Essays on the Econometrics of Option Prices

by

Erik Vogt

Department of Economics
Duke University

Date: ____________________________

Approved:

____
George Tauchen, Supervisor

____
Tim Bollerslev

____
Jia Li

____
Andrew Patton

____
Federico Bugni

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2014

Abstract

Essays on the Econometrics of Option Prices

by

Erik Vogt

Department of Economics
Duke University

Date: ______________________

Approved:

____
George Tauchen, Supervisor

____
Tim Bollerslev

____
Jia Li

____
Andrew Patton

____
Federico Bugni

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2014

Abstract

This dissertation develops new econometric techniques for use in estimating and conducting inference on parameters that can be identified from option prices. The techniques in question extend the existing literature in financial econometrics along several directions.

The first essay considers the problem of estimating and conducting inference on the term structures of a class of economically interesting option portfolios. The option portfolios of interest play the role of functionals on an infinite-dimensional parameter (the option surface indexed by the term structure of state-price densities) that is well-known to be identified from option prices. Admissible functionals in the essay are generalizations of the VIX volatility index, which represent weighted integrals of options prices at a fixed maturity. By forming portfolios for various maturities, one can study their term structure. However, an important econometric difficulty that must be addressed is the illiquidity of options at longer maturities, which the essay overcomes by proposing a new nonparametric framework that takes advantage of asset pricing restrictions to estimate a shape-conforming option surface. In a second stage, the option portfolios of interest are cast as functionals of the estimated option surface, which then gives rise to a new, asymptotic distribution

theory for option portfolios. The distribution theory is used to quantify the estimation error induced by computing integrated option portfolios from a sample of noisy option data. Moreover, by relying on the method of sieves, the framework is nonparametric, adheres to economic shape restrictions for arbitrary maturities, yields closed-form option prices, and is easy to compute. The framework also permits the extraction of the entire term structure of risk-neutral distributions in closed-form. Monte Carlo simulations confirm the framework's performance in finite samples. An application to the term structure of the synthetic variance swap portfolio finds sizeable uncertainty around the swap's true fair value, particularly when the variance swap is synthesized from noisy long-maturity options. A nonparametric investigation into the term structure of the variance risk premium finds growing compensation for variance risk at long maturities.

The second essay, which represents joint work with Jia Li, proposes an econometric framework for inference on parametric option pricing models with two novel features. First, point identification is not assumed. The lack of identification arises naturally when a researcher only has interval observations on option quotes rather than on the efficient option price itself, which implies that the parameters of interest are only partially identified by observed option prices. This issue is solved by adopting a moment inequality approach. Second, the essay imposes no-arbitrage restrictions between the risk-neutral and the physical measures by nonparametrically estimating quantities that are invariant to changes of measures using high-frequency returns data. Theoretical justification for this framework is provided and is based on an asymptotic setting in which the sampling interval of high frequency returns goes to zero as the sampling span goes to infinity. Empirically, the essay shows that inference on risk-neutral parameters becomes much more conservative once the assumption of identification is relaxed. At the same time, however, the conservative inference approach yields new and interesting insights into how option model parameters are related. Finally, the essay shows how the informativeness of the inference can be restored with the use of high frequency observations on the underlying.

The third essay applies the sieve estimation framework developed in this dissertation to estimate a weekly time series of the risk-neutral return distribution's quantiles. Analogous quantiles for the objective-measure distribution are estimated using available methods in the literature for forecasting conditional quantiles from historical data. The essay documents the time-series properties for a range of return quantiles under each measure and further compares the difference between matching return quantiles. This difference is shown to correspond to a risk premium on binary options that pay off when the underlying asset moves below a given quantile. A brief empirical study shows asymmetric compensation for these return risk premia across different quantiles of the conditional return distribution.

Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

Symbols

$\mathbb{P}$	The physical or objective measure.
$\mathbb{Q}$	The risk-neutral measure.

Abbreviations

VIX	Volatility Index, an integrated portfolio of put and call options written on the S&P 500 Index.
CBOE	The Chicago Board Options Exchange.
SVS	Synthetic Variance Swap, a scaled square of the VIX portfolio.
SPD	State-Price Density, a density that, when integrated over a Borel set, gives the price of a contingent claim that pays $1 if an outcome in the Borel set is realized.
CDF	Cumulative Distribution Function.
DGP	Data Generating Process.
BIC	Bayesian Information Criterion.
RHS	Right-Hand Side.
LHS	Left-Hand Side.
ARFIMA	Fractionally Integrated Autoregressive Moving Average, a long-memory time series model.

SV	Stochastic Volatility, a continuous-time stochastic process for returns with diffusive coefficient that is itself a stochastic process.
SVJ	Stochastic Volatility with Jumps, an SV process with discontinuous jumps in the return process.
SVJJ	Stochastic Volatility double-Jump, an SVJ process with discontinuous jumps in the volatility process.
HAC	Heteroskedasticity and Autocorrelation, a covariance matrix estimator.
GMS	Generalized Moment Selection.
MMM	Modified Method of Moments.
RV	Realized Variance, an integral or sum of squared log-returns.

Acknowledgements

I would like to thank my dissertation committee, George Tauchen (chairman), Tim Bollerslev, Federico Bugni, Jia Li, and Andrew Patton, for their invaluable insights and encouragement, as well as for organizing the outstanding Duke Financial Econometrics Lunchgroup. The Lunchgroup was easily one of the most terrifying, yet rewarding experiences in my adult life. The material in this dissertation has greatly benefitted from my presentations to the Lunchgroup, as well as from listening to and interacting with the professors and students. In particular, I am indebted to George and Jia for being my toughest critics and mentors, to Tim and Andrew for always keeping an open door to discuss ideas, and to Federico for his mentorship. I am inspired by their example.

The second essay in my dissertation represents joint work with Jia Li, whose support and ideas substantially improved upon my first prospectus draft on the topic. I would like to thank him for his contribution, as well as Federico Bugni, who brought the partial identification techniques to my attention.

The first essay in my dissertation relied on the techniques developed by Ron Gallant and Doug Nychka. I am grateful to Ron for many fruitful discussions and patient explanations.

I would also like to thank Duncan Thomas. Duncan first got me excited about econometrics when he taught an introductory course at UCLA when I was a sophomore. His influence on me continued when I worked as his research assistant at the Population Center and well beyond college when he encouraged me and supported my effort to come to Duke for grad school.

I gratefully acknowledge that substantial portions of this dissertation were made possible by funding from the Duke Economics program, the Duke Graduate School, and the Stuart Irwin Harris fellowship.

I am grateful also for my colleagues and friends at Duke: Lily Liu, Rosen Valchev, Dave Bielen, Iaryna Grynkiv, and Nick Ingwersen have provided many helpful conversations and comments.

Finally, I have to thank my parents, Emma and Stefan, as well as my fiancée, Ricky Powell, for their immeasurable support and encouragement during this process.

1

Introduction

Options are derivative securities whose payoffs depend on the behavior of a specified underlying asset. This dissertation studies estimation and inference problems on economic models whose parameters can be identified from the prices of options.

Over the last half-century, two major strands of literature have emerged concerning the study of option prices. The first strand of literature, pioneered by the seminal work of Black and Scholes (1973) and Merton (1973), takes the underlying asset and relevant risk factors as primitives and then examines the problem of determining prices for options written on that asset.[1] In this approach, option prices are derived from the behavior of other primitives, the underlying asset and relevant risk factors. The second strand of literature, starting at least since Ross (1976), treats option prices themselves as primitives. The enormous fruitfulness of the latter approach is in no small part due to the remarkable theoretical *spanning* properties of

[1] Further examples of papers that fall into this literature include Cox et al. (1985), Hull and White (1987), Heston (1993), Bakshi et al. (1997), Duffie et al. (2000), Christoffersen et al. (2006), and Christoffersen et al. (2008).

options, which play a fundamental role in completing markets by acting as de facto Arrow–Debreu securities.[2]

The primary focus of this dissertation is a collection of econometric and empirical contributions to the second strand of literature. That is, throughout this dissertation, options and their prices are treated as primitives, in the sense that they are used to extract information about the underlying asset and relevant risk factors. The process of extracting information from observed option price data is fundamentally an econometric problem and requires answers to questions about model identification, estimation, and inference. The chapters ahead consider variations on these three types of questions under varying assumptions on the option data generating process, the structure of the model considered, the dimensionality of parameters, the degree of identifiability of these parameters, and on functionals of the parameters.

1.1 Inference on Option Portfolios

The next chapter of this dissertation considers the problem of estimating the term structures of a class of economically interesting option portfolios. A primary example of such an option portfolio is the VIX, or volatility index, which is synthesized by combining a large number of S&P 500 Index options at a fixed maturity into a single option portfolio. The Chicago Board Options Exchange, or CBOE, publishes the value of the VIX portfolio using short-run, 30-day options, and when these are unavailable, it constructs two VIX portfolios using options that straddle the 30 day maturity and then performs a linear interpolation to arrive at a 30-day VIX approx-

[2] See, for example, Debreu (1959), Arrow (1964), Breeden and Litzenberger (1978), Banz and Miller (1978), Aït-Sahalia and Lo (1998), Britten-Jones and Neuberger (2000), Bakshi and Madan (2000), and Andersen et al. (2012).

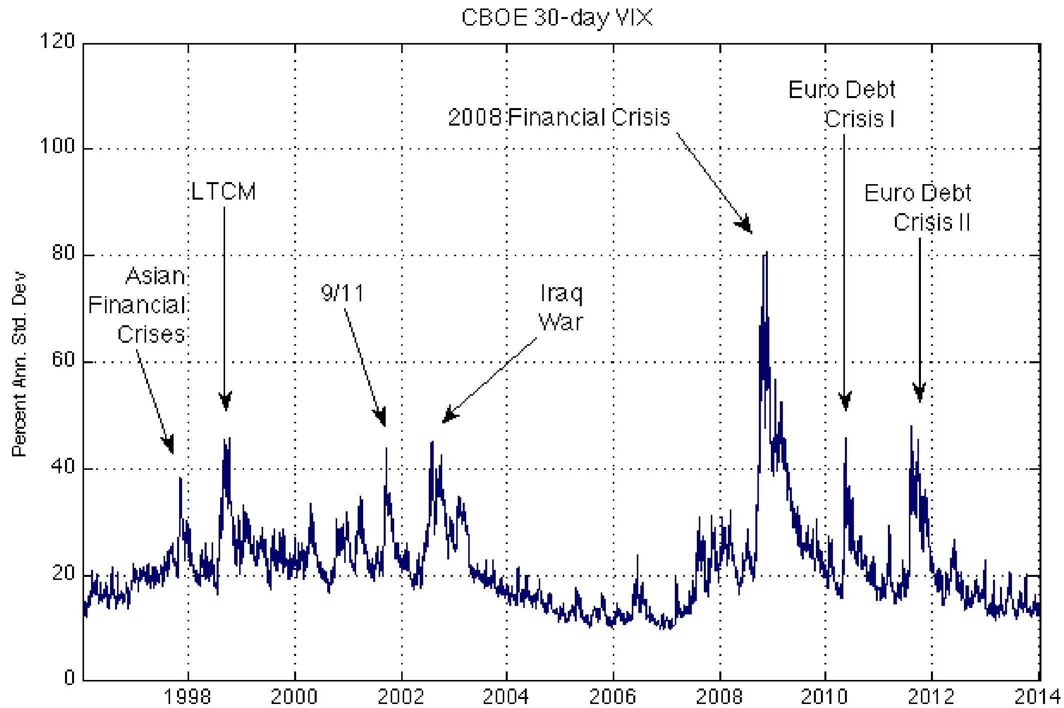

FIGURE 1.1: The daily VIX historical time series, 1996–2014. The Chicago Board Options Exchange's 30-day VIX volatility index represents the value of a portfolio of S&P 500 Index options.

imation. The CBOE's 30-day VIX has garnered an informal interpretation as an investor "fear gauge" because of its tendency to rise during periods associated with high financial market volatility.[3] A daily time series of the CBOE's historical 30-day VIX is plotted in Figure 1.1.

More formally, the CBOE's VIX approximates a model-free measure of implied volatility derived, for example, in Carr and Wu (2009). This measure of implied volatility (or its square) under certain conditions replicates the price of a volatility insurance contract (a so-called variance swap) through the construction of a hedging portfolio consisting of a theoretically infinite number of put and call options on the S&P 500 Index. The variance swap then pays off when future realized volatility

[3] As an example for this use of terminology, see the Wall Street Journal article by Kiernan (2014).

exceeds the value of the index. Depending on the length of the period for which one would like to purchase this volatility insurance, a natural term structure of hedging option portfolios can be formed to value the insurance contract by simply forming the Carr-Wu portfolio using options at different maturities. Thus, pricing the variance swap term structure is one economic motivation for being able to compute the term structure of the VIX portfolio.

The next chapter shows, however, that extending the CBOE's VIX approximation to longer maturity options is not straightforward. This is due to the illiquidity of options at longer maturities, which results in option prices that are more sparse and noisy at long maturities, possess larger bid-ask spreads, and are subject to more egregious synchronization errors than their short-maturity counterparts. It is clear that any VIX portfolio formed from the sparse and noisy long-maturity options should be subject to some form of estimation error. In other words, the study of long-run analogs of the CBOE's 30-day VIX should involve a theory of inference that can quantify the notion of estimation error in a VIX portfolio, which the econometrics literature has hitherto neglected to provide. The provision of such a theory of inference is one of the main contributions of this dissertation.

In particular, the next chapter provides a new nonparametric framework for estimating and conducting inference on the term structure of the VIX and other VIX-like option portfolios. The illiquidity problem of options at longer maturities is partially solved by introducing additional structure in the form of asset pricing theory. Specifically, the risk-neutral valuation equation imposes well-known shape constraints on option prices at all maturities.[4] To take advantage of these constraints, the next chapter proposes a sieve estimation framework that can fully incorporate the structure afforded by the risk-neutral valuation equation and that can, at the same time, remain fully nonparametric.[5] A key step in this framework comes from the observation that the shape constraints on option prices are driven by an expectation against a valid density (the so-called state-price density, or SPD), i.e. one that is nonnegative and integrates to one. The framework proposed below remains nonparametric by treating the term structure of SPDs that generated observed options as an element of an infinite-dimensional parameter space. The infinite-dimensional parameter is then estimated via the method of sieves, which generates shape-conforming option surfaces. This result extends the existing literature on shape-constrained nonparametric option pricing to the entire option surface, rather than exploiting the shape-constraints for single option maturities.[6]

The resulting nonparametrically estimated option surface is then used as an input to computing the term structure of the VIX or other VIX-like portfolios. Specifically, by treating the VIX (an integrated option portfolio) as a functional of the estimated option surface, I rely on some well-known functional delta methods to derive an asymptotic distribution theory for the VIX, pointwise along its term structure. The distribution theory is used to quantify the notion that the noise and sparseness of options at long maturities should induce an estimation error when forming the VIX portfolio from an option surface that was estimated from sparse and noisy

[4] See Duffie (2001) for an introduction to the notion of risk-neutral valuation.

[5] Chen (2007) provides an excellent survey of sieve estimation.

[6] See Aït-Sahalia and Duarte (2003), Jiang and Tian (2005), and Figlewski (2008) for examples of this literature.

options. Moreover, by computing the VIX term structure off of the estimated option surface, one can circumvent the issue of interpolating neighboring VIXs, as is done, for example, in the CBOE's calculation. The central contributions of the next chapter are, therefore, a procedure for estimating the term structure of the VIX (or VIX-like portfolios) at arbitrary maturities without resorting to theoretically unsupported interpolations, as well as the provision of an asymptotic distribution theory that is used to construct confidence intervals around the VIX term structure.

Monte Carlo simulations show that minimizing the Bayesian Information Criterion provides a simple, data-driven method for computing the sieve's expansion terms with good finite sample coverage probabilities. The sieve framework is then tested in three separate empirical applications: the pricing of the term structure of variance swaps, a comparison with the CBOE's interpolation method for generating a VIX term structure, and the estimation of the term structure of the variance risk premium. The latter application is one of a few existing extensions of the well-established literature on the variance risk premium to longer horizons.[7] The findings imply larger magnitudes for the term structure of the variance risk premium than previously found in parametric estimates.[8]

1.2 Inference on Partially Identified Option Pricing Models

The next essay in this dissertation considers the problem of inference on option model parameters that are finite-dimensional, but may only be partially identified.

[7] Bakshi and Madan (2006), Carr and Wu (2009), Bollerslev and Todorov (2011), Bollerslev et al. (2011), Bollerslev et al. (2013), and the references therein.

[8] See Aït-Sahalia et al. (2012) and Fusari and Gonzalez-Perez (2012).

This essay represents joint work with Jia Li and builds on earlier findings from my third-year prospectus. In our setup, the partial identification of option model parameters arises naturally when observed option data come in the form of bid and ask quotes instead of efficient prices. That is, since option pricing models make point predictions on the efficient, equilibrium price of an option, but observed option data come in pairs of quotes, it is not clear how to identify the option model's parameters without specifying a mapping from quotes to efficient prices. This mapping can be interpreted as a microstructure model that specifies the option market maker's price setting schedule relative to the efficient price. Because this mapping is neither revealed by the option pricing model nor the observed data, a lack of identification arises naturally.

To address the lack of a microstructure model that maps quotes to efficient prices, Li and I propose merely bounding the efficient price by the bid and ask quotes. This assumption substantially relaxes the existing practice in the literature of proxying unobserved efficient prices by the option mid-quote, i.e. the arithmetic average of observed bid and ask prices. This is because the mid-quote itself represents one of the simplest possible microstructure models, which posits that observed bid and ask quotes are always symmetric about the efficient price, in every state of the world. While the plausibility of this assumption has been questioned in the literature,[9] our essay represents the first to examine the issue from an identification viewpoint.

We argue that working with bid-ask bounds rather than mid-quotes is especially relevant with option data, since illiquid deep in-the-money options and long-maturity

[9] See Carr and Wu (2009) and Pan (2002).

options tend to have nontrivially wide bid-ask spreads. It is intuitively clear that the wider the bid-ask spread, the stronger the mid-quote assumption becomes. Indeed, since many efficient prices can potentially fit the observed bid-ask spread, choosing the bid-ask mid-point amounts to selecting an arbitrary element among a set of admissible prices. The goal, therefore, is to conduct inference on option model parameters that are set-identified.

Our inference on option pricing models is implemented via a moment inequality framework. To allow for option quotes that are observed with error, our identifying restrictions only require efficient option prices to lie between the observed bid-ask spread on average. The framework relies on methods from the burgeoning moment inequality literature, including the papers by Andrews and Soares (2010) and Andrews and Shi (2014). We extend this literature to accommodate features that are unique to option pricing. In particular, the literature on empirical option pricing (discussed below) has established the need for incorporating stochastic volatility dynamics in order to adequately explain observed patterns in option prices.[10] The implication for option pricing is that a latent spot volatility variable enters into the option pricing equation, and hence, into the moment functions that constitute the identifying restrictions.

To overcome the presence of a latent spot volatility variable within the moment functions, we propose replacing the latent variable with an estimate obtained from high-frequency data on the underlying. In particular, using the jump-robust estimator from Mancini (2009), we arrive at an end-of-day estimate of spot volatility using

[10] SeeHull and White (1987), Heston (1993), and Bakshi et al. (1997) for early stochastic volatility models.

the square root of average integrated variance near the closing minutes of trading. This estimate is then plugged into the option pricer in a second stage and is treated as though it represents an actual observation on spot volatility. We provide a rigorous theoretical justification for replacing latent spot volatility with this high-frequency estimate and further perform Monte Carlo experiments to confirm this practice in empirically realistic settings.

The option pricing models considered fall under the general framework of Duffie et al. (2000) and will therefore admit comparisons between our parameter set estimates and the point estimates obtained in the existing empirical options literature.[11] In particular, our empirical findings reveal large estimated parameter sets when the mid-quote assumption is replaced by the bid-ask quote bounds. These parameter sets reveal new and interesting relationships between option pricing parameters. Finally, by providing further identifying information using high-frequency data on the underlying, we show that the informativeness of inference can be restored even within our partially identified option pricing framework.

1.3 A Sieve Application to Estimating Quantile Risk Premia

The final essay in this dissertation takes the sieve estimation framework proposed in Chapter 2 to examine the time series properties of the option-implied risk-neutral return distribution. An economic motivation for studying this distribution is that it can be used to decompose the familiar equity risk premium into its return quantile constituents. That is, by directly comparing the time series of risk-neutral and

[11] See Bates (1996, 2000), Bakshi et al. (1997), Pan (2002), Eraker (2004), Broadie et al. (2007), Andersen et al. (2012), and Andersen et al. (2013).

objective return distributions, one can shed further light on which features of these distributions are responsible for the time-varying risk premia observed in the data. However, in contrast to previous estimates of higher-order moment risk premia, the third essay in this dissertation proposes comparisons via the risk-neutral and objective return quantiles. The differences in these quantiles are shown to have an interpretation as a risk premium on certain binary options that pay off when the underlying asset moves below a given quantile.

To study such risk premia, an estimate of $\mathbb{P}$-measure return quantiles is needed. To this end, I use the CAViaR forecasting model proposed in Engle and Manganelli (2004) to estimate conditional $\mathbb{P}$-measure return quantiles. However, in order to obtain 30-day ahead return quantiles, the CAViaR model requires monthly return data. To mitigate the loss of intra-month return variation, I augment the CAViaR quantile model with daily (intra-month) realized variance estimates and find that this procedure aids the nonlinear optimization's convergence.

Quantile comparisons similar to those discussed in the third essay of this dissertation are found in Metaxoglou and Smith (2013). My approach differs from theirs primarily through the use of sieves to compute the $\mathbb{Q}$-measure return quantiles, whereas Metaxoglou and Smith (2013) use a mixture of log-normals. The framework presented here can therefore be viewed as complementary to theirs from a nonparametrically motivated perspective.

The empirical findings of the quantile-forecasting approach point to the presence of asymmetric tail risk compensation. This underscores existing findings using moment, rather than quantile, based frameworks.[12] In particular, I find that there is significant compensation for bearing left-tail risk, but that the risk premium changes sign and has a smaller magnitude for bearing right-tail risk.

[12] See, for example, Conrad et al. (2013).

2

Term Structures, Shape-Constraints, and Inference for Option Portfolios

2.1 Introduction

This paper is concerned with nonparametrically estimating a shape-conforming option price surface and quantifying the statistical uncertainty around associated integrated option portfolios. The use of option prices in the extraction of economically significant quantities is linked to their ability to approximate state-contingent claims. This observation is due to the fundamental insights of Ross (1976) and Breeden and Litzenberger (1978), who show that options can be combined into portfolios that replicate the role of Arrow-Debreu securities in spanning or hedging against uncertain future states. More recently, option prices and their portfolios have been used to extract state-price densities,[1] to learn about the market prices of jump risk and

[1] See Jackwerth and Rubinstein (1996), Aït-Sahalia and Lo (1998), Aït-Sahalia et al. (2001), Bondarenko (2003), Yatchew and Härdle (2006), Figlewski (2008) among others.

crash fears,[2] to estimate investor risk aversion and risk-neutral skewness[3], to forecast returns,[4] and to study how investors price and perceive volatility risk.[5] The latter category, in particular, has benefitted from a collection of recently developed model-free implied volatility measures that are obtained by forming integrated portfolios of option prices, the most well-known of which is arguably the synthetic variance swap and its square-root, the VIX volatility index.[6]

However, a central issue with implementing the above theory is the sparseness and noise of option data due to illiquidity. For example, in order to construct the synthetic variance swap portfolio or risk-neutral density at some horizon τ, an infinite continuum of European options expiring in τ periods is required (Carr and Wu (2009)). In reality, option prices are discrete and truncated in strikes and maturity, since there may only be a few dozen observations available from which to infer the infinite option portfolio. The problem is even more severe when the objective is to investigate term structures implied by option prices, since the typical option panel has only a handful of maturities clustered at short horizons. To overcome this mismatch between data sparseness and theory, it has been customary to numerically interpolate observed options and then to treat the resulting estimates as though they represent actual observations on the theoretical object of interest. This approach omits at least two important considerations: first, the replacement of option prices by estimates

[2] See, for example, Bates (2000), Pan (2002), Broadie et al. (2007), and Bollerslev and Todorov (2011).

[3] See Bliss and Panigirtzoglou (2004) and Bakshi et al. (2003).

[4] Bakshi et al. (2011) and Bollerslev et al. (2013).

[5] Carr and Wu (2009), Bollerslev et al. (2011),Drechsler and Yaron (2011)

[6] See Britten-Jones and Neuberger (2000), Jiang and Tian (2005), and Carr and Wu (2009).

should induce an estimation error. How quickly does the estimation error vanish as more options become available? Second, options are frequently observed with microstructure error arising from synchronization issues, bid-ask spreads, and quote staleness. Does the presence and variance of these errors affect the precision of the estimated portfolio, and can this be meaningfully quantified?

To answer these questions, I propose a new nonparametric framework to (1) overcome the discreteness and truncation of option data in both the strike and maturity dimension and (2) additionally provide a distribution theory for option portfolios. The key ingredient in this framework is the nonparametric estimation of an option surface that satisfies certain shape constraints implied by economic theory. Thus, in illiquid regions of the option panel where maturities or strikes are only sparsely available, economic theory guides the estimator to maintain the proper structure. Moreover, the estimator has a number of appealing properties from an empirical perspective: first, option prices can be solved in closed-form. Second, because the option prices are shape-conforming along any maturity of interest, the estimator yields an entire term structure of valid state-price densities (SPDs) and risk-neutral CDFs indexed by maturity. That is, for any maturity of interest, the estimated SPDs always integrate to one, even in finite samples and even off the support of observed options. Third, the term structures of SPDs and risk-neutral CDFs are available in closed-form, avoiding the need to consider numerical differentiation or integration errors. Fourth, while having nonparametric properties, the estimator is easy to implement.

To be specific, given a cross-section of observed options at a fixed point in time, I propose solving a sieve least squares problem involving bivariate Hermite polynomial expansions of the joint risk-neutral density in both the return- and maturity-dimensions. This joint density is divided by its marginal on τ, yielding densities that are conditional versions of the Gallant and Nychka (1987) type (conditioning on τ) and are normalized to be nonnegative and to always integrate to one for any τ. When integrated against the option payoff function, I show that these Hermite densities yield closed-form option prices, SPD term structures, and risk-neutral CDF term structures. This result extends the work of León and Sentana (2009) to the bivariate case involving the τ expansion. The closed-form option prices are indexed by Hermite polynomial coefficients that are chosen to minimize a least squares criterion in a procedure that is numerically equivalent to nonlinear least squares.

The main econometric results of this paper are the consistency of the nonparametric price surface, its rate of convergence, and an asymptotic distribution theory for integrated portfolios of options. In other words, the latter result can be used to put confidence intervals on the synthetic variance swap (SVS) or VIX term structure that quantify the precision of portfolios that are constructed from estimated option prices. Throughout, the focus of this paper will be on the twin problems of (1) producing reliable estimates of option term structure objects (e.g. the SVSs, SPDs, and risk-neutral CDFs), and (2) the quantification of "reliability" as measured by asymptotically valid standard errors on the portfolio term structures. It should be emphasized, however, that the methods presented here are of interest even if the application is not about the option term structure. Indeed, because the present sieve estimator is shape-conforming for a given maturity across all strikes, it can be used to extrapolate option prices into extreme strikes. In light of the recent financial crisis and the renewed interest in studying tails of return distributions, the estimator's ability to estimate risk-neutral tails could be helpful in certain applications.

The paper connects with several strands of the literatures in finance and econometrics. The incorporation of asset pricing information relates to an existing literature on nonparametric shape-constrained estimation for options, which includes the extant papers by Aït-Sahalia and Duarte (2003), Bondarenko (2003), Yatchew and Härdle (2006), Figlewski (2008), as well as the numerical procedures to produce extrapolated option smiles in Bliss and Panigirtzoglou (2004), Jiang and Tian (2005), and Metaxoglou and Smith (2011). This literature has produced estimators that are shape-conforming for a fixed option maturity. In contrast, the framework presented here extends these methods in a new direction by generating shape-conforming surfaces for arbitrary (and even sparsely observed) maturities, while at the same time also offering closed-form term structures of valid state-price densities and risk-neutral CDFs using a single option panel. Finally, the paper's nonparametric distribution theory to quantify the estimation error in option portfolios is novel to this literature. Collectively, these results are obtained by connecting ideas from León and Sentana (2009) to the ongoing literature on sieve estimation, e.g. Gallant and Nychka (1987), Shen (1997), Chen and Shen (1998), Chen (2007), and Chen et al. (2013). In particular, the computationally simple distribution theory for option portfolios in this paper is adapted from Chen et al. (2013).

Simulations and several examples illustrate the framework's flexibility. Monte Carlo simulations show that the sieve estimator can capture the term structure of option prices, risk-neutral CDFs, and state-price densities implied by a variety of continuous-time double jump-diffusion data generating processes (DGPs), including the processes by Black and Scholes (1973), Heston (1993), and Duffie et al. (2000). Moreover, additional simulation exercises demonstrate that the portfolio distribution theory provides good coverage of the term structure of VIX's, regardless of whether the DGP has jumps in price and/or stochastic volatility. This flexibility is due to a key tuning parameter, the number of sieve expansion terms, which is required to grow with the sample size. The simulations show that minimizing the Bayesian Information Criterion (BIC) provides a simple but effective method for selecting the number of expansion terms automatically. Finally, a brief simulation shows that the method can be employed in an "out-of-sample" sense to generate daily or weekly balanced panels option portfolios, risk-neutral CDFs, and SPDs by evaluating the sieve estimator at arbitrary τ.[7]

My empirical applications of the sieve option estimator study the term structure of the synthetic variance swap portfolio and the associated variance risk premia using actual data from S&P 500 Index options from 1996 to 2010. The results show that sampling variation in noisy option prices induces up to 8% uncertainty around the fair value of the long-run variance swap contract when the swap is synthesized from noisy long-maturity options. In contrast, swaps synthesized from short- and medium-maturity options on the S&P 500 Index appear more precisely estimated, which supports the validity of the linearly interpolated approximations at short horizons commonly adopted in the literature. The latter observation is underscored in

[7] This result is relegated to Appendix A for brevity.

empirical comparisons of the sieve-estimated VIX term structure and the CBOE's discretized analog.

The sieve-estimated variance swap term structures are then used to estimate the term structure of the variance risk premium. An active literature in financial economics has documented the existence of a significant and time-varying risk premium that investors demand for bearing return variance risk.[8] Recently, Aït-Sahalia et al. (2012) and Fusari and Gonzalez-Perez (2012) have extended this literature by examining the variance risk premium at longer horizons using a flexible parametric model combined with data on variance swaps. The present paper complements their work from a nonparametric perspective and confirms that the variance risk premium term structure grows with maturity. Moreover, I find that the shape of the term structure depends on current volatility levels by applying a set of novel expectation hypothesis regressions.

The paper is organized as follows. Section 2.2 introduces the sieve least squares estimator for the shape-conforming option surface. Section 2.3 gives the closed-form option pricing formulas that are used in the sieve framework, and Section 2.4 establishes the estimator's consistency and its rate of convergence. Section 2.5 derives the asymptotic distribution theory for integrated option portfolios, which are functionals of the option surface estimated in the preceding sections. The results of Monte Carlo simulations that examine the sieve estimator's properties in finite samples is given in Section 2.6. Section 2.7 studies the term structure of the synthetic variance swap

[8] See Bakshi and Madan (2006), Carr and Wu (2009), Bollerslev and Todorov (2011), Bollerslev et al. (2011), and the related literature exploring parametric estimates of the volatility risk premium, e.g. Pan (2002), Eraker (2004), and Broadie et al. (2007). Equilibrium models that seek to explain the existence and size of the variance risk premium from a preference-based point of view are examined in Bakshi and Madan (2006), Bollerslev et al. (2009), and Drechsler and Yaron (2011).

portfolio and associated variance risk premia, and Section 2.8 concludes.

2.2 A Nonparametric, Shape-Conforming Option Surface

The goal is to provide model-free confidence intervals for the term structure of the VIX or VIX-like portfolios. These portfolio term structures can be cast as functionals of the nonparametric, shape-constrained option surface estimator outlined in this section.

2.2.1 Setup

Under mild restrictions, the current price $P_0(\kappa, \tau)$ of a European put option with strike κ and time-to-maturity τ is given by the well-known risk-neutral valuation equation[9]

$$\begin{aligned} P_0(\kappa, \tau) &\equiv e^{-r\tau} E_0^{\mathbb{Q}}\Big[\big[\kappa - S_\tau\big]_+ \Big| \mathbf{V}\Big] \\ &= e^{-r\tau} \int_0^\kappa [\kappa - S] f_0^{\mathbb{Q}}(S|\tau, \mathbf{V}) dS, \end{aligned} \tag{2.2.1}$$

where $\mathbf{V}$ is a vector of state variables that generate the current information set, $f_0^{\mathbb{Q}}(\,\cdot\,|\tau, \mathbf{V})$ is the unobserved transition or state-price density (SPD), and r is the risk-free rate. The components of $\mathbf{V}$ are left unspecified and can contain any number of variables relevant to pricing options. The Heston model, for example, specifies $\mathbf{V} = (S_0, V_0)$, where S_0 is the current underlying price and V_0 represents spot volatility (see Heston (1993), Duffie et al. (2000)).

Since the goal is to estimate a shape-conforming option surface at a single point in time, $\mathbf{V}$ realizes to some fixed value $\mathbf{v}_0$, so that the density in (2.2.1) becomes

[9] See, for example, Chapters 6 and 8 in Duffie (2001).

$f_0^{\mathbb{Q}}(S|\tau, \mathbf{V} = \mathbf{v}_0)$. To avoid cumbersome notation, I therefore define $f_0^{\mathbb{Q}}(S|\tau) \equiv f_0^{\mathbb{Q}}(S|\tau, \mathbf{V} = \mathbf{v}_0)$, since $\mathbf{v}_0$ is static across the option surface. On the other hand, τ is not static on the option surface because it indexes maturity. In this form, the risk-neutral valuation formula on a single option cross-section becomes

$$P_0(\kappa, \tau) \equiv e^{-r\tau} \int_0^{\kappa} [\kappa - S] f_0^{\mathbb{Q}}(S|\tau) dS. \tag{2.2.2}$$

Letting $\mathbf{Z} = (\kappa, \tau, r, q)$ denote a vector of characteristics containing the contract variables (κ, τ), the risk-free rate r, and the dividend yield q, the dependence of the option price on the SPD $f_0^{\mathbb{Q}}$ and the characteristics $\mathbf{Z}$ can be expressed as

$$P_0(\kappa, \tau) \equiv P(f_0^{\mathbb{Q}}, \mathbf{Z}).$$

The no-arbitrage pricing equation (2.2.2) implies shape restrictions on the option prices. By differentiating $P(f_0^{\mathbb{Q}}, \mathbf{Z})$ repeatedly with respect to the strike price κ, one has

$$\frac{\partial P_0}{\partial \kappa} = e^{-r\tau} F_0^{\mathbb{Q}}(\kappa|\tau), \qquad \frac{\partial^2 P_0}{\partial \kappa^2} = e^{-r\tau} f_0^{\mathbb{Q}}(\kappa|\tau),$$

where $F_0^{\mathbb{Q}}$ is the CDF of $f_0^{\mathbb{Q}}$. These conditions immediately imply that $P(f_0^{\mathbb{Q}}, \mathbf{Z})$ is monotone and convex in κ for any τ, and additionally has slope $e^{-r\tau}$ as $\kappa \to \infty$ and slope 0 as $\kappa \to 0$. Notice that these shape constraints follow directly from the nonnegativity of $f_0^{\mathbb{Q}}$ and the property that $f_0^{\mathbb{Q}}$ integrates to one with respect to S for all τ.[10]

Since the option price's shape constraints are implied by the fact that $f_0^{\mathbb{Q}}$ is a PDF,

[10] These shape constraints have been exploited elsewhere in the nonparametric option pricing literature for a single τ. See, for example, Aït-Sahalia and Lo (1998), Aït-Sahalia and Duarte (2003), Bondarenko (2003), Yatchew and Härdle (2006), and Figlewski (2008).

the strategy I employ to obtain shape-conforming option price estimates is to use approximating densities that are valid PDFs within the context of sieve estimation. However, instead of approximating $f_0^{\mathbb{Q}}$ directly, it turns out to be more convenient to first transform S by a change of variables, and then find approximating densities to a Jacobian transformation of $f_0^{\mathbb{Q}}$. The results of this straightforward change-of-variables are analytically closed-form option prices that are theoretically informative and computationally convenient.

2.2.2 Change of Variables

I propose the following change of variables to obtain closed-form expressions of estimates to the option price in Eq. (2.2.2). Let Y be the τ-measurable random variable that satisfies

$$\log\left(\frac{S}{S_0}\right) = \mu(\mathbf{Z}) + \sigma(\mathbf{Z})Y, \tag{2.2.3}$$

where $Y \sim f_0(\,\cdot\,|\tau)$, and $\mu(\cdot)$ and $\sigma(\cdot) > 0$ are known functions of the characteristics $\mathbf{Z}$, and where $f_0(\cdot|\tau)$ is the unknown density to be nonparametrically estimated from the data.

Under this change of variables, the valuation equation (2.2.2) becomes

$$\begin{aligned} P(f_0^{\mathbb{Q}}, \mathbf{Z}) &= e^{-r\tau}\int_0^{\kappa}\big(\kappa - S\big)f_0^{\mathbb{Q}}(S|\tau)dS \\ &= e^{-r\tau}\int_0^{d(\mathbf{Z})}\Big(\kappa - S_0 e^{\mu(\mathbf{Z})+\sigma(\mathbf{Z})Y}\Big)f_0(Y|\tau)dY \qquad (2.2.4) \\ &\equiv P_Y(f_0, \mathbf{Z}), \end{aligned}$$

where

$$d(\mathbf{Z}) = \frac{\log(\kappa/S) - \mu(\mathbf{Z})}{\sigma(\mathbf{Z})}. \tag{2.2.5}$$

The original SPD of interest evaluated at an arbitrary point s in the domain of S, $f_0^{\mathbb{Q}}(s|\mathbf{Z})$, can then be obtained by the Jacobian transformation

$$f_0^{\mathbb{Q}}(s|\tau) = (s\sigma(\mathbf{Z}))^{-1} f_0(s|\tau). \tag{2.2.6}$$

The sieve framework outlined below will produce consistent estimates $\widehat{f}_n$ of f_0. By a continuous mapping theorem, $\widehat{g}_n$ defined pointwise by $\widehat{g}_n(s|\tau) = (s\sigma(\mathbf{Z}))^{-1}\widehat{f}_n(s|\tau)$ will also converge to $f_0^{\mathbb{Q}}$.

If only the option price and not $\widehat{g}_n$ is needed, then one does not have to perform the Jacobian transformation, since Eq. (2.2.4) says $P(f_0^{\mathbb{Q}}, \mathbf{Z}) = P_Y(f_0, \mathbf{Z})$. This allows the analysis to focus on option pricing equations of the form

$$P_Y(f, \mathbf{Z}) = e^{-r\tau} \int_0^{d(\mathbf{Z})} \left(\kappa - S_0 e^{\mu(\mathbf{Z}) + \sigma(\mathbf{Z})Y}\right) f(Y|\tau) dY. \tag{2.2.7}$$

It is easy to verify that Eq. (2.2.7) contains the same shape restrictions as Eq. (2.2.2) for any f with $\int f(y|\tau)dy = 1$. Proposition 1 below solves this integral in closed-form when f represents a sieve approximation.

2.2.3 Sieve Least Squares Regression

The goal is to obtain a shape-conforming option surface by directly using the structure implied by Eq. (2.2.7). Notice that the true option price $P_Y(f_0, \mathbf{Z})$ is not observed because of the presence of the unknown infinite-dimensional parameter f_0, which is assumed to reside in some general function space $\mathcal{F}$. The space $\mathcal{F}$ consists of a very large class of smooth conditional densities $f(y|\tau)$ and will be described shortly. Thus, given a random sample $\{P_i, \mathbf{Z}_i\}_{i=1}^n$ on put option prices P_i and characteristics $\mathbf{Z}_i$, the idea is to solve problems of the form

$$\hat{f} = \arg \inf_{f \in \mathcal{F}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \Big[P_i - P_Y(f, \mathbf{Z}_i) \Big]^2 W_i \right\}, \tag{2.2.8}$$

where $P_Y(\cdot)$ is the known pricing functional from equation (2.2.7), $\mathbf{Z}_i$ is the vector of observables, and $W_i \equiv W(\mathbf{Z}_i)$ are known weights as a function of $\mathbf{Z}_i$.[11]

The main difficulty with solving the optimization problem in equation (2.2.8) is the infinite dimension of the function space $\mathcal{F}$. In general, optimizing over an infinite-dimensional function space may not be feasible or could even be ill-posed. Instead, it is typical to proceed by the method of sieves, which involves approximating $\mathcal{F}$ by a sequence of finite-dimensional function spaces (the "sieve" spaces)

$$\mathcal{F}_K \subset \mathcal{F}_{K+1} \subset \cdots \subset \mathcal{F} \tag{2.2.9}$$

[see Chen (2007), Chen and Shen (1998), and Shen (1997)]. The crucial property of sieve spaces is that they are much simpler than $\mathcal{F}$ but are sufficiently rich to eventually become dense in $\mathcal{F}$. That is, given any $f \in \mathcal{F}$ and any $\varepsilon > 0$, there is an M such that for all $K > M$, there exists $f_K \in \mathcal{F}_K$ such that $\|f - f_K\| < \varepsilon$.

The sieve space properties – along with mild regularity conditions – ensure that solutions to

$$\widehat{f}_{K_n} = \arg \min_{f \in \mathcal{F}_{K_n}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \Big[P_i - P_Y(f, \mathbf{Z}_i) \Big]^2 W_i \right\} \tag{2.2.10}$$

[11] Call options can be handled analogously in what follows, but for brevity I focus on puts.

are consistent for f_0. I provide conditions in Section 2.4 that ensure that the convergence of $\widehat{f}_{K_n}$ to f_0 also implies the convergence of $P_Y(\widehat{f}, \mathbf{Z})$ to $P_Y(f_0, \mathbf{Z})$ under suitable norms. Also note that the minimum in (2.2.10) is taken over the subspace $\mathcal{F}_{K_n} \subset \mathcal{F}$, where $K_n \to \infty$ slowly as $n \to \infty$. The requirement that $K_n \to \infty$ slowly is crucial and can be interpreted as the sieve analog of a bandwidth selection in kernel estimation and has an intuitive interpretation: as the sample size grows, the approximating spaces $\mathcal{F}_{K_n}$ increasingly resemble the parent space $\mathcal{F}$. The regularity conditions then ensure that optima on $\mathcal{F}_{K_n}$ indeed converge to f_0.

2.2.4 The Definition of $\mathcal{F}$ and its Sieve Spaces $\mathcal{F}_K$

For an option surface to conform to the theoretical shape restrictions of Eq. (2.2.7), $\mathcal{F}$ must be a function space consisting of conditional densities $f(Y|\tau)$ in the sense that $\int f(y|\tau)dy = 1$ for all τ. I construct such functions by first defining a collection of joint densities $\mathcal{F}^{Y,\tau}$ with elements $f^{Y,\tau}(y, \tau)$, and then defining $\mathcal{F}$ to consist of those functions $f(y|\tau)$ such that $f(y|\tau) = f^{Y,\tau}(y, \tau)/\int f^{Y,\tau}(y, \tau)dy$ for some $f^{Y,\tau} \in \mathcal{F}^{Y,\tau}$.

Gallant and Nychka (1987) show that if $\mathcal{F}^{Y,\tau}$ is a Sobolev subspace and $\{\mathcal{F}_K^{Y,\tau}\}_{K=0}^{\infty}$ is a collection of squared and scaled Hermite functions, then $\{\mathcal{F}_K^{Y,\tau}\}_{K=0}^{\infty}$ is a valid sieve for $\mathcal{F}^{Y,\tau}$. I show that the conditional approximating spaces $\{\mathcal{F}_K\}_{K=0}^{\infty}$ consisting of those functions f_K for which $f_K(y|\tau) = f_K^{Y,\tau}(y, \tau)/\int f_K^{Y,\tau}(y, \tau)dy$ for some $f_K^{Y,\tau} \in \mathcal{F}_K^{Y,\tau}$ is also a valid sieve for the conditional parent space $\mathcal{F}$, although the topologies differ. A formal discussion of these technical details is postponed until Section 2.4, when the asymptotic properties of the estimator are examined. For now, it is sufficient to note that when $\mathcal{F}_K$ is constructed from a ratio of two Gallant-Nychka densities, then there exists a norm under which $\mathcal{F}_K$ is a valid sieve for $\mathcal{F}$.

The Gallant-Nychka sieve spaces $\{\mathcal{F}_K^{Y,\tau}\}_{K=0}^{\infty}$ consist of functions of the form

$$\begin{aligned} f_K^{Y,\tau}(y,\tau) &= \left[\sum_{k=0}^{K_y}\left(\sum_{j=0}^{K_\tau}\beta_{kj}H_j(\tau)\right)H_k(y)\right]^2 e^{-\tau^2/2}e^{-y^2/2} \\ &= \left[\sum_{k=0}^{K_y}\alpha_k(B,\tau)H_k(y)\right]^2 e^{-\tau^2/2}e^{-y^2/2}, \end{aligned} \tag{2.2.11}$$

where H_k are Hermite polynomials of degree k, and where B is a matrix of coefficients with kj-entry β_{kj} and $K = (K_y+1)(K_\tau+1)$.[12] This function is clearly non-negative. Then, using orthogonality properties of Hermite polynomials, it can be shown that in order for $\iint f_K^{Y,\tau}(y,\tau)dyd\tau = 1$ for any K, it suffices to impose $\sum_{k=0}^{K_y}\sum_{j=0}^{K_\tau}\beta_{kj}^2 = 1$.

The conditional sieve spaces $\mathcal{F}_K$ will then consist of functions of the form

$$f_K(y|\tau) = f_K^{Y,\tau}(y,\tau)\bigg/\int f_K^{Y,\tau}(y,\tau)dy \tag{2.2.12}$$

for some joint density $f_K^{Y,\tau} \in \mathcal{F}_K^{Y,\tau}$. Notice that because the sieve joint densities $f_K^{Y,\tau}(y,\tau)$ are completely determined by the parameter matrix of coefficients B, then so are the conditional densities in $\mathcal{F}_K$. Therefore, for $\beta \equiv vec(B)$, the least squares

[12] The Hermite polynomials are orthogonalized polynomials. They are defined, for scalars x, by

$$H_K(x) = \frac{xH_{K-1}(x) - \sqrt{K-1}H_{K-2}(x)}{\sqrt{K}}, \quad K \geqslant 2$$

where $H_0(x) = 1$, and $H_1(x) = x$ [see, for example, León and Sentana (2009)]. Note that $H_K(x)$ is a polynomial in x of degree K.

problem in (2.2.10) becomes

$$\widehat{\beta}_n = \arg\min_{\beta \in \mathbb{R}^{K_n}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \Big[P_i - P_Y(\beta, \mathbf{Z}_i) \Big]^2 W_i \right\}$$
$$s.t. \sum_{k=0}^{K_y(n)} \sum_{j=0}^{K_\tau(n)} \beta_{kj}^2 = 1, \tag{2.2.13}$$

which is numerically equivalent to nonlinear least squares estimation for fixed K_n.

As written, $P_Y(\beta, \mathbf{Z}_i)$ is identical to $P_Y(f_K, \mathbf{Z}_i)$ from Eq. (2.2.7), which still requires an integration to obtain a candidate option price. Section 2.3 shows that in fact, $P_Y(f_K, \mathbf{Z}_i)$ is available in closed-form for any $f_K \in \mathcal{F}_K$, which considerably facilitates implementation.

2.2.5 The Sieve Satisfies the Required Shape Constraints

The sieve option prices produce highly structured option surfaces because the resulting option prices are shape-conforming for each τ. To see this, one differentiates with respect to κ to obtain

$$e^{r\tau} \frac{\partial P_Y(f_K, \mathbf{Z})}{\partial \kappa} = \int_0^{d(\mathbf{Z})} f_K(Y|\tau) dY\tau = F_K \left(\frac{\log(\kappa/S_0) - \mu(\mathbf{Z})}{\sigma(\mathbf{Z})} \middle| \tau \right) \tag{2.2.14}$$

where $F_K(\cdot|\tau)$ is the cumulative distribution function of f_K. Hence, because $f_K \geqslant 0$ and integrates to one, one observes that (a) $P_Y(f_K, \mathbf{Z})$ is increasing in the κ dimension (since $F_K >= 0$ as a CDF), (b) $P_Y(f_K, \mathbf{Z})$ is convex (since $\partial F_K / \partial \kappa$ is $f_K/(\kappa\sigma(\mathbf{Z}))$ and $f_K \geqslant 0$), and (c)

$$\lim_{\kappa \nearrow +\infty} e^{r\tau} \frac{\partial P_Y(f_K, \mathbf{Z})}{\partial \kappa} = 1, \qquad \lim_{\kappa \searrow 0} e^{r\tau} \frac{\partial P_Y(f_K, \mathbf{Z})}{\partial \kappa} = 0. \tag{2.2.15}$$

This shows that the sieve option prices satisfy the shape constraints implied by economic theory, for any τ.

2.3 Closed-form Option Prices

I now provide closed-form expressions for the sieve option prices $P_Y(f_K, \mathbf{Z})$ to be used in the regression (2.2.10) and show that $\mu(\mathbf{Z})$ and $\sigma(\mathbf{Z})$ can be chosen so that the sieve option prices have a natural interpretation as expansions around the Black-Scholes model.

2.3.1 Closed-Form Option Prices

To obtain closed-form option prices, it is convenient to first obtain a closed-form expression for the conditional sieve densities f_K of Eq. (2.2.12). This is done by expanding the squared polynomial term in the joint densities of Eq. (2.2.11) using techniques similar to those in León and Sentana (2009).

Lemma 2.3.1. *Any $f_K \in \mathcal{F}_K$ can be expressed in the form*

$$f_K(y|\tau) = \sum_{k=0}^{2K_y} \gamma_k(B,\tau) H_k(y)\phi(y), \tag{2.3.1}$$

where

$$\gamma_k(B,\tau) = \frac{\alpha(B,\tau)' A_k \alpha(B,\tau)}{\alpha(B,\tau)'\alpha(B,\tau)},$$

A_k is a known matrix of constants, and $\alpha(B,\tau)$ is a $(K_y+1)\times 1$ column vector obtained by stacking the $\alpha_k(B,\tau)$ in Eq. (2.2.11)*.*

Proof. Appendix A.2. □

The use of densities $f_K(y|\tau)$ that are linear combinations of functions in y helps with the derivation of closed-form option prices. The following result is an extension of Proposition 9 in León and Sentana (2009) to the case allowing for conditioning on τ.

Proposition 1. *For a candidate SPD $f_K(x|\tau) \in \mathcal{F}_K$ of the form given in equation* (2.3.1)*, the put option price $P_Y(f_K, \mathbf{Z})$ from equation* (2.2.7) *is given by*

$$P_Y(f_K, \mathbf{Z}) = \kappa e^{-r\tau}\left[\Phi(d(\mathbf{Z})) - \sum_{k=1}^{2K_y} \frac{\gamma_k(B,\tau)}{\sqrt{k}} H_{k-1}(d(\mathbf{Z}))\phi(d(\mathbf{Z}))\right]$$
$$- S_0 e^{-r\tau+\mu(\mathbf{Z})}\left[e^{\sigma(\mathbf{Z})^2/2}\Phi(d(\mathbf{Z}) - \sigma(\mathbf{Z})) + \sum_{k=1}^{2K_y} \gamma_k(B,\tau) I_k^*(d(\mathbf{Z}))\right] \quad (2.3.2)$$

where $\Phi(\cdot)$ is the standard normal CDF, $K = (K_y+1)(K_\tau+1)$, and where

$$I_k^*(d(\mathbf{Z})) = \frac{\sigma(\mathbf{Z})}{\sqrt{k}} I_{k-1}^*(d(\mathbf{Z})) - \frac{1}{\sqrt{k}} e^{\sigma(\mathbf{Z})d(\mathbf{Z})} H_{k-1}(d(\mathbf{Z}))\phi(d(\mathbf{Z})), \quad \text{for } k \geqslant 1,$$
$$I_0^*(d(\mathbf{Z})) = e^{\sigma(\mathbf{Z})^2/2}\Phi(d(\mathbf{Z}) - \sigma(\mathbf{Z})),$$

and $\gamma_k(B,\tau)$ is the coefficient function given in equation (2.3.1).

The price of a call option is given by

$$C_Y(f_K, \mathbf{Z}) = S_0 e^{-r\tau+\mu(\mathbf{Z})}\left[e^{\sigma(\mathbf{Z})^2/2}[1 - \Phi(d(\mathbf{Z}) - \sigma(\mathbf{Z}))] + \sum_{k=1}^{2K_y} \gamma_k(B,\tau) I_k^*(d(\mathbf{Z}))\right]$$
$$- \kappa e^{-r\tau}\left[[1 - \Phi(d(\mathbf{Z}))] - \sum_{k=1}^{K_y} \frac{\gamma_k(B,\tau)}{\sqrt{k}} H_{k-1}(d(\mathbf{Z}))\phi(d(\mathbf{Z}))\right]. \quad (2.3.3)$$

Proof. Appendix A.2 □

Remark 2.3.2. The significance of this result is that it makes the sieve regressor function of Eq. (2.2.13) available in closed-form. Indeed, in sharp contrast to the large class of parametric option pricing models of Heston (1993) and Duffie et al. (2000), no numerical integrations are required to compute an option price, which significantly facilitates the optimization problem Eq. (2.2.13). Moreover, Appendix A also provides closed-form gradients and second derivatives of the prices $P_Y(f_K, \mathbf{Z})$. And finally, having closed-form price estimates additionally simplifies the ultimate objective of computing integrated portfolios of $P_Y(f_K, \mathbf{Z})$.

The sieve put option price in Eq. (2.3.2) has an intuitive interpretation. Rearranging equation (2.3.2), one obtains

$$P_Y(f_K, \mathbf{Z}) = \kappa e^{-r\tau}\Phi(d(\mathbf{Z})) - S_0 e^{-r\tau+\mu(\mathbf{Z})} e^{\sigma(\mathbf{Z})^2/2}\Phi(d(\mathbf{Z}) - \sigma(\mathbf{Z}))$$
$$- \sum_{k=1}^{K_y} \gamma_k(B, \tau)\left[\frac{1}{\sqrt{k}} H_{k-1}(d(\mathbf{Z}))\phi(d(\mathbf{Z})) + S_0 e^{-r\tau+\mu(\mathbf{Z})} I_k^*(d(\mathbf{Z}))\right]. \quad (2.3.4)$$

Inspection of equation (2.3.4) shows that choosing

$$\sigma(\mathbf{Z}) \equiv \sigma\sqrt{\tau}, \quad \mu(\mathbf{Z}) \equiv (r - q - \sigma^2/2)\tau \quad (2.3.5)$$

will cause the leading term in equation (2.3.4) to become $P_{BS}(\sigma, \mathbf{Z}) \equiv \kappa e^{-r\tau}\Phi(d(\mathbf{Z})) - S_0 e^{-q\tau}\Phi(d(\mathbf{Z}) - \sigma\sqrt{\tau})$, where q is the dividend yield, and where the function $d(\mathbf{Z})$ from equation (2.2.5) is now $d(\mathbf{Z}) = (\log(\kappa/S_0) - (r - q - \sigma^2/2)\tau)/(\sigma\sqrt{\tau})$. The value σ is a tuning parameter in the sieve framework and is chosen to be equal to the average implied volatility of the observed option cross-section.

This is the familiar option pricing formula of Black and Scholes (1973). Therefore, the choice of $\mu(\mathbf{Z})$ and $\sigma(\mathbf{Z})$ above result in a sieve approximation with leading term given by the Black-Scholes formula, that is,

$$P_Y(f_K, \mathbf{Z}) = P_{BS}(\sigma, \mathbf{Z}) \\ - \sum_{k=1}^{2K_y} \gamma_k(B,\tau) \left[\frac{\kappa e^{-r\tau}}{\sqrt{k}} H_{k-1}(d(\mathbf{Z}))\phi(d(\mathbf{Z})) + S e^{-q\tau - \sigma^2\tau/2} I_k^*(d(\mathbf{Z})) \right]. \tag{2.3.6}$$

This formula can be interpreted as "centering" the sieve at Black-Scholes, and then supplementing it with higher-order "correction" terms.[13] As the sample size n increases, the number of correction terms K_y and K_τ also increase,[14] albeit at a slower rate than n. Thus, the more data one has, the more complex the sieve option pricer is permitted to be relative to Black-Scholes.

If the $\gamma_k(B,\tau)$ terms for $k \geqslant 1$ above are nonzero in the data, then we can regard this as evidence against the Black-Scholes model. In particular, it has been well-documented that conditional distributions of asset prices contain substantial volatility, skewness, and kurtosis that the Black-Scholes model is unable to capture. Modeling techniques to introduce such features into the return distribution includes the addition of stochastic volatility [Heston (1993)], as well as jumps [Bates (1996), Bates (2000), Bakshi et al. (1997), Duffie et al. (2000)]. The simulation study in Section 2.6 explores how these continuous time parametric features feed into the coefficients of the Hermite expansion and shows that low-order expansion terms (order 4 to 6) are quite capable of fitting the conditional distributions implied by complicated stochastic volatility and jump specifications.

[13] Recently, Kristensen and Mele (2011), Xiu (2011), and León and Sentana (2009) have employed Hermite polynomials in a parametric option pricing setting. The formulas derived here differ in that they are the result of a nonparametric sieve least squares framework.

[14] Recall that the $\gamma_k(B,\tau)$ terms also contain expansions.

2.4 Consistency

The critical feature of $P_Y(f, \mathbf{Z})$ in Eq. (2.2.7) is that it generates shape-conforming option prices for any τ. It does so by indexing state-price densities with τ, which appears as a conditioning variable. A straightforward extension of Eq. (2.2.7) is to permit a state-price density with arbitrary conditioning information, $f(Y|\mathbf{X})$, where $\mathbf{X} \subseteq \mathbf{Z}$ and contains τ. For example, one could have $\mathbf{X} = (\tau, r)$ to accommodate a risk-free rate term structure that does not match the maturities of observed option prices. Allowing for general $\mathbf{X}$ is instructive in order to see how the rate of convergence is slowed by the dimension of the conditioning variable $\mathbf{X}$. Therefore, this section establishes the asymptotic theory for the extension that allows for arbitrary conditioning information in the SPD.

A summary of the theoretical results developed in this section is as follows. First, I move from Gallant-Nychka joint density spaces to conditional density spaces (the norm changes), and from conditional spaces to option price spaces (with another norm change). The theoretical contribution of this section is to show that each of these transitions corresponds to a Lipschitz map between function spaces. Thus, the complexity of the option price spaces, as measured by the $L^2(\mathbb{R}^{1+d_x}, \mathbb{P})$ metric entropy from empirical process theory, is completely determined by the complexity of the Gallant-Nychka joint density spaces, which are Sobolev subspaces with known covering numbers. Hence, one can apply existing general theorems from the sieve estimation and inference literature to obtain convergence and asymptotic distribution results.

2.4.1 Consistency of Sieve Option Prices and State-Price Densities

With $\widehat{\beta}_n$ in hand, estimated option prices are simply given by $P_Y(\widehat{\beta}_n, \mathbf{Z}) \equiv \widehat{P}_Y^{K_n}$, which has the closed-form expression stated in Proposition 1. This subsection establishes that $\|\widehat{P}_Y^{K_n} - P_Y^0\|_2 \xrightarrow{p} 0$ as $n \to 0$, where the consistency norm is the $L^2(\mathbb{R}^{d_x}, \mathbb{P})$ norm defined below in Eq. (2.4.1).

The asymptotic results developed in the remainder of this section make use of Sobolev spaces and associated norms. Detailed definitions of these spaces are given in Appendix A.1.1. Under those definitions, the sieve spaces of conditional densities from Secion 2.2.4 are assumed to be subspaces of $W^{m,1}(\mathbb{R}^{d_u})$ stated in Definition A.1.1.

The results in this section refer to the following norms: The option price consistency norm is

$$\|P_{Y,1} - P_{Y,2}\|_2^2 = \mathbb{E}\{[P_{Y,1}(\mathbf{Z}) - P_{Y,2}(\mathbf{Z})]^2 W(\mathbf{Z})\} = \int [P_{Y,1}(\mathbf{Z}) - P_{Y,2}(\mathbf{Z})]^2 W(\mathbf{Z}) \mathbb{P}(d\mathbf{Z}) \tag{2.4.1}$$

i.e. the $L_2(\mathbb{P}, W)$-norm on the space of option prices $\mathcal{P}$ that are obtained by integration against some $f \in \mathcal{F}$. The state-price density consistency norm is $d(f_1, f_2) \equiv \|f_1 - f_2\|_{m,1}^2$.

The consistency proof requires some assumptions and a few preliminary results.

Bounded Stock Prices

When state-price densities are close, then asset prices computed off those densities should be close. This intuition is formalized in the following assumption.

Assumption 2.4.1. *(Locally Uniformly Bounded Stock Prices). Given any* $f_0 \in \mathcal{F}$*, there exists an* $\|\cdot\|_{m,1}-$ *open neighborhood* U *containing* f_0 *and a constant* M *(possibly depending on* U*) such that*

$$\sup_{f \in U} |S(f, \mathbf{Z})| \leqslant M \qquad \mathbb{P} - a.s.,$$

where

$$S(f, \mathbf{Z}) = e^{-r\tau} \int S_0 e^{\mu(\mathbf{Z}) + \sigma(\mathbf{Z})Y} f(Y|\mathbf{X}) dY = e^{-r\tau} \int S_T f(Y|\mathbf{X}) dY$$

denotes the price of a stock given a candidate SPD f*.*

Assumption 2.4.1 is a technical condition that is required in order for certain arguments in the asymptotic theory to go through and has little bearing on practical applications. In particular, it is easy to check within optimization routines that this constraint is never close to being violated.

I also make the following assumption.

Assumption 2.4.2. *Assume*

(i) $\{p_i, \mathbf{z}_i\}_{i=1}^n$ *are i.i.d. draws from* $Y = (P, \mathbf{Z})$ *with* $\mathbb{E}|Y|^{2+\delta} < \infty$ *for some* $\delta > 0$*, and* $\mathbb{E}[W(\mathbf{Z}_i)] < \infty$*.*

(ii) The true state-price density $f_0 \in \mathcal{F}$ *satisfies* $P = \mathbb{E}[P_Y(f_0, \mathbf{Z})|\mathbf{Z}]$*.*

Assumption 2.4.2 is standard and very mild. It says that the options are observed with conditional mean-zero errors with bounded $2 + \delta$ moments.

Taken together, Assumptions 2.4.1 and 2.4.2 imply a number of useful properties that are summarized in several Lemmas that I prove in Appendix A.1.1. These properties are used to establish the following consistency result.

Proposition 2. *(Consistency) Under Assumptions 2.4.1 and 2.4.2,* $d(\widehat{f}_n, f_0) \xrightarrow{p} 0$ *and* $\|\widehat{P}_Y^{K_n} - P_Y^0\|_2 \xrightarrow{p} 0$.

Proof. Appendix A.2 □

2.4.2 Rate of Convergence

The ultimate aim is to derive asymptotic inference procedures for certain option portfolios. To implement such procedures, one requires knowledge of the rate of convergence of $\|\widehat{P}_Y^{K_n}(\mathbf{Z}) - P_Y^0(\mathbf{Z})\|_2 \xrightarrow{p} 0$.

