NBER WORKING PAPER SERIES

# PARAMETRIC INFERENCE AND DYNAMIC STATE RECOVERY FROM OPTION PANELS

Torben G. Andersen
Nicola Fusari
Viktor Todorov

Working Paper 18046
http://www.nber.org/papers/w18046

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2012

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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Parametric Inference and Dynamic State Recovery from Option Panels
Torben G. Andersen, Nicola Fusari, and Viktor Todorov
NBER Working Paper No. 18046
May 2012
JEL No. C51,C52,C58,G12,G13

## ABSTRACT

We develop a new parametric estimation procedure for option panels observed with error which relies on asymptotic approximations assuming an ever increasing set of observed option prices in the moneyness-maturity (cross-sectional) dimension, but with a fixed time span. We develop consistent estimators of the parameter vector and the dynamic realization of the state vector that governs the option price dynamics. The estimators converge stably to a mixed-Gaussian law and we develop feasible estimators for the limiting variance. We provide semiparametric tests for the option price dynamics based on the distance between the spot volatility extracted from the options and the one obtained nonparametrically from high-frequency data on the underlying asset. We further construct new formal tests of the model fit for specific regions of the volatility surface and for the stability of the risk-neutral dynamics over a given period of time. A large-scale Monte Carlo study indicates the inference procedures work well for empirically realistic specifications and sample sizes. In an empirical application to S&P 500 index options we extend the popular double-jump stochastic volatility model to allow for time-varying jump risk premia and a flexible relation between risk premia and the level of risk. Both extensions lead to an improved characterization of observed option prices.

Torben G. Andersen
Kellogg School of Management
Northwestern University
2001 Sheridan Road
Evanston, IL 60208
and NBER
t-andersen@kellogg.northwestern.edu

Viktor Todorov
Department of Finance
Kellogg School of Management
Northwestern University
2001 Sheridan Road
Evanston, IL
v-todorov@kellogg.northwestern.edu

Nicola Fusari
Department of Finance
Kellogg School, Northwestern University
Evanston, IL 60208
n-fusari@northwestern.edu

# 1 Introduction

A voluminous literature spanning several decades has, unambiguously, established that time-varying volatility and jumps are intrinsic features of financial asset price processes. More recently, there has been substantial interest in linking financial risk premiums to the compensation for such factors. Indeed, the evidence indicates that the pricing of volatility and jump risks is critical for understanding the magnitude and variation in both equity and variance risk premiums, see, e.g., Bates (2000), Pan (2002), Broadie et al. (2009) and Bollerslev and Todorov (2011). In parallel, the trading of derivative contracts has grown explosively, in part reflecting a desire among investors to actively manage volatility and jump risk exposures. As a result, ever more comprehensive price data for, in particular, exchange-traded options have become available over time.[1] 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. For the horizons spanned by the surface, we may extract risk-neutral density estimates for the equity index and price any payoff expressed as a smooth function of future index values. Moreover, we may infer the value of various path-dependent payoffs, including the future realized return variation. In other words, a sequence of equity-index option surfaces – which we label an *option panel* – contains valuable information about the dynamic pricing of future economy-wide contingencies.

In this paper, we develop rigorous estimation and inference tools for extracting information from option panels under minimal auxiliary conditions. That is, we develop formal inference techniques for the implied (latent) state vector and the risk-neutral, or pricing, distribution, while avoiding parametric assumptions about the actual measure governing the state vector dynamics. In fact, the latter may be non-stationary. 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 also allow for the option prices to be observed with errors exhibiting limited dependence in the spatial (across strikes and tenors) and time series dimension. We accommodate the case where the strike range and tenor of the option surface, as well as the total number of option quotes, change across time – as in the data – and there is no requirement of stationarity in the pattern of maturity and moneyness. Similarly, we allow for the observation error to 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-squared-error in fitting the observed option prices

[1] For example, on average, over 230 active bid-ask quotes with positive bid prices for out-of-the-money options are reported daily, at the end-of-trading, for our sample of options on the S&P 500 index at the Chicago Board Options Exchange (CBOE) during $1996 - 2010$, and this number is significantly higher in the second part of the sample.

over the estimation period 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 governing the option price dynamics.

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 our estimator is efficient if the option errors are homoskedastic, and it may be rendered efficient otherwise 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.

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 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 estima-

tors, enabling us to devise a formal model specification test based on the distance between the two volatility measures. Intuitively, this is feasible as, even though different volatility states (or jump intensities) are not directly observed, the (total) diffusive volatility may be filtered from the underlying asset data,[2] and this value should coincide with that priced in the observed options.

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 tests if it differs significantly from zero. In essence, the approach allows us to disentangle 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. We find the inference technique to perform admirably within realistically calibrated settings. Furthermore, in an empirical application using an extensive option panel for the S&P 500 index, we estimate the model along with some new extensions. The expanded system affords compensation for rare events through a time-varying jump intensity as well as a two-factor volatility structure that adds flexibility to the link between the level of risk and the associated compensation. The extension provides improvements in explaining the option price dynamics, especially for the longer maturities, and it brings the extracted volatility state closer to a nonparametric estimate constructed from high-frequency S&P 500 index returns. Nevertheless, even our most general representation fails our full battery of specification tests. In particular, it cannot suitably capture the rate of decay in option prices as we move deeper into the out-of-the-money region. Our diagnostics suggest that alternative specifications for the jump distribution, involving a more gradual decay in the Lévy density, are critical for alleviating such shortcomings.

[2]See, e.g., Foster and Nelson (1996) for early work on this subject.

The rest of the paper is organized as follows. Section 2 positions our contribution relative to the dominant paradigms in the literature. Section 3 introduces our formal setup. Section 4 develops our estimators and derives the feasible limit theory. In Section 5, we develop diagnostic tests for the option price dynamics. Section 6 contains a Monte Carlo study of the proposed estimators. In Section 7, we apply our estimation methods to analyze the option price dynamics of the S&P 500 index. Section 8 concludes. All proofs are deferred to the appendix.

## 2 Relation to Existing Option Pricing Paradigms

Two separate empirical approaches are dominant, reflecting different objectives in extracting information from options. One set of studies focuses on point-in-time nonparametric approximation of the pricing functional for basic derivative securities, for example by smoothing the individually observed Black-Scholes implied volatilities into a coherent surface covering all strikes and tenors of interest. Closely related procedures generate nonparametric estimates of the risk-neutral density for the underlying asset at a given maturity using option prices across all available strikes, exploiting ideas of Ross (1976) and Breeden and Litzenberger (1978). Again, by interpolation and extrapolation, this procedure can generate risk-neutral density estimates for a broad range of maturities as long as reliable option prices or quotes are available. However, a defining characteristic of this approach is that option surfaces at different points in time are treated separately. There is no attempt at enforcing coherence across the estimates of the risk-neutral measure, henceforth denoted $\mathbb{Q}$, across trading periods. In other words, the procedure does not seek to infer the dynamic evolution of the risk-neutral probability measure. Instead, the objective of such "calibration" exercises is to provide a basis for coherent valuation of derivative securities at a point in time using the contemporaneous pricing structure extracted from the universe of actively traded options.

In contrast, a separate branch of the literature focuses on the dynamics of option pricing. Here, the primary object of interest is extraction of the state vector dynamics, under the actual probability measure, $\mathbb{P}$, along with coherent modeling of the risk-neutral measure over time. These studies are almost invariably parametric and exploit only a limited set of observations from the option panel along with time series data for the underlying security to simultaneously estimate the return generating process under $\mathbb{P}$ and $\mathbb{Q}$. The approach relies, almost exclusively, on the time series of underlying asset returns for estimation of the $\mathbb{P}$ dynamics, exploiting standard long span asymptotics, while it infers the $\mathbb{Q}$ parameters from the options.[3] Specifically, imposing tight restrictions between the $\mathbb{P}$ and $\mathbb{Q}$ measures, the option prices identify a few auxiliary ($\mathbb{Q}$) risk

[3]Pan (2002) and Pastorello et al. (2003) use also limited number of options (determined by the dimension of the latent state vector) in the inference of the $\mathbb{P}$ dynamics.

premium parameters.[4] In such a setting, specification testing relies critically on the joint hypothesis that the parametric models for the risk-neutral *and* the statistical distribution are well specified. Thus, even if this approach is theoretically more consistent than the point-in-time calibration of the $\mathbb{Q}$ density, it is subject to a number of pitfalls. There is no compelling reason that the two distributions should be identical except for a few shifts in parameter values. It is possible – perhaps even likely – that there are state variables characterizing the broader economic landscape which exert an important impact on the risk premiums, and thus the $\mathbb{Q}$ parameters, but have a negligible impact on the $\mathbb{P}$ dynamics. The existence of such state variables is hard to ascertain via procedures that rely primarily on the asset returns for identification of the system dynamics.[5] Failure to account for the full state vector typically manifests itself through instability in the $\mathbb{Q}$ parameters as the neglected state variables shift over time. Such instability may materialize only gradually if the macroeconomic conditions change slowly, but the environment may also undergo a sudden regime shift, with an immediate and dramatic effect on risk premiums.[6]

Related problems arise from the limited, and inefficient, use of option panels. Relying on only a few options, often representing just one or two separate maturities and a narrow strike range, complicates the identification of critical features of the $\mathbb{Q}$ dynamics. Hence, even when using the correct set of state variables, the power to detect model misspecification is hampered by the systematic exclusion of informative data points. Finally, the presence of transaction costs, such as bid-ask spreads, renders individual option prices noisy indicators of underlying valuations. Using a larger set of options facilitates improved noise filtering and thus more precise inference.[7]

The current work is motivated by a desire to overcome a number of the issues noted above. We seek to conduct parametric inference regarding the $\mathbb{Q}$ parameters *and* the realized values of the state vector, while exploiting the information in the option panel efficiently. In order to accommodate potential instability of the $\mathbb{Q}$ measure, relative to the identifiable state variables, we develop inference techniques that require only a limited time span. Moreover, we accommodate general types of option pricing errors. Finally, we stress the importance of developing powerful diagnostics

[4] There are a few empirical papers which rely solely on a long option panel with a fixed cross-section (i.e., without explicitly modeling the $\mathbb{P}$ law) but the formal treatment of the option error in this case is not clear.

[5] A corresponding argument applies to models for the term structure of interest rates. In this context, Joslin et al. (2010) document important links between (unspanned components of) the macroeconomic environment and the pricing of risk.

[6] For example, one cannot rationalize the permanent shift in the implied volatility skew for equity-index options following October 1987 via state variables estimated solely from stock returns. One way to accommodate this change is by a shift in an underlying state variable impacting only the risk-neutral dynamics. This illustrates the need for flexibility in modeling risk premiums. Similar, albeit more gradually evolving, problems may plague empirical studies based on option panels and return series covering long time spans.

[7] See, e.g., Bates (2003) for a discussion of these issues.

for adequate fit to the option panel. In order for this approach to be widely applicable, it should be valid regardless of the nature of the $\mathbb{P}$ dynamics. We ensure robustness in this dimension by avoiding specific assumptions regarding the objective probability measure.[8]

In summary, we adopt an expansive view of the informativeness of option data. Our premise is that the option panel – in-and-of-itself – suffices for identifying both the risk-neutral distribution and the realized trajectory of the underlying state vector. As a result, we deviate from the traditional approaches along critical dimensions. First, we avoid specifying a parametric model for the underlying asset price process, and we do not rely on long-span asymptotics. This allows us to estimate the $\mathbb{Q}$ distribution consistently in a flexible manner over short intervals of time, independently of the $\mathbb{P}$ dynamics. Moreover, we can test for parameter stability and we may explicitly accommodate shifts in this distribution.[9] Second, we develop formal diagnostics that allow for direct testing of whether the current diffusive volatility state under the $\mathbb{P}$ and $\mathbb{Q}$ measures are identical. In addition, we provide diagnostics that focus on the fit to limited regions of the volatility surface and over short time intervals, thus enhancing the power of these specification tests. Third, prior work conducting inference from options either excludes observation errors or assumes they are Gaussian. By contrast, exploiting the large option price surface, we can be nonparametric about the observation error and allow for their presence in our feasible limit theory. Fourth, while the "calibration" approach uses a flexible functional form to fit the implied volatility surface, typically exploiting an independent calibration at each point in time, we impose temporal stability of the risk-neutral parameters over (limited periods of) time, enabling formal inference.

# 3 The Basic Modeling Framework

## 3.1 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 $\left(\Omega^{(0)}, \mathcal{F}^{(0)}, (\mathcal{F}_t^{(0)})_{t\geq 0}, \mathbb{P}^{(0)}\right)$. It is assumed to have the following very general dynamics (under $\mathbb{P}^{(0)}$)

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

[8] Nonetheless, since our approach produces a time series of estimates for the realization of the state vector, it should facilitate the development of inference tools for the $\mathbb{P}$ dynamics as well.

[9] Gagliardini et al. (2011) also specify only the risk-neutral distribution, while drawing inference regarding the option pricing dynamics, exploiting the extended method of moments. Here, the major difference is that Gagliardini et al. (2011) develop the asymptotic distribution results along an entirely different dimension, as they rely on a long span of data, but only a small – and fixed – cross-section of option prices, assumed to be observed without error.

where $\alpha_t$ and $V_t$ are càdlàg; $W_t$ is a $\mathbb{P}^{(0)}$-Brownian motion; $\mu$ 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), defined over the fixed interval $[0,\ T]$, satisfies:*

*(i)* *For any $s, t \geq 0$, there exists a constant $K > 0$ such that* $\mathbb{E}\left\{|V_t - V_s|^2 \wedge K\right\} \leq K|t - s|$.

