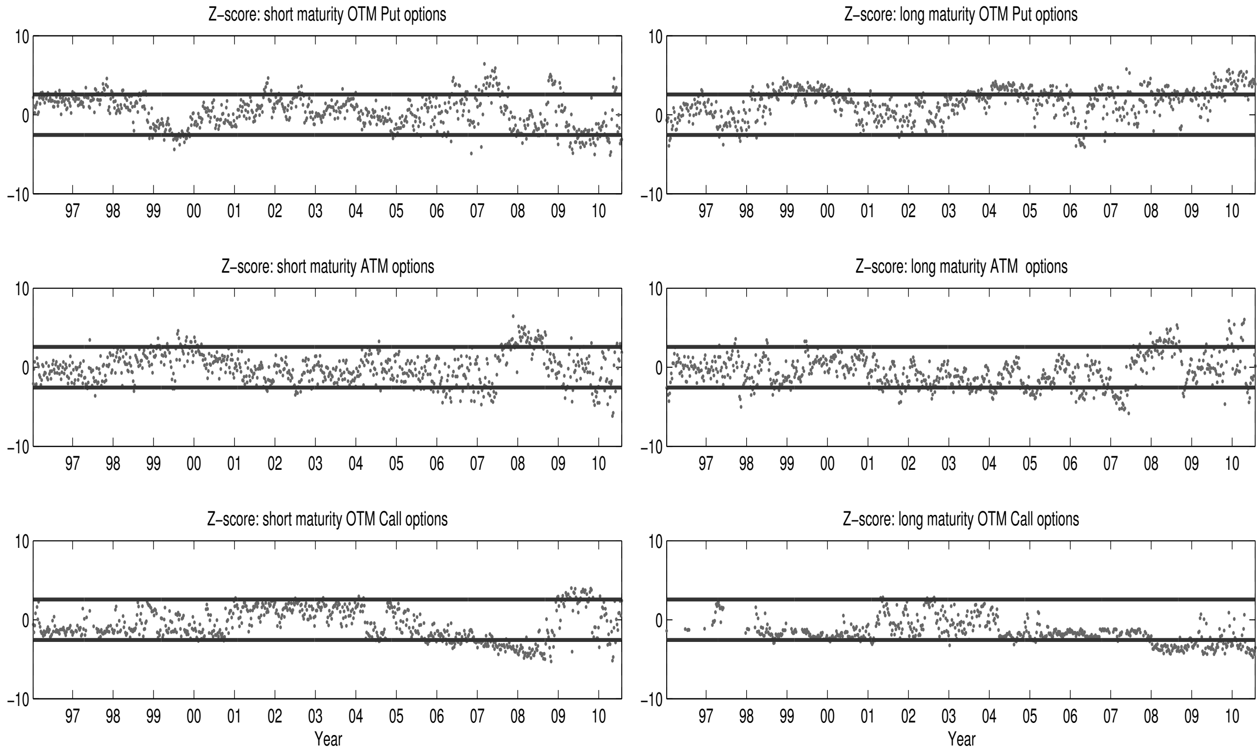

Figure 4: *Option Price Fit for the Three-factor Model.* Notation as for Figure 3.

To gain further insight into the success in capturing the evolution of the option surface over time, we plot, on Figure 5, the fit of the model to the observed model-implied "volatility smirk," i.e., the implied volatility as a function of moneyness, for two years within the sample along with a 95% confidence band. We pick two very different episodes: year 2006 is characterized by quiet conditions while 2009 represents a turbulent time for the market, following on the heels of the Fall 2008 crisis. In both cases the model tracks the short maturity volatility smirk well, showing that the model can

[23]These diagnostics include the tests for parameter stability and equality of the high-frequency and model-implied volatility estimators. In implementing the latter, we further rescale the high-frequency increments to account for the pronounced intraday volatility pattern.

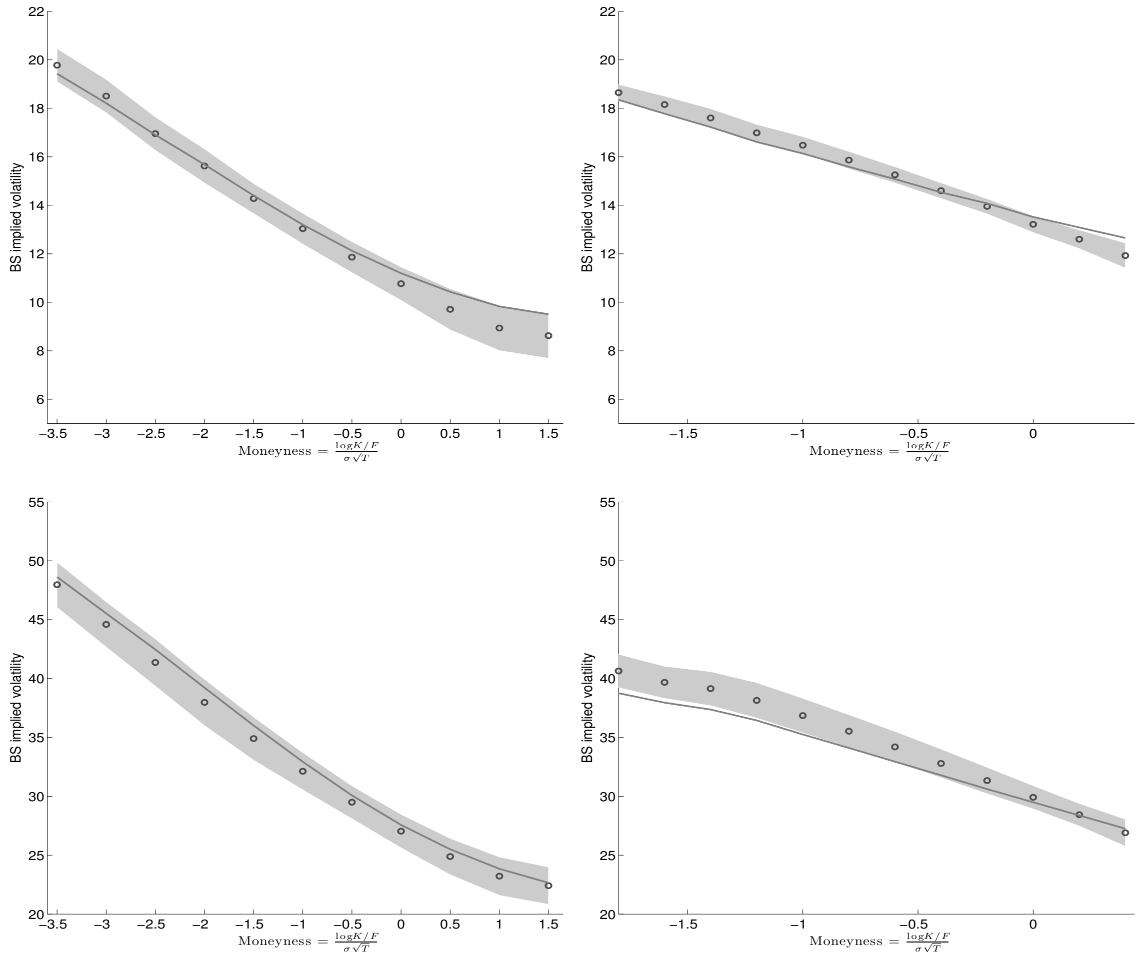

Figure 5: *Average Implied Volatility Curves for 2006 and 2009.* The dots represent averages of the observed Black-Scholes implied volatilities with moneyness within specific ranges. The shaded regions on the plots correspond to a 95% confidence region constructed using Corollary 1 and the three-factor model (22). The sold line represents the fitted values from the model. The top panels represent 2006 and the bottom panels 2009, while the left panels refer to the shorter and the right panels to the longer maturities.

adapt to the dramatic difference in market conditions. The two years are also distinct in terms of the pricing of the longer term options, with the flattening of the volatility smirk being relatively less pronounced during the tranquil conditions of 2006 than for the turbulent times in 2009. In both cases, a slight undervaluation of the long maturity OTM puts is evident but the overall fit seems quite satisfactory. We also note the large difference in the width of the standard error bands across the two years and across the tenors. Considering the discrepancy in scale for the two years, it is evident that bands are much wider for 2009 than 2006. This primarily reflects the heterogeneity of the measurement errors for the option prices across tranquil and volatile market conditions and the associated precision of inference for the state vector over the course of the year. Similarly, the bands are narrower for ATM than deep OTM options, reflecting the relative bid-ask spreads and liquidity across moneyness.[24] Finally, for the longer maturities, the implied volatility curve is often less sensitive to the current value of the state vector and more dependent on the parameter vector, θ. Since the estimation of θ exploits information across the full sample, we infer its value with much greater precision than the current state vector which can lead to narrower confidence bands for the more distant maturities, in spite of the bid-ask spreads typically being larger. In the supplementary appendix, we provide similar plots for all the years in our sample. They illustrate the advantages of our new fixed span asymptotics to dynamically evaluate the success of the model in matching the observed option prices.

Figures 4 and 5 capture complementary aspects of the fit to the option panel. First, note that Figure 4 indicates no systematic mispricing during either 2006 or 2009 for the short- or long-maturity ATM options. This is reflected in Figure 5 by the model-implied prices falling within the gray standard error bands. In contrast, the strongest indications of model failure in Figure 5 is the model-implied overpricing for OTM call options at the long and, less strikingly, short maturity in 2006 as well as underpricing of long-maturity OTM put options in 2009. In Figure 4, the Z-scores are on the lower boundary of the confidence band throughout 2006 for both maturities. Clearly, this persistent bias in the pricing errors – with model prices exceeding market prices – signals deficiencies in the fit, even if most of the individual Z-scores fall within the confidence band. Likewise, from the middle to the end of 2009, we observe large Z-scores for the long maturity OTM puts. Hence, the persistent biases in the model prices manifest themselves in significant pricing errors for the corresponding region of the option surface in Figure 5.

The empirical evidence, summarized in Figure 4 as well as the supplementary appendix, suggests

[24]Time-averaging across the year renders the width slightly non-monotonic, as the number of options within each moneyness category changes over time, leading to some randomness in estimation precision, which is also reflected in the relative width.

that our new three-factor model provides a much more coherent account of the observed option surface dynamics than the standard two-factor model. The key to the improved performance is the introduction of the new intensity factor U that allows for separation of the risk-neutral jump intensity dynamics from the (components of) stochastic volatility. As a further test of this hypothesis, we estimated an alternative three-factor model that only deviates from model (22) by having the third factor enter as a component of stochastic volatility – fully in line with the more traditional approach to the modeling of multi-factor volatility models. We report the results from the estimation of this alternative model in the Supplementary Appendix. The findings confirm that this more "traditional" three-factor specification fares substantially worse than model (22), especially in terms of fitting the critical short-maturity OTM put options.