The rate of convergence of the sieve option prices depends on notions of size or complexity of the space of admissible option pricing functions as measured by the latter's bracketing numbers. Note that each candidate option price $P_Y(f, \mathbf{Z})$ is uniquely identified by the state-price density f (Lemma A.1.5). In turn, $f \in \mathcal{F}$ is the target of a Lipschitz map with preimage $f^{Y,X} = h^2 + \varepsilon_0 h_0$, a Gallant-Nychka density (Lemma A.1.6). The Gallant-Nychka class of densities requires h to reside in $\mathcal{H}$, a closed Sobolev ball of some radius $\mathcal{B}_0$.[15] The rate result obtained below hinges on the observation that the collection of possible option prices,

$$\mathcal{P} \equiv \{P_Y \ : \ P_Y(\mathbf{Z}) = P_Y(f, \mathbf{Z}) \text{ for some } f \in \mathcal{F}\},$$

is ultimately Lipschitz in the index parameter $h \in \mathcal{H}$. Therefore, the size and complexity of $\mathcal{P}$, as measured by its $L^2(\mathbb{R}^{d_x}, \mathbb{P})$ bracketing number, is bounded by the

[15] For further details, see the Online Appendix as well as Gallant and Nychka (1987).

covering number of the Sobolev ball $\mathcal{H}$ (see Van Der Vaart and Wellner (1996)). The following assumptions are used in the proof of the rate result below.

Assumption 2.4.3. *$\sigma(\mathbf{Z}) \equiv \mathbb{E}[e|\mathbf{Z}]$ and $W(\mathbf{Z})$ are bounded, where $e = P - P_Y^0(\mathbf{Z})$.*

Assumption 2.4.4. *The deterministic approximation error rate satisfies*

$$\|h - \pi_{K_n} h\|_{m_0+m,2,\zeta_0} = O(K_n^{-\alpha})$$

for some $\alpha > 0$, where $h \in \mathcal{H}$ and its orthogonal projection $\pi_{K_n} h \in \mathcal{H}_{K_n}$ are defined in Definitions A.1.2 and A.1.3, and where $K_n \equiv [K_y(n)+1][K_{x,1}(n)+1]\ldots[K_{x,d_x}(n)+1]$ denotes the total number of series terms for functions in H_{K_n}.

Assumption 2.4.5. *For state-price densities in $W^{m,1}(\mathbb{R}^{d_u})$, we have $m \geqslant d_u + 2$.*

Assumption 2.4.3 is mild and commonly adopted in the literature (see Chen (2007)). Assumption 2.4.4 takes as given the deterministic approximation error rate, and Assumption 2.4.5 imposes additional smoothness in order to invoke Sobolev imbedding theorems (see Adams and Fournier (2003)).

Proposition 3. *Let $\widehat{P}_Y(\mathbf{Z}) \equiv P_Y(\widehat{f}_n, \mathbf{Z})$, where $\widehat{f}_n$ solves (2.2.10), and let $P_Y^0 \equiv P_Y(f_0, \mathbf{Z})$ denote the true option price. Under Assumptions 2.4.1, 2.4.2, 2.4.3, and 2.4.4,*

$$\|\widehat{P}_Y - P_Y^0\|_2 = O_P(\varepsilon_n), \qquad \textit{where } \varepsilon_n = \max\{n^{-(m_0+m)/(2(m_0+m)+d_u)}, n^{-\alpha d_u/(2(m_0+m)+d_u)}\}.$$

Proof. Appendix A.2. □

Coppejans and Gallant (2002) provide conditions under which $\alpha = (m_0 + m)$ in the univariate density case ($d_u = 1$) using a chi-squared norm. If this rate extends to $\alpha = (m_0 + m)/d_u$ in the multivariate case, the above rate simplifies to the optimal

$$\varepsilon_n = O_P(n^{-\overline{m}/(2\overline{m}+d_u)}),$$

where $\overline{m} \equiv m_0 + m$, implying that the entropy and approximation error rates in Proposition 3 balance out.

2.5 Inference for Option Portfolios

I now turn to quantifying the precision of option portfolios that use the estimated option prices $\widehat{P}_Y^{K_n}$ just derived. Many such portfolios fall into the following class of functionals that take the function P_Y as inputs and return a real number. Therefore, the general sieve functional inference framework of Chen et al. (2013) can readily by applied.

Split $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$. The prime example is $\mathbf{Z}_1 = \kappa$, which includes the large class of functionals used in option hedging that integrate option prices over strikes. While the results in this subsection apply more generally, for concreteness this discussion will consider linear functionals of the form

$$\Gamma(P_Y) \equiv \Gamma_{\mathbf{Z}_2}(P_Y) = c(\mathbf{Z}) + \int_{\mathcal{Z}_1} \omega(\mathbf{Z}_1, \mathbf{Z}_2)[P_Y(\mathbf{Z}_1, \mathbf{Z}_2) + b(\mathbf{Z}_1, \mathbf{Z}_2)]d\mathbf{Z}_1. \qquad (2.5.1)$$

This general functional includes so-called weighted integration functionals as well as evaluation functionals, or combinations of both (this terminology is borrowed from Chen et al. (2013)). The following examples serve to illustrate the flexibility of this functional.

Example 2.5.1. *To compute the Synthetic Variance Swap (SVS) of Carr and Wu*

(2009) at horizon τ, one has for $\mathbf{Z}_1 = \kappa$, $\mathbf{Z}_2 = \mathbf{Z}_{-\kappa}$ and by put-call parity[16]

$$\begin{aligned} SVS(\tau) &= \Gamma_{\mathbf{Z}_2}(P_Y) \\ &= \frac{2}{\tau}\int_{-\infty}^{F(\mathbf{Z})} e^{r\tau}\frac{1}{\kappa^2}P_Y(\mathbf{Z})d\kappa + \frac{2}{\tau}\int_{F(\mathbf{Z})}^{\infty} e^{r\tau}\frac{1}{\kappa^2}C_Y(\mathbf{Z})d\kappa \qquad (2.5.2) \\ &= \int_{\mathcal{K}} \omega(\kappa, \mathbf{Z}_2)[P_Y(\kappa, \mathbf{Z}_2) + b(\mathbf{Z})]d\kappa \\ &= \int_{\mathcal{K}} \omega(\mathbf{Z})[P_Y(\mathbf{Z}) + b(\mathbf{Z})]d\kappa \end{aligned}$$

where $F(\mathbf{Z}) = S_0 e^{(r-q)\tau}$ denotes the forward price, and where

$$\omega(\mathbf{Z}) = e^{-r\tau}\frac{2}{\tau\kappa^2}, \qquad b(\mathbf{Z}) = 1[\kappa > F(\mathbf{Z})][S_0 e^{-q\tau} - \kappa e^{-r\tau}],$$

and $c(\mathbf{Z}) = 0$.

Example 2.5.2. *Bakshi et al. (2011) consider the exponential claim on integrated variance proposed in Carr and Lee (2008) given by*

$$\Gamma(P_Y) \equiv e^{-r\tau}E^{\mathbb{Q}}\left[\exp\left(-\int_0^T \sigma_t^2 dt\right)\Big|\mathbf{Z}\right] \qquad (2.5.3)$$

$$\approx e^{-r\tau} + \int_{-\infty}^{S_0} \omega(\mathbf{Z})P_Y(\mathbf{Z})d\kappa + \int_{S_0}^{\infty} \omega(\mathbf{Z})C_Y(\mathbf{Z})d\kappa \qquad (2.5.4)$$

$$= c(\mathbf{Z}) + \int_{\mathcal{K}} \omega(\mathbf{Z})[P_Y(\mathbf{Z}) + b(\mathbf{Z})]d\kappa \qquad (2.5.5)$$

[16] $\mathbf{Z}_{-\kappa}$ denotes all the values of $\mathbf{Z}$ that exclude the component κ.

where

$$\omega(\mathbf{Z}) = \frac{\frac{8}{\sqrt{14}}\cos\left(\arctan(1/\sqrt{7}) + \frac{\sqrt{7}}{2}\ln\left(\frac{K}{S_0}\right)\right)}{\sqrt{S_0}K^{3/2}}$$

$$b(\mathbf{Z}) = 1[\kappa > S_0][S_0 e^{-q\tau} - \kappa e^{-r\tau}]$$

$$c(\mathbf{Z}) = e^{-r\tau}$$

Many more functionals of option prices fall under this framework. For instance, the large class of so-called "model-free" volatility measures that are widely used in the literature, involve some type of weighted integral of option prices across strikes. See, for example, Carr and Wu (2009), Jiang and Tian (2005), Britten-Jones and Neuberger (2000), and Aït-Sahalia et al. (2012) and the many references therein. The Carr and Wu (2009) SVS-type portfolios will be the subject of this paper's empirical application below. Note that because of put-call parity, the $b(\mathbf{Z})$ term in Eq. (2.5.1) can serve to create call options from the put pricing function.

The goal is now to establish the asymptotic distribution of $\Gamma(\widehat{P}_Y)$. This will permit the construction of (pointwise) confidence intervals on the wide variety of option portfolios described in the preceding examples, which includes the class of model-free option-implied measures. It is natural to think of the estimated option pricing function $\widehat{P}_Y$ as being indexed by a finite-dimensional parameter, i.e. $\widehat{P}_Y(\mathbf{Z}) = P_Y(\widehat{\beta}_n, \mathbf{Z})$ pointwise in $\mathbf{Z}$. Hence $\Gamma(\widehat{P}_Y)$ is indexed by $\widehat{\beta}_n$, which suggests the use of the standard parametric delta-method for the derivation of the asymptotic distribution of $\Gamma(\widehat{P}_Y)$. This intuition turns out to be correct, but only if $K_n = (K_y(n)+1)(K_\tau(n)+1)$ has been chosen appropriately. The Monte Carlo simulations below confirm that incorrect choices of K_n (i.e. either too large or too small) yield incorrect asymptotic distributions. However, the simple procedure of selecting K_n by minimizing the BIC turns out to perform quite well.

With a correct choice of K_n in hand, inference on $\Gamma(\widehat{P}_Y)$ by parametric delta-method is numerically equivalent to nonparametric sieve inference. This is the result of Proposition 4 below, whose proof involves verifying some Donsker properties in order to invoke the theorems of Chen et al. (2013).

Specifically, let $\Xi_i \equiv (P_i, \mathbf{Z}_i)$ denote observations on option prices and characteristics, and define $\ell(\beta, \Xi) \equiv -\frac{1}{2}[P_i - P_Y(\beta, \mathbf{Z}_i)]^2 W_i$. The following assumption is made.

Assumption 2.5.1.

(i) *The smallest and largest eigenvalues of R_{K_n} are bounded and bounded away from zero uniformly for all K_n.*

(ii) $\lim_{K_n \to \infty} \| \frac{\partial \Gamma_{\mathbf{z}}(P_Y^0)}{\partial P_Y} [\frac{\partial P_Y^{K_n}}{\partial \beta}] \|_E^2 < \infty$.

(iii) $\|v_n^* - v^*\|_2 = O(n^{-\beta})$ *for* $\beta > \frac{1}{2} - \frac{2\alpha d_u}{2(m_0+m)+d_u}$, *where* v_n^* *and* v^* *are the Riesz representors defined in Appendix B.2.1.*

(iv) $\|v_n^*\| / \|v_n^*\|_{sd} = O(1)$.

Here, R_{K_n} is the population analog of (2.5.8) below. The main result of this section is the following proposition, which enables the construction of confidence intervals or a wide array of option portfolios that fall under the functional class in Eq. (2.5.1).

Proposition 4. *Assume the conditions of Proposition 3 as well as Assumption 2.5.1. Then*

$$\sqrt{n}\widehat{V}_n^{-1/2}[\Gamma(\widehat{P}_Y) - \Gamma(P_Y^0)] \xrightarrow{d} N(0,1) \tag{2.5.6}$$

where

$$\widehat{V}_n = \widehat{G}'_{K_n}\widehat{R}_{K_n}^{-1}\widehat{\Sigma}_{K_n}\widehat{R}_{K_n}^{-1}\widehat{G}_{K_n} \tag{2.5.7}$$

and where

$$\begin{aligned}
\widehat{G}_{K_n} &= \frac{\partial\Gamma(P_Y(\widehat{\beta}_n, \mathbf{Z}))}{\partial\beta} \\
\widehat{R}_{K_n} &= -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\ell(\widehat{\beta}_n, \Xi_i)}{\partial\beta\partial\beta'} \\
\widehat{\Sigma}_{K_n} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial\ell(\widehat{\beta}_n, \Xi_i)}{\partial\beta}\frac{\partial\ell(\widehat{\beta}_n, \Xi_i)}{\partial\beta}' .
\end{aligned} \tag{2.5.8}$$

Proof. Appendix A.2. □

Remark 2.5.3. The objects in Eq. (2.5.8) are the usual quantities involved in the estimation of the variance matrix in nonlinear least squares problems. For example, if Γ represents the 1-month SVS, i.e. $SVS(1)$ of Example 2.5.1, then Eq. (2.5.6) says that the $SVS(1)$ is asymptotically normally distributed with estimated variance $\widehat{V}_n$ computed above. Moreover, this calculation can be done for any $SVS(\tau)$ for arbitrary τ, which enables the construction of SVS term structures that quantify the estimation error involved with the construction of long-term SVS's. An analysis of SVS term structures and associated estimation errors is conducted in Section 2.7.

Remark 2.5.4. Proposition 4 shows $\sqrt{n}$-consistency of functionals of the option price, whereas Proposition 3 shows a somewhat slower rate for the convergence of $\widehat{P}_Y^{K_n}$ to

P_Y^0. This is because the functionals of interest (Eq. (2.5.1)) belong to the so-called *regular* class of functionals of Chen et al. (2013). Similar $\sqrt{n}$-consistency of well-behaved functionals is obtained in Newey (1997).

Implementation Summary

1. Choose $\mu(\mathbf{Z})$ and $\sigma(\mathbf{Z})$. To center expanions around Black-Scholes, use Eq. (2.3.5).
2. Construct $P_Y(\beta, \mathbf{Z})$ from Proposition 1.
3. Choose $K_n = (K_y(n) + 1)(K_\tau(n) + 1)$ to grow slowly as $n \to \infty$, e.g. by minimizing BIC.
4. Optimize the objective function over sieve coefficients in Eq. (2.2.13), using all options from a given cross-section of options.
5. Form $\widehat{V}_n$ using Eq. (2.5.7) and use critical values from standard normal tables.

2.6 Simulations

Aside from its ease of computation, a key advantage of the estimation and inference framework developed above is its flexibility. Here I show that the sieve estimator performs well in capturing the term structures of option smiles, risk-neutral quantiles, and state-price densities when the data are generated by familiar parametric DGPs.[17] The section concludes with a Monte Carlo experiment showing good finite-sample properties of the functional estimator from Proposition 4.

The simulations in the this section refer to various subcases of the following

[17] I also show how the sieve can be used in the construction of daily or weekly time series of fixed-maturity (e.g. 30-day) implied measures. The latter is relevant to applications that use exchange-traded options with fixed expiration, which often results in daily option surfaces in which the maturities "cycle" deterministically as the expiration date approaches. For brevity, the results of these simulations are relegated to an Online Appendix to this paper.

general data generating process.

$$
\begin{aligned}
dX_t &= \left(r - q - \lambda\overline{\mu} - \frac{1}{2}V_t\right)dt + \rho\sqrt{V_t}dW_t + J_t dN_t \\
dV_t &= \kappa_v(\overline{V} - V_t)dt + \rho v\sqrt{V_t}dW_t + \left(1-\rho^2\right)^{1/2} v\sqrt{V_t}dW_t' + Z_t dN_t
\end{aligned}
\tag{2.6.1}
$$

where V_t is a stochastic volatility process, W_t and W_t' are standard Brownian motions, and κ_v, $\overline{V}$, ρ, v parametrize the volatility process' mean reversion, long-run mean, the leverage effect, and the volatility of volatility, respectively. N_t is a Poisson process with arrival intensity λ and compensator $\lambda\overline{\mu}$, where $\overline{\mu} = \exp(\mu_J + 0.5\sigma_J^2)/(1 - \mu_v - \rho_J\mu_v) - 1$. The variable $J_t|Z_t \sim N(\mu_J + \rho_J Z_t, \sigma_J^2)$ is the price jump component and $Z_t \sim \exp(\mu_v)$ is the volatility jump component. This is the well-known double-jump process, which is a special case of the general affine-jump diffusion processes treated in Duffie et al. (2000) that is nonetheless general enough to nest the celebrated models of Black and Scholes (1973), Heston (1993), and other jump-diffusions commonly used in the option pricing literature. The values of these parameters are set to those used in Andersen et al. (2012) and are given in the Online Appendix to this paper.

2.6.1 Shape-Constrained Fitting

The main paper contributes to an existing literature concerned with shape-constrained option price fitting. Specifically, Eqs. (2.2.14) and (2.2.15) show that the option pricing function P_Y (a.) is monotone in κ, (b.) convex in κ, (c.) has first derivative $e^{r\tau}\frac{\partial P_Y^K(\mathbf{Z})}{\partial\kappa}$ as a CDF, yielding limits of 0 and 1 as κ goes to 0 and $+\infty$, respectively, and (d.) restrictions (a.)-(c.) must hold for arbitrary time-to-maturity τ.

Several papers have proposed estimation and fitting methods that obey a subset of these shape constraints. For example, Yatchew and Härdle (2006) propose a nonparametric shape-constrained estimator for options along a single maturity, as do Aït-Sahalia and Duarte (2003). Garcia and Gençay (2000) use neural network methods and the Black-Scholes formula to impose structure on their option estimates. The attractive feature of these models is that they can be differentiated to obtain risk-neutral CDFs and PDFs.

Another potential value of shape-constrained option estimators is their use in applications that require a continuum of option prices in the strike dimension that extends to infinity as inputs in empirical investigations. These include studies that use the integrated portfolios of the form Eq. (2.5.1) [see e.g. Britten-Jones and Neuberger (2000), Bakshi et al. (2003), Jiang and Tian (2005), Carr and Lee (2008), Carr and Wu (2009), or their uses in e.g. Bollerslev and Zhou (2006), Bollerslev et al. (2011), Bakshi et al. (2011) among many others]. Thus, since it is well-known that option prices are only discretely observed on a truncated interval, one often requires some type of interpolation or smoothing on the range of observed discrete options, and an extrapolation beyond the truncated range of strikes. Jiang and Tian (2005) examine the numerical properties of one such interpolation and extrapolation procedure. The sieve estimator derived above provides a complementary tool that upholds the no-arbitrage shape-constraints across all maturities τ, even for maturities for which there are few or even no observations. This is done by evaluating $\widehat{P}_Y$ at a desired τ.

Figure 2.1 illustrates the estimator's adherence to the shape constraints in Eqs. (2.2.14) and (2.2.15). The figure shows that for each observed maturity in the simu-

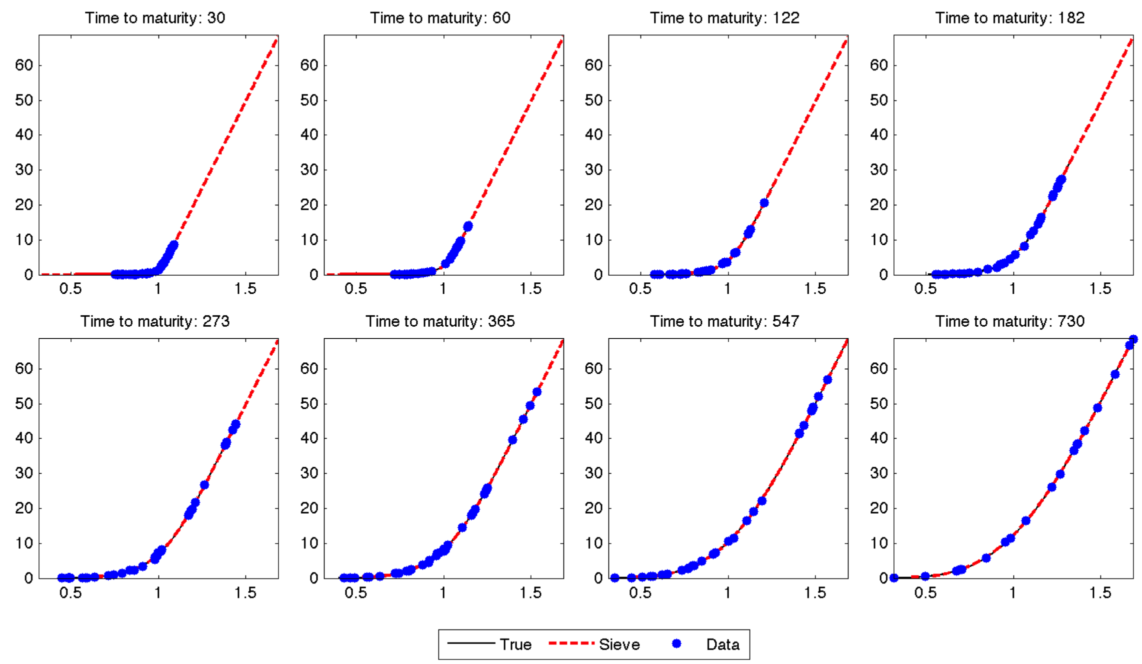

FIGURE 2.1: Shape-conforming option price estimates for multiple maturities. A dense set of true option prices is simulated from the double-jump process in Eq. (2.6.1) and are plotted in solid for eight maturities. A subset of 250 option prices is drawn from this dense set and perturbed with zero-mean measurement error (round dots). The sieve least squares problem in Eq. (2.2.13) is solved with BIC-selected $K_x = 6$ and $K_\tau = 2$ and is plotted (dash).

lated sample, option prices satisfy the shape constraints even beyond the truncated range of observed option data and asymptote to the option's intrinsic value. Thus, the sieve estimator performs the task of both interpolating between observed data, and extrapolating beyond observed data in a single estimation step across all maturities.

2.6.2 Risk-Neutral Quantiles and Densities

The purpose of shape-constrained option fitting is often to differentiate the (scaled) put pricing function once to obtain the option-implied risk neutral CDF, or differentiating it twice to obtain the state-price density. This is the subject of Jackwerth and

Rubinstein (1996), Aït-Sahalia and Lo (1998), Figlewski (2008), Birru and Figlewski (2012), as well as Bondarenko (2003) and the many references therein.

In contrast to methods that require numerical differentiation of the option pricing function, the sieve estimator in Eq. (2.2.13) delivers risk-neutral CDFs and PDFs in closed form. The closed-form expression of the risk-neutral PDF (i.e. the state-price density) is obtained by plugging Eq. (2.3.1) into Eq. (2.2.6) above. The closed-form formula for the risk-neutral CDF can be obtained by integrating the PDF and using properties of Hermite polynomials, yielding

$$\mathbb{Q}_K(S_T \leqslant \kappa|\tau) = \Phi(d(\mathbf{Z})) - \sum_{k=1}^{2K_x} \frac{\gamma_k(B,\tau)}{\sqrt{k}} H_{k-1}(d(\mathbf{Z}))\phi(d(\mathbf{Z})), \tag{2.6.2}$$

where $\mathbb{Q}_K(A) = \int_A f_K(x|\tau)dx$ is the sieve-implied risk-neutral measure obtained by integrating against the sieve state-price density. See Eq. (A.2.2) in the proof of Proposition 1 for a derivation of this expression.

Figures 2.2 and 2.3 show the term structures of risk-neutral CDFs and PDFs, using the estimated sieve coefficients obtained by solving the least squares problem in Eq. (2.2.13) on data generated by the SVJJ process in Eq. (2.6.1) and the last column of Table B.1. The CDFs cannot violate the 0 and 1 bounds at all maturities by construction of the risk-neutral PDF, since it was scaled to integrate to one for expansions of any order K. The true CDF and true PDF are plotted as well and show remarkable fit across all maturities in the panel. However, in the more extreme quantiles of the data, and particularly in the left tail, the sieve estimator begins to oscillate. This is a consequence of the bias-variance tradeoff given in the rate result of Proposition 3. In particular, the rate shows that the estimator's convergence is

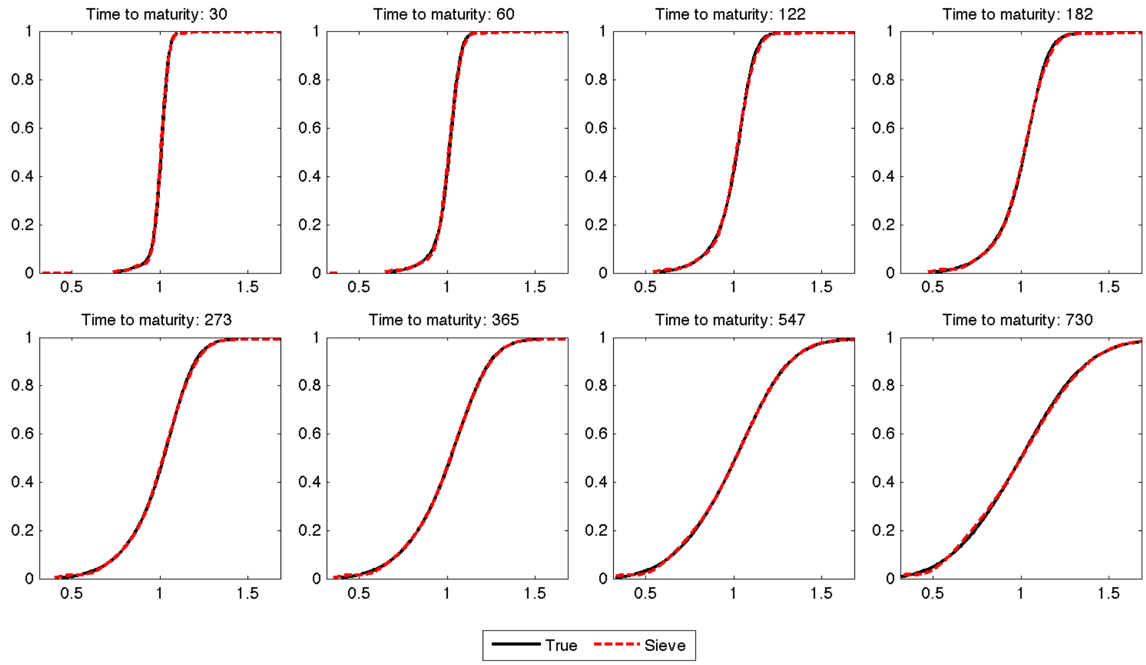

FIGURE 2.2: The term structure of risk-neutral CDFs. A dense set of true option prices is simulated from the double-jump process in Eq. (2.6.1) and are plotted in solid for eight maturities. A subset of 250 option prices is drawn from this dense set and perturbed with zero-mean measurement error (round dots). The sieve least squares problem in Eq. (2.2.13) is solved with BIC-selected $K_x = 6$ and $K_\tau = 2$. Eq. (2.6.2) is then evaluated at the estimated coefficient matrix $\widehat{B}$ (dash).

a function of bias (i.e., how quickly the sieve space fills in the parent space as K_n grows), versus the variability of the approximator, which grows with K_n. Thus, the oscillatory behavior in the plot can be decreased by reducing expansion terms, however at the cost of introducing some bias into the estimate.[18]

Finally, it is worth noting that the sieve provides remarkable fit of the entire term structure of option prices, risk-neutral CDFs, and risk-neutral PDFs, without incorporating any information about the underlying SVJJ parameters and state vectors. It can therefore be considered "model-free" in that it does not require correct

[18] An alternative approach to reducing oscillatory behavior would be to add a penalization or "regularization" term against oscillatory solutions to the least squares problem in Eq. (2.2.10). An approach of this type would fall under the penalized sieve literature and would require techniques that are beyond the scope of this paper.

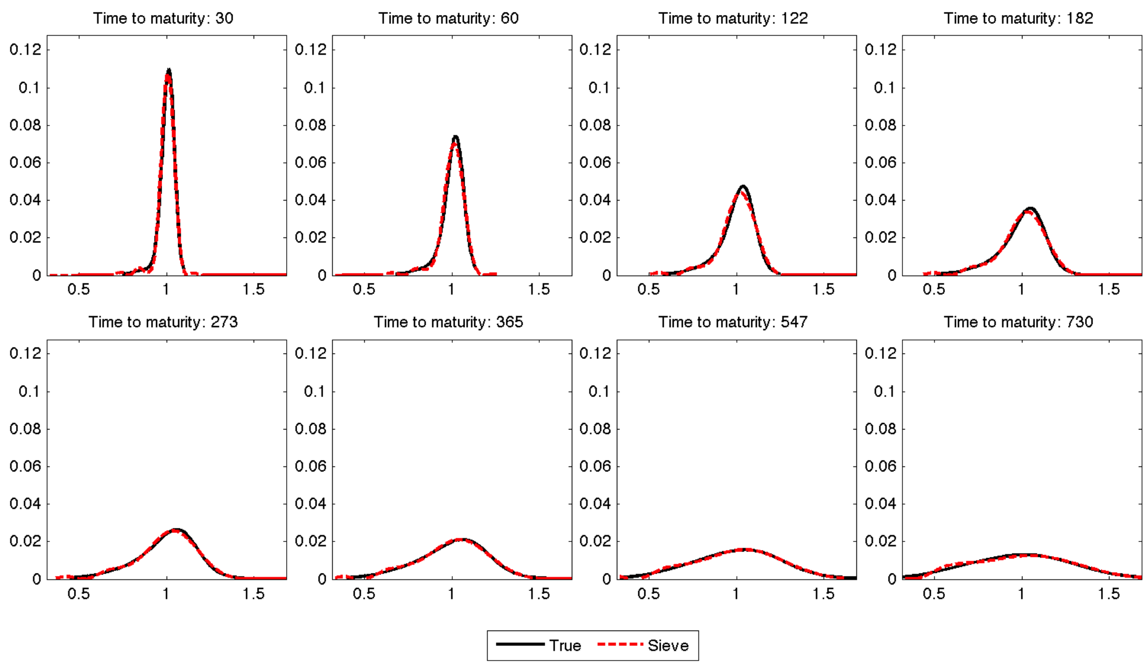

FIGURE 2.3: The term structure of risk-neutral PDFs. A dense set of true option prices is simulated from the double-jump process in Eq. (2.6.1) and are plotted in solid for eight maturities. A subset of 250 option prices is drawn from this dense set and perturbed with zero-mean measurement error (round dots). The sieve least squares problem in Eq. (2.2.13) is solved with BIC-selected $K_x = 6$ and $K_\tau = 2$. Eq. (2.3.1) is then evaluated at the estimated coefficient matrix $\widehat{B}$ (dash).

specification of the underlying dynamics.

2.6.3 Coverage

This section shows that uninformed choices of expansion terms can yield incorrect inference on portfolios of option prices. In particular, the discussion in Section 2.2 demonstrated that the number of sieve expansion terms $K_n = (K_y(n)+1)(K_\tau(n)+1)$ must grow slowly as the sample size n tends to infinity. As mentioned above, the result in Proposition 3 shows that the estimated option price $\widehat{P}_Y^{K_n}$ converges to the true option price P_Y^0 at a rate that trades off two criteria with bias and variance interpretations [see Chen (2007)]. That is, on the one hand, large choices of K_n result in lower bias, as the sieve space has more basis functions available to approximate the parent function space. On the other hand, large K_n will increase the variability of the sieve estimate, leading to oscillatory behavior. The optimal choice of K_n, therefore, balances these two influences. However, an inspection of the rate in Proposition 3 shows that the optimal choice of K_n depends on the degree of smoothness of the true state-price density f_0. Since f_0 is unknown, the rate result does not inform us of the optimal K_n. A formal theory for selecting K_n specifically for the least squares option pricing problem in Eq. (2.2.10) is therefore required, but beyond the scope of the current paper.

Instead, this section shows that selecting $K_y(n)$ and $K_\tau(n)$ by minimizing the Bayesian Information Criterion can yield effective results in terms of coverage probabilities for test statistics of the form in Proposition 4. Moreover, minimizing the BIC is computationally attractive, and has been compared favorably to the more formal cross-validation procedures in Coppejans and Gallant (2002). Given the empirical application in the next section, the focus will be on the term structure of synthetic variance swaps, $SVS(\tau)$, converted to standard deviation units, yielding the $VIX(\tau) = 100\sqrt{SVS(\tau)}$ functionals for various τ ranging from 1 month to 2 years.

The simulation design is as follows: A rich set of put option prices is simulated for each of the Heston, SVJ, and SVJJ models for maturities of 1, 2, 4, 6, 9, 12, 18, and 24 months. That is, for each of these maturities, a collection of 600 option prices is generated with moneyness ranging from 0.3 to 1.7, for a total of $600 \times 8 = 4,800$ option prices. This is considered the panel of "true" option prices within the simulation. From this true option panel, a sample of 250 options is drawn at random and perturbed with i.i.d. noise calibrated from actual options data. To ensure realistic sampling across maturities, I use the option counts across the eight maturities available for S&P 500 Index option prices on a randomly chosen day (in this case, January 5, 2005), which had a distribution of 48, 28, 27, 30, 34, 29, and 20 options at the respective maturities. This design captures the richness of available option prices at short maturities relative to long maturities.

Then, the NLLS problem in Eq. (2.2.13) is solved on the sample of $n = 250$ option prices, which yields $\widehat{\beta}_n$ that is then plugged into Eq. (2.3.2) to yield $\widehat{P}_Y^{K_n}(\mathbf{Z})$. These estimated option prices are then used to construct the Carr-Wu Synthetic Variance Swap at each maturity [see Carr and Wu (2009)], by numerically integrating

$$\widehat{SVS}(\tau) = \frac{2}{\tau}e^{r\tau}\int_0^{F(\mathbf{Z})}\frac{1}{\kappa^2}\widehat{P}_Y^{K_n}(\mathbf{Z})d\kappa + \frac{2}{\tau}e^{r\tau}\int_{F(\mathbf{Z})}^{\infty}\frac{1}{\kappa^2}\widehat{C}_Y^{K_n}(\mathbf{Z})d\kappa, \tag{2.6.3}$$

where the call prices $\widehat{C}_Y^{K_n}(\mathbf{Z})$ are obtained by put-call parity [see Example 2.5.1 above]. The associated sieve estimate of the $VIX(\tau)$ is given by

$$\widehat{VIX}(\tau) = 100\sqrt{\widehat{SVS}(\tau)}. \tag{2.6.4}$$

Because this can be done for each τ from 1 to 24 months, this procedure yields an entire estimated VIX term structure. Because I also observe a rich set of noise-free option prices within the simulation, I can compute the true VIX term structure as well. Finally, for each point along the VIX term structure, 95%-confidence intervals are constructed using the variance matrix in Proposition 4, with $\widehat{G}_{K_n}$ similar to

(2.7.5) below with adjustment for the square root and scale factor 100. The random sampling of 250 options and sieve estimation is then repeated 1,000 times, yielding 1,000 $VIX(\tau)$ confidence intervals for each τ from 1 to 24 months.

Table 2.1 shows the results of this Monte Carlo experiment for each of the Heston, SVJ, and SVJJ DGPs, and with varying expansion lengths K_y and K_τ. By definition of frequentist 95%-confidence intervals, one should expect the true $VIX(\tau)$ to lie inside the estimated confidence intervals for around 95% of the 1,000 simulated samples, for each τ. The table shows this is often the case along the entire VIX term structure and that the best performance is achieved when the BIC is permitted to select the number of expansion terms K_n (BIC selections are shaded). In particular, choices of K_n that are small or large relative to the BIC choice appear to result in overrejections, i.e. confidence intervals that are too biased or narrow to cover the true $VIX(\tau)$ in 95% of samples.

The case for allowing K_n to grow slowly with the sample size is seen strongly in the top panel of Table 2.1, where the sieve was expanded to $K_y = 3$ and $K_\tau = 1$ terms. When the option prices are generated by a Heston-type stochastic volatility process, an expansion to $(K_y, K_\tau) = (3, 1)$ terms provides near 95% coverage. In contrast, if the underlying DGP were instead to include jumps in price and/or volatility, a $(3, 1)$ expansion is clearly inadequate to capture the short end of the $VIX(\tau)$ term structure. For the SVJ DGP in particular, the 1-month true VIX was only inside 6.9% of estimated confidence intervals.

This form of overrejection is significantly improved when the BIC is allowed to choose the expansion terms. The middle panel of Table 2.1 shows expansions to

Table 2.1: Simulated Coverage Probabilities Given Affine Jump-Diffusion DGP. A dense set of true option prices is simulated from the double-jump process in Eq. (2.6.1) with parameters from Table B.1, from which a true $VIX(\tau)$ for each of the τ given in the displayed horizons is computed. Then a sample of 250 option prices is drawn at random from this dense set and perturbed with zero-mean measurement error. The sieve least squares problem in Eq. (2.2.13) is solved with different choices of expansion terms K_y and K_τ. The estimated variance from Proposition 4 is then computed to construct 95% confidence intervals around the estimated $VIX(\tau)$. The process of drawing 250 options prices and computing estimated $\widehat{VIX}(\tau)$ and its confidence intervals is repeated 1,000 times, and the proportion of occasions on which the true $VIX(\tau)$ lies inside the estimated 95% confidence intervals is then recorded at each horizon τ. Shading denotes BIC selection.

	Horizon (months)							
	1	2	4	6	9	12	18	24
$K_y = 3,\ K_\tau = 1$								
Heston	0.939	0.946	0.910	0.915	0.943	0.948	0.941	0.873
SVJ	0.069	0.040	0.434	0.886	0.909	0.802	0.853	0.908
SVJJ	0.076	0.031	0.039	0.056	0.317	0.706	0.861	0.840
$K_y = 6,\ K_\tau = 2$								
Heston	0.930	0.935	0.960	0.923	0.929	0.916	0.919	0.888
SVJ	0.971	0.941	0.962	0.944	0.909	0.897	0.916	0.890
SVJJ	0.961	0.940	0.918	0.941	0.863	0.876	0.878	0.835
$K_y = 7,\ K_\tau = 2$								
Heston	0.958	0.945	0.913	0.942	0.955	0.941	0.908	0.867
SVJ	0.970	0.968	0.922	0.931	0.960	0.934	0.913	0.883
SVJJ	0.934	0.907	0.915	0.939	0.909	0.915	0.921	0.852

$(K_y, K_\tau) = (6, 2)$, which is the BIC choice for SVJ DGPs. Allowing the expansion to go from $(3, 1)$ to $(6, 2)$ improved the coverage rate from 6.9% to 97%, which is much closer to the asymptotic rejection probability of 95%.

The main takeaway from this Monte Carlo experiment is that choosing an expansion length that is too small relative to the BIC yields incorrect inference. Moreover, the rate result in Proposition 3 suggests that this is due to pronounced biases in the sieve estimator. Allowing K_n to increase with the complexity of the model is therefore necessary to avoid misspecification biases. Finally, I note that although more formal methods for selecting K_n are needed, the choice that minimizes BIC performs remarkably well in terms of coverage probabilities.

2.7 The Term Structure of Variance Swaps and Risk Premia

An active literature in financial economics is concerned with studying the variance risk premium, i.e. the compensation that investors demand for bearing return variance risk. This literature has shown that investors are averse to return variation and have historically demanded a significant, but time-varying premium for holding securities that are exposed to such risk. Moreover, the variance risk premium, although correlated with the equity risk premium, appears to identify a source of risk that is unexplained by classic risk factors.[19]

2.7.1 Construction of the VRP Term Structure

The variance risk premium is typically measured by examining the difference between some measure of expected realized variance under the physical measure and a comparable measure of expected realized variance under the risk-neutral measure over a fixed time horizon τ, where the measure of realized variance considered here is defined as follows. If F_t is the futures price of an asset, no-arbitrage and some mild regularity conditions imply that on a risk-neutral probability space $(\Omega, \mathcal{I}, \mathbb{Q})$,

[19] See, for example, Bakshi and Madan (2006), Carr and Wu (2009), Bollerslev and Todorov (2011), Bollerslev et al. (2011), Bollerslev et al. (2013), and the references therein.

F_t solves the stochastic differential equation

$$dF_t = F_{t-}\sigma_{t-}dW_t + \int_{\mathbb{R}} F_{t-}(e^x - 1)[\mu(dx, dt) - \nu_t(x)dxdt], \tag{2.7.1}$$

where σ_{t-} is a stochastic volatility process, F_{t-} is the futures price prior to a jump at time t of size $F_{t-}(e^x - 1)$, $\mu(dx, dt)$ is a counting measure, and $\nu_t(x)dx$ is a compensator. I assume for simplicity that all quantities involved satisfy the usual regularity conditions, including finite jump activity [see e.g. Jacod and Protter (2012)]. This is a very general and commonly adopted specification in the literature [see e.g. Carr and Wu (2009) and Bollerslev and Todorov (2011)]. The realized variance of this process is defined as its annualized quadratic variation, i.e.

$$RV_t(\tau) = \frac{1}{\tau}\int_t^{t+\tau} \sigma_{s-}^2 ds + \frac{1}{\tau}\int_t^{t+\tau}\int_{\mathbb{R}} x^2\mu(dx, ds). \tag{2.7.2}$$

The second term on the right-hand side is variation due to jumps, which is not hedged by the SVS portfolio. Hence, my focus in this application will be on the first term, the truncated variation

$$TV_t(\tau) = \frac{1}{\tau}\int_t^{t+\tau} \sigma_{s-}^2 ds, \tag{2.7.3}$$

which measures the continuous variation in the underlying.

The (continuous) variance risk premium then measures the difference between this quantity's physical and risk-neutral expectations,

$$VRP_t(\tau) \equiv \mathbb{E}_t^{\mathbb{P}}[TV_t(\tau)] - \mathbb{E}_t^{\mathbb{Q}}[TV_t(\tau)]. \tag{2.7.4}$$

Carr and Wu (2009) show that the second term in the right-hand side of Eq. (2.7.4) is spanned by the $SVS_t(\tau)$ portfolio given in Eq. (2.5.2) above. That is, $SVS_t(\tau) = \mathbb{E}_t^{\mathbb{Q}}[TV_{t,\tau}]$.

Portfolios of this type have been studied in recent years, but the focus has generally been on short (e.g. $\tau = 30$ day) horizons.[20] A glance at the $SVS_t(\tau)$ portfolio given in Eq. (2.5.2) above, and the option data counts in Table **??** should reveal why: Beyond $\tau = 90$ days, the availability of option prices to approximate the infinite integral in the SVS portfolio (2.5.2) drops off significantly. Any $SVS_t(\tau)$ portfolio constructed on the sparse long-run portions of the option surface should therefore be less precise than integrated portfolios constructed from the rich short-run data. But this is exactly what Proposition 4 above contributes: it provides a formal way to quantify the precision of estimates of integrated option portfolios that are constructed from sparse and possibly noisy long-run option data.

This section therefore studies the term structure of sieve-estimated $\widehat{SVS}_t(\tau)$ portfolios and its implications for studying the corresponding term structure of the variance risk premium (VRP). I now turn to estimating the two quantities involved in the construction of the $VRP_t(\tau)$ in Eq. (2.7.4).

2.7.2 The $\mathbb{Q}$-Measure: Estimating $\mathbb{E}_t^{\mathbb{Q}}[TV_t(\tau)]$

As discussed above, the $SVS(\tau)$ spans $\mathbb{E}_t^{\mathbb{Q}}[TV_t(\tau)]$ and is computed from actual option price data. The option data used are S&P 500 Index options obtained from OptionMetrics for the time period spanning January 1996 to January 2013. The usual

[20] The notable exceptions are the papers by Aït-Sahalia et al. (2012) and Fusari and Gonzalez-Perez (2012).

filters are applied to the data, i.e. options with zero bid prices are discarded, as are in-the-money options and options with maturity less than a week. This is common practice in the literature in order to mitigate effects arising from price discreteness, liquidity effects, quote staleness, and general microstructure effects. See, for example, Andersen et al. (2012). For the analysis below, I consider the construction of 9 weekly $SVS_t(\tau)$ time series for $\tau = 1, 2, 3, 4, 6, 9, 12, 18$, and 24 months-to-maturity that use options every Wednesday of the week. While a similar construction of daily or monthly time series is also possible, the weekly frequency strikes a balance between providing a sufficiently rich time series of variance swap term structures while avoiding observations that overlap too strongly, since the $SVS_t(\tau)$ is a forward-looking measure.

Because index options are sparse at both long-run maturities and very short-run maturities, and because the maturities vary from week to week, I use the sieve estimator derived in the previous sections to obtain a balanced time series of estimated SVS term structures. Note that the coefficient solution to Eq. (2.2.13) uses options across all maturities in one step and does not require second-stage $\widehat{SVS}_t(\tau)$ interpolations. That is, for each Wednesday, it uses all available option prices and characteristics $\{P_i, \mathbf{Z}_i\}_{i=1}^n$ and solves the NLLS problem in Eq. (2.2.13) for BIC-selected $K_y = 6$ and $K_\tau = 2$, yielding a parameter matrix $\widehat{B}$, from which the estimated option pricing function $\widehat{P}_Y^{K_n}$ is obtained. The call pricing function $\widehat{C}_Y^{K_n}$ is obtained by put-call parity, and the week-t $\widehat{SVS}_t(\tau)$ term structure is then numerically computed via Eq. (2.6.3) by evaluating $\widehat{P}_Y^{K_n}$ and $\widehat{C}_Y^{K_n}$ at $\tau = 1, 2, 3, 4, 6, 9, 12, 18$, and 24 months.

To compute confidence intervals on any $\widehat{SVS}_t(\tau)$, I set

$$\widehat{G}_{K_n} = \frac{2}{\tau} e^{-r\tau} \int_a^b \frac{1}{\kappa^2} \frac{\partial P_Y^{K_n}(\widehat{\beta}_n, \mathbf{Z})}{\partial \beta} d\kappa \tag{2.7.5}$$

and construct the estimated covariance matrix $\widehat{V}_n$ in Eq. (2.5.7). Note that because $\widehat{P}_Y^{K_n}$ is shape-conforming even for strikes that are unobserved, the integral discretization error can be made arbitrarily small. Similarly, because of the sieve-estimator's adherence to shape-constraints, the integration limits a and b can be set arbitrarily wide. However, to facilitate comparisons with the CBOE's VIX, I set the integration limits to exclude option prices that fall below 1 cent.

2.7.3 Measuring the Economic Value of Standard Errors on the Variance Swap Portfolio

The above procedure yields a weekly time series of nonparametrically estimated and balanced synthetic variance swap term structures, $\widehat{SVS}_t(\tau)$, along with corresponding confidence intervals obtained from the inference theory of Proposition 4 above. One way to measure the economic value of sampling uncertainty induced by noisy option prices is to examine the width of the synthetic variance swap confidence intervals relative to the synthetic variance swap itself.

To be specific, for each day t and horizon τ, I compute the 95% confidence intervals of $\widehat{SVS}_t(\tau)$. The long position in a variance swap contract receives the payoff $N(TV_t(\tau) - \widehat{SVS}_t(\tau))$, where N is the variance notional that converts variance units into US Dollar amounts. To keep with an industry standard over-the-counter variance swap, the notional is set to $N = 100,000/(2 \cdot 100 \cdot \sqrt{SVS})$.[21]

[21] See CBOE Futures Exchange (2013).

Table 2.2: Variance swap confidence intervals and ex-post payoffs for the sample period 1996-2010. The synthetic variance swap term structure $\widehat{SVS}_t(\tau)$ along with 95% confidence intervals is estimated using the sieve methods derived in Sections 2-4 in the text. The $\mathbb{P}$-measure ex-post realized analog is computed by truncating jumps from 5-minute return data according to Eq. (B.4.1). The payoff $(TV_t - \widehat{SVS}_t)$ of a hypothetical long position in the continuous variance swap is reported. Profit and loss in US dollars are computed using the variance notional given in the text.

	Variance		USD		Pct.
Maturity	SVS	95%-CI Range	SVS	95%-CI Range	CI Range
1	0.043	0.002	979,692	59,036	5.81
2	0.045	0.002	1,009,857	39,874	4.00
3	0.045	0.002	1,023,792	33,843	3.36
4	0.046	0.001	1,030,537	32,238	3.19
6	0.046	0.001	1,035,922	26,874	2.66
9	0.046	0.001	1,038,554	23,017	2.23
12	0.046	0.001	1,039,255	23,692	2.27
18	0.045	0.001	1,037,955	30,035	2.82
24	0.045	0.004	1,035,320	87,291	8.17

Table 2.2 displays the sample average of the synthetic variance swap contract in both variance units and US Dollars. Column 3 shows the 95% confidence interval width on the estimated fixed leg $\widehat{SVS}_t(\tau)$. The last column of the table shows the proportion of this confidence interval width in relation to the swap's notional value. For short and very long horizons, the 95% confidence intervals command a sizeable fraction of the swap's fixed leg; 5.81% for 1-month swaps and up to 8.17% for two-year swaps. The corresponding Dollar amounts for these values are given in columns 4 and 5. Medium horizon swaps, in contrast, appear very well estimated and account for a relatively smaller fraction of the fixed leg payout over the sample period, compared with their short and long-maturity counterparts.

Table 2.2 suggests that sampling variation can account for about 8% of the observed average payoff on 2-year synthetic variance swaps. This is largely due to the lack of observations on long-maturity options for the first half of the sample, shown in Table ??. If the confidence interval width is interpreted as a gauge of precision for the fair value of the variance swap given observed option information, then this result suggests that available options at long (2-year) horizons relatively less informative hedges of the swap contract's true value and could therefore receive a premium relative to swaps at more liquid option maturities.

2.7.4 Comparing the Sieve and CBOE VIX

For maturities $\tau = 30$ days, the Chicago Board Options Exchange publishes a discretized estimate of the synthetic variance swap, given by

$$VIX^2_{CBOE}(\tau)/10^4 = \frac{2}{\tau}e^{r\tau}\sum_{\kappa_j \leqslant F}\frac{1}{\kappa_j^2}P(\kappa_j,\tau)\Delta\kappa_j + \frac{2}{\tau}e^{r\tau}\sum_{\kappa_j > F}\frac{1}{\kappa_j^2}C(\kappa_j,\tau)\Delta\kappa_j - \frac{1}{\tau}\left[\frac{F}{\kappa_0} - 1\right], \tag{2.7.6}$$

where κ_0 is the largest observed strike below the forward price F. Note that the last term in (2.7.6) is zero when options with $\kappa_0 = F$ are observed. Because S&P 500 Index options expire on the third Friday of each month, options expiring exactly 30 days hence are not available in most instances. In such instances, the CBOE takes the two maturities that straddle 30 days, i.e. $\tau_1 < 30$ and $\tau_2 > 30$, and computes the linear interpolation

$$VIX^2_{CBOE}(30) = \omega_1 VIX^2_{CBOE}(\tau_1) + \omega_2 VIX^2_{CBOE}(\tau_2)$$

for $\omega_1 = (30 - \tau_1)/(\tau_2 - \tau_1)$ and $\omega_2 = (\tau_2 - 30)/(\tau_2 - \tau_1)$.

Table 2.3: The CBOE and Sieve VIX Term Structurs from 1996 to 2013. VIX term structures and corresponding confidence intervals are obtained for each Wednesday of the sample using the sieve estimator from the main text as well as the CBOE's discrete approximation and linear interpolation procedure.

	Mean			Standard Deviation			95% CI
	CBOE	Sieve	Diff.	CBOE	Sieve	Diff.	Frac. Days Signif.
Maturity	(1)	(2)	(1)-(2)	(4)	(5)	(4)-(5)	Diff.
1	21.3	21.3	−0.03	8.4	8.2	0.18	0.56
2	21.7	21.9	−0.18	7.6	7.8	−0.13	0.70
3	22.0	22.3	−0.31	7.5	7.4	0.11	0.74
4	22.3	22.6	−0.26	7.2	7.2	−0.03	0.70
6	22.2	22.8	−0.58	6.4	6.9	−0.52	0.74
9	22.0	23.0	−0.96	6.1	6.6	−0.45	0.82
12	21.9	23.1	−1.16	6.1	6.4	−0.37	0.82
18	22.3	23.2	−0.88	6.3	6.3	0.05	0.84
24	22.2	23.3	−1.15	6.7	6.3	0.41	0.74

It is informative to compare this volatility index with the analogous sieve estimate from Eq. (2.6.4). Using the above interpolation scheme, I compute a term structure of $VIX_t^{CBOE}(\tau)$ at fixed horizons $\tau = 1, 2, 3, 4, 6, 9, 12, 18$, and 24 months-to-maturity for each date t in the weekly sample.

An unconditional comparison of the resulting sieve and CBOE VIX term structures is given in Table 2.3, which shows that the $VIX_t^{CBOE}(\tau)$ term structure is generally lower than the sieve estimator. The difference is negligible at the 1-month (=30 day) horizon (about 3bp on average), but becomes substantial at longer horizons (about 100bp on average). This difference is primarily due to truncation of available strikes, as can be seen by comparing the theoretical formula in Eq. (2.5.2) with the approximation in Eq. (2.7.6). While the theoretical formula in Eq. (2.5.2) extends to infinity for both call and put prices, the approximation in Eq. (2.7.6) sums only over observed, positive option prices. Implicitly, Eq. (2.7.6) has set option prices with theoretical strikes beyond observed strikes to zero, biasing down the synthetic variance swap estimate. In contrast, the sieve estimator permits extrapolation into unobserved strikes in a shape-conforming way. Figure 2.4 illustrates the downward bias of the CBOE VIX for long maturity options by comparing the time series of 30-day to 365-day CBOE and sieve VIX estimates.

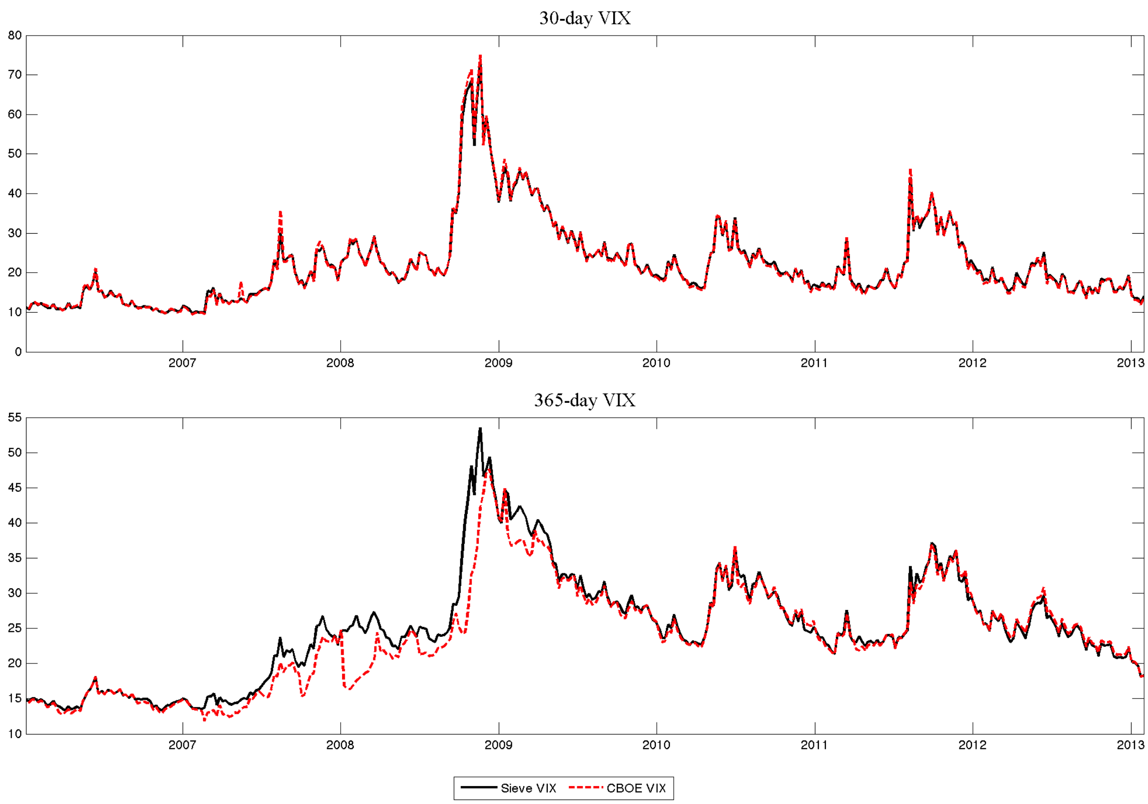

FIGURE 2.4: Short- and long-maturity time series of the sieve VIX from Eq. (2.6.4) and the CBOE VIX from Eq. (2.7.6).

The disagreement between the sieve estimate and the long-run CBOE VIX is further quantified in the last column of Table 2.3, which records the proportion of days in the sample in which the CBOE VIX lies outside the 95% confidence intervals of the sieve VIX. The proportion is clearly largest for long maturity options, suggesting clear differences between the two estimators at long maturities. Figure 2.5 shows that indeed, fluctuations in the CBOE VIX from maturity to maturity are larger than the width implied by the sieve confidence intervals. This should come as no surprise: Since the CBOE VIX only uses information from two neighboring maturities, one only needs a single sparsely observed or noisy maturity to cause the CBOE VIX to lose coherence with the CBOE VIX at other maturities on the same day. In contrast, the sieve VIX estimate uses information on all maturities to construct the term structure, which has the effect of downweighting individual poorly observed maturities.

2.7.5 The Term Structure of Continuous Variance Risk Premia

With estimates $\mathbb{E}_t^{\mathbb{Q}}[TV_t(\tau)]$ in hand, the only object needed to compute the variance risk premium is an objective forecast of $\mathbb{E}_t^{\mathbb{P}}[TV_t(\tau)]$. I follow Andersen et al. (2003) and model the long-memory properties of realized volatilities as an ARFIMA$(5, 0.4, 0)$ process.[22] The variance risk premium is then computed as in (2.7.4).

[22] For brevity, the details of this forecasting model are provided in the Online Appendix.

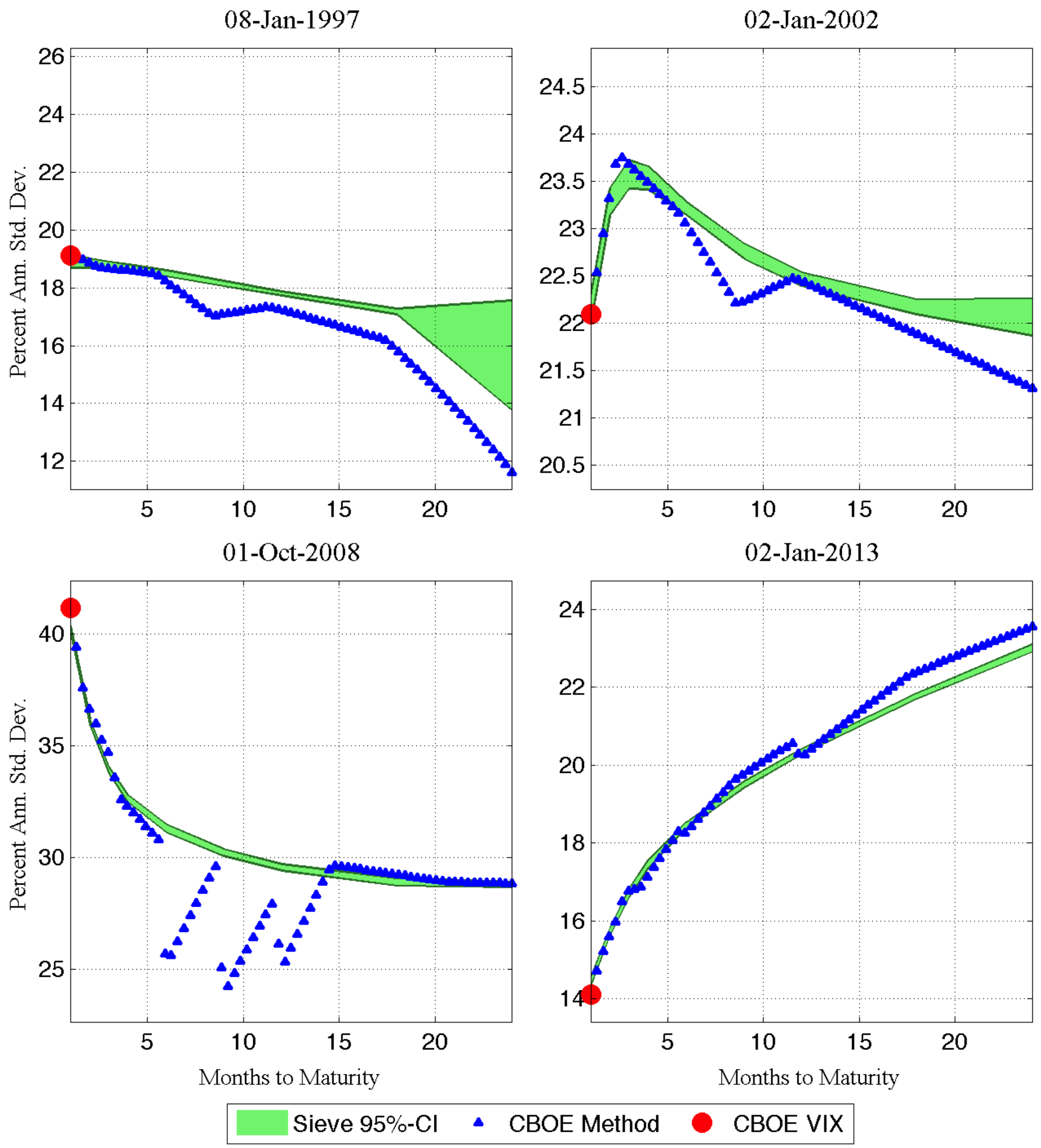

FIGURE 2.5: Term Structures for the sieve VIX and the CBOE VIX. The sieve VIX 95% confidence intervals of the estimate in Eq. (2.6.4) are plotted alongside the CBOE VIX approximations from Eq. (2.7.6) for four sample days with τ ranging from 1 month to 24 months.

Table 2.4: Summary statistics for the term structure of variance risk premia: $\mathbb{E}_t^{\mathbb{P}}[TV_t(\tau)] - \widehat{SVS}_t(\tau)$ for varying τ. The synthetic variance swap term structure $\widehat{SVS}_t(\tau)$ is estimated using the sieve methods derived in Sections 2-4 in the text. The $\mathbb{P}$-measure analog, $\mathbb{E}_t^{\mathbb{P}}[TV_t(\tau)]$, is obtained from an ARFIMA(5,0.401,0) forecast in Eq. (B.4.2).

	1996-2010					2007-2010				
Maturity	Mean	Std. dev.	Skew	Kurt	Auto	Mean	Std. dev.	Skew	Kurt	Auto
1	-0.028	0.036	-5.10	39.50	0.37	-0.045	0.059	-3.48	16.63	0.27
2	-0.031	0.035	-4.19	28.70	0.47	-0.049	0.054	-2.99	13.28	0.37
3	-0.032	0.034	-3.72	23.65	0.54	-0.052	0.051	-2.70	11.49	0.42
4	-0.033	0.033	-3.42	20.58	0.58	-0.053	0.049	-2.49	10.24	0.45
6	-0.034	0.032	-3.02	16.66	0.65	-0.055	0.046	-2.17	8.48	0.50
9	-0.035	0.031	-2.61	13.12	0.70	-0.057	0.043	-1.84	6.91	0.54
12	-0.035	0.030	-2.35	11.14	0.74	-0.058	0.041	-1.62	6.08	0.56
18	-0.035	0.029	-2.04	8.88	0.77	-0.058	0.038	-1.34	5.10	0.58
24	-0.036	0.030	-1.85	7.50	0.73	-0.059	0.036	-1.11	4.37	0.60

Table 2.4 provides summary statistics of the variance risk premium term structure of Eq. (2.7.4). The average volatility risk premium ranges from -0.028 at the 1-month horizon to -0.036 at the 2-year horizon. This finding corroborates the results of Aït-Sahalia et al. (2012) and Fusari and Gonzalez-Perez (2012). When the sample is restricted to the financial crisis period from 2007 to 2010, the variance risk premium widens significantly in magnitude and exhibits a downward-sloping term structure ranging from -0.045 to about -0.059. At the same time, the term structure of variance risk premia is itself more volatile. The decline in skewness and kurtosis and increase in persistence with τ is also consistent with Aït-Sahalia et al. (2012), who employ a different model and data set to back out a VRP term structure. First-order autocorrelations clearly show that the variance risk premium is most persistent at long horizons.