*(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 $\alpha_t$, $V_t$ and $a_t$ are locally bounded.*

Assumption A0 is quite weak. A0(i) is satisfied when the evolution of $V_t$ is given by a (multivariate) stochastic differential equation which is a common modeling approach for jump-diffusive asset returns displaying stochastic volatility. Assumption A0(ii) restricts the so-called Blumenthal-Getoor index of the jumps to be below $\beta$ and some of our results, such as Theorem 3 below, depend on the value of this coefficient. Finally, assumption A0(iii) implies that, at each point in time on $[0, T]$, the price process has a non-vanishing continuous martingale component. This assumption may be relaxed to require only that $\int_0^T V_t dt > 0$, but it is nevertheless very weak and satisfied for the models typically used in asset pricing. 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 standard no-arbitrage restrictions on the price process, see, e.g., 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 $\delta_t$ 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 that the diffusive volatility and the jump process 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; $\xi_1$ and $\xi_2$ are functions in $C^2$, and $\mathbf{S}_t$ denotes the $p \times 1$ state vector. Moreover, we assume that $r_t$ and $\delta_t$ are smooth functions of the state vector $\mathbf{S}_t$, and that the latter follows a jump-diffusive Markov process under $\mathbb{Q}$. This specification nests most of the continuous-time models used in empirical work, including, e.g., the affine jump-diffusion class of Duffie et al. (2000).

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,

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

where $\tau$ is the time-to-maturity, $K$ is the strike price, $F_{t,t+\tau}$ is the price of the futures contract on the underlying asset at time $t$ expiring at time $t+\tau$, and $k = \ln(K/F_{t,t+\tau})$ is 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 time-to-maturity, the state vector, and the 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 $[t, t+\tau]$. We denote the Black-Scholes implied volatility corresponding to the option price $O_{t,k,\tau}$ by $\kappa_{t,k,\tau}$. This merely represents an alternative, and convenient, pricing convention for the options, as the Black-Scholes implied volatility is a deterministic and strictly monotone transformation of the ratio $\frac{e^{r_{t,t+\tau}}O_{t,k,\tau}}{F_{t,t+\tau}}$.

### 3.2 The Parametric Option Pricing Framework

Henceforth, we assume a parametric model for the risk-neutral distribution, characterized by the $q \times 1$ parameter vector $\theta$, with $\theta_0$ signifying the true value. The option panel has a fixed time span, but contains a large cross-section spanning a significant range of $k$ and $\tau$ values. This is a natural assumption for active and liquid derivatives markets. Moreover, this section focuses on the ideal case of no errors in the observed option prices. The critical extension to the case involving such errors is provided in Section 4.2. The theoretical value of the Black-Scholes implied volatility under the risk-neutral model is denoted $\kappa(k, \tau, \mathbf{S}_t, \theta)$.[10] For each trading day $t$, we have a cross-section of option prices $\{O_{t,k_j,\tau_j}\}_{j=1,...,N_t}$ for some integer $N_t$, where the index $j$ runs across the full set of strike and time-to-maturity combinations available on day $t$. The number of options for the maturity $\tau$ is denoted by $N_t^\tau$ with $N_t = \sum_\tau N_t^\tau$. We invoke the following condition to formally capture the notion of a large, yet potentially heterogeneous, cross-section of options.

**Assumption A1.** *Fix $T > 0$. For each $t = 1, ..., T$ and each moneyness $\tau$, the number of options $N_t^\tau \uparrow \infty$ with $N_t^\tau / N_t \to \pi_t^\tau$ and $N_t / \sum_{t=1}^T N_t \to \varsigma_t$, for some positive numbers $\pi_t^\tau$ and $\varsigma_t$. Further, for each pair $(t, \tau)$, $\underline{k}(t,\tau)$ and $\overline{k}(t,\tau)$ denote the minimum and maximum log-moneyness, respectively, on day $t$ with maturity $\tau$ and the grid of moneyness of the options is given by $\underline{k}(t,\tau) = k_{t,\tau}(0) < k_{t,\tau}(1) ... < k_{t,\tau}(N_t^\tau) = \overline{k}(t,\tau)$. We assume that the sequence of moneyness grids is nested, and further if we denote $\Delta_{t,\tau}(i) = k_{t,\tau}(i) - k_{t,\tau}(i-1)$, then $N_t\,\Delta_{t,\tau}(i) \to \psi_{t,\tau}(k)$ uniformly on the interval $(\underline{k}(t,\tau),\ \overline{k}(t,\tau))$, where $\psi_{t,\tau}(\cdot)$ takes on finite and strictly positive values.*

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

Assumption A1 allows for a great deal of intertemporal heterogeneity in the observation scheme. For example, the times-to-maturity 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.

We supplement the large cross-section of options with an identification requirement:

**Assumption A2.** *For every $\epsilon > 0$ and $T > 0$ finite, we have*

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

*with $\theta \in \Theta$ for some compact set $\Theta$.*

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. 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 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 5, covering the empirically relevant case of noisy option price observations.

# 4 Inference for Option Panels with a Fixed Time Span

## 4.1 Insights from Error-Free Option Panels

In a frictionless market with continuous trading, no arbitrage opportunities and, consequently, no pricing errors, the observed option prices should perfectly match those generated by a correctly specified parametric model. This immediately suggests calibrating the risk-neutral parameters by matching the actual option prices to the model-implied prices, given as a function of the unknown parameter and state vectors, and then "inverting" the system to solve for the unknown parameters and state variables. In fact, since there are $q$ unknown $\mathbb{Q}$ parameters and the state vector is $p$-dimensional, we need – subject to suitable regularity conditions – only $p + q$ different option prices to infer the unknown parameters and states. The only issue is whether the inverse mapping from a

given value of model prices, expressed in Black-Scholes implied volatilities, $\kappa(k, \tau, \mathbf{S}_t, \theta)$, back into the parameter and state vector, $\theta$ and $\mathbf{S}_t$, is one-to-one.

In the error-free case, it is therefore, in principle, straightforward to identify the risk-neutral distribution and the state vector realization at any point in time. However, anticipating the empirically relevant case with observation error, it is clear that the quality of the inference will depend on the specific set of options used in the analysis. In particular, prices for options with similar strikes and maturities are highly correlated, so it is beneficial for identification to exploit options spanning a wide range of moneyness and tenors. Intuitively, options with diverse characteristics load very differently on the volatility and jump risks. Specifically, close-to-maturity, deep out-of-the-money (OTM) puts and calls load, respectively, mostly on the negative and positive jumps. Similarly, short and long term options load differently on volatility state variables which differ in their degree of persistence. Thus, the information content is also enhanced if the panel covers periods with a wide variation in the realization of the state vector.[11]

In summary, if we exploit a broad cross-section of options and manage to limit the impact of observation error on the inference, the qualitative insights from the ideal frictionless modeling framework regarding the informativeness of the option panel should carry over to the noisy setting. We conclude this section with a few remarks regarding our setup.

**Remark 1**. *In a setting where the time span $T$ of the option panel increases, one may exploit the time series of the recovered state vector, $\mathbf{S}_t$ to estimate, parametrically or nonparametrically, the associated $\mathbb{P}$ law. Hence, an option panel with increasing time span (and wide enough cross-section) suffices for estimation of both the $\mathbb{Q}$ and $\mathbb{P}$ measures, and thus also the risk premiums associated with dynamics of the state vector. There is, in principle, no need for direct observations on the underlying asset.*

**Remark 2**. *Our emphasis on the superior information content of option panels is related to the intuition behind the Recovery Theorem of Ross (2011). Under certain conditions regarding the asset price dynamics, he shows that the projection of the pricing kernel on the space of returns of the underlying asset can be recovered from observing the derivatives prices alone. In our setting, we are interested in the pricing kernel (respectively risk-neutral probability) on a wider filtration that incorporates multiple, possibly latent, state variables, but we "reduce the dimension of the kernel" by parameterizing it. In this setting, we also find that the derivatives prices encapsulate critical information for the pricing of the risks associated with the dynamics of the state variables.*

[11] Of course, this hinges on the stability of the risk-neutral distribution over time, but this assumption may be explicitly tested, as we demonstrate in Section 5.

**Remark 3**. *There are marked differences in the information content of the option panel (with a fixed time grid) versus the price path of the underlying asset. A continuous record of the X process lets us estimate the aggregate diffusive volatility, without error, from a local neighborhood of the current time in nonparametric and model-free fashion, and we can identify the timing and exact size of any price jumps. In contrast, the option data enable us to directly observe the state vector,* $\mathbf{S}_t$. *If the state vector consists of a single (diffusive) volatility factor,* $V_t$, *as is commonly 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 allows us to infer the possibly time-varying risk-neutral jump intensities along with the associated jump distributions, but they do not reveal actual jump realizations. On the contrary, the price path for X identifies the actual jumps, but it does not identify the jump distribution. Finally, if the model is generalized to a multi-factor volatility setting, as implied by an extensive body of empirical research, then the options data become even more valuable for inference. For example, suppose there are two volatility factors, so that* $V_t = V_{1,t} + V_{2,t}$. *The high-frequency data for X directly informs us about the aggregate value of* $V_t$ *only, while the option data enable us to separately identify the two components,* $V_{1,t}$ *and* $V_{2,t}$.

## 4.2 Estimation based on Noisy Option Panels

We continue next with the empirically relevant case of noisy observations. Figure 1 depicts a non-parametric kernel regression estimate of the relative bid-ask spread in the quotes for S&P 500 index options, in units of Black-Scholes implied volatility, as a function of the moneyness, normalized by volatility. The spread is non-trivial and increases quite sharply for deep OTM calls. However, even if the noise in any individual option price is significant, the impact may be mitigated by exploiting an extensive cross-section of option prices.[12]

In the remainder of this section, we develop inference procedures for the parameter vector, $\theta$, governing the risk-neutral distribution and the realized trajectory of the state vector $\{\mathbf{S}_t\}_{t=1,...,T}$ 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.

### 4.2.1 The Option Error Process

We stipulate that option prices, quoted in terms of Black-Scholes implied volatility, are observed with error, i.e., we observe $\widehat{\kappa}_{t,k,\tau}$,

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

[12] A similar perspective underlies the 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.

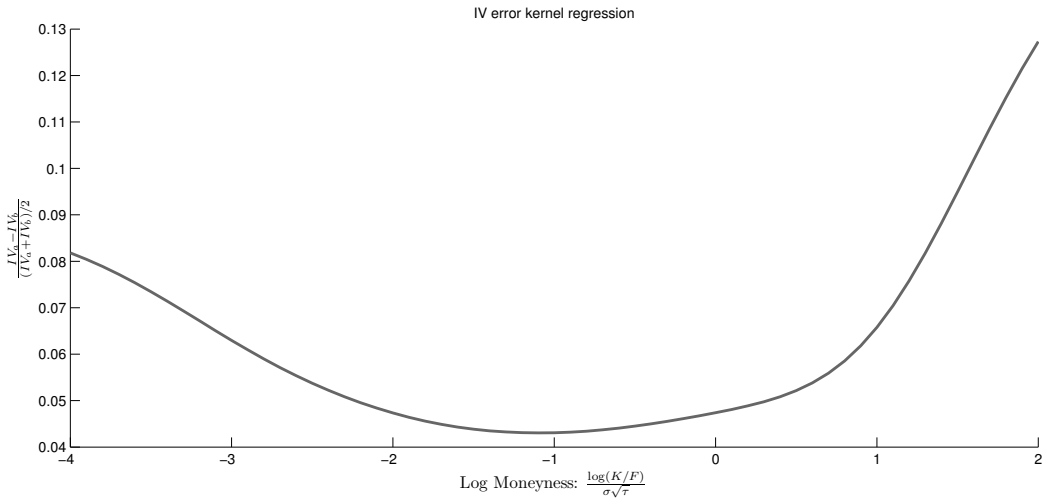


Figure 1: *Kernel regression estimate of the bid-ask spread of option implied volatility as a function of moneyness.* The estimates are based on the best bid and ask quotes for the S&P 500 options on the CBOE at the end-of-trading for each Wednesday during January 1, 1996 – July 21, 2010. We use all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

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 $\Gamma$ denoting the set of all possible tenors. $\Omega^{(1)}$ is equipped with the product Borel $\sigma$-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 $\mathcal{F}_t^{(1)} = \sigma\left(\epsilon_{s,k,\tau} : s \leq t\right)$. Then the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ 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 $\Omega$ as well, e.g., $W_t$ continues to be a Brownian motion on $\Omega$. We henceforth adopt this perspective without further mention.

Given the presence of observation error in the option prices, we cannot identify the parameters and state vector simply by inverting the option pricing formula, as discussed in Section 4.1. We must explicitly accommodate the impact of noise on the inference. In particular, if a limited set of options is included in the analysis, 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 will suffice for establishing consistency of our estimator.

**Assumption A3.** *For every* $\epsilon > 0$ *and* $T > 0$ *finite, we have*

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

*when* $\min_{t=1,\ldots,T} N_t \to \infty$ *for all* $\theta \in \Theta$.

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) and the references therein. Of course, for typical models of the asset price dynamics, the (stochastic) volatility process, and thus $\mathbf{S}_t$, has unbounded support. In this scenario one may establish the validity of Assumption 3 in one of two ways. First, one may apply a nonparametric estimator for $V_t$ based on high-frequency data on the underlying asset. For the vast majority of applications, the entire $\mathbf{S}_t$ vector determines $V_t$, and this will constrain $\mathbf{S}_t$ to reside within a compact set with probability approaching 1. Alternatively, whenever the state vector $\mathbf{S}_t$ increases sharply (towards infinity), the option price will, for all empirically relevant models, diverge as well, and this will drive the ratio in assumption A3 toward zero. Therefore, the asymptotic negligibility result in assumption A3 may be derived as a consequence of uniform convergence on a space of functions vanishing at infinity; see, e.g., Theorem 21 in Ibragimov and Has'minskii (1981).

We need additional regularity conditions to characterize the limiting distribution for the state and parameter vector estimates. We state those in the following.

**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 $\mathbf{S}_t$. Hence, it is considerably weaker than unconditional independence. 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.

