Federal Reserve Bank of New York
Staff Reports

Option-Implied Term Structures

Erik Vogt

Staff Report No. 706
December 2014
Revised January 2016

Option-Implied Term Structures
Erik Vogt
Federal Reserve Bank of New York Staff Reports, no. 706
December 2014; revised January 2016
JEL classification: G12, G17, C58

Abstract

This paper proposes a nonparametric sieve regression framework for pricing the term structure of option spanning portfolios. The framework delivers closed-form, nonparametric option pricing and hedging formulas through basis function expansions that grow with the sample size. Novel confidence intervals quantify term structure estimation uncertainty. The framework is applied to estimating the term structure of variance risk premia and finds that a short-run component dominates market excess return predictability. This finding is inconsistent with existing asset pricing models that seek to explain the variance risk premium's predictive content.

Key words: variance risk premium, term structures, options, return predictability, nonparametric regression.

Vogt: Federal Reserve Bank of New York (e-mail: erik.vogt@ny.frb.org). The author is especially thankful to his dissertation chair, George Tauchen, and his dissertation committee members, Tim Bollerslev, Federico Bugni, Jia Li, and Andrew Patton, for their guidance and encouragement. He is also grateful to Tobias Adrian, Torben Andersen, Gurdip Bakshi, Nina Boyarchenko, Michael Brandt, Peter Christoffersen, Richard Crump, Ron Gallant, Steve Heston, Roger Koenker, Ralph Koijen, Pete Kyle, Simon (Sokbae) Lee, Adam McCloskey, Eric Renault, Matthew Richardson, Alberto Rossi, Enrique Sentana, Peter Van Tassel, and Dacheng Xiu, as well as seminar participants at Duke, the Federal Reserve Bank of New York, the Federal Reserve Board, Brown University, London Business School, the University of Illinois at Urbana-Champaign, the University of Maryland, the 2013 Triangle Econometrics Conference, AQR, Citadel, PDT Partners, QMS Capital Management, and the 2015 Econometric Society World Congress, for valuable comments and suggestions. All errors are those of the author. The views expressed in this paper are those of the author and are not necessarily reflective of views at the Federal Reserve Bank of New York or the Federal Reserve System.

1 Introduction

When an asset's cash flows are expected to occur at multiple future dates, the prices of each individual cash flow form a term structure. While term structures are traditionally studied for assets with a predetermined sequence of cash flows (as with risk-free bonds), a flurry of recent research has expanded into studying the term structures of a wider array of risky assets.[1] This paper studies term structures of risky assets whose cash flows depend on future realized moments of another reference security, with a particular focus on realized variances, their prices, and risk premia.

Measuring the prices of long-run variance and other realized moments is challenging. While theoretical hedging arguments show that integrated portfolios of call and put options price (or "span") these moments, long-dated options are generally illiquid, meaning that the valuations obtained from integrated hedging portfolios must be inferred from a handful of noisy options. How well do illiquid, long-dated option-spanning portfolios approximate risk-neutral expectations of long-run realized moments? Understanding the answer to this question is of first-order importance for interpreting risk premia in the term structures of realized moments.

This paper makes a methodological contribution toward measuring long-run risk-neutral expectations of variance and other moments from options. By forming basis function expansions of the state-price density, the paper derives new closed-form option prices that can be interpreted as nonparametric extensions of the Black and Scholes (1973) formula. While these option prices are potentially of independent interest (for example for hedging purposes), they are particularly suited to the problem of estimating the term structures of risk-neutral moments: First, they inject theoretical structure into estimated option prices while remaining nonparametric. This theoretical structure helps connect estimates of long-run option spanning portfolios to the information contained in liquid short-maturity options. Second, they allow for a novel theory of inference that allows one to obtain confidence intervals for the term structures of realized moments, thereby quantifying the precision with which long-run risk-neutral moments can be priced with illiquid options. In a first application of the inference result, we find that there is substantially higher estimation uncertainty around long-run risk-neutral skewness and kurtosis than for long-run risk-neutral variance.[2] Moreover, we

[1] See Van Binsbergen and Koijen (2015) for a survey.