In economic terms, the magnitudes of the variance risk premium suggest that investors demand significant compensation for bearing return-variance risk and that this compensation must increase with maturity. In turbulent times, the premium is widens to about 1.6 times the average premium over the sample period. Figure 2.6 shows that this premium cannot be solely accounted for by sampling variation in option prices. The top two panels show that the variance risk premium, visualized as the gap between the displayed $\mathbb{P}$- and $\mathbb{Q}$-measure variance term structures, widens with longer maturities. Table 2.4 suggests that this is standard behavior for generic variance term structures. However, the bottom right panel suggests that on certain high-volatility days, there also appears to be significant uncertainty about the long-run variance swap price itself, although it still cannot account for the entire risk

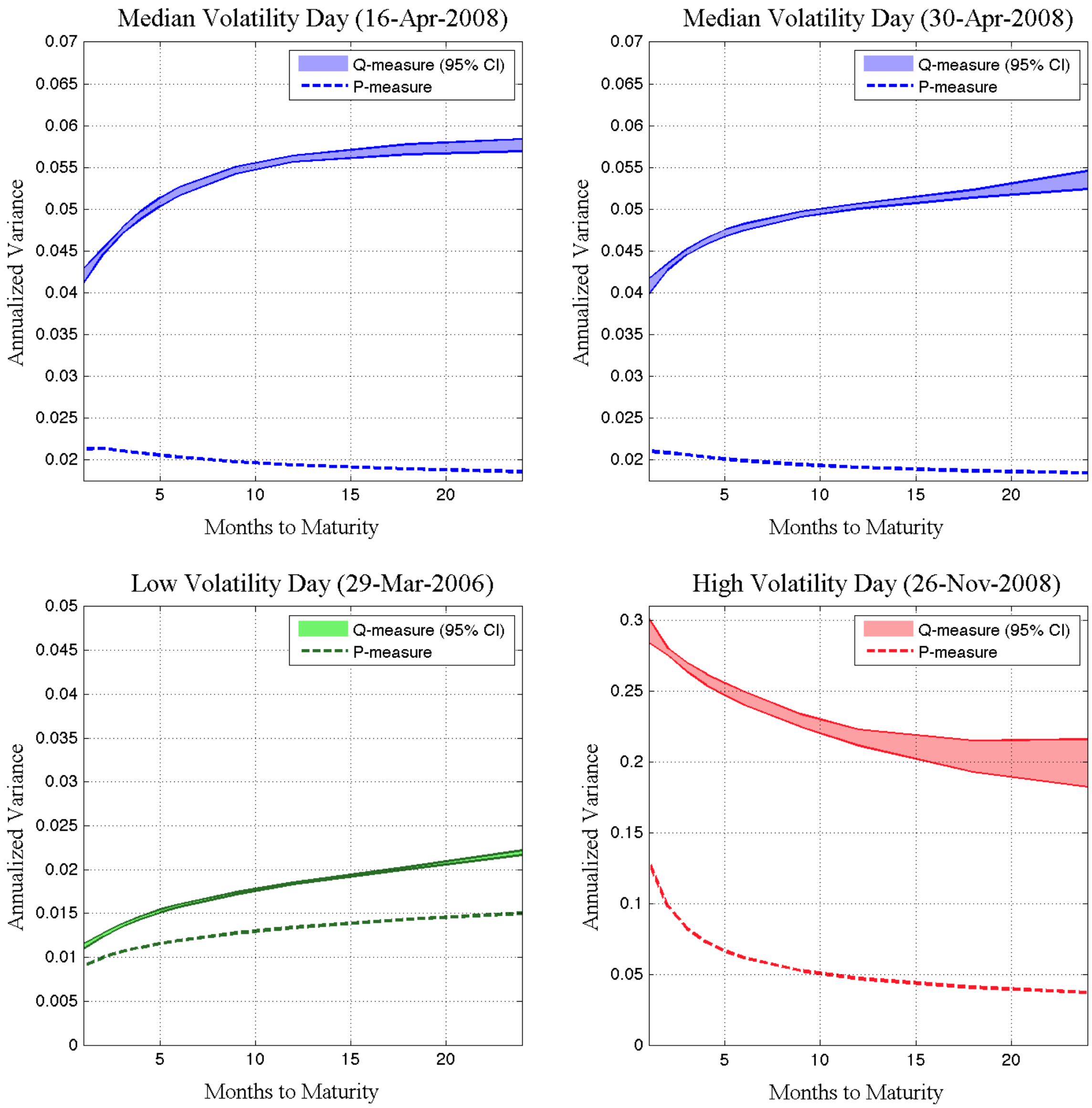

FIGURE 2.6: Term Structures for High-, Medium-, and Low-Volatility Days. Trading days are sorted by 1-month volatility as measured by the synthetic variance swap, SVS. A high-volatility day is chosen to exceed the 95th percentile of all 1-month SVS values, and a low-volatility trading day is chosen to lie below the 5th percentile of 1-month VIX values. $\mathbb{Q}$-measure SVS 95% confidence intervals and $\mathbb{P}$-measure forecasts of truncated variation, $\mathbb{E}_t^{\mathbb{P}}[TV_t(\tau)]$ are plotted for τ ranging from 1 month to 24 months.

premium.

2.7.6 Expectation Hypothesis Regressions

The balanced time series of sieve-estimated $\widehat{SVS}_t(\tau)$ can also be used to test the expectation hypothesis. Specifically, for stochastic discount factor $m_t(\tau)$, since

$$\begin{aligned} SVS_t(\tau) &= \mathbb{E}_t^{\mathbb{Q}}[TV_t(\tau)] = \mathbb{E}_t^{\mathbb{P}}[m_t(\tau)TV_t(\tau)] \\ &= \mathbb{E}_t^{\mathbb{P}}[TV_t(\tau)] + Cov_t^{\mathbb{P}}[m_t(\tau), TV_t(\tau)], \end{aligned} \tag{2.7.7}$$

the null hypothesis of no variance risk premium is equivalent to testing the null hypothesis $H_0 : Cov_t^{\mathbb{P}}[m_t(\tau), TV_t(\tau)] = 0$, i.e. no covariance between the stochastic discount factor $m_t(\tau)$ and the continuous variation of the market portfolio. That is, under H_0, one has $SVS_t(\tau) = \mathbb{E}_t^{\mathbb{P}}[TV_t(\tau)]$, so that for $\varepsilon_t(\tau)$ with $\mathbb{E}_t^{\mathbb{P}}[\varepsilon_t(\tau)] = 0$,

$$TV_t(\tau) = SVS_t(\tau) + \varepsilon_t(\tau).$$

Therefore, H_0 is equivalent to the joint hypothesis $a = 0$ and $b = 1$ in the regressions

$$TV_t(\tau) = a(\tau) + b(\tau)SVS_t(\tau) + \varepsilon_t(\tau). \tag{2.7.8}$$

The special case of $\tau = 1$ month is considered, for example, in Carr and Wu (2009). More recently, Aït-Sahalia et al. (2012) examine this regression for general τ from a model-based perspective and derive interesting interpretations of the coefficients a and b in terms of Heston model coefficients. The sieve estimate $\widehat{SVS}_t(\tau)$ can complement their approach from a nonparametric perspective.

The results of the regression (2.7.8) on the weekly sample from 1996-2010 are given in Table 2.5. Note that the $\hat{b}(\tau)$ is monotonically declining in τ and uniformly below 1. The joint hypothesis of $a(\tau) = 0 \cap b(\tau) = 1$ is firmly rejected for $\tau = 1$ month, with p-values less than 0.000 at all horizons. This result corresponds to the model-based implications of Aït-Sahalia et al. (2012). The results from Table 2.4, however, suggest that the variance risk-premium behaves differently when conditioning on different volatility regimes. To this end, I also perform the augmented regression

$$TV_t(\tau) = a(\tau) + b(\tau)SVS_t(\tau) + c(\tau)SVS_t(\tau) \times 1_{\{t \text{ is High Volatility Day}\}} + \varepsilon_t(\tau). \quad (2.7.9)$$

The interaction of the synthetic variance swap $SVS_t(\tau)$ with a dummy variable that is one during high-volatility periods and zero otherwise allows the slope coefficient on $SVS_t(\tau)$ to change according to volatility regimes. For this exercise, a trading day was considered "high-volatility" if the 30-day VIX exceeded its 67% sample quantile. The results of this regression do not change materially for different cutoffs ranging from 60% to 90% quantiles.[23] The estimates of this augmented expectation hypthesis regressions are given in the bottom panel of Table 2.5. The results are quite surprising. The magnitudes of swap coefficient are uniformly higher and are significantly closer to one.[24] In particular, the expectation hypothesis cannot be rejected for maturities ranging from 1 to 4 months during normal times, since $\hat{b}(\tau)$ cannot be distinguished from one for these maturities. However, during high volatility periods, the slope coefficient is given by $\hat{b}(\tau) + \hat{c}(\tau)$. The significantly negative sign on $\hat{c}(\tau)$ is evidence of a sizeable risk premium on high-volatility days, which drives the wedge

[23] Above 90% quantiles, the long-maturity regressions samples had relatively few observations to identify c.

[24] The sole exception is the 2-year maturity regression, which has zero explanatory power.

Table 2.5: Expectation Hypothesis Regressions. The OLS regressions $TV_t(\tau) = a(\tau) + b(\tau)SVS_t(\tau) + \varepsilon_t(\tau)$ from Eq. (2.7.8) of realized continuous variation on synthetic variance swaps are estimated for each of the horizons $\tau = 1, 2, 3, 4, 6, 9, 12, 18$, and 24 and are reported in the top panel. The same regressions are augmented in Eq. (2.7.9) to incorporate conditioning information on volatility, i.e. $TV_t(\tau) = a(\tau) + b(\tau)SVS_t(\tau) + c(\tau)SVS_t(\tau) \times 1_{\{t \text{ is High Volatility Day}\}} + \varepsilon_t(\tau)$, and are reported in the bottom panel. t-statistics on $a(\tau)$ and $c(\tau)$ are centered at 0, whereas the t-statistic on $b(\tau)$ is centered at 1. p-values report the outcome of the joint tests $a(\tau) = 0 \cap b(\tau) = 1$.

τ	$\hat{a}(\tau)$	t-stat	$\hat{b}(\tau)$	t-stat	$\hat{c}(\tau)$	t-stat	p-val	R^2	N
1	-0.001	-0.18	0.760	-1.96			0.000	0.51	764
2	0.008	2.07	0.573	-4.69			0.000	0.29	758
3	0.013	2.80	0.479	-6.20			0.000	0.21	752
4	0.018	3.52	0.409	-7.50			0.000	0.16	745
6	0.022	4.53	0.330	-9.34			0.000	0.12	733
9	0.027	5.54	0.256	-10.26			0.000	0.08	713
12	0.030	6.06	0.206	-11.47			0.000	0.06	694
18	0.031	4.27	0.216	-6.19			0.000	0.03	657
24	0.040	7.19	0.003	-12.20			0.000	0.00	619

τ	$\hat{a}(\tau)$	t-stat	$\hat{b}(\tau)$	t-stat	$\hat{c}(\tau)$	t-stat	p-val	R^2	N
1	-0.006	-1.36	0.979	-0.15	-0.193	-1.92	0.012	0.51	764
2	-0.001	-0.23	0.918	-0.49	-0.298	-2.56	0.673	0.30	758
3	0.002	0.54	0.879	-0.69	-0.341	-2.74	0.461	0.23	752
4	0.005	1.15	0.838	-0.87	-0.363	-2.60	0.190	0.19	745
6	0.012	2.16	0.685	-1.68	-0.301	-2.13	0.001	0.14	733
9	0.018	2.87	0.546	-2.68	-0.243	-2.05	0.000	0.10	713
12	0.025	3.53	0.388	-4.15	-0.154	-1.57	0.000	0.07	694
18	0.031	2.84	0.235	-3.58	-0.017	-0.19	0.000	0.03	657
24	0.040	5.26	-0.005	-7.88	0.008	0.11	0.000	0.00	619

between $\mathbb{P}$- and $\mathbb{Q}$-measure expected variation. This wedge is not detected for the shorter horizon, medium- to low-volatility days.

The compensation for variance risk in the long-run, however, does not appear as sensitive. The t-statistics report that the $\hat{b}(\tau)$ coefficients are significantly different from one, suggesting that the expectation hypothesis is rejected for those horizons even in medium- to low-volatility days. The risk premium widens even for longer maturities on high-volatility days.

2.8 Conclusion

This paper presented a nonparametric framework to help estimate option portfolios at sparsely observed maturities. The framework involved Hermite polynomial expansions of the state-price density conditional on maturity that yielded shape-conforming option surfaces in closed-form. The coefficients of the sieve option prices are computationally easy to obtain by solving a simple sieve least squares problem.

In addition, I provided a new asymptotic theory for the sieve option prices and showed them to be consistent for the true option price. I further derived its rate of convergence in terms of the deterministic sieve approximation error rate of Gallant and Nychka (1987) densities. Finally, the paper provides an asymptotic distribution theory for certain integrated portfolios of options, enabling the computation of pointwise confidence intervals for the synthetic variance swap (or VIX) term structure and related measures.

In addition to providing closed-form option prices, the framework also produced closed-form term structures of state-price densities and risk-neutral CDFs. Simulations showed that the term structures of sieve option prices, SPDs, and risk-neutral CDFs can capture a variety of data-generating processes well, and that confidence intervals obtained from the aforementioned distribution theory provide good coverage of the VIX term structure in finite samples.

An application to the term structure of the synthetic variance swap portfolios and the associated variance risk premia embedded in S&P 500 Index options and high-frequency index returns was also presented. The results showed that sampling variation in option prices can account for significant uncertainty around the variance swap's true fair value, particularly when the variance swap is synthesized from noisy long-maturity options. The term structure of variance risk premia was found to be downward-sloping and sizeable, especially on high-volatility trading days. This finding is corroborated within novel expectation hypothesis regressions that condition on volatility level information.

3

Inference on Option Pricing Models under Partial Identification

Introduction

It is widely recognized that transaction and quote data represent imperfect observations on an asset's efficient price. Available data on financial instruments might suffer from any combination of irregularities concerning time lags, sampling frequencies, price discreteness, market microstructure frictions, and a positive spread between quoted bid and ask prices. Thus, if a model imposes testable restrictions on asset prices, but asset prices do not directly correspond to observed data, then it may not be possible to recover the model's features from the data.

We examine the mismatch between model-implied asset pricing restrictions and available data in the option quote setting. While available option pricing models map underlying state variables to efficient prices, observed option bid and ask quotes effectively deliver only interval information on efficient prices. While the prevailing practice in empirical option pricing is to simply average these bids and asks and then to fit option prices to the resulting mid-quotes, we argue that the illiquidity of many deep in-the-money options induces substantial bid-ask spreads that prevent the point-identification of option model parameters. Indeed, if, as Carr and Wu (2009) remark, "[t]he mid-quote may not reflect the fair price if the bid and ask quotes are not symmetric around the fair price," then fitting option models to the mid-quote introduces a joint hypothesis problem, in the sense that rejections of option pricing models can either come from option model misspecification or mid-quote (microstructure) misspecification.

This paper takes a new approach to inference on option pricing models in the bid-ask quote setting. Rather than assuming knowledge of the structure that equates untestable functions of observed quotes to the efficient price process (of which the mid-quote is the most prominent example), we take a conservative approach of bounding moments of efficient option prices by observed bid and ask quotes. We then proceed with inference on option model parameters by leaving the relationship between the option's efficient price and the quotes otherwise unspecified. Because this relationship is a function of the market maker's price-setting schedule, we are in essence proceeding with inference without having to model the (unobserved) pricing practices of the market maker. The cost of leaving the market maker's pricing schedule unspecified is a loss of point-identification. This implies in particular that the information contained in the data-generating process (DGP) is only able to restrict the option model's parameters to a set. However, while there is a large and growing literature on inference in this type of partial identification setting,[1] option pricing models often depend on latent variables (like spot volatility) that cannot be accommodated using existing econometric techniques.

Our main theoretical contribution, therefore, is a theory of set-inference in which the moment function depends on latent spot volatility. We solve this dependence by using high-frequency data on the underlying asset to nonparametrically estimate spot volatility near the close of the trading day in a first stage. We then plug the spot volatility estimate into the pricing model and proceed with the familiar moment inequality framework of Andrews and Soares (2010) and Andrews and Shi (2014). We provide rigorous justification of this two-step inference procedure in an asymptotic setting in which the high-frequency sampling interval goes to zero sufficiently quickly as the sample size of options grows. In particular, we establish the rates at which the spot volatility estimator must converge relative to the length of the option panel. Under these rate conditions, we provide asymptotic coverage results for familiar test statistics and critical values from the partial identification literature.[2]

The framework for inference is also flexible enough to allow for additional moment equalities, which enable us to sharpen inference with information obtained from high-frequency observations on the underlying state variables. Incorporating restrictions based on information on the underlying is natural, given that certain parameters are invariant to changing the measure from the objective to the risk-neutral

[1] See Chernozhukov et al. (2007), Romano and Shaikh (2008), Andrews and Soares (2010), Andrews and Guggenberger (2009), Bugni (2010), Andrews and Shi (2014), and the many references therein.

[2] Our results pertain to the modified method of moments statistic and critical values introduced in Andrews and Soares (2010).

one. This idea has been exploited in the existing empirical options literature.[3] This literature, however, has made the empirically puzzling observation that the theoretically motivated invariance to changes of measure does not seem to hold in the data, a phenomenon dubbed the *time-series inconsistency* of option data (Broadie et al. (2007)). Motivated by these observations, we extend our moment inequality framework to allow for objective measure restrictions on leverage and volatility-of-volatility–quantities that are theoretically invariant to changes of measure. We find that time-series inconsistency is partially mitigated by relaxing the mid-quote assumption, because fitting to the option bid-ask spread allows for a wider range of option-implied leverage and volatility-of-volatility than would be possible by merely fitting to the option mid-quote.

To examine the finite-sample properties of our asymptotic coverage results, we conduct Monte Carlo simulations using a flexible stochastic volatility jump-diffusion model (Duffie et al. (2000)). The simulations confirm that replacing latent spot volatility with a high-frequency estimate preserves the finite-sample coverage probabilities of the proposed confidence sets. We also examine the tradeoff between long-span asymptotics and infill asymptotics that invariably affects our moment conditions by considering simulated option samples of different lengths, as well as the effects of incorporating objective measure restrictions on the underlying. Finally, we consider the effect of assuming that the efficient price equals the mid-quote proxy under several quote-setting DGPs and find that doing so can result in sometimes severe option model overrejections.

[3] See, for example Bakshi et al. (1997), Bates (2000), and Broadie et al. (2007).

Finally, we conduct inference on the aforementioned stochastic volatility jump-diffusion model (SVJ) and its specialization, the Heston (1993) model, using actual data on S&P 500 Index options from 1996 to 2010. We find that relaxing the mid-quote assumption results in large estimated parameter sets that leave us unable to reject the SVJ model, which contrasts with the existing findings of the empirical options literature. However, the same model is rejected when the mid-quote assumption is imposed. In addition, the estimated parameter sets reveal new and interesting relationships among option model parameters. For example, the parameter sets clearly indicate a tradeoff between the jump component's intensity and the jump-size mean and variance; option quotes appear supportive of either high-intensity-small-jumps or low-intensity-large-jump models. We also find a positive relationship between volatility's speed of mean reversion and volatility-of-volatility. Taken together, these findings point toward a greater need to study the precise identifying information contained in option panels.

The remainder of the paper is organized as follows: Section 3.1 lays down our moment inequality framework, our two-step inference procedure, the asymptotic theory, and the incorporation of additional objective measure identifying restrictions. Sections 3.2 and 3.3 present our Monte Carlo simulation and empirical results, and Section 3.4 concludes.

3.1 The Framework

3.1.1 Moment inequality restrictions on option pricing models

For each day $t \in \{1, \ldots, T\}$, we observe an option panel with N_t options, where N_t is a random integer-valued variable taking values in $\left[1, \bar{N}\right]$ for some constant $\bar{N} \geqslant 1$. Let $Z_{i,t}$ denote the characteristics of option $i \in \{1, \ldots, N_t\}$ and $Z_t = \{Z_{i,t} : 1 \leqslant i \leqslant N_t\}$. We denote the domain of $Z_{i,t}$ and Z_t by $\mathcal{Z}$ and $\tilde{\mathcal{Z}}$ respectively.[4] In our empirical application, we consider European style call and put options with $Z_{i,t} = (K_{i,t}, \tau_{i,t})$ where $K_{i,t}$ is the strike price and $\tau_{i,t}$ is the time to maturity. We suppose that Z_t is realized at the market opening and is constant throughout each trading day. At the market close, we observe the quotes, the ask price $A_{i,t}$ and the bid price $B_{i,t}$, of option i, along with the log price of the underlying X_t. We denote the midquote by $M_{i,t} = (A_{i,t} + B_{i,t})/2$.

The econometric problem here is to conduct inference for parameters of an option pricing model. We consider models with two state variables: the underlying log price X_t and the spot variance V_t of the underlying asset, taking values in $\mathcal{X} \subset \mathbb{R}$ and $\mathcal{V} \subset (0, \infty)$ respectively. We denote the option pricing function by $f(X_t, V_t, Z_{i,t}; \theta)$, $\theta \in \Theta$, where $\Theta \subset \mathbb{R}^{\dim(\Theta)}$ is a compact parameter space. The unknown parameter θ governs the risk-neutral dynamics of the underlying asset and is referred to as the risk-neutral parameter below. We call a model correctly specified if for some $\theta_0 \in \Theta$, the true option price $p^*_{i,t}$ equals $f(X_t, V_t, Z_{i,t}; \theta_0)$ for each option i and day t. In an ideal market without any friction, the observed option price should coincide with $p^*_{i,t}$ generated by a correctly specified model. However, in reality, the observed price may

[4] Note that $\tilde{\mathcal{Z}} = \bigcup_{n=1}^{\bar{N}} \mathcal{Z}^n$, where $\mathcal{Z}^n$ is the n-tuple of $\mathcal{Z}$.

deviate from the model prediction due to various frictions, such as the existence of a bid-ask spread, data synchronization errors, or simply idiosyncratic errors specific to each transaction. In contrast to the existing empirical options literature, we treat the true option price $p^*_{i,t}$ as a latent variable.

To close the econometric model, we need to link the latent true price $p^*_{i,t}$ with observed quantities. A standard approach in the literature is to consider a noisy proxy $p_{i,t}$, such as the midquote or the actual transaction price, of the latent price $p^*_{i,t}$.[5] Moreover, statistical assumptions on the pricing error $\varepsilon_{i,t} = p_{i,t} - p^*_{i,t}$ are often imposed so that standard econometric procedures are valid. Perhaps the most popular estimation procedure is the nonlinear least squares,[6] which in turn can be cast in a general conditional moment equality framework:

$$\mathbb{E}\left[\varepsilon_{i,t}|\mathcal{I}_t\right] = \mathbb{E}\left[p_{i,t} - p^*_{i,t}|\mathcal{I}_t\right] = \mathbb{E}\left[p_{i,t} - f\left(X_t, V_t, Z_{i,t}; \theta_0\right)|\mathcal{I}_t\right] = 0, \tag{3.1.1}$$

where $\mathcal{I}_t$ denotes an information set (i.e. σ-field) generated by $\{X_s, V_s, Z_s : s \leqslant t\}$ or its subsets.[7] Likelihood-based procedures, such as the one considered by Bates (2000), often impose stronger parametric assumptions which imply (3.1.1) as a consequence.

While the option pricing model is designed to describe the relationship between true option price $p^*_{i,t}$ and the state variables (X_t, V_t), (3.1.1) imposes an extra restriction on the microstructure of the option market, which links the latent value $p^*_{i,t}$

[5] For example, Bates (2000) and Broadie et al. (2007) use transaction data for S&P 500 futures options, whereas Bakshi et al. (1997), Pan (2002), Christoffersen and Jacobs (2004), Eraker (2004), Andersen et al. (2012) use midquote data for S&P 500 index options.

[6] See, for example, Bakshi et al. (1997), Bates (1991), Christoffersen and Jacobs (2004), Broadie et al. (2007).

[7] See Hayashi (2000) for a discussion in textbook form.

with its observed proxy. Conditional moment equality restrictions like (3.1.1) are often needed to justify standard econometric procedures, but it is not automatically clear on which economic ground this restriction should hold. Indeed, as Bates (2000) page 195 points out, "a fundamental difficulty with implicit parameter estimation is the absence of an appropriate statistical theory of option pricing errors." If the inference for the option pricing model also depends on the validity of (3.1.1), which is typically the case in the literature, it is important to investigate the economic content underlying (3.1.1).

To make the discussion precise, we consider a simple empirical microstructure model. We suppose that the option quotes are observed with noise, that is,

$$A_{i,t} = A^*_{i,t} + \varepsilon^A_{i,t}, \quad B_{i,t} = B^*_{i,t} + \varepsilon^B_{i,t}, \tag{3.1.2}$$

where $A^*_{i,t}$ and $B^*_{i,t}$ are the latent "efficient" ask and bid prices which would be consistent with the behavior of a rational market maker equipped with the information set $\mathcal{I}_t$.[8]

The pricing errors $\varepsilon^A_{i,t}$ and $\varepsilon^B_{i,t}$ are introduced here because we do not expect every pair of quotes in the data to reflect precisely the prediction of a theoretical model. Instead, we assume that the pricing error is zero on average:

$$\mathbb{E}\left[\varepsilon^A_{i,t}|\mathcal{I}_t\right] = \mathbb{E}\left[\varepsilon^B_{i,t}|\mathcal{I}_t\right] = 0. \tag{3.1.3}$$

[8] Since $A^*_{i,t}$ and $B^*_{i,t}$ are determined based on the information in $\mathcal{I}_t$, they are (tautologically) $\mathcal{I}_t$-measurable.

When we take the proxy $p_{i,t}$ to be the midquote, (3.1.1) holds if and only if

$$M_{i,t}^{*} = p_{i,t}^{*}, \tag{3.1.4}$$

where $M_{i,t}^{*} = (A_{i,t}^{*} + B_{i,t}^{*})/2$ is the efficient midquote. That is, the mid-quote coincides with the efficient price. However, it is well-known that this condition is violated in general. For example, based on the sequential trade model of Glosten and Milgrom (1985), Hasbrouck (2006) page 46 shows that the mid-quote generically deviates from the efficient price.

Taking $p_{i,t}$ to be the transaction price is unlikely to fulfill the condition (3.1.1) either. Following Hasbrouck (2006), we consider a simple model in which transactions are treated as a sampling mechanism of the quote price. Suppose that the transaction price $p_{i,t}$ is given by

$$p_{i,t} = (1 - \delta_{i,t})\, A_{i,t} + \delta_{i,t} B_{i,t},$$

where $\delta_{i,t} = 1$ (resp. 0) for a sell (resp. buy) order. If, conditionally on $\mathcal{I}_t$, the direction of the order $\delta_{i,t}$ is uncorrelated with the pricing errors $\varepsilon_{i,t}^{A}$ and $\varepsilon_{i,t}^{B}$, then (3.1.1) is equivalent to

$$\text{Prob}\,(\text{sell}\,|\,\mathcal{I}_t)\, A_{i,t}^{*} + \text{Prob}\,(\text{buy}\,|\,\mathcal{I}_t)\, B_{i,t}^{*} = p_{i,t}^{*}. \tag{3.1.5}$$

Hence, when using the transaction price as a proxy for $p_{i,t}^{*}$, the condition (3.1.1) is satisfied if and only if the buying propensity Prob(buy$|\mathcal{I}_t$)$= (p_{i,t}^{*} - A_{i,t}^{*})/(B_{i,t}^{*} - A_{i,t}^{*})$. At least in this simple setup, there appears to be no compelling reason for this condition, and thus (3.1.1), to hold.

Of course, the discussion above does not imply that the condition (3.1.1) is false or irrelavant. Indeed, even if this condition is violated in a strict sense, it might be a decent approximation to reality and lead to useful empirical results, as demonstrated by the existing literature. Instead, we only argue that this identification condition is likely to impose nontrivial restrictions on the microstructure of the option market; the setup specified in the above two paragraphs only serve the purpose of making this argument precise. In our view, such extra restrictions are orthogonal to the specification of risk-neutral option pricing models as well as the inference based on these models. This concern motivates us to propose an inference framework for option pricing models without assuming (3.1.1). We deem our approach below as complementary to the existing toolbox for empirical asset pricing.

We propose an alternative econometric framework based on moment inequality restrictions. We maintain (3.1.2) and (3.1.3) and suppose that

$$B^*_{i,t} \leqslant p^*_{i,t} \leqslant A^*_{i,t}, \tag{3.1.6}$$

i.e., the true option price falls in the efficient bid-ask bracket. Compared with conditions such as (3.1.4) and (3.1.5), (3.1.6) is very mild and should be satisfied by any reasonable microstructure model. Given (3.1.2) and (3.1.6), a correctly specified pricing function $f(\cdot;\theta_0)$ satisfies the following conditional moment inequalities:

$$\begin{cases} \mathbb{E}[A_{i,t} - f(X_t, V_t, Z_{i,t}; \theta_0)|\mathcal{I}_t] & \geqslant & 0 \\ \mathbb{E}[f(X_t, V_t, Z_{i,t}; \theta_0) - B_{i,t}|\mathcal{I}_t] & \geqslant & 0. \end{cases} \tag{3.1.7}$$

We observe that we do not require the observed quotes to bracket $p^*_{i,t}$ for every realization of the data, but rather the weaker condition that the bracketing only holds in expectation.

These conditional inequalities further imply a set of unconditional moment inequalities: for any nonnegative function $G(\cdot)$ on $\mathcal{X} \times \mathcal{V} \times \mathcal{Z} \times \tilde{\mathcal{Z}}$,

$$\begin{cases} \mathbb{E}\left[\sum_{i=1}^{N_t}(A_{i,t} - f(X_t, V_t, Z_{i,t}; \theta_0))G(X_t, V_t, Z_{i,t}, Z_t)\right] & \geqslant \; 0 \\ \mathbb{E}\left[\sum_{i=1}^{N_t}(f(X_t, V_t, Z_{i,t}; \theta_0) - B_{i,t})G(X_t, V_t, Z_{i,t}, Z_t)\right] & \geqslant \; 0. \end{cases} \tag{3.1.8}$$

The conditional moment equality (3.1.1), if it holds, also implies a set of moment equalities

$$\mathbb{E}\left[\sum_{i=1}^{N_t}(p_{i,t} - f(X_t, V_t, Z_{i,t}; \theta_0))G(X_t, V_t, Z_{i,t}, Z_t)\right] = 0, \tag{3.1.9}$$

while here the function $G(\cdot)$ does not have to be nonnegative. Below, we refer to the function $G(\cdot)$ as a weighting function. We sometimes write $G(x, v, z, \tilde{z})$ for $x \in \mathcal{X}$, $v \in \mathcal{V}$, $z \in \mathcal{Z}$ and $\tilde{z} \in \tilde{\mathcal{Z}}$ in order to make its arguments explicit.

We finish this section by introducing a standard moment inequality/equality framework encompassing both (3.1.8) and (3.1.9). While our main proposal is to conduct inference based on the inequalities (3.1.8), incorporting (3.1.9) into the same framework facilitates the comparison between results based on bid-ask brackets and those based on midquotes or transaction data in a unified manner.

Let $Q_t = (Q_{i,t})_{1 \leqslant i \leqslant N_t}$, $Q_{i,t} = (A_{i,t}, B_{i,t}, p_{i,t})$, denote the collection of option price data; recall that $p_{i,t}$ can be either the midquote or the transaction price depending on the application. We consider $k_I + k_E$ weighting functions $G_j(\cdot)$, $1 \leqslant j \leqslant k_I + k_E$, and suppose $G_j(\cdot)$ is nonnegative for $1 \leqslant j \leqslant k_I$. Let $k = 2k_I + k_E$. We consider an

$\mathbb{R}^k$-valued function:

$$m\left(X_t, V_t, Q_t, Z_t; \theta\right) = \left(m_1\left(X_t, V_t, Q_t, Z_t; \theta\right), \ldots, m_k\left(X_t, V_t, Q_t, Z_t; \theta\right)\right)^\top,$$

where

$$\begin{aligned} & m_j\left(X_t, V_t, Q_t, Z_t; \theta\right) \\ = & \begin{cases} \sum_{i=1}^{N_t} (A_{i,t} - f(X_t, V_t, Z_{i,t}; \theta)) G_{(j+1)/2}(X_t, V_t, Z_{i,t}, Z_t) & \text{if } \ 1 \leqslant j \leqslant 2k_I, \ j \text{ odd} \\ \sum_{i=1}^{N_t} (f(X_t, V_t, Z_{i,t}; \theta) - B_{i,t}) G_{j/2}(X_t, V_t, Z_{i,t}, Z_t) & \text{if } \ 1 \leqslant j \leqslant 2k_I, \ j \text{ even} \\ \sum_{i=1}^{N_t} (p_{i,t} - f(X_t, V_t, Z_{i,t}; \theta)) G_{j-k_I}(X_t, V_t, Z_{i,t}, Z_t) & \text{if } \ 2k_I + 1 \leqslant j \leqslant k. \end{cases} \end{aligned}$$

We then consider a collection moment inequalities and equalities given by

$$\begin{cases} \mathbb{E}[m_j(X_t, V_t, Q_t, Z_t; \theta_0)] & \geqslant & 0, \quad j = 1, \ldots, 2k_I, \\ \mathbb{E}[m_j(X_t, V_t, Q_t, Z_t; \theta_0)] & = & 0, \quad j = 2k_I + 1, \ldots, k. \end{cases} \tag{3.1.10}$$

Here, the $2k_I$ inequalities correponds to (3.1.8) and weighting functions G_j, $1 \leqslant j \leqslant k_I$; the k_E inequalities correspond to (3.1.9) associated with weight functions G_j, $k_I + 1 \leqslant j \leqslant k$.

The model (3.1.10) may only be partially identified, i.e., the solution to (3.1.10) on the parameter space Θ may not be unique. Indeed, if the equality restrictions in (3.1.10) are absent, the model is not point identified in general. The situation is illustrated in Figure 3.1 This feature of the model is in sharp contrast with aforementioned work in empirical option pricing. Moreover, even in cases where parameters are point identified by the moment equalities, the identification may be weak and could render standard inference unreliable (Stock and Wright (2000), Andrews and Cheng (2012)). Our inference procedure below does not rely on identification. Thus,

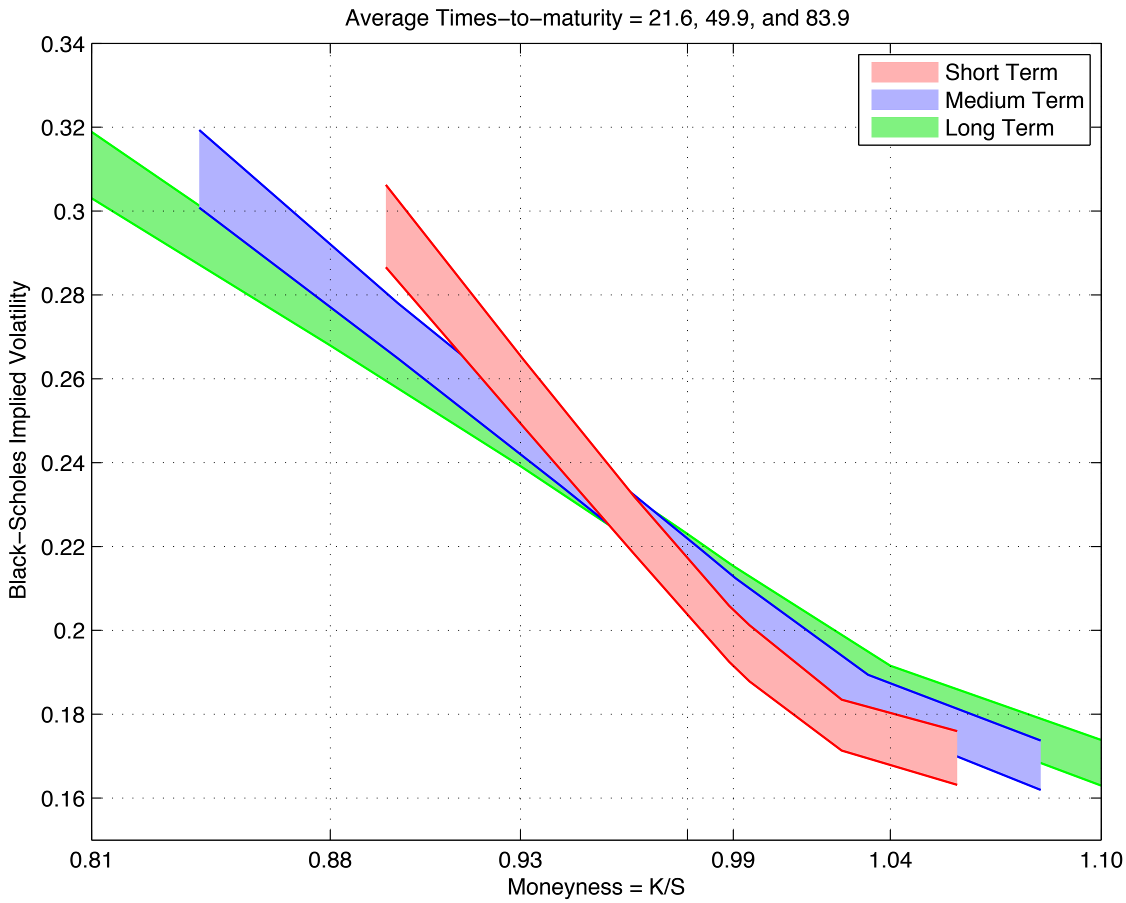

FIGURE 3.1: Set-Inference framework illustration. Time-averaged bid and ask quotes for seven strike categories and three maturity categories are plotted. Under correct specification in our framework, a true option model parameter is assumed to generate option prices that lie between the shaded bounds for all three maturities. The inference procedure delivers confidence sets that cover the true parameter with a prespecified probability.

it is naturally immune to the weak identification problem. As the weak identification problem is rarely, if ever, dealt with in prior works in empirical option pricing, our empirical analysis based on the moment equalities also complements the literature in an interesting way.

Below, we discuss the construction of confidence sets for the true risk-neutral parameter θ_0 based on the moment inequality/equality model (3.1.10).

3.1.2 A two-step inference procedure

Inference based on the moment inequality/equality model like (3.1.10) has been extensively studied in the recent econometrics literature (see, for example, Andrews and Soares (2010), Andrews and Shi (2014), Romano and Shaikh (2008), Bugni (2010), Andrews and Guggenberger (2009), Chernozhukov et al. (2007)). However, here we face a complication that is unique to financial econometric applications—the spot variance process V_t is not observable. In a nutshell, we solve this problem in a two-stage procedure. In the first stage, we nonparametrically estimate the spot variance V_t of the log-price at the market close using intraday high frequency data. In the second-stage, we construct confidence sets by using the generalized moment selection (GMS) method of Andrews and Soares (2010), treating the first-stage estimate of the spot variance as if it were the true spot variance. We provide a rigorous theoretical justification for the validity of this two-stage procedure.

To further the discussion, we need to introduce more notations to decribe the intraday data. We suppose that, on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$, the log-price of the underlying asset follows an Itô semimartingale with the form:

$$X_t = X_0 + \int_0^t b_s\, ds + \int_0^t \sigma_s\, dW_s + \int_0^t \int_{\mathbb{R}} \delta(s,z) \underline{\mu}(ds, dz), \tag{3.1.11}$$

where b_t is locally bounded predictable process, $\delta : \Omega \times \mathbb{R}_+ \times \mathbb{R} \mapsto \mathbb{R}$ a predictable function, W_t a standard Brownian motion, $\underline{\mu}$ is a Poisson measure with compensator $\underline{\nu}(dt, dz) = dt \otimes \lambda(dz)$ for some σ-finite measure λ on $\mathbb{R}$. We refer to the process σ_t as the spot volatility, which is associated with the spot variance V_t by $V_t = \sigma_t^2$. We refer to trading day t, $t = 1, \ldots, T$, as the interval $[\underline{t}, t]$, where $\underline{t}$ is the opening time of the market. We suppose that intraday data of the log price are regularly sampled at discrete times $\underline{t} + i\Delta_n$, $i = 0, \ldots, n$, where $\Delta_n = (t - \underline{t})/n$.

We estimate the spot variance V_t nonparametrically with the following jump-robust estimator (Mancini (2009)):

$$\widehat{V}_{n,t} = \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} (X_{t-(j-1)\Delta_n} - X_{t-j\Delta_n})^2 1_{\{|X_{t-(j-1)\Delta_n} - X_{t-j\Delta_n}| \leqslant \alpha \Delta_n^{\varpi}\}}, \tag{3.1.12}$$

where $1_{\{\cdot\}}$ denotes the indicator function, $\alpha > 0$ and $\varpi \in (0, 1/2)$ are constants, and k_n is a sequence of integers satisfying $k_n \to \infty$ and $k_n \Delta_n \to 0$. The asymptotic behavior of this estimator under the fill-in asymptotics (i.e. $\Delta_n \to 0$) is well known: under mild regularity conditions, $\widehat{V}_{n,t}$ consistently estimates V_t for each t as the sampling interval goes to 0.[9]

We construct confidence sets by inverting tests of the null hypothesis that θ is the true value for each $\theta \in \Theta$, as is standard in the partial identification literature. Let $S_{n,T}(\theta)$ be a test statistic and $c_{n,T}(\theta, 1-\alpha)$ be a corresponding critical value for a test with nominal significance level α. A nominal level $1 - \alpha$ confidence set (CS) for the true value θ_0 is then given by

$$CS_{n,T}(1-\alpha) = \{\theta \in \Theta : S_{n,T}(\theta) \leqslant c_{n,T}(\theta, 1-\alpha)\}, \tag{3.1.13}$$

To define the test statistic, we set

$$\begin{aligned} \widehat{m}_{n,t}(\theta) &= (\widehat{m}_{1,n,t}(\theta), \ldots, \widehat{m}_{k,n,t}(\theta))^\top, \quad \text{where} \\ \widehat{m}_{j,n,t}(\theta) &= m_j\left(X_t, \widehat{V}_{n,t}, Q_t, Z_t; \theta\right), \quad j = 1, \ldots, k. \end{aligned} \tag{3.1.14}$$

[9] See Theorem 9.3.2 in Jacod and Protter (2012) for details.

The sample moments are given by

$$\overline{m}_{n,T}(\theta) = \frac{1}{T}\sum_{t=1}^{T}\widehat{m}_{n,t}(\theta), \quad \overline{m}_{j,n,T}(\theta) = \frac{1}{T}\sum_{t=1}^{T}\widehat{m}_{j,n,t}(\theta), \quad j = 1, \ldots, k. \tag{3.1.15}$$

Let $\widehat{\Sigma}_{n,T}(\theta)$ be a heteroskedasticity and autocorrelation consistent (HAC) estimator associated with $\overline{m}_{n,T}(\theta)$. To be concrete, we follow Andrews (1991) by considering

$$\widehat{\Sigma}_{n,T}(\theta) = \sum_{l=-L_T}^{L_T}\omega(l/L_T)\,\widehat{\Gamma}_{l,n,T}(\theta), \quad \text{where for } |l| \leqslant T-1, \tag{3.1.16}$$

$$\widehat{\Gamma}_{l,n,T}(\theta) = \frac{1}{T}\sum_{1\leqslant t,t+l\leqslant T}\left(\widehat{m}_{n,t}(\theta) - \overline{m}_{n,T}(\theta)\right)\left(\widehat{m}_{n,t+l}(\theta) - \overline{m}_{n,T}(\theta)\right)^{\top}, \tag{3.1.17}$$

where $\omega(\cdot)$ and L_T are the kernel function and the bandwidth parameter, respectively.[10] The test statistic $S_{n,T}(\theta)$ is defined to be

$$S_{n,T}(\theta) = \sum_{j=1}^{2k_I}\left[T^{1/2}\overline{m}_{j,n,T}(\theta)/\widehat{D}_{j,n,T}^{1/2}(\theta)\right]_{-}^{2} + \sum_{j=2k_I+1}^{k}\left(T^{1/2}\overline{m}_{j,n,T}(\theta)/\widehat{D}_{j,n,T}^{1/2}(\theta)\right)^{2}, \tag{3.1.18}$$

where $[x]_{-} = \max\{-x, 0\}$ for $x \in \mathbb{R}$, $\widehat{D}_{n,T}(\theta) = Diag(\widehat{\Sigma}_{n,T}(\theta))$ the diagonal matrix collecting the diagonal elements of $\widehat{\Sigma}_{n,T}(\theta)$, and $\widehat{D}_{j,n,T}(\theta)$ is the jth diagonal element of $\widehat{D}_{n,T}(\theta)$. This test statistic has been studied, for example, in Andrews and Soares (2010) and quantifies the extent to which the moment inequalities/equalities in (3.1.10) are violated in the sample.[11] Each sample moment $\overline{m}_{j,n,T}$ is normalized

[10] For technical convenience, we only consider HAC estimators associated with compactly supported kernel functions, hence the estimator $\widehat{\Sigma}_{n,T}(\theta)$ only involves autocovariances up to order L_T. Examples of such kernel functions include Bartlett's kernel, Parzen's kernel and the Tukey-Hanning kernel. The quadratic spectral kernel considered by Andrews (1991) is excluded.

[11] In the terminology of Andrews and Soares (2010), the test statistic (3.1.18) is called the modified method of moments (MMM) statistic. Andrews and Soares (2010) also other possible test statistics

with respect to its standard error to ensure that the test statistic is scale invariant; this is clearly a desirable property in applications.

We choose the critical value based on the GMS procedure of Andrews and Soares (2010). Let $\widehat{\Omega}_{n,T}(\theta) = \widehat{D}_{n,T}^{-1/2}(\theta)\widehat{\Sigma}_{n,T}(\theta)\widehat{D}_{n,T}^{-1/2}(\theta)$ denote the correlation matrix associated with $\widehat{\Sigma}_{n,T}(\theta)$. We consider an $\mathbb{R}^k$-valued random variable $\widehat{Y}_{n,T}(\theta) = \widehat{\Omega}_{n,T}(\theta)^{1/2}Y^*$, where $Y^* \sim \mathcal{N}(0_k, I_k)$ is independent of the data; the jth element $\widehat{Y}_{n,T}(\theta)$ is denoted by $\widehat{Y}_{j,n,T}(\theta)$. Conditionally on the data, we set $c_{n,T}(\theta, 1-\alpha)$ as the $(1-\alpha)$-quantile of

$$\phi_{n,T}(\theta) = \sum_{j=1}^{2k_I} \left[\widehat{Y}_{j,n,T}\right]_-^2 1_{\left\{T^{1/2}\overline{m}_{j,n,T}(\theta) \leqslant \log(T)^{1/2}\widehat{D}_{j,n,T}^{1/2}(\theta)\right\}} + \sum_{j=2k_I+1}^{k} \widehat{Y}_{j,n,T}(\theta)^2 . \quad (3.1.19)$$

The indicator function in the above display selects moment inequalities that are “almost”binding at 0, where the factor $\log(T)^{1/2}$ corresponds to the BIC criterion.

3.1.3 Asymptotic results

We now turn to the asymptotic property of the confidence set $CS_{n,T}(1-\alpha)$. We first collect and discuss our assumptions, starting with those for the underlying processes.

Assumption A. For some constants $C > 0$ and $k \geqslant 2$, we have the following.

A1. The process X_t is an Itô semimartingale given by (3.1.11) with $\lambda(\mathbb{R}) < \infty$. The processes $\sigma_t > 0$ is also an Itô semimartingale with the form

$$\sigma_t = \sigma_0 + \int_0^t \tilde{b}_s\, ds + \int_0^t \tilde{\sigma}_s\, dW_s + \int_0^t \tilde{\sigma}'_s\, dW'_s + \int_0^t \int_{\mathbb{R}} \tilde{\delta}(s,z)(\underline{\mu} - \underline{\nu})(ds, dz) \quad (3.1.20)$$

including the Gaussian quasilikelihood ratio statistic and the generalized empirical likelihood ratio statistics. In the current paper, we only consider the MMM statistic for the concreteness and simplicity of exposition. We choose the MMM statistic because it is a direct generalization of the mean square error commonly used in empirical option pricing.

where $\tilde{b}_t$ is a locally bounded predictable process, W'_t is a standard Brownian motion orthogonal to W_t, $\tilde{\sigma}_t$ and $\tilde{\sigma}'_t$ are locally bounded càdlàg adapted processes and $\tilde{\delta}(\cdot)$ is a predictable function.

A2. For any $t \in \mathbb{R}_+$,

$$\mathbb{E}\left(|b_t|^k + |\tilde{b}_t|^k + |\sigma_t|^k + |\tilde{\sigma}_t|^k + |\tilde{\sigma}'_t|^k + \int_{\mathbb{R}} |\tilde{\delta}\,(t,z)\,|^k \lambda\,(dz)\right) \leqslant C. \tag{3.1.21}$$

A3. For any $t, s \in \mathbb{R}_+$,

$$\mathbb{E}\left(|b_t - b_s|^2 + |\tilde{b}_t - \tilde{b}_s|^2\right) \leqslant C\,|t-s|\,.$$

Assumption A1 provides a standard setup for analyzing high-frequency data. Jumps are allowed in both the price process and the spot volatility process. The leverage effect is also allowed here as dX_t and $d\sigma_t$ are both loaded on the Brownian shock dW_t. Moreover, by Itô's lemma, the spot variance $V_t = \sigma_t^2$ is also an Itô semimartingale, which can be expressed as

$$V_t = V_0 + \int_0^t b_{V,s} ds + \int_0^t \sigma_{V,s}\, dW_s + \int_0^t \sigma'_{V,s}\, dW'_s + \int_0^t \int_{\mathbb{R}} \delta_V(s,z)(\underline{\mu} - \underline{\nu})(ds, dz), \tag{3.1.22}$$

where $b_{V,t}$, $\sigma_{V,t}$, $\sigma'_{V,t}$ and $\delta_V(\cdot)$ are determined by the coefficients in (3.1.20), see (C.2.2) for explicit expressions. By assuming $\lambda(\mathbb{R}) < \infty$, we restrict our analysis to the case with finitely active jumps. While imposing finite activity appears strong for analyzing high-frequency financial data, it is not overly restrictive in the study of option pricing models. Indeed, many empirical works assume that the jumps are finitely active, typically compound Poisson, under the risk-neutral measure; by the equavalence of the risk-neutral and the physical measure, jumps are also finitely active under the physical measure in such settings. Assumption A2 is useful in analyzing the asymptotic behavior of various bias terms when $T \to \infty$.[12] Assumption A3 imposes some additional smoothness for the drift coefficient, but again is very mild.

We now state the assumptions that relate to option prices. Below, we denote by $f_V(\cdot)$ the partial derivative of $f(\cdot)$ with respect to spot variance.

Assumption B. We suppose for each fixed $\theta \in \Theta$, there exists some constant $C > 0$ such that the following hold for some $k' > 0$ and all t.

B1. The variables N_t and $Z_t = \{Z_{i,t} : 1 \leqslant i \leqslant N_t\}$ are $\mathcal{F}_{\underline{t}}$-measurable and $Z_s = Z_{\underline{t}}$ for all $s \in [\underline{t}, t]$.

B2. $\mathbb{E}\left(|f_V(X_t, V_t, Z_{i,t}; \theta)|^4\right) \leqslant C$.

B3. $\mathbb{E}\left(|f(X_t, V_t, Z_{i,t}; \theta)|^{2k'} + \|Q_{i,t}\|^{2k'}\right) \leqslant C$, where $Q_{i,t} = (A_{i,t}, B_{i,t}, p_{i,t})$ and $\|\cdot\|$ is the Euclidean norm.

B4. For some $\eta > 0$ and all $s \in [t-\eta, t]$, $\mathbb{E}\left(|f_V(X_t, V_t, Z_{i,\underline{t}}; \theta) - f_V(X_s, V_s, Z_{i,\underline{t}}; \theta)|^2\right) \leqslant C|t-s|$.

B5. We have $|f(X_t, \widehat{V}_{n,t}, Z_{i,t}; \theta) - f(X_t, V_t, Z_{i,t}; \theta)| \leqslant \chi_{n,t}|\widehat{V}_{n,t} - V_t|$ for some random variable $\chi_{n,t}$ such that $\mathbb{E}(|\chi_{n,t}|^{2k'}) \leqslant C$ for all n.

B6. We have $|f(X_t, \widehat{V}_{n,t}, Z_{i,t}; \theta) - f(X_t, V_t, Z_{i,t}; \theta) - f_V(X_t, V_t, Z_{i,t}; \theta)(\widehat{V}_{n,t} - V_t)| \leqslant$

[12] In a typical fill-in asymptotic setting, the processes b_t, $\tilde{b}_t$, σ_t, $\tilde{\sigma}_t$ and $\tilde{\delta}(\cdot)$ can be assumed bounded without loss of generality with the help of a standard localization argument; see Jacod and Protter (2012) for a comprehensive treatment. In such cases, (3.1.21) is satisfied trivially for the localized processes. In the current paper, we consider a setting in which the time span T also goes to infinity. This setting prevents the use of localization and requires explicit integrability conditions as imposed in (3.1.21).

$\xi_{n,t}|\widehat{V}_{n,t}-V_t|^2$ for some random variable $\xi_{n,t}$ such that for some $k>0$, $\mathbb{E}(|\xi_{n,t}|^{k'})\leqslant C$ for all n.

Assumption B1 formalizes the notion that N_t and Z_t are realized at the market opening and constant over the trading day. This assumption is realistic. It is also techinically convenient, because Assumption B1 allows us to ignore the theoretically possible intraday movement of the "menu" of quoted options. Assumptions B2 and B3 are primitive in nature and easy to interpret. Assumption B4 holds if for each day t, the process $s\mapsto f_V\left(X_s,V_s,Z_{i,\underline{t}};\theta\right)$ is Hölder continuous with exponent 1/2 in the L^2-space. This high-level assumption can be easily reduced into primitive conditions. Indeed, under the assumption that both X_t and σ_t are Itô semimartigales (Assumption A), the process $s\mapsto f_V\left(X_s,V_s,Z_{i,\underline{t}};\theta\right)$ is also an Itô semimartingale by Itô's formula, provided that $f_V\left(X_t,V_t,Z_{i,\underline{t}};\theta\right)$ is twice continuously differentiable in (X_t,V_t). The 1/2-Hölder continuity is a standard estimate for Itô semimartingales.

Assumptions B5 and B6 are high-level in nature and are used to impose, respectively, the first- and the second-order smoothness of the pricing function $f(\cdot)$ in the spot variance. We first discuss Assumption B5. By the mean value theorem, we have $|f(X_t,\widehat{V}_{n,t},Z_{i,t};\theta)-f(X_t,V_t,Z_{i,t};\theta)|=|f_V\left(X_t,\tilde{V}_{n,t},Z_{i,t};\theta\right)||\widehat{V}_{n,t}-V_t|$, where $\tilde{V}_{n,t}$ is the mean value between V_t and $\widehat{V}_{n,t}$. With $\chi_{n,t}=|f_V\left(X_t,\tilde{V}_{n,t},Z_{i,t};\theta\right)|$, Assumption B5 effectively imposes that the slope $\chi_{n,t}$ is bounded in a stochastic sense. The interpretation of Assumption B6 is similar; there, the variable $\xi_{n,t}$ plays the role of the second derivative of $f(\cdot)$ with respect to the spot variance evaluated at the mean value in a second order Taylor expansion. Although we deem Assumptions B5 and B6 to be intuitively appealing, we note that these assumptions may be difficult to verify under primitive conditions because option pricing functions typically do not have closed-form expressions. This being said, we are able to provide primitive conditions for B5 and B6 under the assumption that the derivatives of $f(\cdot)$ have polynomial growth in its arguments.

We assume either of the following assumption for the weighting functions and conditioning information set $\mathcal{I}_t$ in (3.1.3).

Assumption C. The condition (3.1.3) holds for $\mathcal{I}_t = \sigma\{Z_s : s \leqslant t\}$. For each $j \in \{1, \ldots, K\}$, the function $G_j(x, v, z, \tilde{z})$ does not depend on (x, v).

Assumption C'. The condition (3.1.3) holds for $\mathcal{I}_t = \sigma\{X_s, V_s, Z_s : s \leqslant t\}$. For each $j \in \{1, \ldots, K\}$, the function $G_j(x, v, z, \tilde{z})$ is continuously differentiable in $v \in \mathcal{V}$ with bounded partial derivative.

Assumption C is fairly mild. Assumption C' requires (3.1.3) to hold for a larger information set; this assumption is stronger but commonly adopted in empirical work with the benefit of improving the identification power of the model. In Assumption C', we require the weighting function to be smooth in the spot variance because we need to control the approximation error when using $\widehat{V}_{n,t}$ as a proxy of V_t.

A simple but important example of a weighting function $G(\cdot)$ satisfying Assumption C is $G(X_t, V_t, Z_{i,t}, Z_t) = 1\{\underline{K}_t \leqslant K_{i,t} \leqslant \overline{K}_t, \underline{\tau}_t \leqslant \tau \leqslant \overline{\tau}_t\}$, where the thresholds $\left(\underline{K}_t, \overline{K}_t\right)$ and $(\underline{\tau}_t, \overline{\tau}_t)$ are allowed to be dependent on Z_t.[13] By varying the thresholds, we effectively sort the options into various groups, or "boxes", indexed by

[13] For example, one may take $\underline{K}_t$ and $\overline{K}_t$ as the 10th-percentile and the 25th-percentile of the list of observed strikes $\{K_{i,t} : 1 \leqslant i \leqslant N_t\}$, and take $\underline{\tau}_t = 7$ days and $\overline{\tau}_t = 30$ days; in this case, the weighting function $G(\cdot)$ plays the role of selecting a group of in-the-money call and/or out-of-the-money put options with short maturity.

the strike price and the time-to-maturity. The inequality/equality system (3.1.10) simply consists of restrictions for group-wise moments. Clearly, the identification power of this system can be improved by using finer boxes. Indeed, Andrews and Shi (2014) show that when the number of boxes goes to infinity, trasforming conditional moment equalities and inequalities into unconditional ones incurs no loss of identification power.[14] Although this box construction is simple and intuitive, the smoothness requirement in Assumption C' prevents us from using the discontinuous indicator function on the dimension of the spot variance. In this case, a smooth approximation of the indicator function can be used for the spot variance.

We now state the key approximation result characterizing the approximation error for using $\widehat{V}_{n,t}$ as a proxy for V_t in the computation of sample moments and the HAC estimator. To state the result, we complement the notations in (3.1.14), (3.1.15) and (3.1.16) by defining $m_t^*(\theta)$, $\overline{m}_T^*(\theta)$, and $\widehat{\Sigma}_T^*(\theta)$ in the same way as $\widehat{m}_{n,t}(\theta)$, $\overline{m}_{n,T}(\theta)$, and $\widehat{\Sigma}_{n,T}(\theta)$ but with V_t in place of $\widehat{V}_{n,t}$.

Theorem 3.1.1. *Let $\varpi \in (3/8, 1/2)$. Suppose that Assumptions A and B hold for some $k \geqslant 2/(1-2\varpi)$ and $k' \geqslant 2/(8\varpi - 3)$.*

(a) Under Assumption C, we have $T^{1/2}\left(\overline{m}_{n,T}(\theta) - \overline{m}_T^*(\theta)\right) = O_p(T^{1/2}\Delta_n^{1/2}) + o_p(1)$.

(b) Under Assumption C', we have $T^{1/2}\left(\overline{m}_{n,T}(\theta) - \overline{m}_T^*(\theta)\right) = O_p(T^{1/2}\Delta_n^{1/4}) + o_p(1)$.

[14] Andrews and Shi's theory requires transforming the conditional moment (in)equalities into infinitely many unconditional moment equality/inequalities corresponding to infinitely many boxes in the space spanned by the conditioning variables. In finite samples, only a finite number of boxes can be used. Andrews and Shi show that the identification power of the conditional moment (in)equalities is still preserved if the number of boxes goes to infinity asymptotically. In the current paper, we maintain the simpler setting where the number of unconditional moment (in)equalities is fixed in the asymptotic theory and chosen to be relatively small in the numerical work. While we believe that the intuition underlying Andrews and Shi's theory is valid in our setting, we do not attempt to make a contribution in this direction in the current paper.

(c) Suppose either Assumption C or C'. If $L_T \Delta_n^{1/4} \to 0$, then $\widehat{\Sigma}_{n,T}(\theta) - \widehat{\Sigma}_T^*(\theta) = o_p(1)$.

Comments. (i) Part (a) and part (b) establishes the order of magnitude of the approximation error of the feasible sample moment function $\overline{m}_{n,T}(\theta)$ to its infeasible couterpart $\overline{m}_T^*(\theta)$. Qualitatively, both (a) and (b) ensures that the approximation error of the sample moments is negligible in the analysis of the asymptotic distributions of sample moments if $\Delta_n \to 0$ sufficiently fast relative to $T \to \infty$. Quantitatively, part (a) provides a stronger result than part (b); this improvement is achieved by extracting the martingale component from the approximation error and relies on the simple structure of the weighting function.

(ii) Part (c) shows that the feasible HAC estimator approximates its infeasible counterpart with a first-order negligible error, as soon as the bandwidth L_T in the HAC estimation grows to infinity sufficiently slow.

(iii) The assertions of the theorem holds for fixed truncation parameter ϖ and for sufficiently large k and k' which are needed for the elimination of jumps. These conditions can be equivalently formulated as

$$\varpi \in \left[\frac{3k'+2}{8k'}, \frac{k-2}{2k}\right], \quad k' > 2, \quad k > \frac{8k'}{k'-2} > 8.$$

For fixed k and k', the condition in Theorem 3.1.1 hence specifies a range of admissible ϖ for which jumps can be eliminated effectively enough as if they did not exist. This admissible range widens as k and k' increase; in particular, it converges to $(3/8, 1/2)$ as k and k' approaches infinity.

The significance of Theorem 3.1.1 is that it helps reduce the inference problem at hand, which involves the latent spot variance, to a standard moment inequality/equality problem involving only observables. More precisely, Theorem 3.1.1 implies that when $\Delta_n \to 0$ sufficiently fast, the feasible and the infeasible versions of the sample moments, as well as the HAC estimators, have the same asymptotic behavior. This result provides a theoretical justification for treating the latent spot variance V_t as if it is observed to be $\widehat{V}_{n,t}$. This formulation allows us to borrow the full strength of the existing results in the moment inequality/equality literature with no additional cost.

We introduce assumptions that are needed for the inference on the moment inequality/equality model in the second stage.

Assumption D. For each $\theta \in \Theta$, we have the following:

D1. The $\mathbb{R}^k$-valued process $m_t^*(\theta)$, $t = 1, \ldots, T$, is stationary.