**Remark 4**. *Assumption A4 is analogous to the conditions imposed on the microstructure noise process for asset prices observed at high frequencies in Jacod et al. (2009) and subsequent papers. We stress that part (i) is critical for the subsequent results, although it may be weakened by allowing for a bias which vanishes asymptotically. Part (iii) excludes correlation in the error across strikes, but we can accommodate week (spatial) dependence in the errors, at the cost of more complicated notation (and proof). On the other hand, if the option errors contain a common component across all strikes, then this error, obviously, cannot be "averaged out" by spatial integration in the moneyness dimension, as we do below in our estimation. For example, Bates (2000) assumes that option prices on a given day, for moneyness within prespecified ranges, may contain such a common error component. He interprets this as a model specification error while, in our setting, such features should be included in the theoretical value $\kappa(k, \tau, \mathbf{S}_t, \theta_0)$ rather than being treated as errors.*

**Remark 5**. *If one deems the unbiasedness property of the error to be more appropriate for the option price level instead of the Black-Scholes implied volatility – which constitutes a nonlinear transformation of the price – then one should instead minimize the distance between observed and model-implied option prices. For our empirical application, we find the implied volatilities to be approximately linear in prices across the range of moneyness we exploit, so the distinction between unbiasedness of the implied volatilities or the prices is not a practical concern; see, e.g., Christoffersen and Jacobs (2004) for a discussion of the impact of the error specification.*

### 4.2.2 Consistency

In order to more formally define our inference procedure, we first introduce an arbitrary consistent nonparametric 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 5). 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,

$$\left(\{\widehat{\mathbf{S}}_t^n\}_{t=1,\ldots,T},\, \widehat{\theta}^n\right) = \underset{\{\mathbf{Z}_t\}_{t=1,\ldots,T},\, \theta\in\Theta}{\operatorname{argmin}} \sum_{t=1}^{T} \left\{ \frac{1}{N_t} \sum_{j=1}^{N_t} \left(\widehat{\kappa}_{t,k,\tau} - \kappa(k_j, \tau_j, \mathbf{Z}_t, \theta)\right)^2 + \lambda_n \left(\widehat{V}_t^n - \xi_1(\mathbf{Z}_t)\right)^2 \right\}, \quad (4)$$

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 as a regularization device that helps identify the parameter vector by penalizing values that imply "unreasonable" volatility levels.

**Remark 6**. *The presence of the penalization term in (4) is reminiscent of the inclusion of information regarding the $\mathbb{P}$ dynamics of the state variables in the option-based estimation, e.g., Bates (2000) and Pan (2002). There is, however, a fundamental difference in our approach. We do not model the $\mathbb{P}$ dynamics and the penalization in (4) concerns the pathwise behavior of the option surface, not its $\mathbb{P}$ law. This is therefore, a more robust (we do not parameterize the $\mathbb{P}$ law) and stronger (it is pathwise) restriction on the option dynamics.*

The procedure in (4) is akin to nonlinear least squares (NLS) estimation but, unlike the usual case in econometrics, the conditional mean equation changes across the option observations indexed by the triple $(t, k, \tau)$. The consistency of $\left(\widehat{\mathbf{S}}_t^n, \widehat{\theta}_t^n\right)$ follows from the next theorem.

**Theorem 1** *Suppose assumptions A1-A3 hold for some $T \in \mathbb{N}$ fixed and that $\{\widehat{V}_t^n\}_{t=1,...,T}$ is consistent for $\{V_t\}_{t=1,...,T}$ as $n \to \infty$. Then, if $\min_{t=1,...,T} N_t \to \infty$ 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,*

$$||\widehat{\mathbf{S}}_t^n - \mathbf{S}_t|| \xrightarrow{\mathbb{P}} 0, \quad ||\widehat{\theta}^n - \theta_0|| \xrightarrow{\mathbb{P}} 0, \quad t = 1, ..., T. \tag{5}$$

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.

### 4.2.3 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 is the case in our setting. For formal analysis of this concept, see, e.g., Jacod and Shiryaev (2003) and the references therein. 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 5.

**Theorem 2** *Assume A1-A4 are satisfied for $T \in \mathbb{N}$ fixed and $\kappa(t,\tau,\mathbf{Z},\theta)$ is twice continuously-differentiable in its arguments. Then, if $\min_{t=1,...,T} N_t \to \infty$ and $\lambda_n^2 \min_{t=1,...,T} N_t \to 0$, for $n \to \infty$, we have:*

$$\begin{pmatrix} \sqrt{N_1}(\widehat{\mathbf{S}}_1^n - \mathbf{S}_1) \\ \vdots \\ \sqrt{N_T}(\widehat{\mathbf{S}}_T^n - \mathbf{S}_T) \\ \sqrt{\frac{N_1+...+N_T}{T}}(\widehat{\theta}^n - \theta_0) \end{pmatrix} \xrightarrow{\mathcal{L}-s} \mathbf{H}_T^{-1} \left(\mathbf{\Omega}_T\right)^{1/2} \begin{pmatrix} \mathbf{E}_1 \\ \vdots \\ \mathbf{E}_T \\ \mathbf{E}' \end{pmatrix}, \tag{6}$$

*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 being i.i.d. with standard normal distribution, and we define*

$$\mathbf{\Phi} = \begin{pmatrix} \mathbf{\Phi}_T^{1,1} & \dots & \mathbf{0}_{p\times p} & \mathbf{\Phi}_T^{1,T+1} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{p\times p} & \dots & \mathbf{\Phi}_T^{T,T} & \mathbf{\Phi}_T^{T,T+1} \\ \mathbf{\Phi}_T^{T+1,1} & \dots & \mathbf{\Phi}_T^{T+1,T} & \mathbf{\Phi}_T^{T+1,T+1} \end{pmatrix}, \quad \mathbf{\Phi} = \mathbf{H},\ \mathbf{\Omega}, \tag{7}$$

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

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

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

*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 (7), we set*

$$\begin{cases} \widehat{\mathbf{H}}_T^{t,t} = \frac{1}{N_t} \sum_{j=1}^{N_t} \nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t,\widehat{\theta}) \nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t,\widehat{\theta})', \\ \widehat{\mathbf{H}}_T^{T+1,T+1} = \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \nabla_\theta\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t,\widehat{\theta}) \nabla_\theta\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t,\widehat{\theta})', \\ \widehat{\mathbf{H}}_T^{t,T+1} = \left(\widehat{\mathbf{H}}_T^{T+1,t}\right)' = \frac{1}{N_t} \sum_{j=1}^{N_t} \nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t,\widehat{\theta}) \nabla_\theta\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t,\widehat{\theta})', \end{cases} \tag{8}$$

$$\begin{cases} \widehat{\mathbf{\Omega}}_T^{t,t} = \frac{1}{N_t}\sum_{j=1}^{N_t}\left(\widehat{\kappa}_j - \kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t,\widehat{\theta})\right)^2 \nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t,\widehat{\theta})\,\nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t,\widehat{\theta})', \\ \widehat{\mathbf{\Omega}}_T^{T+1,T+1} = (\frac{\sum_{t=1}^T N_t}{T})\sum_{t=1}^T \frac{1}{N_t^2}\sum_{j=1}^{N_t}\left(\widehat{\kappa}_j - \kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t,\widehat{\theta})\right)^2 \nabla_{\theta}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t,\widehat{\theta})\,\nabla_{\theta}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t,\widehat{\theta})', \\ \widehat{\mathbf{\Omega}}_T^{t,T+1} = \left(\widehat{\mathbf{\Omega}}_T^{T+1,t}\right)' = \sqrt{\frac{\sum_{t=1}^T N_t}{TN_t^3}}\sum_{j=1}^{N_t}\left(\widehat{\kappa}_j - \kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t,\widehat{\theta})\right)^2 \nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t,\widehat{\theta})\,\nabla_{\theta}\kappa(k_j,\tau_j,\widehat{\mathbf{S}}_t,\widehat{\theta})'. \end{cases} \tag{9}$$

Several comments are in order. First, we reiterate that the limit result in (6) holds stably conditional on the filtration of the original probability space. The limit is mixed-Gaussian, with a mixing variable, $\mathbf{H}_T^{-1}\left(\mathbf{\Omega}_T\right)^{1/2}$, that is adapted to $\mathcal{F}_T$.[13] The random asymptotic variance of the estimator implies that the precision in recovering the state vector and the risk-neutral parameters varies from period to period, depending on the level of the state variables and underlying asset prices as well as the number and characteristics of the options used for estimation (maturity and moneyness). This provides important flexibility as the features of the option data do 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 also 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 of the results established above mimic features of the limiting distributional theory for volatility estimators based on high-frequency data, see, e.g., Barndorff-Nielsen et al. (2006).

It is straightforward to implement feasible inference using the estimates in equations (8)-(9). We may obtain the requisite consistent estimate of the asymptotic variance for $\left(\{\widehat{\mathbf{S}}_t^n\}_{t=1,\ldots,T},\widehat{\theta}^n\right)$ from a consistent estimator of the option error, $\epsilon_{t,k,\tau}$. Moreover, based on equations (8) and (9), pivotal tests such as $t$-tests for the parameters are readily constructed. This is a by-product of the stable convergence in equation (6), 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 heteroscedasticity in the option price error. In analogy with the standard NLS, when the observation errors are conditionally homoscedastic, i.e., $\mathbb{E}(\epsilon_{t,j,k}^2|\mathcal{F}^{(0)}) = \phi$, for $\phi$ non-random and constant, then $\Omega_T = \phi H_T$ (recall $\Omega_T$ and $H_T$ are $\mathcal{F}_T^{(0)}$-adapted random variables) and the estimator is asymptotically efficient. More generally, when conditional homoskedasticity is violated, we may restore efficiency by weighting the option squared errors in (4) appropriately over the space $(t,k,\tau)$. The weights depend on a consistent estimate for the dependence of the conditional variance of $\epsilon_{t,k,\tau}$ on $k$ (parametric or nonparametric), for each pair $(t,\tau)$, in a manner similar to the standard weighted least squares estimators, see, e.g.,

[13] 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.

Robinson (1987) and Newey and McFadden (1994).

**Remark 7**. *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 are i.i.d., but conditional only on an $\mathcal{F}_0$-adapted random variable. Within that setting, stronger result may be derived compared to our stable convergence. 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 obtained for a much wider $\sigma$-field than (a subfield of) $\mathcal{F}_0$ – as in Andrews (2005) – enabling feasible inference from the observed options. Further, in Theorem 3 of the next section, we show that the stable convergence of Theorem 2 holds jointly with that of a high-frequency estimator for spot volatility. The general stable convergence result in Theorem 2 is crucial for establishing this property.*

**Remark 8**. *The assumption, $\lambda_n^2 \min_{t=1,\ldots,T} N_t \to 0$, in Theorem 2 ensures that the penalty term in (4) has no first-order asymptotic effect in the estimation. We can readily extend the analysis to cover scenarios in which the penalty term is reflected in the limiting distribution. The requirement is that 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 (6) – is known. This is satisfied for the nonparametric jump-robust realized volatility estimators we invoke in the following section.*

# 5 Pathwise Risk-Neutral Model Tests

As noted in Section 2, the model for the risk-neutral dynamics has numerous testable implications. The previous section provides the necessary limit theory to develop formal tests for these restrictions. We propose a battery of diagnostics, grouped into three separate categories: the first gauges the quality of the model fit to the option surface, the second checks for stability of the risk-neutral parameters over time, and the third assesses the equality between the option-implied volatility and a nonparametric volatility estimate based on high-frequency data. These tests are all pathwise in nature as they involve restrictions on the *observed paths* of the option surface and the underlying asset prices. Importantly, they do not require any assumptions regarding the statistical law of the process $X$, beyond what is implied by the risk-neutral law. As such, they do not rely on a joint hypothesis of correctly specified models under both the $\mathbb{P}$ and $\mathbb{Q}$ measures.

### 5.1 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 the observation error in option prices from the model misspecification error in fitting 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 maturity.

**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 time to maturity $\tau^*$ and log-moneyness belonging to the set $\mathcal{K}$. Then, under the assumptions of Theorem 2, we have,*

$$\frac{\sum_{j:k_j\in\mathcal{K}}\left(\widehat{\kappa}_{t,k_j,\tau^*}-\kappa(k_j,\tau^*,\widehat{\mathbf{S}}_t^n,\widehat{\theta}^n)\right)}{\sqrt{\widehat{\Pi}_T'\widehat{\Xi}_T\widehat{\Pi}_T}} \xrightarrow{\mathcal{L}-s} \mathcal{N}(0,1), \quad \widehat{\Xi}_T=\begin{pmatrix}\widehat{\mathbf{H}}_T^{-1}\widehat{\mathbf{\Omega}}_T(\widehat{\mathbf{H}}_T^{-1})' & \widehat{\mathbf{H}}_T^{-1}\widehat{\mathbf{\Upsilon}}_{1,T}\\ \widehat{\mathbf{\Upsilon}}_{1,T}'(\widehat{\mathbf{H}}_T^{-1})' & \widehat{\mathbf{\Upsilon}}_{2,T}\end{pmatrix}, \quad (10)$$

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

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

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

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 (10) 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 (10) resemble the conditional moment tests proposed by Newey (1985) and Tauchen (1985).

The asymptotic variance of the option fit $\sum_{j:k_j\in\mathcal{K}}\left(\widehat{\kappa}_{t,k_j,\tau^*}-\kappa(k_j,\tau^*,\widehat{\mathbf{S}}_t^n,\widehat{\theta}^n)\right)$, 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 (6) holds stably so that the standardization of the model fit in equation (10) yields a variable with a limiting standard normal distribution. The test in equation (10) 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 (10). 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.

### 5.2 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.[14]

**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,*

$$\left(\widehat{\theta}_1^n - \widehat{\theta}_2^n\right)' \left(\widehat{Avar}(\widehat{\theta}_1^n) + \widehat{Avar}(\widehat{\theta}_2^n)\right)^{-1} \left(\widehat{\theta}_1^n - \widehat{\theta}_2^n\right) \xrightarrow{\mathcal{L}-s} \chi^2(q), \tag{11}$$

*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 (8)-(9) in Theorem 2, 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 (11).*

[14] 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.