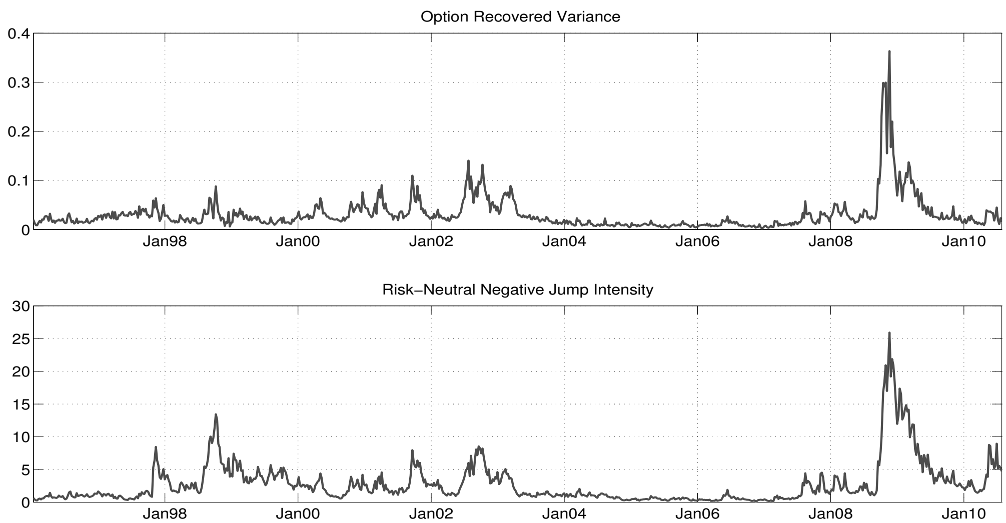

Figure 6: *Volatility and Negative Jump Intensity from the three factor model (22).*

Figure 6 depicts the recovered stochastic volatility as well as the risk-neutral intensity of the negative jumps from our three-factor model (22). It is evident that the volatility process shares common low frequency features with the intensity of the left risk-neutral jump tail. At the same time, the figure reveals pronounced differences as well. In particular, over the period following the 1997 Asian and the 1998 Russian crises, the intensity of the left jump tail remains elevated for an extended period and mean-reverts much more slowly than the volatility. A similar pattern emerges following the Fall 2008 crisis as well as the European crisis in August 2010. On the other hand, following the collapse of the 2002 tech bubble and during the quiet period of 2004-2007, the market volatility and the intensity of the left risk-neutral jump tail are quite closely aligned. Thus, overall, there is a striking heterogeneity in the dynamic relation between the two processes across time.

How do we synthesize and interpret the empirical evidence? As mentioned, a tight link between volatility (components) and the risk-neutral jump intensity also arises naturally in equilibrium models with a representative agent, equipped with Epstein-Zin preferences, who faces jump risk either directly in consumption growth or in its conditional mean and variance. Our findings suggest that, following periods of crises like in 1998 or the Fall of 2008, the compensation demanded for bearing negative jump tail risk rises by a disproportional amount, rendering it incompatible with the relatively fast reversion of stock volatility towards its pre-crises level. Intuitively, OTM put options are too expensive for too long following such episodes to be rationalized by the concurrent level of risk as captured by market volatility. Thus, our evidence suggests that the dynamic pricing of jump tail risk is complex and falls outside the standard mean-variance paradigm, for which prices of risk are proportional to volatility and its factors. Consequently, the latter is unable to rationalize the joint dynamics of the stock market and the options written on it.

7 Conclusion

In this paper we consider the problem of estimating the parameters of the risk-neutral distribution and the latent state variables from a panel of options, observed with error, with fixed time span and increasing cross-sectional dimension. We prove consistency of the estimators and show that they converge stably to mixed Gaussian laws. We further propose and implement feasible inference based on the developed limit theory. We design novel tests for the observed option trajectories by evaluating the option price fit and the parameter stability over time as well as the pathwise distance between the volatility implied by the option-based estimation and a nonparametric estimate constructed from high-frequency data for the underlying asset.

An extensive Monte Carlo study confirms that the inference techniques work well over relatively short time spans for realistically calibrated parameter settings. In an empirical application to S&P 500 equity-index options, we extend existing one- and two-factor stochastic volatility models by allowing the left jump tail intensity to depend on an additional factor that is not a component of the stochastic volatility process. This extension of the traditional asset pricing models is crucial for explaining the observed dynamic dependencies between the short maturity OTM puts and ATM options. Our estimation results imply that priced jump tail risks have a much more persistent reaction to large negative shocks in the economy than the market volatility.

8 Appendix

We first establish some preliminary results and then provide proofs for the theorems and corollaries.

8.1 Preliminary results

Lemma 1 *Under the conditions of Theorem 2, we have*

$$\begin{pmatrix} \frac{1}{\sqrt{N_1}}\sum_{j=1}^{N_1} \nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\mathbf{S}_1,\theta_0)\epsilon_{1,k_j,\tau_j} \\ \vdots \\ \frac{1}{\sqrt{N_T}}\sum_{j=1}^{N_T} \nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\mathbf{S}_T,\theta_0)\epsilon_{T,k_j,\tau_j} \\ \frac{1}{\sqrt{N_1}}\sum_{j=1}^{N_1} \nabla_{\theta}\kappa(k_j,\tau_j,\mathbf{S}_1,\theta_0)\epsilon_{1,k_j,\tau_j} \\ \vdots \\ \frac{1}{\sqrt{N_T}}\sum_{j=1}^{N_T} \nabla_{\theta}\kappa(k_j,\tau_j,\mathbf{S}_T,\theta_0)\epsilon_{T,k_j,\tau_j} \end{pmatrix} \xrightarrow{\mathcal{L}-s} \left(\widetilde{\mathbf{\Omega}}_T\right)^{1/2} \begin{pmatrix} \mathbf{E}_1 \\ \vdots \\ \mathbf{E}_T \\ \mathbf{E}'_1 \\ \vdots \\ \mathbf{E}'_T \end{pmatrix}, \tag{23}$$

where $\{\mathbf{E}_t\}_{t\geq 1}$ are defined in Theorem 2, $\{\mathbf{E}'_t\}_{t\geq 1}$ are vectors of standard normal variables, each of size $q\times 1$, independent of each other and of the filtration $\mathbf{F}$ as well as the vector $\{\mathbf{E}_t\}_{t\geq 1}$ and

$$\widetilde{\mathbf{\Omega}} = \begin{pmatrix} \widetilde{\mathbf{\Omega}}_T^{1,1} & \dots & \mathbf{0}_{p\times p} & \widetilde{\mathbf{\Omega}}_T^{1,T+1} & \dots & \mathbf{0}_{p\times q} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{p\times p} & \dots & \widetilde{\mathbf{\Omega}}_T^{T,T} & \mathbf{0}_{p\times q} & \dots & \widetilde{\mathbf{\Omega}}_T^{T,2T} \\ \widetilde{\mathbf{\Omega}}_T^{T+1,1} & \dots & \mathbf{0}_{q\times p} & \widetilde{\mathbf{\Omega}}_T^{T+1,T+1} & \dots & \mathbf{0}_{q\times q} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{q\times p} & \dots & \widetilde{\mathbf{\Omega}}_T^{2T,2T} & \mathbf{0}_{q\times q} & \dots & \widetilde{\mathbf{\Omega}}_T^{2T,T} \end{pmatrix}.$$

The block components of $\widetilde{\mathbf{\Omega}}_T$ are defined as follows:

$$\widetilde{\mathbf{\Omega}}_T^{t,t} = \mathbf{\Omega}_T^{t,t}, \quad \widetilde{\mathbf{\Omega}}_T^{t,T+t} = (\widetilde{\mathbf{\Omega}}_T^{T+t,t})' = \sum_{\tau} \pi_t^{\tau} \int_{\underline{k}(t,\tau)}^{\overline{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)} \phi_{t,k,\tau} \nabla_{\mathbf{S}}\kappa(k,\tau,\mathbf{S}_t,\theta_0)\nabla_{\theta}\kappa(k,\tau,\mathbf{S}_t,\theta_0)'dk,$$

$$\widetilde{\mathbf{\Omega}}_T^{T+t,T+t} = \sum_{\tau} \pi_t^{\tau} \int_{\underline{k}(t,\tau)}^{\overline{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)} \phi_{t,k,\tau} \nabla_{\theta}\kappa(k,\tau,\mathbf{S}_t,\theta_0)\nabla_{\theta}\kappa(k,\tau,\mathbf{S}_t,\theta_0)'dk, \qquad t=1,...,T.$$

Proof of Lemma 1. We denote

$$\chi_j^{(t)} = \begin{pmatrix} \nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0)\epsilon_{t,k_j,\tau_j} \\ \nabla_{\theta}\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0)\epsilon_{t,k_j,\tau_j} \end{pmatrix}, \quad t=1,...,T, \tag{24}$$

which is a $(p+q)\times 1$ vector. We further denote the filtration $\widetilde{\mathcal{F}}_j = \sigma\left(\{\epsilon_{t,k_i,\tau_i}\}_{t=1,...,T,i=1,...,j}\right) \cup \mathcal{F}_T^{(0)}$ for $j=0,1,...$ (recall from A1 that the sequence of observation grids on the moneyness dimension is nested). With this notation we will show

$$\Upsilon_t^{N_t} = \frac{1}{\sqrt{N_t}} \sum_{j=1}^{N_t} \chi_j^{(t)} \xrightarrow{\mathcal{L}-s} \Upsilon_t, \quad t=1,...,T, \tag{25}$$

where Υ_t is a $\mathcal{F}_T^{(0)}$-conditionally centered Gaussian process with $\mathcal{F}_T^{(0)}$-conditional variance of

$$\begin{pmatrix} \widetilde{\mathbf{\Omega}}_T^{t,t} & \widetilde{\mathbf{\Omega}}_T^{t,T+t} \\ \widetilde{\mathbf{\Omega}}_T^{T+t,t} & \widetilde{\mathbf{\Omega}}_T^{T+t,T+t} \end{pmatrix}.$$

We have that $\frac{1}{\sqrt{N_t}}\sum_{j=1}^{N_t}\chi_j^{(t)}$ is adapted to $\widetilde{\mathcal{F}}_{N_t}$, for $N_t \in \mathbb{N}$. Moreover, we have that the nesting property of the filtration, $\widetilde{\mathcal{F}}_N \subset \widetilde{\mathcal{F}}_{N+1}$ for $N \in \mathbb{N}$, and furthermore $\mathcal{F}_T = \bigvee_j \widetilde{\mathcal{F}}_j$. Therefore, we can apply Theorem VIII.5.42 of Jacod and Shiryaev (2003). To establish (25) above it now suffices to prove,

$$\mathbb{E}\left(\chi_j^{(t)}|\widetilde{\mathcal{F}}_{j-1}\right) = 0,\ \ \frac{1}{N_t}\sum_{j=1}^{N_t}\mathbb{E}\left(\chi_j^{(t)}\chi_j^{(t)'}|\widetilde{\mathcal{F}}_{j-1}\right) \stackrel{\mathbb{P}}{\longrightarrow} \begin{pmatrix} \widetilde{\mathbf{\Omega}}_T^{t,t} & \widetilde{\mathbf{\Omega}}_T^{t,T+t} \\ \widetilde{\mathbf{\Omega}}_T^{T+t,t} & \widetilde{\mathbf{\Omega}}_T^{T+t,T+t} \end{pmatrix},\ \ \frac{1}{N_t^2}\sum_{j=1}^{N_t}\mathbb{E}||\chi_j^{(t)}|\widetilde{\mathcal{F}}_{j-1}||^4 \stackrel{\mathbb{P}}{\longrightarrow} 0. \tag{26}$$

The first and the third result of (26) follow immediately upon making use of A4(i) and A4(iv). To prove the second claim we apply assumption A4(ii) and A4(iii) as well as assumption A1, concerning the mesh of the grid in the log-moneyness dimension of the options, the smoothness of the $\phi_{t,k,\tau}$ function in A4(iii), and the smoothness of $\kappa(k,\tau,\mathbf{Z},\theta)$ in its first argument. In fact, we even have the convergence holding almost surely.