[2] See Conrad, Dittmar, and Ghysels (2013) and Bakshi, Kapadia, and Madan (2003) for studies of skewness and

find that the term structures of second, third, and fourth risk-neutral moments have a strong factor structure that loads heavily on variance factors (i.e. the VIX term structure).

The paper's empirical focus on variance is therefore motivated by the factor structure result on higher order moments, as well as recent research on the term structure of variance risk premia: by examining Sharpe ratios of trading strategies that effectively sell variance along its term structure, Dew-Becker, Giglio, Le, and Rodriguez (2015) have found that only news about short-run realized variance is priced. In related work, Andries, Eisenbach, Schmalz, and Wang (2015) use the Heston (1993) model and a short option straddle to estimate the term structure of variance risk premia, an object also studied in Aït-Sahalia, Karaman, and Mancini (2015), who find a significant price-jump component in their own estimated term structure of variance risk premia. These papers join a growing literature on understanding risk premia in the term structures of assets beyond fixed income which includes van Binsbergen, Brandt, and Koijen (2012) and van Binsbergen, Hueskes, Koijen, and Vrugt (2013), who decompose the equity premium into its term structure components. All of these papers find that risk premia are downward sloping in absolute value, that is, that a short-run component dominates equity and variance risk premia. Since several leading asset pricing benchmarks have counterfactual implications for the term structure of risk premia, these findings present something of a puzzle to current models in finance.[3]

This paper presents new findings on the relationship between variance risk premia and equity market return predictability. Motivated by Bollerslev, Tauchen, and Zhou (2009)'s and Drechsler and Yaron (2011)'s findings that the one-month variance risk premium (VRP) predicts equity market excess returns, we examine the predictive ability of long-run variance risk premia. We find that while long-run variance risk premia strongly predict returns, the predictability is largely generated by a short-run component. This is the result of predictive regressions that additively decompose long-run VRP into the sum of one-month VRP and a *forward* VRP that represents compensation that investors earn for exposure to variance risk beyond the first month. The predictive results connect the findings of Dew-Becker, Giglio, Le, and Rodriguez (2015) and Andries, Eisenbach, Schmalz, and Wang (2015) to the findings of van Binsbergen, Brandt, and Koijen (2012) by showing that

kurtosis option-spanning portfolios that are based on the spanning theorems of Bakshi and Madan (2000).

[3]For instance, the long-run risk model of Bansal and Yaron (2004) and the habit formation model of Campbell and Cochrane (1999) have equity risk premia that are counterfactually upward sloping in the term structure, and the long-run risk model of Drechsler and Yaron (2011) has a variance risk premium term structure that does not account for the steep risk premium at the one month horizon and subsequent flatness for remaining horizons.

discount rates for short-term variance cash flows are tightly linked to those of the equity term structure, which in turn are tied to discount rates of short-term dividend cash flows.

The predictive results present a challenge to the return predictability theories of Bollerslev, Tauchen, and Zhou (2009) and Drechsler and Yaron (2011). In the Bollerslev, Tauchen, and Zhou (2009) equilibrium framework, for example, recursive preferences combined with time-varying volatility of consumption volatility (consumption vol-of-vol) imply that the equity premium and the variance risk premium are linked by the same consumption vol-of-vol process q_t. Similarly, in the extended long-run risk model of Drechsler and Yaron (2011), this link is established by a time-varying aggregate stochastic volatility process σ_t^2 that is common to all state variables in their economy. Because it can be shown that the same process (q_t for Bollerslev, Tauchen, and Zhou (2009) and σ_t^2 for Drechsler and Yaron (2011)) also governs variation in *forward* VRP, both models suggest that forward VRP should also predict equity market excess returns. However, the predictive results presented in Section 4 below show that this is not the case, since the forward VRP by itself does not predict returns. The predictive results suggest that the factors driving time variation in the one-month VRP must be distinct from those that drive forward VRP and, furthermore, that only the one-month VRP factors are priced in equity returns. Therefore, the empirical results presented in this paper provide guidance for equilibrium asset pricing models that seek to explain the link between discount rates for variance-linked cash flows and discount rates for equity cash flows.