### 5.3 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 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 (4).

To render the test feasible, we need an estimate of $V_t$. We exploit two high-frequency based nonparametric jump-robust estimators, defined as follows,

$$\widehat{V}_t^{\pm,n} = \frac{n}{k_n} \sum_{i \in I^{\pm,n}} (\Delta_i^{t,n} X)^2 \, 1\left(|\Delta_i^{t,n} X| \leq \alpha n^{-\varpi}\right), \quad \Delta_i^{t,n} X = \log\left(X_{t+\frac{i}{n}}\right) - \log\left(X_{t+\frac{i-1}{n}}\right), \tag{12}$$

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 variance from the left and right, respectively. If we denote the set of jump times for the variance process by $\mathbf{J} = \{s : \Delta V_s > 0\}$, then, under regularity conditions weaker than those in Assumption A0, we can show that both $V_t^{+,n}$ and $V_t^{-,n}$ are consistent for $V_t$, provided $t \notin \mathbf{J}$. We only need to estimate the spot volatility on a finite set of times. Since the jump compensator controlling the jumps in $V_t$ is absolutely continuous in time, the probability of having jumps at any of these discrete times is zero, as, almost surely, $t \notin \mathbf{J}$.

The theorem below provides the joint limit distribution of $\widehat{V}_t^{\pm,n}$ and the option-based $\widehat{\mathbf{S}}_t^n$.

**Theorem 3** *Under assumption A0, provided* $k_n \to \infty$ *with* $\frac{\sqrt{k_n}}{n^{(2-\beta)\varpi \bigwedge [(\frac{1}{\beta} \wedge 1) - 1/2 + (1-\beta)\varpi \vee 0] \bigwedge 1/4}} \to 0$, *and with* $\beta$ *defined as in A0(iii), we have for* $T \in \mathbb{N}$,

$$\sqrt{k_n} \begin{pmatrix} \widehat{V}_1^{+,n} - V_1 \\ \vdots \\ \widehat{V}_T^{+,n} - V_T \end{pmatrix} \xrightarrow{\mathcal{L}-s} \begin{pmatrix} \sqrt{2}V_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sqrt{2}V_T \end{pmatrix} \begin{pmatrix} \widetilde{E}_1 \\ \vdots \\ \widetilde{E}_T \end{pmatrix}, \tag{13}$$

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

*Theorem 2 are satisfied, then 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 (6).*

*If further* $(1, ..., T) \cap \mathbf{J} = \emptyset$*, then the above results continue to hold when* $\widehat{V}_t^{+,n}$ *is replaced by* $\widehat{V}_t^{-,n}$ *for* $t = 1, ..., T$*.*

The asymptotic distribution of $\widehat{V}_t^{\pm,n}$ on the random set of jump times in $X$ is derived in Jacod and Todorov (2010). It is optimal to choose $\varpi$ close to 1/2 and, next, let $k_n$ be close to $n^{1/2}$, provided the jumps are not too active, i.e., their activity index satisfies $\beta < 4/3$ (recall assumption A0(ii)). This is a relatively mild restriction.[15] 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 when conditioned on the filtration of the original probability space. This allows us to derive the asymptotic behavior of the difference $\widehat{V}_t^{\pm,n} - \xi_1(\widehat{\mathbf{S}}_t^n)$. We state this important result as a corollary to Theorem 3.

**Corollary 3** *Under the same conditions as in Theorem 3, we have for* $k_n \to \infty$*,* $\min_{t=1,...,T} N_t \to \infty$ *and* $\lambda_n^2 \min_{t=1,...,T} N_t \to 0$*,*

$$\left\{ \frac{\xi_1(\widehat{\mathbf{S}}_t^n) - \widehat{V}_t^{+,n}}{\sqrt{\frac{\nabla_{\mathbf{S}}\xi_1(\widehat{\mathbf{S}}_t^n)'\widehat{\chi}_t\nabla_{\mathbf{S}}\xi_1(\widehat{\mathbf{S}}_t^n)}{N_t} + \frac{2(\widehat{V}_t^{+,n})^2}{k_n}}} \right\}_{t=1,...,T} \xrightarrow{\mathcal{L}-s} \begin{pmatrix} \breve{E}_1 \\ \vdots \\ \breve{E}_T \end{pmatrix}, \tag{14}$$

*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 3. 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.

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

[15] The relative speed condition between $k_n$ and $n$ in Theorem 3 can be slightly weakened in the case $\beta \geq 1$ at the cost of more lengthy derivations. The improvement is relatively small, so we abstain from this generalization here.

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.[16]

Finally, we emphasize that, to increase the power to detect model misspecification, the three tests in Corollary 1-3 should be applied in parallel. For example, a misspecified model might generate implied spot volatility estimates that are close to the model-free ones, but in doing so it will provide a poor fit to the observed option prices or induce parameter instability. On the other hand, a faulty model may occasionally provide a good fit for the option panel – the vector $\{\mathbf{S}_t\}_{t=1,...,T}$ provides flexibility in fitting option prices – but this will typically be at the expense of implausible volatility estimates.

# 6 Numerical Experiments

## 6.1 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,

$$\frac{dX_t}{X_{t-}} = \sqrt{V_t}dW_t + dL_{x,t}\,, \qquad dV_t = \kappa_d(\overline{v} - V_t)dt + \sigma_d\sqrt{V_t}dB_t + dL_{v,t}\,, \tag{15}$$

where $(W_t, B_t)$ is a two-dimensional Brownian motion with correlation $\text{corr}\,(B_t, W_t) = \rho_d \cdot t$; $(L_{x,t}, L_{v,t})$ is a compound Poisson jumps process with intensity $\lambda_j$ 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 $\mu_v$, while, conditional on $Z_v$, $\log(Z_x + 1)$ is Gaussian with mean $\mu_x + \rho_j Z_v$ and standard deviation $\sigma_x$, and, finally, $L_{x,t}$ is a jump martingale. The model also involves the cross-parameter restriction, $\sigma_d \leq \sqrt{2\kappa_d\overline{v}}$. 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

[16]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.

adopt the standard approach of the empirical option pricing literature, see, e.g., Singleton (2006), 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 $\rho_j = 0$, thus 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 |
| $\rho_d$ | $-0.4600$ | $\lambda_j$ | 1.0080 | $\rho_d$ | $-0.4600$ | $\lambda_j$ | 1.0080 |
| $\overline{v}$ | 0.0144 | $\mu_x$ | $-0.0284$ | $\overline{v}$ | 0.0144 | $\mu_x$ | $-0.0501$ |
| $\kappa_d$ | 4.0320 | $\sigma_x$ | 0.0490 | $\kappa_d$ | 4.0320 | $\sigma_x$ | 0.0751 |
| $\sigma_d$ | 0.2000 | $\mu_v$ | 0.0315 | $\sigma_d$ | 0.2000 | $\mu_v$ | 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 as the short maturity options are determined largely by the current volatility state. The same logic applies to the identification of the mean reversion parameter, $\kappa_d$. Not surprisingly, the sensitivity of options with respect to $\kappa_d$ 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, $\mu_x$ 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, $\sigma_x$, 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.

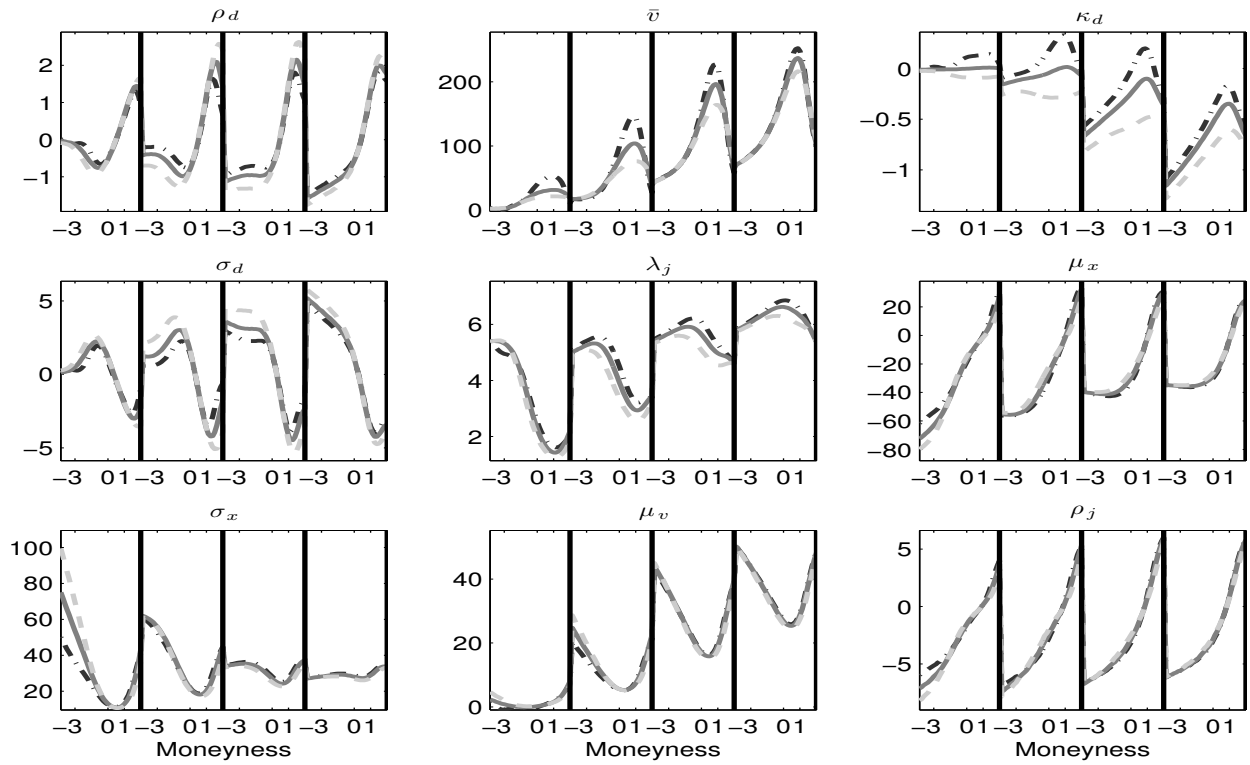


Figure 2: *Option Sensitivity to Parameters in Double-Jump Model.* The figure plots the first derivatives of options (in terms of implied volatility) with respect to the parameters of the double-jump model. The parameters are set at the values reported in Table 1. Moneyness is reported in terms of volatility 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.

### 6.2 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.[17]

The Monte Carlo setting is aimed at broadly mimicking the features of the models that have been estimated in prior work as well as the data used in the subsequent empirical application.

[17]The computational burden associated with this study is highly non-trivial, but a variety of improvements over the traditional approach to the computation of the option prices within this framework has rendered the exercise feasible. To the best of our knowledge, this constitutes the first comprehensive Monte Carlo study of the efficacy of inference procedures for the double-jump model at this level of relative complexity. An account of our computational approach may be found in a supplementary web-appendix.

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).[18] 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 $\sigma$ is the at-the-money Black-Scholes implied volatility 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. 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 $\psi_k$ 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 both on the current level of volatility and the moneyness. Finally, we set $\lambda_n = 0$ in (4) and, for the computation of 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. $\alpha$ and $\varpi$ were calculated as in Bollerslev and Todorov (2011) and we refer to that source 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 $\rho_d$ and $\sigma_d$. 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 5. Generally, the small sample behavior is quite satisfactory. The tests for the fit to the 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.

[18] More precise inference would be garnered from using a longer sample, but the one-year horizon provides a sensible compromise, in practice, between the joint objectives of accuracy in estimation and minimization of the time 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 |
|---|---|---|---|---|---|---|---|
| $\rho_d$ | $-0.4600$ | $-0.4564$ | 0.3911 | $\lambda_j$ | 1.0080 | 1.0080 | 0.1798 |
| $\overline{v}$ | 0.0144 | 0.0144 | 0.0035 | $\mu_x$ | $-0.0501$ | $-0.0501$ | 0.0134 |
| $\kappa_d$ | 4.0320 | 4.0321 | 0.1939 | $\sigma_x$ | 0.0751 | 0.0751 | 0.0057 |
| $\sigma_d$ | 0.2000 | 0.2022 | 0.1240 | $\mu_v$ | 0.0935 | 0.0935 | 0.0055 |

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 to 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., $\rho_d$ and $\sigma_d$, 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%** |
| $\rho_d$ | 0.00% | 0.42% | 1.90% | $\lambda_j$ | 1.90% | 6.96% | 10.97% |
| $\overline{v}$ | 1.48% | 6.96% | 10.97% | $\mu_x$ | 0.63% | 6.33% | 11.39% |
| $\kappa_d$ | 1.48% | 6.96% | 12.66% | $\sigma_x$ | 0.84% | 5.06% | 8.65% |
| $\sigma_d$ | 0.63% | 4.43% | 8.86% | $\mu_v$ | 0.63% | 5.70% | 9.49% |

*Note:* The parameter stability test is given in equation (11).

# 7 Empirical Application

We apply our inference procedure to data on the European style S&P 500 equity-index (SPX) options traded at the CBOE. We use the closing bid and ask prices reported by OptionMetrics.[19] We apply standard filters and discard all in-the-money options, options with time-to-maturity of less than 7 days, as well as options with zero bid prices. We then compute the mid bid-ask Black-Scholes implied volatilities. The data spans the period January 1, 1996, till 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 in the estimation. The nonparametric estimate of volatility 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 time-span of the options data. The construction of the high-frequency estimate follows the exact steps outlined in describing the Monte Carlo study. In the estimation we set $\lambda_n = 0.2$.