Now we will prove that the convergence in (25) holds jointly for $t = 1,...,T$ with the limits being $\mathcal{F}_T^{(0)}$-conditionally independent. For this it suffices show

$$\mathbb{E}\left(Y\prod_{t=1}^{T}f_t\left(\Upsilon_t^{N_t}\right)\right) \longrightarrow \mathbb{E}\left(Y\prod_{t=1}^{T}\mathbb{E}\left(f_t\left(\Upsilon_t\right)|\mathcal{F}_T\right)\right), \tag{27}$$

for $f_t(\cdot)$ being Lipschitz functions on $\mathbb{R}^{p+q}$, Y denoting a bounded random variable on $\mathcal{F}_T$, and Υ_t indicating the limits in (25) (this follows from Corollary 1.4.5 of van der Vaart and Wellner (1996)).

We look first at the case when Y is adapted to $\mathcal{F}_T^{(0)}$. In this case, using A4(iii), we have

$$\mathbb{E}\left(Y\prod_{t=1}^{T}f_t\left(\Upsilon_t^{N_t}\right)\right) = \mathbb{E}\left(Y\prod_{t=1}^{T}\mathbb{E}\left(f_t\left(\Upsilon_t^{N_t}\right)|\mathcal{F}_T^{(0)}\right)\right).$$

Next, using Theorem VIII.5.25 of Jacod and Shiryaev (2003), we have

$$\widehat{Q}^N \stackrel{\mathbb{P}}{\longrightarrow} Q, \tag{28}$$

where $\widehat{Q}^N(\omega^{(0)},\cdot)$ is the conditional distribution of the process $\Upsilon_t^{N_t}$, which is a transitional probability kernel from $(\Omega^{(0)},\mathcal{F}^{(0)})$ into $(\mathbb{R}^{p+q},\mathcal{B}(\mathbb{R}^{p+q}))$ and Q is the conditional probability associated with the limiting process Υ_t (conditional on the event $\omega^{(0)} \in \Omega^{(0)}$). This convergence is in the space of probability measures equipped with the weak topology, therefore we have

$$\mathbb{E}\left(f_t(\Upsilon_t^{N_t})\big|\mathcal{F}_T^{(0)}\right) \stackrel{\mathbb{P}}{\longrightarrow} \mathbb{E}\left(f_t(\Upsilon_t)\big|\mathcal{F}_T^{(0)}\right). \tag{29}$$

From here, since the functions $f_t(\cdot)$ and the variable Y are bounded, we have

$$\mathbb{E}\left(Y\prod_{t=1}^{T}\mathbb{E}\left(f_t\left(\Upsilon_t^{N_t}\right)|\mathcal{F}_T^{(0)}\right)\right) \longrightarrow \mathbb{E}\left(Y\prod_{t=1}^{T}\mathbb{E}\left(f_t(\Upsilon_t)|\mathcal{F}_T^{(0)}\right)\right), \tag{30}$$

and therefore (27) holds when Y is $\mathcal{F}_T^{(0)}$-adapted.

We are left with the case when Y is adapted to $\mathcal{F}_T^{(1)}$. Due to separability of the σ-field $\mathcal{F}_T^{(1)}$, we can proceed exactly as in step 4 of the proof of Theorem IX.7.28 in Jacod and Shiryaev (2003) and look only at the case when $Y = h(\{\epsilon_m\}_{m\in\overline{M}})$ where $\overline{M}$ is a *finite* set of triplets (t, k, τ). Now, we let $\widetilde{\Upsilon}_t^{N_t}$ denote a variable constructed from $\Upsilon_t^{N_t}$ by excluding the options corresponding to the triplets $(t, k, \tau) \in \overline{M}$. Since this is a finite number, the differences $\widetilde{\Upsilon}_t^{N_t} - \Upsilon_t^{N_t}$ are obviously negligible. So, we only need to verify (27) for the case where $\widetilde{\Upsilon}_t^{N_t}$ is replaced by $\Upsilon_t^{N_t}$. However, due to assumption A4(iii), we have that $\widetilde{\Upsilon}_t^{N_t}$ and Y are independent conditional on $\mathcal{F}_T^{(0)}$. From here, we can proceed exactly as in the proof for the case when Y is adapted to $\mathcal{F}_T^{(0)}$. □

Lemma 2 *Under the conditions of Theorem 5, we have*

$$\sqrt{k_n}\begin{pmatrix} V_1\left\{\frac{n}{k_n}\sum_{i\in I^{+,n}}\left(\Delta_i^{1,n}W\right)^2-1\right\} \\ \vdots \\ V_T\left\{\frac{n}{k_n}\sum_{i\in I^{+,n}}\left(\Delta_i^{T,n}W\right)^2-1\right\}\end{pmatrix} \xrightarrow{\mathcal{L}-s} \begin{pmatrix}\sqrt{2}V_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sqrt{2}V_T\end{pmatrix}\begin{pmatrix}\widetilde{E}_1 \\ \vdots \\ \widetilde{E}_T\end{pmatrix}, \tag{31}$$

where $\{\widetilde{E}_t\}_{t\geq 1}$ are defined in Theorem 5.

Proof of Lemma 2. Since $k_n/n \to 0$, it is no limitation to assume $k_n/n < 1$, and we do so henceforth. In particular, this implies that the sets $\{t + \frac{i}{n}\}_{i\in I^{\pm,n}}$ for $t = 1, ..., T$ are disjoint. The validity of the Lemma now follows if we can show,

$$\sqrt{k_n}\begin{pmatrix}\left\{\frac{n}{k_n}\sum_{i\in I^{+,n}}\left(\Delta_i^{1,n}W\right)^2-1\right\} & \dots & \left\{\frac{n}{k_n}\sum_{i\in I^{+,n}}\left(\Delta_i^{T,n}W\right)^2-1\right\}\end{pmatrix} \xrightarrow{\mathcal{L}-s} \sqrt{2}\begin{pmatrix}\widetilde{E}_1 & \dots & \widetilde{E}_T\end{pmatrix}.$$

The convergence in law follows from a standard central limit theorem, as $\left\{\frac{n}{k_n}\sum_{i\in I^{+,n}}\left(\Delta_i^{t,n}W\right)^2-1\right\}$ are independent of each other for different values of t, and further each of them equals $\left\{\frac{1}{k_n}\sum_{i=1}^{k_n}(Z_i)^2-1\right\}$ in probability, where Z_i are i.i.d. standard normal variables. Thus we only need show that the convergence holds stably, and for this it suffices to consider at bounded variables adopted to the filtration generated by the Brownian motion W_t. We are now in position to directly apply Steps 3 and 4 of the proof of Proposition 8.2 of Jacod and Todorov (2010) to establish the result. □

Lemma 3 *If the conditions of Theorem 2 and Theorem 5 hold, then the convergence in Lemma 1 and Lemma 2 holds jointly and further the vectors $(\mathbf{E}_1, ..., \mathbf{E}_T, \mathbf{E}'_1, ..., \mathbf{E}'_T)'$ and $(\widetilde{E}_1, ..., \widetilde{E}_T)'$ in Lemma 1 and Lemma 2 are independent.*

Proof of Lemma 3. Exploiting the same notation as in the proof of Lemmas 1 and 2, we further denote,

$$X_1^{\underline{N}} = \begin{pmatrix}\Upsilon_1^{N_1} & \dots & \Upsilon_T^{N_T}\end{pmatrix}',$$
$$X_2^n = \sqrt{k_n}\begin{pmatrix}V_1\left\{\frac{n}{k_n}\sum_{i\in I^{+,n}}\left(\Delta_i^{1,n}W\right)^2-1\right\} & \dots & V_T\left\{\frac{n}{k_n}\sum_{i\in I^{+,n}}\left(\Delta_i^{T,n}W\right)^2-1\right\}\end{pmatrix}',$$

where recall $\underline{N} = \min_{t=1,...,T} N_t$. In this notation, we must prove

$$\mathbb{E}\left(Yf(X_1^{\underline{N}})g(X_2^n)\right) \longrightarrow \mathbb{E}\left(Y\mathbb{E}\left(f(X_1)|\mathcal{F}_T\right)\mathbb{E}\left(g(X_2)|\mathcal{F}_T\right)\right), \tag{32}$$

for $f(\cdot)$ and $g(\cdot)$ being Lipschitz functions on $\mathbb{R}^{T(p+q)}$ and $\mathbb{R}^T$, respectively, Y denoting a bounded random variable on $\mathcal{F}_T$, and X_1 and X_2 representing the limits in (23) and (31), respectively.

First we consider at the case where Y is adapted to $\mathcal{F}_T^{(0)}$. Exactly as in the proof of Lemma 1, we can show

$$\mathbb{E}\left(f(X_1^{\underline{N}})\middle|\mathcal{F}_T^{(0)}\right) \xrightarrow{\mathbb{P}} \mathbb{E}\left(f(X_1)\middle|\mathcal{F}_T^{(0)}\right). \tag{33}$$

From here, for every sufficiently small $\epsilon > 0$, there exists $\widetilde{N} > 0$ such that for $\underline{N} > \widetilde{N}$, we have

$$\mathbb{E}\left\{Yg(X_2^n)\left|\mathbb{E}\left(f(X_1^{\underline{N}})\middle|\mathcal{F}_T^{(0)}\right) - \mathbb{E}\left(f(X_1)\middle|\mathcal{F}_T^{(0)}\right)\right|\right\} \le K\epsilon, \tag{34}$$

for some positive constant K (that does not depend on ϵ and $\widetilde{N}$), where we also exploited the boundedness of Y, $f(X_1^{\underline{N}})$ and $g(X_2^n)$. Next, using the fact that $\mathbb{E}\left(f(X_1)\middle|\mathcal{F}_T^{(0)}\right)Y$ is $\mathcal{F}_T^{(0)}$-adapted, the definition of stable convergence, and the result of Lemma 2, we obtain the limit result in (32) for the case where Y is $\mathcal{F}_T^{(0)}$-adapted.

We are left with the case where Y is adapted to $\mathcal{F}_T^{(1)}$. The proof is identical to the analogous case for the proof of Lemma 1. Hence, the proof is omitted. $\square$

Lemma 4 *Under the conditions of Theorem 5 we have*

$$\frac{n}{\sqrt{k_n}}\sum_{t=1}^{T}\sum_{i\in I^{+,n}}\left|(\Delta_i^{t,n}X)^2 1\left(|\Delta_i^{t,n}X| \le \alpha n^{-\varpi}\right) - V_t(\Delta_i^{t,n}W)^2\right| \xrightarrow{\mathbb{P}} 0. \tag{35}$$

Proof of Lemma 4. First, via a localization argument similar to that in, e.g., Lemma 4.6 of Jacod (2008), it suffices to consider the case where the processes α_t, V_t and a_t, as well as the jumps of the process X, are bounded. Thus, we impose this condition for the remainder of this proof.