To study the long-run properties of risk-neutral realized moments, this paper presents the first sieve application to option pricing. The method of sieves is a nonparametric alternative to kernel methods that have been used in option pricing (e.g. Aït-Sahalia and Lo (1998)) and can therefore be viewed as complementary. One advantage of sieve-based methods is that many quantities in finance are concerned with computing expectations where the only object missing is the density over possible future states, but the remaining objects involved in the expectation are known. Sieve-based methods allow the direct expansion of the unknown density, and hence should present a useful alternative tool to computing the many expectations that are prevalent in finance. Moreover, since the method of sieve allows some discretion in the choice of basis function expansion, the results are often closed-form expressions. One such expression is presented in this paper, in which option prices appear as Black-Scholes prices plus higher order expansion terms that account for non-lognormality of the underlying asset. Formulas of this type additionally have implications for hedging, since the expansion terms correct local misspecification of Black-Scholes option prices, to the extent that non-lognormalities are present.

The outline of the paper is as follows. Section 2 presents the methodological contribution for measuring option-implied term structures. Section 3 shows its validity in a brief simulation exercise, and Section 4 presents the main empirical findings.

2 Measuring Long-Run Risk-Neutral Expectations

This section develops a framework for estimating long-run risk-neutral expectations of realized moments that explicitly recognizes illiquidity issues in long-maturity options. For exposition, the focus is on risk-neutral variance; higher order moments are presented at the end of this section. The general idea in what follows is to exploit existing option-spanning relationships together with shape information in options.

Standing at time 0, consider the problem of measuring the risk-neutral expectation of realized variance $E_0^Q[RV_{0,\tau}]$, where $RV_{0,\tau}$ measures the sum of squared returns of a reference asset from now until time τ. Well-known spanning results imply that the synthetic variance swap (SVS) provides a good hedge, i.e. $SVS_0(\tau) = E_0^Q[RV_{0,\tau}]$ up to a third-order approximation error (Carr and Wu (2009)). For a given time horizon τ, the SVS is obtained by combining European put and call options with different strikes κ and common maturity τ into a single portfolio. Letting $\mathbf{Z} = (\kappa, \tau, r, q)$, where $r \equiv r(\tau)$ and $q \equiv q(\tau)$ correspond to the risk-free rate and underlying dividend yield at the maturity τ of interest, the SVS term structure is the function

$$SVS_0(\tau) = \frac{2}{\tau} e^{r\tau} \int_0^{F(\mathbf{Z})} \frac{1}{\kappa^2} P_0(\mathbf{Z}) d\kappa + \frac{2}{\tau} e^{r\tau} \int_{F(\mathbf{Z})}^{\infty} \frac{1}{\kappa^2} C_0(\mathbf{Z}) d\kappa, \tag{1}$$

where $F(\mathbf{Z}) = S_0 e^{(r-q)\tau}$ is the forward price, S_0 is the current (fixed) stock price, $P_0(\mathbf{Z})$ is the put option price with characteristics $\mathbf{Z}$, and $C_0(\mathbf{Z}) = P_0(\mathbf{Z}) + S_0 e^{-q\tau} - \kappa e^{-r\tau}$ is the call option price by put-call parity. We therefore need to evaluate $P_0(\mathbf{Z})$ at arbitrary τ across an infinite continuum of κ in order to get at the portfolio term structure $SVS_0(\tau)$ and thereby measure $E_0^Q[RV_{0,\tau}]$.

Because $P_0(\mathbf{Z})$ (and therefore $C_0(\mathbf{Z})$ by put-call parity) is unobserved, it must be estimated from a sample of put option prices and characteristics $\{P_i, \mathbf{Z}_i\}_{i=1}^n$. Table 1 shows that a typical

Table 1: Sample Averages for Monthly S&P 500 Index Options, 1996-2013.