We estimate the double-jump model in (15) as well as an extension that allows for the jumps to have time-varying intensity and the stochastic volatility to contain two independent factors. The extended model (under the risk-neutral measure) is given by,

$$\begin{aligned} &\frac{dX_t}{X_{t-}} = (r_t - \delta_t)dt + \sqrt{V_{1,t}}dW_{1,t} + \sqrt{V_{2,t}}dW_{2,t} + dL_{x,t}, \\ &dV_{1,t} = \kappa_{d,1}(\overline{v}_1 - V_{1,t})dt + \sigma_{d,1}\sqrt{V_{1,t}}dB_{1,t} + dL_{v,t}, \quad dV_{2,t} = \kappa_{d,2}(\overline{v}_2 - V_{2,t})dt + \sigma_{d,2}\sqrt{V_{2,t}}dB_{2,t}, \end{aligned} \tag{16}$$

where $(W_{1,t}, W_{2,t}, B_{1,t}, B_{2,t})$ is a four-dimensional Brownian motion with $W_{1,t} \perp W_{2,t}$, $W_{1,t} \perp B_{2,t}$ and $\text{corr}\,(W_{1,t}, B_{1,t}) = \rho_{d,1} \cdot t$ as well as $W_{2,t} \perp B_{1,t}$ and $\text{corr}\,(W_{2,t}, B_{2,t}) = \rho_{d,2} \cdot t$. 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) distribu-

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

tion $(Z_1,\ Z_2)$, where the marginal distribution of $Z_2$ is exponential with mean $\mu_v$ and, conditional on $Z_2$, $\log(Z_1+1)$ is normal with mean $\mu_x+\rho_j Z_2$, and variance $\sigma_x^2$. For simplicity, we assume that $r_t$ and $\delta_t$ are deterministic, where the LIBOR rate for the corresponding maturity represents the interest rate while the dividend yield is obtained from OptionMetrics.

From model (16) we obtain the original one-factor double-jump model if we impose $\lambda_{j,1}=0$ as well as $\sigma_{d,2}=\overline{v}_2=0$. We first present results based on this simpler benchmark model. It is a mildly generalized version of the Broadie et al. (2009) setting used for our numerical experiments, as we allow $\rho_j\neq 0$. The parameter estimates and associated standard errors are reported in Table 5.

Table 5: Parameter Estimates for the One-Factor Model

| Parameter | Estimate | Standard Error | Parameter | Estimate | Standard Error |
|---|---|---|---|---|---|
| $\rho_d$ | $-0.9586$ | 0.0059 | $\lambda_j$ | 0.0150 | 0.0006 |
| $\overline{v}$ | 0.0356 | 0.0004 | $\mu_x$ | 0.5833 | 0.0417 |
| $\kappa_d$ | 1.4350 | 0.0198 | $\sigma_x$ | 0.3220 | 0.0430 |
| $\sigma_d$ | 0.2965 | 0.0032 | $\mu_v$ | 1.6162 | 0.1912 |
| | | | $\rho_j$ | $-0.8579$ | 0.1572 |

*Note:* Parameter estimates of the one-factor model (15) for S&P 500 equity-index option data sampled every Wednesday over the period January 1996-July 2010.

A few comments are in order. Focusing first on the broader features, the risk-neutral mean of volatility implied by the point estimates in Table 5 equals 24.3% (quoted in annualized volatility), while the sample mean of volatility under the statistical distribution, obtained by cumulating the intraday five-minute and overnight S&P 500 futures returns, is 21.35%. This is consistent with the presence of a (negative) volatility risk premium, as also established in much prior work. Moreover, the contribution of jumps to the overall (risk-neutral) volatility is non-trivial. Second, the point estimates deviate markedly from those commonly reported in the literature, as exemplified by Table 1. For example, the (negative) correlation between the Gaussian shocks in volatility and price is close to $-1$, while the implied mean price jump is huge, around $-80\%$, and with high volatility, but occurring with a frequency of only 1.5% per year. Similarly, the mean of the volatility jump is extremely large, at an imposing 162%. In summary, our parameters are at odds with studies that estimates the risk-neutral parameters from time series observations on the underlying asset returns along with a limited number of options. This is likely due to the relatively large weight assigned to the asset returns in the estimation as well as the tight link imposed on the $\mathbb{P}-\mathbb{Q}$ relation. Indeed, our findings are consistent with studies that calibrate models exclusively to the option surface; see, e.g., Christoffersen et al. (2009). In essence, the large negative jumps accommodate the strong

negative skew over short maturities, while the strong negative correlation between diffusive return and volatility shocks help rationalize the negative skew for long maturity options. Finally, the negative association between return and volatility jumps further accentuate the skewness.

Figure 3 depicts the weekly series of Z-scores from Corollary 1. The model faces difficulties in several dimensions. First, not surprisingly, it struggles to fit options over the crisis period, 2008-2010. Second, the fit to deep OTM short-maturity options is relatively good with the exception of the crisis and the tranquil period of 2003-2007. Third, the fit to short-maturity OTM calls is poor throughout, as they undergo episodes of persistent over- or under-valuation. Fourth, long maturity options are fit poorly with alternating periods of over- and undervaluation. These problems are also evident from Table 7 which reports the rejection rates from the diagnostic tests.

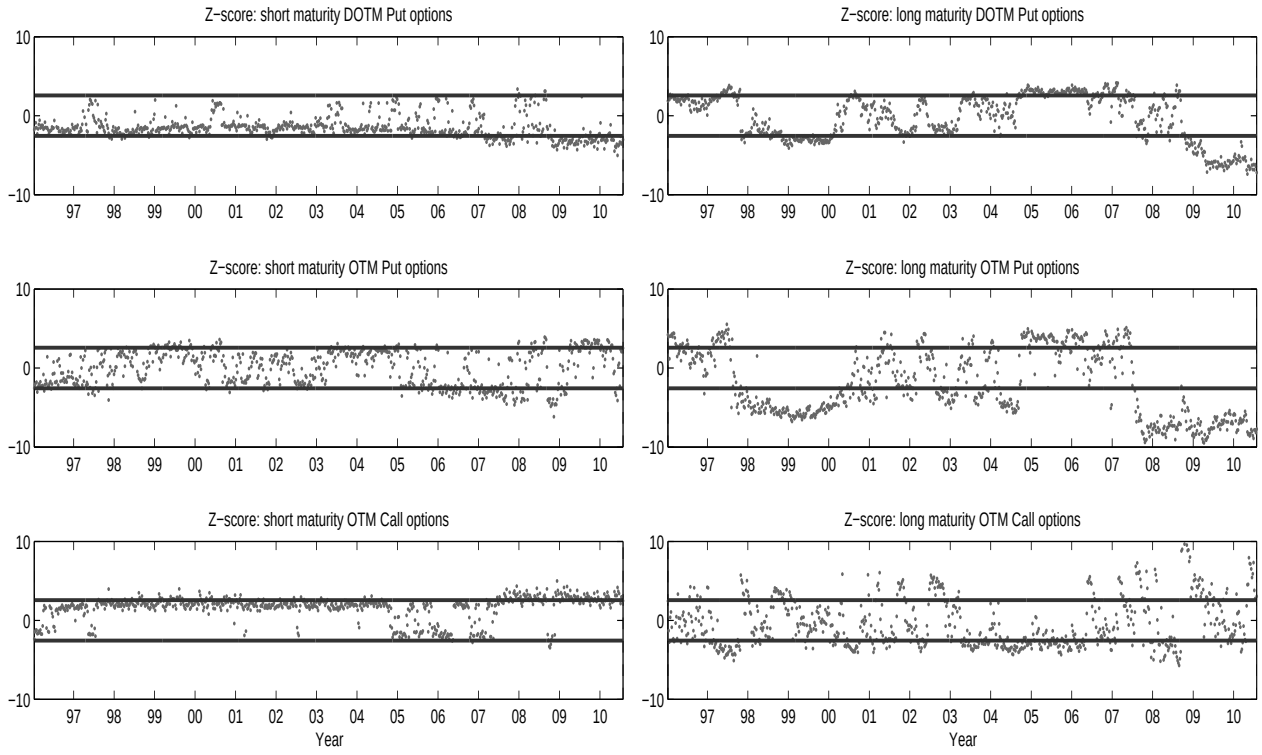


Figure 3: *Option Panel Fit for the One-factor Model.* The short-maturity options are those with the two shortest time-to-maturity available on a given day and the long-maturity options are all remaining options with time-to-maturity less than one year. DOTM signifies deep out-of-the-money and OTM is out-of-the-money. DOTM puts are those with moneyness $[-4,-2]\times\sigma\sqrt{\tau}$, OTM puts are those with moneyness $[-2,0]\times\sigma\sqrt{\tau}$, and OTM calls are those with moneyness $[0,1]\times\sigma\sqrt{\tau}$, where $\sigma$ is the at-the-money Black-Scholes implied volatility on the given day.

Figure 3 and Table 7 suggest that the one-factor model has several shortcomings. First, the

constant jump intensity prevents a steepening of the smirk (the slope of implied volatility as a function of moneyness) in response to high volatility. Second, the one-factor volatility structure generates too much dependence between short and long term options. Third, the shifts in over- and undervaluation of OTM calls is indicative of a misspecified jump distribution, likely reflecting the dearth of smaller (positive) jumps. This conjecture is consistent with the nonparametric evidence in Bollerslev and Todorov (2011). Inspired by these observations, we now estimate the two-factor model in (16). It contains features that should help mitigate the first two deficiencies of the original model. Note also 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 6.

Table 6: 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 | $\mu_x$ | $-0.0145$ | 0.0136 |
| $\sigma_{d,1}$ | 0.2640 | 0.0113 | $\sigma_x$ | 0.0877 | 0.0082 |
| $\rho_{d,2}$ | $-0.1824$ | 0.0388 | $\mu_v$ | 0.1501 | 0.0124 |
| $\overline{v}_2$ | 0.0066 | 0.0001 | $\rho_j$ | $-0.7756$ | 0.0718 |
| $\kappa_{d,2}$ | 29.8797 | 0.5951 | | | |
| $\sigma_{d,2}$ | 0.2341 | 0.0569 | | | |

*Note:* Parameter estimates for the two-factor model in (16), obtained using S&P 500 equity-index options sampled every Wednesday over the period January 1996-July 2010.

The parameter estimates for the diffusive part of the first volatility factor are similar to those for the one-factor model in Table 5. The second volatility factor has a much lower mean and much shorter half-life. Interestingly, once we free up the $\lambda_{j,1} = 0$ constraint, 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). Moreover, the mean jump intensity is now 28.4%, i.e., a jump occurs about every 3-4 years, and the mean jump size is approximately $-13\%$. Likewise, the mean volatility jump is a more palatable 15%. These jump estimates are now roughly in line with those from studies exploiting both asset returns and options, see, e.g, Singleton (2006), chapter 15, for a review. As before, the correlation between price and volatility jumps is strongly negative. This is consistent with the nonparametric evidence in Todorov and Tauchen (2011) and Andersen et al. (2011). The

fact that $\lambda_{j,1}$ and $\rho_j$ 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 increase the future level of volatility. Finally, the estimated mean volatility is now 22.8%, which remains consistent with a negative volatility risk premium.

Figure 4 depicts the weekly Z-scores for the fit to the volatility surface afforded by the two-factor model. There are still nontrivial rejections of the model over the sample period. In particular, the out-of-the-money call options continue to be mispriced, but there appears to be less persistent shifts in the Z-scores for the short-maturity out-of-the-money calls. The period 2008-2010 continues to be a significant challenge and we find systematic overpricing of the out-of-the-money short-maturity puts over the tranquil period prior to the Fall of 2007. Similarly, the persistence of the mispricing appears to be somewhat lower for the long-maturity options than was the case for the one-factor model. Nonetheless, it is hard to formally gauge the relative quality of fit by ocular inspection, so we turn to the diagnostic tests provided in Table 7. Panel A suggests that the extension to a two-factor model allows for increased flexibility in fitting the long term put options and the short term call options, but only by doing somewhat worse on the short term put and long term call options. Likely factors in the improved fit are the more persistent first volatility factor, the increased probability of smaller negative jumps and an even more negative correlation between returns and volatility for the first factor. All these features enhance the (risk-neutral) probability of a large negative move over long horizons. At the same time, the ability to fit the long term call options deteriorates, and the lack of extremely large negative jumps render the short term put options harder to price. The presence of the second much less persistent and less volatile factor also helps with the shorter term calls, but it cannot offset the long term return-volatility asymmetry that renders the long term calls hard to price. Overall, the two-factor model improves the average fit, measured in terms of the root-mean-squared-error, by close to 25%.

The more intuitive and reasonable parameter values and the better average fit to the option prices noted in Figure 4 as well as Panels A and B of Table 7 suggest that the two-factor model provides a non-trivial improvement. However, especially the worsening fit for the longer maturity out-of-the-money calls indicates that there is scope for significant improvement in future work, mainly via the specification of a more flexible jump distribution that allows for more frequent small, positive as well as negative, jumps and a more gradual decay of the jump tails.