D2. $T^{1/2}\left(\overline{m}_T^*(\theta) - \overline{m}(\theta)\right) \xrightarrow{d} \mathcal{N}\left(0_k, \Sigma(\theta)\right)$ for some k-dimensional positive definite matrix $\Sigma(\theta)$, where $\overline{m}(\theta) = \mathbb{E}\left[m_t^*(\theta)\right]$.

D3. $\widehat{\Sigma}_T^*(\theta) \xrightarrow{\mathbb{P}} \Sigma(\theta)$.

Assumption D1 imposes stationarity. We note that we only need $m_t^*(\theta)$ to be stationary, instead of the (unrealistically) strong assumption that X_t, V_t and other processes are jointly stationary. This assumption is reasonable when the option price is quoted in Black-Scholes implied volatility terms, rather than in currency terms. While it is possible to allow some mild heterogeneity in the time series by using the proper limiting theorem, such a generalization is beyond the interest of the current paper.[15] Assumption D2 describes the asymptotic distribution of the infeasible sample moment $\overline{m}_T^*(\theta_0)$. Observe that we assume the existence and the nonsingularity of the asymptotic variance-covariance matrix. This assumption is nevertheless quite mild. Assumption D3 imposes the consistency of the infeasible HAC estimator $\widehat{\Sigma}_T^*(\theta)$. Primitive conditions for D2 and D3 are well known in econometrics; see e.g. Davidson (1994).

We are now ready to state the asymptotic coverage property of $CS_{n,T}(1-\alpha)$. Below, we denote by $\Omega(\theta)$ the correlation matrix associated with $\Sigma(\theta)$.

Theorem 3.1.2. *Let $\alpha \in (0,1)$. Suppose either (i) the conditions of Theorem 3.1.1 with $T\Delta_n \to 0$ or (ii) the conditions of Theorem 3.1.1(b) with $T\Delta_n^{1/2} \to 0$. Also suppose (iii) Assumption D holds and $L_T\Delta_n^{1/4} \to 0$. Then we have*

$$\liminf_{\Delta_n \to 0, T \to \infty} \mathbb{P}\left(\theta_0 \in CS_{n,T}(1-\alpha)\right) \geqslant 1-\alpha. \tag{3.1.23}$$

If additionally, we have (iv) the distribution function of the sum $\sum_{j=1}^{2k_I}[U_j]_-^2 1_{\{\bar{m}_j(\theta_0)=0\}} + \sum_{j=2k_I+1}^{k}(U_j)^2$ is continuous at its $(1-\alpha)$-quantile, where $U \sim \mathcal{N}(0,\Omega(\theta_0))$, then

$$\lim_{\Delta_n \to 0, T \to \infty} \mathbb{P}\left(\theta_0 \in CS_{n,T}(1-\alpha)\right) = 1-\alpha. \tag{3.1.24}$$

Theorem 3.1.2 shows that $CS_{n,T}(1-\alpha)$ has asymptotically correct coverage for the true parameter θ_0. The proof is a straightforward consequence of Theorem 3.1.1. This result only asserts the pointwise validity of the CS. A stronger, uniform coverage result is beyond the scope of the paper.

[15] We note that Andrews and Soares (2010) also impose stationarity in their time-series setting. As some of our proofs rely on their results, we maintain this assumption for the convenience of reference.

3.1.4 No-arbitrage restrictions for affine jump-diffusion models

We now specialize the pricing function $f(\cdot)$ in a risk-neutral pricing setting. In this section, we propose an additional testable restriction based on a no-arbitrage argument. Here we extend the above inference procedure to accommodate this new restriction.

In the absence of arbitrage opportunities, there exists a risk-neutral measure $\mathbb{Q}$, which is locally equivalent to the physical measure $\mathbb{P}$ and transforms the discounted asset prices into local martingales. We suppose that under the $\mathbb{Q}$-measure, the log-price and spot variance processes have the following dynamics:

$$\begin{aligned} dX_t &= b_t^{\mathbb{Q}} dt + \sqrt{V_t} dW_t^{\mathbb{Q}} + dJ_t^{\mathbb{Q}}, \\ dV_t &= b_{V,t}^{\mathbb{Q}} dt + \rho \upsilon \sqrt{V_t} dW_t^{\mathbb{Q}} + \left(1-\rho^2\right)^{1/2} \upsilon \sqrt{V_t} dW_t'^{\mathbb{Q}}, \end{aligned} \tag{3.1.25}$$

where $J_t^{\mathbb{Q}}$ is a pure jump process with finite activity, $W_t^{\mathbb{Q}}$ and $W_t'^{\mathbb{Q}}$ are mutually orthogonal standard $\mathbb{Q}$-Brownian motions, and the constant $\rho \in [-1, 1]$ captures the "leverage effect" (Black (1976)), and the constant $\upsilon > 0$ is often referred to as the volatility of volatility. The specific forms of the drift processes $b_t^{\mathbb{Q}}$ and $b_{V,t}^{\mathbb{Q}}$ are not relevant for the discussion in this section.

We observe an important parametric restriction imposed by (3.1.25), namely that the spot variance process follows a square-root diffusion under $\mathbb{Q}$. This parametric form is commonly imposed in the empirical option pricing literature as it, when combined with other simplications, admits a semi closed-form solution for option prices. The resulting computational efficiency gain is desirable and often necessary for analyzing large data sets of options.

Next, note that since the covariation process between X_t and V_t is invariant with respect to equivalent change of measure, (3.1.25) imposes the following restriction:

$$CV_t = \rho_0 v_0 IV_t, \quad \text{where} \quad CV_t = \int_{\underline{t}}^{t} \sigma_{V,s}\sigma_s ds, \quad IV_t = \int_{\underline{t}}^{t} V_s ds, \tag{3.1.26}$$

where ρ_0 and v_0 denote the true value of ρ and v; "CV" and "IV" stand, respectively, for "covariation" and "integrated volatility".

To have a feasible procedure, we approximate the latent variables CV_t and IV_t with estimators based on high frequency data. For IV_t, we consider the well-known jump-robust estimator proposed by Mancini (2001):

$$\widehat{IV}_{n,t} = \sum_{i=1}^{n} \left|\Delta_i^{t,n} X\right|^2 1_{\left\{\left|\Delta_i^{t,n} X\right| \leqslant \alpha \Delta_n^{\varpi}\right\}}. \tag{3.1.27}$$

For CV_t, Wang and Mykland (2011) propose an estimator in the absence of price and volatility jumps. To accomodate the price jumps, we consider a truncated version of Wang and Mykland's estimator defined as follows. For each day t, we group the n intraday returns into B_n blocks, where block $i \in \{1, \ldots, B_n\}$ collects the $k_n = n/B_n$ returns within the interval $(\tau(t,n,i-1), \tau(t,n,i)]$, $\tau(t,n,i) = \underline{t} + ik_n\Delta_n$. For simplicity, we have implicitly assumed that k_n is an integer, but this assumption has little effect on the asymptotic theory. For day t, we denote the jth return in the ith block to be $\Delta_{i,j}^{t,n} X = X_{\tau(t,n,i-1)+j\Delta_n} - X_{\tau(t,n,i-1)+(j-1)\Delta_n}$, and define

$$
\begin{aligned}
\widehat{CV}_{n,t} &= 2\sum_{i=2}^{B_n}\left(\sum_{j=1}^{k_n}\Delta_{i,j}^{t,n}X\,1_{\left\{\left|\Delta_{i,j}^{t,n}X\right|\leqslant\alpha\Delta_n^{\varpi}\right\}}\right)\left(\widehat{V}_{n,\tau(t,n,i)} - \widehat{V}_{n,\tau(t,n,i-1)}\right), \quad \text{where} \\
\widehat{V}_{n,\tau(t,n,i)} &= \frac{1}{k_n\Delta_n}\sum_{j=1}^{k_n}\left(\Delta_{i,j}^{t,n}X\right)^2 1_{\left\{\left|\Delta_{i,j}^{t,n}X\right|\leqslant\alpha\Delta_n^{\varpi}\right\}}, \quad i \in \{1,\ldots,B_n\}.
\end{aligned}
\tag{3.1.28}
$$

We observe that Wang and Mykland's orginal estimator corresponds the case with $\alpha = \infty$ (i.e. no truncation), as price jumps are not considered in their paper.

To motivate the use of $\widehat{CV}_{n,t}$, we provide an auxiliary result describing its asymptotic behavior.

Lemma 3.1.3. *Suppose that Assumption A1 holds and σ_t is continuous. Then for each fixed t, as $n \to \infty$, $\Delta_n^{-1/4}\left(\widehat{CV}_{n,t} - CV_t\right)$ converges stably in law to a random variable defined on an extension of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which conditionally on $\mathcal{F}$, is centered Gaussian with strictly positive variance.*

Comments. (i) Lemma 3.1.3 establishes the $n^{1/4}$-rate of convergence of $\widehat{CV}_{n,t}$ towards CV_t and shows that the rate is sharp. This result is related to Wang and Mykland's Theorem 1, which derives the stable convergence in law of $\widehat{CV}_{n,t}$ for continuous processes (hence without truncation). In our proof, we show that for fixed $\alpha \in (0, \infty)$ and $\varpi \in (5/12, 1/2)$, the truncation in $\widehat{CV}_{n,t}$ eleminates price jumps "effectively enough", so that the asymptotic theory in the presence of price jumps can be reduced to the continuous setting as considered by Wang and Mykland.[16]

[16] In particular, the asympototic variance of $\widehat{CV}_{n,t}$ is identical to that given in Theorem 1 of Wang and Mykland (2013). Its (rather complicated) expression is suppressed here as we do not make use of its exact form in the current paper.

(ii) An important consequence is that $\Delta_n^{-1/4}(\widehat{CV}_{n,t} - \rho v \widehat{IV}_{n,t})$ is nondegenerate asympotically. To see this, first recall the well-known result that $\widehat{IV}_{n,t} - IV_t = O_p(\Delta_n^{1/2})$. Under (3.1.26), $\Delta_n^{-1/4}(\widehat{CV}_{n,t} - \rho_0 v_0 \widehat{IV}_{n,t}) = \Delta_n^{-1/4}(\widehat{CV}_{n,t} - CV_t) + o_p(1)$. Lemma 3.1.3 hence implies that $\Delta_n^{-1/4}(\widehat{CV}_{n,t} - \rho_0 v_0 \widehat{IV}_{n,t})$ is asymptotically nondegenerate, i.e., its asymptotic variance is strictly positive. For $\rho v \neq \rho_0 v_0$, it is easily to see that $\Delta_n^{-1/4}(\widehat{CV}_{n,t} - \rho v \widehat{IV}_{n,t}) = (\rho_0 v_0 - \rho v)\Delta_n^{-1/4} IV_t + O_p(1)$, which diverges to infinity and is clearly nondegenerate.

In practice, it turns out that it is helpful to consider an equivalent formulation of (3.1.26) as follows. Let $a_{IV} \geqslant 0$ and $b_{IV} > 0$ be constants fixed *a priori*. We can express (3.1.26) equivalently as

$$\frac{CV_t}{a_{IV} IV_{t-1} + b_{IV}} - \rho_0 v_0 \frac{IV_t}{a_{IV} IV_{t-1} + b_{IV}} = 0. \tag{3.1.29}$$

The idea is that by normalizing CV_t and IV_t by $1/(a_{IV} IV_{t-1} + b_{IV})$, both series tend to become less volatile, leading to better numerical performance in finite-samples. In our simulation and empirical work, we take $a_{IV} = 1$ and $b_{IV} = 0.0001$.[17] We associate (3.1.29) with its sample moment,

$$\begin{aligned} \bar{m}_{n,T}^{+}(\theta) &= \frac{1}{T}\sum_{t=1}^{T} \hat{m}_{n,t}^{+}(\theta), \quad \text{where} \\ \hat{m}_{n,t}^{+}(\theta) &= \Delta_n^{-1/4}\left(\frac{\widehat{CV}_{n,t}}{a_{IV}\widehat{IV}_{n,t-1} + b_{IV}} - \rho v \frac{\widehat{IV}_{n,t}}{a_{IV}\widehat{IV}_{n,t-1} + b_{IV}}\right), \end{aligned} \tag{3.1.30}$$

[17] Setting b_{IV} to be strictly positive, instead of being zero, ensures that the normalizing factor $1/(a_{IV} IV_t + b_{IV})$ is bounded above. The boundedness is clearly convenient for asymptotic arguments; it also enhances the stability in numerical works (CV_t/IV_t may be very large when IV_t is small). We note that b_{IV} is a fixed constant and plays no role in the asymptotics.

(note that ρ and v are elements of θ) and let $\widehat{\Sigma}_{n,T}^{+}(\theta)$ be a HAC estimator of long-run variance of the sequence $\hat{m}_{n,t}^{+}(\theta)$, $1 \leqslant t \leqslant T$. We include the factor $\Delta_n^{-1/4}$ in the definition of $\widehat{m}_{n,t}^{+}(\theta)$ to prevent degeneracy; see comment (ii) of Lemma 3.1.3.

The new test statistic is given by

$$S'_{n,T}(\theta) = S_{n,T}(\theta) + \left(T^{1/2}\bar{m}_{n,T}^{+}(\theta)\right)^2 / \widehat{\Sigma}_{n,T}^{+}(\theta),$$

Here, we incorporate the feasible version of the equality restriction (3.1.29) as an extra moment equality. Although $\mathbb{E}\left[\hat{m}_{n,t}^{+}(\theta)\right] = 0$ only holds "approximately" (as $\Delta_n \to 0$), this complication can be readily accomodated in the asymptotic theory as shown below. For $\alpha \in (0,1)$, a $1-\alpha$ level confidence set based on $S'_{n,T}(\theta)$ is given by

$$CS'_{n,T}(1-\alpha) = \left\{\theta \in \Theta : S'_{n,T} \leqslant c'_{n,T}(\theta, 1-\alpha)\right\}, \tag{3.1.31}$$

where the critical value $c'_{n,T}(\theta, 1-\alpha)$ is determined as follows. We set

$$\phi'_{n,T}(\theta) = \phi_{n,T}(\theta) + (Y^{**})^2,$$

where $\phi_{n,T}(\theta)$ is given by (3.1.19) and Y^{**} is a generic standard Gaussian random variable independent of $\phi_{n,T}(\theta)$. We then set $c'_{n,T}(\theta, 1-\alpha)$ as the $(1-\alpha)$-quantile of $\phi'_{n,T}(\theta)$ conditionally on the data. Under regularity conditions, we can show that $S'_{n,T}(\theta_0) - S_{n,T}(\theta_0)$ is asymptotically chi-square distributed with degree of freedom one and asymptotically independent of $S_{n,T}(\theta_0)$; this motivates using the distribution of $\phi'_{n,T}(\theta_0)$ to approximate the distribution of $S'_{n,T}(\theta_0)$. Our formal results demand the following assumption.

Assumption E. Denote $\psi_{n,t}(\theta) = \left(T^{-1/2}m_t^*(\theta)^\top, T^{-1/2}\hat{m}_{n,t}^{+}(\theta)\right)^\top$. Let $\Sigma_{\psi,n,T}(\theta)$

and $\Sigma^{+}_{n,T}(\theta)$ be the variance-covariance matrices of $\sum_{t=1}^{T}\psi_{n,t}(\theta)$ and $\sum_{t=1}^{T}T^{-1/2}\hat{m}^{+}_{n,t}(\theta)$, respectively. Let c_1 and $c_2 \in (0,1)$ be strictly positive constants.

E1. $\Delta_n \asymp T^{-c_1}$.

E2. $\Sigma_{\psi,n,T}(\theta)^{-1/2}\sum_{t=1}^{T}(\psi_{n,t}(\theta) - \mathbb{E}[\psi_{n,t}(\theta)]) \xrightarrow{d} \mathcal{N}(0_{k+1}, I_{k+1})$.

E3. $\widehat{\Sigma}^{+}_{n,T}(\theta) - \Sigma^{+}_{n,T}(\theta) \xrightarrow{\mathbb{P}} 0$ as $T \to \infty$.

E4. $\Sigma^{+}_{n,T}(\theta)^{-1} = O_p(1)$ as $T \to \infty$.

E5. The process $\psi_{n,t}(\theta)$, $1 \leqslant t \leqslant T$, is stationary and α-mixing with mixing coefficient $\alpha_{mix,l}$ satisfying $\sum_{l=1}^{\infty}\alpha_{mix,l}^{1-c_2} < \infty$.

Assumption E1 links the asymptotic behavior of Δ_n with that of T so that the double asymptotic nesting is reduced to a simpler one indexed only by T. This simplification has little effect in practice, but allows us to use well-known limit theorems for dependent triangular arrays, and hence greatly simplifies the exposition of results. Under Assumption E1, $\psi_{n,t}(\theta_0) - \mathbb{E}[\psi_{n,t}(\theta_0)]$ forms a zero-mean triangular array by construction. Assumptions E2 can then be verified under primitive conditions via central limit theorems for dependent triangular arrays; see e.g. de Jong (1997). Assumption E2 partially overlaps with Assumption D2 in that they both imposes the weak convergence of the sample moment $\overline{m}^{*}_{n,T}(\theta)$. Assumption D2 is actually stronger on this regard, because it imposes the existence of the asymptotic variance. Hence, the additional regularity from Assumption E2 is for the weak convergence of the additional moment $\overline{m}^{+}_{n,T}(\theta)$, joint with $\bar{m}^{*}_{n,T}(\theta)$, without imposing the existence of their asymptotic variance-covariance. Nevertheless, we show that the asymptotic covariance between $\overline{m}^{+}_{n,T}(\theta)$ and $\bar{m}^{*}_{n,T}(\theta)$ exists and is zero. This result is a key step for the proof of Theorem 3.1.4 below. Assumptions E3 imposes the consistency of the HAC estimator for $\Sigma^{+}_{n,T}(\theta_0)$. This assumption can be verified under primitive conditions via results on the HAC estimation for dependent triangular arrays; see e.g. Davidson and de Jong (2002). Assumption E4 states that the variance of the sequence $\hat{m}^{+}_{n,t}(\theta)$ is nondegenerate asymptotically. This assumption is not very restrictive. As discussed in comment (ii) of Lemma 3.1.3, $\Delta_n^{-1/4}$ is the proper scaling factor for $\widehat{CV}_{n,t} - \rho v \widehat{IV}_{n,t}$ that prevents its variance from degenerating asympototically.

Theorem 3.1.4. *Let $\varpi \in (5/12, 1/2)$. Suppose (i) Assumption A holds for some $k \geqslant \max\{8, 2/(1-2\varpi)\}$ and V_t satisfies (3.1.25); (ii) Assumption B holds for some $k' > 2/(12\varpi - 5)$; (iii) either Assumption C or Assumption C'; (iv) Assumption D holds with $L_T \Delta_n^{1/4} = o(1)$; (v) Assumption E holds for some $c_1 > 2/(12\varpi - 5)$ and $c_2 = 1/(2k') + (9 - 12\varpi)/4$. Then*

$$\lim_{T \to \infty} \mathbb{P}\left(\theta_0 \in CS'_{n,T}(1-\alpha)\right) = 1 - \alpha.$$

We finish this section by pointing out several limitations of the approach above, as well as possible extensions on these directions. Firstly, (3.1.25) excludes the presence of volatility jumps. Although the truncated estimator $\widehat{CV}_{n,t}$ is likely to remain $n^{1/4}$-consistent for CV_t with volatility jumps, the characterizations of its limiting distribution under the fill-in asymptotics (cf. Lemma 3.1.3) and its bias under the long-span asymptotics are still open questions. We exclude volatility jumps here mainly for this reason. Secondly, the approach above only incorporates the restriction on the covariation between the log-price and the spot variance. Clearly, a similar restriction can be imposed on the quadratic variation process of V_t. Indeed, (3.1.25) implies that the quadratic variation of V_t (or more generally the continuous part of V_t) is $v_0 \int_{\underline{t}}^{t} V_s ds$, yielding an extra restriction on the volatility-of-volatility parameter. If a nonparametric estimator for the quadratic variation of V_t is available, we may incorporate this additional restriction in a similar fashion as we have done for (3.1.26). This being said, nonparameteric estimation for the quadratic variation of V_t is, to the best of our knowledge, still an open question, especially when price jumps and/or volatility jumps are present and treated nonparametrically. Hence, extensions on these two directions demand further high-frequency estimation results for the covariation CV_t and/or the diffusive quadratic variation of V_t in a general setting with price and volatilty jumps under both fill-in and long-span asymptotics. These results are interesting on their own and technically challenging, and beyond the scope of the current paper. We hope our discussion here may motivate future research on this direction, which in turn may be incorporated to generalize the approach consider here.

3.2 Simulation Study

We now examine the finite-sample performance of the above asymptotic theory.

3.2.1 Simulation Design

We consider the stochastic volatility model with jumps under the risk-neutral measure:

$$\begin{aligned} dX_t &= (r_t - \delta_t)\, dt + \sqrt{V_t} dW_t^{\mathbb{Q}} + dJ_t, \\ dV_t &= \kappa^{\mathbb{Q}} \left(V_t - \bar{V}^{\mathbb{Q}}\right) dt + \rho^{\mathbb{Q}} v^{\mathbb{Q}} \sqrt{V_t} dW_t^{\mathbb{Q}} + \left(1 - \left(\rho^{\mathbb{Q}}\right)^2\right)^{1/2} v^{\mathbb{Q}} \sqrt{V_t} dW_t'^{\mathbb{Q}} \end{aligned} \tag{3.2.1}$$

where J_t is a compound Poisson process with intensity $\lambda^{\mathbb{Q}}$ and jump size distributed as $\mathcal{N}(\mu_J^{\mathbb{Q}}, (\sigma_J^{\mathbb{Q}})^2)$. For simplicity, we suppose the interest rate and the dividend yield are zero in the simulation. The risk neutral parameter θ is given by the vector $(\kappa^{\mathbb{Q}}, \bar{V}^{\mathbb{Q}}, \rho^{\mathbb{Q}}, v^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \mu_J^{\mathbb{Q}}, \sigma_J^{\mathbb{Q}})$. We set

$$\left(\kappa^{\mathbb{Q}}, \bar{V}^{\mathbb{Q}}, \rho^{\mathbb{Q}}, v^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \mu_J^{\mathbb{Q}}, \sigma_J^{\mathbb{Q}}\right) = (5.00, 0.05, -0.50, 0.60, 0.50, -0.05, 0.20).$$

Under the physical measure, the data is generated using the same model but with parameters

$$\left(\kappa^{\mathbb{P}}, \bar{V}^{\mathbb{P}}, \rho^{\mathbb{P}}, v^{\mathbb{P}}, \lambda^{\mathbb{P}}, \mu_J^{\mathbb{P}}, \sigma_J^{\mathbb{P}}\right) = (6.33, 0.04, -0.50, 0.60, 0.50, -0.05, 0.20).$$

We have imposed $\rho^{\mathbb{P}} = \rho^{\mathbb{Q}}$ and $v^{\mathbb{P}} = v^{\mathbb{Q}}$ to ensures the equivalence between the risk-neutral measure and the physical measure. The continuous-time process is simulated via the Euler scheme discretized at 5 seconds. The high-frequency data used in our estimation is then resampled at 1-minute interval.

We consider European call options. Quoted in terms of the Black-Scholes implied volatility, the true price of an option with strike price κ and time-to-maturity τ is given by

$$f(x, v, \kappa, \tau; \theta) = BS^{-1}_{x,\kappa,\tau}(\mathbb{E}^{\mathbb{Q}}\left[\left(e^{X_\tau} - \kappa\right)^+ | X_0 = x, V_0 = v\right]),$$

where $BS^{-1}_{x,\kappa,\tau}(\cdot)$ is the functional inversion of the Black-Scholes formula. We adopt the method of Duffie et al. (2000) and Fang and Oosterlee (2008) for computing the true option price. For each day, we generate 7 options with strike prices $\{\kappa_{i,t} : 1 \leqslant i \leqslant 7\}$ and the same times-to-maturity $\tau_{i,t}$ = 30, 60, and 90 days. We set $\kappa_{i,t} = \kappa_i \exp(X_t)$, where $(\kappa_1, \ldots, \kappa_7) = (0.8, 0.85, 0.9, 0.95, 1, 1.05, 1.1)$.

We consider three settings for generating the efficient quotes $\left(A^*_{i,t}, B^*_{i,t}\right)$:

$$\text{Case 1} \quad : \quad \begin{cases} A^*_{i,t} &= f(X_t, V_t, \kappa_{i,t}, \tau_{i,t}; \theta_0) + SPRD/2 \\ B^*_{i,t} &= f(X_t, V_t, \kappa_{i,t}, \tau_{i,t}; \theta_0) - SPRD/2 \end{cases}$$

$$\text{Case 2} \quad : \quad \begin{cases} A^*_{i,t} &= f(X_t, V_t, \kappa_{i,t}, \tau_{i,t}; \theta_0) + (1 - \kappa_i/2)\, SPRD \\ B^*_{i,t} &= f(X_t, V_t, \kappa_{i,t}, \tau_{i,t}; \theta_0) - (\kappa_i/2)\, SPRD, \end{cases}$$

$$\text{Case 3} \quad : \quad \begin{cases} A^*_{i,t} &= f(X_t, V_t, \kappa_{i,t}, \tau_{i,t}; \theta_0) + (1 - (i-1)/3)\, SPRD \\ B^*_{i,t} &= f(X_t, V_t, \kappa_{i,t}, \tau_{i,t}; \theta_0) - ((i-1)/3)\, SPRD, \end{cases}$$

where the bid-ask spread $SPRD$ is calibrated from actual data and corresponds to about 1-2% implied volatility units, depending on the strike-maturity combination. In case 1, the true option price coincides with the efficient mid-quote. In case 2, the true option price is lower (resp. higher) than the mid-quote for in-the-money (resp. out-of-the-money) options. The observed quotes are then generated according to (3.1.2) where ε^A_{it} and ε^B_{it} are drawn independently from a uniform distribution supported on $[-\sigma_\varepsilon, \sigma_\varepsilon]$. With $\sigma_\varepsilon < SPRD/2$, this design is a simple way of ensuring the natural ordering $A_{i,t} > B_{i,t}$. We calibrate σ_ε according to $Var(A_{i,t} - B_{i,t}) =$

$Var\left(\varepsilon_{i,t}^{A}-\varepsilon_{i,t}^{B}\right)=2Var\left(\varepsilon_{i,t}^{A}\right)=\frac{2}{3}\sigma_{\varepsilon}^{2}$, so $\sigma_{\varepsilon}=\sqrt{(3/2)}Se\left(A_{i,t}-B_{i,t}\right)$ and then further refine the specification to $\sigma_{\varepsilon}=\min\left\{\sqrt{(3/2)}Se\left(A_{i,t}-B_{i,t}\right),0.99*SPRD/2\right\}$. Case 3 considers the case of boundary misspecification, i.e. when the true option price lies on either the bid or ask as opposed to uniformly in the interior of the spread.

We consider Monte Carlo samples of length $T=2$ years and 15 years, consisting of daily end-of-day simulated option prices. The purpose of looking at different length samples is to examine the effect of T, given the convergence rate interaction with the sampling interval Δ_n derived previously. To obtain end-of-day prices, we simulate the stochastic process in (3.2.1) and record the values of the state variables at the close of the simulated trading day and use those to compute closing "true" option prices. We then generate noisy bid and ask quotes following the framework outlined above. For each of the resulting daily option samples, we compute the finite-sample coverage probabilities given in the $\mathbb{Q}$-measure setup from (3.1.13), as well as the joint $\mathbb{P}$- and $\mathbb{Q}$-measure setup in (3.1.31), under the three DGP cases described above. Specifically, we examine the coverage probabilities under correct mid-quote specification, interior misspecification, and boundary misspecification.

3.2.2 Simulation Results

The results of the simulation exercises are given in Tables 3.1 through 3.6. Tables 3.1 through 3.3 show coverage probabilities for 15 year samples, whereas Tables 3.4 through 3.6 show the corresponding results for the 2 year samples.

Table 3.1 represents the DGP under which the mid-quote and the latent efficient option price coincide exactly. The left panel shows the coverage probabilities using the joint $\mathbb{P}$- and $\mathbb{Q}$-measure conditions (which includes an equality restriction on the

Table 3.1: Monte Carlo Simulation: Size Control, 15 Year Sample, Case 1. This table shows the results of a Monte Carlo experiment, in which 15 years of options bid and ask quotes were simulated 2,000 times by pricing options on an underlying SVJ model generated via an Euler scheme. The Andrews-Soares (2010) confidence sets were computed each time, and the rejection frequencies were recorded for the nominal sizes given in column 1. The left panel shows simulation results using both $\mathbb{P}$ and $\mathbb{Q}$ measure restrictions; the right panel shows $\mathbb{Q}$ measure restrictions only. Within each panel, the table shows the effect of using estimated spot volatility $\widehat{V}_{n,t}$ under varying widths of the bid-ask spread.

	$\mathbb{P}$ and $\mathbb{Q}$ measure restrictions				$\mathbb{Q}$ measure restrictions only			
	Mid-quote		Full Spread		Mid-quote		Full Spread	
α	V_t	$\widehat{V}_{n,t}$	V_t	$\widehat{V}_{n,t}$	V_t	$\widehat{V}_{n,t}$	V_t	$\widehat{V}_{n,t}$
0.01	0.010	0.017	0.029	0.028	0.011	0.015	0.000	0.000
0.05	0.062	0.069	0.109	0.108	0.059	0.062	0.000	0.000
0.10	0.124	0.122	0.171	0.172	0.107	0.109	0.000	0.000
0.15	0.171	0.172	0.228	0.228	0.162	0.167	0.000	0.000
0.20	0.226	0.220	0.290	0.289	0.212	0.218	0.000	0.000
0.25	0.278	0.270	0.338	0.338	0.260	0.269	0.000	0.000
0.30	0.324	0.321	0.390	0.391	0.306	0.319	0.000	0.000
0.35	0.373	0.382	0.432	0.434	0.358	0.371	0.000	0.000
0.40	0.429	0.438	0.483	0.483	0.410	0.417	0.000	0.000
0.45	0.482	0.499	0.538	0.538	0.465	0.459	0.000	0.000
0.50	0.533	0.544	0.587	0.588	0.513	0.512	0.000	0.000

product of the leverage effect and vol-of-vol), whereas the right panel shows coverage probabilities under only $\mathbb{Q}$-measure restrictions (using only moment inequalities). Not surprisingly, the best coverage is achieved by fitting the model to mid-quotes when the DGP is in fact the mid-quote. This holds whether one uses the infeasible true spot volatility V_t to price options or uses its high-frequence estimate $\widehat{V}_{n,t}$ (columns 2 and 3). In contrast, using the more conservative approach of bounding the efficient price by the bid-ask spread (columns titled "Full Spread") results in

Table 3.2: Monte Carlo Simulation: Size Control, 15 Year Sample, Case 2. This table shows the results of a Monte Carlo experiment, in which 15 years of options bid and ask quotes were simulated 2,000 times by pricing options on an underlying SVJ model generated via an Euler scheme. The Andrews-Soares (2010) confidence sets were computed each time, and the rejection frequencies were recorded for the nominal sizes given in column 1. The left panel shows simulation results using both $\mathbb{P}$ and $\mathbb{Q}$ measure restrictions; the right panel shows $\mathbb{Q}$ measure restrictions only. Within each panel, the table shows the effect of using estimated spot volatility $\widehat{V}_{n,t}$ under varying widths of the bid-ask spread.

	$\mathbb{P}$ and $\mathbb{Q}$ measure restrictions				$\mathbb{Q}$ measure restrictions only			
	Mid-quote		Full Spread		Mid-quote		Full Spread	
α	V_t	$\widehat{V}_{n,t}$	V_t	$\widehat{V}_{n,t}$	V_t	$\widehat{V}_{n,t}$	V_t	$\widehat{V}_{n,t}$
0.01	0.045	0.034	0.027	0.026	0.040	0.029	0.000	0.000
0.05	0.229	0.164	0.109	0.107	0.232	0.154	0.000	0.000
0.10	0.427	0.336	0.165	0.165	0.423	0.323	0.000	0.000
0.15	0.592	0.487	0.221	0.222	0.586	0.481	0.000	0.000
0.20	0.709	0.614	0.283	0.282	0.709	0.605	0.000	0.000
0.25	0.796	0.722	0.333	0.332	0.788	0.713	0.000	0.000
0.30	0.861	0.795	0.384	0.385	0.851	0.788	0.000	0.000
0.35	0.901	0.853	0.429	0.429	0.892	0.844	0.000	0.000
0.40	0.931	0.886	0.478	0.478	0.926	0.883	0.000	0.000
0.45	0.953	0.917	0.531	0.531	0.945	0.918	0.000	0.000
0.50	0.972	0.941	0.582	0.583	0.960	0.946	0.000	0.000

slight overrejection. Under $\mathbb{Q}$-measure restrictions only, using the full spread given by bid ask quotes now results in conservative inference, whereas using the mid-quote directly results in correct finite-sample coverage.

The most striking results are given in Tables 3.2 and 3.3. Table 3.2 shows coverage probabilities under the first type of mid-quote misspecification (Case 2), i.e. the efficient price is not the mid-quote, but it is also not on the boundary given by the bid and ask quotes. In the left panel, the columns labeled "Mid-quote" show that fitting an option pricing model to the mid-quote when the DGP in fact deviates from the mid-quote can result in severe overrejection of the option pricing model. We interpret this result as evidence that incorrect assumptions on the option market's microstructure can have profound effects on inference on frictionless option pricing models. On the other hand, taking the more conservative approach of merely bounding the efficient price by the bid and ask quotes results in comparatively mild overrejection. For example, for nominal level 50% confidence sets, the bid-ask bound approach results in 58.3% rejections compared to 94.1% rejections from the erroneous mid-quote point-identifying assumption. For $\mathbb{Q}$-measure-only confidence sets, the bid-ask bound approach results in conservative inference, whereas the mid-quote assumption again results in severe overrejection.

Table 3.3 shows results for the Case 3 DGPs, i.e. when the efficient price sometimes coincides with the bid and ask quotes and generally deviates from the mid-quote. In this scenario, fitting option prices to the mid-quote results in automatic 100% rejections. In contrast, bounding the efficient prices by the bid and ask quotes yields slightly conservative and sometimes correct inference. This observation holds whether one is using the $\mathbb{P}$-measure equality restriction (confirming Theorem 3.1.4) or not (Theorem 3.1.2, Equation (3.1.24)), since Case 3 corresponds to the boundary parameter case.

Comparing Table 3.4 with Table 3.1, we observe that the effect of a smaller sample size T mitigates the mild overrejections found on the 15-year sample. This agrees well with the result in Theorem 3.1.1, which required the sampling interval $\Delta_n \to 0$ sufficiently fast relative to the growth in T. This improvement in inference is carried

Table 3.3: Monte Carlo Simulation: Size Control, 15 Year Sample, Case 3. This table shows the results of a Monte Carlo experiment, in which 15 years of options bid and ask quotes were simulated 2,000 times by pricing options on an underlying SVJ model generated via an Euler scheme. The Andrews-Soares (2010) confidence sets were computed each time, and the rejection frequencies were recorded for the nominal sizes given in column 1. The left panel shows simulation results using both $\mathbb{P}$ and $\mathbb{Q}$ measure restrictions; the right panel shows $\mathbb{Q}$ measure restrictions only. Within each panel, the table shows the effect of using estimated spot volatility $\widehat{V}_{n,t}$ under varying widths of the bid-ask spread.

	$\mathbb{P}$ and $\mathbb{Q}$ measure restrictions				$\mathbb{Q}$ measure restrictions only			
	Mid-quote		Full Spread		Mid-quote		Full Spread	
α	V_t	$\widehat{V}_{n,t}$	V_t	$\widehat{V}_{n,t}$	V_t	$\widehat{V}_{n,t}$	V_t	$\widehat{V}_{n,t}$
0.01	1.000	1.000	0.001	0.001	1.000	1.000	0.002	0.001
0.05	1.000	1.000	0.020	0.021	1.000	1.000	0.014	0.016
0.10	1.000	1.000	0.057	0.051	1.000	1.000	0.044	0.044
0.15	1.000	1.000	0.111	0.107	1.000	1.000	0.089	0.088
0.20	1.000	1.000	0.166	0.169	1.000	1.000	0.136	0.137
0.25	1.000	1.000	0.216	0.226	1.000	1.000	0.187	0.182
0.30	1.000	1.000	0.275	0.284	1.000	1.000	0.246	0.240
0.35	1.000	1.000	0.337	0.339	1.000	1.000	0.299	0.298
0.40	1.000	1.000	0.398	0.405	1.000	1.000	0.350	0.350
0.45	1.000	1.000	0.461	0.466	1.000	1.000	0.404	0.414
0.50	1.000	1.000	0.504	0.507	1.000	1.000	0.454	0.473

forward into the misspecification Cases 2 and 3 shown in Tables 3.5 and 3.6, where the rejection frequencies are close to their nominal levels.

3.3 Empirical Results

We examine our set inference framework on actual S&P 500 Index Options and high-frequency observations on the underlying.

Table 3.4: Monte Carlo Simulation: Size Control, 2 Year Sample, Case 1. This table shows the results of a Monte Carlo experiment, in which 2 years of options bid and ask quotes were simulated 2,000 times by pricing options on an underlying SVJ model generated via an Euler scheme. The Andrews-Soares (2010) confidence sets were computed each time, and the rejection frequencies were recorded for the nominal sizes given in column 1. The left panel shows simulation results using both $\mathbb{P}$ and $\mathbb{Q}$ measure restrictions; the right panel shows $\mathbb{Q}$ measure restrictions only. Within each panel, the table shows the effect of using estimated spot volatility $\widehat{V}_{n,t}$ under varying widths of the bid-ask spread.

	$\mathbb{P}$ and $\mathbb{Q}$ measure restrictions				$\mathbb{Q}$ measure restrictions only			
	Mid-quote		Full Spread		Mid-quote		Full Spread	
α	V_t	$\widehat{V}_{n,t}$	V_t	$\widehat{V}_{n,t}$	V_t	$\widehat{V}_{n,t}$	V_t	$\widehat{V}_{n,t}$
0.01	0.008	0.069	0.014	0.009	0.012	0.072	0.000	0.000
0.05	0.045	0.234	0.054	0.041	0.057	0.240	0.000	0.000
0.10	0.102	0.357	0.098	0.073	0.115	0.363	0.000	0.000
0.15	0.149	0.452	0.131	0.104	0.166	0.454	0.000	0.000
0.20	0.199	0.536	0.171	0.145	0.220	0.519	0.000	0.000
0.25	0.256	0.607	0.216	0.179	0.268	0.582	0.000	0.000
0.30	0.304	0.659	0.254	0.213	0.319	0.635	0.000	0.000
0.35	0.365	0.701	0.294	0.251	0.375	0.684	0.000	0.001
0.40	0.413	0.740	0.337	0.290	0.417	0.723	0.000	0.001
0.45	0.469	0.777	0.378	0.322	0.476	0.759	0.000	0.001
0.50	0.510	0.809	0.413	0.366	0.526	0.796	0.001	0.001

3.3.1 The Data

Our data are obtained from OptionMetrics and represent daily observations on S&P 500 Index Options spanning January 1996 to December 2010. In light of our discussion following Assumption C, a relevant set of instruments for our moment inequality approach is given by the indicator functions that categorize options according to strike and time-to-maturity. Andrews and Shi (2014) show that there is no loss of identification power when the number of categories (or "boxes") goes to infinity.

Table 3.5: Monte Carlo Simulation: Size Control, 2 Year Sample, Case 2. This table shows the results of a Monte Carlo experiment, in which 2 years of options bid and ask quotes were simulated 2,000 times by pricing options on an underlying SVJ model generated via an Euler scheme. The Andrews-Soares (2010) confidence sets were computed each time, and the rejection frequencies were recorded for the nominal sizes given in column 1. The left panel shows simulation results using both $\mathbb{P}$ and $\mathbb{Q}$ measure restrictions; the right panel shows $\mathbb{Q}$ measure restrictions only. Within each panel, the table shows the effect of using estimated spot volatility $\widehat{V}_{n,t}$ under varying widths of the bid-ask spread.

	$\mathbb{P}$ and $\mathbb{Q}$ measure restrictions				$\mathbb{Q}$ measure restrictions only			
	Mid-quote		Full Spread		Mid-quote		Full Spread	
α	V_t	$\widehat{V}_{n,t}$	V_t	$\widehat{V}_{n,t}$	V_t	$\widehat{V}_{n,t}$	V_t	$\widehat{V}_{n,t}$
0.01	0.028	0.106	0.011	0.010	0.033	0.108	0.000	0.000
0.05	0.163	0.378	0.047	0.031	0.181	0.367	0.000	0.000
0.10	0.304	0.578	0.082	0.064	0.330	0.577	0.000	0.000
0.15	0.430	0.703	0.112	0.089	0.451	0.690	0.000	0.000
0.20	0.536	0.790	0.148	0.117	0.552	0.775	0.000	0.001
0.25	0.628	0.850	0.185	0.148	0.637	0.838	0.000	0.001
0.30	0.708	0.890	0.217	0.178	0.704	0.891	0.000	0.001
0.35	0.762	0.916	0.249	0.215	0.772	0.918	0.001	0.001
0.40	0.816	0.945	0.286	0.251	0.828	0.944	0.001	0.001
0.45	0.861	0.964	0.324	0.285	0.868	0.966	0.001	0.002
0.50	0.896	0.976	0.365	0.322	0.896	0.979	0.003	0.002

However, for the purposes of empirical tractability, we settle on categorizing options into seven strike categories and three maturity categories, for a total of twenty-one option categories.

Summary statistics for each category of options is presented in Table 3.7. The categories were chosen to be representative of the full option smile observed on the three closest maturities for a given trading day and therefore make use of information on both deeply in-the-money (ITM) and deeply out-of-the-money (OTM) options.

Table 3.6: Monte Carlo Simulation: Size Control, 2 Year Sample, Case 3. This table shows the results of a Monte Carlo experiment, in which 2 years of options bid and ask quotes were simulated 2,000 times by pricing options on an underlying SVJ model generated via an Euler scheme. The Andrews-Soares (2010) confidence sets were computed each time, and the rejection frequencies were recorded for the nominal sizes given in column 1. The left panel shows simulation results using both $\mathbb{P}$ and $\mathbb{Q}$ measure restrictions; the right panel shows $\mathbb{Q}$ measure restrictions only. Within each panel, the table shows the effect of using estimated spot volatility $\widehat{V}_{n,t}$ under varying widths of the bid-ask spread.

	$\mathbb{P}$ and $\mathbb{Q}$ measure restrictions				$\mathbb{Q}$ measure restrictions only			
	Mid-quote		Full Spread		Mid-quote		Full Spread	
α	V_t	$\widehat{V}_{n,t}$	V_t	$\widehat{V}_{n,t}$	V_t	$\widehat{V}_{n,t}$	V_t	$\widehat{V}_{n,t}$
0.01	1.000	1.000	0.001	0.002	1.000	1.000	0.000	0.001
0.05	1.000	1.000	0.005	0.029	1.000	1.000	0.003	0.022
0.10	1.000	1.000	0.019	0.065	1.000	1.000	0.014	0.057
0.15	1.000	1.000	0.031	0.111	1.000	1.000	0.035	0.112
0.20	1.000	1.000	0.057	0.164	1.000	1.000	0.052	0.170
0.25	1.000	1.000	0.087	0.227	1.000	1.000	0.086	0.222
0.30	1.000	1.000	0.133	0.288	1.000	1.000	0.120	0.278
0.35	1.000	1.000	0.178	0.354	1.000	1.000	0.159	0.336
0.40	1.000	1.000	0.224	0.406	1.000	1.000	0.198	0.394
0.45	1.000	1.000	0.281	0.474	1.000	1.000	0.240	0.451
0.50	1.000	1.000	0.338	0.536	1.000	1.000	0.296	0.517

The table shows that the bid-ask spread is on average 200bp wide for the less liquid ITM options. Wider bid-ask spreads are actually observed when OTM calls are not converted to ITM via put-call parity, as we have done following practices in the existing literature (see for example Andersen et al. (2012)). These option categories result in an effective option sample of $7 \times 3 \times 3,462 = 72,702$ pairs of quotes. Over the same sample period, we also have 1-minute observations on S&P 500 futures that we use to construct our high-frequency estimates of spot volatility $\widehat{V}_{n,t}$ in (3.1.12).[18]

[18] We thank Sophia Zhengzi Li for providing us with the data.

The data was filtered following the standard practices in the literature (Andersen et al. (2012)). In particular, options with time-to-maturity less than 7 days are discarded, as are options with zero bids. The riskfree rate is interpolated from the observed LIBOR term structure and the dividend yield process is the one supplied by OptionMetrics.

3.3.2 Results

We present empirical results for several specifications. Table 3.8 shows various confidence set estimates for the general stochastic jump-diffusion model in (3.1.25). The top panel of the table shows results using only option data (the $\mathbb{Q}$-measure model), whereas the bottom panel shows results using additional identifying information about leverage and volatility-of-volatility from the high-frequency record on the underlying (the joint $\mathbb{P}$-$\mathbb{Q}$ model).

The column labeled PNT describes point estimates of the parameters obtained under the mid-quote point-identifying assumption under squared loss. The estimates accord well with findings in the existing literature, e.g. Eraker et al. (2003), Andersen et al. (2002), Chernov et al. (2003), Eraker (2004), Broadie et al. (2007). In particular, the options are clearly pricing volatility mean-reversion ($\kappa > 0$) and are implying strong negative skewness (ρ close to -1). The long-run variance $\overline{V}$ also corresponds well to nonparametric estimates of long-run risk-neutral variance (Bollerslev et al. (2011)). The option data are also clearly pricing jumps with a negative jump size mean ($\mu_J < 0$), with one jump occurring approximately every 1.3 years. Comparisons between the top and bottom panels of the PNT column show that the estimates

Table 3.7: Summary of Options Quote Data. This table shows time-averaged strike-to-spot ratios (K_j/S), and bid and ask quotes for strike categories K_j, $j = 1, \ldots, 7$, for a total of $7 \times 3 \times 3,462 = 72,702$ pairs of quotes. The data represent weekly observations on S&P500 index call options from January 2, 1996 to December 30, 2010. In-the-money call options were replaced by out-of-the-money puts converted by put-call parity. Days with fewer than seven strikes or three maturities were dropped.

Short Term	K/S	Bid-IV	*SD*	Ask-IV	*SD*	Spread	τ	*SD*	T
K_1	0.88	0.29	0.12	0.31	0.13	0.02	24.3	9.3	3,462
K_2	0.93	0.25	0.10	0.27	0.11	0.02	24.3	9.3	3,462
K_3	0.96	0.22	0.09	0.24	0.09	0.01	24.3	9.3	3,462
K_4	0.99	0.20	0.08	0.21	0.08	0.01	24.3	9.3	3,462
K_5	1.00	0.19	0.07	0.20	0.08	0.01	24.3	9.3	3,462
K_6	1.03	0.17	0.07	0.19	0.07	0.01	24.3	9.3	3,462
K_7	1.06	0.16	0.06	0.18	0.07	0.01	24.3	9.3	3,462
Med. Term									
K_1	0.83	0.30	0.11	0.32	0.12	0.02	52.5	11.3	3,462
K_2	0.89	0.27	0.10	0.28	0.10	0.01	52.5	11.3	3,462
K_3	0.94	0.24	0.08	0.25	0.09	0.01	52.5	11.3	3,462
K_4	0.98	0.21	0.07	0.22	0.08	0.01	52.5	11.3	3,462
K_5	0.99	0.20	0.07	0.21	0.08	0.01	52.5	11.3	3,462
K_6	1.03	0.18	0.06	0.19	0.07	0.01	52.5	11.3	3,462
K_7	1.09	0.16	0.06	0.17	0.06	0.01	52.5	11.3	3,462
Long Term									
K_1	0.80	0.30	0.11	0.32	0.11	0.02	88.7	29.5	3,462
K_2	0.87	0.27	0.09	0.28	0.09	0.01	88.7	29.5	3,462
K_3	0.93	0.24	0.08	0.25	0.08	0.01	88.7	29.5	3,462
K_4	0.98	0.21	0.07	0.22	0.07	0.01	88.7	29.5	3,462
K_5	0.99	0.21	0.07	0.22	0.07	0.01	88.7	29.5	3,462
K_6	1.04	0.18	0.06	0.19	0.06	0.01	88.7	29.5	3,462
K_7	1.10	0.16	0.05	0.17	0.06	0.01	88.7	29.5	3,462

Table 3.8: SVJ Baseline Set-Inference. Estimation and inference for SVJ model. PNT, E50, I50, E95, and I95 denote point estimates, mid-quote moment equality CS at size 50%, full bid-ask spread CS at size 50%, mid-quote moment equality CS at size 95%, and full bid-ask spread CS at size 95%, respectively. "Obj. fn." denotes value of the objective function $(S(\theta) - c(\theta, 1-\alpha))_+$.

$\mathbb{Q}$	PNT	E50	I50	E95	I95
κ	3.81	3.44	[1.45, 59.95]	[3.85, 131.64]	[1.26, 157.7]
$\overline{V}$	0.04	0.04	[0.02, 0.08]	[0.02, 0.06]	[0.01, 0.07]
ρ	-1.00	-1.00	[-1.00, -0.44]	[-1.00, -0.38]	[-1.00, -0.00]
v	0.56	0.51	[0.19, 2.11]	[0.34, 3.03]	[0.00, 3.40]
λ	0.76	0.69	[0.03, 0.94]	[0.02, 0.61]	[0.00, 1.11]
μ_J	-0.07	-0.07	[-0.78, -0.05]	[-0.98, -0.05]	[-1.47, 0.06]
σ_J	0.14	0.15	[0.00, 0.37]	[0.00, 0.45]	[0.00, 0.78]
Obj. fn.	2.93	12.11	0.00	0.00	0.00
$\mathbb{Q}$ and $\mathbb{P}$	PNT	E50	I50	E95	I95
κ	4.77	3.34	[1.45, 41.89]	[3.11, 86.17]	[1.33,139.2]
$\overline{V}$	0.04	0.04	[0.02, 0.08]	[0.02, 0.06]	[0.01,0.08]
ρ	-0.98	-0.99	[-1.00, -0.41]	[-1.00, -0.38]	[-1.00,-0.00]
v	0.60	0.51	[0.21, 1.51]	[0.34, 2.26]	[0.00,3.4]
λ	0.67	0.67	[0.07, 0.94]	[0.01, 0.72]	[0.00,1.11]
μ_J	-0.07	-0.07	[-0.52, -0.04]	[-0.97, -0.04]	[-1.47, 0.06]
σ_J	0.14	0.15	[0.00, 0.35]	[0.00, 0.45]	[0.00, 0.78]
Obj fn.	2.98	11.54	0.00	0.00	0.00

are little changed by the additional identifying restriction.

The remaining columns of Table 3.8 show estimated 50% and 95% confidence sets under both mid-quote equality restrictions as well as full bound width restrictions. That is, the moment inequality framework in Section 3.1 is specialized to the case when the bid and ask quotes are artificially collapsed to the mid-quote (columns labeled E50 and E95), whereas the columns labeled I50 and I95 correspond to the full moment inequality setup and represent 50% and 95% CS's, respectively. The results show a rejection of the SVJ model at the 50% level when the mid-quote equality is artificially imposed. The result disappears when the mid-quote assumption is relaxed. More generally, Table 3.8 reveals large confidence sets, as indicated by the intervals on each parameter. In interpreting the results, however, one should remain cautious of the fact that indicated intervals represent projections of a 7-dimensional confidence set onto individual parameters.

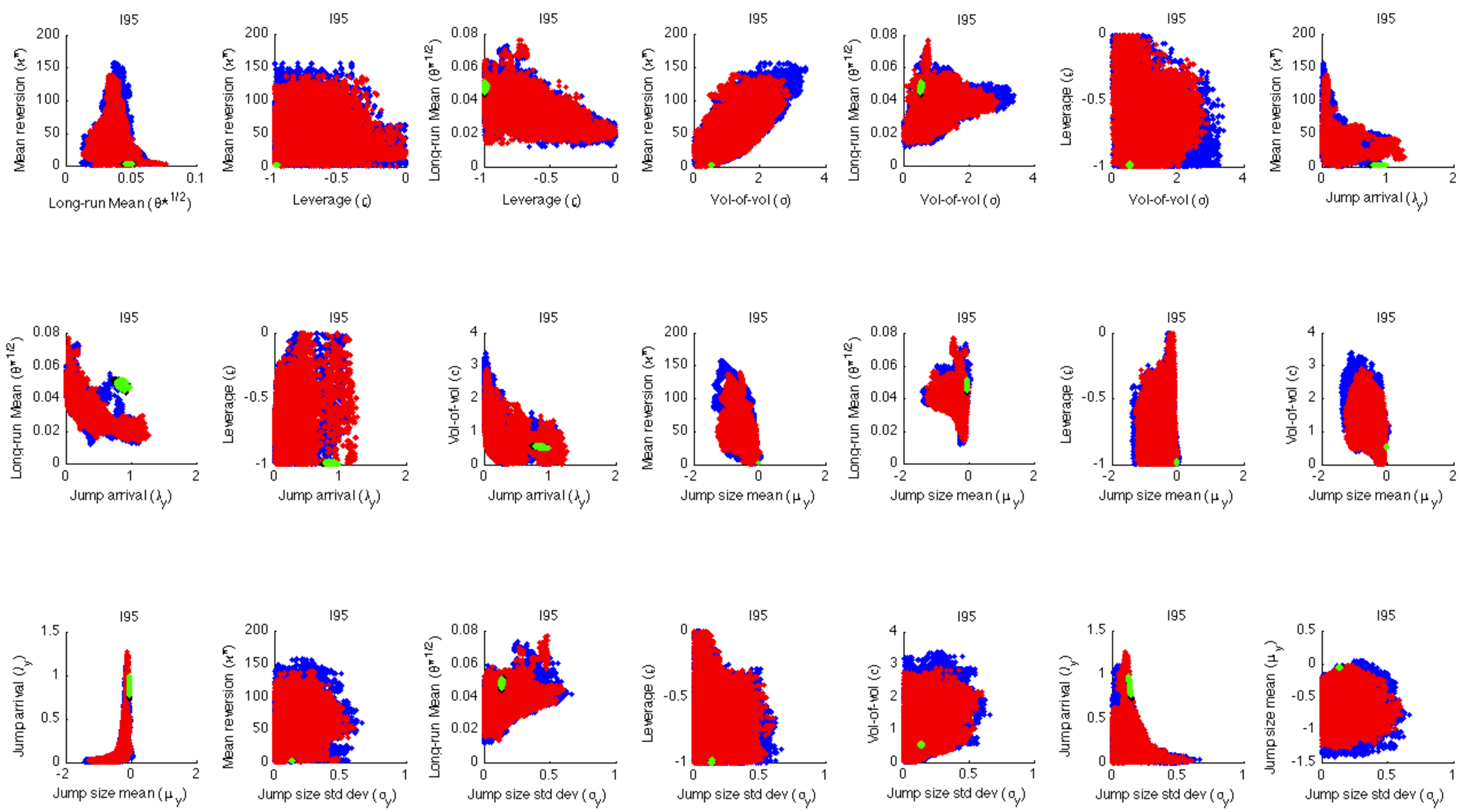

FIGURE 3.2: Confidence Set Projections onto Coordinate Pairs. The confidence sets (3.1.13) are projected onto pairs of parameters and plotted for the SVJ model in (3.1.25). Dark (blue) shading corresponds to confidence sets using only $\mathbb{Q}$-measure bid-ask bound restrictions, whereas red (light) shading corresponds to confidence sets using joint $\mathbb{P}$-$\mathbb{Q}$ restrictions from Subsection 3.1.4.

To better understand the shape of the 7-dimensional CS, we plot its projections onto pairs of parameter coordinates in Figure 3.2. The figure reveals new and interesting tradeoffs between option model parameters in the SVJ model. That is, observed option quotes appear to admit a variety of parameter configurations. For example, the fourth panel in the first row of Figure 3.2 suggests that option prices do not appear to distinguish {high mean-reversion, high vol-of-vol} from {low mean-reversion, low vol-of-vol}. Similarly, the first panel in the third row shows that option prices can be well represented by either a {high-intensity, small jump-size} specification or a {low-intensity, large jump-size}. The figure also reveals how the volatility and jump process interact: A model with high long-run variance is consistent with observed option quotes, as long as the jump-intensity goes to zero (second row, first column). Conversely, the long-run variance is permitted to assume smaller values when the jump arrival intensity goes up. A similar relationship appears to hold for the vol-of-vol parameter as well.

In general, Table 3.8 and Figure 3.2 show large estimated confidence sets. To sharpen the inference, we introduce additional moment restrictions of the form in Assumption C'. In particular, rather than using only test functions corresponding to indicator functions in the strike and maturity dimension, we introduce a smoothed indicator function in the volatility space that will require option prices to satisfy the bid-ask bounds in three distinct volatility states: high, medium, and low volatility periods as distinguished by their sample tertiles. These additional test functions effectively triple the number of moment inequalities used for inference and (perhaps not surprisingly) result in smaller estimated confidence sets.

Table 3.9: Set estimates with volatility conditioning information. Estimation and inference for SVJ model with volatility weighting function. PNT, E50, I50, E95, and I95 denote point estimates, mid-quote moment equality CS at size 50%, full bid-ask spread CS at size 50%, mid-quote moment equality CS at size 95%, and full bid-ask spread CS at size 95%, respectively. "Obj. fn." denotes value of the objective function $(S(\theta) - c(\theta, 1 - \alpha))_+$.

$\mathbb{Q}$	PNT	E50	I50	E95	I95
κ	3.30	3.71	3.02	3.20	[2.78, 4.05]
$\overline{V}$	0.05	0.05	0.05	0.05	[0.044, 0.053]
ρ	-1.00	-1.00	-1.00	-1.00	[-1.00, -0.98]
v	0.56	0.61	0.55	0.56	[0.51, 0.64]
λ	0.93	0.96	0.81	0.96	[0.68, 0.97]
μ_J	-0.04	-0.04	-0.04	-0.04	[-0.05, -0.04]
σ_J	0.13	0.13	0.14	0.13	[0.13, 0.15]
Obj fn.	18.26	390.94	78.59	270.22	0.00

$\mathbb{Q}$ and $\mathbb{P}$	PNT	E50	I50	E95	I95
κ	3.47	3.53	3.090	3.570	[2.71, 3.34]
$\overline{V}$	0.05	0.05	0.050	0.050	[0.045,0.052]
ρ	-0.96	-1.00	-1.000	-1.000	[-1.00, -0.98]
v	0.54	0.57	0.550	0.570	[0.50, 0.58]
λ	0.78	1.03	0.800	1.000	[0.73, 0.94]
μ_J	-0.05	-0.04	-0.040	-0.040	[-0.05, -0.04]
σ_J	0.14	0.12	0.140	0.130	[0.13, 0.15]
Obj fn.	18.26	397.33	77.56	267.91	0.00

To illustrate, Figure 3.2 also shows estimated 95% CS's for the volatility test function case in green (light shading). The resulting CS's display significant reductions in the size of the new CS's, suggesting that the volatility test functions are ruling out many of the observationally similar parameter estimates from the original model. Figure 3.3 zooms into the parameter estimates for the volatility test function confidence sets and reveals parameter sets that are close to the point estimates obtained under the mid-quote assumption. However, in a striking deviation from existing findings in the literature (Andersen et al. (2012)), the SVJ model is not rejected at the 5% level when the full width of bid and ask quotes is used, whereas point estimates specializing to the option mid-quote are rejected (Table 3.9). Our results suggest that the mid-quote assumption imports enough information into an option model that it may overturn conclusions obtained under more conservative assumptions on the data generating process.

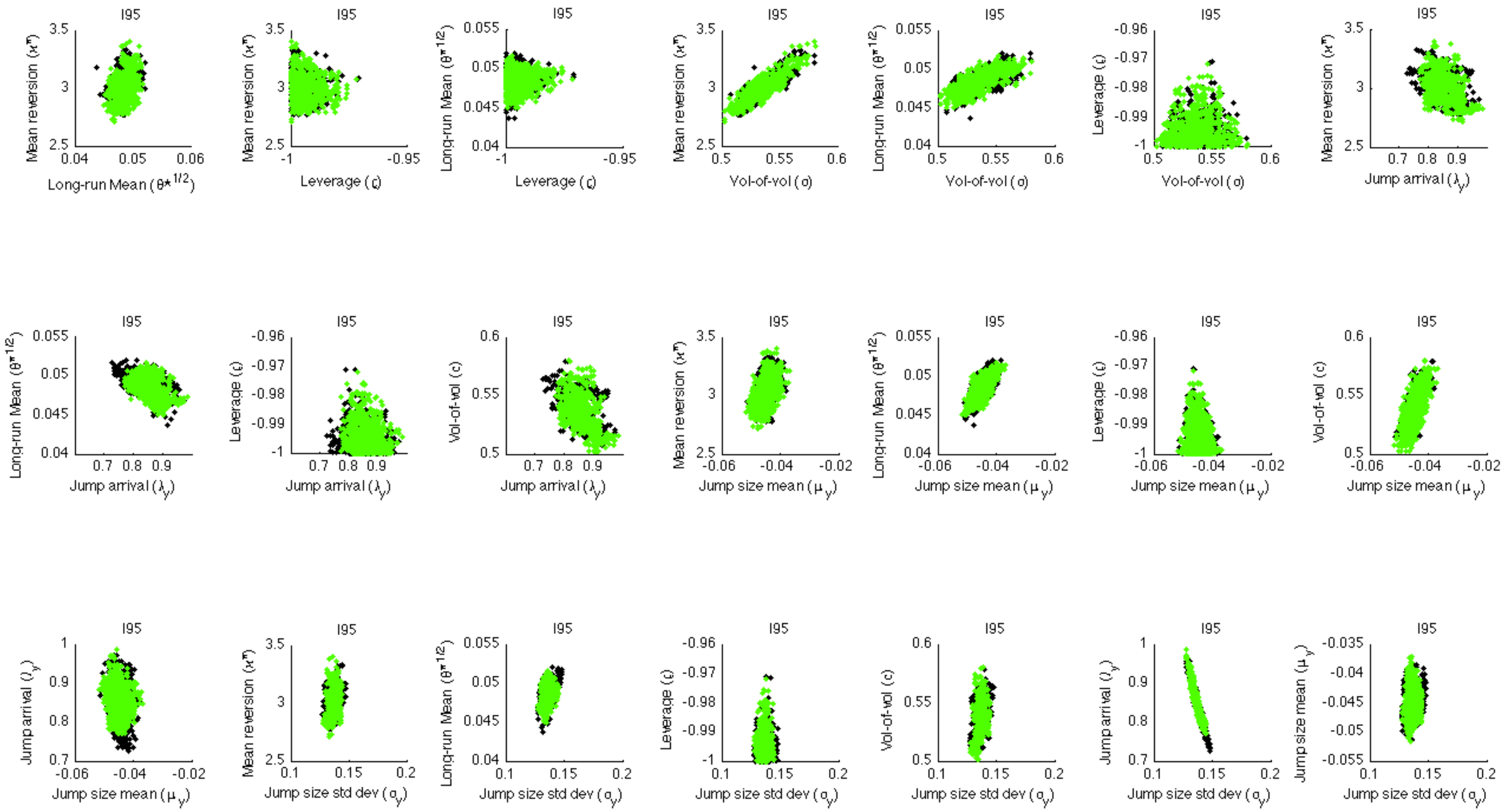

FIGURE 3.3: Estimated confidence sets using volatility test functions from Assumption C'. The confidence sets (3.1.13) are projected onto pairs of parameters and plotted for the SVJ model in (3.1.25). Dark (black) shading corresponds to confidence sets using only $\mathbb{Q}$-measure bid-ask bound restrictions, whereas green (light) shading corresponds to confidence sets using joint $\mathbb{P}$-$\mathbb{Q}$ restrictions from Subsection 3.1.4.