Applying Itô lemma, we have

$$d\log(X_t) = \left(\alpha_t - \frac{1}{2}V_t\right)dt + \sqrt{V_t}dW_t + \int_{x>-1} x\widetilde{\mu}(dt,dx) + \int_{x>-1}\left(\log(1+x) - x\right)\mu(dt,dx), \tag{36}$$

note that $\log(1+x) - x \sim x^2$ for $x \to 0$ and therefore the last integral above is well defined in the usual Riemann-Stieltjes sense. We further denote

$$A_t = \int_0^t \left(\alpha_s - \frac{1}{2}V_s\right)ds,\ \ Z_t = \int_0^t \sqrt{V_s}dW_s,\ \ Y_t = \int_0^t\int_{x>-1} x\widetilde{\mu}(ds,dx) + \int_0^t\int_{x>-1}\left(\log(1+x)-x\right)\mu(ds,dx).$$

In this notation, we obtain the following decomposition for any $i \in I^{+,n}$ and $t = 1,...,T$,

$$\begin{aligned}
&(\Delta_i^{t,n}X)^2 1_{\{|\Delta_i^{t,n}X|\le\alpha n^{-\varpi}\}} - V_t\left(\Delta_i^{t,n}W\right)^2 \\
&\quad= (\Delta_i^{t,n}X)^2 1_{\{|\Delta_i^{t,n}X|\le\alpha n^{-\varpi}\}} - (\Delta_i^{t,n}X)^2 1_{\{|\Delta_i^{t,n}Y|\le\frac{\alpha}{2}n^{-\varpi}\}} + |\Delta_i^{t,n}X - \Delta_i^{t,n}Z|^2 1_{\{|\Delta_i^{t,n}Y|\le\frac{\alpha}{2}n^{-\varpi}\}} \\
&\qquad+ 2(\Delta_i^{t,n}X - \Delta_i^{t,n}Z)\Delta_i^{t,n}Z 1_{\{|\Delta_i^{t,n}Y|\le\frac{\alpha}{2}n^{-\varpi}\}} - (\Delta_i^{t,n}Z)^2 1_{\{|\Delta_i^{t,n}Y|>\frac{\alpha}{2}n^{-\varpi}\}} \\
&\qquad+ (\Delta_i^{t,n}Z)^2 - V_{t+\frac{i-1}{n}}\left(\Delta_i^{t,n}W\right)^2 + (V_{t+\frac{i-1}{n}} - V_t)\left(\Delta_i^{t,n}W\right)^2.
\end{aligned} \tag{37}$$

We now derive bounds for moments of each of the terms of the decomposition. These bounds will, in combination, prove convergence of their (scaled) sums either in the L^1 or L^2 norms. Henceforth, in the proof of the lemma, K denotes a positive constant which is independent of n and typically will take on different values across the different equations. First, we have,

$$|(\Delta_i^{t,n}X)^2 1_{\{|\Delta_i^{t,n}X|\le \alpha n^{-\varpi}\}} - (\Delta_i^{t,n}X)^2 1_{\{|\Delta_i^{t,n}Y|\le \frac{\alpha}{2}n^{-\varpi}\}}| \le (\Delta_i^{t,n}X)^2 1_{\{|\Delta_i^{t,n}X-\Delta_i^{t,n}Y|\ge \frac{\alpha}{2}n^{-\varpi}\}}, \quad (38)$$

and then using Hölder inequality, the boundedness of α_t and V_t, and the Burkholder-Davis-Gundy inequality, we get

$$\mathbb{E}\left|(\Delta_i^{t,n}X)^2 1_{\{|\Delta_i^{t,n}X|\le \alpha n^{-\varpi}\}} - (\Delta_i^{t,n}X)^2 1_{\{|\Delta_i^{t,n}Y|\le \frac{\alpha}{2}n^{-\varpi}\}}\right| \le Kn^{-\zeta}, \quad \forall \zeta > 0. \quad (39)$$

Next, by applying the Burkholder-Davis-Gundy inequality and/or the algebraic inequality $|\sum_i |a_i||^p \le \sum_i |a_i|^p$ for any $p \in (0,1]$, we obtain

$$\mathbb{E}|\Delta_i^{t,n}Y|^\zeta \le Kn^{-1}, \quad \forall \zeta \ge \beta. \quad (40)$$

Exploiting the above inequality, we deduce

$$\mathbb{E}\left\{|\Delta_i^{t,n}X - \Delta_i^{t,n}Z|^2 1_{\{|\Delta_i^{t,n}Y|\le \frac{\alpha}{2}n^{-\varpi}\}}\right\} \le Kn^{-1-(2-\beta)\varpi}. \quad (41)$$

Next, we decompose Y_t depending on the value of β in Assumption A0. $Y_t = \int_0^t \int_{x>-1} \log(1+x)\widetilde{\mu}(ds,dx) + \int_0^t a_s ds \int_{x>-1}(\log(1+x)-x)\nu^{\mathbb{P}}(dx)$ when β in assumption A0 cannot be chosen less than 1 and $Y_t = \int_0^t \int_{x>-1} \log(1+x)\mu(ds,dx) - \int_0^t a_s ds \int_{x>-1} x\nu^{\mathbb{P}}(dx)$ otherwise. Then, upon making use of the elementary inequality $1(|a+b| < c) \le 1(|a| > c) + 1(|b| < 2c)$ for any $a,\ b \in \mathbb{R}$ and $c > 0$, we get

$$\mathbb{E}\left|(\Delta_i^{t,n}X - \Delta_i^{t,n}Z)\Delta_i^{t,n}Z 1_{\{|\Delta_i^{t,n}Y|\le \frac{\alpha}{2}n^{-\varpi}\}}\right| \le Kn^{-1/2-\frac{1}{\beta}\wedge 1-(1-\beta)\varpi\vee 0}. \quad (42)$$

Moreover, applying the Hölder inequality and the inequality (40) yields,

$$\mathbb{E}\left\{(\Delta_i^{t,n}Z)^2 1_{\{|\Delta_i^{t,n}Y|> \frac{\alpha}{2}n^{-\varpi}\}}\right\} \le Kn^{-1-(1-\beta\varpi)}. \quad (43)$$

Finally, using the boundedness of V_t from both below and above, along with Aassumption A0(i), and the Itô isometry, we obtain,

$$\mathbb{E}\left|(\Delta_i^{t,n}Z)^2 - V_{t+\frac{i-1}{n}}\left(\Delta_i^{t,n}W\right)^2\right| \le Kn^{-1/2}\sqrt{\int_{t+\frac{i-1}{n}}^{t+\frac{i}{n}} \mathbb{E}(V_s - V_{t+\frac{i-1}{n}})^2 ds} \le Kn^{-3/2}. \quad (44)$$

$$\mathbb{E}\left\{\left(V_{t+\frac{i-1}{n}} - V_t\right)(\Delta_i^{t,n}W)^2\right\} \le Kn^{-3/2}\sqrt{i-1}. \quad (45)$$

Combining these findings, we validate the convergence result in (35). □

8.2 Proof that Assumption A3′ implies A3

The proof consists of verifying the conditions of Theorem 21 of Ibragimov and Has'minskii (1981) for uniform convergence on the space of functions vanishing at infinity equipped with the uniform topology. We use the shorthand notation

$$\eta_\theta^{N_t}(\mathbf{Z}_t) = \frac{\sum_{j=1}^{N_t} \zeta_t(k_j,\tau_j)\left[\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0) - \kappa(k_j,\tau_j,\mathbf{Z}_t,\theta)\right]\epsilon_{t,k_j,\tau_j}}{\sum_{j=1}^{N_t}\left[\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0) - \kappa(k_j,\tau_j,\mathbf{Z}_t,\theta)\right]^2},$$

and we further fix some $\epsilon > 0$. Using the continuous differentiability of $\kappa(k, \tau, \mathbf{S}_t, \theta)$ in its last two arguments, as well as the integrability of $\epsilon_{t,k,\tau}$ (conditional on $\mathcal{F}_T^{(0)}$) together with the Burkholder-Davis-Gundy inequality for discrete martingales (applied successively), as well as the conditional integrability conditions for the error $\epsilon_{t,k,\tau}$ in A3$'$(iii), we have for N_t sufficiently large

$$\mathbb{E}\left(|\eta_\theta^{N_t}(\mathbf{Z}_t) - \eta_\theta^{N_t}(\mathbf{Z}_t + h)|^{p+\iota}\middle|\mathcal{F}_T^{(0)}\right) \leq K||h||^{p+\iota}, \quad \mathbb{E}\left(|\eta_\theta^{N_t}(\mathbf{Z}_t)|^{p+\iota}\middle|\mathcal{F}_T^{(0)}\right) \leq K, \tag{46}$$

for either $||\theta - \theta_0|| > \epsilon$ or $||\mathbf{Z}_t - \mathbf{S}_t|| > \epsilon$ and where K is $\mathcal{F}_T^{(0)}$-adapted finite-valued random variable (recall that p is the dimension of the state vector $\mathbf{S}_t$). Next, for every $\mathbf{Z}_t$, we have

$$\mathbb{E}\left(\sup_{\theta\in\Theta:\ ||\theta-\theta_0||>\epsilon} |\eta_\theta^{N_t}(\mathbf{Z}_t)| \bigwedge 1 \middle|\mathcal{F}_T^{(0)}\right) \longrightarrow 0, \quad \text{a.s.}, \tag{47}$$

using a criterion for uniform convergence on compact sets.

By Cauchy-Schwartz inequality and the fact that $\zeta_t(k, \tau)$ is finite-valued, we have

$$|\eta_\theta^{N_t}(\mathbf{Z}_t)| \leq K\sqrt{\frac{\frac{1}{N_t}\sum_{j=1}^{N_t} \epsilon_{t,k_j,\tau_j}^2}{\frac{1}{N_t}\sum_{j=1}^{N_t} \left[\kappa(k_j, \tau_j, \mathbf{S}_t, \theta_0) - \kappa(k_j, \tau_j, \mathbf{Z}_t, \theta)\right]^2}},$$

for K some $\mathcal{F}_T^{(0)}$-adapted random variable. From here and our assumption for the behavior of $\kappa(k, \tau, \mathbf{S}_t, \theta)$ as $||\mathbf{S}_t|| \to \infty$, together with Chebyshev's inequality, provided N_t is sufficiently big so that we have at least one observation in the range for which assumption A3$'$(ii) holds, we have

$$\lim_{y\to\infty} \sup_{N_t,\theta} \mathbb{P}\left(\sup_{||\mathbf{Z}_t||>y} |\eta_\theta^{N_t}(\mathbf{Z}_t)| > |\log(y)|^{1/2-\iota}\middle|\mathcal{F}_T^{(0)}\right) = 0. \tag{48}$$

Combining (46)-(48), we have by an application of Theorem 21 of Ibragimov and Has'minskii (1981)

$$\mathbb{E}\left(\sup_{\{||\mathbf{Z}_t-\mathbf{S}_t||>\epsilon\}\ \cup\ \{||\theta-\theta_0||>\epsilon\}} |\eta_\theta^{N_t}(\mathbf{Z}_t)| \bigwedge 1 \middle|\mathcal{F}_T^{(0)}\right) \to 0. \tag{49}$$

From here we have the asymptotic negligibility of Assumption A3. $\square$

8.3 Proof of Theorem 1

We fix an arbitrarily small $\epsilon > 0$. We make use of the following decomposition, for $t = 1, ..., T$ and $j = 1, ..., N_t$:

$$\begin{aligned}\left(\widehat{\kappa}_{t,k_j,\tau_j} - \kappa(k_j, \tau_j, \mathbf{Z}_t, \theta)\right)^2 =& \epsilon_{t,k_j,\tau_j}^2 + \left(\kappa(k_j, \tau_j, \mathbf{S}_t, \theta_0) - \kappa(k_j, \tau_j, \mathbf{Z}_t, \theta)\right)^2 \\ &+ 2\epsilon_{t,k_j,\tau_j}\left(\kappa(k_j, \tau_j, \mathbf{S}_t, \theta_0) - \kappa(k_j, \tau_j, \mathbf{Z}_t, \theta)\right).\end{aligned} \tag{50}$$