It is natural to associate some of the problems for the one-factor model with an inability to flexibly accommodate shifts in the term structure of volatility, see e.g., Bates (2000) and Christoffersen et al. (2009). Figure 5 displays the fit for at-the-money implied volatilities over short and

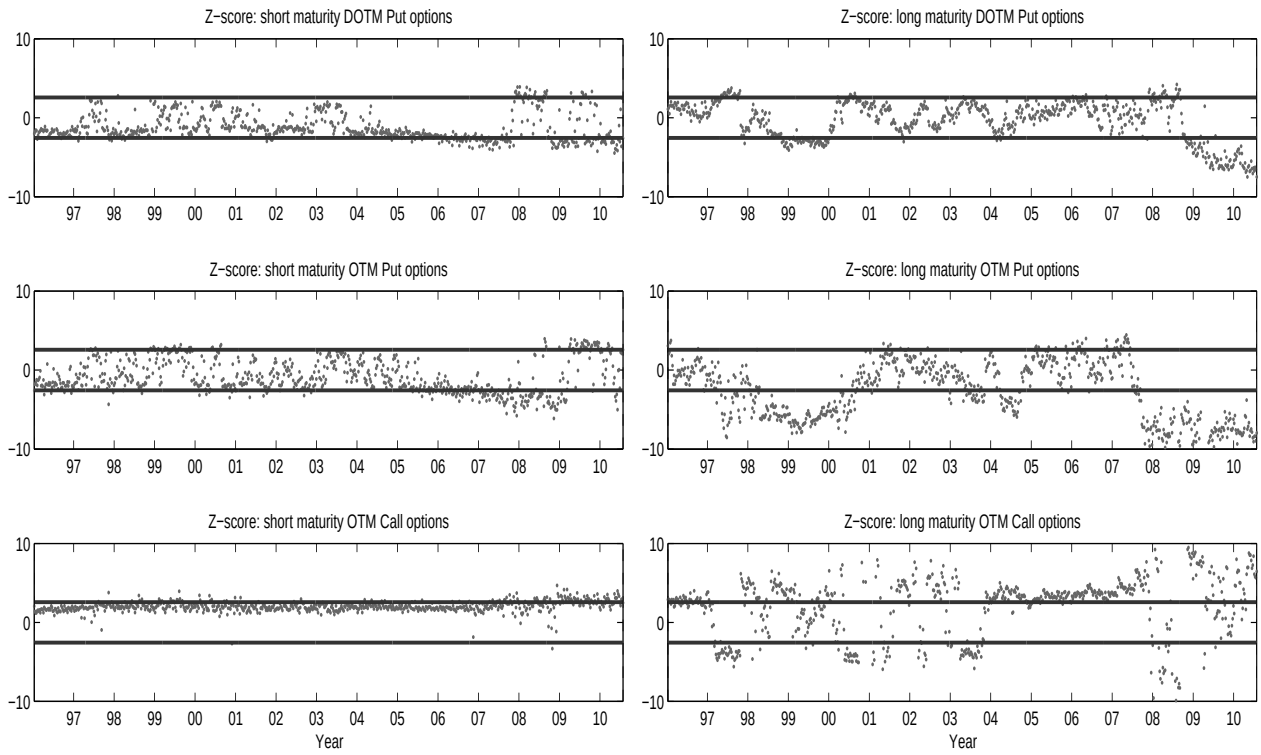


Figure 4: *Option Price Fit for the Two-factor Model.* Notation as for Figure 3.

long maturities. Since we do not average across moneyness, the series are inevitably somewhat noisy, but the plots are nevertheless indicative of systematic mispricing. First, it is apparent that the two-factor model tracks the term structure of implied volatility much better. In particular, unlike the two-factor model, the one-factor model performs poorly from the end of 1998 till the end of 2000 and from the beginning of 2009 till the end of 2010. Both periods feature quickly mean-reverting volatility and a term structure of volatility that turns negative. The one-factor model cannot generate this behavior due to a one-to-one mapping between the term structure and the level of volatility. Introducing a second factor breaks this link and allows for some independent variation in the term structure at a given volatility level. This evidence also suggests that risk premia may have a more complicated relation with the level of volatility than typically assumed in equilibrium-based asset pricing models.

Table 8 reports rejection rates for tests of individual parameter stability across consecutive one-year samples. The table reveals significant variation in the parameter estimates, but also a dramatic improvement in the temporal stability for some parameters as we move from the one-factor to the two-factor model. This is indicative of improved model specification. Nevertheless, some parameters

Table 7: Diagnostic Tests for S&P 500 Option Data

| Test | One-factor Model Nominal Size | | Two-factor Model Nominal Size | |
|---|---|---|---|---|
| | 1% | 5% | 1% | 5% |
| **Panel A: Fit to the Option Panel** | | | | |
| DOTM, short-maturity puts | 18.16% | 39.74% | 21.84% | 45.53% |
| OTM, short-maturity puts | 24.87% | 53.03% | 27.89% | 50.39% |
| OTM, short-maturity calls | 20.53% | 55.00% | 16.58% | 48.55% |
| DOTM, long-maturity puts | 41.45% | 62.11% | 27.50% | 41.97% |
| OTM, long-maturity puts | 72.63% | 80.79% | 52.89% | 60.53% |
| OTM, long-maturity calls | 53.16% | 65.53% | 79.47% | 86.45% |
| **Panel B: Root-Mean Squared Error of IV Option Fit** | | | | |
| | 3.09% | | 2.32% | |
| **Panel C: Equality of Implied and Nonparametric Volatility** | | | | |
| | 54.08% | 65.66% | 49.74% | 61.58% |

*Note:* Panel A reports rejection frequencies across the full sample for the option fit to specific portions of the option surface at the end of trading on Wednesdays. This test is based on Corollary 1, using the first two maturities for the three initial tests and all remaining options with maturity less than one year for the last three tests. DOTM puts, OTM puts and calls, and short- versus long-maturity options are defined in Figure 3. Panel B provides the root-mean-squared-error of the model-implied Black-Scholes volatilities relative to the market mid-quote implied Black-Scholes volatilities across all options used for estimation over the full sample. The test in Panel C is defined in Table 3.

in the two-factor model are quite unstable, most notably the persistence parameters $\kappa_{d,1}$ and $\kappa_{d,2}$. Again, it is evident that neither model is correctly specified. In fact, for both models the joint test for stability of the full parameter vector across any two consecutive years has a 100% rejection rate.

Finally, Figure 7 depicts the nonparametric and two alternative option extracted volatility series. It is evident that they all are highly correlated. Nonetheless, the formal test for equality between the option-implied and the nonparametric diffusive volatility estimates rejects the null hypothesis for a nontrivial number of days in both models, as may be confirmed from Panel C of Table 7. The rejection rates for the two-factor model are somewhat lower than those for the one-factor model, once more suggesting it captures the option dynamics better. This is corroborated by the serial correlation in the discrepancy between the option implied and nonparametric volatility estimates

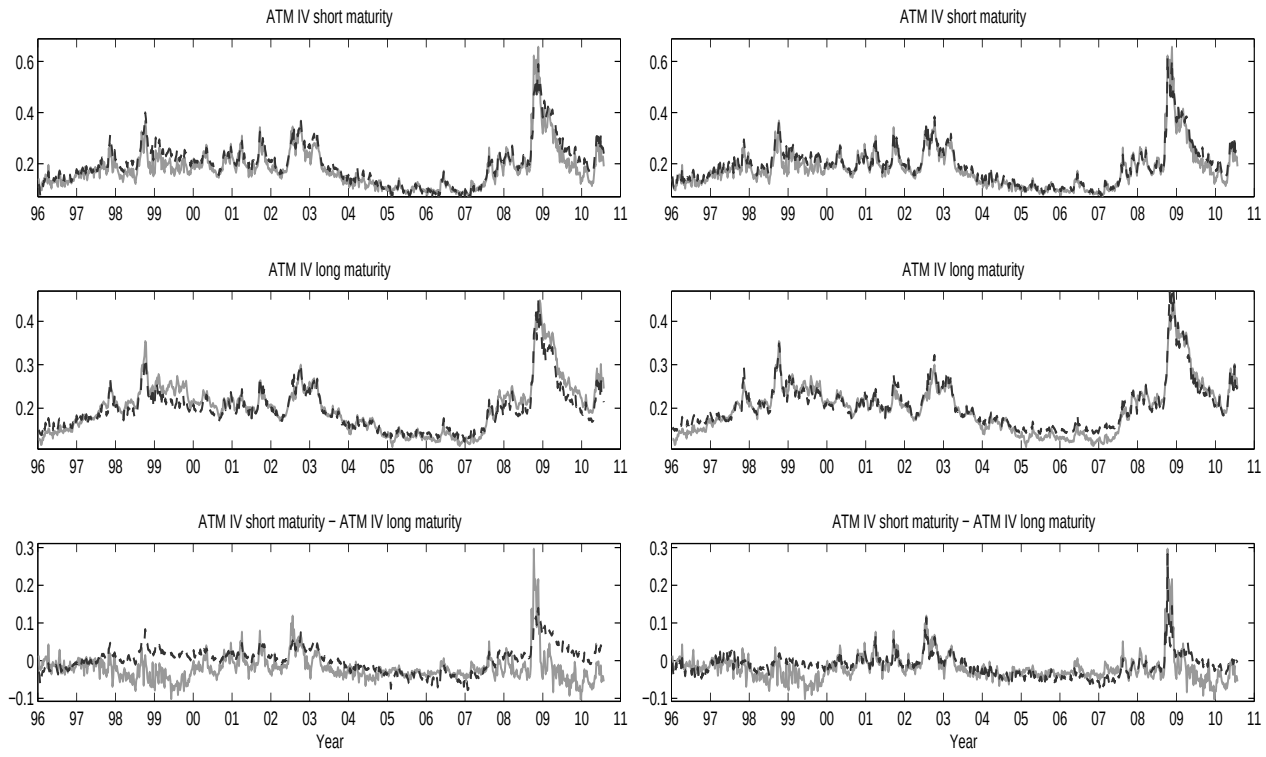


Figure 5: *Fit of the Term Structure of Implied Volatility.* The left side corresponds to the one-factor model and the right-hand side to the two-factor one. On each plot, the grey line represents the at-the-money implied volatility for the option with moneyness closest to 1, while the dashed line indicates the model implied volatility.

plotted in the bottom panels of Figure 7. Under correct model specification these series should not display significant autocorrelation. However, very strong temporal dependence is evident for the one-factor model, while it drops substantially for the two-factor model.

# 8 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

Table 8: Parameter Stability Tests on S&P 500 options data

| Parameter | Nominal size of test | | Parameter | Nominal size of test | |
|---|---|---|---|---|---|
| | 1% | 5% | | 1% | 5% |
| **Panel A: One-Factor Model** | | | | | |
| $\rho_d$ | 62.86% | 70.48% | $\lambda_j$ | 56.19% | 67.62% |
| $\overline{v}$ | 71.43% | 73.33% | $\mu_x$ | 20.00% | 25.71% |
| $\kappa$ | 91.43% | 93.33% | $\sigma_x$ | 49.52% | 61.91% |
| $\sigma_d$ | 77.14% | 80.95% | $\mu_v$ | 31.43% | 36.19% |
| | | | $\rho_j$ | 13.33% | 17.14% |
| **Panel B: Two-Factor Model** | | | | | |
| $\rho_{d,1}$ | 8.57% | 16.19% | $\lambda_{j,0}$ | 37.14% | 49.52% |
| $\overline{v}_1$ | 7.62% | 12.38% | $\lambda_{j,1}$ | 21.91% | 33.33% |
| $\kappa_{d,1}$ | 72.38% | 70.05% | $\mu_x$ | 0.00% | 4.76% |
| $\sigma_{d,1}$ | 42.86% | 53.33% | $\sigma_x$ | 9.52% | 17.14% |
| $\rho_{d,2}$ | 7.62% | 16.19% | $\mu_v$ | 23.81% | 34.29% |
| $\overline{v}_2$ | 76.19% | 80.95% | $\rho_j$ | 1.91% | 2.86% |
| $\kappa_{d,2}$ | 69.52% | 77.14% | | | |
| $\sigma_{d,2}$ | 0.95% | 0.95% | | | |

*Note:* Tests based on parameter estimates of the models over consecutive calendar years in the sample. The test is based on Corollary 2.

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 the double-jump model of Duffie et al. (2000) to allow for time-varying compensation for jumps and a more flexible relation between risk compensation and (stochastic) volatility. We find nontrivial improvements for the extension of the model relative to the standard one-factor representation. Nonetheless, the model is rejected when confronted with our full battery of specification tests. This demonstrates the power of the developed tools for model diagnostics. Our findings point towards the use of a more flexible jump distribution as the most promising avenue for improving the fit to the option panel.

# 9 Appendix

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

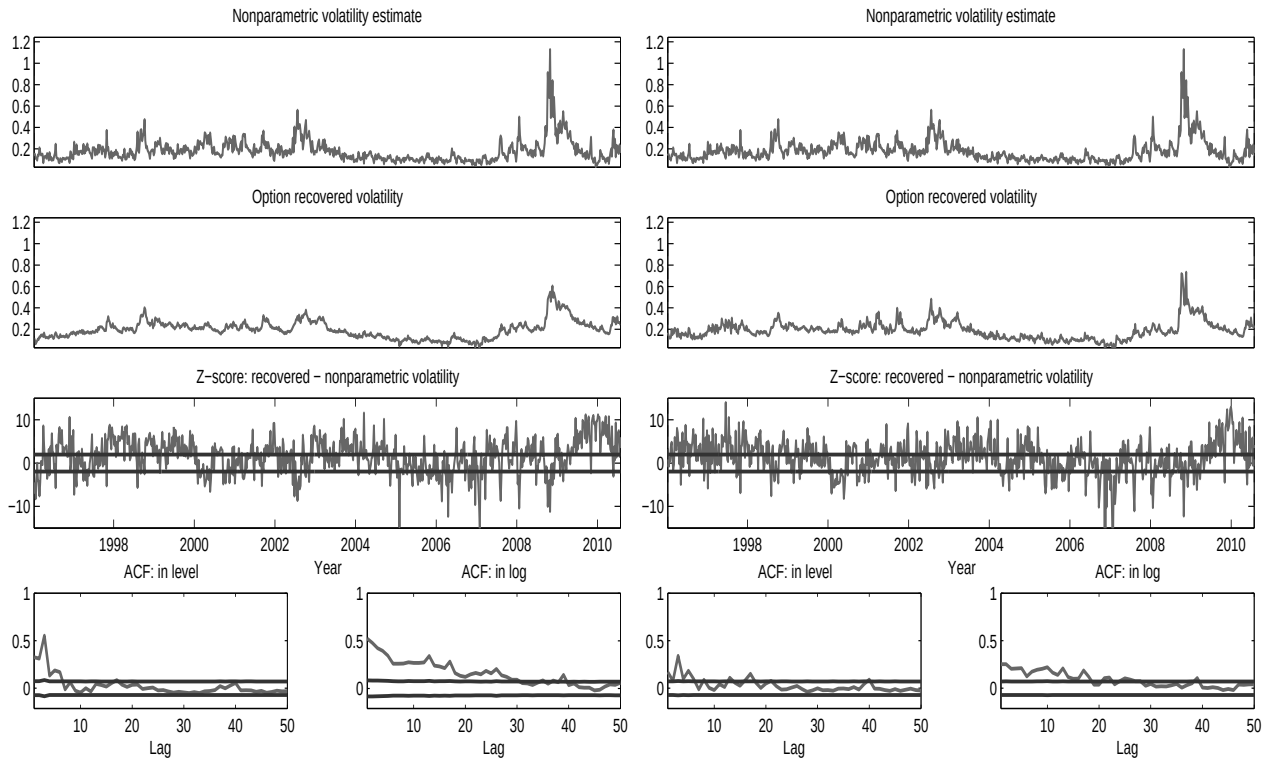


Figure 6: *Volatility Estimates.* The left side corresponds to the one-factor model and the right-hand side to the two-factor one. The bottom plots of the figure are the autocorrelations in $\xi_1(\widehat{\mathbf{S}}_t^n) - \widehat{V}_t^n$ and $\log(\xi_1(\widehat{\mathbf{S}}_t^n)) - \log(\widehat{V}_t^n)$.