3.4 Conclusion

This paper examines inference on option model parameters in the bid-ask quote setting. Specifically, we propose an econometric framework that explicitly recognizes bid and ask quotes as interval observations on the efficient option price, which naturally gives rise to a lack of point-identification: that is, a situation in which multiple option model parameters are consistent with observed option quotes. Our framework relies on moment inequalities that bound model-implied option prices between observed bid and ask quotes, thus avoiding the untestable microstructure restriction of equating efficient option prices to the option mid-quote. We argue that the mid-quote point-identifying assumption is especially relevant in an empirical option pricing setting, where the illiquidity of certain deep in-the-money options can induce significant bid-ask spreads.

Our framework extends the existing econometric literature on set-inference by admitting moment functions that depend on a latent variable (spot variance) that must be estimated in a first stage from high-frequency data on the option's underlying. This extension allows us to conduct inference on a general affine stochastic volatility jump-diffusion model of Duffie et al. (2000) within a partial identification setting. By construction, the inference is robust to microstructure misspecifications by allowing us to remain agnostic about the relationship between bid-ask quotes and the efficient price. We also illustrate that the framework is general enough to accommodate additional identifying restrictions on certain option pricing parameters that are invariant to the change of measure.

Monte Carlo simulations show that the use of estimated spot variance in place of latent spot variance provides accurate coverage of the true pricing parameter under empirically realistic sample sizes. Our empirical exercise shows that relaxing the mid-quote assumption results in large estimated parameter sets that reveal novel relationships among option model parameters. We also show that the informativeness of inference can be restored by incorporating certain variance test functions into the moment inequality framework.

4

A Sieve Application to Estimating Quantile Risk Premia

Introduction

This chapter uses the sieve framework developed in Chapter 2 to examine the time series properties of risk-neutral return quantiles. An economic motivation for examining these return quantiles is that they are informative about tail risk premia. To this end, I estimate the objective-measure counterparts to the risk-neutral quantiles and show that their difference is related to a risk-premium on binary options that pay off \$1 in case of moves in the underlying asset of a given size.

The quantities defined in this chapter are related to the state-price of conditional quantiles (SPOCQ), defined in Metaxoglou and Smith (2013). The analysis differs here in the $\mathbb{Q}$-measure estimation, since they employed a lognormal mixture to obtain estimates of the risk-neutral distribution, whereas here I rely on the sieve methods of Chapter 2 to provide a complementary view on return quantiles. The methods discussed here also differ from Metaxoglou and Smith (2013) in that I examine the risk-premium interpretation associated with their SPOCQ and furthermore make direct comparisons of the quantiles of the $\mathbb{P}$- and $\mathbb{Q}$-measure distributions. Lastly, I also propose a regression framework for forecasting excess returns based on a decomposition of the equity risk-premium into its quantile constituents.

In what follows, $\mathbb{P}$-measure return quantiles were estimated using the CAViaR methods of Engle and Manganelli (2004), augmented with intra-month information on realized daily squared returns. The empirical findings of the chapter suggest a pronounced presence of risk premia in the extremes of the return distribution, underscoring existing results on the drivers of the equity risk premium (Bollerslev and Todorov (2011)). However, in contrast to the variance risk premium literature discussed in Chapter 2, the compensation for these distributional risks appear asymmetric across quantiles.

4.1 Motivating Quantile Risk Premia

I define the equity premium as the τ-period ahead expected excess return $E_t^{\mathbb{P}}[r_{t,t+\tau}] - r^f_{t,t+\tau}$ on the aggregate market index S_t, where $r_{t,t+\tau} = (S_{t+\tau}/S_t) - 1$ and $r^f_{t,t+\tau}$ denotes the riskfree rate over the corresponding horizon. The expectation operator for the market return in this definition is under the objective measure, or $\mathbb{P}$-measure.

On the other hand, it is well-known that the riskfree rate $r^f_{t,t+\tau}$ can also be written as a conditional expectation of returns, but against the risk-neutral measure $\mathbb{Q}$, that is

$$r^f_{t,t+\tau} = E_t^{\mathbb{Q}}[r_{t,t+\tau}], \tag{4.1.1}$$

giving rise to the definition

$$ERP_{t,t+\tau} \equiv E_t^{\mathbb{P}}[r_{t,t+\tau}] - E_t^{\mathbb{Q}}[r_{t,t+\tau}]. \tag{4.1.2}$$

Equation (4.1.2) shows that the equity premium depends critically on two expectations of future returns, $E_t^{\mathbb{P}}[r_{t,t+\tau}]$ and $E_t^{\mathbb{Q}}[r_{t,t+\tau}]$. Each of these, in turn, can be viewed as functionals on the conditional distributions that generated the expectation. That is, letting $F_{t+\tau|t}$ denote a conditional CDF and defining

$$T\Big(F_{t+\tau|t}\Big) = \int_{-\infty}^{\infty} r\, dF_{t+\tau|t}(r),$$

we see that the equity premium is given

$$ERP_{t,t+\tau} = T\left(F_{t+\tau|t}^{\mathbb{P}}\right) - T\left(F_{t+\tau|t}^{\mathbb{Q}}\right). \tag{4.1.3}$$

The focus of this paper is on direct comparisons of the distributions $F_{t+\tau|t}^{\mathbb{P}}$ and $F_{t+\tau|t}^{\mathbb{Q}}$ through their quantiles, rather than through the lens of the operator $T(\cdot)$.

4.2 Forecasting Objective and Risk-Neutral Quantiles

Since the distributions $F_{t+\tau|t}^{\mathbb{P}}$ and $F_{t+\tau|t}^{\mathbb{Q}}$ in (4.1.3) are unobserved, they must be estimated from historical data. I discuss the estimation of each in turn.

4.2.1 Forecasting $\mathbb{Q}$-measure Return Quantiles: The Method of Sieves

I use the method of sieves proposed in Chapter 2 to estimate quantiles of $F_{t+\tau|t}^{\mathbb{Q}}$. The method of sieves exploits the structure embedded in the risk-neutral valuation equation and allows for the estimation of the entire function $F_{t+\tau|t}^{\mathbb{Q}}$ for a fixed τ, even if options with maturity τ are unobserved.

Following the same arguments from Chapter 2, the price of a put option with strike κ and maturity τ is given by[1]

$$P_0(\kappa, \tau) = e^{-r\tau} \int_0^{\kappa} [\kappa - S] f_0^{\mathbb{Q}}(S|\tau) dS.$$

Next, the following convenient change of variables is adopted,

$$\log\left(\frac{S}{S_0}\right) = \mu(\mathbf{Z}) + \sigma(\mathbf{Z})Y,$$

where $\mathbf{Z} = (\kappa, \tau, r, q)$ holds the strike, maturity, risk-free rate, and dividend yield, and where $\mu(\mathbf{Z}) = (r - q - \sigma^2/2)\tau$ and $\sigma(\mathbf{Z}) = \sigma\sqrt{\tau}$. For empirical implementation, σ is chosen as an interpolated implied volatility of an at-the-money τ-maturity option. Under this change of variables, the valuation equation becomes

$$P_0(\kappa, \tau) \equiv P_Y(f_0, \mathbf{Z}) = e^{-r\tau} \int_0^{d(\mathbf{Z})} \left(\kappa - S_0 e^{\mu(\mathbf{Z}) + \sigma(\mathbf{Z})Y}\right) f_0(Y|\tau) dY, \qquad (4.2.1)$$

where $d(\mathbf{Z}) = \frac{\log(\kappa/S) - \mu(\mathbf{Z})}{\sigma(\mathbf{Z})}$. The procedure in Chapter 2 delivers estimates for $f_0(Y|\tau)$, which are related to the density of interest by the simple Jacobian transformation

$$f_0^{\mathbb{Q}}(s|\tau) = (s\sigma(\mathbf{Z}))^{-1} f_0(s|\tau).$$

To estimate $f_0^{\mathbb{Q}}(s|\tau)$, one constructs Hermite polynomial expansions using candidate densities of the form

$$f_K(y|\tau) = \sum_{k=0}^{2K_y} \gamma_k(B, \tau) H_k(y) \phi(y), \qquad (4.2.2)$$

[1] Exactly as in Chapter 2, the dependence of $f_0^{\mathbb{Q}}(S|\tau)$ on a possibly unobserved state vector $\mathbf{V}$ that generated the time-t information set, i.e. $f_0^{\mathbb{Q}}(S|\tau, \mathbf{V} = \mathbf{v}_0)$, is suppressed, since the strategy here is to estimated $f_0^{\mathbb{Q}}(S|\tau)$ separately for each option cross section.

where $K = (K_y + 1)(K_\tau + 1)$ and where

$$\gamma_k(B, \tau) = \frac{\alpha(B, \tau)' A_k \alpha(B, \tau)}{\alpha(B, \tau)' \alpha(B, \tau)},$$

and $\alpha(B, \tau) = (\sum_{j=0}^{K_\tau} \beta_{0j} H_j(\tau),\ \ldots,\ \sum_{j=0}^{K_\tau} \beta_{K_y j} H_j(\tau))'$. A_k are matrices of constants derived in León and Sentana (2009). $H_k(y)$ are Hermite polynomials of order k that are orthonormal with respect to $\exp(-y^2/2)\, dy$. B is a $(K_y + 1) \times (K_\tau + 1)$ matrix of coefficients. I denote its vectorized representation as $\beta = vec(B)$ and then estimate β by solving the least squares problem

$$\begin{aligned} \widehat{\beta} &= \arg \min_{\beta \in \mathbb{R}^K} \left\{ \frac{1}{n} \sum_{i=1}^{n} \Big[P_i - P_Y(\beta, \mathbf{Z}_i) \Big]^2 W_i \right\} \\ s.t. &\sum_{k=0}^{K_y} \sum_{j=0}^{K_\tau} \beta_{kj}^2 = 1. \end{aligned} \tag{4.2.3}$$

The contribution of Chapter 2 was to show that if $K \to \infty$ as the cross-section of options grew ($n \to \infty$), then $d(\widehat{f}_K, f_0) \to 0$, where $\widehat{f}_K(y|\tau) = \sum_{k=0}^{2K_y} \gamma_k(\widehat{B}, \tau) H_k(y) \phi(y)$. The simulations in Chapter 2 also demonstrate that selecting K by minimizing the BIC is effective for obtaining nonparametric coverage of certain portfolios of options that depend on f_K.

Because the focus here is on f_K and its CDF and not option portfolio inference, I fix $K_y = 5$ and $K_\tau = 1$ for computational ease. The low order on τ expansions is supported by using a restricted option panel, e.g. by only using the first three available maturities in the option panel in order to estimate the 30-day ahead return distribution. The minimization in (4.2.3) is then conducted on a weekly sample of S&P 500 Index options spanning 1996 to 2013. A total of 883 optimizations

Table 4.1: Time-Averaged Sieve Estimates. The sieve least-squares problem in (4.2.3) with $K_y = 5$ and $K_\tau = 1$ is solved on a weekly sample of S&P 500 Index options spanning 1996 to 2013, resulting in 883 coefficient estimates for $\widehat{B}$, where $\widehat{\beta} = vec(\widehat{B})$. The table reports time-averages and standard deviations for the squares of the coefficient matrices $\widehat{B}$.

	Mean $(B \circ B)$		Std $(B \circ B)$	
$K_y \backslash K_\tau$	τ^0	τ^1	τ^0	τ^1
0	0.127	0.660	0.221	0.297
1	0.001	0.017	0.004	0.042
2	0.005	0.054	0.011	0.054
3	0.001	0.024	0.003	0.044
4	0.001	0.023	0.004	0.040
5	0.007	0.081	0.015	0.069

corresponding to the number of weeks in the sample are performed for this sample, and time-averaged squares of coefficient estimates are reported in Table 4.1. The table reports squares of coefficients because of the constraint in (4.2.3), which gives an indication of how much weight the option data place on each Hermite polynomial term. The table clearly shows that on average, most of the weight is placed on the leading expansion term, and that this leading weight changes significantly over time. The higher-order Hermite polynomial terms are nonzero on average, with more emphasis placed on the first-order τ expansion. With estimates of $\widehat{B}$ in hand, one can easily construct estimates of $F^{\mathbb{Q}}_{t+\tau|t}$ using the closed-form relation

$$\begin{aligned} \widehat{\mathbb{Q}}_K(S_{t+\tau} \leqslant \kappa|\tau) &= \int_0^{d(\mathbf{Z})} \widehat{f}_K(x|\tau)\,dx \\ &= \Phi(d(\mathbf{Z})) - \sum_{k=1}^{2K_x} \frac{\gamma_k(\widehat{B},\tau)}{\sqrt{k}} H_{k-1}(d(\mathbf{Z}))\phi(d(\mathbf{Z})), \end{aligned} \tag{4.2.4}$$

$\widehat{\mathbb{Q}}_K(A) = \int_A \widehat{f}_K(x|\tau)dx$ is the estimated risk-neutral measure.[2] One then obtains the return distribution estimate

$$\widehat{F}^{\mathbb{Q}}_{t+\tau|t}(r) = \widehat{\mathbb{Q}}_K\left(\frac{S_{t+\tau}-S_t}{S_t} \leqslant \frac{\kappa - S_t}{S_t}\middle|\tau\right), \tag{4.2.5}$$

where $r = (\kappa - S_t)/S_t$. The above procedure yields a time-series of $\tau = 30$-day ahead distributions $\widehat{F}^{\mathbb{Q}}_{t+\tau|t}(r)$, from which quantiles are readily computed by inversion.

4.2.2 Forecasting $\mathbb{P}$-measure Return Quantiles: CAViaR

To estimate the quantiles of $F^{\mathbb{P}}_{t+\tau|t}$, I apply the CAViaR model of Engle and Manganelli (2004). To ease notation, let $\tau = 1$ month, and let $Quant^{\mathbb{P}}_{r_{t+1}|t}(\alpha)$ denote the level α quantile of the return distribution $F^{\mathbb{P}}_{t+\tau|t}$. That is, let

$$Quant^{\mathbb{P}}_{r_{t+1}|t}(\alpha) = \inf\{r : \alpha \leqslant F^{\mathbb{P}}_{t+1|t}(r)\}. \tag{4.2.6}$$

The CAViaR model specifies dynamics for conditional quantiles $Quant^{\mathbb{P}}_{r_{t+1}|t}(\alpha)$ in a manner analogous to the GARCH specification of Bollerslev (1986) for conditional variances. Thus, given returns and observables $\{r_t, x_t\}_{t=1}^{T}$, one defines

$$f_{t+1}(\beta) \equiv f_{t+1}(x_t, \beta_\alpha) \equiv Quant^{\mathbb{P}}_{r_{t+1}|t}(\alpha). \tag{4.2.7}$$

[2] See Chapter 2 for a derivation of this expression.

Then a generic CAViaR specification takes the form

$$f_{t+1}(\beta) = \beta_0 + \sum_{i=0}^{q} \beta_i f_{t-i}(\beta) + \sum_{j=0}^{p} \beta_j \ell_{t-j}(x_{t-j})$$

$$r_t = f_t(\beta^0) + \varepsilon_{\alpha t} \tag{4.2.8}$$

$$Quant^{\mathbb{P}}_{\varepsilon_{\alpha t}|x_{t-1}}(\alpha) = 0.$$

Thus, conditional return quantiles are allowed to depend on its own lags as well as lags of covariates x_t. For the empirics below, I consider the specialization

$$f_{t+1}(\beta) = \beta_0 + \beta_1 f_t(\beta) + \beta_2 |r_t| + \beta_3 RV_{t-1,t}, \tag{4.2.9}$$

where $RV_{t-1,t}$ denotes the realized variance, or sum of squared daily (intra-month) returns between $t-1$ and t. The idea behind this specification is to let large intra-month variances affect next period's quantiles.

Then, the model in (4.2.8) with quantile dynamics (4.2.9) is optimized to yield the coefficient vector

$$\widehat{\beta}_\alpha \equiv \arg\min_{\beta} \frac{1}{T} \sum_{t=1}^{T} \big[\alpha - 1[r_t \leqslant f_t(\beta)]\big][r_t - f_t(\beta)], \tag{4.2.10}$$

which one then uses to forecast the return quantile $f_{t+1}(\widehat{\beta}_\alpha)$. To obtain multiple α quantiles, I re-estimate this model for each α of interest to obtain a forecast of a different conditional quantile.

It is worth noting that in order for $t+1$ to denote a 30-day ahead $\mathbb{P}$-measure quantile, one must use a historical time series with 30-day increments. Since this amounts to a loss of intra-month information, the inclusion of $RV_{t-1,t}$ is designed to compensate by incorporating intra-month variation. Furthermore, for my empirical analysis, I require weekly observations on the 30-day ahead $\mathbb{P}$-measure quantiles. I therefore estimate a CAViaR model on a weekly expanding-window sample, which produces a new coefficient vector estimate $\widehat{\beta}_\alpha$ for each week. The results of these estimations are summarized in Table 4.2, which shows the average of coefficients obtained for each of the 30-day ahead forecasting problems. A noteworthy feature of Table 4.2 is the higher persistence in upper quantiles than for lower quantiles.

To aid in the estimation, the coefficients on realized variance and absolute returns were constrained to be negative for quantiles below the median. Furthermore, the coefficient on lagged quantiles was constrained to be positive to enforce continuity. These constraints were imposed in order to help the numerical optimizer converge to stable coefficient estimates. Indeed, Engle and Manganelli (2004) themselves provide a lengthy discussion on optimizing the CAViaR model, whose numerical instability can easily lead to different optima that depend on the starting location of the initial parameter vector. The results reported here were obtained using a genetic algorithm.

4.3 Quantile Risk Premia

I first document the time series properties of the $\mathbb{P}$- and $\mathbb{Q}$-measure quantiles estimated in the preceding section. I then show how the differences between corresponding $\mathbb{P}$- and $\mathbb{Q}$-measure quantiles is related to a risk premium on a binary option that pays \$1 for a given return value.

Table 4.2: Averaged CAViaR Estimates. The model in (4.2.8) with quantile dynamics (4.2.9) is estimated on an expanding-window sample by solving the minimization problem in (4.2.10). The sample consists of weekly forecasts of a 30-day ahead return quantile. Average coefficient estimates are shown alongside the unconditional sample probability of returns realizing below the indicated quantile.

α	$\mathbb{P}(r_t \leqslant f_t(\beta))$	const	$f_t(\beta)$	$\lvert r_t \rvert$	$RV_{t-1,t}$
0.01	0.01	-0.02	0.33	-0.22	-1.20
0.05	0.06	-0.02	0.46	-0.12	-0.34
0.10	0.12	-0.01	0.39	-0.11	-0.24
0.20	0.22	-0.01	0.21	-0.10	-0.18
0.30	0.31	0.00	0.11	-0.10	-0.13
0.40	0.38	0.01	0.10	-0.10	-0.10
0.50	0.51	0.00	0.35	0.10	0.10
0.60	0.61	0.00	0.42	0.10	0.12
0.70	0.70	0.00	0.53	0.10	0.15
0.80	0.81	0.01	0.55	0.10	0.22
0.90	0.93	0.01	0.52	0.10	0.36
0.95	0.97	0.01	0.46	0.10	0.58
0.99	1.00	0.02	0.45	0.20	0.70

4.3.1 $\mathbb{P}$- and $\mathbb{Q}$-measure Quantile Time Series

Figure 4.1 displays the time series of quantile estimates associated to 13 different probabilities, 0.01, 0.05, 0.1, 0.2, ..., 0.8, 0.9, 0.95, and 0.99. The time-series are clearly capturing the volatility-clustering effects commonly reported in the realized variance literature. That is, high- and low-volatility periods appear to persist in the quantile data as they do in volatility estimation literature (see, for example, Andersen et al. (2003) and Corsi (2009)). More strikingly, however, are the apparent quantile asymmetries that emerage in particularly volatile times. The financial crisis period from 2008 to 2010 is displaying significant skewness in the left-tail of both $\mathbb{P}$- and $\mathbb{Q}$-measure return quantiles. Not surprisingly, however, these extreme quantile estimates are each displaying significantly higher own-volatility that is likely to be affected by a lack of power in the estimation step.

These qualitative observations are underscored in Table 4.3, which highlights several interesting features. First, extreme $\mathbb{Q}$-measure quantiles appear more pronounced than their $\mathbb{P}$-measure counterpart, a feature that can be interpreted as evidence for a quantile risk premium (discussed further below). Moreover, these differences appear clearly more pronounced in the left tail than in the right. Second, the further one travels from the median, the more volatile each time series becomes, with a drop in autocorrelation at the furthest extremes. Third, and perhaps not surprisingly, each conditional distribution is showing pronounced signs of left-skewness.

Table 4.3: Summary statistics for the return quantile time series.

	$\mathbb{P}$-measure			$\mathbb{Q}$-measure			$\mathbb{P}-\mathbb{Q}$		
	Mean	Std	Auto	Mean	Std	Auto	Mean	Std	Auto
0.01	-0.14	0.06	0.82	-0.18	0.07	0.74	0.04	0.05	0.19
0.05	-0.08	0.03	0.89	-0.10	0.04	0.88	0.02	0.02	0.59
0.10	-0.05	0.02	0.90	-0.07	0.03	0.90	0.02	0.02	0.67
0.20	-0.03	0.01	0.84	-0.04	0.02	0.89	0.01	0.01	0.63
0.30	-0.02	0.01	0.80	-0.02	0.01	0.77	0.01	0.01	0.37
0.40	0.00	0.01	0.81	-0.01	0.01	0.48	0.00	0.01	0.32
0.50	0.01	0.01	0.82	0.01	0.01	0.31	0.01	0.01	0.52
0.60	0.02	0.01	0.83	0.02	0.01	0.78	0.00	0.01	0.34
0.70	0.03	0.01	0.86	0.03	0.01	0.92	0.00	0.01	0.37
0.80	0.05	0.02	0.91	0.05	0.02	0.92	0.00	0.01	0.48
0.90	0.07	0.02	0.93	0.07	0.03	0.83	0.00	0.02	0.36
0.95	0.09	0.03	0.91	0.09	0.04	0.79	0.00	0.02	0.34
0.99	0.12	0.04	0.84	0.15	0.07	0.71	-0.02	0.05	0.41

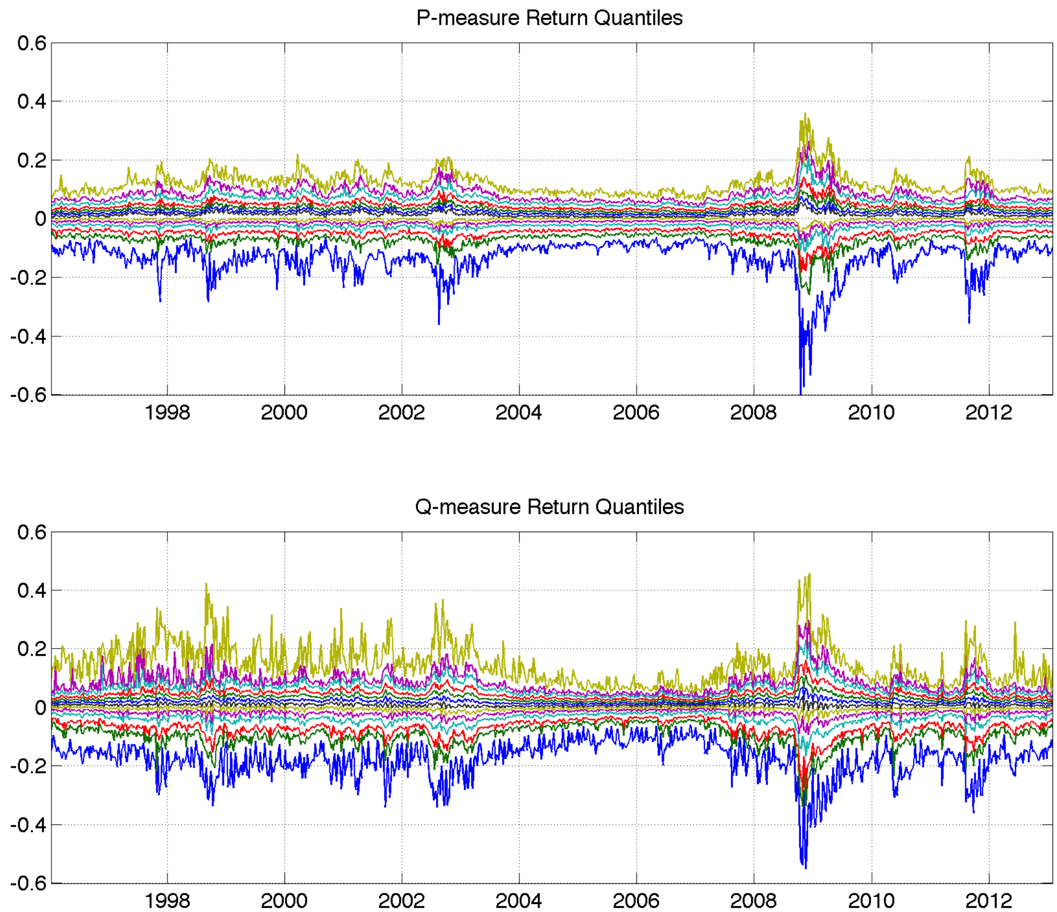

FIGURE 4.1: Weekly time-series of 30-day-ahead $\mathbb{P}$ and $\mathbb{Q}$-measure return quantiles, 1996–2013. The procedure in Section 4.2 is implemented to obtain quantiles estimates of the $\tau = 30$-day-ahead return distributions, $F^{\mathbb{P}}_{t+\tau|t}(r)$ and $F^{\mathbb{Q}}_{t+\tau|t}(r)$, introduced in Section 4.1. The quantiles displayed correspond to 13 probabilities of 0.01, 0.05, 0.1, 0.2, ..., 0.8, 0.9, 0.95, and 0.99.

4.3.2 Relation to the Risk Premium on a Binary Option

Differences in the quantiles of the $\mathbb{P}$- and $\mathbb{Q}$-measure distributions are related to risk premia on certain binary options. That is, let $r^{\mathbb{P}}_{\alpha}$ denote the α-quantile of $F^{\mathbb{P}}_{t+\tau|t}$. Then, by definition, one has

$$\alpha = F^{\mathbb{P}}_{t+\tau|t}(r^{\mathbb{P}}_{\alpha}) = \mathbb{P}_t(r_{t,t+\tau} \leqslant r^{\mathbb{P}}_{\alpha}) = \mathbb{P}_t(S_{t+\tau} \leqslant S_t(r^{\mathbb{P}}_{\alpha} + 1)).$$

The $\mathbb{Q}$-measure quantile of the same $r_\alpha^{\mathbb{P}}$ can also be obtained, yielding

$$\alpha = F^{\mathbb{Q}}_{t+\tau|t}(r_\alpha^{\mathbb{P}}) = \mathbb{Q}_t(r_{t,t+\tau} \leqslant r_\alpha^{\mathbb{P}}) = \mathbb{Q}_t(S_{t+\tau} \leqslant S_t(r_\alpha^{\mathbb{P}} + 1)),$$

The quantity on the right-hand side can be identified by the price of a binary option with strike $\kappa = S_t(r_\alpha^{\mathbb{P}} + 1)$, since $E_t^{\mathbb{Q}}\big(1[S_{t+\tau} \leqslant S_t(r_\alpha^{\mathbb{P}} + 1)]\big) = \mathbb{Q}_t(S_{t+\tau} \leqslant S_t(r_\alpha^{\mathbb{P}} + 1))$. In the absence of such options, one can use the sieve procedure outlined above to estimate $\mathbb{Q}_t(S_{t+\tau} \leqslant S_t(r_\alpha^{\mathbb{P}} + 1))$ from a panel of plain vanilla European options.

In light of this, I adopt the term *quantile risk premium* to refer to the following quantity,

$$QRP_{t+\tau|t}(\alpha) = e^{-r\tau}\mathbb{P}_t(S_{t+\tau} \leqslant S_t(r_\alpha^{\mathbb{P}} + 1)) - e^{-r\tau}\mathbb{Q}_t(S_{t+\tau} \leqslant S_t(r_\alpha^{\mathbb{P}} + 1)) \tag{4.3.1}$$

$$= e^{-r\tau}\alpha - e^{-r\tau}\mathbb{Q}_t(S_{t+\tau} \leqslant S_t(r_\alpha^{\mathbb{P}} + 1)). \tag{4.3.2}$$

The situation is illustrated in Figure 4.2. For $\alpha = 20\%$, the quantile risk premium $QRP_{t+\tau|t}(\alpha)$ is plotted for hypothetical CDFs and corresponds to vertical differences between the $\mathbb{P}$- and $\mathbb{Q}$-measure CDFs at the $\mathbb{P}$-measure α-quantile.

The time series of various quantile risk premia are plotted in Figure 4.3. The time series show that the $\mathbb{Q}$-measure distribution consistently lies above the corresponding $\mathbb{P}$-measure distribution at the $\alpha = 1\%$ $\mathbb{P}$-measure quantile, providing evidence for a tail risk premium. A similar sign on the tail risk premium is reserved at the 5% and 10% quantiles. On the other hand, the risk premium at the $\alpha = 90\%$, 95%, and 99% quantiles appears to have the opposite sign. Taken together, these results suggest an asymmetric compensation for right- and left-tail risk that contrast with the symmetric compensation for variance risk discussed in Chapter 2.

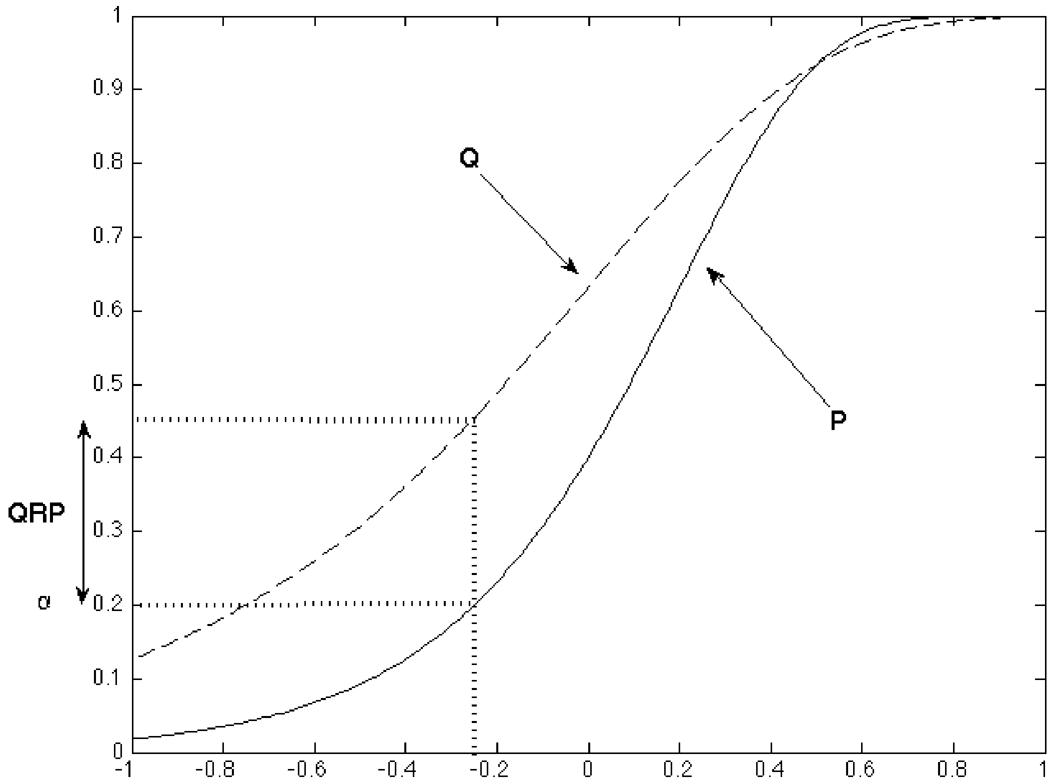

FIGURE 4.2: Illustration of an $\alpha = 20\%$ Quantile Risk Premium. The $\mathbb{Q}$-measure CDF (dash) and $\mathbb{P}$-measure CDF (solid) are plotted alongside an $\alpha = 20\%$ quantile risk premium.

4.4 Future Directions: Return Forecastability

I obtain a simple decomposition of the equity risk premium in terms of the distributions associated with the two expectation operators. Specifically, write (4.1.2) as

$$\begin{aligned} ERP_{t,t+\tau} = E_t^{\mathbb{P}}[r_{t,t+\tau}] - r^f_{t,t+\tau} &= \int_{-\infty}^{\infty} r dF^{\mathbb{P}}_{t+\tau|t}(r) - \int_{-\infty}^{\infty} r dF^{\mathbb{Q}}_{t+\tau|t}(r) \\ &= \int_{-\infty}^{\infty} r \Big[dF^{\mathbb{P}}_{t+\tau|t}(r) - dF^{\mathbb{Q}}_{t+\tau|t}(r) \Big]. \end{aligned} \tag{4.4.1}$$

Equation (4.4.1) makes explicit that any risk premium on returns, as measured

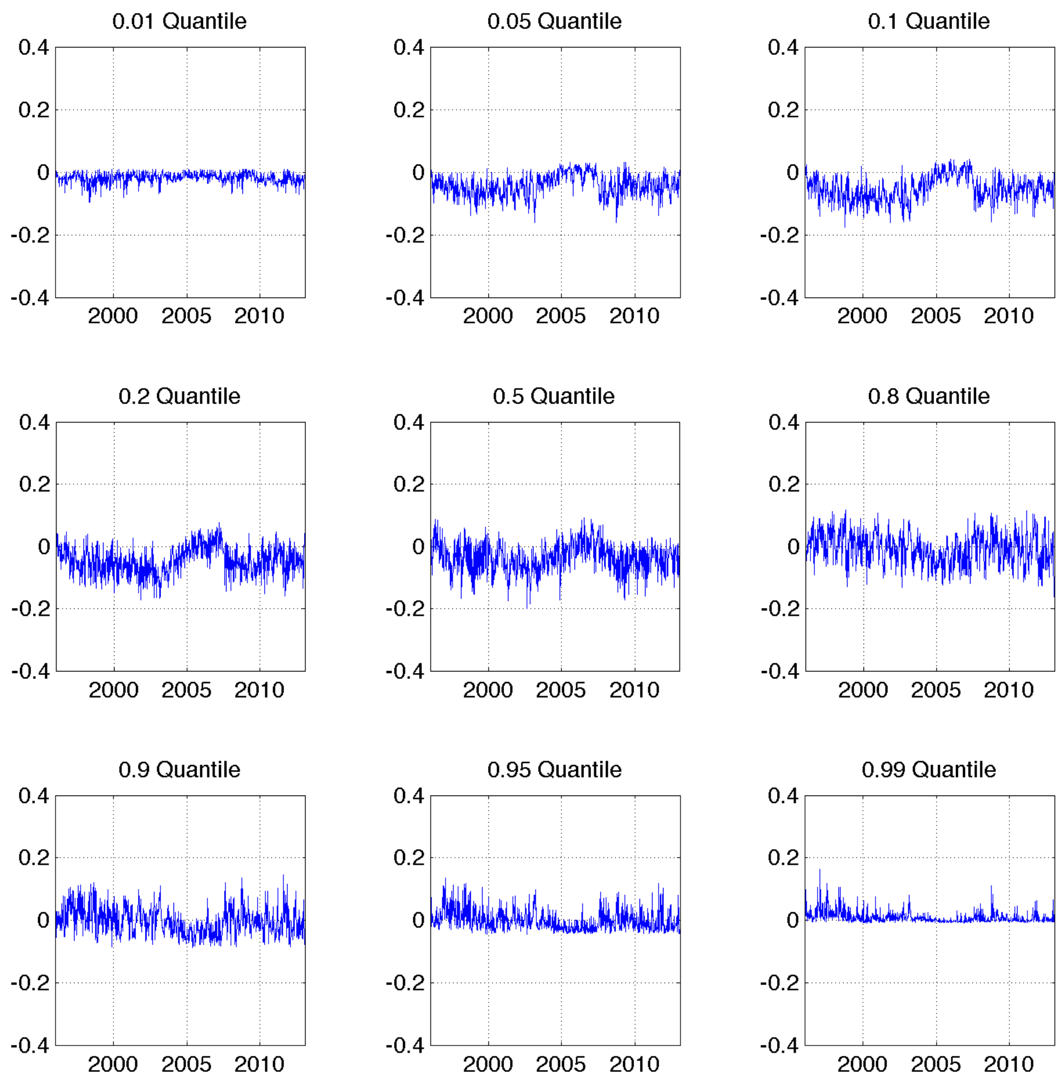

FIGURE 4.3: Time series of quantile risk premia. The quantile risk premium in (4.3.1) is plotted for the α-quantiles listed at the top of each panel.

by differences in objective and risk-neutral expectations, is driven by differences in increments of the distribution functions. To make this intuition empirically implementable, one can interpret the right-hand side of (4.4.1) as a Riemann-Stieltjes integral. That is, given a sequence of ordered partitions P_n of the return domain, the integral in (4.4.1) can be interpreted as a limit,

$$\int_{-\infty}^{\infty} r\Big[dF^{\mathbb{P}}_{t+\tau|t}(r) - dF^{\mathbb{Q}}_{t+\tau|t}(r)\Big]$$
$$= \lim_{n\to\infty} \sum_{r_i \in P_n} r_{i+1}\Big[F^{\mathbb{P}}_{t+\tau|t}(r_{i+1}) - F_{t+\tau|t}(r_i) - [F^{\mathbb{Q}}_{t+\tau|t}(r_{i+1}) - F^{\mathbb{Q}}_{t+\tau|t}(r_i)]\Big].$$

A natural example of a partition P_n is given by the return quantiles $r_i = F^{-1}_{t+\tau|t}(\theta_i)$, where $\theta_i \in [0,1]$. Using this partition, one has

$$\int_{-\infty}^{\infty} r\Big[dF^{\mathbb{P}}_{t+\tau|t}(r) - dF^{\mathbb{Q}}_{t+\tau|t}(r)\Big] = \lim_{n\to\infty} \sum_{r_i \in P_n} r_{i+1}\Big[[\theta_{i+1} - \theta_i] - F^{\mathbb{Q}}_{t+\tau|t}(r_{i+1}) - F^{\mathbb{Q}}_{t+\tau|t}(r_i)\Big]. \tag{4.4.2}$$

Equation (4.4.2) motivates a simple regressor framework for forecasting the equity risk-premium, obtained by regressing excess returns on differenced objective and risk-neutral quantiles of the return distribution.

4.5 Conclusion

This chapter outlined methods for making direct comparisons of $\mathbb{P}$- and $\mathbb{Q}$-measure return distributions by examining the time series of their conditional quantile estimates. By relying on the method of sieves, I extracted a balanced time series of $\mathbb{Q}$-measure return quantiles using the methods outlined in Chapter 2. Corresponding estimates of $\mathbb{P}$-measure quantiles were obtained using the CAViaR methods of Engle and Manganelli (2004). The findings showed evidence for the existence of pronounced risk premia in the extremes of the respective return distributions, but with asymmetric compensation for left- and right-tail risk. I further outlined a regression framework that motivated the use of quantile risk premia for forecasting returns by decomposing the equity risk premium into its return quantile components.

Appendix A

Definitions and Proofs for Chapter 2

A.1 Technical Results and Definitions

A.1.1 Sobolev Sieve Spaces

Establishing consistency and asymptotic normality of functionals requires a precise definition of the sieve approximation spaces. The final sieve spaces of interest are collections of *conditional* densities that we obtain by first defining a space of joint densities, and whose future payoff component can be integrated out to yield marginals. As mentioned above, the space of joint densities is the Gallant-Nychka class of densities first defined in Gallant and Nychka (1987). This class of densities is reviewed here.

Let $\mathbf{u} = (y, \mathbf{x}) \in \mathbb{R}^{d_u}$, where $d_u \equiv 1 + d_x$, and define the following notation for higher order derivatives,

$$D^{\lambda} f(\mathbf{u}) = \frac{\partial^{\lambda_1} \partial^{\lambda_2} \dots \partial^{\lambda_{d_u}}}{\partial u_1^{\lambda_1} \partial u_2^{\lambda_2} \dots \partial u_{d_u}^{\lambda_{d_u}}} f(\mathbf{u}),$$

with $\lambda = (\lambda_1, \dots, \lambda_{d_u})'$ consisting of nonnegative integer elements. The order of the derivative is $|\lambda| = \sum_{i=1}^{d_u} |\lambda_i|$, and $D^0 f = f$.

Definition A.1.1. (Sobolev norms). For $1 \leqslant p < \infty$, define the Sobolev norm of f with respect to the nonnegative weight function $\zeta(\mathbf{u})$ by

$$\|f\|_{m,p,\zeta} = \left(\sum_{|\lambda| \leqslant m} \int |D^{\lambda} f(\mathbf{u})|^p \zeta(\mathbf{u}) d\mathbf{u} \right)^{1/p}.$$

For $p = \infty$ and f with continuous partial derivatives to order m, define

$$\|f\|_{m,\infty,\zeta} = \max_{|\lambda| \leqslant m} \sup_{\mathbf{u} \in \mathbb{R}^{d_u}} |D^{\lambda} f(\mathbf{u})| \zeta(\mathbf{u}).$$

If $\zeta(\mathbf{u}) = 1$, simply write $\|f\|_{m,p}$ and $\|f\|_{m,\infty}$. Associated with each of these norms are the weighted Sobolev spaces

$$W^{m,p,\zeta}(\mathbb{R}^{d_u}) \equiv \{ f \in L^p(\mathbb{R}^{d_u}) : D^{\lambda} f \in L^p(\mathbb{R}^{d_u}) \},$$

where $1 \leqslant p \leqslant \infty$.

The following definitions are precisely the same as the collections $\mathcal{H}$ and $\mathcal{H}_K$ in Gallant and Nychka (1987).

Definition A.1.2. (The Gallant-Nychka Joint Density Space $\mathcal{F}^{Y,X}$). Let m denote the number of derivatives that characterize the degree of smoothness of the true joint SPD. Then for some integer $m_0 > d_u/2$, some bound $\mathcal{B}_0$, some small $\varepsilon_0 > 0$, some $\delta_0 > d_u/2$, and some probability density function $h_0(\mathbf{u})$ with zero mean and $\|h_0\|_{m_0+m,2,\zeta_0} \leqslant \mathcal{B}_0$, let $\mathcal{F}^{Y,X}$ consist of those probability density functions $f(\mathbf{u})$ with zero mean that have the form

$$f^{Y,X}(\mathbf{u}) = h(\mathbf{u})^2 + \varepsilon h_0(\mathbf{u})$$

with $\|h\|_{m_0+m,2,\zeta_0} \leqslant \mathcal{B}_0$ and $\varepsilon > \varepsilon_0$, where

$$\zeta_0(\mathbf{u}) = (1 + \mathbf{u}'\mathbf{u})^{\delta_0}.$$

Let

$$\mathcal{H} \equiv \{h \in W^{m_0+m,2,\zeta_0} : \|h\|_{m_0+m,2,\zeta_0} \leqslant \mathcal{B}_0\}.$$

The collection $\mathcal{F}^{Y,X}$ is the parent space of densities from which the conditional class of densities of interest are derived. Similarly, the sieve spaces that approximate the conditional parent space are obtained from joint density sieve spaces that approximate $\mathcal{F}^{Y,X}$.

Definition A.1.3. (The Gallant-Nychka Sieve Space $\mathcal{F}^{Y,X}_K$). Let $\phi(\mathbf{u}) = \exp(-\mathbf{u}'\mathbf{u}/2)$, and let $P_K(\mathbf{u})$ denote a Hermite polynomial of degree K. $\mathcal{F}^{Y,X}_K$ consists of those probability density functions with zero mean that are of the form

$$f^{Y,X}_K(\mathbf{u}) = [P_K(\mathbf{u} - \tau)]^2\phi(\mathbf{u} - \tau) + \varepsilon h_0(\mathbf{u})$$

with $\|P_K(\mathbf{u} - \tau)\phi(\mathbf{u} - \tau)^{1/2}\|_{m_0+m,2,\zeta_0} \leqslant \mathcal{B}_0$ and $\varepsilon > \varepsilon_0$.

The Conditional Density Spaces

The state-price density of interest, f_0, is a conditional density that resides in some parent function space of conditional densities. The associated sieve spaces are subspaces constructed to approximate this parent function space. The conditional density spaces of interest are obtained by simply dividing each member of $\mathcal{F}^{Y,X}$ by a marginal in $\mathbf{x}$, after having integrated out the first component in y.

Definition A.1.4. (The Sieve Spaces $\mathcal{F}$ and $\mathcal{F}_K$). Define

$$\mathcal{F}^{Y|X} \equiv \left\{ f \in W^{m,1}(\mathbb{R}^{d_u}) : f(y|\mathbf{x}) = \frac{f^{Y,X}(y,\mathbf{x})}{\int f^{Y,X}(y,\mathbf{x})dx} \text{ some } f^{Y,X} \in \mathcal{F}^{Y,X} \right\} \quad \text{and}$$

$$\mathcal{F}_K^{Y|X} \equiv \left\{ f_K \in W^{m,1}(\mathbb{R}^{d_u}) : f_K(y|\mathbf{x}) = \frac{f_K^{Y,X}(y,\mathbf{x})}{\int f_K^{Y,X}(y,\mathbf{x})dx} \text{ some } f_K^{Y,X} \in \mathcal{F}_K^{Y,X} \right\}.$$

This definition says that to each joint density in $\mathcal{F}^{Y,X}$, one can associate its corresponding conditional density. This association naturally gives rise to map $\Lambda : \mathcal{F}^{Y,X} \to \mathcal{F}$ with the following continuity property. Note that the densities in $\mathcal{F}$ are related to the return distribution via the change of variables formula in Eq. (2.2.3).

A.1.2 Intermediate Results

Lemma A.1.5. $P_X(f_1, \mathbf{Z}) = P_X(f_2, \mathbf{Z})$ *if and only if* $f_1 = f_2$ *almost everywhere.*

Proof. If $f_1 = f_2$ a.e., then by definition $P_X(f_1, \mathbf{Z}) = P_X(f_2, \mathbf{Z})$. Conversely, suppose $P_X(f_1, \mathbf{Z}) = P_X(f_2, \mathbf{Z})$. Then differentiating the option price with respect to strike twice yields

$$e^{r\tau} \frac{\partial^2 P_X(f_1, \mathbf{Z})}{\partial \kappa^2}\bigg|_\kappa = e^{r\tau} \frac{\partial^2 P_X(f_2, \mathbf{Z})}{\partial \kappa^2}\bigg|_\kappa \quad \Longrightarrow \quad f_1(\kappa|\mathbf{Z}) = f_2(\kappa|\mathbf{Z}).$$

Since this holds for every κ, the result follows. $\square$

Lemma A.1.6. *The map $\Lambda : \mathcal{F}^{Y,X} \to \mathcal{F}$ taking joint densities to their conditional counterparts in $\mathcal{F}$, i.e. $\Lambda(f^{Y,X}) = f$, is $\|\cdot\|_{m,\infty,\zeta} - \|\cdot\|_{m,1}$ Lipschitz continuous, where f is defined pointwise by*

$$f(y|\mathbf{x}) \equiv \Lambda(f^{Y,X}(y,\mathbf{x})) = \frac{f^{Y,X}(y,\mathbf{x})}{\int f^{Y,X}(y,\mathbf{x})dx}$$

and where $\zeta(\mathbf{u}) = (1 + \mathbf{u}'\mathbf{u})^{\delta}$ and $\delta \in (d_u/2, \delta_0)$.

Proof. Let $f_0(\mathbf{x}) = \int_{\mathbf{R}} f_0^{Y,X}(y,\mathbf{x})dx$ and $f_K(\mathbf{x}) = \int_{\mathbf{R}} f_K^{Y,X}(y,\mathbf{x})dx$ denote the marginal distributions of $\mathbf{X}$ of generic $f_0^{Y,X} \in \mathcal{F}^{Y,X}$ and $f_K^{Y,X} \in \mathcal{F}_K^{Y,X}$, and let $g_0(\mathbf{x}) = 1/f_0(\mathbf{x})$ and $g_K(\mathbf{x}) = 1/f_K(\mathbf{x})$ denote their reciprocals. In this notation, the conditional densities become $f_0(y|\mathbf{x}) = f_0^{Y,X}(y,\mathbf{x})g_0(\mathbf{x})$ and $f_K(y|\mathbf{x}) = f_K^{Y,X}(y,\mathbf{x})g_K(\mathbf{x})$. Let $\mathcal{X} = \mathbb{R}^{d_x}$ and $\mathcal{Y} = \mathbb{R}$.

The goal of the proof is to show that $\|f_K(y|\mathbf{x}) - f_0(y|\mathbf{x})\|_{m,1}$ is small whenever the corresponding joint distribution error $\|f_K^{Y,X}(y,\mathbf{x}) - f_0^{Y,X}(y,\mathbf{x})\|_{m_0+m,\infty,\zeta}$ is small.

Note first that by definition,

$$\begin{aligned}\|f_K(y|\mathbf{x}) - f_0(y|\mathbf{x})\|_{m,1} &= \sum_{|\lambda|\leqslant m} \int_{\mathcal{X}} \int_{\mathcal{Y}} |D^{\lambda} f_K(y|\mathbf{x}) - D^{\lambda} f_0(y|\mathbf{x})| dx d\mathbf{x} \\ &= \sum_{|\lambda|\leqslant m} \|D^{\lambda} f_K(y|\mathbf{x}) - D^{\lambda} f_0(y|\mathbf{x})\|_{0,1}, \qquad \text{(A.1.1)}\end{aligned}$$

so we can focus on the $\|D^{\lambda} f_K(y|\mathbf{x}) - D^{\lambda} f_0(y|\mathbf{x})\|_{0,1}$ terms on the RHS.

$$
\begin{aligned}
\|D^\lambda f_K(y|\mathbf{x}) - D^\lambda f_0(y|\mathbf{x})\|_{0,1} &= \int_{\mathcal{X}}\int_{\mathcal{Y}} \Big| D^\lambda\{f_K(y|\mathbf{x})\} - D^\lambda\{f_0(y|\mathbf{x})\}\Big| dx d\mathbf{x} \\
&= \int_{\mathcal{X}}\int_{\mathcal{Y}} \Big| D^\lambda\{f_K(y,\mathbf{x})g_K(\mathbf{x})\} - D^\lambda\{f_0(y,\mathbf{x})g_0(\mathbf{x})\}\Big| dx d\mathbf{x} \\
&= \int_{\mathcal{X}}\int_{\mathcal{Y}} \Big| D^\lambda\{[f_K(y,\mathbf{x}) - f_0(y,\mathbf{x})]g_K(\mathbf{x})\} \\
&+ D^\lambda\{[g_K(\mathbf{x}) - g_0(\mathbf{x})]f_0(y,\mathbf{x})\}\Big| dx d\mathbf{x} \\
&\leqslant \int_{\mathcal{X}}\int_{\mathcal{Y}} \Big| D^\lambda\{[f_K(y,\mathbf{x}) - f_0(y,\mathbf{x})]g_K(\mathbf{x})\}\Big| dx d\mathbf{x} \\
&+ \int_{\mathcal{X}}\int_{\mathcal{Y}} \Big| D^\lambda\{[g_K(\mathbf{x}) - g_0(\mathbf{x})]f_0(y,\mathbf{x})\}\Big| dx d\mathbf{x} \\
&= \|D^\lambda\{[f_K(y,\mathbf{x}) - f_0(y,\mathbf{x})]g_K(\mathbf{x})\}\|_{0,1} \\
&+ \|D^\lambda\{[g_K(\mathbf{x}) - g_0(\mathbf{x})]f_0(y,\mathbf{x})\}\|_{0,1}. \qquad \text{(A.1.2)}
\end{aligned}
$$

To bound the two terms on the RHS, we first need to establish bounds on the marginals. Observe that since the marginal densities have one fewer component than the joint densities, for $|\lambda| \leqslant m$, define the multi-index $\alpha = (0, \lambda_2, \ldots, \lambda_{d_u})$. Then

$$
\begin{aligned}
\|D^\alpha f_K(\mathbf{x}) - D^\alpha f_0(\mathbf{x})\|_{0,1} &= \int_{\mathcal{X}} |D^\alpha f_K(\mathbf{x}) - D^\alpha f_0(\mathbf{x})| d\mathbf{x} \\
&= \int_{\mathcal{X}} |D^\alpha \int_{\mathcal{Y}} f_K^{Y,X}(y,\mathbf{x}) dx - D^\alpha \int_{\mathcal{Y}} f_0^{Y,X}(y,\mathbf{x}) dx | d\mathbf{x} \\
&= \int_{\mathcal{X}} |\int_{\mathcal{Y}} D^\alpha f_K^{Y,X}(y,\mathbf{x}) dx - \int_{\mathcal{Y}} D^\alpha f_0^{Y,X}(y,\mathbf{x}) dx | d\mathbf{x} \quad \text{(A.1.3)} \\
&\leqslant \int_{\mathcal{X}}\int_{\mathcal{Y}} |D^\alpha f_K^{Y,X}(y,\mathbf{x}) dx - D^\alpha f_0^{Y,X}(y,\mathbf{x})| dx d\mathbf{x} \\
&\leqslant \|f_K^{Y,X} - f_0^{Y,X}\|_{m,\infty,\zeta} \int_{\mathbb{R}^{d_u}} (1 + \mathbf{u}'\mathbf{u})^{-\delta} d\mathbf{u},
\end{aligned}
$$

where the last two inequalities are the triangle and Hölder's inequality, respectively. The interchange between the integration and differentiation operator on the third line is due to the dominated convergence theorem with dominating function derived from the norm bound on functions in $\mathcal{F}^{Y,X}$ as follows. By assumption,

$$f^{Y,X} = h^2 + \varepsilon_0 h_0,$$

where $\|h\|_{m_0+m,2,\zeta_0} < \mathcal{B}_0$. Thus,

$$\begin{aligned} |\zeta_0(\mathbf{u})^{1/2} h(\mathbf{u})| &\leqslant \max_{|\lambda| \leqslant m} \sup_{\mathbf{u} \in \mathbb{R}^{d_u}} |D^\lambda \zeta_0(\mathbf{u})^{1/2} h(\mathbf{u})| \\ &= \|\zeta_0^{1/2} h\|_{m,\infty} \\ &\leqslant M_2 \|h\|_{m_0+m,2,\zeta_0} \quad \text{by Gallant and Nychka (1987) Lemma A.1(b)} \\ &< M_2 \mathcal{B}_0. \end{aligned}$$

Therefore,

$$\begin{aligned} \zeta_0(\mathbf{u}) h(\mathbf{u})^2 &\leqslant (M_2 \mathcal{B}_0)^2 \\ h(\mathbf{u})^2 &\leqslant (M_2 \mathcal{B}_0)^2 (1 + \mathbf{u}'\mathbf{u})^{-\delta_0} \leqslant (M_2 \mathcal{B}_0)^2 (1 + \mathbf{u}'\mathbf{u})^{-\delta}. \end{aligned}$$

Similar reasoning establishes a bound on h_0, so we have

$$f^{Y,X} \leqslant const.(1 + \mathbf{u}'\mathbf{u})^{-\delta},$$

where the RHS is integrable. By dominated convergence, this establishes the validity of interchanging the differentiation and integration operator in Eq. (A.1.3).

Therefore,

$$\begin{aligned}
\|f_K(\mathbf{x}) - f_0(\mathbf{x})\|_{m,1} &= \sum_{|\alpha|\leqslant m} \int_{\mathcal{X}} |D^\alpha\{f_K(\mathbf{x})\} - D^\alpha\{f_0(\mathbf{x})\}| d\mathbf{x} \\
&= \sum_{|\alpha|\leqslant m} \|D^\alpha f_K(\mathbf{x}) - D^\alpha f_0(\mathbf{x})\|_{0,1} \\
&\leqslant const. \|f_K^{Y,X} - f_0^{Y,X}\|_{m,\infty,\zeta} \int_{\mathbb{R}^{d_u}} (1 + \mathbf{u}'\mathbf{u})^{-\delta} d\mathbf{u},
\end{aligned}$$

which implies that if $\|f_K^{Y,X} - f_0^{Y,X}\|_{m,\infty,\zeta} \to 0$, then $\|f_K(\mathbf{x}) - f_0(\mathbf{x})\|_{m,1} \to 0$. Next, observe that this type of convergence holds for the reciprocal marginals $g = 1/f$ too, due to the continuity of the operator $f \mapsto 1/f$ (since f has a lower density bound of order $\varepsilon_0 h_0$). Thus, $\|g_K(\mathbf{x}) - g_0(\mathbf{x})\|_{m,1} \to 0$ as well.

We are now ready to examine the two terms on the RHS of Eq. (A.1.2). Apply Leibniz' formula (see Adams and Fournier (2003)) to get

$$D^\lambda\{[g_K(\mathbf{x}) - g_0(\mathbf{x})]f_0(y, \mathbf{x})\} = \sum_{\beta\leqslant\lambda} \begin{bmatrix} \lambda \\ \beta \end{bmatrix} D^{\lambda-\beta}\{g_K(\mathbf{x}) - g_0(\mathbf{x})\} D^\beta f_0(y, \mathbf{x}),$$

so that by the triangle inequality, Hölder's inequality, and the definitions of Sobolev norms,

$$\|D^{\lambda}\{[g_K(\mathbf{x}) - g_0(\mathbf{x})]f_0(y,\mathbf{x})\}\|_{0,1} \leqslant$$

$$\sum_{\beta \leqslant \lambda} \begin{bmatrix} \lambda \\ \beta \end{bmatrix} \|D^{\lambda-\beta}\{g_K(\mathbf{x}) - g_0(\mathbf{x})\}D^{\beta}f_0(y,\mathbf{x})\|_{0,1}$$

$$\leqslant \sum_{\beta \leqslant \lambda} \begin{bmatrix} \lambda \\ \beta \end{bmatrix} \|D^{\lambda-\beta}\{g_K(\mathbf{x}) - g_0(\mathbf{x})\}\|_{0,1}\|D^{\beta}f_0(y,\mathbf{x})\|_{0,\infty}$$

$$\leqslant const.\|g_K(\mathbf{x}) - g_0(\mathbf{x})\|_{m,1}\|f_0(y,\mathbf{x})\|_{m,\infty}$$

$$\leqslant const.\|g_K(\mathbf{x}) - g_0(\mathbf{x})\|_{m,1}. \tag{A.1.4}$$

Similarly,

$$\|D^{\lambda}\{[f_K(y,\mathbf{x}) - f_0(y,\mathbf{x})]g_K(\mathbf{x})\}\|_{0,1} \leqslant$$

$$\sum_{\beta \leqslant \lambda} \begin{bmatrix} \lambda \\ \beta \end{bmatrix} \|D^{\lambda-\beta}\{f_K(y,\mathbf{x}) - f_0(y,\mathbf{x})\}D^{\beta}g_K(\mathbf{x})\|_{0,1}$$

$$\leqslant \sum_{\beta \leqslant \lambda} \begin{bmatrix} \lambda \\ \beta \end{bmatrix} \|D^{\lambda-\beta}\{f_K(y,\mathbf{x}) - f_0(y,\mathbf{x})\}\|_{0,\infty}\|D^{\beta}g_K(\mathbf{x})\|_{0,1}$$

$$\leqslant const.\|f_K(y,\mathbf{x}) - f_0(y,\mathbf{x})\|_{m,\infty}\|g_K(\mathbf{x})\|_{m,1}$$

$$\leqslant const.\|f_K(y,\mathbf{x}) - f_0(y,\mathbf{x})\|_{m,\infty,\zeta_0}\|g_K(\mathbf{x})\|_{m,1}$$

$$\leqslant const.\|f_K(y,\mathbf{x}) - f_0(y,\mathbf{x})\|_{m,\infty,\zeta_0}\|g_K(\mathbf{x}) - g_0(\mathbf{x}) + g_0(\mathbf{x})\|_{m,1}$$

$$\leqslant const.\|f_K(y,\mathbf{x}) - f_0(y,\mathbf{x})\|_{m,\infty,\zeta_0}\|g_K(\mathbf{x}) - g_0(\mathbf{x})\|_{m,1}$$

$$+ const.\|f_K(y,\mathbf{x}) - f_0(y,\mathbf{x})\|_{m,\infty,\zeta_0}\|g_0(\mathbf{x})\|_{m,1}. \tag{A.1.5}$$

Plugging Eqs. (A.1.4) and (A.1.5) into the RHS of Eq. (A.1.2), we see that $\|D^{\lambda}f_K(y|\mathbf{x}) - D^{\lambda}f_0(y|\mathbf{x})\|_{0,1} \to 0$ whenever $\|f_K(y,\mathbf{x}) - f_0(y,\mathbf{x})\|_{m,\infty,\zeta_0} \to 0$. By Eq.

(A.1.1), this means that we have $\|f_K(y|\mathbf{x}) - f_0(y|\mathbf{x})\|_{m,1} \to 0$ whenever $\|D^\lambda f_K(y|\mathbf{x}) - D^\lambda f_0(y|\mathbf{x})\|_{0,1} \to 0$, which is the statement of the lemma. □

This result formalizes an intuitive notion: when two joint densities in $\mathcal{F}^{Y,X}$ are close, then so are the conditional densities in $\mathcal{F}$. The Lemma provides the Sobolev norms for which this intuition is correct. Furthermore, this property will be used below to regulate the complexity of the space of option pricing functions that are obtained by integrating the option payoff against a candidate from $\mathcal{F}$. Note that the map in Lemma A.1.6 is also surjective by definition.

The final sieve spaces on which the asymptotic theory is built are of the following form.

Definition A.1.7. (Sieve Spaces). The sieve spaces of interest are denoted $\mathcal{F} \equiv \overline{\Lambda}[cl(\mathcal{F}^{Y,X})]$ and $\mathcal{F}_K^{Y|X} \equiv \overline{\Lambda}[cl(\mathcal{F}_K^{Y,X})]$, where $cl(\cdot)$ denotes the closure.

The following is a consequence of Lemma A.1.6.

Corollary A.1.8. *There exists a continuous extension of* Λ *to a mapping* $\overline{\Lambda}$: $cl(\mathcal{F}^{Y,Y}) \to cl(\mathcal{F}^{Y|Y})$*, where* $cl(\cdot)$ *denotes the closure.*

Proof. Note that $cl(\mathcal{F}^{Y|X})$ is a closed subset of a (complete) Sobolev space and is therefore complete (p. 194 Royden and Fitzpatrick (2010)). In addition, Lemma A.1.6 shows that $\Lambda : \mathcal{F}^{Y,X} \to \mathcal{F}^{Y|X}$ is Lipschitz continuous and is therefore uniformly continuous. Therefore, this map has a unique uniformly continuous extension $\overline{\Lambda}$ from $\mathcal{F}^{Y,X}$ to $cl(\mathcal{F}^{Y,X})$ (p. 196 Royden and Fitzpatrick (2010)). This extension sends $cl(\mathcal{F}^{Y|X})$ into $cl(\mathcal{F}^{Y|X})$. □

To establish the asymptotic properties of the sieve estimator, the following two conditions are required.

Lemma A.1.9. *The sieve spaces $\mathcal{F}_K$ satisfy the following conditions:*

(i) $\mathcal{F}_K$ is compact in the topology generated by $\|\cdot\|_{m,1}$ for all $K \geqslant 0$.