Assumption A2 implies that

$$\delta = \inf_{\left(\bigcap_{t=1}^T\ \{||\mathbf{Z}_t-\mathbf{S}_t||\leq\epsilon\}\ \cap\ \{||\theta-\theta_0||\leq\epsilon\}\right)^c} \sum_{t=1}^T \sum_\tau \int_{\underline{k}(t,\tau)}^{\overline{k}(t,\tau)} \left(\kappa(k, \tau, \mathbf{S}_t, \theta_0) - \kappa(k, \tau, \mathbf{Z}_t, \theta)\right)^2 dk$$

is almost surely positive (δ depends on the realization ω). The fact that $\delta > 0$ implies there is at least one time and maturity pairing for which there is a range of moneyness within the interval $(\underline{k}(t,\tau), \overline{k}(t,\tau))$ with positive Lebesgue measure (this range can be different for different ω in the probability space) over which the difference $(\kappa(k,\tau,\mathbf{S}_t,\theta_0) - \kappa(k,\tau,\mathbf{Z}_t,\theta))^2$ is strictly positive. But then, since the mesh of the moneyness for each $t = 1,...,T$ and each time-to-maturity is of size $O(1/N_t)$, we have that the number of pairs (k_j,τ_j) within this range of log-moneyness and time-to-maturity is of order $O(N_t)$ (note that $\psi_{t,\tau}(k)$ of assumption A1 is strictly positive). Next, because κ is continuously differentiable in its log-moneyness argument, there exists an $\widetilde{N}_1 > 0$, almost surely ($\widetilde{N}_1$ depends on ω, of course), such that for $\underline{N} > \widetilde{N}_1$:

$$\inf_{\left(\bigcap_{t=1}^{T} \{||\mathbf{Z}_t - \mathbf{S}_t|| \le \epsilon\} \cap \{||\theta-\theta_0||\le\epsilon\}\right)^c} \sum_{t=1}^{T} \frac{1}{N_t} \sum_{j=1}^{N_t} \left(\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0) - \kappa(k_j,\tau_j,\mathbf{Z},\theta)\right)^2 > \frac{\delta}{2}. \tag{51}$$

Now, assumption A3 implies that, for every infinite subsequence of $\underline{N}$, there exists a further subsequence, denoted N', along which we have,

$$\sup_{\left(\bigcap_{t=1}^{T} \{||\mathbf{Z}_t - \mathbf{S}_t|| \le \epsilon\} \cap \{||\theta-\theta_0||\le\epsilon\}\right)^c} \left| \frac{\sum_{t=1}^{T} \frac{1}{N_t} \sum_{j=1}^{N_t} \left(\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0) - \kappa(k_j,\tau_j,\mathbf{Z},\theta)\right) \epsilon_{t,k_j,\tau_j}}{\sum_{t=1}^{T} \frac{1}{N_t} \sum_{j=1}^{N_t} \left(\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0) - \kappa(k_j,\tau_j,\mathbf{Z},\theta)\right)^2} \right| \longrightarrow 0, \;\; a.s,$$

where $\underline{N}$ in the above almost sure convergence is an element of the subsequence N'. Therefore, for some $\iota \in (0,1)$, there exists $\widetilde{N}_2$, such that for $\underline{N} > \max\{\widetilde{N}_1, \widetilde{N}_2\}$, along the subsequence N', we have on ω

$$\sup_{\left(\bigcap_{t=1}^{T} \{||\mathbf{Z}_t - \mathbf{S}_t|| \le \epsilon\} \cap \{||\theta-\theta_0||\le\epsilon\}\right)^c} \sum_{t=1}^{T} \frac{1}{N_t} \sum_{j=1}^{N_t} \left(\widehat{\kappa}_{t,k_j,\tau_j} - \kappa(k_j,\tau_j,\mathbf{Z}_t,\theta)\right)^2 > \sum_{t=1}^{T} \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon^2_{t,k_j,\tau_j} + \frac{\delta(1-\iota)}{2}, \tag{52}$$

where, again, $\underline{N}$ is an element of the subsequence N' and we also exploited (51). Next, since $\{\widehat{V}_t^n\}_{t=1,...,T}$ is consistent for $\{V_t\}_{t=1,...,T}$, we have, for every infinite subsequence of n, a further subsequence, denoted n', along which we have almost sure convergence, i.e.,

$$\widehat{V}_t^{n'} \to V_t, \quad t = 1,...,T, \quad \text{a.s.}$$

Then, since λ_n converges to a finite λ, there exists $\overline{n} > 0$ such that, for $n > \overline{n}$ along the subsequence n' and on the same ω for which (52) is true, we have,

$$\lambda_n \sup_{t=1,...,T} |\widehat{V}_t^{n'} - V_t|^2 \le \frac{\delta(1-\iota)}{4}. \tag{53}$$

Combining (52) and (53) implies that, along the subsequences N' and n', we have for $\underline{N}$ and n sufficiently high,

$$\sup_{t=1,...,T} ||\widehat{\mathbf{S}}_t^n - \mathbf{S}_t|| < \epsilon \;\text{ and }\; ||\widehat{\theta}^n - \theta_0|| < \epsilon, \quad \text{a.s.} \tag{54}$$

Therefore, since convergence in probability is equivalent to almost sure convergence on a subsequence to any infinite subsequence of the original series, see, e.g., Lemma 3.2 of Kallenberg (1997), we have from (54),

$$\mathbb{P}\left(\sup_{t=1,...,T} ||\widehat{\mathbf{S}}_t^n - \mathbf{S}_t|| > \epsilon \;\cup\; ||\widehat{\theta}^n - \theta_0|| > \epsilon\right) \;\longrightarrow\; 0.$$

Since, the choice of ϵ was arbitrary, the above proves the consistency result of the theorem. $\square$

8.4 Proof of Theorem 2

Exploiting that the implied volatility function is differentiable with respect to the state variables and the parameters of the risk-neutral distribution along with the consistency result in Theorem 1, we have that $\{\widehat{\mathbf{S}}_t^n\}_{t=1,\dots,T}$ and $\widehat{\theta}^n$, with probability approaching 1, solve

$$\begin{cases} \frac{1}{N_1}\sum_{j=1}^{N_1}\left(\widehat{\kappa}_{1,k_j,\tau_j}-\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_1^n,\widehat{\theta}^n)\right)\nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_1^n,\widehat{\theta}^n)+\lambda_n\nabla_{\mathbf{S}}\xi_1(\widehat{\mathbf{S}}_1^n)(\widehat{V}_1^n-\xi_1(\widehat{\mathbf{S}}_1^n))=\mathbf{0}, \\ \vdots \\ \frac{1}{N_T}\sum_{j=1}^{N_T}\left(\widehat{\kappa}_{T,k_j,\tau_j}-\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_T^n,\widehat{\theta}^n)\right)\nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_T^n,\widehat{\theta}^n)+\lambda_n\nabla_{\mathbf{S}}\xi_1(\widehat{\mathbf{S}}_T^n)(\widehat{V}_T^n-\xi_1(\widehat{\mathbf{S}}_T^n))=\mathbf{0}, \\ \sum_{t=1}^{T}\frac{1}{N_t}\sum_{j=1}^{N_t}\left(\widehat{\kappa}_{t,k_j,\tau_j}-\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t^n,\widehat{\theta}^n)\right)\nabla_{\theta}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t^n,\widehat{\theta}^n)=\mathbf{0}. \end{cases} \tag{55}$$

Using a first-order Taylor expansion we obtain,

$$\widetilde{H}_T\begin{pmatrix} \widehat{\mathbf{S}}_1^n-\mathbf{S}_1 \\ \vdots \\ \widehat{\mathbf{S}}_T^n-\mathbf{S}_T \\ \widehat{\theta}^n-\theta_0 \end{pmatrix}=\begin{pmatrix} \frac{1}{N_1}\sum_{j=1}^{N_1}\epsilon_{1,k_j,\tau_j}\nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\mathbf{S}_1,\theta_0) \\ \vdots \\ \frac{1}{N_T}\sum_{j=1}^{N_T}\epsilon_{T,k_j,\tau_j}\nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\mathbf{S}_T,\theta_0) \\ \sum_{t=1}^{T}\frac{1}{N_t}\sum_{j=1}^{N_t}\epsilon_{t,k_j,\tau_j}\nabla_{\theta}\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0) \end{pmatrix}+o_p\left(\frac{1}{\sqrt{\underline{N}}}\right),$$

where $\widetilde{H}_T$ denotes the analogue of $\widehat{H}_T$ in which $\{\widehat{\mathbf{S}}_t^n\}_{t=1,\dots,T}$ is replaced by $\{\widetilde{\mathbf{S}}_t\}_{t=1,\dots,T}$ and $\widehat{\theta}^n$ with $\widetilde{\theta}$ for $\{\widetilde{\mathbf{S}}_t\}_{t=1,\dots,T}$ lying between $\{\widehat{\mathbf{S}}_t^n\}_{t=1,\dots,T}$ and $\{\mathbf{S}_t\}_{t=1,\dots,T}$ and $\widetilde{\theta}$ residing in the interval between $\widehat{\theta}^n$ and θ_0. The o_p term in the expansion stems from the presence of terms depending on $\widehat{V}_t^n$ in the first-order conditions, the fact that $\left(\widehat{\mathbf{S}}_t^n,\widehat{\theta}_t^n\right)$ is consistent (and hence asymptotically bounded in probability), the fact that $\{\widehat{V}_t^n\}_{t=1,\dots,T}$ is asymptotically bounded in probability, and the assumed relation $\lambda_n^2\underline{N}\to 0$ in the theorem.

Since the mesh of the grid on the log-moneyness of the options decreases with $N_t^{\tau}\Delta_{t,\tau}(i)\to\psi_{t,\tau}(k)$ uniformly on the interval $(\underline{k}(t,\tau),\ \overline{k}(t,\tau))$, we trivially have, pathwise,

$$\frac{1}{N_t}\sum_{j=1}^{N_t}\nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\mathbf{Z},\theta)\nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\mathbf{Z},\theta)'\to\sum_{\tau}\int_{\underline{k}(t,\tau)}^{\overline{k}(t,\tau)}\frac{1}{\psi_{t,\tau}(k)}\nabla_{\mathbf{S}}\kappa(k,\tau,\mathbf{Z},\theta),\nabla_{\mathbf{S}}\kappa(k,\tau,\mathbf{Z},\theta)'dk,$$

for any finite $\mathbf{Z}$ and θ. Moreover, since $\nabla_{\mathbf{S}}\kappa(k,\tau,\mathbf{Z},\theta)$ is continuous in the arguments $\mathbf{Z}$ and θ, the above convergence also holds locally uniformly in $\mathbf{Z}$ and θ. Therefore, since $\widetilde{\theta}\overset{\mathbb{P}}{\longrightarrow}\theta_0$ and $\widetilde{\mathbf{S}}_t\overset{\mathbb{P}}{\longrightarrow}\mathbf{S}_t$ for $t=1,\dots,T$, we have $\widetilde{H}_T\overset{\mathbb{P}}{\longrightarrow}H_T$. Combining this result with the limit result in Lemma 1, we establish the asymptotic distribution result in (7). □