## 9.1 Preliminary results

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

$$\begin{pmatrix} \frac{1}{\sqrt{N_1}} \sum_{j=1}^{N_1} \nabla_{\mathbf{S}} \kappa(k_j, \tau_j, \mathbf{S}_1, \theta_0) \epsilon_{1,k_j,\tau_j} \\ \vdots \\ \frac{1}{\sqrt{N_T}} \sum_{j=1}^{N_T} \nabla_{\mathbf{S}} \kappa(k_j, \tau_j, \mathbf{S}_T, \theta_0) \epsilon_{T,k_j,\tau_j} \\ \frac{1}{\sqrt{N_1}} \sum_{j=1}^{N_1} \nabla_{\theta} \kappa(k_j, \tau_j, \mathbf{S}_1, \theta_0) \epsilon_{1,k_j,\tau_j} \\ \vdots \\ \frac{1}{\sqrt{N_T}} \sum_{j=1}^{N_T} \nabla_{\theta} \kappa(k_j, \tau_j, \mathbf{S}_T, \theta_0) \epsilon_{T,k_j,\tau_j} \end{pmatrix} \xrightarrow{\mathcal{L}-s} \left(\widetilde{\mathbf{\Omega}}_T\right)^{1/2} \begin{pmatrix} \mathbf{E}_1 \\ \vdots \\ \mathbf{E}_T \\ \mathbf{E}_1' \\ \vdots \\ \mathbf{E}_T' \end{pmatrix}, \tag{17}$$

*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

$$\chi_j^{(t)} = \begin{pmatrix} \nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0)\epsilon_{t,k_j,\tau_j} \\ \nabla_\theta\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0)\epsilon_{t,k_j,\tau_j} \end{pmatrix}, \quad t=1,...,T, \tag{18}$$

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

$$\Upsilon_t^{N_t} = \frac{1}{\sqrt{N_t}}\sum_{j=1}^{N_t}\chi_j^{(t)} \overset{\mathcal{L}-s}{\longrightarrow} \Upsilon_t, \quad t=1,...,T, \tag{19}$$

where $\Upsilon_t$ 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 (19) above it now suffices to prove,

$$\mathbb{E}\left(\chi_j^{(t)}|\widetilde{\mathcal{F}}_{j-1}\right)=0, \ \frac{1}{N_t}\sum_{j=1}^{N_t}\mathbb{E}\left(\chi_j^{(t)}\chi_j^{(t)'}|\widetilde{\mathcal{F}}_{j-1}\right) \overset{\mathbb{P}}{\longrightarrow} \begin{pmatrix} \widetilde{\mathbf{\Omega}}_T^{t,t} & \widetilde{\mathbf{\Omega}}_T^{t,T+t} \\ \widetilde{\mathbf{\Omega}}_T^{T+t,t} & \widetilde{\mathbf{\Omega}}_T^{T+t,T+t} \end{pmatrix}, \ \frac{1}{N_t^2}\sum_{j=1}^{N_t}\mathbb{E}||\chi_j^{(t)}|\widetilde{\mathcal{F}}_{j-1}||^4 \overset{\mathbb{P}}{\longrightarrow} 0. \tag{20}$$

The first and the third result of (20) 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 (19) holds jointly for $t = 1, ..., T$ with the limits being $\mathcal{F}_T^{(0)}$-conditionally independent. That is, we will show

$$\mathbb{E}\left(Y\prod_{t=1}^{T} f_t\left(\Upsilon_t^{N_t}\right)\right) \longrightarrow \mathbb{E}\left(Y\prod_{t=1}^{T}\mathbb{E}\left(f_1\left(\Upsilon_1\right)|\mathcal{F}_T\right)\right), \tag{21}$$

for $f_t(\cdot)$ being Lipschitz functions on $\mathbb{R}^{p+q}$, $Y$ denoting a bounded random variable on $\mathcal{F}_T$, and $\Upsilon_t$ indicating the limits in (19).

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

$$\widehat{Q}^N \xrightarrow{\mathbb{P}} Q, \tag{22}$$

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 $\Upsilon_t$ (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

$$\mathbb{E}\left(f_t(\Upsilon_t^{N_t})\big|\mathcal{F}_T^{(0)}\right) \xrightarrow{\mathbb{P}} \mathbb{E}\left(f_t(\Upsilon_t)\big|\mathcal{F}_T^{(0)}\right). \tag{23}$$

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

$$\mathbb{E}\left(Y\prod_{t=1}^{T}\mathbb{E}\left(f_t\left(\Upsilon_t^{N_t}\right)|\mathcal{F}_T^{(0)}\right)\right) \longrightarrow \mathbb{E}\left(Y\prod_{t=1}^{T}\mathbb{E}\left(f_t(\Upsilon_t)|\mathcal{F}_T^{(0)}\right)\right), \tag{24}$$

and therefore (21) 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 $\sigma$-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 $(t,k,\tau)$. 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 (21) 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 3, we have*

$$\sqrt{k_n}\begin{pmatrix} V_1\left\{\frac{n}{k_n}\sum_{i\in I^{+,n}}\left(\Delta_i^{1,n}W\right)^2 - 1\right\} \\ \vdots \\ V_T\left\{\frac{n}{k_n}\sum_{i\in I^{+,n}}\left(\Delta_i^{T,n}W\right)^2 - 1\right\}\end{pmatrix} \xrightarrow{\mathcal{L}-s} \begin{pmatrix}\sqrt{2}V_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sqrt{2}V_T\end{pmatrix}\begin{pmatrix}\widetilde{E}_1 \\ \vdots \\ \widetilde{E}_T\end{pmatrix}, \tag{25}$$

*where* $\{\widetilde{E}_t\}_{t\geq 1}$ *are defined in Theorem 3.*

**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,

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

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 3 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,

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

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

$$\mathbb{E}\left(Y f(X_1^N) g(X_2^n)\right) \longrightarrow \mathbb{E}\left(Y \mathbb{E}\left(f(X_1)|\mathcal{F}_T\right) \mathbb{E}\left(g(X_2)|\mathcal{F}_T\right)\right), \tag{26}$$

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 (17) and (25), 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

$$\mathbb{E}\left(f(X_1^N)\middle|\mathcal{F}_T^{(0)}\right) \xrightarrow{\mathbb{P}} \mathbb{E}\left(f(X_1)\middle|\mathcal{F}_T^{(0)}\right). \tag{27}$$

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

$$\mathbb{E}\left\{ Y g(X_2^n) \left| \mathbb{E}\left(f(X_1^N)\middle|\mathcal{F}_T^{(0)}\right) - \mathbb{E}\left(f(X_1)\middle|\mathcal{F}_T^{(0)}\right) \right| \right\} \leq K\epsilon, \tag{28}$$

for some positive constant $K$ (that does not depend on $\epsilon$ and $\overline{N}$), where we also exploited the boundedness of $Y$, $f(X_1^N)$ 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 (26) 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. □

**Lemma 4** *Under the conditions of Theorem 3 we have*

$$\frac{n}{\sqrt{k_n}}\sum_{t=1}^{T}\sum_{i\in I^{+,n}}\left|(\Delta_i^{t,n}X)^2 1\left(|\Delta_i^{t,n}X|\leq \alpha n^{-\varpi}\right)-V_t(\Delta_i^{t,n}W)^2\right| \xrightarrow{\mathbb{P}} 0. \tag{29}$$

**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 $\alpha_t$, $V_t$ 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

$$d\log(X_t)=\left(\alpha_t-\frac{1}{2}V_t\right)dt+\sqrt{V_t}dW_t+\int_{x>-1}x\widetilde{\mu}(dt,dx)+\int_{x>-1}(\log(1+x)-x)\,\mu(dt,dx), \tag{30}$$

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

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

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

$$\begin{aligned}
&(\Delta_i^{t,n}X)^2 1_{\{|\Delta_i^{t,n}X|\leq\alpha n^{-\varpi}\}}-V_t\left(\Delta_i^{t,n}W\right)^2\\
&\quad=(\Delta_i^{t,n}X)^2 1_{\{|\Delta_i^{t,n}X|\leq\alpha n^{-\varpi}\}}-(\Delta_i^{t,n}X)^2 1_{\{|\Delta_i^{t,n}Y|\leq\frac{\alpha}{2}n^{-\varpi}\}}+|\Delta_i^{t,n}X-\Delta_i^{t,n}Z|^2 1_{\{|\Delta_i^{t,n}Y|\leq\frac{\alpha}{2}n^{-\varpi}\}}\\
&\qquad+2(\Delta_i^{t,n}X-\Delta_i^{t,n}Z)\Delta_i^{t,n}Z 1_{\{|\Delta_i^{t,n}Y|\leq\frac{\alpha}{2}n^{-\varpi}\}}-(\Delta_i^{t,n}Z)^2 1_{\{|\Delta_i^{t,n}Y|>\frac{\alpha}{2}n^{-\varpi}\}}\\
&\qquad+(\Delta_i^{t,n}Z)^2-V_{t+\frac{i-1}{n}}\left(\Delta_i^{t,n}W\right)^2+(V_{t+\frac{i-1}{n}}-V_t)\left(\Delta_i^{t,n}W\right)^2.
\end{aligned} \tag{31}$$

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,

$$|(\Delta_i^{t,n}X)^2 1_{\{|\Delta_i^{t,n}X|\leq\alpha n^{-\varpi}\}}-(\Delta_i^{t,n}X)^2 1_{\{|\Delta_i^{t,n}Y|\leq\frac{\alpha}{2}n^{-\varpi}\}}|\leq(\Delta_i^{t,n}X)^2 1_{\{|\Delta_i^{t,n}X-\Delta_i^{t,n}Y|\geq\frac{\alpha}{2}n^{-\varpi}\}}, \tag{32}$$

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

$$\mathbb{E}\left|(\Delta_i^{t,n}X)^2 1_{\{|\Delta_i^{t,n}X|\leq\alpha n^{-\varpi}\}}-(\Delta_i^{t,n}X)^2 1_{\{|\Delta_i^{t,n}Y|\leq\frac{\alpha}{2}n^{-\varpi}\}}\right|\leq Kn^{-\zeta},\ \ \forall\zeta>0. \tag{33}$$

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

$$\mathbb{E}|\Delta_i^{t,n}Y|^\zeta\leq Kn^{-1},\quad\forall\zeta\geq\beta. \tag{34}$$

Exploiting the above inequality, we deduce

$$\mathbb{E}\left\{|\Delta_i^{t,n}X-\Delta_i^{t,n}Z|^2 1_{\{|\Delta_i^{t,n}Y|\leq\frac{\alpha}{2}n^{-\varpi}\}}\right\}\leq Kn^{-1-(2-\beta)\varpi}. \tag{35}$$

Next, we decompose $Y_t$ depending on the value of $\beta$ 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 $\beta$ 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 $1(|a+b|<c) \leq 1(|a|>c) + 1(|b|<2c)$ for any $a,\ b \in \mathbb{R}$ and $c>0$, we get

$$\mathbb{E}\left|(\Delta_i^{t,n}X - \Delta_i^{t,n}Z)\Delta_i^{t,n}Z 1_{\{|\Delta_i^{t,n}Y| \leq \frac{\alpha}{2}n^{-\varpi}\}}\right| \leq Kn^{-1/2-\frac{1}{\beta}\wedge 1-(1-\beta)\varpi\vee 0}. \tag{36}$$

Moreover, applying the Hölder inequality and the inequality (34) yields,

$$\mathbb{E}\left\{(\Delta_i^{t,n}Z)^2 1_{\{|\Delta_i^{t,n}Y| > \frac{\alpha}{2}n^{-\varpi}\}}\right\} \leq Kn^{-1-(1-\beta\varpi)}. \tag{37}$$

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

$$\mathbb{E}\left|(\Delta_i^{t,n}Z)^2 - V_{t+\frac{i-1}{n}}\left(\Delta_i^{t,n}W\right)^2\right| \leq Kn^{-1/2}\sqrt{\int_{t+\frac{i-1}{n}}^{t+\frac{i}{n}} \mathbb{E}(V_s - V_{t+\frac{i-1}{n}})^2 ds} \leq Kn^{-3/2}. \tag{38}$$

$$\mathbb{E}\left\{\left(V_{t+\frac{i-1}{n}} - V_t\right)(\Delta_i^{t,n}W)^2\right\} \leq Kn^{-3/2}\sqrt{i-1}. \tag{39}$$

Combining these findings, we validate the convergence result in (29). □

## 9.2 Proof of Theorem 1

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

$$\begin{aligned}\left(\widehat{\kappa}_{t,k_j,\tau_j} - \kappa(k_j,\tau_j,\mathbf{Z}_t,\theta)\right)^2 =& \epsilon_{t,k_j,\tau_j}^2 + \left(\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0) - \kappa(k_j,\tau_j,\mathbf{Z}_t,\theta)\right)^2 \\ &+ 2\epsilon_{t,k_j,\tau_j}\left(\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0) - \kappa(k_j,\tau_j,\mathbf{Z}_t,\theta)\right).\end{aligned} \tag{40}$$

Assumption A2 implies that

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

is almost surely positive ($\delta$ depends on the realization $\omega$). 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 $\omega$ in the probability space) over which the difference $\left(\kappa(k,\tau,\mathbf{S}_t,\theta_0) - \kappa(k,\tau,\mathbf{Z}_t,\theta)\right)^2$ 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 $(k_j, \tau_j)$ 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 $\kappa$ is continuously differentiable in its log-moneyness argument, there exists an $\overline{N}_1 > 0$, almost surely ($\overline{N}_1$ depends on $\omega$, of course), such that for $\min_{t=1,...,T} N_t > \overline{N}_1$:

$$\inf_{t=1,...,T:\ \cup\ ||\mathbf{Z}_t - \mathbf{S}_t|| > \epsilon\ \cup\ ||\theta-\theta_0|| > \epsilon} \sum_{t=1}^{T}\frac{1}{N_t}\sum_{j=1}^{N_t}\left(\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0) - \kappa(k_j,\tau_j,\mathbf{Z},\theta)\right)^2 > \frac{\delta}{2}. \tag{41}$$