(ii) $\cup_{K=0}^{\infty}\mathcal{F}_K$ is dense in $\mathcal{F}$ with the topology generated by $\|\cdot\|_{m,1}$.

Proof. Continuity of the $\overline{\Lambda}$ map above means that the spaces $\mathcal{F}_K$ inherit the topological properties of their preimages under $\overline{\Lambda}$. Since Theorem 1 of Gallant and Nychka (1987) says that $cl(\mathcal{F}^{Y,X})$ is compact in $\|\cdot\|_{m,\infty,\zeta}$, we have that $cl(\mathcal{F}_K^{Y,X})$ are compact as well. By continuity of $\overline{\Lambda}$, this means that $\mathcal{F}_K$ is compact in $\|\cdot\|_{m,1}$, which shows (i). Similarly, Theorem 2 of Gallant and Nychka (1987) shows that $\cup_{K=0}^{\infty}\mathcal{F}_K^{Y,X}$ is a dense subset of $cl(\mathcal{F}^{Y,X})$, so $\cup_{K=0}^{\infty}cl(\mathcal{F}_K^{Y,X})$ is as well. Next, note that the definition of $\mathcal{F}$ says that $\overline{\Lambda}$ is surjective. Because the image of a dense set is again dense under a continuous surjective map, we have that $\overline{\Lambda}(\cup_{K=0}^{\infty}cl(\mathcal{F}_K^{Y,X})) = \cup_{K=0}^{\infty}\Lambda(cl(F_K^{Y,X})) = \cup_{K=0}^{\infty}\mathcal{F}_K$ is dense in $\mathcal{F}$ under $\|\cdot\|_{m,1}$, showing (ii). □

Finally, the densities are related to option prices via the following result.

Lemma A.1.10. *Under Assumption 2.4.1, the option pricing functional $P_Y(f, \mathbf{Z})$ is*

(i) almost surely locally $|\cdot|$, $\|\cdot\|_{m,1}$-Lipschitz continuous in f.

(ii) locally $\|\cdot\|_2$, $\|\cdot\|_{m,1}$-Lipschitz continuous in f.

Proof. (i) Let $\varepsilon > 0$ be given, and fix an $f_0 \in \mathcal{F}$. Under Assumption 2.4.1, there exists an $\|\cdot\|_{m,1}-$open ball $B_\delta(f_0)$ of radius δ such that

$$\sup_{f \in B_\delta(f_0)} |S(f, \mathbf{Z})| \leqslant M \qquad \mathbb{P} - a.s.$$

Then choose $\eta = \min\{\varepsilon, \delta\}/(2M)$, and consider any $f \in B_\eta(f_0)$. Using put-call parity,

$$\begin{aligned}
|P_Y(f,\mathbf{Z}) - P_Y(f_0,\mathbf{Z})| &= |C_Y(f,\mathbf{Z}) - S_0 e^{-q\tau} + \kappa e^{-r\tau} - C_Y(f_0,\mathbf{Z}) + S_0 e^{-q\tau} - \kappa e^{-r\tau}| \\
&= |C_Y(f,\mathbf{Z}) - C_Y(f_0,\mathbf{Z})| \\
&\leqslant \sup_{g\in(f,f_0)} |\frac{\partial C_Y(g,\mathbf{Z})}{\partial f}| \|f - f_0\|_{m,1} \\
&= \sup_{g\in(f,f_0)} |C_Y(g,\mathbf{Z})| \|f - f_0\|_{m,1} \\
&\leqslant \sup_{g\in(f,f_0)} |S_T(g,\mathbf{Z})| \|f - f_0\|_{m,1} \\
&\leqslant M\|f - f_0\|_{m,1} \quad a.s., \\
&\leqslant \varepsilon/2
\end{aligned}$$

so that $|P_Y(f,\mathbf{Z}) - P_Y(f_0,\mathbf{Z})| < \varepsilon$. The third line in the preceding display is due to the functional mean value theorem, the fourth due to the linearity of $C_Y(f,\mathbf{Z})$ in f, the fifth is a consequence of no arbitrage bounds on option prices, and the final inequality is due to our assumption.

(ii) follows from (i), after observing that

$$\sup_{f\in B_\delta(f_0)} \|S_T(f,\mathbf{Z})\|_2^2 = \sup_{f\in B_\delta(f_0)} \mathbb{E}[S_T(f,\mathbf{Z})^2 W(\mathbf{Z})] \leqslant \mathbb{E}[\sup_{f\in B_\delta(f_0)} S_T(f,\mathbf{Z})^2 W(\mathbf{Z})] \leqslant const.$$

by combining Assumption 2.4.1 and Assumption 2.4.2 (i). Choose η similar to (i) above, but depending on *const*. Then perform the derivations under (i) above, replacing $|\cdot|$ with $\|\cdot\|_2$ to obtain the desired result. □

A.2 Appendix: Proofs

Proof of Lemma 2.1 Let $\alpha(B,\tau) = (\sum_{j=0}^{K_\tau} \beta_{0j} H_j(\tau),\ \ldots,\ \sum_{j=0}^{K_\tau} \beta_{K_y j} H_j(\tau))'$. Then

$$\begin{aligned}
\int f_K^{X,Z}(y,\tau)dy &= \int \left[\sum_{k=0}^{K_y} \alpha_k(B,\tau)H_k(y)\right]^2 \phi(\tau)\phi(y)dy \\
&= \phi(\tau)\int \sum_{k=0}^{K_y} \alpha_k(B,\tau)^2 H_k(y)^2\phi(y)dy = \phi(\tau)\sum_{k=0}^{K_y}\alpha_k(B,\tau)^2 \\
&= \alpha(B,\tau)'\alpha(B,\tau)\phi(\tau),
\end{aligned}$$

where the second and third equality follow from the orthonormality of the Hermite polynomials. Then,

$$\begin{aligned}
f_K(y|\tau) = \frac{f_K^{X,Z}(y,\tau)}{\int f_K^{X,Z}(y,\tau)dy} &= \frac{\left[\sum_{k=0}^{K_y} \alpha_k(B,\tau)H_k(y)\right]^2 \phi(\tau)\phi(y)}{\alpha(B,\tau)'\alpha(B,\tau)\phi(\tau)} \\
&= \frac{\sum_{k=0}^{2K_y} \alpha(B,\tau)'A_k\alpha(B,\tau)H_k(y)\phi(y)}{\alpha(B,\tau)'\alpha(B,\tau)}
\end{aligned}$$

where the last equality and the definition of A_k follow by applying Proposition 1 of Leon, Mencia, and Sentana (2009). The result follows.

Proof of Proposition 1 I follow the derivation of León and Sentana (2009), which differs due to the conditioning on τ. The plug-in estimator of the population option price in equation (2.2.4), is given by

$$P_Y(f_K, \mathbf{Z}) = e^{-r\tau} \int_{-\infty}^{d(\mathbf{Z})} \left(\kappa - Se^{\mu(\mathbf{Z})+\sigma(\mathbf{Z})Y}\right) f_K(Y|\tau)dY$$

$$= \kappa e^{-r\tau} \int_{-\infty}^{d(\mathbf{Z})} f_K(Y|\tau)dY - Se^{-r\tau+\mu(\mathbf{Z})} \int_{-\infty}^{d(\mathbf{Z})} e^{\sigma(\mathbf{Z})Y} f_K(Y|\tau)dY. \quad \text{(A.2.1)}$$

The integral in the first term becomes

$$\int_{-\infty}^{d(\mathbf{Z})} f_K(Y|\tau)dY = \int_{-\infty}^{d(\mathbf{Z})} \left[\sum_{k=0}^{2K_y} \gamma_k(B,\tau) H_k(Y)\phi(Y)\right] dY$$

$$= \sum_{k=0}^{2K_y} \gamma_k(B,\tau) \int_{-\infty}^{d(\mathbf{Z})} H_k(Y)\phi(Y)dY = \Phi(d(\mathbf{Z})) - \sum_{k=1}^{2K_y} \frac{\gamma_k(B,\tau)}{\sqrt{k}} H_{k-1}(d(\mathbf{Z}))\phi(d(\mathbf{Z})), \quad \text{(A.2.2)}$$

where the last equality follows from integration properties of the Hermite functions. The integral in the second term on the right-hand side (RHS) of equation (A.2.1) can further be simplified by integrating by parts. Let

$$I_k^*(d(\mathbf{Z})) = \int_{-\infty}^{d(\mathbf{Z})} e^{\sigma(\mathbf{Z})Y} H_k(Y)\phi(Y)dY.$$

For $k = 0$,

$$I_0^*(d(\mathbf{Z})) = \int_{-\infty}^{d(\mathbf{Z})} e^{\sigma(\mathbf{Z})Y}\phi(Y)dY = e^{\sigma(\mathbf{Z})^2/2} \int_{-\infty}^{d(\mathbf{Z})-\sigma(\mathbf{Z})} \phi(u)du = e^{\sigma(\mathbf{Z})^2/2}\Phi(d(\mathbf{z}) - \sigma(\mathbf{z}))$$

by a change of variables. For $k \geqslant 1$,

$$
\begin{aligned}
I_k^*(d(\mathbf{Z})) &= \int_{-\infty}^{d(\mathbf{Z})} e^{\sigma(\mathbf{Z})Y} H_k(Y)\phi(Y)dY \\
&= \left[-\frac{1}{\sqrt{k}} e^{\sigma(\mathbf{Z})Y} H_{k-1}(Y)\phi(Y)\right]_{-\infty}^{d(\mathbf{Z})} + \frac{\sigma(\mathbf{Z})}{\sqrt{k}} \int_{-\infty}^{d(\mathbf{Z})} e^{\sigma(\mathbf{Z})Y} H_{k-1}(Y)\phi(Y)dY \\
&= -\frac{1}{\sqrt{k}} e^{\sigma(\mathbf{Z})d(\mathbf{Z})} H_{k-1}(d(\mathbf{Z}))\phi(d(\mathbf{Z})) + \frac{\sigma(\mathbf{Z})}{\sqrt{k}} I_{k-1}^*(d(\mathbf{Z})).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\int_{-\infty}^{d(\mathbf{Z})} e^{\sigma(\mathbf{Z})Y} f_K(Y|\tau)dY = \int_{-\infty}^{d(\mathbf{Z})} e^{\sigma(\mathbf{Z})Y} \left[\sum_{k=0}^{2K_y} \gamma_k(B,\tau) H_k(Y)\phi(Y)\right] dY \\
&= \sum_{k=0}^{2K_y} \gamma_k(B,\tau) \int_{-\infty}^{d(\mathbf{Z})} e^{\sigma(\mathbf{Z})Y} H_k(Y)\phi(Y)dY = \sum_{k=0}^{2K_y} \gamma_k(B,\tau) I_k^*(d(\mathbf{Z})) \\
&= \gamma_0(B,\mathbf{Z}) e^{\sigma(\mathbf{Z})^2/2} \Phi(d(\mathbf{z}) - \sigma(\mathbf{z})) + \sum_{k=1}^{2K_y} \gamma_k(B,\tau) I_k^*(d(\mathbf{Z})) \\
&= 1 \cdot e^{\sigma(\mathbf{Z})^2/2} \Phi(d(\mathbf{z}) - \sigma(\mathbf{z})) + \sum_{k=1}^{2K_y} \gamma_k(B,\tau) I_k^*(d(\mathbf{Z})) \qquad \text{(A.2.3)}
\end{aligned}
$$

Plugging equations (A.2.2) and (A.2.3) into (A.2.1) obtains the desired result. The proof for call options is analogous and is therefore omitted.

Proof of Proposition 2 Let $L(f) = \mathbb{E}\{-\frac{1}{2}[P - P_Y(f,\mathbf{Z})]^2 W\} \equiv \mathbb{E}\{\ell(f,Y)\}$, where $Y \equiv (P,\mathbf{Z})$, and $W = W(\mathbf{Z})$ is a strictly positive weighting function. ℓ is concave in f, and L is strictly concave in f. The goal is to estimate the unknown

$P_Y^0(\mathbf{Z}) = \mathbb{E}[P|\mathbf{Z}]$ by invoking the general sieve consistency theorem in Chen (2007) (i.e. her Theorem 3.1). This requires verification of her Conditions 3.1' - 3.3', 3.4, and 3.5(i), which adapts to the present notation as follows:

Condition 3.1'.

(i) $L(f)$ is continuous at $f_0 \in \mathcal{F}$, $L(f_0) > -\infty$.

(ii) for all $\varepsilon > 0$, $L(f_0) > \sup_{\{f \in \mathcal{F}: d(f, f_0) \geqslant \varepsilon\}} L(f)$

Condition 3.2'.

(i) $\mathcal{F}_K \subseteq \mathcal{F}_{K+1} \subseteq \cdots \subseteq \mathcal{F}$, for all $K \geqslant 1$.

(ii) For any $f \in \mathcal{F}$, there exists $\pi_K f \in \mathcal{F}_K$ such that $d(f, \pi_K f) \to 0$ as $K \to \infty$.

Condition 3.3'.

(i) $L_n(f)$ is a measurable function of the data $\{Y_i\}_{i=1}^n$ for all $f \in \mathcal{F}_K$

(ii) For any data $\{Y_i\}_{i=1}^n$, $L_n(f)$ is upper semicontinuous on $\mathcal{F}_K$ under $d(\cdot, \cdot)$.

Condition 3.4. The sieve spaces $\mathcal{F}_K$ are compact under $d(\cdot, \cdot)$.

Condition 3.5.

(i) For all $K \geqslant 1$, $\sup_{f \in \mathcal{F}_K} |L_n(f) - L(f)| = 0$.

I verify each of these conditions in turn.

Condition 3.1': Assumption 2.4.2 (ii) implies $L(f_0) = 0 > -\infty$. Also,

$$
\begin{aligned}
&L(f_0) - L(f) \\
&= -\mathbb{E}\{\frac{1}{2}[P - P_Y(f_0, \mathbf{Z})]^2 W(\mathbf{Z})\} + \mathbb{E}\{\frac{1}{2}[P - P_Y(f, \mathbf{Z})]^2 W(\mathbf{Z})\} \\
&= \frac{1}{2}\mathbb{E}\{[P^2 - 2PP_Y(f, \mathbf{Z}) + P_Y(f, \mathbf{Z})^2 - P^2 + 2PP_Y(f_0, \mathbf{Z}) - P_Y(f_0, \mathbf{Z})]W(\mathbf{Z})\} \\
&= \frac{1}{2}\mathbb{E}\{[P_Y(f, \mathbf{Z}) - P_Y(f_0, \mathbf{Z})][-2P + P_Y(f, \mathbf{Z}) + P_Y(f_0, \mathbf{Z})]W(\mathbf{Z})\} \\
&= -\mathbb{E}\{[P_Y(f, \mathbf{Z}) - P_Y(f_0, \mathbf{Z})][(P - P_Y(f_0, \mathbf{Z})) - \frac{1}{2}(P_Y(f, \mathbf{Z}) - P_Y(f_0, \mathbf{Z}))]W(\mathbf{Z})\} \\
&= \frac{1}{2}\mathbb{E}\{[P_Y(f, \mathbf{Z}) - P_Y(f_0, \mathbf{Z})]^2 W(\mathbf{Z})\} \\
&= \frac{1}{2}\|P_Y(f, \mathbf{Z}) - P_Y(f_0, \mathbf{Z})\|_2^2.
\end{aligned}
$$

As $d(f_n, f_0) \to 0$, the local Lipschitz continuity condition derived in Lemma A.1.10 implies that the RHS tends to zero, i.e. $L(f_0) - L(f) = |L(f_0) - L(f)| \to 0$. This establishes Condition 3.1'(i). As for Condition 3.1'(ii), note that continuity of $L(f)$ at f_0 implies that for any $\eta > 0$, there exists a $\varepsilon > 0$ such that for all f satisfying $d(f, f_0) < \varepsilon$, we have $\|P_Y(f, \mathbf{Z}) - P_Y(f_0, \mathbf{Z})\|_2 < \eta$. The contrapositive of this statement reads: Given any $\varepsilon > 0$, there exists $\eta > 0$ such that if $d(f, f_0) \geqslant \varepsilon$, then $\|P_Y(f, \mathbf{Z}) - P_Y(f_0, \mathbf{Z})\|_2 \geqslant \eta$. Now let $\varepsilon > 0$ be given as in Condition 3.1'(ii), and consider any $f \in \{f \in \mathcal{F} : d(f, f_0) \geqslant \varepsilon\}$. By the previous derivations,

$$L(f_0) - L(f) = \frac{1}{2}\|P_Y(f, \mathbf{Z}) - P_Y(f_0, \mathbf{Z})\|_2^2 \geqslant \frac{1}{2}\eta^2,$$

so

$$L(f_0) - \sup_{\{f \in \mathcal{F}: d(f, f_0) \geqslant \varepsilon\}} L(f) = \inf_{\{f \in \mathcal{F}: d(f, f_0) \geqslant \varepsilon\}} [L(f_0) - L(f)] \geqslant \frac{1}{4}\eta^2 > 0,$$

which establishes Condition 3.1'(ii).

Condition 3.2': Condition 3.2'(i) follows readily from the orthogonality of Hermite polynomials. Condition 3.2'(ii) is shown in Lemma A.1.9 (ii).

Condition 3.3': First note that Chen's Theorem 3.1 still goes through if we only require $L_n(f)$'s upper semi-continuity to hold almost surely. To this end, observe that Assumption 2.4.2 (i) implies that P_i is almost surely finite, i.e. $\exists$ a Borel set Ω_F with $P_i(\omega) < \infty$ for all $\omega \in \Omega_F$,[1] and Assumption 2.4.1 with no arbitrage imposed implies $P_Y(f, \mathbf{Z}_i)$ is locally bounded $\mathbb{P}-a.s.$ on $\mathcal{F}$. Therefore $P_i - P_Y(f, \mathbf{Z}_i)$ is finite on Ω_F.

Next, fix $\omega \in \Omega_F$. Given any sequence $f_j \in \mathcal{F}_K$ with $\|f_j - f\|_{m,1} \to 0$,

$$
\begin{aligned}
|L_n(f_j) - L_n(f)| \leqslant \frac{1}{n}\sum_{i=1}^{n}\Big|&[P_Y(f_j, \mathbf{Z}_i(\omega)) - P_Y(f, \mathbf{Z}_i(\omega))]\\
&[(P_i(\omega) - P_Y(f, \mathbf{Z}_i(\omega))) - \frac{1}{2}(P_Y(f_j, \mathbf{Z}_i(\omega)) - P_Y(f, \mathbf{Z}_i(\omega)))]W(\mathbf{Z}_i(\omega))\Big|\\
&\leqslant const.\frac{1}{n}\sum_{i=1}^{n}\{|[P_Y(f_j, \mathbf{Z}_i(\omega)) - P_Y(f, \mathbf{Z}_i(\omega))]^2 W(\mathbf{Z}_i(\omega))|\\
&+ |[(P_i(\omega) - P_Y(f, \mathbf{Z}_i(\omega)))(P_Y(f_j, \mathbf{Z}_i(\omega)) - P_Y(f, \mathbf{Z}_i(\omega)))]W(\mathbf{Z}_i(\omega))|\}.\\
&\leqslant const.\frac{1}{n}\sum_{i=1}^{n}\{\sup_{g\in(f_j,f)}|P_Y(g, \mathbf{Z}_i(\omega))|^2\|f_j - f\|_{m,1}^2\\
&+ [(P_i(\omega) - P_Y(f, \mathbf{Z}_i(\omega)))\sup_{g\in(f_j,f)}|P_Y(g, \mathbf{Z}_i(\omega))|\|f_j - f\|_{m,1}\}\\
&\to 0
\end{aligned}
$$

where the last inequality follows from the mean value theorem, and Assumption

[1] To see this, note by Markov's inequality that $\mathbb{P}(|P_i| > M) \leqslant Var(P_i)/M^2$. Applying the Borel-Cantelli Lemma then shows that P_i is almost surely finite. See Billingsley (1995).

2.4.1 implies that the suprema are bounded for sufficiently large j. Hence $L_n(f)$ is almost surely continuous and therefore upper semi-continuous. On the other hand, $L_n(f) = \frac{1}{n}\sum_{i=1}^{n} -\frac{1}{2}[P_i - P_Y(f, \mathbf{Z}_i)]^2 W(\mathbf{Z}_i)$ is continuous in $\mathbf{Z}_i$ for each $f \in \mathcal{F}$ and is therefore measurable. Thus Condition 3.3'(i) is satisfied.

Condition 3.4: Compactness of the $\mathcal{F}_K$ is the result of Lemma A.1.9 (i).

Condition 3.5(i): Finally, we require the uniform convergence of the empirical criterion $L_n(f) = \frac{1}{n}\sum_{i=1}^{n} -\frac{1}{2}[P_i - P_Y(f, \mathbf{Z})]^2 W_i$ over sieves, i.e. $\forall K \geqslant 1$, $\sup_{f\in\mathcal{F}_K}|L_n(f) - L(f)| \xrightarrow{p} 0$ as $n \to \infty$. First, note that by Assumption 2.4.2 (i) and the law of large numbers, $L_n(f) - L(f) = o_p(1)$ pointwise in f on $\mathcal{F}_K$. Second, standard arguments show

$$
\begin{aligned}
\sup_{f\in\mathcal{F}_K} |L'_n(f)| &\leqslant \sup_{f\in\mathcal{F}_K} \frac{1}{n}\sum_{i=1}^{n} |P_i - P_Y(f, \mathbf{Z}_i)||W(\mathbf{Z}_i)| \\
&\leqslant \frac{1}{n}\sum_{i=1}^{n} |P_i W(\mathbf{Z}_i)| + \sup_{g\in\mathcal{F}_K} |P_Y(g, \mathbf{Z}_i)| \left(\frac{1}{n}\sum_{i=1}^{n} |W(\mathbf{Z}_i)|\right) \\
&\leqslant \left(\frac{1}{n}\sum_{i=1}^{n} |P_i|^2\right)^{1/2} \left(\frac{1}{n}\sum_{i=1}^{n} |W(\mathbf{Z}_i)|^2\right)^{1/2} + \sup_{g\in\mathcal{F}_K} |P_Y(g, \mathbf{Z}_i)| \left(\frac{1}{n}\sum_{i=1}^{n} |W(\mathbf{Z}_i)|\right).
\end{aligned}
$$

The first term is $O_p(1)$ by Assumption 2.4.2 (i). The second term is also $O_p(1)$ by the following arguments. By Lemma A.1.9 (i), the $\mathcal{F}_K$ are compact. Next, cover each point in $\mathcal{F}_K$ with balls of radius small enough to make the local boundedness Assumption 2.4.1 hold. By compactness of $\mathcal{F}_K$, there exists a finite subcover $\{U_i\}_{i=1}^{N}$ of $\mathcal{F}_K$ where for each set U_i in the subcover, $\sup_{f\in U_i} S(f, \mathbf{Z}) \leqslant M_i$ $\mathbb{P} - a.s.$ Then $M = \max\{M_1, \ldots, M_N\}$ is a bound on $\sup_{g\in\mathcal{F}_K}|P_Y(g, \mathbf{Z}_i)|$, so the second term in the above display is $O_p(1)$ under Assumption 2.4.2 (i). Hence, by the mean value theorem, for $f_1, f_2 \in \mathcal{F}_K$,

$$|L_n(f_1) - L_n(f_2)| \leqslant O_p(1)\|f_1 - f_2\|_{m,1}.$$

This Lipschitz condition, the compactness of $\mathcal{F}_K$, and the pointwise convergence of $L_n(f)$ to $L(f)$ mean that the conditions for Corollary 2.2 in Newey (1991) are met, so that $\sup_{h\in\mathcal{H}_K}|L_n(h) - L(h)| \xrightarrow{p} 0$, as required. Since the conditions for Chen's Theorem 3.1 are met, we conclude that $d(\hat{f}_n, f_0) = o_p(1)$. Applying Lemma A.1.10 gives $\|P_Y(\widehat{f}_n, \mathbf{Z}) - P_Y(f_0, \mathbf{Z})\|_2 \xrightarrow{p} 0$. □

Proof of Proposition 3 Recall that the option prices $\mathcal{P}_Y(\mathbf{Z})$ are generated by a conditional density, i.e. $\mathcal{P}_Y(\mathbf{Z}) \equiv P_Y(f, \mathbf{Z})$, where $f \in \mathcal{F}$ is the target of a Lipschitz map with preimage $f^{Y,X} = h^2 + \varepsilon_0 h_0$. The function $h \in \mathcal{H}$ lives in a Sobolev ball of radius $\mathcal{B}_0$. The complexity of the space of possible option prices $\mathcal{P}$ is then firmly linked to the complexity of the Sobolev ball $\mathcal{H}$. The proof strategy is therefore to establish this link, and then to apply Theorem 3.2 in Chen (2007) once we have a handle on the complexity of $\mathcal{P}$.

Application of Theorem 3.2 in Chen (2007) requires verification of her Conditions 3.6, 3.7, and 3.8, reproduced here for the current notation. It also requires the computation of a certain bracketing entropy integral, which is undertaken below. Condition 3.6 requires an i.i.d. sample, which we have already assumed in Assumption 2.4.2. It remains to check Conditions 3.7 and 3.8 and to compute the bracketing entropy integral.

Condition 3.7. There exists $C_1 > 0$ such that $\forall \varepsilon > 0$ small,

$$\sup_{P_Y \in B_\varepsilon(P_Y^0)} Var\Big(\ell(P_Y, Y_i) - \ell(P_Y^0, Y_i)\Big) \leqslant C_1 \varepsilon^2.$$

Condition 3.8. For all $\delta > 0$, there exists a constant $s \in (0, 2)$ such that

$$\sup_{P_Y \in B_\delta(P_Y^0)} \Big|\ell(P_Y, Y_i) - \ell(P_Y^0, Y_i)\Big| \leqslant \delta^s U(Y_i),$$

with $\mathbb{E}[U(Y_i)^\gamma] \leqslant C_2$ for some $\gamma \geqslant 2$.

First, note that $\ell(P_Y, Y_i) - \ell(P_Y^0, Y_i) = W(\mathbf{Z}_i)[P_Y(\mathbf{Z}_i) - P_Y^0(\mathbf{Z}_i)]\{e_i + \frac{1}{2}[P_Y(\mathbf{Z}_i) - P_Y^0(\mathbf{Z}_i)]\}$. Then

$$\begin{aligned}
&\mathbb{E}\{[\ell(P_Y, Y_i) - \ell(P_Y^0, Y_i)]^2\} \\
&\quad = \mathbb{E}\{W(\mathbf{Z}_i)^2[P_Y(\mathbf{Z}_i) - P_Y^0(\mathbf{Z}_i)]^2\{e_i + \frac{1}{2}[P_Y(\mathbf{Z}_i) - P_Y^0(\mathbf{Z}_i)]\}^2 \\
&\quad = \mathbb{E}\{W(\mathbf{Z}_i)^2[P_Y(\mathbf{Z}_i) - P_Y^0(\mathbf{Z}_i)]^2 e_i^2\} + \mathbb{E}\{\frac{1}{4}W(\mathbf{Z}_i)^2[P_Y(\mathbf{Z}_i) - P_Y^0(\mathbf{Z}_i)]^4\} \\
&\quad = \mathbb{E}\{W(\mathbf{Z}_i)^2[P_Y(\mathbf{Z}_i) - P_Y^0(\mathbf{Z}_i)]^2 \sigma(\mathbf{Z}_i)\} + \frac{1}{4}\mathbb{E}\{W(\mathbf{Z}_i)^2[P_Y(\mathbf{Z}_i) - P_Y^0(\mathbf{Z}_i)]^4\} \\
&\quad \leqslant const. \|P_Y - P_Y^0\|_2^2 + \frac{1}{4}\mathbb{E}\{W(\mathbf{Z}_i)^2[P_Y(\mathbf{Z}_i) - P_Y^0(\mathbf{Z}_i)]^4\}
\end{aligned}$$

where the last inequality uses the bound from Assumption 2.4.3. The second term on the RHS can be further bounded,

$$\begin{aligned}
&\mathbb{E}\{W(\mathbf{Z}_i)^2[P_Y(\mathbf{Z}_i) - P_Y^0(\mathbf{Z}_i)]^4\} \\
&\qquad \leqslant C \sup_{\mathbf{Z} \in \mathcal{Z}} [P_Y(\mathbf{Z}) - P_Y^0(\mathbf{Z})]^2 \mathbb{E}\{[P_Y(\mathbf{Z}_i) - P_Y^0(\mathbf{Z}_i)]^2 W(\mathbf{Z}_i)\} \\
&\qquad = C\|P_Y - P_Y^0\|_\infty^2 \|P_Y - P_Y^0\|_2^2
\end{aligned}$$

The smoothness of P_Y and P_Y^0 can be used to bound $\|P_Y - P_Y^0\|_\infty^2$ as follows. First, let

$$C^{j,\eta}(cl(\mathbb{R}^{d_u})) = \Big\{ f \in C^m(cl(\mathbb{R}^{d_u})) : \max_{|\lambda| \leqslant j} \sup_{\mathbf{u} \in \mathbb{R}^{d_u}} |D^\lambda f(\mathbf{u})| \leqslant L$$
$$\max_{|\lambda| = j} \sup_{\mathbf{u}_1 \neq \mathbf{u}_2 \in \mathbb{R}^{d_u}} \frac{|D^\lambda f(\mathbf{u}_1) - D^\lambda f(\mathbf{u}_2)|}{|\mathbf{u}_1 - \mathbf{u}_2|^\eta} \leqslant L \Big\}$$

denote a Hölder space. Let $m = j + k$, $\eta = 1$, $k = d_u + 1$. If the domain $\mathcal{Z}$ satisfies some mild regularity conditions, the Sobolev Embedding Theorem (Theorem 4.12 Adams and Fournier (2003) Part II) implies that $W^{m,1}(\mathbb{R}^{d_u}) \hookrightarrow C^{j,\eta}(cl(\mathbb{R}^{d_u}))$, where $j = m - k = m - d_u - 1 \geqslant 1$ by Assumption 2.4.5. Thus Assumption 2.4.5 ensures that $W^{m,1}(\mathbb{R}^{d_u})$ can be embedded in a Hölder space consisting of functions that are at least once continuously differentiable and therefore Lipschitz. Now, since the smoothness of

$$P_Y(\mathbf{Z}) = e^{-r\tau} \int_{-\infty}^{d(\mathbf{Z})} [\kappa - S_0 e^{\mu(\mathbf{Z}) + \sigma(\mathbf{Z})Y}] f(Y|\mathbf{X}) dY$$

depends on that of f (apply Leibniz' formula), which is continuously differentiable, then by Lemma 2 in Chen and Shen (1998), one has $\|P_Y - P_Y^0\|_\infty \leqslant \|P_Y - P_Y^0\|^{2/(2+d_x)}$. Therefore

$$\mathbb{E}\{W(\mathbf{Z}_i)^2 [P_Y(\mathbf{Z}_i) - P_Y^0(\mathbf{Z}_i)]^4\} \leqslant C \|P_Y - P_Y^0\|_2^{2+4/(2+d_x)},$$

and one has

$$\mathbb{E}\{[\ell(P_Y, Y_i) - \ell(P_Y^0, Y_i)]^2\} \leqslant const. \|P_Y - P_Y^0\|_2^2 + \frac{C}{4} \|P_Y - P_Y^0\|_2^{2+4/(2+d_x)}.$$

This implies that Condition 3.7 is satisfied for all $\varepsilon \leqslant 1$.

To show Condition 3.8, note that

$$|\ell(P_Y, Y_i) - \ell(P_Y^0, Y_i)| = |[P_Y(\mathbf{Z}_i) - P_Y^0(\mathbf{Z}_i)]\big[e_i + \frac{1}{2}[P_Y^0(\mathbf{Z}_i) - P_Y(\mathbf{Z}_i)]\big]|$$

$$\leqslant const.\|P_Y - P_Y^0\|_\infty\{|e_i| + \frac{1}{2}\|P_Y^0\|_\infty + \frac{1}{2}\|P_Y\|_\infty\}.$$

The terms involving $\|P_Y^0\|_\infty$ and $\|P_Y\|_\infty$ are bounded as a consequence of the local boundedness of the stock price in Assumption 2.4.1 as well as the compactness of the sieve space (Lemma A.1.9).[2] Thus Lemma 2 in Chen and Shen (1998) and another appeal to the Sobolev Embedding Theorem imply that

$$|\ell(P_Y, Y_i) - \ell(P_Y^0, Y_i)| \leqslant const.\|P_Y - P_Y^0\|_\infty U(Y_i)$$

$$\leqslant const.U(Y_i)\|P_Y - P_Y^0\|_2^{2/(2+d_x)}$$

for $U(Y_i) = |e_i| + const.$ Thus $s = 2/(2 + d_x)$ is the required modulus of continuity, and $\gamma = 2$ by Assumption 2.4.2. This establishes Condition 3.8.

An appeal to Chen (2007)'s Theorem 3.2 requires the computation of δ_n satisfying

$$\delta_n = \inf\left\{\delta \in (0,1) : \frac{1}{\sqrt{n}\delta^2}\int_{b\delta^2}^{\delta}\sqrt{H_{[\,]}(w, \mathcal{G}_n, \|\cdot\|_2)}dw\right\},$$

for the bracketing entropy $H_{[\,]}(w, \mathcal{G}_n, \|\cdot\|_2)$, where

$$\mathcal{G}_n = \{\ell(P_Y, Y_i) - \ell(P_Y^0, Y_i) : \|P_Y - P_Y^0\|_2 \leqslant \delta,\ P_Y \in \mathcal{P}_{K_n}\}.$$

[2] See also the argument in the proof of Proposition 2.

Consider the following chain of inequalities

$$
\begin{aligned}
|\ell(P_Y, Y_i) - \ell(P_Y^0, Y_i)| &= |[P_Y(\mathbf{Z}_i) - P_Y^0(\mathbf{Z}_i)||e_i + \frac{1}{2}[P_Y^0(\mathbf{Z}_i) - P_Y(\mathbf{Z}_i)]| \\
&\leqslant M_1 \|f - f_0\|_{m,1} U(Y_i) \\
&\leqslant M_2 U(Y_i) \|f^{Y,X} - f_0^{Y,X}\|_{m,\infty,\zeta_0} \quad \text{by Lemma A.1.6} \\
&= M_2 U(Y_i) \|(h^{Y,X})^2 - (h_0^{Y,X})^2\|_{m,\infty,\zeta_0} \quad \text{by Def. A.1.2} \\
&\leqslant M_3 U(Y_i) \|h^{Y,X} - h_0^{Y,X}\|_{m_0+m,2,\zeta_0} \qquad \text{(A.2.4)}
\end{aligned}
$$

To see the last inequality, observe that

$$
\begin{aligned}
\|(h^{Y,X})^2 - (h_0^{Y,X})^2\|_{m,\infty,\zeta_0} &\leqslant C \|h^{Y,X} + h_0^{Y,X}\|_{m,\infty,\zeta_0^{1/2}} \|h^{Y,X} - h_0^{Y,X}\|_{m,\infty,\zeta_0^{1/2}} \\
&\leqslant C_1 \|\zeta_0^{1/2}(h^{Y,X} + h_0^{Y,X})\|_{m,\infty} C_2 \|\zeta_0^{1/2}(h^{Y,X} - h_0^{Y,X})\|_{m,\infty} \\
&\leqslant C_3 \|h^{Y,X} + h_0^{Y,X}\|_{m_0+m,2,\zeta_0} C_4 \|h^{Y,X} - h_0^{Y,X}\|_{m_0+m,2,\zeta_0} \\
&\leqslant C_3 (2\mathcal{B}_0) C_4 \|h^{Y,X} - h_0^{Y,X}\|_{m_0+m,2,\zeta_0}.
\end{aligned}
$$

for some constants M_j and C_j, and where the first inequality follows from Gallant and Nychka (1987) Lemma A.3, the second from Gallant and Nychka (1987) Lemma A.1(d), the third from Gallant and Nychka (1987) Lemma A.1(b), and the fourth by the definition of $\mathcal{H}_n$ as a bounded Sobolev ball.

Theorem 2.7.11 in Van Der Vaart and Wellner (1996) implies that the bracketing number for $\mathcal{G}_n$ can be bounded

$$
N_{[\,]}(w, \mathcal{G}_n, \|\cdot\|_2) \leqslant N\left(\frac{w}{2CM_3}, \mathcal{H}_n, \|\cdot\|_{m_0+m,2,\zeta_0}\right),
$$

where the RHS is by the covering number of a Sobolev ball with dimension $K_n \equiv [K_y(n)+1][K_{x,1}(n)+1]\ldots[K_{x,d_x}(n)+1]$. By Lemma 2.5 in Van De Geer (2000), we can further bound the RHS, giving

$$N_{[\,]}(w, \mathcal{G}_n, \|\cdot\|_2) \leqslant N\left(\frac{w}{2CM_3}, \mathcal{H}_n, \|\cdot\|_{m_0+m,2,\zeta_0}\right) \leqslant \left(1 + \frac{8\mathcal{B}_0 CM_3}{w}\right)^{K_n}.$$

Therefore,

$$\begin{aligned}\frac{1}{\sqrt{n}\delta_n^2}\int_{b\delta_n^2}^{\delta_n}\sqrt{H_{[\,]}(w, \mathcal{G}_n, \|\cdot\|_2)}dw &\leqslant \frac{1}{\sqrt{n}\delta_n^2}\int_{b\delta_n^2}^{\delta_n}\sqrt{K_n \log\left(1 + \frac{8\mathcal{B}_0 CM_3}{w}\right)}dw \\ &\leqslant C\frac{1}{\sqrt{n}\delta_n^2}\sqrt{K_n}\delta_n,\end{aligned}$$

which is less than or equal to a constant for the choice $\delta_n \asymp \sqrt{K_n/n}$. Put $K_y(n) \asymp K_{x,1}(n) \asymp \cdots \asymp K_{x,d_x}(n) \asymp n^{1/(2(m_0+m)+d_u)}$, so that $K_n \asymp n^{d_u/(2(m_0+m)+d_u)}$, yielding

$$\delta_n \asymp \frac{\sqrt{K_n}}{\sqrt{n}} \asymp n^{d_u/[2(2(m_0+m)+d_u)]} n^{-1/2} = n^{\frac{-(m_0+m)}{2(m_0+m)+d_u}}.$$

On the other hand, this choice of K_n combined with Assumption 2.4.4 yields the approximation error rate

$$\|[P_Y(\mathbf{Z}_i) - P_Y^0(\mathbf{Z}_i)\|_2 \leqslant const.\|h^{Y,X} - h_0^{Y,X}\|_{m_0+m,2,\zeta_0} = O(K_n^{-\alpha}) = O\left(n^{\frac{-\alpha d_u}{2(m_0+m)+d_u}}\right),$$

where the inequality follows from the ones in Eq. (A.2.4). Applying Chen (2007)'s Theorem 3.2 yields the stated result. □

Proof of Proposition 4 I verify Assumptions 3.1 - 3.4 of Chen et al. (2013) (CLS).

Linearity of $v \mapsto \frac{\partial\Gamma(P_Y^0)}{\partial P_Y}[v]$ is satisfied for the linear functional Γ in Eq. (2.5.1), since $\frac{\partial\Gamma(P_Y^0)}{\partial P_Y}[av_1 + bv_2] = \int\omega(\mathbf{Z})[av_1 + bv_2]d\mathbf{Z}_1 = a\int\omega(\mathbf{Z})v_1(\mathbf{Z})d\mathbf{Z}_1 + b\int\omega(\mathbf{Z})v_2(\mathbf{Z})d\mathbf{Z}_1 =$

$a\frac{\partial\Gamma(P_Y^0)}{\partial P_Y}[v_1] + b\frac{\partial\Gamma(P_Y^0)}{\partial P_Y}[v_2]$. CLS Assumption 3.1(ii) is also trivially satisfied, since Γ is a linear functional. By Assumption 2.5.1 (iii) and the rate in Proposition 3, we also have $\|v_n^* - v^*\| \times \|P_Y^{0,n} - P_Y^0\| = o(n^{-1/2})$. Thus CLS Assumption 3.1 is satsified.

CLS Assumption 3.2 is directly assumed under our Assumption 2.5.1 (iv).

CLS Assumption 3.3(i) follows from the linearity of $\ell'(P_Y^0, \Xi)[av_1 + bv_2] = [P - P_Y^0(\mathbf{Z})]W(\mathbf{Z})(av_1(\mathbf{Z}) + bv_2(\mathbf{Z})) = a\ell'(P_Y^0, \Xi)[v_1] + b\ell'(P_Y^0, \Xi)[v_2]$. To show CLS Assumption 3.3(ii), we invoke Lemma 4.2 of Chen (2007). Take

$$\sup_{\{\|P_Y - P_Y^0\|_2 \leqslant \delta\}} |[P_Y^0(\mathbf{Z}) - P_Y(\mathbf{Z})]\frac{\partial P_Y(\overline{\beta})}{\partial \beta} R_{K_n}^{-1} G_{K_n}| \leqslant M\delta.$$

by Assumption 2.5.1 (ii). Lemma B.2.1 implies that the entropy integral of Chen (2007) (4.2.2) is satisfied. CLS Assumption 3.3(ii) therefore follows after invoking Lemma 4.2 of Chen (2007). Next, note that CLS Assumption 3.3.(iii) follows by definition of the least squares objective function (see e.g. Shen (1997) Example 1).

Finally, define the empirical process $\mu\{g(\Xi)\} = \frac{1}{n}\sum_{i=1}^n g(\Xi_i) - \mathbb{E}g(\Xi_i)$, and let $u_n^* = v_n^*/\|v_n^*\|_{sd}$. Then

$$\begin{aligned}
\sqrt{n}\mu\{\ell'(P_Y^0, \Xi)[u_n^*]\} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\ell'(P_Y^0, \Xi)[v_n^*]}{Var(\ell'(P_Y^0, \Xi_i)[v_n^*])} \\
&= \left(G'_{K_n} R_{K_n}^{-1} \Sigma_{K_n} R_{K_n}^{-1} G_{K_n}\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \ell'(P_Y^0, \Xi)[v_n^*]\right) \\
&= \left(G'_{K_n} R_{K_n}^{-1} \Sigma_{K_n} R_{K_n}^{-1} G_{K_n}\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i \frac{\partial P_Y(\overline{\beta})}{\partial \beta} R_{K_n}^{-1} G_{K_n}\right) \\
&\xrightarrow{d} N(0,1)
\end{aligned}$$

by continuous mapping theorem and a standard central limit theorem for i.i.d.

samples. CLS Assumption 3.4 then follows.

Chen et al. (2013) Theorem 3.1 then implies

$$\sqrt{n}\frac{\Gamma(\widehat{P}_Y) - \Gamma(P_Y^0)}{\|v_n^*\|_{sd}} \xrightarrow{d} N(0,1).$$

Next, we want to replace $\|v_n^*\|_{sd}$ with its estimate

$$\|\widehat{v}_n^*\|_{sd,n}^2 \equiv \widehat{Var}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \ell'(\widehat{P}_Y, \Xi_i)[\widehat{v}_n^*]\right). \tag{A.2.5}$$

First note that Corollary B.2.2 implies that

$$|\frac{\|\widehat{v}_n^*\|}{\|v_n^*\|} - 1| = O_p(\epsilon_n^*), \qquad \frac{\|\widehat{v}_n^* - v_n^*\|}{\|v_n^*\|} = O_p(\epsilon_n^*)$$

by application of CLS Lemma 5.1. Then following the proof of CLS Theorem 5.1 with modifications to the weight functions, consider

$$\begin{aligned}
\widehat{\Lambda} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \ell'(\widehat{P}_Y, \Xi_i)[\widehat{v}_n^*] \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big\{\ell'(\widehat{P}_Y, \Xi_i)[\widehat{v}_n^*] - \mathbb{E}[\ell'(\widehat{P}_Y, \Xi_i)[\widehat{v}_n^*]] - \ell'(P_Y^0, \Xi_i)[\widehat{v}_n^*] + \mathbb{E}[\ell'(P_Y^0, \Xi_i)[\widehat{v}_n^*]]\Big\} \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big\{\mathbb{E}[\ell'(\widehat{P}_Y, \Xi_i)[\widehat{v}_n^*]] - \mathbb{E}[\ell'(P_Y^0, \Xi_i)[\widehat{v}_n^*]] - \mathbb{E}[r(P_Y^0, \Xi_i)[\widehat{v}_n^*, \widehat{P}_Y - P_Y^0]]\Big\} \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big\{\mathbb{E}[r(P_Y^0, \Xi_i)[\widehat{v}_n^*, \widehat{P}_Y - P_Y^0]]\Big\} \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^n \ell'(P_Y^0, \Xi_i)[\widehat{v}_n^*] \\
&= \hat{I}_1 + \hat{I}_2 + \hat{I}_3 + \hat{I}_4.
\end{aligned}$$

Corollary B.2.2 (iii) and (iv) implies that $\hat{I}_1 = o_p(\|\widehat{v}_n^*\|)$ and $\hat{I}_2 = O_p(\sqrt{n}\epsilon_n^*\varepsilon_n\|\widehat{v}_n^*\|)$, which makes

$$\begin{aligned}\widehat{\Lambda} &= o_p(\|\widehat{v}_n^*\|) + O_p(\sqrt{n}\epsilon_n^*\varepsilon_n\|\widehat{v}_n^*\|)\\ &+ \frac{1}{\sqrt{n}}\sum_{i=1}^n\{\langle\widehat{v}_n^* - v_n^*, \widehat{P}_Y - P_Y^0\rangle + \langle v_n^*, \widehat{P}_Y - P_Y^0\rangle\}\\ &+ \frac{1}{\sqrt{n}}\sum_{i=1}^n\ell'(P_Y^0, \Xi_i)[\widehat{v}_n^* - v_n^*] + \frac{1}{\sqrt{n}}\sum_{i=1}^n\ell'(P_Y^0, \Xi_i)[v_n^*].\end{aligned}$$

By arguments similar to the proof of CLS Theorem 5.1, we have

$$\sqrt{n}\|v_n^*\|_{sd}^{-1}\langle v_n^*, \widehat{P}_Y - P_Y^0\rangle = \|v_n^*\|_{sd}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\ell'(P_Y^0, \Xi_i)[v_n^*] + o_p(1)$$

and

$$|\sqrt{n}\langle\widehat{v}_n^* - v_n^*, \widehat{P}_Y - P_Y^0\rangle| \leqslant \sqrt{n}\|\widehat{v}_n^* - v_n^*\|\|\widehat{P}_Y - P_Y^0\| = O_p(\sqrt{n}\|v_n^*\|\epsilon_n^*\varepsilon_n)$$

and

$$|\frac{1}{\sqrt{n}}\sum_{i=1}^n\ell'(P_Y^0, \Xi_i)[\widehat{v}_n^* - v_n^*]| \leqslant \|\widehat{v}_n^* - v_n^*\|\sup_{v\in\mathcal{W}_n}|\frac{1}{\sqrt{n}}\sum_{i=1}^n\ell'(P_Y^0, \Xi_i)[v]| = O_p(\|v_n^*\|\epsilon_n^*).$$

Since $\sqrt{n}\epsilon_n^*\varepsilon_n = o(1)$, we combine these results to obtain

$$\begin{aligned}\frac{\|\widehat{v}_n^*\|_{sd,n}}{\|v_n^*\|_{sd}} &= \|v_n^*\|_{sd}^{-1}\widehat{Var}\left(\widehat{\Lambda}\right)\\ &= \|v_n^*\|_{sd}^{-1}\widehat{Var}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\ell'(P_Y^0, \Xi_i)[v_n^*] + o_p(1)\right)\\ &\xrightarrow{p} \|v_n^*\|_{sd}^{-1}\|v_n^*\|_{sd} = 1.\end{aligned}$$

Therefore

$$\frac{\sqrt{n}[\Gamma(\widehat{P}_Y) - \Gamma(P_Y)]}{\|\widehat{v}_n^*\|_{sd,n}} = \frac{\sqrt{n}[\Gamma(\widehat{P}_Y) - \Gamma(P_Y)]}{\|v_n^*\|_{sd,n}} \frac{\|v_n^*\|_{sd,n}}{\|\widehat{v}_n^*\|_{sd,n}} \quad \xrightarrow{d} \quad N(0,1)$$

□

Appendix B

Implementation Details, Riesz Representors, and Technical Lemmas

B.1 Implementation Details

B.1.1 Gradients and Hessian

The objective function to be minimized is

$$L_n(\beta) \equiv \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\Big[p_i - P_X(\beta, \mathbf{z}_i)\Big]^2 W_i \equiv \frac{1}{n}\sum_{i=1}^{n}\ell(\beta, Y_i). \tag{B.1.1}$$

The gradient and Hessian of the objective function are

$$\frac{\partial L_n(\widehat{\beta})}{\partial \beta} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \ell(\widehat{\beta}, Y_i)}{\partial \beta} = \frac{1}{n}\sum_{i=1}^{n} -\Big[p_i - P_X(\widehat{\beta}, \mathbf{z}_i)\Big] W_i \frac{\partial P_X(\widehat{\beta}, \mathbf{z}_i)}{\partial \beta} \tag{B.1.2}$$

$$\frac{\partial^2 L_n(\widehat{\beta})}{\partial\beta\partial\beta'} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \ell(\widehat{\beta}, Y_i)}{\partial\beta\partial\beta'}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{-\Big[p_i - P_X(\widehat{\beta}, \mathbf{z}_i)\Big]W_i\frac{\partial^2 P_X(\widehat{\beta}, \mathbf{z}_i)}{\partial\beta\partial\beta'} + W_i\frac{\partial P_X(\widehat{\beta}, \mathbf{z}_i)}{\partial\beta}\frac{\partial P_X(\widehat{\beta}, \mathbf{z}_i)}{\partial\beta}'\right\}. \tag{B.1.3}$$

The pricing function $P_X(\beta, \mathbf{z}_i)$ is found in Proposition 1,

$$P_X(\beta, \mathbf{z}_i) = \kappa e^{-r_i\tau_i}\left[\Phi(d(\mathbf{z}_i)) - \sum_{k=1}^{2K_x(n)}\frac{\gamma_k(\beta,\tau_i)}{\sqrt{k}}H_{k-1}(d(\mathbf{z}_i))\phi(d(\mathbf{z}_i))\right]$$
$$- S_i e^{-r_i\tau_i+\mu(\mathbf{z}_i)}\left[e^{\sigma(\mathbf{z}_i)^2/2}\Phi(d(\mathbf{z}_i) - \sigma(\mathbf{z}_i)) + \sum_{k=1}^{2K_x(n)}\gamma_k(\beta,\tau_i)I_k^*(d(\mathbf{z}_i))\right] \tag{B.1.4}$$

where $\Phi(\cdot)$ is the standard normal CDF, $K_n \equiv (K_x(n), K_\tau(n))$, and where

$$I_k^*(d(\mathbf{z}_i)) = \frac{\sigma(\mathbf{z}_i)}{\sqrt{k}}I_{k-1}^*(d(\mathbf{z}_i)) - \frac{1}{\sqrt{k}}e^{\sigma(\mathbf{z}_i)d(\mathbf{z}_i)}H_{k-1}(d(\mathbf{z}_i))\phi(d(\mathbf{z}_i)), \quad \text{for } k \geqslant 1,$$
$$I_0^*(d(\mathbf{z}_i)) = e^{\sigma(\mathbf{z}_i)^2/2}\Phi(d(\mathbf{z}_i) - \sigma(\mathbf{z}_i)),$$

and $\gamma_k(\beta, \tau_i)$, $\beta = vec(B)$ is the coefficient function

$$\gamma_k(\beta,\tau_i) = \frac{\alpha(B,\tau)'A_k\alpha(B,\tau)}{\alpha(B,\tau)'\alpha(B,\tau)},$$

A_k is the known matrix of constants in Leon, Mencia, Sentana (2009) Prop. 1, and

$$\alpha(B,\tau) = \begin{bmatrix} \sum_{j=0}^{Kz}\beta_{0j}H_j(\tau) \\ \vdots \\ \sum_{j=0}^{Kz}\beta_{K_xj}H_j(\tau) \end{bmatrix} = B \cdot H$$

for

$$B \equiv \begin{bmatrix} \beta_{00} & \beta_{01} & \dots & \beta_{0K_\tau} \\ \beta_{10} & \beta_{11} & \dots & \beta_{1K_\tau} \\ \vdots & & & \\ \beta_{K_x0} & \beta_{K_x1} & \dots & \beta_{K_xK_\tau} \end{bmatrix}, \quad H \equiv \begin{bmatrix} H_0(\tau) \\ \vdots \\ H_{K_\tau}(\tau) \end{bmatrix}.$$

The only place where β shows up in $P_X(\beta, \mathbf{z})$ is through each of the $\gamma_k(\beta, \tau)$, $k = 1, \dots, K_x$. Hence to find first and second derivatives of $P_X(\cdot, \mathbf{z})$ we must find them for $\gamma_k(\cdot, \tau)$.

Note that

$$\gamma_k(B, \tau) = [H'B'B'H]^{-1}H'B'A_kBH.$$

I suppress the k subscript on A in subsequent derivations. Using the matrix differential conventions in Fackler (2005), it can be shown that

$$\begin{aligned} \frac{\partial\gamma_k(B,\tau)}{\partial B} &\equiv \frac{\partial vec\{\gamma_k(B,\tau)\}}{\partial vec\{B\}} = \frac{\partial\gamma_k(\beta,\tau)}{\partial\beta} \\ &= [H'B'BH]^{-1}\Big\{(H' \otimes H'B'A) + (H'B'A' \otimes H')T_{(K_x+1),(K_\tau+1)}\Big\} \\ &- [H'B'BH]^{-2}[H'B'ABH]\Big\{(H' \otimes H'B') + (H'B' \otimes H')T_{(K_x+1),(K_\tau+1)}\Big\}, \end{aligned} \tag{B.1.5}$$

where $T_{m,n}$ is an $mn \times mn$ permutation matrix satisfying for any matrix $C_{m\times n}$, $vec\{C'\} = T_{m,n}vec\{C\}$. Then,

$$\begin{aligned} \frac{\partial P_X(\widehat{\beta}, \mathbf{z}_i)}{\partial\beta} &= \kappa e^{-r_i\tau_i}\left[\Phi(d(\mathbf{z}_i)) - \sum_{k=1}^{2K_x(n)} \frac{1}{\sqrt{k}}\frac{\partial\gamma_k(\beta,\tau)}{\partial\beta}H_{k-1}(d(\mathbf{z}_i))\phi(d(\mathbf{z}_i))\right] \\ &- S_ie^{-r_i\tau_i+\mu(\mathbf{z}_i)}\left[e^{\sigma(\mathbf{z}_i)^2/2}\Phi(d(\mathbf{z}_i) - \sigma(\mathbf{z}_i)) + \sum_{k=1}^{2K_x(n)} \frac{\partial\gamma_k(\beta,\tau)}{\partial\beta}I_k^*(d(\mathbf{z}_i))\right]. \end{aligned}$$

To obtain the Hessian, decompose the expression in Eq. (B.1.5) into its four terms,

$$f_1(B) = [H'B'BH]^{-1}(H' \otimes H'B'A)$$

$$f_2(B) = [H'B'BH]^{-1}(H'B'A' \otimes H')T_{(K_x+1),(K_\tau+1)}$$

$$f_3(B) = [H'B'BH]^{-2}[H'B'ABH](H' \otimes H'B')$$

$$f_4(B) = [H'B'BH]^{-2}[H'B'ABH](H'B' \otimes H')T_{(K_x+1),(K_\tau+1)}$$

To avoid clutter, let $u_1(B) = [H'B'BH]^{-1}$, $v_1(B) = (H' \otimes H'B'A)$. Then

$$u_1'(B) = -[H'H'BH]^{-2}\Big\{(H' \otimes H'B') + (H'B' \otimes H')T_{(K_x+1),(K_\tau+1)}\Big\}$$

$$v_1'(B) = H' \otimes (H \otimes A')$$

so

$$f_1'(B) = v_1(B)'u_1'(B) + [I_{(K_x+1)(K_\tau+1)} \otimes u_1(B)]v_1'(B).$$

Next, for $v_2(B) \equiv (H'B'A' \otimes H')T_{(K_x+1),(K_\tau+1)}$,

$$v_2'(B) = H' \otimes T'_{(K_x+1),(K_\tau+1)}(A \otimes H),$$

so

$$f_2'(B) = v_2(B)'u_1'(B) + [I_{(K_x+1)(K_\tau+1)} \otimes u_1(B)]v_2'(B).$$

Next, note that $f_3(B)$ and $f_4(B)$ are products of three functions of B, so I make use of the product rule for $i = 3, 4$:

$$f_i'(B) = Du(B)_{m\times p}v(B)_{p\times q}w(B)_{q\times n} = [w(B)'v(B)' \otimes I_m]u'(B)$$
$$+ (I_n \otimes u(B))\Big\{(w(B)' \otimes I_p)v'(B) + (I_n \otimes v(B)w'(B))\Big\}.$$

In this setup, $m = p = q = 1$ and $n = (K_x + 1)(K_\tau + 1)$, the number of free coefficients. To obtain $f_3(B)$, set

$$u(B) = [H'B'BH]^{-2}, \quad v(B) = H'B'ABH, \quad w(B) = H' \otimes H'B'.$$

Then

$$u'(B) = -2[H'B'BH]^{-3}\Big\{(H' \otimes H'B') + (H'B' \otimes H')T_{(K_x+1),(K_\tau+1)}\Big\}$$

$$v'(B) = (H' \otimes H'B'A) + (H'B'A' \otimes H')T_{(K_x+1),(K_\tau+1)}$$

$$w'(B) = H' \otimes (H \otimes I_{K_x+1}).$$

The expression for $f_4'(B)$ is obtained similarly by using the same $u(B)$ and $v(B)$ as above but by changing $w(B)$ to

$$w(B) = (H'B' \otimes H')T_{(K_x+1),(K_\tau+1)}$$

$$w'(B) = H' \otimes T'_{(K_x+1),(K_\tau+1)}(I_{K_x+1} \otimes H).$$

Combining these gives

$$\frac{\partial^2 \gamma_k(\beta, \tau_i)}{\partial\beta\partial\beta'} = f_1'(B) + f_2'(B) + f_3'(B) + f_4'(B).$$

Then,

$$\frac{\partial^2 P_X(\hat{\beta}, \mathbf{z}_i)}{\partial\beta\partial\beta'} = \kappa e^{-r_i\tau_i}\left[\Phi(d(\mathbf{z}_i)) - \sum_{k=1}^{2K_x(n)} \frac{1}{\sqrt{k}} \frac{\partial^2 \gamma_k(\beta, \tau_i)}{\partial\beta\partial\beta'} H_{k-1}(d(\mathbf{z}_i))\phi(d(\mathbf{z}_i))\right]$$
$$- S_i e^{-r_i\tau_i + \mu(\mathbf{z}_i)}\left[e^{\sigma(\mathbf{z}_i)^2/2}\Phi(d(\mathbf{z}_i) - \sigma(\mathbf{z}_i)) + \sum_{k=1}^{2K_x(n)} \frac{\partial^2 \gamma_k(\beta, \tau_i)}{\partial\beta\partial\beta'} I_k^*(d(\mathbf{z}_i))\right].$$

B.2 Technical Definitions and Results

B.2.1 Riesz Representors

The aim is to connect the sieve asymptotic theory with simple non-linear least squares implementations. It is therefore helpful to adopt the notation $P_X^{K_n}(\mathbf{Z}) = P_X(\beta_n, \mathbf{Z})$ and $\ell(P_X^{K_n}, \mathbf{Z}) = \ell(\beta_n, \mathbf{Z})$ for the purposes of this section. Then following Chen et al. (2013), one can define the inner product

$$\langle P_X^1 - P_X^0, P_X^2 - P_X^0 \rangle \equiv -\mathbb{E}\{r(P_X^0, Y)[P_X^1 - P_X^0, P_X^2 - P_X^0]\},$$

where

$$r(P_X^0, Y)[P_X^1 - P_X^0, P_X^2 - P_X^0] \equiv \left. \frac{\partial \ell'(P_X^0 + \eta(P_X^2 - P_X^0), Y)[P_X^1 - P_X^0]}{\partial \eta} \right|_{\eta=0}$$

can be interpreted as a second-order Gateaux derivative in the directions $P_X^1 - P_X^0$ and $P_X^2 - P_X^0$. The associated norm is given by

$$\|P_X - P_X^0\|^2 = -\mathbb{E}\{r(P_X^0, Y)[P_X - P_X^0, P_X - P_X^0]\}.$$

Heuristically, this norm measures deviations of the objective function from its linear approximation and will have a Hessian interpretation later on.

In light of the consistency and rate results in Propositions 2 and 3, one can confine the analysis to the local setting of Chen et al. (2013). That is, the convergence rate ε_n in Proposition 3 implies that $\widehat{P}_X \in \mathcal{B}_n$ with probability approaching one, where

$$\mathcal{B}_n \equiv \mathcal{B}_0 \cap \mathcal{P}_{K_n}, \qquad \text{where } \mathcal{B}_0 \equiv \{P_X \in \mathcal{P}_{K_n} : \|P_X - P_X^0\|_2 \leqslant \varepsilon_n \log\log n\}.$$

Let $\mathcal{V} \equiv clsp(\mathcal{B}_0) - \{P_X^0\}$ and $\mathcal{V}_n \equiv clsp(\mathcal{B}_n) - \{P_X^{0,n}\}$, where $clsp(\cdot)$ denotes the closed linear span and where $P_X^{0,n} = \pi_{K_n} P_X^0$ denotes the orthogonal projection of P_X^0 onto the sieve space $\mathcal{P}_{K_n}$.

$\mathcal{V}_n$ is a finite-dimensional Hilbert space, which implies that the functional $\Gamma(P_X)$ in Eq. (2.5.1) has a Riesz representer $v_n^* \in \mathcal{V}_n$ such that the Gateaux derivative in the direction $v \in \mathcal{V}_n$ can be expressed as an inner product

$$\frac{\partial \Gamma(P_X^0)}{\partial P_X}[v] \equiv \left.\frac{\partial \Gamma(P_X^0 + \eta v)}{\partial \eta}\right|_{\eta=0} = \langle v_n^*, v \rangle$$

and

$$\frac{\partial \Gamma(P_X^0)}{\partial P_X}[v_n^*] = \|v_n^*\|^2 = \sup_{v \in \mathcal{V}_n, v \neq 0} |\frac{\partial \Gamma(P_X^0)}{\partial P_X}[v]|^2 / \|v\|^2. \qquad \text{(B.2.1)}$$

To get a step closer to familiar expressions from non-linear least squares asymptotic theory, one linearizes the option pricing function P_X. Since any $v \in \mathcal{V}_n$ has the form $v = P_X^{K_n} - P_X^{0,n}$, one has by mean value theorem $v = \frac{\partial P_X(\overline{\beta}, \mathbf{Z})}{\partial \beta}(\beta_n - \beta_{0,n})$ for $\overline{\beta}$ between β_n and the coefficients of the projection $P_X^{0,n}$. Thus $v_n^* = \frac{\partial P_X(\overline{\beta}, \mathbf{Z})}{\partial \beta}(\beta_n^* - \beta_{0,n})$ for some β_n^* that depends on the functional $\Gamma(P_X^{K_n})$.