8.5 Proof of Theorem 3

We only show the consistency of the block $\widehat{\Omega}_T^{t,t}$, as the proofs for the other blocks of $\widehat{\Omega}_T$ and $\widehat{H}_T$ proceed in an identical fashion. Moreover, it suffices to prove consistency for each of the elements of $\widehat{\Omega}_T^{t,t}$. First, using Assumptions A4(ii), A4(iii) and A4(iv), we have,

$$\begin{aligned} &\mathbb{E}\left\{\left[\frac{1}{N_t}\sum_{j=1}^{N_t}\left(\epsilon_{t,k_j,\tau_j}^2-\phi_{t,k_j,\tau_j}\right)A^{(\iota_1,\iota_2)}(k_j,\tau_j,\mathbf{S}_t,\theta_0)\right]^2\middle|\mathcal{F}^{(0)}\right\} \\ &=\frac{1}{N_t^2}\sum_{j=1}^{N_t}\varrho_{t,k_j,\tau_j}\left(A^{(\iota_1,\iota_2)}(k_j,\tau_j,\mathbf{S}_t,\theta_0)\right)^2\longrightarrow 0,\quad\text{a.s.},\quad \iota_1,\ \iota_2=1,\dots,p, \end{aligned} \tag{56}$$

where $A^{(\iota_1,\iota_2)}(k,\tau,\mathbf{Z},\theta)$ denotes the (ι_1,ι_2) element of the matrix $\nabla_{\mathbf{S}}\kappa(k,\tau,\mathbf{Z},\theta)\nabla_{\mathbf{S}}\kappa(k,\tau,\mathbf{Z},\theta)'$, and where $\varrho_{t,k,\tau} = \mathbb{E}\left((\epsilon^2_{t,k,\tau} - \phi_{t,k,\tau})^2|\mathcal{F}^{(0)}\right)$, which by Assumption A4(iv) is finite. The almost sure convergence in (56) follows because of the pathwise boundedness of the functions $A^{(\iota_1,\iota_2)}(k,\tau,\mathbf{Z},\theta)$ and $\varrho_{t,k,\tau}$ on the range of moneyness $(\underline{k}(t,\tau),\ \overline{k}(t,\tau))$. The convergence in (56) implies,

$$\mathbb{E}\left\{\mathbb{E}\left\{\left[\frac{1}{N_t}\sum_{j=1}^{N_t}\left(\epsilon^2_{t,k_j,\tau_j} - \phi_{t,k_j,\tau_j}\right)A^{(\iota_1,\iota_2)}(k_j,\tau_j,\mathbf{S}_t,\theta_0)\right]^2\Big|\mathcal{F}^{(0)}\right\}\bigwedge 1\right\} \longrightarrow 0. \tag{57}$$

Using Jensen's inequality and (57), as well as law of iterated expectations, we further have,

$$\mathbb{E}\left\{\left[\frac{1}{N_t}\sum_{j=1}^{N_t}\left(\epsilon^2_{t,k_j,\tau_j} - \phi_{t,k_j,\tau_j}\right)A^{(\iota_1,\iota_2)}(k_j,\tau_j,\mathbf{S}_t,\theta_0)\right]^2\bigwedge 1\right\} \longrightarrow 0, \tag{58}$$

which is equivalent to

$$\frac{1}{N_t}\sum_{j=1}^{N_t}\left(\epsilon^2_{t,k_j,\tau_j} - \phi_{t,k_j,\tau_j}\right)\nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0)\nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0)' \stackrel{\mathbb{P}}{\longrightarrow} 0. \tag{59}$$

Finally, using the fact that $\kappa(k,\tau,\mathbf{Z},\theta)$ is twice continuously differentiable in all its arguments, we have for some $\epsilon > 0$, $\mathcal{S} = \{\mathbf{S}_1,...,\mathbf{S}_T\}$, $\underline{k} = \min_{t=1,...,T}\min_\tau \underline{k}(t,\tau)$ and $\overline{k} = \max_{t=1,...,T}\max_\tau \overline{k}(t,\tau)$,

$$\sup_{k\in(\underline{k},\overline{k})}\ \sup_{\mathbf{Z}\in(\mathcal{S}-\epsilon,\mathcal{S}+\epsilon)}\ \sup_{\theta\in(\theta_0-\epsilon,\theta_0+\epsilon)}\left\{\,||\nabla_{\mathbf{S}}\kappa(k,\tau,\mathbf{Z},\theta)|| + ||\nabla_{\mathbf{SS}}\kappa(k,\tau,\mathbf{Z},\theta)||\,\right\} < \infty\,,\quad a.s. \tag{60}$$

Therefore, since for N_t sufficiently high, we will have $\widehat{\mathbf{S}}_t^n$ sufficiently close to $\mathbf{S}_t$ and $\widehat{\theta}^n$ sufficiently close to θ_0, so that we have,

$$\widehat{\Omega}_T^{t,t} - \frac{1}{N_t}\sum_{j=1}^{N_t}\epsilon^2_{t,k_j,\tau_j}\nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0)\,\nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0)' \stackrel{\mathbb{P}}{\longrightarrow} 0. \tag{61}$$

Combining (59) and (61) with the the smoothness of $\phi_{t,k,\tau}$ in its second argument and the smoothness of $\nabla_{\mathbf{S}}\kappa(k,\tau,\mathbf{Z},\theta)$ in the first argument, we conclude, $\widehat{\Omega}_T^{t,t} \stackrel{\mathbb{P}}{\longrightarrow} \Omega_T^{t,t}$. □

8.6 Proof of Theorem 4

Consistency. The proof proceeds in exactly the same way as that of Theorem 1 once the following is established

$$\sup_{||\{\mathbf{Z}_t-\mathbf{S}_t||>\epsilon\}\,\cup\,\{||\theta-\theta_0||>\epsilon\}}\left\{\frac{\sum_{j=1}^{N_t}\widehat{\phi}^{-1}_{t,k_j,\tau_j}\left[\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0) - \kappa(k_j,\tau_j,\mathbf{Z}_t,\theta)\right]\epsilon_{t,k_j,\tau_j}}{\sum_{j=1}^{N_t}\widehat{\phi}^{-1}_{t,k_j,\tau_j}\left[\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0) - \kappa(k_j,\tau_j,\mathbf{Z}_t,\theta)\right]^2} - \frac{\sum_{j=1}^{N_t}\phi^{-1}_{t,k_j,\tau_j}\left[\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0) - \kappa(k_j,\tau_j,\mathbf{Z}_t,\theta)\right]\epsilon_{t,k_j,\tau_j}}{\sum_{j=1}^{N_t}\phi^{-1}_{t,k_j,\tau_j}\left[\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0) - \kappa(k_j,\tau_j,\mathbf{Z}_t,\theta)\right]^2}\right\} \stackrel{\mathbb{P}}{\longrightarrow} 0. \tag{62}$$

The asymptotic negligibility result in (62) follows from combining the following results:

(a) Using the uniform convergence in (11), we have

$$\sup_{j=1,\dots,N_t} |\widehat{\phi}^{-1}_{t,k_j,\tau_j} - \phi^{-1}_{t,k_j,\tau_j}| \xrightarrow{\mathbb{P}} 0, \quad \text{as } N_t \to \infty.$$

(b) By Cauchy-Schwartz inequality

$$\frac{1}{N_t}\sum_{j=1}^{N_t} \phi^{-1}_{t,k_j,\tau_j} \left[\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0) - \kappa(k_j,\tau_j,\mathbf{Z}_t,\theta)\right] \epsilon_{t,k_j,\tau_j}$$

$$\leq \sqrt{\frac{1}{N_t}\sum_{j=1}^{N_t} \phi^{-2}_{t,k_j,\tau_j} \left[\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0) - \kappa(k_j,\tau_j,\mathbf{Z}_t,\theta)\right]^2} \sqrt{\frac{1}{N_t}\sum_{j=1}^{N_t} \epsilon^2_{t,k_j,\tau_j}}.$$

(c) Using Assumption A4, exactly as in (56)-(59), we have

$$\frac{1}{N_t}\sum_{j=1}^{N_t} \epsilon^2_{t,k_j,\tau_j} \xrightarrow{\mathbb{P}} \sum_{\tau} \pi_\tau \int_{\underline{k}(t,\tau)}^{\overline{k}(t,\tau)} \frac{\phi_{t,k,\tau}}{\psi_{t,\tau}(k)} dk.$$

(d) From Assumption A1 and A2, we have

$$\sup_{||\{\mathbf{Z}_t - \mathbf{S}_t||>\epsilon\} \,\cup\, \{||\theta-\theta_0||>\epsilon\}} \left(\frac{1}{N_t}\sum_{j=1}^{N_t} \left[\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0) - \kappa(k_j,\tau_j,\mathbf{Z}_t,\theta)\right]^2\right)^{-1},$$

is almost surely bounded in probability, i.e., for $\omega \in \widetilde{\Omega}$ with $\mathbb{P}(\widetilde{\Omega}) = 1$, there exists $\widetilde{N}$ and $K > 0$ (depending on ω) such that the above is smaller than K for $\underline{N} > \overline{N}$.

Combining the results in (a)-(d) above, we readily have the asymptotic negligibility in (62) and from here the consistency of $\left(\{\overline{\mathbf{S}}^n_t\}_{t=1,\dots,T},\, \overline{\theta}^n\right)$.

Asymptotic Normality. We proceed exactly as in the proof of Theorem 2. A first-order Taylor expansion of the first-order conditions yields

$$(\widetilde{\Lambda}_T + \widetilde{R}_T)\begin{pmatrix} \overline{\mathbf{S}}^n_1 - \mathbf{S}_1 \\ \vdots \\ \overline{\mathbf{S}}^n_T - \mathbf{S}_T \\ \overline{\theta}^n - \theta_0 \end{pmatrix} = \frac{1}{N}\begin{pmatrix} \sum_{j=1}^{N_1} \phi^{-1}_{1,k_j,\tau_j} \epsilon_{1,k_j,\tau_j} \nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\mathbf{S}_1,\theta_0) \\ \vdots \\ \sum_{j=1}^{N_T} \phi^{-1}_{T,k_j,\tau_j} \epsilon_{T,k_j,\tau_j} \nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\mathbf{S}_T,\theta_0) \\ \sum_{t=1}^{T}\sum_{j=1}^{N_t} \phi^{-1}_{t,k_j,\tau_j} \epsilon_{t,k_j,\tau_j} \nabla_{\theta}\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0) \end{pmatrix} + o_p\left(\underline{N}^{-1/2}\right), \quad (63)$$

where $\widetilde{\Lambda}_T$ denotes the analogue of $\widehat{\Lambda}_T$ in which $\{\overline{\mathbf{S}}^n_t\}_{t=1,\dots,T}$ is replaced by $\{\widetilde{\mathbf{S}}_t\}_{t=1,\dots,T}$ and $\overline{\theta}^n$ with $\widetilde{\theta}$ for $\{\widetilde{\mathbf{S}}_t\}_{t=1,\dots,T}$ lying between $\{\overline{\mathbf{S}}^n_t\}_{t=1,\dots,T}$ and $\{\mathbf{S}_t\}_{t=1,\dots,T}$ and $\widetilde{\theta}$ residing in the interval between $\overline{\theta}^n$ and θ_0. $\widetilde{R}_T$ is a matrix containing second-order derivatives of $\kappa(k,\tau,\mathbf{Z},\theta)$. From the assumption of the theorem, the latter are continuous and since $\widetilde{\theta} \xrightarrow{\mathbb{P}} \theta_0$ and $\widetilde{\mathbf{S}}_t \xrightarrow{\mathbb{P}} \mathbf{S}_t$ for $t = 1,\dots,T$, we easily have $\widetilde{R}_T \xrightarrow{\mathbb{P}} \mathbf{0}$. Finally, the o_p term in the expansion stems from the presence of terms depending on $\widehat{V}^n_t$ in the first-order conditions, the fact that $\left(\overline{\mathbf{S}}^n_t, \overline{\theta}^n_t\right)$ is consistent (and hence asymptotically bounded in probability), the fact that $\{\widehat{V}^n_t\}_{t=1,\dots,T}$ is asymptotically bounded in probability, and the assumed relation $\lambda_n^2 \underline{N} \to 0$ in the theorem as well as the consistency result for $\widehat{\phi}_{t,k,\tau}$ in (11) and the asymptotic result in (64) below.