Maturity Range (Days)	Number of Options	Option Volume	Open Interest	Bid-Ask Spread ($)
0 - 90	196.2	188 952.6	2 403 231.7	0.96
90 - 180	59.6	20 325.3	702 530.5	1.19
180 - 270	45.4	8626.8	434 287.5	1.38
270 - 360	42.4	4770.0	260 077.3	1.61
360 - 450	16.8	2180.6	161 567.9	1.97
450 - 540	16.5	1131.8	122 241.0	2.13
540 - 630	12.6	751.0	83 417.1	2.35
630 - 720	12.3	810.5	63 685.1	2.42
720 - ∞	18.7	1521.4	90 235.1	4.46
0 - ∞	420.9	229 072.0	4 321 320.0	1.35

cross-section of S&P 500 index options, some of the most liquid option contracts available, contains about $n = 420$ prices, with most of the observations concentrated at short maturities. The thinning of available option quotes for increasing τ is also associated with smaller trading volumes, less open interest, and widening bid-ask spreads.[4] The widening spreads introduce varying levels of uncertainty about $P_0(\mathbf{Z})$ for different τ, which we allow for by letting $P_i = P_0(\mathbf{Z}_i) + \varepsilon_i$ for ε_i a conditionally mean-zero, heteroskedastic measurement error. In this context, we refer to $P_0(\mathbf{Z}_i)$ as representing the true option price for an option with characteristics $\mathbf{Z}_i$. Collectively, we follow the literature and refer to the thinning of prices and increased noise as manifestations of illiquidity at longer maturities.[5]

2.1 Theoretical Structure to Inform Long-Run Expectations

To preserve the SVS's interpretation as a model-free spanning portfolio, $P_0(\mathbf{Z})$ is estimated nonparametrically. However, the illiquidity of long-maturity options suggests that they are less informative about long-run SVS, requiring the need for additional information. For added structure, the proposed framework therefore relies on the risk-neutral valuation equation, which states that there exists a conditional density f_0 such that $P_0(\mathbf{Z}) = P(f_0, \mathbf{Z})$ for known $P(\cdot, \mathbf{Z})$.

[4]Note that Table 1 reports dollar bid-ask spreads because dollar values enter the integral in (1). The data set follows the CBOE data filters and is discussed further in Section 4.

[5]See, e.g. Aït-Sahalia, Karaman, and Mancini (2015) and Driessen, Maenhout, and Vilkov (2009) for papers that express concerns about this illiquidity when studying long-run risk-neutral moments.

Formally, for a vector of characteristics $\mathbf{Z} = (\kappa, \tau, r, q)$, the true option price is modeled as

$$P_0(\mathbf{Z}) \equiv e^{-r\tau} E_0^{\mathbb{Q}}\Big[\big[\kappa - S\big]_+ \Big| \tau, \mathbf{V} = \mathbf{v}_0\Big] = e^{-r\tau} \int_0^{\kappa} [\kappa - S] f_0^{\mathbb{Q}}(S|\tau, \mathbf{V} = \mathbf{v}_0) dS, \tag{2}$$

where $\mathbf{V}$ is a vector of state variables that generate the current information set, $f_0^{\mathbb{Q}}(\,\cdot\,|\tau, \mathbf{V} = \mathbf{v}_0)$ is the unobserved state-price density (SPD), r is the risk-free rate, and S is the random (future) value of the underlying. 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, Pan, and Singleton (2000)).

Since the data represent an option cross-section at a single point in time, $\mathbf{V}$ realizes to some fixed value $\mathbf{V} = \mathbf{v}_0$. To simplify notation, we 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

$$P_0(\mathbf{Z}) \equiv e^{-r\tau} \int_0^{\kappa} [\kappa - S] f_0^{\mathbb{Q}}(S|\tau)\, dS. \tag{3}$$

The dependence of the option price on the SPD $f_0^{\mathbb{Q}}$ and the characteristics $\mathbf{Z}$ can be expressed as $P_S(f_0^{\mathbb{Q}}, \mathbf{Z}) \equiv P_0(\mathbf{Z})$.