Now, for each $\gamma_n = (\beta_n - \beta_{0,n})$, define the associated

$$G_{K_n} \equiv \frac{\partial \Gamma_{\mathbf{z}}(P_X^0)}{\partial P_X}' \frac{\partial P_X(\overline{\beta}, \mathbf{Z})}{\partial \beta}, \qquad R_{K_n} \equiv \mathbb{E}\left\{ -\frac{\partial^2 \ell(\overline{\beta}, Y)}{\partial \beta \partial \beta'} \right\}.$$

In this notation, the problem in Eq. (B.2.1) translates to finding the solution

$$\gamma_n^* = \arg \sup_{\gamma_n \in \mathbb{R}^{K_n}, \gamma_n \neq 0} \frac{\gamma_n' G_{K_n} G_{K_n}' \gamma_n}{\gamma_n' R_{K_n} \gamma_n},$$

which, following derivations similar to Chen et al. (2013), is given by

$$\gamma_n^* = R_{K_n}^{-1} G_{K_n}.$$

Therefore,

$$v_n^* = \frac{\partial P_X(\overline{\beta}, \mathbf{Z})}{\partial \beta}(\beta_n^* - \beta_{0,n}) = \frac{\partial P_X(\overline{\beta}, \mathbf{Z})}{\partial \beta}\gamma_n^* = \frac{\partial P_X(\overline{\beta}, \mathbf{Z})}{\partial \beta} R_{K_n}^{-1} G_{K_n},$$

which by definition implies the norm

$$\|v_n^*\|^2 = G'_{K_n} R_{K_n}^{-1} G_{K_n}.$$

Finally, the score process

$$\begin{aligned}
\ell'(P_X^0, Y_i)[v_n^*] &= [P_i - P_X^0(\mathbf{Z}_i)]W(\mathbf{Z}_i)v_n^* \\
&= [P_i - P_X^0(\mathbf{Z}_i)]W(\mathbf{Z}_i)\frac{\partial P_X(\overline{\beta}, \mathbf{Z}_i)}{\partial \beta}\gamma_n^* \\
&= e_i W(\mathbf{Z}_i)\frac{\partial P_X(\overline{\beta}, \mathbf{Z}_i)}{\partial \beta}\gamma_n^*
\end{aligned}$$

is required, with so-called standard deviation norm

$$\begin{aligned}
\|v_n^*\|_{sd}^2 &= Var\Big(\ell'(P_X^0, Y_i)[v_n^*]\Big) \\
&= \gamma_n'^* \mathbb{E}\left[e_i^2 W(\mathbf{Z}_i)^2 \frac{\partial P_X(\overline{\beta}, \mathbf{Z}_i)}{\partial \beta}\frac{\partial P_X(\overline{\beta}, \mathbf{Z}_i)}{\partial \beta}'\right]\gamma_n^* \\
&= G'_{K_n} R_{K_n}^{-1} \Sigma_{K_n} R_{K_n}^{-1} G_{K_n}. \qquad \text{(B.2.2)}
\end{aligned}$$

This object can be estimated by replacing the Riesz representer v_n^* with an estimate $\widehat{v}_n^*$. Define

$$\widehat{R}_{K_n} = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\ell(\widehat{\beta}_n, Y)}{\partial\beta\partial\beta'} \qquad \widehat{\Sigma}_{K_n} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial\ell(\widehat{\beta}_n, Y)}{\partial\beta}\frac{\partial\ell(\widehat{\beta}_n, Y)}{\partial\beta}'$$

$$\widehat{G}_{K_n} = \int_{\mathcal{Z}_1}\omega(\mathbf{Z})\frac{\partial P_X(\widehat{\beta}_n, \mathbf{Z})}{\partial\beta}d\mathbf{Z}_1 \qquad \widehat{v}_n^* = \frac{\partial P_X(\overline{\beta}, \mathbf{Z})}{\partial\beta}\widehat{R}_{K_n}^{-1}\widehat{G}_{K_n}.$$

Then

$$\|\widehat{v}_n^*\|_{sd,n}^2 = \widehat{G}'_{K_n}\widehat{R}_{K_n}^{-1}\widehat{\Sigma}_{K_n}\widehat{R}_{K_n}^{-1}\widehat{G}_{K_n} \equiv \widehat{V}_n \tag{B.2.3}$$

corresponds to the usual variance estimator using the familiar parametric Delta method.

B.2.2 Technical Lemmas

Lemma B.2.1. *For small $\delta > 0$, the subset of option pricing functions $\mathcal{G}(\delta) \equiv \{P_X^1, P_X^2 \in \mathcal{P} : \|P_X^1 - P_X^2\|_2 \leqslant \delta\}$ is $\mathbb{P}$-Donsker.*

Proof. By Lemma A.1.10,

$$\begin{aligned}
\|P_X^1 - P_X^2\|_2 &\leqslant M_1\|f_1 - f_2\|_{m,1} \\
&\leqslant M_2\|f^{X,Z} - f_0^{X,Z}\|_{m,\infty,\zeta_0} \quad \text{by Lemma A.1.6} \\
&= M_2\|(h^{X,Z})^2 - (h_0^{X,Z})^2\|_{m,\infty,\zeta_0} \quad \text{by Def. A.1.2} \\
&\leqslant M_3\|h^{X,Z} - h_0^{X,Z}\|_{m_0+m,2,\zeta_0} \qquad \text{(B.2.4)} \\
&\leqslant 2M_3\mathcal{B}_0. \qquad \text{(B.2.5)}
\end{aligned}$$

Therefore $\mathcal{F}(\delta) \equiv \{f \in \mathcal{F} : P_X(f, \mathbf{Z}) \in \mathcal{P}(\delta)\}$ is a bounded subset of the weighted Sobolev space $W^{m_0+m,2,\zeta_0}(\mathbb{R}^{d_u})$. Therefore we can think of $\mathcal{P}(\delta)$ as being Lipschitz in an index parameter that is a bounded subset of $W^{m_0+m,2,\zeta_0}(\mathbb{R}^{d_u})$.

Theorem 2.7.11 in Van Der Vaart and Wellner (1996) then implies that the bracketing number for $\mathcal{G}(\delta)$ can be bounded, i.e.

$$N_{[\,]}(w, \mathcal{G}(\delta), \|\cdot\|_2) \leqslant N\left(\frac{w}{4\mathcal{B}_0 M_3}, \mathcal{H}^{X,Z}, \|\cdot\|_{m_0+m,2,\zeta_0}\right) \leqslant N\left(\frac{w}{4\mathcal{B}_0 M_3}, \mathcal{H}^{X,Z}, \|\cdot\|_{\infty}\right),$$

where the second inequality follows from Gallant and Nychka (1987) Lemma A.1(c). Therefore,

$$H_{[\,]}(w, \mathcal{G}(\delta), \|\cdot\|_2) \leqslant C_2 w^{-d_u/m}$$

by Corollary 4 of Nickl and Pötscher (2007). Because $m > d_u/2$ by assumption on the Gallant-Nychka spaces, we have that

$$\int_0^{\infty} H_{[\,]}^{1/2}(w, \mathcal{G}(\delta), \|\cdot\|_2) dw < \infty,$$

which is a sufficient condition for $\mathcal{G}(\delta)$ to be $\mathbb{P}$-Donsker (see Van Der Vaart and Wellner (1996) p. 129). □

Corollary B.2.2. *For $\mathcal{W}_n \equiv \{v \in \mathcal{V}_n : \|v\| = 1\}$, $\epsilon_n^* = o(1)$, and $\sqrt{n}\epsilon_n^* \varepsilon_n = o(1)$, the following conditions are satisfied:*

(i) $\sup_{P_X \in \mathcal{B}_n v_1, v_2 \in \mathcal{W}_n} \mu_n\{r(P_X, Y)[v_1, v_2]\} = O_p(\epsilon_n^*)$.

(ii) $\sup_{P_X \in \mathcal{B}_n v \in \mathcal{W}_n} |\frac{\partial \Gamma(P_X)}{\partial P_X}[v] - \frac{\partial \Gamma(P_X^0)}{\partial P_X}[v]| = O(\epsilon_n^*)$

(iii) $\sup_{P_X \in \mathcal{B}_n v \in \mathcal{W}_n} \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big[\ell'(P_X, Y_i)[v] - \ell'(P_X^0, Y_i)[v] - \mathbb{E}\{\ell'(P_X, Y_i)[v]\}\Big] = o(1)$

(iv) $\sup_{P_X \in \mathcal{B}_n v \in \mathcal{W}_n} \mathbb{E}\Big\{\ell'(P_X, Y_i)[v] - \ell'(P_X^0, Y_i)[v] - r(P_X^0, Y_i)[v, P_X - P_X^0]\Big\} = O(\epsilon_n^* \varepsilon_n)$

(v) $\sup_{v\in\mathcal{W}_n}|\frac{1}{\sqrt{n}}\sum_{i=1}^n \ell'(P_X^0, Y_i)[v]| = O_p(1)$.

Proof. (ii) holds trivially for the weighted integral functionals in Eq. (2.5.1). Direct calculation of Gateaux derivatives shows $\ell'(P_X, Y_i)[v_1] = [P - P_X(\mathbf{Z})]W(\mathbf{Z})v_1(\mathbf{Z})$ and $r(P_X, Y)[v_1, v_2] = v_1(\mathbf{Z})v_2(\mathbf{Z})W(\mathbf{Z})$, where $v_j(\mathbf{Z}) = [P_X^j(\mathbf{Z}) - P_X^{0,n}(\mathbf{Z})]$ for some $P_X^j \in \mathcal{P}_X^{K_n}$, which only involve objects from the Donsker class in Lemma B.2.1. The remaining results hold by stochastic equicontinuity and/or application of ULLN and UCLT. □

B.3 Further Simulations and Examples

B.3.1 Simulation Parameters

Section 2.6 in the main paper simulates a double-jump process whose parameter values correspond to those from Andersen et al. (2012). For completeness, the parameter values used in my simulations are given in Table B.1.

The headers of the table reflect the fact that the first three columns represent models that can be viewed as special cases of the stochastic volatility price-jump and volatility-jump model shown in the fourth column.

B.3.2 Bayesian Information Criterion Selection

I examine how the BIC selection relates to the familiar parametric data generating processes in Eq. (2.6.1). To this end, I simulated a dense panel of option prices for eight different maturities from the continuous time process in Eq. (2.6.1) for each of the parameter specifications in Table B.1. Then, I drew a random sample of 250 options to mimic features of the option prices of a given trading day (in this case January 5, 2005). Finally I performed the NLLS optimization in Eq. (2.2.13) for

Table B.1: Parameter values used in the simulation exercises in Section 2.3.

	Black-Scholes	Heston	SVJ	SVJJ
V_0	0.014	0.014	0.014	0.014
κ		4.032	4.032	4.032
$\overline{V}$		0.014	0.014	0.014
ρ		-0.460	-0.460	-0.460
v		0.200	0.200	0.200
λ			1.008	1.008
μ_J			-0.050	-0.050
σ_J			0.075	0.075
μ_v				0.100
ρ_J				-0.500

K_x ranging from 1 to 9, and K_τ ranging from 0 to 2. The largest model involved $(9+1)(2+1) = 30$ parameters.

Table B.2 records the squared coefficient values and the BIC expansion choices for each of the Black-Scholes, Heston, SVJ, and SVJJ models. Recall that the sieve coefficients are normalized to make their squares sum to one. Thus the squared coefficients represent the share of weight, or loading, onto individual expansion terms. Blank rows in the table reflect the fact that the BIC did not use expansions in x of that order. Similarly, blank columns illustrate that the minimized BIC did not select expansions in the τ dimension of that order. Thus, the fact that the BIC correctly chose 0 expansion terms in x and 0 in τ means that the BIC correctly chose the Black-Scholes model when faced with a Black-Scholes DGP. The next panel of Table B.2 then shows that as stochastic volatility is added to the DGP, more expansion terms ($K_x = 3$ and $K_\tau = 1$) are required to fit the newly generated option surface. This trend continues in the last two panels of Table B.2: the more complexity is added to the DGP, the more expansion terms are required to provide adequate fit.

Table B.2: BIC Selection Given Affine Jump-Diffusion DGP.

	Black-Scholes			Heston			SVJ			SVJJ		
$K_x \backslash K_\tau$	τ^0	τ^1	τ^2	τ^0	τ^1	τ^2	τ^0	τ^1	τ^2	τ^0	τ^1	τ^2
0	1.0			0.8	0.2		0.4	0.1	0.5	0.0	0.8	0.0
1				0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0
2				0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0
3				0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0
4							0.0	0.0	0.0	0.0	0.0	0.0
5							0.0	0.0	0.0	0.0	0.0	0.0
6							0.0	0.0	0.0	0.0	0.0	0.0
7										0.0	0.0	0.0
8												
9												

B.3.3 30-day Measures

The fit of the sieve option pricer extends beyond observed maturities. Because the estimation problem in Eq. (2.2.13) uses the entire option panel in a single step (i.e. options of all available maturities), and because the sieve expansion of the state-price density in (2.3.1) is bivariate in the return-maturity space, it is possible to simply evaluate $\widehat{P}_X^{K_n}(\kappa, \tau, S_0, r, q)$ at arbitrary maturities $\tau = \tau^*$.

The ability to evaluate $\widehat{P}_X^{K_n}(\kappa, \tau, S_0, r, q)$ at arbitrary maturities has applications in the construction of time series of balanced option panels and term structures. For instance, in the application to the variance risk premium in the next section, I consider the expectation hypothesis regression, which requires weekly observations on the $VIX(\tau)$ for, say, $\tau = 30$ days to maturity. However, exchange traded options like the S&P 500 Index Options traded on the Chicago Board Options Exchange

(CBOE) have a fixed maturity date on the third Friday of each month. Hence, from week to week, the observed options' time-to-maturity shortens by one week, which complicates the construction of a balanced time series. Moreover, as short-maturity options expire, new long maturity options are added, causing a type of "cycling" in the time series of observed maturities. Such deterministic cycling can induce non-stationarities in the constructed option-implied time series [see Pan (2002)]. To address this issue, the CBOE's VIX is computed by interpolating two VIX's that straddle the 30-day maturity. However, if the VIX is designed to approximate the forward-looking risk-neutral expectation of realized variance, $\mathbb{E}_t^{\mathbb{Q}}[RV_{t,t+\tau}]$, it is unclear how such an interpolation maps to the true $VIX(30)$ if 30-day time-to-maturity options were actually observed.

Figure B.1 shows the result of evaluating the estimated sieve pricer on a maturity that was not available for estimation – in this case 30 days – and comparing it with true values that are known inside the simulation. That is, I simulated a dense set of put option prices (i.e. 600 per maturity times 9 maturities, totalling 5,400 true option prices) from the Heston submodel of Eq. (2.6.1) with days-to-maturity 17, 30, 45, 73, 164, 255, 346, 528, and 710. I then drew a random sample of 250 observations from this dense set of true prices but omitted the options maturing at 30 days. I then perturbed this sample with random errors to generate microstructure noise as calibrated from actual data. The sieve estimator was then estimated on the 250 noisy options.

In this setup, even though the sieve estimator was not permitted to "see" the information contained in the 30-day maturity options, it was nonetheless able to

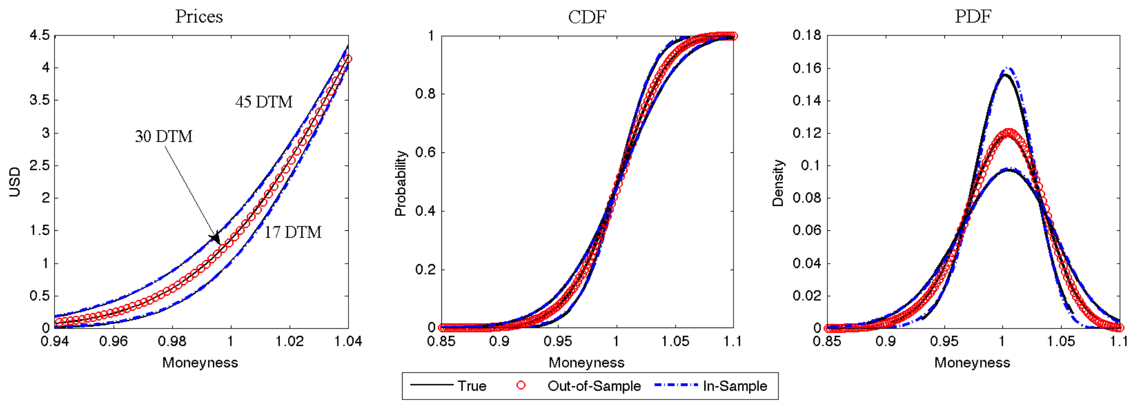

FIGURE B.1: Projecting the Sieve onto Arbitrary Maturities. A dense set of true option prices is simulated from the Heston submodel Eq. (2.6.1) with days-to-maturity 17, 30, 45, 73, 164, 255, 346, 528, 710. A subset of 250 option prices is drawn from this dense set, but omitting the 30-day maturity, and is perturbed with zero-mean measurement error. The sieve least squares problem in Eq. (2.2.13) is solved with BIC-selected $K_x = 3$ and $K_\tau = 1$. Eqs. (2.3.2), (2.6.2), and (2.3.1) are then evaluated at the estimated coefficient matrix $\widehat{B}$ for $\tau = 17$, 30, and 45. The 30-day horizon (circles) is out-of-sample, since the estimation omitted data at the 30-day horizon.

accurately predict what those option values were, given information on options at other maturities. This is seen by comparing the circles (corresponding to the sieve) with the solid line (true prices) in the left panel of Figure B.1. Moreover, the sieve estimator does remarkably well in estimating the 30-day risk-neutral CDF (center panel) and the 30-day risk-neutral PDF. It is useful to emphasize that all quantities in the figure are available in closed-form via Eqs. (2.3.2), (2.6.2), and (2.3.1).

B.4 The $\mathbb{P}$-Measure: Estimating $\mathbb{E}_t^{\mathbb{P}}[TV_t(\tau)]$

The variance risk premium requires the forecast of continuous variation, that is, $\mathbb{E}_t^{\mathbb{P}}[TV_t(\tau)]$, which I obtain as follows. The data used for estimating $\mathbb{E}_t^{\mathbb{P}}[TV_t(\tau)]$ are last-tick-sampled 5-minute S&P 500 futures prices from the TAQ database. The 5-minute sampling frequency is chosen to mitigate the well-known effects arising from market microstructure noise in ultra high-frequency realizations of the continuous-time process in Eq. (2.7.1).

The advantage of using high-frequency observations for estimating quadratic variation and its components has been well-established in the last decade [see, for example, Andersen et al. (2003)]. Reliable model-free estimators for the continuous variation in Eq. (2.7.3) have since been developed, including the bipower variation estimator of Barndorff-Nielsen and Shephard (2004), the threshold estimator by Mancini (2009), and the threshold multipower variation estimator by Corsi et al. (2010). For brevity I present results only for Mancini's estimator, although conclusions are not materially affected by the particular choice of continuous variation estimator. In particular, Mancini (2009) shows that

$$\widehat{TV}_t(\tau) = \sum_{i=1}^{[\tau/\Delta_n]} |\Delta_i X|^2 1_{\{|\Delta_i X|^2 \leqslant \upsilon_n\}} \quad \xrightarrow{p} TV_t(\tau), \tag{B.4.1}$$

where n is the number of 5-minute observations from t to $t+\tau$, Δ_n is the 5-minute sampling interval, $\Delta_i X = \log(F_{i\Delta_n}) - \log(F_{(i-1)\Delta_n})$, and υ_n is a thresholding sequence that can depend on a local estimate of σ_{s-} [see Jacod and Protter (2012) p. 248].

To compute $\mathbb{E}_t^{\mathbb{P}}[TV_t(\tau)]$, however, only time-$t$ information may be used. Hence, a forecast of $TV_t(\tau)$ is required. Let TV_t denote the intraday continuous variation for one trading day. It is well-known that the realized volatility measures are persistent and exhibit long-memory properties. In particular, Andersen et al. (2003) show that the realized volatilities are well-described as integrated order d processes, i.e. $I(d)$ for $d \in (0, \frac{1}{2})$. Thus, to obtain forecasts of TV_t, it is reasonable to parsimoniously model TV_t as a fractionally differenced ARMA process, ARFIMA(p, d, q). That is,

$$(1-L)^d \phi(L)(y_t - \mu) = \theta(L)\varepsilon_t \tag{B.4.2}$$

where $y_t = log(TV_t^{1/2})$, $\phi(L)$ and $\theta(L)$ are the standard AR and MA lag polynomials, and $(1-L)^d$ is the fractional differencing filter. A plot of autocorrelations (not shown for brevity) of TV_t and $(1-L)^d TV_t$ reveals that setting $d = 0.401$ as in Andersen et al. (2003) suffices to isolate the low-frequency movements in the TV_t time series, which is a key requirement for obtaining long-run forecasts. To obtain forecasts of $TV_t(\tau)$, one forecasts y_{t+h} for $h = 1, \ldots, \tau$ days out using Eq. (B.4.2) and then inverts for $TV_t(h)$. Then $TV_t(\tau)$ is taken as the annualized sum of one day forecasts between t and $t + \tau$.[1]

[1] An analogous forecast using an HAR$(1, 5, 22)$ process developed in Corsi (2009) was also considered and yields similar results.

Appendix C

Proofs for Inference on Option Pricing Models under Partial Identification

We prove the main theorems in Appendix C.1 based on some technical lemmas which in turn are proved in Appendix C.2. Throughout the proof, we use K to denote a generic constant which may change from line to line.

C.1 Proofs of main theorems

C.1.1 Proof of Theorem 3.1.1

The proof requires the following technical lemma, the proof of which is in Appendix C.2.

Lemma C.1.1. *Suppose Assumption A holds for some* $k \geqslant 2$*. Let* $\varpi \in (0, 1/2)$ *and* $q \in [1, k/2]$*. Then we can decompose* $\widehat{V}_{n,t} - V_t = \zeta_{n,t} + \zeta'_{n,t}$ *such that* $\zeta_{n,t}$ *and* $\zeta'_{n,t}$ *are*

$\mathcal{F}_t$-measurable and

$$\mathbb{E}\left|\zeta_{n,t}\right|^q \leqslant K\big(\Delta_n^{1-q(1-2\varpi)} + \Delta_n^{1-q/k-q(1/2-\varpi)} + \Delta_n^{(k-2q)(1/2-\varpi)}$$

$$+\Delta_n^{(q/2)\wedge(1-q/k)}\big), \tag{C.1.1}$$

$$\mathbb{E}\left[\zeta'_{n,t}|\mathcal{F}_{t-k_n\Delta_n}\right] = 0, \quad \mathbb{E}\left|\zeta'_{n,t}\right|^q \leqslant K\Delta_n^{q/4}. \tag{C.1.2}$$

If, in addition, $k \geqslant 2/(1-2\varpi)$, then

$$\mathbb{E}\left|\zeta_{n,t}\right|^q \leqslant K\Delta_n^{(1-q(1-2\varpi))\wedge(q/2)}. \tag{C.1.3}$$

Proof of Theorem 3.1.1. Step 1. We prove part (a) in this step. By arguing compoment by component, we can assume that $m(\cdot)$ is 1-dimensional without loss of generality. Moreover, we only prove the case for the ask price, so

$$m(X_t, v, Q_{i,t}, Z_{i,t}, Z_t;\theta) = \left(A_{i,t} - f\left(X_t, v, Z_{i,t};\theta\right)\right) G\left(X_t, v, Z_{i,t}, Z_t\right); \tag{C.1.4}$$

the other cases involving $B_{i,t}$ or $p_{i,t}$ follow the same argument with only notational changes. We denote

$$f_{i,t} = f\left(X_t, V_t, Z_{i,t};\theta\right), \quad \widehat{f}_{i,t} = f\left(X_t, \widehat{V}_{n,t}, Z_{i,t};\theta\right), \quad f_{V,i,t} = f_V\left(X_t, V_t, Z_{i,t};\theta\right),$$
$$G_{i,t} = G\left(X_t, V_t, Z_{i,t}, Z_t\right), \quad \widehat{G}_{i,t} = G\left(X_t, \widehat{V}_{n,t}, Z_{i,t}, Z_t\right). \tag{C.1.5}$$

Then we can rewrite

$$\overline{m}^*_{n,T}(\theta) = T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N_t}\left(A_{i,t} - f_{i,t}\right)G_{i,t}, \quad \overline{m}_{n,T}(\theta) = T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N_t}(A_{i,t} - \widehat{f}_{i,t})\widehat{G}_{i,t}.$$

We shall use the following estimates repeatedly. Since $k \geqslant 2/(1-2\varpi)$, by Lemma C.1.1, for each $q \in [1, k/2]$,

$$\begin{aligned} \left(\mathbb{E}|\widehat{V}_{n,t} - V_t|^q\right)^{1/q} &\leqslant K\Delta_n^{(1/q-(1-2\varpi))\wedge(1/2)} + K\Delta_n^{1/4} \\ \left(\mathbb{E}|\widehat{V}_{n,t} - V_t|^{2q}\right)^{1/q} &\leqslant K\Delta_n^{1/q-2(1-2\varpi)} + K\Delta_n^{1/2}. \end{aligned} \tag{C.1.6}$$

Recalling the decomposition $\widehat{V}_{n,t} - V_t = \zeta_{n,t} + \zeta'_{n,t}$ as described in Lemma C.1.1, we consider the following decomposition

$$T^{1/2}\left(\overline{m}^*_{n,T}(\theta) - \overline{m}_{n,T}(\theta)\right) = \sum_{j=1}^{4} R^{(j)}_{n,T}, \tag{C.1.7}$$

where

$$\begin{aligned} R^{(1)}_{n,T} &= T^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{N_t}(\widehat{f}_{i,t} - f_{i,t} - f_{V,i,t}(\widehat{V}_{n,t} - V_t))G_{i,t}, \\ R^{(2)}_{n,T} &= T^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{N_t}\left(f_{V,i,t}G_{i,t} - f_{V,i,t-k_n\Delta_n}G_{i,t-k_n\Delta_n}\right)(\widehat{V}_{n,t} - V_t), \\ R^{(3)}_{n,T} &= T^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{N_t} f_{V,i,t-k_n\Delta_n}G_{i,t-k_n\Delta_n}\zeta_{n,t}, \\ R^{(4)}_{n,T} &= T^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{N_t} f_{V,i,t-k_n\Delta_n}G_{i,t-k_n\Delta_n}\zeta'_{n,t}. \end{aligned}$$

We now provide estimates for $R^{(j)}_{n,T}$, $1 \leqslant j \leqslant 4$. First consider $R^{(1)}_{n,T}$. Observe that,

$$\begin{aligned}\mathbb{E}\left|\widehat{f}_{i,t}-f_{i,t}-f_{V,i,t}(\widehat{V}_{n,t}-V_t))\right| &\leqslant \mathbb{E}\left|\xi_{n,t}\left(\widehat{V}_{n,t}-V_t\right)^2\right| \\ &\leqslant K\left(\mathbb{E}\left|\xi_{n,t}\right|^{k_3}\right)^{1/k_3}\left(\mathbb{E}\left[\left|\widehat{V}_{n,t}-V_t\right|^{2k_3/(k_3-1)}\right]\right)^{1-1/k_3} \\ &\leqslant K\Delta_n^{1-1/k_3-2(1-2\varpi)}+K\Delta_n^{1/2} \\ &\leqslant K\Delta_n^{1/2},\end{aligned}$$

where the first inequality follows Assumption B, the second inequality is by Hölder's inequality, the third inequality follows (C.1.6) and the last inequality holds because $\varpi \in (3/8, 1/2)$ and $k_3 \geqslant 2/(8\varpi - 3)$. Since N_t and $G_{i,t}$ are bounded, we have

$$\mathbb{E}\left|R_{n,T}^{(1)}\right| \leqslant KT^{1/2}\Delta_n^{1/2}. \tag{C.1.8}$$

Next, consider $R_{n,T}^{(2)}$. By Assumption B and the Cauchy-Schwarz inequality,

$$\begin{aligned}&\mathbb{E}\left|\left(f_{V,i,t}G_{i,t}-f_{V,i,t-k_n\Delta_n}G_{i,t-k_n\Delta_n}\right)\left(\widehat{V}_{n,t}-V_t\right)\right| \\ &\qquad\leqslant\left(\mathbb{E}\left[\left(f_{V,i,t}G_{i,t}-f_{V,i,t-k_n\Delta_n}G_{i,t-k_n\Delta_n}\right)^2\right]\right)^{1/2}\left(\mathbb{E}\left[\left(\widehat{V}_{n,t}-V_t\right)^2\right]\right)^{1/2}\end{aligned} \tag{C.1.9}$$

By the assumption of part (a), $G(x, v, z, \tilde{z})$ does not depend on (x, v). Moreover, recall that $Z_{i,t}$ and Z_t are constant within each day. Hence, we have $G_{i,t} = G_{i,t-k_n\Delta_n}$. Then by Assumption B4, $\mathbb{E}\left|f_{V,i,t}G_{i,t} - f_{V,i,t-k_n\Delta_n}G_{i,t-k_n\Delta_n}\right|^2 \leqslant K\Delta_n^{1/2}$. Moreover, by (C.1.6) and $\varpi > 3/8$,

$$\left(\mathbb{E}\left[\left(\widehat{V}_{n,t}-V_t\right)^2\right]\right)^{1/2} \leqslant K\Delta_n^{1/2-(1-2\varpi)}+K\Delta_n^{1/4} \leqslant K\Delta_n^{1/4}. \tag{C.1.10}$$

Therefore, the majorant side of (C.1.9) can be further bounded by $K\Delta_n^{1/2}$; thus,

$$\mathbb{E}\left|R_{n,T}^{(2)}\right| \leqslant KT^{1/2}\Delta_n^{1/2}. \tag{C.1.11}$$

We now turn to $R_{n,T}^{(3)}$. We observe

$$\begin{aligned}
\mathbb{E}\left|f_{V,i,t-k_n\Delta_n}G_{i,t-k_n\Delta_n}\zeta_{n,t}\right| &\leqslant K\left(\mathbb{E}\left[(f_{V,i,t-k_n\Delta_n})^4\right]\right)^{1/4}\left(\mathbb{E}\left[|\zeta_{n,t}|^{4/3}\right]\right)^{3/4} \\
&= K\Delta_n^{(2\varpi-1/4)\wedge(1/2)} \\
&\leqslant K\Delta_n^{1/2},
\end{aligned}$$

where the first inequality is by Hölder's inequality, the second inequality follows Lemma C.1.1 and the last inequality holds because $\varpi > 3/8$. Hence,

$$\mathbb{E}\left|R_{n,T}^{(3)}\right| \leqslant K\Delta_n^{1/2}. \tag{C.1.12}$$

Now, consider $R_{n,T}^{(4)}$. Observe that, by Lemma C.1.1, $\mathbb{E}\left[\zeta'_{n,t}|\mathcal{F}_{t-k_n\Delta_n}\right] = 0$. By assumption B1, N_t is $\mathcal{F}_{\underline{t}}$-measurable. It is then easy to see that the quantity $\sum_{t=1}^{N_t} f_{V,i,t-k_n\Delta_n}G_{i,t-k_n\Delta_n}\zeta'_{n,t}$ forms a martingale difference sequence with respect to the filtration $(\mathcal{F}_t)_{t=1,\ldots,T}$. By the Cauchy-Schwarz inequality, the boundedness of N_t and $G_{i,t-k_n\Delta_n}$, and Lemma C.1.1 with $q = 4$, we derive

$$\begin{aligned}
\mathbb{E}\left|R_{n,T}^{(4)}\right|^2 &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left(\sum_{i=1}^{N_t} f_{V,i,t-k_n\Delta_n}G_{i,t-k_n\Delta_n}\zeta'_{n,t}\right)^2\right] \\
&\leqslant K\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\sum_{i=1}^{N_t}\left(f_{V,i,t-k_n\Delta_n}\zeta'_{n,t}\right)^2\right] \\
&\leqslant K\Delta_n^{1/2}.
\end{aligned} \tag{C.1.13}$$

Finally, combining (C.1.7), (C.1.8), (C.1.11), (C.1.12) and (C.1.13), we readily derive the assertion of part (a).

Step 2. In this step, we prove part (b). As argued in step 1, we consider the 1-dimensional case and suppose (C.1.4) without loss of generality. We also maintain the notations in (C.1.5). Observe that

$$\begin{aligned}&\mathbb{E}\left|(A_{i,t}-f_{i,t})\,G_{i,t}-\left(A_{i,t}-\widehat{f}_{i,t}\right)\widehat{G}_{i,t}\right|\\&\qquad\leqslant \mathbb{E}\left|A_{i,t}\left(\widehat{G}_{i,t}-G_{i,t}\right)\right|+\mathbb{E}\left|f_{i,t}\left(\widehat{G}_{i,t}-G_{i,t}\right)\right|+K\mathbb{E}\left|\left(\widehat{f}_{i,t}-f_{i,t}\right)\right|.\end{aligned} \tag{C.1.14}$$

Since $G(x,v,z,\tilde{z})$ is continuously differentiable in v with bounded derivative, we have

$$|G_{i,t}-\widehat{G}_{i,t}|\leqslant K|\widehat{V}_{n,t}-V_t|. \tag{C.1.15}$$

Since $\mathbb{E}|A_{i,t}|^2+\mathbb{E}|f_{i,t}|^2\leqslant K$ by Assumption B3, we apply the Cauchy-Schwarz inequality and (C.1.10) to show that the first two terms on the majorant side of (C.1.14) can be further bounded by $K\Delta_n^{1/4}$. Similarly, since $|\widehat{f}_{i,t}-f_{i,t}|\leqslant \chi_{n,t}|\widehat{V}_{n,t}-V_t|$ and $\mathbb{E}\,|\chi_{n,t}|^2\leqslant K$, we can bound the third term on the majorant side of (C.1.14) by $K\Delta_n^{1/4}$. Hence, we derive

$$\mathbb{E}\left|(A_{i,t}-f_{i,t})\,G_{i,t}-\left(A_{i,t}-\widehat{f}_{i,t}\right)\widehat{G}_{i,t}\right|\leqslant K\Delta_n^{1/4}.$$

The assertion of part (b) then readily follows.

Step 3. We now show part (c). We first consider the one-dimensional case. Without loss, we consider the case for the ask price, so (C.1.4) is in force. We observe

$$\mathbb{E}\left|\left(A_{i,t}-f_{i,t}\right)G_{i,t}-\left(A_{i,t}-\widehat{f}_{i,t}\right)\widehat{G}_{i,t}\right|^{2}$$
$$\leqslant K\mathbb{E}\left|A_{i,t}\left(G_{i,t}-\widehat{G}_{i,t}\right)\right|^{2}+K\mathbb{E}\left|f_{i,t}\left(\widehat{G}_{i,t}-G_{i,t}\right)\right|^{2}+K\mathbb{E}\left|\widehat{G}_{i,t}\left(\widehat{f}_{i,t}-f_{i,t}\right)\right|^{2}. \tag{C.1.16}$$

By Hölder's inequality and (C.1.15),

$$\begin{aligned}\mathbb{E}\left|A_{i,t}\left(G_{i,t}-\widehat{G}_{i,t}\right)\right|^{2} &\leqslant \left(\mathbb{E}\left[\left(A_{i,t}\right)^{2k'}\right]\right)^{1/k'}\left(\mathbb{E}\left[\left|\widehat{V}_{n,t}-V_{t}\right|^{2k'/(k'-1)}\right]\right)^{1-1/k'}\\ &\leqslant K\Delta_{n}^{1-1/k'-2(1-2\varpi)}+K\Delta_{n}^{1/2}\\ &\leqslant K\Delta_{n}^{1/2},\end{aligned}$$

where the second inequality follows (C.1.6) and the third inequality holds because $\varpi\in(3/8,1/2)$ and $k'\geqslant 2/\left(8\varpi-3\right)$. Similarly, we can also bound the last two terms on the right-hand side of (C.1.16) by $K\Delta_{n}^{1/2}$. Hence, by Cauchy-Schwarz,

$$\mathbb{E}\left|\tilde{m}_{t}\left(V_{t};\theta\right)-\tilde{m}_{t}\left(\widehat{V}_{n,t};\theta\right)\right|^{2}\leqslant K\Delta_{n}^{1/2}, \tag{C.1.17}$$

which further implies that

$$\mathbb{E}\left|\bar{m}_{n,T}^{*}\left(\theta\right)-\bar{m}_{n,T}\left(\theta\right)\right|^{2}\leqslant K\Delta_{n}^{1/2}. \tag{C.1.18}$$

Since $\tilde{m}_{t}\left(V_{t};\theta\right)$ and $\bar{m}_{n,T}^{*}\left(\theta\right)$ are L^{2}-bounded by Assumption B3, it is then easy to see that

$$\mathbb{E}\left|\tilde{m}_{t}\left(\widehat{V}_{n,t};\theta\right)\right|^{2}\leqslant K,\quad \mathbb{E}\left|\bar{m}_{n,T}\left(\theta\right)\right|^{2}\leqslant K. \tag{C.1.19}$$

Let $\widehat{\Gamma}_{l,T}^{*}\left(\theta\right)$ be defined as $\widehat{\Gamma}_{l,n,T}\left(\theta\right)$ in (3.1.17) but with V_{t} in place of $\widehat{V}_{n,t}$. Observe that, by the triangle inequality,

$$
\begin{aligned}
&\left|\widehat{\Gamma}_{l,n,T}(\theta)-\widehat{\Gamma}^{*}_{l,T}(\theta)\right| \\
\leqslant\ & \frac{1}{T}\sum_{1\leqslant t,t+l\leqslant T}\left|\left(\tilde{m}_t\left(\widehat{V}_{n,t};\theta\right)-\tilde{m}_t\left(V_t;\theta\right)\right)\left(\tilde{m}_{t+l}\left(\widehat{V}_{n,t+l};\theta\right)-\bar{m}_{n,T}(\theta)\right)\right| \\
&+\frac{1}{T}\sum_{1\leqslant t,t+l\leqslant T}\left|\left(\bar{m}_{n,T}(\theta)-\bar{m}^{*}_{T}(\theta)\right)\left(\tilde{m}_{t+l}\left(\widehat{V}_{n,t+l};\theta\right)-\bar{m}_{n,T}(\theta)\right)\right| \\
&+\frac{1}{T}\sum_{1\leqslant t,t+l\leqslant T}\left|\left(\tilde{m}_t\left(V_t;\theta\right)-\bar{m}^{*}_{T}(\theta)\right)\left(\tilde{m}_{t+l}\left(\widehat{V}_{n,t+l};\theta\right)-\tilde{m}_{t+l}\left(V_{t+l};\theta\right)\right)\right| \\
&+\frac{1}{T}\sum_{1\leqslant t,t+l\leqslant T}\left|\left(\tilde{m}_t\left(V_t;\theta\right)-\bar{m}^{*}_{T}(\theta)\right)\left(\bar{m}_{n,T}(\theta)-\bar{m}^{*}_{T}(\theta)\right)\right|.
\end{aligned}
$$

By the Cauchy-Schwarz inequality and (C.1.17), (C.1.18), (C.1.19), we derive

$$\mathbb{E}\left|\widehat{\Gamma}_{l,n,T}(\theta)-\widehat{\Gamma}^{*}_{l,T}(\theta)\right|\leqslant K\Delta_n^{1/4}.$$

Since $L_{n,T}=o(\Delta_n^{1/4})$ by assumption, we derive $\mathbb{E}|\widehat{\Sigma}_{n,T}(\theta)-\widehat{\Sigma}^{*}_{T}(\theta)|\leqslant KL_{n,T}\Delta_n^{1/4}=o(1)$. Hence, $\widehat{\Sigma}_{n,T}(\theta)-\widehat{\Sigma}^{*}_{T}(\theta)=o_p(1)$.

For the multivariate case, we can use exactly the same argument to show that each element of $\widehat{\Sigma}_{n,T}(\theta)-\widehat{\Sigma}^{*}_{T}(\theta)$ is $o_p(1)$, as asserted. □

C.1.2 Proof of Theorem 3.1.2

Proof. By Theorem 3.1.1 and Assumption D, we have

$$T^{1/2}\left(\bar{m}_{n,T}(\theta_0)-\bar{m}(\theta_0)\right)\xrightarrow{d}\mathcal{N}\left(0_K,\Sigma(\theta_0)\right),\quad \widehat{\Sigma}_{n,T}(\theta_0)\xrightarrow{\mathbb{P}}\Sigma(\theta_0). \tag{C.1.20}$$

The proof for (3.1.23) and (3.1.24) then follows Theorem 1(a) and 1(b) in the supplement to Andrews and Soares (2010) by specializing their proof to a fixed sequence of data generating process. To make the proof self-contained, we prove a direct proof below.

Let $Z^* = \Omega(\theta_0)^{1/2} Y^*$ with the same Y^* in the definition of $\phi_{n,T}(\theta)$ (recall (3.1.19)). By the continuous mapping theorem,

$$S_{n,T}(\theta_0) \xrightarrow{d} \sum_{j=1}^{2K_I} \left[Z_j^*\right]_-^2 1_{\{\overline{m}_j(\theta_0)=0\}} + \sum_{j=2K_I+1}^{K} \left(Z_j^*\right)^2 \tag{C.1.21}$$

Moreover, for any (fixed) realization of Y^* (hence Z^*), by the continuous mapping theorem (recall the notation $\phi_{n,T}(\cdot)$ from (3.1.19)),

$$\phi_{n,T}(\theta_0) \xrightarrow{\mathbb{P}} \sum_{j=1}^{2K_I} \left[Z_j^*\right]_-^2 1_{\{\overline{m}_j(\theta_0)=0\}} + \sum_{j=2K_I+1}^{K} \left(Z_j^*\right)^2. \tag{C.1.22}$$

Below, let $\mathcal{L}$ denote the distribution function of the variable on the right-hand side of (C.1.21) and denote by $c(\theta_0, 1-\alpha)$ the $(1-\alpha)$-quantile of $\mathcal{L}$.

We first prove (3.1.24), so condition (iv) is in force. Since $c(\theta_0, 1-\alpha)$ is a continuity point of $\mathcal{L}$, we have $c(\theta_0, 1-\alpha) > 0$. Indeed, if $c(\theta_0, 1-\alpha) = 0$ were true, the continuity of $\mathcal{L}$ at $c(\theta_0, 1-\alpha)$ would imply $\mathcal{L}(c(\theta_0, 1-\alpha)) = \mathcal{L}(0) = 0$; this would be a contradiction to the fact that $c(\theta_0, 1-\alpha)$ is $(1-\alpha)$-quantile of $\mathcal{L}$ for some $\alpha \in (0,1)$. Furthermore, we observe that $\mathcal{L}$ is strictly increasing at $c(\theta_0, 1-\alpha)$ and thus $1-\alpha$ is a continuity point of the quantile function associated with $\mathcal{L}$. Combining this with (C.1.22), we derive

$$c_{n,T}\left(\theta_0,1-\alpha\right)\xrightarrow{\mathbb{P}}c\left(\theta_0,1-\alpha\right). \tag{C.1.23}$$

We then finish the proof by observing

$$\mathbb{P}\left(\theta_0\in CS_{n,T}\left(1-\alpha\right)\right)=\mathbb{P}\left(S_{n,T}\left(\theta_0\right)\leqslant c_{n,T}\left(\theta_0,1-\alpha\right)\right)\to 1-\alpha,$$

where the convergence follows (C.1.21) and (C.1.23), as well as the continuity of $\mathcal{L}\left(\cdot\right)$ at $c\left(\theta_0,1-\alpha\right)$.

We now turn to (3.1.23). By (3.1.24), it remains to consider the case in which $\mathcal{L}$ is discontinuous at $c\left(\theta_0,1-\alpha\right)$. Note that this is possible only if $c\left(\theta_0,1-\alpha\right)=0$ and $k_E=0$. Hence,

$$\begin{aligned}
&\liminf_{\Delta_n\to0,T\to\infty}\mathbb{P}\left(\theta_0\in CS_{n,T}\left(1-\alpha\right)\right)\\
=&\liminf_{\Delta_n\to0,T\to\infty}\mathbb{P}\left(S_{n,T}\left(\theta_0\right)\leqslant c_{n,T}\left(\theta_0,1-\alpha\right)\right)\\
\geqslant&\liminf_{\Delta_n\to0,T\to\infty}\mathbb{P}\left(S_{n,T}\left(\theta_0\right)=0\right)\\
=&\liminf_{\Delta_n\to0,T\to\infty}\mathbb{P}\left(\widehat{D}_{j,n,T}^{-1/2}\left(\theta_0\right)T^{1/2}\overline{m}_{j,n,T}\left(\theta_0\right)\geqslant0\text{ for }1\leqslant j\leqslant 2k_I\right)\\
=&\mathcal{L}\left(0\right)\geqslant1-\alpha.
\end{aligned}$$

This finishes the proof of (3.1.23). □

C.1.3 Proof of Theorem 3.1.4

The proof of Theorem 3.1.4 relies on the following technical lemma, whose proof is given in Appendix C.2.

Lemma C.1.2. *Fix some constant $\varpi \in (5/12, 1/2)$. Suppose (i) Assumption A holds for some $k \geqslant \max\{8, 2/(1-2\varpi)\}$ and V_t follows (3.2.1); (ii) Assumption B holds for some $k' > 2/(12\varpi - 5)$. Then*

(a) $\left|\mathbb{E}\left[\widehat{m}_{n,t}^{+}(\theta_0)\right]\right| \leqslant K\Delta_n^{3\varpi-5/4}$.

(b) $\mathbb{E}\left|\widehat{m}_{n,t}^{+}(\theta_0)\right|^{4/(9-12\varpi)} \leqslant K$.

(c) For each s,t, the sequences $(m_s^*(\theta_0)\widehat{m}_{n,t}^{+}(\theta_0))_{n\geqslant 1}$ and $(\widehat{m}_{n,t}^{+}(\theta_0))_{n\geqslant 1}$ are uniformly integrable.

Proof of Theorem 3.1.4. Step 1. In view of Assumption E1, we can and will state limiting results under $T \to \infty$ and $n \to \infty$ interchangeably. We consider the $k \times 1$ vector

$$\Xi_{n,T} = Cov\left(\sum_{t=1}^{T} T^{-1/2} m_t^*(\theta_0), \sum_{t=1}^{T} T^{-1/2}\widehat{m}_{n,t}^{+}(\theta_0)\right).$$

In this step, we show that as $T \to \infty$,

$$\Xi_{n,T} \to 0. \tag{C.1.24}$$

First, observe that

$$\Xi_{n,T} = \sum_{l=-(T-1)}^{T-1} \frac{1}{T} \sum_{1\leqslant t,t+l\leqslant T} Cov\left(m_t^*(\theta_0), \widehat{m}_{n,t+l}^{+}(\theta_0)\right). \tag{C.1.25}$$

Note that for each t, $\widehat{IV}_{n,t} - IV_t = O_p(\Delta_n^{1/2})$. Combining this with Lemma 3.1.3, we deduce that for each t, $\widehat{m}_{n,t}^{+}(\theta_0)$ converges stably in law to some variable ζ_t which, conditionally on $\mathcal{F}$, is centered Gaussian. Hence, by the property of stable convergence, we have for any fixed t and s,

$$\left(m_s^*(\theta_0), \widehat{m}_{n,t}^+(\theta_0)\right) \xrightarrow{d} \left(m_s^*(\theta_0), \zeta_t\right).$$

By Lemma C.1.2(c), $(m_s^*(\theta_0)\widehat{m}_{n,t}(\theta_0))_{n\geqslant 1}$ and $(\widehat{m}_{n,t}(\theta_0))_{n\geqslant 1}$ are both uniformly integrable. Hence, for fixed s and t, as $n \to \infty$,

$$Cov\left(m_s^*(\theta_0), \widehat{m}_{n,t}^+(\theta_0)\right) \to Cov\left(m_s^*(\theta_0), \zeta_t\right) = 0. \tag{C.1.26}$$

Let $\|\cdot\|_p$ denote the L_p-norm and $r = 4/(9 - 12\varpi)$. By the mixing inequality,

$$\left|Cov\left(m_t^*(\theta_0), \widehat{m}_{n,t+l}^+(\theta_0)\right)\right| \leqslant K\alpha_{mix,|l|}^{1-1/(2k')-1/r} ||m_t^*(\theta_0)||_{2k'} ||\widehat{m}_{n,t+l}^+(\theta_0)||_r.$$

By Assumptions B3, E5 and Lemma C.1.2(b), the majorant side of the above display is summable over $\{l : |l| < \infty\}$. By dominated convergence, (C.1.25) and (C.1.26) readily imply (C.1.24).

Step 2. Recall $\Sigma(\theta_0)$ from Assumption D and let $\Omega(\theta_0)$ be the associated correlation matrix. We introduce some notations:

$$\widehat{\Sigma}_{\psi,n,T}(\theta_0) = \begin{pmatrix} \widehat{\Sigma}_{n,T}(\theta_0) & 0_k \\ 0_k^\top & \widehat{\Sigma}_{n,T}^+(\theta_0) \end{pmatrix}, \quad \widehat{D}_{\psi,n,T}(\theta_0) = Diag\left(\widehat{\Sigma}_{\psi,n,T}(\theta_0)\right),$$
$$\Omega_\psi(\theta_0) = \begin{pmatrix} \Omega(\theta_0) & 0_k \\ 0_k^\top & 1 \end{pmatrix}, \quad \bar{\psi} = \left(\mathbb{E}\left[m_t^*(\theta_0)\right]^\top, 0\right)^\top.$$

In this step, we show that

$$\widehat{D}_{\psi,n,T}^{-1/2}(\theta_0) \sum_{t=1}^{T} \left(\psi_{n,t}(\theta_0) - \bar{\psi}\right) \xrightarrow{d} \mathcal{N}\left(0_{k+1}, \Omega_\psi(\theta_0)\right). \tag{C.1.27}$$

By (C.1.24), Assumptions D3 (combined with Theorem 3.1.1(c)) and E3, we derive $\widehat{\Sigma}_{\psi,n,T}(\theta_0) - \Sigma_{\psi,n,T}(\theta_0) = o_p(1)$; then by continuous mapping, $\widehat{\Sigma}^{-1/2}_{\psi,n,T}(\theta_0) - \Sigma^{-1/2}_{\psi,n,T}(\theta_0) = o_p(1)$. In particular, $\widehat{\Sigma}^{-1/2}_{\psi,n,T}(\theta_0) = O_p(1)$ by Assumptions D2 and E4. By Theorem 3.1.1 (note that the conditions here are stronger than those in Theorem 3.1.1) and Lemma C.1.2(a), $\sum_{t=1}^{T}\left(\mathbb{E}\left[\psi_{n,t}(\theta_0)\right] - \bar{\psi}\right) = O(T^{1/2}\Delta_n^{3\varpi-5/4}) = o(1)$, where the second equality is due to Assumption E1 and our choice of c_1. Hence, by Assumption E2,

$$\widehat{\Sigma}^{-1/2}_{\psi,n,T}(\theta_0)\sum_{t=1}^{T}\left(\psi_{n,t}(\theta_0) - \bar{\psi}\right) \xrightarrow{d} \mathcal{N}(0, I_{k+1}). \tag{C.1.28}$$

It is also easy to see that

$$\begin{aligned}\widehat{D}^{-1/2}_{\psi,n,T}(\theta_0)\,\widehat{\Sigma}_{\psi,n,T}(\theta_0)\,\widehat{D}^{-1/2}_{\psi,n,T}(\theta_0) &= \begin{pmatrix} \widehat{D}^{-1/2}_{n,T}(\theta_0)\,\widehat{\Sigma}_{n,T}(\theta_0)\,\widehat{D}^{-1/2}_{n,T}(\theta_0) & 0 \\ 0 & 1 \end{pmatrix} \\ &\xrightarrow{\mathbb{P}} \Omega_{\psi}(\theta_0). \end{aligned} \tag{C.1.29}$$

Combining (C.1.28) and (C.1.29), we derive (C.1.27) as claimed.

Step 3. We now prove the assertion of the theorem. Let $Z^* = \Omega^{1/2}(\theta_0)\,Y^*$. By (C.1.27) and the continuous mapping theorem,

$$S'_{n,T}(\theta_0) \xrightarrow{d} \sum_{j=1}^{2k_I}\left\{\left[Z_j^*\right]_-^2 1_{\{\overline{m}_j(\theta_0)=0\}}\right\} + \sum_{j=2k_I+1}^{k}\left(Z_j^*\right)^2 + (Y^{**})^2.$$

Let $\mathcal{L}'$ denote the distribution function of the right-hand side of the above convergence. Observe (i) the $(1-\alpha)$-quantile of $\mathcal{L}'$ is strictly positive for any $\alpha \in (0,1)$, and (ii) $\mathcal{L}'$ is continuous at its $(1-\alpha)$-quantile. The proof then follows a similar argument as the proof of Theorem 3.1.2. (The only difference is that we do not need condition (v) in Theorem 3.1.2, because the continuity of the limiting distribution at its quantiles automatically holds in the presence of equality restrictions, as shown above.) □

C.2 Proofs of technical lemmas

In this appendix, we prove technical lemmas used in Appendix C.1. Throughout the appendix, we set

$$X'_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s, \quad X''_t = X_t - X'_t. \tag{C.2.1}$$

C.2.1 Proof of Lemma C.1.1

Step 1. We first introduce some notations and some preliminary estimates.Let $t\,(n,i) = t - k_n\Delta_n + i\Delta_n$, $i = 0,\ldots,k_n$. For any process Y, we denote $\Delta_i^{t,n}Y = Y_{t(n,i)} - Y_{t(n,i-1)}$. We then set $\widehat{V}'_{n,t} = (k_n\Delta_n)^{-1}\sum_{i=1}^{k_n}|\Delta_i^{t,n}X'|^2$ and note that $\widehat{V}_{n,t}$ can be written as $\widehat{V}_{n,t} = (k_n\Delta_n)^{-1}\sum_{i=1}^{k_n}|\Delta_i^{t,n}X|^2 1\{|\Delta_i^{t,n}X| \leqslant \alpha\Delta_n^{\varpi}\}$. We denote for $i = 1,\ldots,k_n$,

$$\begin{aligned}
\lambda_i^{t,n} &= \frac{1}{\Delta_n^{1/2}}\left(\int_{t(n,i-1)}^{t(n,i)} b_s ds + \int_{t(n,i-1)}^{t(n,i)} \left(\sigma_s - \sigma_{t(n,i-1)}\right) dW_s\right) \\
\beta_i^{t,n} &= \sigma_{t(n,i-1)}\Delta_i^{t,n}W/\Delta_n^{1/2}.
\end{aligned}$$

Note that $\Delta_i^{t,n}X'/\Delta_n^{1/2} = \lambda_i^{t,n} + \beta_i^{t,n}$.

Recalling $V_t = \sigma_t^2$ and (3.1.20), by Itô's formula, we can represent V_t as

$$V_t = V_0 + \int_0^t b_{V,s} ds + \int_0^t \sigma_{V,s} dW_s + \int_0^t \sigma'_{V,s} dW'_s + \int_0^t \int_{\mathbb{R}} \delta_V(s,z)\, (\underline{\mu} - \underline{\nu})(ds, dz),$$

where

$$\begin{aligned} & b_{V,s} = 2\sigma_s \tilde{b}_s + \tilde{\sigma}_s^2 + \tilde{\sigma}_s'^2 + \textstyle\int_{\mathbb{R}} \tilde{\delta}(s,z)^2 \lambda(dz), \\ & \sigma_{V,s} = 2\sigma_s \tilde{\sigma}_s, \quad \sigma'_{V,s} = 2\sigma_s \tilde{\sigma}'_s, \quad \delta_V(s,z) = 2\sigma_{s-} \tilde{\delta}(s,z) + \tilde{\delta}(s,z)^2. \end{aligned} \tag{C.2.2}$$

We then set $B_{V,t} = \int_0^t b_{V,s} ds$ and $M_{V,t} = V_t - V_0 - B_{V,t}$. By Hölder's inequality, $\mathbb{E}\,|b_{V,s}|^q \leqslant K$ and hence for $t > s \geqslant 0$,

$$\mathbb{E}\,|B_{V,t} - B_{V,s}|^q \leqslant K\,|t-s|^q. \tag{C.2.3}$$

By the Burkholder-Davis-Gundy inequality, as well as Lemma 2.1.5 of Jacod and Protter (2012), we derive, for any $0 \leqslant s < t$ with $|t-s| \leqslant 1$,

$$\mathbb{E}\,|M_{V,t} - M_{V,s}|^q \leqslant K. \tag{C.2.4}$$

The decomposition asserted in the lemma is given by

$$\widehat{V}_{n,t} - V_t = \zeta_{n,t} + \zeta'_{n,t},$$

where

$$\zeta_{n,t} = \sum_{j=1}^{4} \zeta_{n,t}^{(j)}, \quad \zeta_{n,t}^{(1)} = \widehat{V}_{n,t} - \widehat{V}'_{n,t}, \quad \zeta_{n,t}^{(2)} = \frac{1}{k_n} \sum_{i=1}^{k_n} \left(\lambda_i^{t,n}\right)^2,$$

$$\zeta_{n,t}^{(3)} = \frac{2}{k_n} \sum_{i=1}^{k_n} \lambda_i^{t,n} \beta_i^{t,n}, \quad \zeta_{n,t}^{(4)} = \frac{1}{k_n} \sum_{i=1}^{k_n} \left(B_{V,t(n,i-1)} - B_{V,t}\right),$$

$$\zeta'_{n,t} = \frac{1}{k_n} \sum_{i=1}^{k_n} V_{t(n,i-1)} \left(\left(\Delta_i^{t,n} W / \Delta_n^{1/2}\right)^2 - 1\right) + \frac{1}{k_n} \sum_{i=1}^{k_n} \left(M_{V,t(n,i-1)} - M_{V,t}\right).$$

It is clear that $\zeta_{n,t}$ and $\zeta'_{n,t}$ are $\mathcal{F}_t$-measurable and $\mathbb{E}\left[\zeta'_{n,t} | \mathcal{F}_{t-k_n\Delta_n}\right] = 0$. It remains to show the inequalities in the assertion.

Step 2. In this step, we show (C.1.1) and (C.1.3). It is easy to see that (C.1.1) follows

$$\mathbb{E}\left|\zeta_{n,t}^{(1)}\right|^q \leqslant K\left(\Delta_n^{1-q(1-2\varpi)} + \Delta_n^{1-q/k-q(1/2-\varpi)} + \Delta_n^{(k-2q)(1/2-\varpi)}\right) \tag{C.2.5}$$

$$\mathbb{E}\left|\zeta_{n,t}^{(2)}\right|^q \leqslant K\Delta_n \tag{C.2.6}$$

$$\mathbb{E}\left|\zeta_{n,t}^{(3)}\right|^q \leqslant K\Delta_n^{(q/2)\wedge(1-q/k)} \tag{C.2.7}$$

$$\mathbb{E}|\zeta_{n,t}^{(4)}|^q \leqslant K\Delta_n^q. \tag{C.2.8}$$

Moreover, when $k \geqslant 2/(1-2\varpi)$, (C.1.3) follows (C.1.1) as an elementary consequence. It remains to establish the estimates in the above display.

First consider (C.2.5). We use the following elementary inequality: for any $w > 0$, $u > 0$, there exists some constant $K > 0$, such that for all $x, y \in \mathbb{R}$,

$$\left||x+y|^2 1_{\{|x+y|\leqslant u\}} - x^2\right| \leqslant K\left((|y| \wedge u)^2 + |x|(|y| \wedge u) + \frac{|x|^{2+w}}{u^w}\right). \tag{C.2.9}$$

Using (C.2.9) with $x = \Delta_i^{t,n} X'$, $y = \Delta_i^{t.n} X''$, $u = \alpha \Delta_n^{\varpi}$, and $w = k/q - 2$, we derive

$$\begin{aligned}
& \mathbb{E}\left| \left|\Delta_i^{t,n} X\right|^2 1_{\left\{\left|\Delta_i^{t,n} X\right| \leqslant \alpha \Delta_n^{\varpi}\right\}} - \left|\Delta_i^{t,n} X'\right|^2 \right|^q \\
\leqslant \quad & K \Delta_n^{2\varpi q} \mathbb{E} \left| \left(\frac{\Delta_i^{t,n} X''}{\Delta_n^{\varpi}} \right) \wedge 1 \right|^{2q} \\
& + K \Delta_n^{\varpi q} \mathbb{E} \left[\left|\Delta_i^{t,n} X'\right|^q \left(\frac{\Delta_i^{t,n} X''}{\Delta_n^{\varpi}} \wedge 1 \right)^q \right] + K \Delta_n^{-\varpi(k-2q)} \mathbb{E} \left[\left|\Delta_i^{t,n} X'\right|^k \right].
\end{aligned}$$

Observe that $(\Delta_i^{t,n} X''/\Delta_n^{\varpi}) \wedge 1$ is bounded and is non-zero only if jumps occur during $(t\,(n, i-1)\,, t\,(n,i)]$. Hence, for any $r \geqslant 0$, $\mathbb{E}[|(\Delta_i^{t,n} X''/\Delta_n^{\varpi}) \wedge 1|^r] \leqslant K\Delta_n$. Moreover, by Hölder's inequality and the Burkholder-Davis-Gundy inequality, $\mathbb{E}|\Delta_i^{t,n} X'|^k \leqslant K \Delta_n^{k/2}$. Combining these estimates and using Hölder's inequality, we derive

$$\begin{aligned}
\mathbb{E}\left[\left|\Delta_i^{t,n} X'\right|^q \left(\frac{\Delta_i^{t,n} X''}{\Delta_n^{\varpi}} \wedge 1 \right)^q \right] \quad & \leqslant \quad \left(\mathbb{E} \left|\Delta_i^{t,n} X'\right|^k \right)^{q/k} \left(\mathbb{E} \left(\frac{\Delta_i^{t,n} X''}{\Delta_n^{\varpi}} \wedge 1 \right)^{qk/(k-q)} \right)^{1-q/k} \\
& \leqslant \quad K \Delta_n^{1+q/2-q/k},
\end{aligned}$$

and hence

$$\begin{aligned}
\mathbb{E}\left| \left|\Delta_i^{t,n} X\right|^2 1_{\left\{\left|\Delta_i^{t,n} X\right| \leqslant \alpha \Delta_n^{\varpi}\right\}} - \left|\Delta_i^{t,n} X'\right|^2 \right|^q & \leqslant K \Delta_n^{1+2\varpi q} + K \Delta_n^{1+q(1/2+\varpi-1/k)} \\
& \quad + K \Delta_n^{2\varpi q - \varpi k + k/2}. \qquad \text{(C.2.10)}
\end{aligned}$$

By Hölder's inequality,

$$\mathbb{E}\left|\zeta_{n,t}^{(1)}\right|^q \leqslant k_n^{-1} \Delta_n^{-q} \sum_{i=1}^{k_n} \mathbb{E} \left| \left|\Delta_i^{t,n} X\right|^2 1_{\left\{\left|\Delta_i^{t,n} X\right| \leqslant \alpha \Delta_n^{\varpi}\right\}} - \left|\Delta_i^{t,n} X'\right|^2 \right|^q. \qquad \text{(C.2.11)}$$

The claim (C.2.5) readily follows (C.2.10) and (C.2.11).