The consistency of $\widetilde{\Lambda}_T$ is proved in exactly the same way as the consistency of $\widetilde{H}_T$ in Theorem 2. Hence, to prove (13), we need to show stable convergence of the (scaled) term on the right-hand side of (63). The latter is given by

$$\frac{1}{\sqrt{N}} \begin{pmatrix} \sum_{j=1}^{N_1} \phi_{1,k_j,\tau_j}^{-1} \epsilon_{1,k_j,\tau_j} \nabla_{\mathbf{S}} \kappa(k_j, \tau_j, \mathbf{S}_1, \theta_0) \\ \vdots \\ \sum_{j=1}^{N_T} \phi_{T,k_j,\tau_j}^{-1} \epsilon_{T,k_j,\tau_j} \nabla_{\mathbf{S}} \kappa(k_j, \tau_j, \mathbf{S}_T, \theta_0) \\ \sum_{t=1}^{T} \sum_{j=1}^{N_t} \phi_{t,k_j,\tau_j}^{-1} \epsilon_{t,k_j,\tau_j} \nabla_{\theta} \kappa(k_j, \tau_j, \mathbf{S}_t, \theta_0) \end{pmatrix} \xrightarrow{\mathcal{L}-s} \mathbf{\Lambda}_T^{-1/2} \begin{pmatrix} \mathbf{E}_1 \\ \vdots \\ \mathbf{E}_T \\ \mathbf{E}' \end{pmatrix}, \tag{64}$$

and is proved exactly the same way as Lemma 1 since $\phi_{t,k,\tau}$ is $\mathcal{F}_T^{(0)}$-adapted.

The consistency of $\widehat{\Lambda}_T$ is shown exactly in the same way as the consistency of $\widehat{H}_T$ in the proof of Theorem 3. □

8.7 Proof of Theorem 5

The results follows directly from combining Lemmas 2-4. □

8.8 Proof of Theorem 6

The proof of the theorem follows the same steps as the proof of Theorem 2. Using a first-order Taylor expansion of the first-order condition in (55) we obtain,

$$(\widetilde{H}_T + \widetilde{\mathbf{D}}_T + \widetilde{R}_T) \begin{pmatrix} \widehat{\mathbf{S}}_1^n - \mathbf{S}_1 \\ \vdots \\ \widehat{\mathbf{S}}_T^n - \mathbf{S}_T \\ \widehat{\theta}^n - \theta_0 \end{pmatrix} = \begin{pmatrix} \frac{1}{N_1} \sum_{j=1}^{N_1} \epsilon_{1,k_j,\tau_j} \nabla_{\mathbf{S}} \kappa(k_j, \tau_j, \mathbf{S}_1, \theta_0) + \lambda_n \nabla_{\mathbf{S}} \xi_1(\mathbf{S}_1)(\widehat{V}_1^n - V_1) \\ \vdots \\ \frac{1}{N_T} \sum_{j=1}^{N_T} \epsilon_{T,k_j,\tau_j} \nabla_{\mathbf{S}} \kappa(k_j, \tau_j, \mathbf{S}_T, \theta_0) + \lambda_n \nabla_{\mathbf{S}} \xi_1(\mathbf{S}_T)(\widehat{V}_T^n - V_T) \\ \sum_{t=1}^{T} \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{t,k_j,\tau_j} \nabla_{\theta} \kappa(k_j, \tau_j, \mathbf{S}_t, \theta_0) \end{pmatrix}, \tag{65}$$

where $\widetilde{H}_T$ and $\widetilde{D}_T$ denote the analogues of $\widehat{H}_T$ and $\widehat{D}_T$ in which $\{\widehat{\mathbf{S}}_t^n\}_{t=1,...,T}$ is replaced by $\{\widetilde{\mathbf{S}}_t\}_{t=1,...,T}$ and $\widehat{\theta}^n$ with $\widetilde{\theta}$ for $\{\widetilde{\mathbf{S}}_t\}_{t=1,...,T}$ lying between $\{\widehat{\mathbf{S}}_t^n\}_{t=1,...,T}$ and $\{\mathbf{S}_t\}_{t=1,...,T}$ and $\widetilde{\theta}$ residing in the interval between $\widehat{\theta}^n$ and θ_0; $\widetilde{R}_T$ is a matrix containing second-order derivatives of $\kappa(k, \tau, \mathbf{Z}, \theta)$ and $\xi_1(\mathbf{Z})$. From the assumption of the theorem, the latter are continuous and since $\widetilde{\theta} \xrightarrow{\mathbb{P}} \theta_0$ and $\widetilde{\mathbf{S}}_t \xrightarrow{\mathbb{P}} \mathbf{S}_t$ for $t = 1, ..., T$, we easily have $\widetilde{R}_T \xrightarrow{\mathbb{P}} \mathbf{0}$. From the proof of Theorem 2, we have $\widetilde{H}_T \xrightarrow{\mathbb{P}} \mathbf{H}_T$ and we can prove analogously $\widetilde{D}_T \xrightarrow{\mathbb{P}} \mathbf{D}_T$.

Hence to prove the convergence in (16), we are left with the term on the right-hand side of (65). First, since the convergence result in Lemma 2 is stable, we also have (making use in addition of Lemma 4)

$$\sqrt{k_n} \begin{pmatrix} \nabla_{\mathbf{S}} \xi_1(\mathbf{S}_1)(\widehat{V}_1^{-,n} - V_1) \\ \vdots \\ \nabla_{\mathbf{S}} \xi_1(\mathbf{S}_T)(\widehat{V}_T^{-,n} - V_T) \end{pmatrix} \xrightarrow{\mathcal{L}-s} \begin{pmatrix} \sqrt{2} V_1 \nabla_{\mathbf{S}} \xi_1(\mathbf{S}_1) \widetilde{E}_1 \\ \vdots \\ \sqrt{2} V_T \nabla_{\mathbf{S}} \xi_1(\mathbf{S}_T) \widetilde{E}_T \end{pmatrix}, \tag{66}$$

where $\{\widetilde{E}_t\}_{t \geq 1}$ are defined in Theorem 5. Now exactly as in Lemma 3, we have the stable convergence in (66) to hold jointly with that in (23) of Lemma 1. Combining this with the convergence in probability of $\widetilde{H}_T + \widetilde{\mathbf{D}}_T + \widetilde{R}_T$, together with the properties of the stable convergence (Proposition VIII.5.33 in Jacod and Shiryaev (2003)), we have the result in (16).

Finally, the consistency of $\widehat{\mathbf{D}}_T$ and $\widehat{\Sigma}_T$ can be proved exactly in the same way as the consistency of $\widehat{\mathbf{H}}_T$ and $\widehat{\Omega}_T$ in the proof of Theorem 3. $\square$

8.9 Proof of Corollary 1

The proof follows from a trivial extension of Lemma 1, the smoothness of the function $\kappa(k,\tau,\mathbf{Z},\theta)$ in its arguments, the consistency of $\{\widehat{\mathbf{S}}_t^n\}_{t=1,\dots,T}$ and $\widehat{\theta}^n$, as well as the properties of the stable convergence. First, a Taylor expansion yields

$$\sum_{j:k_j\in\mathcal{K}}\left(\widehat{\kappa}_{t,k_j,\tau^*} - \kappa(k_j,\tau^*,\widehat{\mathbf{S}}_t^n,\widehat{\theta})\right) = \sum_{j:k_j\in\mathcal{K}}\epsilon_{t,k_j,\tau^*} - \left(\sum_{j:k_j\in\mathcal{K}}\nabla_{\mathbf{S}}\kappa(k_j,\tau^*,\widetilde{\mathbf{S}}_t,\widetilde{\theta})\right)\left(\widehat{\mathbf{S}}_t^n - \mathbf{S}_t\right)$$
$$- \left(\sum_{j:k_j\in\mathcal{K}}\nabla_{\theta}\kappa(k_j,\tau^*,\widetilde{\mathbf{S}}_t,\widetilde{\theta})\right)\left(\widehat{\theta}^n - \theta_0\right),$$

where $\{\widetilde{\mathbf{S}}_t\}_{t=1,\dots,T}$ is between $\{\widehat{\mathbf{S}}_t^n\}_{t=1,\dots,T}$ and $\{\mathbf{S}_t\}_{t=1,\dots,T}$ and $\widetilde{\theta}$ lies between $\widehat{\theta}^n$ and θ_0. Therefore, using the consistency of $\{\widehat{\mathbf{S}}_t^n\}_{t=1,\dots,T}$ and $\widehat{\theta}$, as well as the smoothness of $\kappa(k,\tau,\mathbf{Z},\theta)$ (and its derivatives with respect to $\mathbf{Z}$ and θ) in the log-moneyness, we obtain,

$$\frac{1}{\sqrt{N_t^{\mathcal{K}}}}\sum_{j:k_j\in\mathcal{K}}\left(\widehat{\kappa}_{t,k_j,\tau^*} - \kappa(k_j,\tau^*,\widehat{\mathbf{S}}_t^n,\widehat{\theta})\right) = \mathbf{\Pi}_T\begin{pmatrix}\sqrt{N_1}(\widehat{\mathbf{S}}_1^n - \mathbf{S}_1)\\ \vdots \\ \sqrt{N_T}(\widehat{\mathbf{S}}_T^n - \mathbf{S}_T)\\ \sqrt{\frac{N_1+\ldots+N_T}{T}}(\widehat{\theta}^n - \theta_0)\\ \frac{1}{\sqrt{N_t^{\mathcal{K}}}}\sum_{j:k_j\in\mathcal{K}}\epsilon_{t,k_j,\tau^*}\end{pmatrix} + o_p, \tag{67}$$

where

$$\mathbf{\Pi}_T = \begin{pmatrix}\mathbf{0}_{1\times(t-1)p} & -\int_{\mathcal{K}}\frac{1}{\psi_{t,\tau}(k)}\nabla_{\mathbf{S}}\kappa(k,\tau,\mathbf{S}_t,\theta)'dk & \mathbf{0}_{1\times(T-t+1)p} & -\sqrt{T\varsigma_t}\int_{\mathcal{K}}\frac{1}{\psi_{t,\tau}(k)}\nabla_{\theta}\kappa(k,\tau,\mathbf{S}_t,\theta)'dk & 1\end{pmatrix}.$$