The no-arbitrage pricing equation (3) implies shape restrictions on the option prices. Differentiating $P_S(f_0^{\mathbb{Q}}, \mathbf{Z})$ repeatedly with respect to the strike price κ yields the conditions $\frac{\partial P_S}{\partial \kappa} = e^{-r\tau} F_0^{\mathbb{Q}}(\kappa|\tau)$ and $\frac{\partial^2 P_S}{\partial \kappa^2} = e^{-r\tau} f_0^{\mathbb{Q}}(\kappa|\tau)$, where $F_0^{\mathbb{Q}}$ is the CDF of $f_0^{\mathbb{Q}}$. These conditions immediately imply that $P_S(f_0^{\mathbb{Q}}, \mathbf{Z})$ 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.[6]

Since the option price's shape constraints are implied by the fact that $f_0^{\mathbb{Q}}$ is a PDF, the strategy employed for obtaining shape-conforming option price estimates is the use of basis function

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

expansions that are valid PDFs. However, instead of approximating $f_0^{\mathbb{Q}}$ directly, it turns out that a convenient change of variables will lead to theoretically appealing closed-form option prices. To this end, let Y be the random variable that satisfies

$$\log\left(\frac{S}{S_0}\right) = \mu(\mathbf{Z}) + \sigma(\mathbf{Z})Y, \tag{4}$$

where $Y|\tau$ has density $f_0(y|\tau)$, 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. This change of variables is always possible for $S > 0$ because for any $\mu(\cdot)$ and $\sigma(\cdot)$, Y simply is the variable that makes (4) hold.

Under this change of variables, the valuation equation (3) becomes

$$P_S(f_0^{\mathbb{Q}}, \mathbf{Z}) = e^{-r\tau}\int_0^{\kappa}\big(\kappa - S\big)f_0^{\mathbb{Q}}(S|\tau)\,dS = e^{-r\tau}\int_0^{d(\mathbf{Z})}\Big(\kappa - S_0e^{\mu(\mathbf{Z})+\sigma(\mathbf{Z})Y}\Big)f_0(Y|\tau)\,dY \equiv P(f_0, \mathbf{Z}), \tag{5}$$

where

$$d(\mathbf{Z}) = \frac{\log(\kappa/S_0) - \mu(\mathbf{Z})}{\sigma(\mathbf{Z})} \tag{6}$$

and

$$f_0^{\mathbb{Q}}(\kappa|\tau) = (\kappa\sigma(\mathbf{Z}))^{-1}f_0(d(\mathbf{Z})|\tau) \tag{7}$$

follow from a Jacobian transformation.

Since (5) says $P_S(f_0^{\mathbb{Q}}, \mathbf{Z}) = P(f_0, \mathbf{Z})$, one can focus on option pricing equations of the form

$$P(f, \mathbf{Z}) = e^{-r\tau}\int_0^{d(\mathbf{Z})}\Big(\kappa - S_0e^{\mu(\mathbf{Z})+\sigma(\mathbf{Z})Y}\Big)f(Y|\tau)\,dY. \tag{8}$$

It is easy to verify that (8) satisfies the same shape restrictions as (3) for any f with $f(y|\tau) \geq 0$ and $\int f(y|\tau)dy = 1$.

2.2 Estimation and Option Pricing

Because the shape information comes from the fact that f_0 is a density for any τ, we only need to consider $P(f, \mathbf{Z})$ for candidates f that are valid densities. Denote the collection of these candidate densities by $\mathcal{F}$. The method of sieves then implies that option prices can be estimated by solving

$$\widehat{f}_{K_n} = \arg \min_{f \in \mathcal{F}_{K_n}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \Big[P_i - P(f, \mathbf{Z}_i) \Big]^2 W(\mathbf{Z}_i) \right\}, \tag{9}$$

where $\mathcal{F}_{K_n} \subset \mathcal{F}$ is a member of a sequence of compact approximating spaces $\{\mathcal{F}_K\}_{K=1}^{\infty}$ that grow slowly in dimension ($K_n \to \infty$) as the sample size $n \to \infty$, eventually becoming dense in $\mathcal{F}$. Informally, as the sample size grows, the approximating spaces $\mathcal{F}_{K_n}$ increasingly resemble the parent space $\mathcal{F}$, so that the solution on $\mathcal{F}_{K_n}$ should converge to f_0. Proposition 3 in the Appendix makes this notion precise and further shows that the option price estimates $\widehat{P}(\mathbf{Z}) \equiv P(\widehat{f}_{K_n}, \mathbf{Z})$ also converge to the true option price $P(f_0, \mathbf{Z})$.[7]