Next, observe that by Hölder's inequality and the Burkholder-Davis-Gundy inequality, we have for any $r \in [0, k]$, $\mathbb{E}|\beta_i^{t,n}|^r \leqslant K$ and $\mathbb{E}|\lambda_i^{t,n}|^r \leqslant K\Delta_n^{(r\wedge 2)/2}$. By using Hölder's inequality again, we derive (C.2.6) as $\mathbb{E}|\zeta_{n,t}^{(2)}|^q \leqslant Kk_n^{-1}\sum_{i=1}^{k_n}\mathbb{E}|\lambda_i^{t,n}|^{2q} \leqslant K\Delta_n$. Similarly, we derive (C.2.7) as

$$\begin{aligned}
\mathbb{E}\left|\zeta_{n,t}^{(3)}\right|^q &\leqslant Kk_n^{-1}\sum_{i=1}^{k_n}\mathbb{E}\left|\lambda_i^{t,n}\beta_i^{t,n}\right|^q \\
&\leqslant Kk_n^{-1}\sum_{i=1}^{k_n}\left(\mathbb{E}\left|\beta_i^{t,n}\right|^k\right)^{q/k}\left(\mathbb{E}\left|\lambda_i^{t,n}\right|^{qk/(k-q)}\right)^{1-q/k} \\
&\leqslant K\Delta_n^{(q/2)\wedge(1-q/k)}. \qquad \text{(C.2.12)}
\end{aligned}$$

Finally, by Hölder's inequality and (C.2.3), we derive (C.2.8).

Step 3. We now turn to (C.1.2). Observe that $\zeta'_{n,t}$ is a sum of martingale differences. When $q \geqslant 2$, by the Burkholder-Davis-Gundy inequality and Hölder's inequality,

$$\mathbb{E}\left|\frac{1}{k_n}\sum_{i=1}^{k_n}V_{t(n,i-1)}\left(\left(\Delta_i^{t,n}W/\Delta_n^{1/2}\right)^2-1\right)\right|^q \leqslant Kk_n^{-q/2}. \qquad \text{(C.2.13)}$$

The estimate still holds when $q \in [1, 2]$, by Jensen's inequality. Combining the same argument with (C.2.4), we derive

$$\mathbb{E}\left|\frac{1}{k_n}\sum_{i=1}^{k_n}\left(M_{V,t(n,i-1)}-M_{V,t}\right)\right|^q \leqslant Kk_n^{-q/2}. \qquad \text{(C.2.14)}$$

By (C.2.13) and (C.2.14), we derive (C.1.2). □

C.2.2 Proof of Lemmas 3.1.3 and C.1.2

We complement the definitions in (3.1.28) for the continuous component X' with the following. Recall $\tau(t,n,i) = \underline{t} + ik_n\Delta_n$ and denote $\Delta_{i,j}^{t,n}X' = X'_{\tau(t,n,i-1)+j\Delta_n} - X'_{\tau(t,n,i-1)+(j-1)\Delta_n}$. We set

$$\begin{aligned} \widehat{CV}'_{n,t} &= 2\sum_{i=2}^{B_n}\left(\sum_{j=1}^{k_n}\Delta_{i,j}^{t,n}X'\right)\left(\widehat{V}'_{n,\tau(t,n,i)} - \widehat{V}'_{n,\tau(t,n,i-1)}\right), \quad \text{where} \\ \widehat{V}'_{n,\tau(t,n,i)} &= \frac{1}{k_n\Delta_n}\sum_{j=1}^{k_n}\left(\Delta_{i,j}^{t,n}X'\right)^2, \quad i \in \{1,\ldots,B_n\}. \end{aligned} \tag{C.2.15}$$

We also complement (3.1.27) with

$$\widehat{IV}'_{n,t} = \sum_{i=1}^{n}\left|\Delta_i^{t,n}X'\right|^2. \tag{C.2.16}$$

We first prove two auxiliary lemmas, Lemma C.2.1 and Lemma C.2.2, and then prove Lemmas 3.1.3 and C.1.2. The proof of Lemma C.1.2 makes full use of these auxiliary lemmas. The proof of Lemma 3.1.3 makes partial use of Lemma C.2.1, with much to spare.

Lemma C.2.1. *Let $\varpi \in \left(\frac{5}{12},\frac{1}{2}\right)$ and $r \in \left[1,\frac{4}{9-12\varpi}\right]$ be constants. Suppose (i) Assumption A holds for some $k \geqslant 8 \vee \frac{2}{1-2\varpi}$ and (ii) the process σ_t is continuous. Then*

$$\begin{aligned} \mathbb{E}\left|\Delta_n^{-1/4}\left(\widehat{CV}_{n,t} - \widehat{CV}'_{n,t}\right)\right| + \mathbb{E}\left|\Delta_n^{-1/4}\left(\widehat{IV}_{n,t} - \widehat{IV}'_{n,t}\right)\right| &\leqslant K\Delta_n^{3\varpi-5/4}, \\ \mathbb{E}\left|\Delta_n^{-1/4}\left(\widehat{CV}_{n,t} - \widehat{CV}'_{n,t}\right)\right|^r + \mathbb{E}\left|\Delta_n^{-1/4}\left(\widehat{IV}_{n,t} - \widehat{IV}'_{n,t}\right)\right|^r &\leqslant K. \end{aligned}$$

Proof. Step 1. In this step, we derive the following preliminary estimates: for any

$q \in [0, k]$,

$$\mathbb{E}\left|\sum_{j=1}^{k_n}\left(\left(\Delta_{i,j}^{t,n} X\right) 1_{\left\{\left|\Delta_{i,j}^{t,n} X\right| \leqslant \alpha \Delta_n^{\varpi}\right\}} - \Delta_{i,j}^{t,n} X'\right)\right|^q \leqslant K\Delta_n^{1-q(1/2-\varpi)} + K\Delta_n^{(k-q)(1/2-\varpi)}, \tag{C.2.17}$$

$$\mathbb{E}\left|\widehat{V}'_{\tau(t,n,i)} - V_{\tau(t,n,i)}\right|^{k/2} \leqslant K\Delta_n^{k/8}, \tag{C.2.18}$$

$$\mathbb{E}\left|\hat{V}'_{\tau(t,n,i)} - \hat{V}'_{\tau(t,n,i-1)}\right|^{k/2} \leqslant K\Delta_n^{k/8}. \tag{C.2.19}$$

We start with (C.2.17). Observe that for any $v > 0$, $w > 0$, there exists a constant $K > 0$ such that for all $x, y \in \mathbb{R}$,

$$\left|(x+y)\, 1_{\{|x+y| \leqslant v\}} - x\right| \leqslant K\left(|y| \wedge v + \frac{|x|^{1+w}}{v^w}\right). \tag{C.2.20}$$

For any $q \in [0, k]$, applying (C.2.20) with $x = \Delta_{i,j}^{t,n} X'$, $y = \Delta_{i,j}^{t,n} X''$, $v = \alpha\Delta_n^{\varpi}$, and $w = k/q - 1$ yields

$$\begin{aligned}
&\mathbb{E}\left|\left(\Delta_{i,j}^{t,n} X\right) 1_{\left\{\left|\Delta_{i,j}^{t,n} X\right| \leqslant \alpha \Delta_n^{\varpi}\right\}} - \Delta_{i,j}^{t,n} X'\right|^q \\
&\leqslant K\mathbb{E}\left[\left(\left|\Delta_{i,j}^{t,n} X''\right| \wedge \Delta_n^{\varpi}\right)^q\right] + K\mathbb{E}\left[\frac{\left|\Delta_{i,j}^{t,n} X'\right|^k}{\Delta_n^{\varpi(k-q)}}\right] \\
&\leqslant K\Delta_n^{1+\varpi q} + K\Delta_n^{k/2-\varpi(k-q)}.
\end{aligned}$$

By Hölder's inequality and $k_n \leqslant K\Delta_n^{-1/2}$, we readily derive (C.2.17).

We now show (C.2.18). We denote a discretized version of the quantity X' by $X'_{n,s} = X'_{\tau(t,n,i-1)+(j-1)\Delta_n}$, for $s \in (\tau(t,n,i-1) + (j-1)\Delta_n, \tau(t,n,i-1) + j\Delta_n]$. By

Itô's formula, we can decompose

$$\widehat{V}'_{n,\tau(t,n,i)} = \frac{1}{k_n\Delta_n}\int_{\tau(t,n,i-1)}^{\tau(t,n,i)} 2\left(X'_s - X'_{n,s}\right)(b_s ds + \sigma_s dW_s) + \frac{1}{k_n\Delta_n}\int_{\tau(t,n,i-1)}^{\tau(t,n,i)} V_s ds, \tag{C.2.21}$$

Recalling the classical estimate that $\mathbb{E}\left|X'_s - X'_{n,s}\right|^k \leqslant K\Delta_n^{k/2}$, by Hölder's inequality and the Burkholder-Davis-Gundy inequality, we derive

$$\begin{cases} \mathbb{E}\left|\frac{1}{k_n\Delta_n}\int_{\tau(t,n,i-1)}^{\tau(t,n,i)} 2\left(X'_s - X'_{n,s}\right) b_s ds\right|^{k/2} & \leqslant K\Delta_n^{k/4} \\ \mathbb{E}\left|\frac{1}{k_n\Delta_n}\int_{\tau(t,n,i-1)}^{\tau(t,n,i)} 2\left(X'_s - X'_{n,s}\right) \sigma_s dW_s\right|^{k/2} & \leqslant K\Delta_n^{k/8}. \end{cases} \tag{C.2.22}$$

Since σ_t (and hence V_t) is continuous, we have $\mathbb{E}\left|V_t - V_s\right|^{k/2} \leqslant K\left|t-s\right|^{k/4}$ by a standard estimate for continuous Itô processes. By Hölder's inequality, we derive

$$\mathbb{E}\left|\frac{1}{k_n\Delta_n}\int_{\tau(t,n,i-1)}^{\tau(t,n,i)} V_s ds - V_{\tau(t,n,i)}\right|^{k/2} \leqslant K\Delta_n^{k/8}. \tag{C.2.23}$$

Combining (C.2.21)-(C.2.23), we have (C.2.18), which further implies (C.2.19).

Step 2. In this step, we show that

$$\begin{aligned} \mathbb{E}\left|\widehat{CV}_{n,t} - \widehat{CV}'_{n,t}\right|^r + \mathbb{E}\left|\widehat{IV}_{n,t} - \widehat{IV}'_{n,t}\right|^r \leqslant & K\Delta_n^{1+r(\varpi-3/4-2/k)} + K\Delta_n^{1-r/2-3r(1/2-\varpi)} \\ & + K\Delta_n^{1-r/k-r/2-2r(1/2-\varpi)} \\ & + K\Delta_n^{(k-3r)(1/2-\varpi)-r/2} \end{aligned} \tag{C.2.24}$$

We consider the following decomposition:

$$\widehat{CV}_{n,t} - \widehat{CV}'_{n,t} = R^{(1)}_{n,t} + R^{(2)}_{n,t}, \tag{C.2.25}$$

where

$$R_{n,t}^{(1)} = 2\sum_{i=2}^{B_n}\left[\sum_{j=1}^{k_n}\left(\left(\Delta_{i,j}^{t,n}X\right)1_{\left\{\left|\Delta_{i,j}^{t,n}X\right|\leqslant\alpha\Delta_n^{\varpi}\right\}} - \Delta_{i,j}^{t,n}X'\right)\right]\left(\widehat{V}'_{n,\tau(t,n,i)} - \widehat{V}'_{n,\tau(t,n,i-1)}\right)$$

$$R_{n,t}^{(2)} = 2\sum_{i=2}^{B_n}\left(\sum_{j=1}^{k_n}\left(\Delta_{i,j}^{t,n}X\right)1_{\left\{\left|\Delta_{i,j}^{t,n}X\right|\leqslant\alpha\Delta_n^{\varpi}\right\}}\right)\left(\widehat{V}_{n,\tau(t,n,i)} - \widehat{V}_{n,\tau(t,n,i-1)}\right.$$
$$\left. - \left(\widehat{V}'_{n,\tau(t,n,i)} - \widehat{V}'_{n,\tau(t,n,i-1)}\right)\right).$$

We first consider $R_{n,t}^{(1)}$. Let q solve $1/q + 2r/k = 1$. The conditions on ϖ, r and k imply $qr \leqslant k$. Then we have

$$\mathbb{E}\left|\left(\sum_{j=1}^{k_n}\left(\left(\Delta_{i,j}^{t,n}X\right)1_{\left\{\left|\Delta_{i,j}^{t,n}X\right|\leqslant\alpha\Delta_n^{\varpi}\right\}} - \Delta_{i,j}^{t,n}X'\right)\right)\left(\hat{V}'_{\tau(t,n,i)} - \hat{V}'_{\tau(t,n,i-1)}\right)\right|^r$$
$$\leqslant \left(\mathbb{E}\left|\sum_{j=1}^{k_n}\left(\left(\Delta_{i,j}^{t,n}X\right)1_{\left\{\left|\Delta_{i,j}^{t,n}X\right|\leqslant\alpha\Delta_n^{\varpi}\right\}} - \Delta_{i,j}^{t,n}X'\right)\right|^{qr}\right)^{1/q}\left(\mathbb{E}\left|\hat{V}'_{\tau(t,n,i)} - \hat{V}'_{\tau(t,n,i-1)}\right|^{k/2}\right)^{2r/k}$$
$$\leqslant K\Delta_n^{1+r(\varpi-1/4-2/k)} + K\Delta_n^{k(1/2-\varpi)+r(3\varpi-5/4)},$$

where the first inequality is due to Hölder's inequality and the second inequality follows from (C.2.17) and (C.2.19). Since $B_n \leqslant K\Delta_n^{-1/2}$, by the triangle inequality,

$$\mathbb{E}\left|R_{n,t}^{(1)}\right|^r \leqslant K\Delta_n^{1+r(\varpi-3/4-2/k)} + K\Delta_n^{(k-3r)(1/2-\varpi)-r/4}. \tag{C.2.26}$$

Now, we turn to $R_{n,t}^{(2)}$. By (C.2.5),

$$\mathbb{E}\left|\widehat{V}_{\tau(t,n,i)} - \widehat{V}'_{\tau(t,n,i)}\right|^r \leqslant K\left(\Delta_n^{1-r(1-2\varpi)} + \Delta_n^{1-r/k-r(1/2-\varpi)} + \Delta_n^{(k-2r)(1/2-\varpi)}\right).$$

Since $|\sum_{j=1}^{k_n}(\Delta_{i,j}^{t,n}X)1\{|\Delta_{i,j}^{t,n}X| \leqslant \alpha\Delta_n^{\varpi}\}| \leqslant K\Delta_n^{\varpi-1/2}$, we derive

$$\mathbb{E}\left|R_{n,t}^{(2)}\right|^r \leqslant K\Delta_n^{1-r/2-3r(1/2-\varpi)} + K\Delta_n^{1-r/k-r/2-2r(1/2-\varpi)} + K\Delta_n^{(k-3r)(1/2-\varpi)-r/2}. \tag{C.2.27}$$

Since $\widehat{IV}_{n,t} - \widehat{IV}'_{n,t} = k_n\Delta_n\sum_{i=1}^{B_n}(\hat{V}_{\tau(t,n,i)} - \hat{V}'_{\tau(t,n,i)})$, it is easy to see that $\mathbb{E}|\widehat{IV}_{n,t} - \widehat{IV}'_{n,t}|^r$ can also be bounded by the right-hand side of (C.2.27), with much to spare. Combining (C.2.25), (C.2.26) and (C.2.27), we derive (C.2.24).

Step 3. We now prove the assertion of the lemma. Using $k \geqslant 8 \vee \frac{2}{1-2\varpi}$, it is elementary to show that the second term on the majorant side of (C.2.24) dominates other terms asymptotically. The first assertion follows (C.2.24) with $r = 1$. For the second assertion, we observe that by (C.2.24),

$$\mathbb{E}\left|\Delta_n^{-1/4}\left(\widehat{CV}_{n,t} - \widehat{CV}'_{n,t}\right)\right|^r + \mathbb{E}\left|\Delta_n^{-1/4}\left(\widehat{IV}_{n,t} - \widehat{IV}'_{n,t}\right)\right|^r \leqslant K\Delta_n^{1-9r/4+3\varpi r}.$$

Under the assumption that $\varpi \in (5/12, 1/2)$ and $1 \leqslant r \leqslant 4/(9-12\varpi)$, the right-hand side of the display is bounded by a constant, as asserted. □

Lemma C.2.2. *Suppose Assumption A holds for some $k \geqslant 8$. In addition, suppose (3.2.1). Then*

$$\mathbb{E}\left|\mathbb{E}\left[\Delta_n^{-1/4}\left(\widehat{CV}'_{n,t} - CV_t\right)\middle|\mathcal{F}_{t-1}\right]\right| + \mathbb{E}\left|\mathbb{E}\left[\Delta_n^{-1/4}\left(\widehat{IV}'_{n,t} - IV_t\right)\middle|\mathcal{F}_{t-1}\right]\right| \leqslant K\Delta_n^{1/4}, \tag{C.2.28}$$

$$\mathbb{E}\left|\Delta_n^{-1/4}\left(\widehat{CV}'_{n,t} - CV_t\right)\right|^2 + \mathbb{E}\left|\Delta_n^{-1/4}\left(\widehat{IV}'_{n,t} - IV_t\right)\right|^2 \leqslant K. \tag{C.2.29}$$

Proof. Step 1. In this step, we introduce some notations and preliminary estimates.

Recall $\tau(t,n,i) = \underline{t} + ik_n\Delta_n$. For notational simplicity, we denote $\tau(i) = \tau(t,n,i)$. By (C.2.21), we can decompose

$$\begin{aligned}
\widehat{V}'_{n,\tau(t,n,i)} &= 2\sum_{j=1}^{3}\zeta^{(j)}_{t,n,i} + \bar{V}'_{\tau(t,n,i)}, \quad \text{where} \\
\zeta^{(1)}_{t,n,i} &= \frac{1}{k_n\Delta_n}\int_{\tau(i-1)}^{\tau(i)}\left(X'_s - X'_{n,s}\right)b_{\tau(i-1)}ds, \\
\zeta^{(2)}_{t,n,i} &= \frac{1}{k_n\Delta_n}\int_{\tau(i-1)}^{\tau(i)}\left(X'_s - X'_{n,s}\right)\left(b_s - b_{\tau(i-1)}\right)ds \\
\zeta^{(3)}_{t,n,i} &= \frac{1}{k_n\Delta_n}\int_{\tau(i-1)}^{\tau(i)}\left(X'_s - X'_{n,s}\right)\sigma_s dW_s, \quad \bar{V}'_{n,\tau(t,n,i)} = \frac{1}{k_n\Delta_n}\int_{\tau(i-1)}^{\tau(i)} V_s ds.
\end{aligned} \tag{C.2.30}$$

Observe that $\bar{V}'_{n,\tau(i)} - \bar{V}'_{n,\tau(i-1)} = (k_n\Delta_n)^{-1}\int_{\tau(i-1)}^{\tau(i)}\left(V_s - V_{s-k_n\Delta_n}\right)ds$. By (3.2.1), we decompose

$$\bar{V}'_{n,\tau(i)} - \bar{V}'_{\tau(i-1)} = \sum_{j=1}^{3} z^{(j)}_{t,n,i}, \quad \text{where}$$

$$\begin{aligned}
z^{(1)}_{t,n,i} &= \frac{1}{k_n\Delta_n}\int_{\tau(i-1)}^{\tau(i)}\left(\int_{s-k_n\Delta_n}^{s} b_{V,u}du\right)ds \\
z^{(2)}_{t,n,i} &= \frac{\rho v}{k_n\Delta_n}\int_{\tau(i-1)}^{\tau(i)}\left(\int_{s-k_n\Delta_n}^{s}\sigma_u dW_u\right)ds, \\
z^{(3)}_{t,n,i} &= \frac{\left(1-\rho^2\right)^{1/2} v}{k_n\Delta_n}\int_{\tau(i-1)}^{\tau(i)}\left(\int_{s-k_n\Delta_n}^{s}\sigma_u dW'_u\right)ds.
\end{aligned}$$

We then rewrite

$$\widehat{V}'_{n,\tau(t,n,i)} - \widehat{V}'_{n,\tau(t,n,i-1)} = \sum_{j=1}^{3}\zeta^{(j)}_{t,n,i} - \sum_{j=1}^{3}\zeta^{(j)}_{t,n,i-1} + \sum_{j=1}^{3} z^{(j)}_{t,n,i}. \tag{C.2.31}$$

We also decompose$X'_{\tau(i)} - X'_{\tau(i-1)} = x_{t,n,i} + y_{t,n,i}$, where

$$x_{t,n,i} = \int_{\tau(i-1)}^{\tau(i)} b_s ds, \quad y_{t,n,i} = \int_{\tau(i-1)}^{\tau(i)} \sigma_s dW_s. \tag{C.2.32}$$

We now collect some obvious estimates: for $q \in [0, k]$,

$$\begin{cases} \mathbb{E}\left|X'_s - X'_{n,s}\right|^q \leqslant K\Delta_n^{q/2}, \quad \mathbb{E}\left|x_{t,n,i}\right|^q \leqslant K\Delta_n^{q/2}, \quad \mathbb{E}\left|y_{t,n,i}\right|^q \leqslant K\Delta_n^{q/4}, \\ \mathbb{E}\left|\zeta_{t,n,i}^{(1)}\right|^q + \mathbb{E}\left|\zeta_{t,n,i}^{(2)}\right|^q \leqslant K\Delta_n^{q/2}, \quad \mathbb{E}\left|\zeta_{t,n,i}^{(2)}\right|^{4/3} \leqslant K\Delta_n, \quad \mathbb{E}\left|\zeta_{t,n,i}^{(3)}\right|^q \leqslant K\Delta_n^{q/4}, \end{cases} \tag{C.2.33}$$

and for $q \in [0, k/2]$,

$$\mathbb{E}\left|z_{t,n,i}^{(1)}\right|^q \leqslant K\Delta_n^{q/2}, \quad \mathbb{E}\left|z_{t,n,i}^{(2)}\right|^q + \mathbb{E}\left|z_{t,n,i}^{(3)}\right|^q \leqslant K\Delta_n^{q/4}. \tag{C.2.34}$$

By Itô's lemma, (3.2.1) implies that $\tilde{\sigma}_t = \rho v/2$, $\tilde{\sigma}'_t = (1-\rho^2)^{1/2} v/2$ and $\tilde{\delta}(s,z) = 0$. By (C.2.2), $b_{V,u} - b_{V,s} = 2\left(\sigma_u \tilde{b}_u - \sigma_s \tilde{b}_s\right)$ for any $u, s \geqslant 0$; hence,

$$\begin{aligned} \mathbb{E}\left|y_{t,n,i}\left(b_{V,u} - b_{V,s}\right)\right| &\leqslant K\mathbb{E}\left|y_{t,n,i}\tilde{b}_u\left(\sigma_u - \sigma_s\right)\right| + K\mathbb{E}\left|y_{t,n,i}\sigma_s\left(\tilde{b}_u - \tilde{b}_s\right)\right| \\ &\leqslant K\Delta_n^{1/4}\left|u - s\right|^{1/2}, \end{aligned} \tag{C.2.35}$$

where the first inequality follows the triangle inequality; the second inequality follows from Hölder's inequality and $\mathbb{E}(\left|\sigma_u - \sigma_s\right|^2 + |\tilde{b}_u - \tilde{b}_s|^2) \leqslant K|u - s|$, the latter in turn is implied by Assumption A.

Below, steps 2-5 are devoted to proving (C.2.28); step 6 contains the proof of (C.2.29).

Step 2. In this step, we show that

$$\mathbb{E}\left|\mathbb{E}\left[x_{t,n,i}\left(\widehat{V}'_{n,\tau(i)}-\widehat{V}'_{n,\tau(i-1)}\right)\middle|\mathcal{F}_{t-1}\right]\right|\leqslant K\Delta_n. \tag{C.2.36}$$

By the Cauchy-Schwarz inequality and (C.2.33),

$$\mathbb{E}\left|x_{t,n,i}\zeta^{(j)}_{t,n,i}\right|+\mathbb{E}\left|x_{t,n,i}\zeta^{(j)}_{t,n,i-1}\right|+\mathbb{E}\left|x_{t,n,i}z^{(1)}_{t,n,i}\right|\leqslant K\Delta_n,\quad j=1,2.$$

Let $x'_{t,n,i}=\int_{\tau(i-1)}^{\tau(i)}\left(b_s-b_{\tau(i-2)}\right)ds$. By Cauchy-Schwarz, $\mathbb{E}|x'_{t,n,i}|^2\leqslant K\Delta_n^{3/2}$. We then observe

$$\mathbb{E}\left|\mathbb{E}\left[x_{t,n,i}\zeta^{(3)}_{t,n,i-1}\middle|\mathcal{F}_{t-1}\right]\right|=\mathbb{E}\left|\mathbb{E}\left[x'_{t,n,i}\zeta^{(3)}_{t,n,i-1}\middle|\mathcal{F}_{t-1}\right]\right|\leqslant\mathbb{E}\left|x'_{t,n,i}\zeta^{(3)}_{t,n,i-1}\right|\leqslant K\Delta_n,$$

where the equality holds because $\mathbb{E}[\zeta^{(3)}_{t,n,i-1}|\mathcal{F}_{\tau(i-2)}]=0$ and $x_{t,n,i}-x'_{t,n,i}$ is $\mathcal{F}_{\tau(i-2)}$ measurable; the first inequality follows Jensen's inequality and repeated conditioning; the second inequality follows (C.2.33) and Cauchy-Schwarz. Using a similar argument, we can also show

$$\mathbb{E}\left|\mathbb{E}\left[x_{t,n,i}\left(\zeta^{(3)}_{t,n,i}+z^{(2)}_{t,n,i}+z^{(3)}_{t,n,i}\right)\middle|\mathcal{F}_{t-1}\right]\right|\leqslant K\Delta_n.$$

Combining the displayed estimates in this step, we readily derive (C.2.36).

Step 3. In this step, we show that

$$\mathbb{E}\left|\mathbb{E}\left[y_{t,n,i}\left(\sum_{j=1}^{3}\zeta^{(j)}_{t,n,i}-\sum_{j=1}^{3}\zeta^{(j)}_{t,n,i-1}\right)\middle|\mathcal{F}_{t-1}\right]\right|\leqslant K\Delta_n. \tag{C.2.37}$$

Denote $\tau\left(t,n,i-1,j\right)=\tau\left(t,n,i-1\right)+j\Delta_n$, for $j\in\{0,\ldots,k_n\}$. For notational simplicity, we write $\tau(i,j)$ in place of $\tau(t,n,i,j)$ below. By definition (see (C.2.30)),

$$\begin{aligned}&\mathbb{E}\left|\mathbb{E}\left[y_{t,n,i}\zeta_{t,n,i}^{(1)}\middle|\mathcal{F}_{t-1}\right]\right|\\&\quad\leqslant\frac{1}{k_n\Delta_n}\mathbb{E}\left|\mathbb{E}\left[y_{t,n,i}\sum_{j=1}^{k_n}\int_{\tau(i-1,j-1)}^{\tau(i-1,j)}\left(\int_{\tau(i-1,j-1)}^{s}b_udu\right)b_{\tau(i-1)}ds\middle|\mathcal{F}_{t-1}\right]\right|\\&\qquad+\frac{1}{k_n\Delta_n}\mathbb{E}\left|\mathbb{E}\left[y_{t,n,i}\sum_{j=1}^{k_n}\int_{\tau(i-1,j-1)}^{\tau(i-1,j)}\left(\int_{\tau(i-1,j-1)}^{s}\sigma_udW_u\right)b_{\tau(i-1)}ds\middle|\mathcal{F}_{t-1}\right]\right|.\end{aligned} \tag{C.2.38}$$

By (C.2.33) and properties of conditional expectation, it is easy to see that the first term on the right-hand side of (C.2.38) can be bounded by $K\Delta_n^{5/4}$. Moreover, the second term can be bounded by $K\Delta_n$. To see this, it suffices to note

$$\begin{aligned}&\mathbb{E}\left[y_{t,n,i}\sum_{j=1}^{k_n}\int_{\tau(i-1,j-1)}^{\tau(i-1,j)}\left(\int_{\tau(i-1,j-1)}^{s}\sigma_udW_u\right)b_{\tau(i-1)}ds\middle|\mathcal{F}_{t-1}\right]\\&=\mathbb{E}\Bigg[\sum_{j=1}^{k_n}\int_{\tau(i-1,j-1)}^{\tau(i-1,j)}\mathbb{E}\left[\left(\int_{\tau(i-1)}^{\tau(i)}\sigma_udW_u\right)\left(\int_{\tau(i-1,j-1)}^{s}\sigma_udW_u\right)\middle|\mathcal{F}_{\tau(i-1)}\right]\\&\quad b_{\tau(i-1)}ds|\mathcal{F}_{t-1}\Bigg]\\&=\mathbb{E}\left[\sum_{j=1}^{k_n}\int_{\tau(i-1,j-1)}^{\tau(i-1,j)}\left(\int_{\tau(i-1,j-1)}^{s}\sigma_u^2du\right)b_{\tau(i-1)}ds\middle|\mathcal{F}_{t-1}\right],\end{aligned}$$

where the first equality follows from the definition of $y_{t,n,i}$ and repeated conditioning; the second equality follows properties of stochastic integrals and repeated conditioning. Hence, (C.2.38) further implies

$$\mathbb{E}\left|\mathbb{E}\left[y_{t,n,i}\zeta_{t,n,i}^{(1)}\middle|\mathcal{F}_{t-1}\right]\right|\leqslant K\Delta_n. \tag{C.2.39}$$

Next, we observe

$$\mathbb{E}\left|\mathbb{E}\left[y_{t,n,i}\zeta^{(2)}_{t,n,i}\middle| \mathcal{F}_{t-1}\right]\right| \leqslant \mathbb{E}\left|y_{t,n,i}\zeta^{(2)}_{t,n,i}\right| \leqslant \left(\mathbb{E}\left|y_{t,n,i}\right|^4\right)^{1/4}\left(\mathbb{E}\left|\zeta^{(2)}_{t,n,i}\right|^{4/3}\right)^{3/4} \leqslant K\Delta_n. \tag{C.2.40}$$

where the first inequality is obvious; the second inequality is by Hölder's inequality; the last inequality follows (C.2.33).

Now, note that

$$\begin{aligned}
&\mathbb{E}\left|\mathbb{E}\left[y_{t,n,i}\zeta^{(3)}_{t,n,i}\middle| \mathcal{F}_{t-1}\right]\right| \\
&\quad = \mathbb{E}\left|\mathbb{E}\left[\frac{1}{k_n\Delta_n}\int_{\tau(i-1)}^{\tau(i)}\left(X'_s - X'_{n,s}\right)V_s ds\middle|\mathcal{F}_{t-1}\right]\right| \\
&\quad \leqslant \mathbb{E}\left[\frac{1}{k_n\Delta_n}\sum_{j=1}^{k_n}\left|\int_{\tau(i-1,j-1)}^{\tau(i-1,j)}\left(\int_{\tau(i-1,j-1)}^{s} b_u du\right)V_{\tau(i-1,j-1)}ds\right|\right] \\
&\qquad + \mathbb{E}\left|\mathbb{E}\left[\frac{1}{k_n\Delta_n}\sum_{j=1}^{k_n}\int_{\tau(i-1,j-1)}^{\tau(i-1,j)}\left(\int_{\tau(i-1,j-1)}^{s}\sigma_u dW_u\right)V_{\tau(i-1,j-1)}ds\middle|\mathcal{F}_{t-1}\right]\right| \\
&\qquad + \mathbb{E}\left[\frac{1}{k_n\Delta_n}\sum_{j=1}^{k_n}\left|\int_{\tau(i-1,j-1)}^{\tau(i-1,j)}\left(X'_s - X'_{\tau(i-1,j-1)}\right)\left(V_s - V_{\tau(i-1,j-1)}\right)\right|ds\right] \\
&\quad \leqslant K\Delta_n,
\end{aligned} \tag{C.2.41}$$

where the equality follows from properties of stochastic integrals; the first inequality is obtained by using the triangle inequality, as well as properties of conditional expection. Observing that the second term on the majorant side of the first inequality is zero, we readily derive the second inequality.

Finally, observe that $\mathbb{E}[y_{t,n,i}|\mathcal{F}_{\tau(i-1)}] = 0$ and $\zeta^{(j)}_{t,n,i-1} \in \mathcal{F}_{\tau(i-1)}$ for $j = 1, 2, 3$; hence, $\mathbb{E}[y_{t,n,i}\zeta^{(j)}_{t,n,i-1}|\mathcal{F}_{t-1}] = 0$. Combining this with (C.2.39), (C.2.40) and (C.2.41), we have (C.2.37).

Step 4. In this step, we show that

$$\mathbb{E}\left|\mathbb{E}\left[y_{t,n,i}\left(z_{t,n,i}^{(1)}+z_{t,n,i}^{(3)}\right)\middle|\mathcal{F}_{t-1}\right]\right| \leqslant K\Delta_n \tag{C.2.42}$$

$$\mathbb{E}\left|\mathbb{E}\left[2\sum_{i=1}^{B_n} y_{t,n,i}z_{t,n,i}^{(2)} - \rho v\int_{\underline{t}}^{t} V_s ds\middle|\mathcal{F}_{t-1}\right]\right| \leqslant K\Delta_n^{1/2}. \tag{C.2.43}$$

Since the Brownian motions W and W' are orthogonal, $\mathbb{E}[y_{t,n,i}z_{t,n,i}^{(3)}|\mathcal{F}_{t-1}] = 0$. Since $\mathbb{E}\left[y_{t,n,i}|\mathcal{F}_{\tau(i-1)}\right] = 0$, by repeated conditioning,

$$\begin{aligned}
&\mathbb{E}\left|\mathbb{E}\left[y_{t,n,i}z_{t,n,i}^{(1)}\middle|\mathcal{F}_{t-1}\right]\right| \\
&= \mathbb{E}\left|\mathbb{E}\left[y_{t,n,i}\frac{1}{k_n\Delta_n}\int_{\tau(i-1)}^{\tau(i)}\left(\int_{s-k_n\Delta_n}^{s}\left(b_{V,u}-b_{V,\tau(i-1)}\right)du\right)ds\middle|\mathcal{F}_{t-1}\right]\right| \\
&\leqslant K\Delta_n,
\end{aligned}$$

where the inequality follows from Jensen's inequality, repeated conditioning and (C.2.35). Then (C.2.42) is obvious.

We now consider (C.2.43). Under (3.2.1), we have the following decomposition

$$\begin{aligned}
2y_{t,n,i}z_{t,n,i}^{(2)} &= \lambda_{t,n,i}^{(1)} + \lambda_{t,n,i}^{(2)} + \lambda_{t,n,i}^{(3)} + \lambda_{t,n,i}^{(4)}, \quad \text{where} \\
\lambda_{t,n,i}^{(1)} &= \rho v V_{\tau(i-1)}k_n\Delta_n \\
\lambda_{t,n,i}^{(2)} &= \frac{2\rho v}{k_n\Delta_n}\int_{\tau(i-1)}^{\tau(i)}\left(\int_{\tau(i-1)}^{s}\left(\int_{\tau(i-1)}^{u} b_{V,r}dr\right)du\right)ds \\
\lambda_{t,n,i}^{(3)} &= \frac{2\rho v}{k_n\Delta_n}\int_{\tau(i-1)}^{\tau(i)}\left(\int_{\tau(i-1)}^{s}\left(\rho v\int_{\tau(i-1)}^{u}\sigma_r dW_r + \left(1-\rho^2\right)^{1/2} v\int_{\tau(i-1)}^{u}\sigma_r dW'_r\right)du\right)ds \\
\lambda_{t,n,i}^{(4)} &= 2y_{t,n,i}z_{t,n,i}^{(2)} - \frac{2\rho v}{k_n\Delta_n}\int_{\tau(i-1)}^{\tau(i)}\left(\int_{\tau(i-1)}^{s} V_u du\right)ds.
\end{aligned} \tag{C.2.44}$$

Note that $\mathbb{E}[\lambda^{(3)}_{t,n,i}|\mathcal{F}_{t-1}] = 0$, $\mathbb{E}[\lambda^{(4)}_{t,n,i}|\mathcal{F}_{t-1}] = 0$ and $\mathbb{E}|\lambda^{(2)}_{t,n,i}| \leqslant K\Delta_n$. Hence,

$$\mathbb{E}\left|\mathbb{E}\left[2\sum_{i=1}^{B_n} y_{t,n,i} z^{(2)}_{t,n,i} - \rho v \int_{\underline{t}}^{t} V_s ds \middle| \mathcal{F}_{t-1}\right]\right| \leqslant \mathbb{E}\left|\mathbb{E}\left[\sum_{i=2}^{B_n} \lambda^{(1)}_{t,n,i} - \rho v \int_{\underline{t}}^{t} V_s ds \middle| \mathcal{F}_{t-1}\right]\right| + K\Delta_n^{1/2}. \tag{C.2.45}$$

The first term on the right-hand side of the above can be further bounded by

$$\begin{aligned} & K\mathbb{E}\left|\mathbb{E}\left[\sum_{i=2}^{B_n}\int_{\tau(i-1)}^{\tau(i)} \left(V_s - V_{\tau(i-1)}\right) ds \middle| \mathcal{F}_{t-1}\right]\right| + K\mathbb{E}\left|\int_{\underline{t}}^{\underline{t}+k_n\Delta_n} V_s ds\right| \\ & \leqslant K\mathbb{E}\left|\mathbb{E}\left[\sum_{i=2}^{B_n}\int_{\tau(i-1)}^{\tau(i)} \left(\int_{\tau(i-1)}^{s} b_{V,u} du\right) ds \middle| \mathcal{F}_{t-1}\right]\right| + K k_n \Delta_n \\ & \leqslant K\Delta_n^{1/2}. \end{aligned} \tag{C.2.46}$$

Combining (C.2.45) and (C.2.46), we derive (C.2.43).

Step 5. We prove (C.2.28) in this step. By (C.2.37), (C.2.42) and (C.2.43), we have

$$\mathbb{E}\left|\mathbb{E}\left[2\sum_{i=2}^{B_n} y_{t,n,i}\left(\widehat{V}'_{n,\tau(i)} - \widehat{V}'_{n,\tau(i-1)}\right) - \rho v \int_{\underline{t}}^{t} V_s ds \middle| \mathcal{F}_{t-1}\right]\right| \leqslant K\Delta_n^{1/2}.$$

Then by (C.2.36), we derive $\mathbb{E}|\mathbb{E}[\widehat{CV}'_{n,t} - CV_t|\mathcal{F}_{t-1}]| \leqslant K\Delta_n^{1/2}$. Moreover, we observe $\widehat{IV}_{n,t} = k_n\Delta_n \sum_{i=1}^{B_n} \widehat{V}_{n,\tau(i)}$, $IV_t = k_n\Delta_n \sum_{i=1}^{B_n} \bar{V}_{n,\tau(i)}$. Since $\mathbb{E}[\zeta^{(3)}_{t,n,i}|\mathcal{F}_{t-1}] = 0$, we derive $\mathbb{E}|\mathbb{E}[\widehat{IV}_{n,t} - IV_t|\mathcal{F}_{t-1}]| \leqslant K\Delta_n^{1/2}$ by using (C.2.30) and (C.2.33). Combining the estimates in this step, we have (C.2.28).

Step 6. We show (C.2.29) in this step. By (C.2.30) and (C.2.33), it is easily seen that $\mathbb{E}|\Delta_n^{-1/4}(\widehat{IV}'_{n,t} - IV_t)|^2 \leqslant K$; the details are omitted for brevity. Below, we prove the more complicated part of the assertion, i.e.,

$$\mathbb{E}\left|\Delta_n^{-1/4}(\widehat{CV}'_{n,t} - CV_t)\right|^2 \leqslant K. \tag{C.2.47}$$

First, by Cauchy-Schwarz, (C.2.31)-(C.2.34),

$$\begin{cases} \mathbb{E}\left|x_{t,n,i}\left(\widehat{V}'_{n,\tau(i)} - \widehat{V}'_{n,\tau(i-1)}\right)\right|^2 \leqslant K\Delta_n^{3/2}, \\ \mathbb{E}\left|y_{t,n,i}z^{(1)}_{t,n,i}\right|^2 + \mathbb{E}\left|y_{t,n,i}\left(\zeta^{(1)}_{t,n,i} + \zeta^{(2)}_{t,n,i}\right)\right|^2 + \mathbb{E}\left|y_{t,n,i}\left(\zeta^{(1)}_{t,n,i-1} + \zeta^{(2)}_{t,n,i-1}\right)\right|^2 \leqslant K\Delta_n^{3/2}. \end{cases} \tag{C.2.48}$$

We claim for the moment that

$$\begin{cases} \mathbb{E}\left|\sum_{i=2}^{B_n} y_{t,n,i}\zeta^{(3)}_{t,n,i}\right|^2 \leqslant K\Delta_n^{1/2} \\ \mathbb{E}\left|\sum_{i=2}^{B_n} y_{t,n,i}\zeta^{(3)}_{t,n,i-1}\right|^2 \leqslant K\Delta_n^{1/2} \\ \mathbb{E}\left|2\sum_{i=2}^{B_n} y_{t,n,i}\left(z^{(2)}_{t,n,i} + z^{(3)}_{t,n,i}\right) - \rho v\int_{\underline{t}}^{t} V_s ds\right|^2 \leqslant K\Delta_n^{1/2}. \end{cases} \tag{C.2.49}$$

Then (C.2.47) readily follows (C.2.48) and (C.2.49).

It remains to show (C.2.49), starting with the first inequality there. Observe that

$$\begin{aligned} \sum_{i=2}^{B_n} y_{t,n,i}\zeta^{(3)}_{t,n,i} &= \sum_{i=2}^{B_n}\left(y_{t,n,i}\zeta^{(3)}_{t,n,i} - \frac{1}{k_n\Delta_n}\int_{\tau(i-1)}^{\tau(i)}\left(X'_s - X'_{n,s}\right)V_s ds\right) \\ &\quad + \sum_{i=2}^{B_n}\frac{1}{k_n\Delta_n}\int_{\tau(i-1)}^{\tau(i)}\left(X'_s - X'_{n,s}\right)\left(V_s - V_{\tau(i-1)}\right)ds \\ &\quad + \sum_{i=2}^{B_n}\frac{1}{k_n\Delta_n}\sum_{j=1}^{k_n}\int_{\tau(i-1,j-1)}^{\tau(i-1,j)}\left(\int_{\tau(i-1,j-1)}^{s} b_u du\right)V_{\tau(i-1)}ds \\ &\quad + \sum_{i=2}^{B_n}\frac{1}{k_n\Delta_n}\sum_{j=1}^{k_n}\int_{\tau(i-1,j-1)}^{\tau(i-1,j)}\left(\int_{\tau(i-1,j-1)}^{s} \sigma_u dW_u\right)V_{\tau(i-1)}ds. \end{aligned}$$

By construction, the first term and the last term on the right-hand side of the above equation are sums of martingale differences. It is then easy to see

$$\mathbb{E}\left|\sum_{i=2}^{B_n}\left(y_{t,n,i}\zeta_{t,n,i}^{(3)}-\frac{1}{k_n\Delta_n}\int_{\tau(i-1)}^{\tau(i)}\left(X'_s-X'_{n,s}\right)V_s ds\right)\right|^2 \leqslant K\Delta_n^{1/2},$$

$$\mathbb{E}\left|\sum_{i=2}^{B_n}\frac{1}{k_n\Delta_n}\sum_{j=1}^{k_n}\int_{\tau(i-1,j-1)}^{\tau(i-1,j)}\left(\int_{\tau(i-1,j-1)}^{s}\sigma_u dW_u\right)V_{\tau(i-1)}ds\right|^2 \leqslant K\Delta_n.$$

Observing that $\mathbb{E}\left|V_t-V_s\right|^4\leqslant K\left|t-s\right|^2$ and $\mathbb{E}\left|X'_s-X'_{n,s}\right|^4\leqslant K\Delta_n^2$, we use Cauchy-Schwarz to derive

$$\mathbb{E}\left|\sum_{i=2}^{B_n}\frac{1}{k_n\Delta_n}\int_{\tau(i-1)}^{\tau(i)}\left(X'_s-X'_{n,s}\right)\left(V_s-V_{\tau(i-1)}\right)ds\right|^2\leqslant K\Delta_n^{1/2}.$$

It is also easy to see that

$$\mathbb{E}\left|\sum_{i=2}^{B_n}\frac{1}{k_n\Delta_n}\sum_{j=1}^{k_n}\int_{\tau(i-1,j-1)}^{\tau(i-1,j)}\left(\int_{\tau(i-1,j-1)}^{s}b_u du\right)V_{\tau(i-1)}ds\right|^2\leqslant K\Delta_n.$$

We combine the above estimates to derive $\mathbb{E}\left|\sum_{i=2}^{B_n}y_{t,n,i}\zeta_{t,n,i}^{(3)}\right|^2\leqslant K\Delta_n^{1/2}$, as asserted in (C.2.49). The second inequality in (C.2.49) can be proved in a similar (but simpler) way.

Now, recall the decomposition in (C.2.44). Observe that

$$\mathbb{E}\left|\sum_{i=2}^{B_n}\lambda_{t,n,i}^{(1)}-\rho v\int_{\underline{t}}^{t}V_s ds\right|^2 \leqslant K\Delta_n, \quad \mathbb{E}\left|\sum_{i=2}^{B_n}\lambda_{t,n,i}^{(2)}\right|^2\leqslant K\Delta_n$$

$$\mathbb{E}\left|\sum_{i=2}^{B_n}\lambda_{t,n,i}^{(3)}\right|^2 \leqslant K\Delta_n, \quad \mathbb{E}\left|\sum_{i=2}^{B_n}\lambda_{t,n,i}^{(4)}\right|^2\leqslant K\Delta_n^{1/2},$$

where the first inequality is a simple estimate for the Riemann approximation error for the Itô process V_t; the second inequality is obvious; the third inequality is obtained by using the fact that $\lambda_{t,,n,i}^{(3)}$ forms a martingale difference sequence; the fourth inequality is derived by using $\mathbb{E}[\lambda_{t,n,i}^{(4)}|\mathcal{F}_{\tau(i-2)}] = 0$ (hence the odd- and even-indexed terms respectively form martingale difference sequences). These estimates further imply $\mathbb{E}|2\sum_{i=2}^{B_n} y_{t,n,i} z_{t,n,i}^{(2)} - \rho v \int_{\underline{t}}^{t} V_s ds|^2 \leqslant K\Delta_n^{1/2}$. A similar (but simpler) argument yields $\mathbb{E}|\sum_{i=2}^{B_n} y_{t,n,i} z_{t,n,i}^{(3)}|^2 \leqslant K\Delta_n^{1/2}$. The third inequality of (C.2.49) then readily follows. This finishes the proof. □

Proof of Lemma 3.1.3. By localization, we can suppose that b_t, σ_t, $\tilde{b}_t$, $\tilde{\sigma}_t$ and $\tilde{\sigma}_t'$ are bounded without loss of generality. By Lemma C.2.1, $\Delta_n^{-1/4}(\widehat{CV}_{n,t} - \widehat{CV}'_{n,t}) = o_p(1)$. By Theorem 1 of Wang and Mykland (2013), $\Delta_n^{-1/4}(\widehat{CV}'_{n,t} - CV_{n,t})$ converges stably in law to some variable ζ, which conditionally on $\mathcal{F}$, is centered Gaussian with strictly positive variance (note that $\sigma_t > 0$ by assumption). The assertion of the lemma then readily follows. □

Proof of Lemma C.1.2. Applying Lemma C.2.1 with $r = 4/(9-12\varpi)$ and Lemma C.2.2, we have

$$\begin{cases} \mathbb{E}\left|\mathbb{E}\left[\Delta_n^{-1/4}\left(\widehat{CV}_{n,t} - \rho_0 v_0 \widehat{IV}_{n,t}\right)\Big|\mathcal{F}_{t-1}\right]\right| \leqslant K\Delta_n^{3\varpi-5/4}, \\ \mathbb{E}\left|\Delta_n^{-1/4}\left(\widehat{CV}_{n,t} - \rho_0 v_0 \widehat{IV}_{n,t}\right)\right|^r \leqslant K. \end{cases} \tag{C.2.50}$$

We then derive

$$\begin{aligned} \left|\mathbb{E}\left[\hat{m}_{n,t}^{+}(\theta_0)\right]\right| &= \left|\mathbb{E}\left[\frac{\mathbb{E}\left[\Delta_n^{-1/4}\left(\widehat{CV}_{n,t}-\rho_0 v_0 \widehat{IV}_{n,t}\right)\Big|\mathcal{F}_{t-1}\right]}{a_{IV}\widehat{IV}_{n,t-1}+b_{IV}}\right]\right| \\ &\leqslant K\mathbb{E}\left|\mathbb{E}\left[\Delta_n^{-1/4}\left(\widehat{CV}_{n,t}-\rho_0 v_0 \widehat{IV}_{n,t}\right)\Big|\mathcal{F}_{t-1}\right]\right| \\ &\leqslant K\Delta_n^{3\varpi-5/4}, \end{aligned}$$

where the equality follows definition (3.1.30) and repeated conditioning; the first inequality follows Jensen's inequality and $a_{IV}\widehat{IV}_{n,t-1}+b_{IV}\geqslant b_{IV}>0$; the second inequality is due to the first line of (C.2.50). This finishes the proof of part (a).

By the second inequality of (C.2.50), it is easy to see that $\mathbb{E}\left|\hat{m}_{n,t}^{+}(\theta_0)\right|^r\leqslant K$, as asserted in part (b). Moreover, since $r>1$, part (b) also implies that $\{\hat{m}_{n,t}(\theta_0)\}_{n\geqslant 1}$ is uniformly integrable. To show part (c), we define some constants:

$$q=\left(\frac{1}{r}+\frac{1}{2k'}\right)^{-1}, \quad p=\frac{2k'}{q}>1, \quad p'=\frac{r}{q}>1.$$

We then observe $\mathbb{E}\left|m_s^*(\theta_0)\,\hat{m}_{n,t}^{+}(\theta_0)\right|^q\leqslant\left(\mathbb{E}\left|m_s^*(\theta_0)\right|^{2k'}\right)^{1/p}\left(\mathbb{E}\left|\hat{m}_{n,t}^{+}(\theta_0)\right|^r\right)^{1/p'}\leqslant K$, where the first inequality follows Hölder's inequality, and the second inequality follows part (b) and Assumption B3. Finally, note that $k'>2/(12\varpi-5)$ implies $q>1$. Hence, $\{m_s^*(\theta_0)\,\hat{m}_{n,t}^{+}(\theta_0)\}_{n\geqslant 1}$ is also uniformly integrable as asserted in part (c). □

Bibliography

Adams, R. and Fournier, J. (2003), *Sobolev spaces*, vol. 140, Access Online via Elsevier.

Aït-Sahalia, Y. and Duarte, J. (2003), "Nonparametric option pricing under shape restrictions," *Journal of Econometrics*, 116, 9–47.

Aït-Sahalia, Y. and Lo, A. W. (1998), "Nonparametric estimation of state-price densities implicit in financial asset prices," *The Journal of Finance*, 53, 499–547.

Aït-Sahalia, Y., Wang, Y., and Yared, F. (2001), "Do option markets correctly price the probabilities of movement of the underlying asset?" *Journal of Econometrics*, 102, 67–110.

Aït-Sahalia, Y., K., M., and Mancini, L. (2012), "The term structure of variance swaps, risk premia and the expectation hypothesis," *Working Paper, Princeton University*.

Andersen, T., Benzoni, L., and Lund, J. (2002), "An Empirical Investigation of Continuous-Time Equity Return Models," *The Journal of Finance*, 57, 1239–1284.

Andersen, T. G., Bollerslev, T., Diebold, F., and Labys, P. (2003), "Modeling and forecasting realized volatility," *Econometrica*, 71, 579–625.

Andersen, T. G., Fusari, N., and Todorov, V. (2012), "Parametric inference and dynamic state recovery from option panels," Tech. rep., National Bureau of Economic Research.

Andersen, T. G., Fusari, N., and Todorov, V. (2013), "The Risk Premia Embedded in Index Options," Tech. rep., Working paper, Northwstern University.

Andrews, D. W. K. (1991), "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica: Journal of the Econometric Society*, pp. 817–858.

Andrews, D. W. K. and Cheng, X. (2012), "Estimation and Inference With Weak, Semi-Strong, and Strong Identification," *Econometrica*, 80, 2153–2211.

Andrews, D. W. K. and Guggenberger, P. (2009), "Validity of subsampling and plug-in asymptotic inference for parameters defined by moment inequalities," *Econometric Theory*, 25, 669–709.

Andrews, D. W. K. and Shi, X. (2014), "Nonparametric inference based on conditional moment inequalities," *Journal of Econometrics*, 179, 31–45.

Andrews, D. W. K. and Soares, G. (2010), "Inference for parameters defined by moment inequalities using generalized moment selection," *Econometrica*, 78, 119–157.

Arrow, K. J. (1964), "The role of securities in the optimal allocation of risk-bearing," *The Review of Economic Studies*, pp. 91–96.

Bakshi, G. and Madan, D. (2000), "Spanning and derivative-security valuation," *Journal of Financial Economics*, 55, 205–238.

Bakshi, G. and Madan, D. (2006), "A theory of volatility spreads," *Management Science*, 52, 1945–1956.

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

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

Bakshi, G., Panayotov, G., and Skoulakis, G. (2011), "Improving the predictability of real economic activity and asset returns with forward variances inferred from option portfolios," *Journal of Financial Economics*, 100, 475–495.

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

Barndorff-Nielsen, O. E. and Shephard, N. (2004), "Power and bipower variation with stochastic volatility and jumps," *Journal of Financial econometrics*, 2, 1–37.

Bates, D. (1996), "Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche Mark options," *Review of Financial Studies*, 9, 69–107.

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

Bates, D. S. (1991), "The crash of '87: Was it expected? The evidence from options markets," *The journal of Finance*, 46, 1009–1044.

Billingsley, P. (1995), "Probability and measure. 1995," *John Wiley&Sons, New York*.

Birru, J. and Figlewski, S. (2012), "Anatomy of a Meltdown: The Risk Neutral Density for the S&P 500 in the Fall of 2008," *Journal of Financial Markets*, 15, 151–180.

Black, F. (1976), "Studies of stock price volatility changes," *In: Proceedings of the 1976 Meetings of the American Statistical Association*, pp. 171–181.

Black, F. and Scholes, M. (1973), "The pricing of options and corporate liabilities," *The Journal of Political Economy*, pp. 637–654.

Bliss, R. and Panigirtzoglou, N. (2004), "Option-implied risk aversion estimates," *The Journal of Finance*, 59, 407–446.

Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity," *Journal of econometrics*, 31, 307–327.

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

Bollerslev, T. and Zhou, H. (2006), "Volatility puzzles: a simple framework for gauging return-volatility regressions," *Journal of Econometrics*, 131, 123–150.

Bollerslev, T., Tauchen, G., and Zhou, H. (2009), "Expected stock returns and variance risk premia," *Review of Financial Studies*, 22, 4463–4492.

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

Bollerslev, T., Osterrieder, D., Sizova, N., and Tauchen, G. (2013), "Risk and return: Long-run relations, fractional cointegration, and return predictability," *Journal of Financial Economics*.

Bondarenko, O. (2003), "Estimation of risk-neutral densities using positive convolution approximation," *Journal of Econometrics*, 116, 85–112.

Breeden, D. and Litzenberger, R. H. (1978), "Prices of state-contingent claims implicit in option prices," *Journal of business*, pp. 621–651.

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

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

Bugni, F. A. (2010), "Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set," *Econometrica*, 78, 735–753.

Carr, P. and Lee, R. (2008), "Robust replication of volatility derivatives," in *PRMIA award for Best Paper in Derivatives, MFA 2008 Annual Meeting*.

Carr, P. and Wu, L. (2009), "Variance risk premiums," *Review of Financial Studies*, 22, 1311–1341.

Chen, X. (2007), "Large sample sieve estimation of semi-nonparametric models," *Handbook of Econometrics*, 6, 5549–5632.

Chen, X. and Shen, X. (1998), "Sieve extremum estimates for weakly dependent data," *Econometrica*, pp. 289–314.

Chen, X., Liao, Z., and Sun, Y. (2013), "Sieve Inference on Semi-Nonparametric Time Series Models," *Cowles Foundation Discussion Paper*.

Chernov, M., Gallant, A., Ghysels, E., and Tauchen, G. (2003), "Alternative models for stock price dynamics," *Journal of Econometrics*, 116, 225–257.

Chernozhukov, V., Hong, H., and Tamer, E. (2007), "Estimation and confidence regions for parameter sets in econometric models1," *Econometrica*, 75, 1243–1284.

Christoffersen, P. and Jacobs, K. (2004), "The importance of the loss function in option valuation," *Journal of Financial Economics*, 72, 291–318.

Christoffersen, P., Heston, S., and Jacobs, K. (2006), "Option valuation with conditional skewness," *Journal of Econometrics*, 131, 253–284.

Christoffersen, P., Jacobs, K., Ornthanalai, C., and Wang, Y. (2008), "Option valuation with long-run and short-run volatility components," *Journal of Financial Economics*, 90, 272–297.

Conrad, J., Dittmar, R. F., and Ghysels, E. (2013), "Ex ante skewness and expected stock returns," *The Journal of Finance*, 68, 85–124.

Coppejans, M. and Gallant, A. R. (2002), "Cross-validated SNP density estimates," *Journal of Econometrics*, 110, 27–65.

Corsi, F. (2009), "A simple approximate long-memory model of realized volatility," *Journal of Financial Econometrics*, 7, 174–196.

Corsi, F., Pirino, D., and Reno, R. (2010), "Threshold bipower variation and the impact of jumps on volatility forecasting," *Journal of Econometrics*, 159, 276–288.

Cox, J. C., Ingersoll Jr, J. E., and Ross, S. A. (1985), "A theory of the term structure of interest rates," *Econometrica: Journal of the Econometric Society*, pp. 385–407.

Davidson, J. (1994), *Stochastic Limit Theory: An Introduction for Econometricians: An Introduction for Econometricians*, Oxford university press.

Debreu, G. (1959), *Theory of value: An axiomatic analysis of economic equilibrium*, no. 17, Yale University Press.

Drechsler, I. and Yaron, A. (2011), "What's vol got to do with it," *Review of Financial Studies*, 24, 1–45.

Duffie, D. (2001), *Dynamic asset pricing theory*, Princeton Univ Pr.

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

Engle, R. and Manganelli, S. (2004), "CAViaR: Conditional autoregressive value at risk by regression quantiles," *Journal of Business & Economic Statistics*, 22, 367–381.

Eraker, B. (2004), "Do stock prices and volatility jump? Reconciling evidence from spot and option prices," *The Journal of Finance*, 59, 1367–1404.

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

Exchange, C. F. (2013), "S&P 500 Variance Futures: Exchange-Traded/OTC Conventions," *CBOE Risk Management Conference, Presented by J. Hiatt.*

Fackler, P. (2005), "Notes on matrix calculus," *North Carolina State University.*

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

Figlewski, S. (2008), "Estimating the Implied Risk Neutral Density," *Volatility and Time Series Econometrics: Essay in Honor of R. F. Engle. Editors: T. Bollerslev, J. R. Russell, M. Watson, Oxford, UK: Oxford University Press.*

Fusari, N. and Gonzalez-Perez, M. T. (2012), "Volatility Dynamics and the Term Structure of the Variance Risk Premium," *Northwestern University.*

Gallant, A. R. and Nychka, D. W. (1987), "Semi-nonparametric maximum likelihood estimation," *Econometrica: Journal of the Econometric Society*, pp. 363–390.

Garcia, R. and Gençay, R. (2000), "Pricing and hedging derivative securities with neural networks and a homogeneity hint," *Journal of Econometrics*, 94, 93–115.

Glosten, L. R. and Milgrom, P. (1985), "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders," *Journal of Financial Economics*, 14, 71–100.

Hasbrouck, J. (2006), *Empirical market microstructure: The institutions, economics, and econometrics of securities trading*, Oxford University Press.

Hayashi, F. (2000), *Econometrics*, Princeton University Press. Section.

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

Hull, J. and White, A. (1987), "The pricing of options on assets with stochastic volatilities," *The Journal of Finance*, 42, 281–300.

Jackwerth, J. C. and Rubinstein, M. (1996), "Recovering probability distributions from option prices," *The Journal of Finance*, 51, 1611–1631.

Jacod, J. and Protter, P. E. (2012), *Discretization of processes*, vol. 67, Springer.

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

Kiernan, K. (2014), "Investors Arm for Imminent Tremors," *The Wall Street Journal*, Feb. 6, 2014.

Kristensen, D. and Mele, A. (2011), "Adding and subtracting Black-Scholes: a new approach to approximating derivative prices in continuous-time models," *Journal of Financial Economics*, 102, 390–415.

León, Á.and Mencía, J. and Sentana, E. (2009), "Parametric properties of semi-nonparametric distributions, with applications to option valuation," *Journal of Business & Economic Statistics*, 27, 176–192.

Mancini, C. (2009), "Non-parametric Threshold Estimation for Models with Stochastic Diffusion Coefficient and Jumps," *Scandinavian Journal of Statistics*, 36, 270–296.

Merton, R. C. (1973), "Theory of rational option pricing," *The Bell Journal of Economics and Management Science*, 4, 141–183.

Metaxoglou, K. and Smith, A. (2011), "State Prices of Conditional Quantiles: New Evidence on Time-Varying Expected Returns," *Available at SSRN 1950263.*

Metaxoglou, K. and Smith, A. (2013), "State Prices of Conditional Quantiles: New Evidence on Time-Varying Expected Returns," *Available at SSRN 1950263.*

Newey, W. K. (1991), "Uniform convergence in probability and stochastic equicontinuity," *Econometrica*, 59, 1161–1167.

Newey, W. K. (1997), "Convergence rates and asymptotic normality for series estimators," *Journal of Econometrics*, 79, 147–168.

Nickl, R. and Pötscher, B. M. (2007), "Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov-and Sobolev-type," *Journal of Theoretical Probability*, 20, 177–199.

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

Romano, J. P. and Shaikh, A. M. (2008), "Inference for identifiable parameters in partially identified econometric models," *Journal of Statistical Planning and Inference*, 138, 2786–2807.

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

Royden, H. and Fitzpatrick, P. (2010), "Real Analysis. 2010," .

Shen, X. (1997), "On methods of sieves and penalization," *The Annals of Statistics*, 25, 2555–2591.

Stock, J. H. and Wright, J. H. (2000), "GMM with weak identification," *Econometrica*, 68, 1055–1096.

Van De Geer, S. (2000), *Empirical processes in M-estimation*, vol. 45, Cambridge university press Cambridge.

Van Der Vaart, A. W. and Wellner, J. A. (1996), *Weak Convergence*, Springer.

Wang, C. D. and Mykland, P. A. (2013), "The estimation of leverage effect with high frequency data," *Journal of the American Statistical Association*, p. forthcoming.

Xiu, D. (2011), "Dissecting and Deciphering European Option Prices using Closed-Form Series Expansion," *Chicago Booth Research Paper*.

Yatchew, A. and Härdle, W. (2006), "Nonparametric state price density estimation using constrained least squares and the bootstrap," *Journal of Econometrics*, 133, 579–599.

Biography

Erik Vogt was born in Manila, Philippines, on March 13, 1985. He attended the *Deutsche Schule Manila* before moving to Hesperange, Luxembourg, in 1993 and then on to Los Angeles, California, in 1999, where he attended the Palos Verdes Peninsula High School. After spending two years (2003-2005) at the University of California, Los Angeles, Erik transferred to the London School of Economics to earn his BSc. in Mathematics and Economics in 2007. From 2007 to 2009, Erik worked as an Associate Economist at the Federal Reserve Bank of Chicago's Macro Research Group. From 2009 to 2014, Erik lived in Durham, North Carolina, where he obtained his MA in Economics at Duke University in 2010 and his PhD in Economics in 2014. He will be joining the Federal Reserve Bank of New York as an Economist in the Capital Markets Function of the Research and Statistics Group in the summer of 2014.