Next, using the Taylor expansion in (56) and the consistency of $\{\widehat{\mathbf{S}}_t^n\}_{t=1,\dots,T}$ and $\widehat{\theta}^n$, we get,

$$\frac{1}{\sqrt{N_t^{\mathcal{K}}}}\sum_{j:k_j\in\mathcal{K}}\left(\widehat{\kappa}_{t,k_j,\tau^*} - \kappa(k_j,\tau^*,\widehat{\mathbf{S}}_t^n,\widehat{\theta})\right) = \mathbf{\Pi}_T\begin{pmatrix}\mathbf{H}_T^{-1} & \mathbf{0}_{pT+q\times 1}\\ \mathbf{0}_{1\times pT+q} & 1\end{pmatrix}\begin{pmatrix}\zeta_{\{N_t\}_{t=1,\dots,T}}\\ \frac{1}{\sqrt{N_t^{\mathcal{K}}}}\sum_{j:k_j\in\mathcal{K}}\epsilon_{t,k_j,\tau^*}\end{pmatrix} + o_p, \tag{68}$$

where we denote

$$\zeta_{\{N_t\}_{t=1,\dots,T}} = \begin{pmatrix}\frac{1}{\sqrt{N_1}}\sum_{j=1}^{N_1}\epsilon_{1,k_j,\tau_j}\nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\mathbf{S}_1,\theta_0)\\ \vdots \\ \frac{1}{\sqrt{N_T}}\sum_{j=1}^{N_T}\epsilon_{T,k_j,\tau_j}\nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\mathbf{S}_T,\theta_0)\\ \sqrt{\frac{\sum_{t=1}^T N_t}{T}}\sum_{t=1}^T\frac{1}{N_t}\sum_{j=1}^{N_t}\epsilon_{t,k_j,\tau_j}\nabla_{\theta}\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0)\end{pmatrix}.$$

Then, upon following exactly the same steps as in the proof of Lemma 1, we obtain,

$$\begin{pmatrix}\zeta_{\{N_t\}_{t=1,\dots,T}}\\ \frac{1}{\sqrt{N_t^{\mathcal{K}}}}\sum_{j:k_j\in\mathcal{K}}\epsilon_{t,k_j,\tau^*}\end{pmatrix} \xrightarrow{\mathcal{L}-s} \mathbf{Z}_T, \tag{69}$$

where $\mathbf{Z}_T$, defined on an extension of the original probability space, is $\mathcal{F}_T^{(0)}$-Gaussian with a $\mathcal{F}_T^{(0)}$-conditional covariance matrix given by

$$\begin{pmatrix} \mathbf{\Omega}_T & \mathbf{\Upsilon}_{1,T} \\ \mathbf{\Upsilon}'_{1,T} & \mathbf{\Upsilon}_{2,T} \end{pmatrix}, \quad \mathbf{\Upsilon}_{1,T} = \begin{pmatrix} \mathbf{0}_{(t-1)p\times 1} \\ \frac{1}{\sqrt{\int_{\mathcal{K}} \frac{1}{\psi_{t,\tau}(k)}dk}} \int_{\mathcal{K}} \frac{1}{\psi_{t,\tau}(k)} \phi_{t,k,\tau} \nabla_{\mathbf{S}} \kappa(k,\tau,\mathbf{S}_t,\theta_0)dk \\ \mathbf{0}_{(T-t+1)p\times 1} \\ \frac{1}{\sqrt{T\varsigma_t}} \frac{1}{\sqrt{\int_{\mathcal{K}} \frac{1}{\psi_{t,\tau}(k)}dk}} \int_{\mathcal{K}} \frac{1}{\psi_{t,\tau}(k)} \phi_{t,k,\tau} \nabla_{\theta} \kappa(k,\tau,\mathbf{S}_t,\theta_0)dk \end{pmatrix},$$

$$\mathbf{\Upsilon}_{2,T} = \frac{1}{\int_{\mathcal{K}} \frac{1}{\psi_{t,\tau}(k)}dk} \int_{\mathcal{K}} \frac{1}{\psi_{t,\tau}(k)} \phi_{t,k,\tau} dk.$$

Then, using the same techniques as in the proof for consistency of $\widehat{\mathbf{\Omega}}_T$ (see (56)-(61) above), we can show that, $\frac{1}{N_t^{\mathcal{K}}} \widehat{\Pi}'_T \, \widehat{\Xi}_T \, \widehat{\Pi}_T \stackrel{\mathbb{P}}{\longrightarrow} \Pi'_T \, \Xi_T \, \Pi_T$, where Ξ_T is defined exactly as $\widehat{\Xi}_T$, but with each matrix replaced by its analogue matrix without the hat. Now, using the definition of stable convergence, we obtain the requisite limit result in (17). □

8.10 Proof of Corollary 2

The proof follows from an extension of Lemma 1 and Theorem 2 for two panels with disjoint time intervals. □

8.11 Proof of Corollary 3

By Lemma 3, we have that $\{\frac{1}{\sqrt{N_t}}\left(\widehat{\mathbf{S}}_t^n - \mathbf{S}_t\right)\}_{t=1,\dots,T}$ and $\sqrt{k_n}\{\widehat{V}_t^{-,n} - V_t\}_{t=1,\dots,T}$ are, conditionally on $\mathcal{F}_T$, asymptotically independent and normally distributed. Since the limit results of Lemma 3 hold stably with respect to $\mathcal{F}_T$, and from Theorem 2, the corresponding part of the matrix $\widehat{\mathbf{H}}_T^{-1}\widehat{\mathbf{\Omega}}_T(\widehat{\mathbf{H}}_T^{-1})'$ provides a consistent estimator for the asymptotic variance of $\{\frac{1}{\sqrt{N_t}}\left(\widehat{\mathbf{S}}_1^n - \mathbf{S}_1\right)\}_{t=1,\dots,T}$, then by applying the Delta method (recall that ξ_1 is continuously differentiable), we obtain the limit result of the Corollary. □

References

Andersen, T. G. and T. Bollerslev (1998). Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts. *International Economic Review 39*, 885–905.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and Forecasting Realized Volatility. *Econometrica 71*, 579–625.

Andrews, D. (2005). Cross-Section Regression with Common Shocks. *Econometrica 71*, 1551–1585.

Barndorff-Nielsen, O. and N. Shephard (2002). Econometric Analysis of Realized Volatility and its Use in Estimating Stochastic Volatility Models. *Journal of the Royal Statistical Society Series B, 64*, 253–280.

Barndorff-Nielsen, O. E., S. E. Graversen, J. Jacod, and N. Shephard (2006). Limit Theorems for Realised Bipower Variation in Econometrics. *Econometric Theory 22*, 677–719.

Barndorff-Nielsen, O. E. and N. Shephard (2006). Econometrics of Testing for Jumps in Financial Economics using Bipower Variation. *Journal of Financial Econometrics 4*, 1–30.

Bates, D. S. (2000). Post-'87 Crash Fears in S&P 500 Future Options. *Journal of Econometrics 94*, 181–238.

Bates, D. S. (2003). Empirical Option Pricing: A Retrospection. *Journal of Econometrics 116*, 387–404.

Bollerslev, T. and V. Todorov (2011). Tails, Fears and Risk Premia. *Journal of Finance 66*, 2165–2211.

Broadie, M., M. Chernov, and M. Johannes (2009). Specification and Risk Premiums: The Information in S&P 500 Futures Options. *Journal of Finance 62*, 1453–1490.

Christoffersen, P. and K. Jacobs (2004). The Importance of the Loss Function in Option Valuation. *Journal of Financial Economics 72*, 291–318.

Dellacherie, C. and P. Meyer (1978). *Probabilities and Potential*. Paris: Hermann.

Drechsler, I. (2013). Uncertainty, Time-Varying Fear, and Asset Prices. *Journal of Finance, forthcoming*.

Drechsler, I. and A. Yaron (2011). What's Vol Got to Do with It? *Review of Financial Studies 24*, 1–45.

Du, D. (2010). General Equilibrium Pricing of Options with Habit Formation and Event Risks. *Journal of Financial Economics 99*, 400–426.

Duffie, D. (2001). *Dynamic Asset Pricing Theory* (3rd ed.). Princeton University Press.

Duffie, D., D. Filipović, and W. Schachermayer (2003). Affine Processes and Applications in Finance. *Annals of Applied Probability 13(3)*, 984–1053.

Duffie, D., J. Pan, and K. Singleton (2000). Transform Analysis and Asset Pricing for Affine Jump-Diffusions. *Econometrica 68*, 1343–1376.

Eraker, B. (2004). Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices. *Journal of Finance 59*, 1367–1403.

Gagliardini, P., C. Gourieroux, and E. Renault (2011). Efficient Derivative Pricing by the Extended Method of Moments. *Econometrica 79*, 1181–1232.

Garcia, R., E. Ghysels, and E. Renault (2010). The Econometrics of Option and Pricing. In Y. Aït-Sahalia and L. Hansen (Eds.), *Handbook of Financial Econometrics*, Volume 1, pp. 479–552. Oxford and Amsterdam: North-Holland.

Gourieroux, C., A. Monfort, and A. Trognon (1984). Pseudo Maximum Likelihood Methods: Theory. *Econometrica 59*, 681–700.

Hausman, J. (1978). Specification Tests in Econometrics. *Econometrica 46*, 1251–1271.

Ibragimov, I. and R. Has'minskii (1981). *Statistical Estimation: Asymptotic Theory*. Berlin: Springer.

Jacod, J. (2008). Asymptotic Properties of Power Variations and Associated Functionals of Semimartingales. *Stochastic Processes and their Applications 118*, 517–559.

Jacod, J., Y. Li, P. Mykland, M. Podolskij, and M. Vettter (2009). Microstructure Noise in the Continuous Case: The Pre-Averaging Approach. *Stochastic Processes and their Applications 119*, 2249–2276.

Jacod, J. and P. Protter (2012). *Discretization of Processes*. Berlin: Springer-Verlag.

Jacod, J. and A. N. Shiryaev (2003). *Limit Theorems For Stochastic Processes* (2nd ed.). Berlin: Springer-Verlag.

Jacod, J. and V. Todorov (2010). Do Price and Volatility Jump Together? *Annals of Applied Probability 20*, 1425–1469.

Jones, C. (2006). Analysis of Index Option Returns. *Journal of Finance LXI*, 2325–2363.

Kallenberg, O. (1997). *Foundations of Modern Probability.* New York: Springer-Verlag.

Kuersteiner, G. and I. Prucha (2011). Limit Theory for Panel Data Models with Cross Sectional Dependence and Sequential Exogenuity. Working paper, Georgetown University and University of Maryland.

Mancini, C. (2001). Disentangling the Jumps of the Diffusion in a Geometric Brownian Motion. *Giornale dell'Istituto Italiano degi Attuari LXIV*, 19–47.

Newey, W. (1985). Maximum Likelihood Specification Testing and Conditional Moment Tests. *Econometrica 53*, 1047–1070.

Newey, W. (1991). Uniform Convergence in Probability and Stochastic Equicontinuity. *Econometrica 59*, 1161–1167.

Newey, W. and D. McFadden (1994). Large Sample Estimation and Hypothesis Testing. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, pp. 2113–2241. Amsterdam: North-Holland.

Pan, J. (2002). The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study. *Journal of Financial Economics 63*, 3–50.

Renault, E. (1997). Econometric Models of Option Pricing. In D. Kreps and K. Wallis (Eds.), *Advances in Economics and Econometrics, Seventh World Congress*, pp. 223–278. Cambridge University Press.

Robinson, P. (1987). Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unkwown Form. *Econometrica 55*, 875–891.

Singleton, K. (2006). *Empirical Dynamic Asset Pricing.* Princeton University Press.

Tauchen, G. (1985). Diagnostic Testing and Evaluation of Maximum Likelihood Models. *Journal of Econoemtrics 30*, 415–443.

van der Vaart, A. and J. Wellner (1996). *Weak Convergence and Empirical Processes.* New York: Springer.

Wachter, J. A. (2013). Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility? *Journal of Finance, forthcoming*.

White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica 50*, 1–25.