The elements of $\mathcal{F}_{K_n}$ are obtained as follows. Since Gallant and Nychka (1987) have shown that squared Hermite polynomials are suitable approximations for smooth joint densities, we can use their densities,

$$f_K^{Y,\tau}(y, \tau) = \left[\sum_{k=0}^{K_y} \left(\sum_{j=0}^{K_\tau} \beta_{kj} H_j(\tau) \right) H_k(y) \right]^2 e^{-\tau^2/2} e^{-y^2/2} = \left[\sum_{k=0}^{K_y} \alpha_k(B, \tau) H_k(y) \right]^2 e^{-\tau^2/2} e^{-y^2/2}, \tag{10}$$

where B is a matrix of coefficients with kj-entry β_{kj} and $K = (K_y + 1)(K_\tau + 1)$.[8] Thus, we can form ratios to obtain closed-form conditional densities

$$f_K(y|\tau) = f_K^{Y,\tau}(y, \tau) \Big/ \int f_K^{Y,\tau}(y, \tau) dy = \sum_{k=0}^{2K_y} \gamma_k(B, \tau) H_k(y) \phi(y), \tag{11}$$

[7] As written, the objective function in (9) is in dollar levels, which weights at-the-money options highest. To see this, let $I(\mathbf{Z}_i) = e^{r\tau} \max[\kappa_i - S_0, 0]$ denote the intrinsic value of option i. Then $[P_i - P(f, \mathbf{Z}_i)]^2 = [\{P_i - I(\mathbf{Z}_i)\} - \{P(f, \mathbf{Z}_i) - I(\mathbf{Z}_i)\}]^2$. Because $\{P_i - I(\mathbf{Z}_i)\}$ and $\{P(f, \mathbf{Z}_i) - I(\mathbf{Z}_i)\}$ assume their largest values at-the-money, the objective function in (14) is most sensitive to deviations at-the-money. If a different weighting is desired, the function $W(\mathbf{Z}_i)$ can be used instead. For example, by setting $W(\mathbf{Z}_i)$ to the inverse of option i's squared vega, one can approximate implied volatility errors (e.g., Christoffersen, Fournier, and Jacobs (2013)). However, for inference problems related to option-implied term structures, the Monte Carlo simulations provided in the Online Appendix suggest that simply setting $W(\mathbf{Z}_i) = 1$ yields superior coverage properties.

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

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

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

where the second equality is derived in Lemma B.1 and $\gamma_k(B,\tau)$ is a known function. Consequences of this choice of basis functions are closed-form option prices.

Proposition 1. *For a candidate SPD $f_K(y|\tau) \in \mathcal{F}_K$ of the form given in equation (40), the put option price $P(f_K, \mathbf{Z})$ from equation (8) is given by*

$$\begin{aligned} P(f_K,\mathbf{Z}) = \kappa e^{-r\tau}&\left[\Phi(d(\mathbf{Z})) - \sum_{k=1}^{2K_y}\frac{\gamma_k(B,\tau)}{\sqrt{k}}H_{k-1}(d(\mathbf{Z}))\phi(d(\mathbf{Z}))\right] \\ &- S_0 e^{-r\tau+\mu(\mathbf{Z})}\left[e^{\sigma(\mathbf{Z})^2/2}\Phi(d(\mathbf{Z})-\sigma(\mathbf{Z})) + \sum_{k=1}^{2K_y}\gamma_k(B,\tau)I_k^*(d(\mathbf{Z}))\right] \end{aligned} \tag{12}$$

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

$$\begin{aligned} I_k^*(d(\mathbf{Z})) &= \frac{\sigma(\mathbf{Z})}{\sqrt{k}}I_{k-1}^*(d(\mathbf{Z})) - \frac{1}{\sqrt{k}}e^{\sigma(\mathbf{Z})d(\mathbf{Z})}H_{k-1}(d(\mathbf{Z}))\phi(d(\mathbf{Z})), \qquad \textit{for } k \geq 1, \\ I_0^*(d(\mathbf{Z})) &= e^{\sigma(\mathbf{Z})^2/2}\Phi(d(\mathbf{Z})-\sigma(\mathbf{Z})), \end{aligned}$$

and $\gamma_k(B,\tau)$ is the coefficient function given in equation (40).