Now, assumption A3 implies that, for every infinite subsequence of $\min_{t=1,...,T} N_t$, there exists a further subsequence, denoted $N'$, along which we have,

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

where $\min_{t=1,...,T} N_t$ in the above almost sure convergence is an element of the subsequence $N'$. Therefore, for some $\iota \in (0,1)$, there exists $\overline{N}_2$, such that for $\min_{t=1,...,T} N_t > \max\{\overline{N}_1, \overline{N}_2\}$, along the subsequence $N'$, we have on $\omega$

$$\sup_{t=1,...,T:\ \cup\ ||\mathbf{Z}_t-\mathbf{S}_t||>\epsilon\ \cup\ ||\theta-\theta_0||>\epsilon} \sum_{t=1}^{T} \frac{1}{N_t} \sum_{j=1}^{N_t} \left(\widehat{\kappa}_{t,k_j,\tau_j} - \kappa(k_j,\tau_j,\mathbf{Z}_t,\theta)\right)^2 > \sum_{t=1}^{T} \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{t,k_j,\tau_j}^2 + \frac{\delta(1-\iota)}{2}, \tag{42}$$

where, again, $\min_{t=1,...,T} N_t$ is an element of the subsequence $N'$ and we also exploited (41). 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 $\lambda_n$ converges to a finite $\lambda$, there exists $\overline{n} > 0$ such that, for $n > \overline{n}$ along the subsequence $n'$ and on the same $\omega$ for which (42) is true, we have,

$$\lambda_n \sup_{t=1,...,T} |\widehat{V}_t^{n'} - V_t|^2 \leq \frac{\delta(1-\iota)}{4}. \tag{43}$$

Combining (42) and (43) implies that, along the subsequences $N'$ and $n'$, we have for $\min_{t=1,...,T} N_t$ and $n$ sufficiently high,

$$\sup_{t=1,...,T} ||\widehat{\mathbf{S}}_t^n - \mathbf{S}_t|| < \epsilon \ \text{ and } \ ||\widehat{\theta}^n - \theta_0|| < \epsilon, \quad \text{a.s.} \tag{44}$$

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 (44),

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

Since, the choice of $\epsilon$ was arbitrary, the above proves the consistency result of the theorem. □

### 9.3 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 $\{\widehat{\mathbf{S}}_t^n\}_{t=1,...,T}$ and $\widehat{\theta}^n$, with probability approaching 1, solve

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

Using a first-order Taylor expansion we obtain,

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

where $\widetilde{H}_T$ denotes the analogue of $\widehat{H}_T$ 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 $\theta_0$. 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 $\{\widehat{V}_t^n\}_{t=1,...,T}$ is asymptotically bounded in probability, and the assumed relation $\lambda_n^2 \min_{t=1,...,T} N_t \to 0$ in the theorem.

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

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

for any finite $\mathbf{Z}$ and $\theta$. Moreover, since $\nabla_{\mathbf{S}} \kappa(k, \tau, \mathbf{Z}, \theta)$ is continuous in the arguments $\mathbf{Z}$ and $\theta$, the above convergence also holds locally uniformly in $\mathbf{Z}$ and $\theta$. Therefore, since $\widetilde{\theta} \xrightarrow{\mathbb{P}} \theta_0$ and $\widetilde{\mathbf{S}}_t \xrightarrow{\mathbb{P}} \mathbf{S}_t$ for $t = 1, ..., T$, we have $\widetilde{H}_T \xrightarrow{\mathbb{P}} H_T$. Combining this result with the limit result in Lemma 1, we establish the asymptotic distribution result in (6).

The remaining part of the proof is now the consistency of $\widehat{\Omega}_T$ for $\Omega_T$ and of $\widehat{H}_T$ for $H_T$. 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,

$$\begin{aligned} &\mathbb{E}\left\{ \left[ \frac{1}{N_t}\sum_{j=1}^{N_t} \left(\epsilon_{t,k_j,\tau_j}^2 - \phi_{t,k_j,\tau_j}\right) A^{(\iota_1,\iota_2)}(k_j, \tau_j, \mathbf{S}_t, \theta_0) \right]^2 \middle| \mathcal{F}^{(0)} \right\} \\ &= \frac{1}{N_t^2}\sum_{j=1}^{N_t} \varrho_{t,k_j,\tau_j} \left(A^{(\iota_1,\iota_2)}(k_j, \tau_j, \mathbf{S}_t, \theta_0)\right)^2 \longrightarrow 0, \quad \text{a.s.}, \quad \iota_1,\ \iota_2 = 1, ..., p, \end{aligned} \tag{45}$$

where $A^{(\iota_1,\iota_2)}(k, \tau, \mathbf{Z}, \theta)$ denotes the $(\iota_1, \iota_2)$ element of the matrix $\nabla_{\mathbf{S}} \kappa(k, \tau, \mathbf{Z}, \theta) \nabla_{\mathbf{S}} \kappa(k, \tau, \mathbf{Z}, \theta)'$, and where $\varrho_{t,k,\tau} = \mathbb{E}\left((\epsilon_{t,k,\tau}^2 - \phi_{t,k,\tau})^2 | \mathcal{F}^{(0)}\right)$, which by Assumption A4(iv) is finite. The almost sure convergence in (45) 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 moneness $(\underline{k}(t,\tau),\ \overline{k}(t,\tau))$. The convergence in (45) implies,

$$\mathbb{E}\left\{ \mathbb{E}\left\{ \left[ \frac{1}{N_t}\sum_{j=1}^{N_t} \left(\epsilon_{t,k_j,\tau_j}^2 - \phi_{t,k_j,\tau_j}\right) A^{(\iota_1,\iota_2)}(k_j, \tau_j, \mathbf{S}_t, \theta_0) \right]^2 \middle| \mathcal{F}^{(0)} \right\} \bigwedge 1 \right\} \longrightarrow 0. \tag{46}$$

Using Jensen's inequality and (46), as well as law of iterated expectations, we further have,

$$\mathbb{E}\left\{ \left[ \frac{1}{N_t}\sum_{j=1}^{N_t} \left(\epsilon_{t,k_j,\tau_j}^2 - \phi_{t,k_j,\tau_j}\right) A^{(\iota_1,\iota_2)}(k_j, \tau_j, \mathbf{S}_t, \theta_0) \right]^2 \bigwedge 1 \right\} \longrightarrow 0, \tag{47}$$

which is equivalent to

$$\frac{1}{N_t}\sum_{j=1}^{N_t}\left(\epsilon^2_{t,k_j,\tau_j}-\phi_{t,k_j,\tau_j}\right)\nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0)\nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0)' \xrightarrow{\mathbb{P}} 0. \tag{48}$$

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 $\overline{k}=\max_{t=1,...,T}\max_\tau \overline{k}(t,\tau)$,

$$\sup_{k\in(\underline{k},\overline{k})}\sup_{\mathbf{Z}\in(\mathcal{S}-\epsilon,\mathcal{S}+\epsilon)}\sup_{\theta\in(\theta_0-\epsilon,\theta_0+\epsilon)}\{\,||\nabla_{\mathbf{S}}\kappa(k,\tau,\mathbf{Z},\theta)||+||\nabla_{\mathbf{SS}}\kappa(k,\tau,\mathbf{Z},\theta)||\,\}<\infty,\quad a.s. \tag{49}$$

Therefore, since for $N_t$ sufficiently high, we will have $\widehat{\mathbf{S}}^n_t$ sufficiently close to $\mathbf{S}_t$ and $\widehat{\theta}^n$ sufficiently close to $\theta_0$, so that we have,

$$\widehat{\Omega}^{t,t}_T-\frac{1}{N_t}\sum_{j=1}^{N_t}\epsilon^2_{t,k_j,\tau_j}\nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0)\nabla_{\mathbf{S}}\kappa(k_j,\tau_j,\mathbf{S}_t,\theta_0)' \xrightarrow{\mathbb{P}} 0. \tag{50}$$

Combining (48) and (50) 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, $\widehat{\Omega}^{t,t}_T \xrightarrow{\mathbb{P}} \Omega^{t,t}_T$. □

### 9.4 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 $\{\widehat{\mathbf{S}}^n_t\}_{t=1,...,T}$ and $\widehat{\theta}^n$, as well as the properties of the stable convergence. First, a Taylor expansion yields

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

where $\{\widetilde{\mathbf{S}}_t\}_{t=1,...,T}$ is between $\{\widehat{\mathbf{S}}^n_t\}_{t=1,...,T}$ and $\{\mathbf{S}_t\}_{t=1,...,T}$ and $\widetilde{\theta}$ lies between $\widehat{\theta}^n$ and $\theta_0$. Therefore, using the consistency of $\{\widehat{\mathbf{S}}^n_t\}_{t=1,...,T}$ and $\widehat{\theta}$, as well as the smoothness of $\kappa(k,\tau,\mathbf{Z},\theta)$ (and its derivatives with respect to $\mathbf{Z}$ and $\theta$) in the log-moneyness, we obtain,

$$\frac{1}{\sqrt{N^{\mathcal{K}}_t}}\sum_{j:k_j\in\mathcal{K}}\left(\widehat{\kappa}_{t,k_j,\tau^*}-\kappa(k_j,\tau^*,\widehat{\mathbf{S}}^n_t,\widehat{\theta})\right)=\mathbf{\Pi}_T\begin{pmatrix}\sqrt{N_1}(\widehat{\mathbf{S}}^n_1-\mathbf{S}_1)\\\vdots\\\sqrt{N_T}(\widehat{\mathbf{S}}^n_T-\mathbf{S}_T)\\\sqrt{\frac{N_1+...+N_T}{T}}(\widehat{\theta}^n-\theta_0)\\\frac{1}{\sqrt{N^{\mathcal{K}}_t}}\sum_{j:k_j\in\mathcal{K}}\epsilon_{t,k_j,\tau^*}\end{pmatrix}+o_p, \tag{51}$$

where

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

Next, using the Taylor expansion in (45) and the consistency of $\{\widehat{\mathbf{S}}_t^n\}_{t=1,\ldots,T}$ and $\widehat{\theta}^n$, we get,

$$\frac{1}{\sqrt{N_t^{\mathcal{K}}}}\sum_{j:k_j\in\mathcal{K}}\left(\widehat{\kappa}_{t,k_j,\tau^*}-\kappa(k_j,\tau^*,\widehat{\mathbf{S}}_t^n,\widehat{\theta})\right)=\mathbf{\Pi}_T\begin{pmatrix}\mathbf{H}_T^{-1} & \mathbf{0}_{pT+q\times 1}\\ \mathbf{0}_{1\times pT+q} & 1\end{pmatrix}\begin{pmatrix}\zeta_{\{N_t\}_{t=1,\ldots,T}}\\ \frac{1}{\sqrt{N_t^{\mathcal{K}}}}\sum_{j:k_j\in\mathcal{K}}\epsilon_{t,k_j,\tau^*}\end{pmatrix}+o_p, \quad (52)$$

where we denote

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

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

$$\begin{pmatrix}\zeta_{\{N_t\}_{t=1,\ldots,T}}\\ \frac{1}{\sqrt{N_t^{\mathcal{K}}}}\sum_{j:k_j\in\mathcal{K}}\epsilon_{t,k_j,\tau^*}\end{pmatrix}\xrightarrow{\mathcal{L}-s}\mathbf{Z}_T, \quad (53)$$

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 (45)-(50) above), we can show that, $\frac{1}{N_t^{\mathcal{K}}}\widehat{\Pi}_T'\widehat{\Xi}_T\widehat{\Pi}_T\xrightarrow{\mathbb{P}}\Pi_T'\Xi_T\Pi_T$, where $\Xi_T$ 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 (10). □

### 9.5 Proof of Corollary 2

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

### 9.6 Proof of Theorem 3

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

### 9.7 Proof of Corollary 3

By Lemma 3, we have that $\{\frac{1}{\sqrt{N_t}}\left(\widehat{\mathbf{S}}_t^n-\mathbf{S}_t\right)\}_{t=1,\ldots,T}$ and $\sqrt{k_n}\{\widehat{V}_t^{+,n}-V_t\}_{t=1,\ldots,T}$ 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 $\{\frac{1}{\sqrt{N_t}}\left(\widehat{\mathbf{S}}_1^n-\mathbf{S}_1\right)\}_{t=1,\ldots,T}$, then by applying the Delta method (recall that $\xi_1$ 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.

Andersen, T. G., O. Bondarenko, and M. T. Gonzalez-Perez (2011). Coherent Model-Free Implied Volatility: A Corridor Fix for High-Frequency VIX. Working paper, Northwestern University.

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.

Breeden, D. and R. Litzenberger (1978). Prices of State Contingent Claims Implicit in Option Prices. *Journal of Business 51*, 621–652.

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., S. Heston, and K. Jacobs (2009). The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work so Well? *Management Science 55*, 1914–1932.

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

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

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.

Foster, D. and D. Nelson (1996). Continuous Record Asymptotics for Rolling Sample Variance Estimators. *Econometrica 64*, 139–174.

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

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

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 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.

Joslin, S., M. Priebsch, and K. Singleton (2010). Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks. Working paper, Stanford University.

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.

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.

Pastorello, S., V. Patilea, and E. Renault (2003). Iterative and Recursive Estimation in Structural Non-adaptive Models. *Journal of Business and Economic Statistics 21*, 449–482.

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

Ross, S. (1976). Options and efficiency. *Quarterly Journal of Economics 90*, 75–89.

Ross, S. (2011). The Recovery Theorem. Working paper, MIT.

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.

Todorov, V. and G. Tauchen (2011). Volatility Jumps. *Journal of Business and Economic Statistics 29*, 356–371.

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