The price of a call option is given by

$$\begin{aligned} C(f_K,\mathbf{Z}) = S_0 e^{-r\tau+\mu(\mathbf{Z})}&\left[e^{\sigma(\mathbf{Z})^2/2}[1-\Phi(d(\mathbf{Z})-\sigma(\mathbf{Z}))] - \sum_{k=1}^{2K_y}\gamma_k(B,\tau)I_k^*(d(\mathbf{Z}))\right] \\ &- \kappa e^{-r\tau}\left[[1-\Phi(d(\mathbf{Z}))] - \sum_{k=1}^{2K_y}\frac{\gamma_k(B,\tau)}{\sqrt{k}}H_{k-1}(d(\mathbf{Z}))\phi(d(\mathbf{Z}))\right]. \end{aligned} \tag{13}$$

Proof. Appendix B. □

With this result in hand, (9) becomes computationally equivalent to nonlinear least squares with known regressors,

$$\widehat{\beta}_n = \arg\min_{\beta\in\mathbb{R}^{K_n}}\left\{\frac{1}{n}\sum_{i=1}^{n}\Big[P_i - P(\beta,\mathbf{Z}_i)\Big]^2 W(\mathbf{Z}_i)\right\}, \tag{14}$$

where $\beta \equiv vec(B)$ and where $P(\beta,\mathbf{Z}) \equiv P(f_K,\mathbf{Z})$ is the option pricer in (12) as a function of the basis function coefficients. Finally, it can be shown that by imposing $\sum_{k=0}^{K_y(n)} \sum_{j=0}^{K_\tau(n)} \beta_{kj}^2 = 1$ during estimation, the $\hat{f}_K(y|\tau)$ will integrate to one for all τ, resulting in option prices $\widehat{P}(\mathbf{Z})$ that incorporate shape information (as required for long maturities).

2.3 Inference for Risk-Neutral Term Structures

Estimates of the SVS term structure are then obtained by evaluating

$$\widehat{SVS}(\tau) = \frac{2}{\tau} e^{r\tau} \int_0^{F(\mathbf{Z})} \frac{1}{\kappa^2} \widehat{P}(\mathbf{Z}) d\kappa + \frac{2}{\tau} e^{r\tau} \int_{F(\mathbf{Z})}^{\infty} \frac{1}{\kappa^2} \widehat{C}(\mathbf{Z}) d\kappa, \tag{15}$$

point-wise in τ, where $\widehat{C}(\mathbf{Z})$ is obtained from $\widehat{P}(\mathbf{Z})$ by put-call parity. Moreover, because $\widehat{P}(\mathbf{Z})$ can be evaluated for any $\mathbf{Z} = (\kappa, \tau, r, q)$, one can obtain projections for $\widehat{SVS}(\tau)$ for unobserved τ.

It is worth emphasizing that the SVS term structure is a special case of the proposed framework, which applies to the general class of portfolios

$$\Gamma(\widehat{P}) = g\left(\int_{\mathcal{Z}_1} a(\mathbf{Z}_1, \mathbf{Z}_2) \widehat{P}(\mathbf{Z}_1, \mathbf{Z}_2) d\mathbf{Z}_1 + \int_{\mathcal{Z}_1^c} b(\mathbf{Z}_1, \mathbf{Z}_2) \widehat{C}(\mathbf{Z}_1, \mathbf{Z}_2) d\mathbf{Z}_1 \right), \tag{16}$$

where $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)'$ and $\mathcal{Z}_1$ is a subset of the domain of $\mathbf{Z}_1$.[9] This class of portfolios encompasses many objects of interest beyond the SVS, and can include e.g. the skewness and kurtosis portfolios of Bakshi, Kapadia, and Madan (2003), in which case $\mathbf{Z}_1 = \kappa$. Because portfolios of this form represent regular functionals in the sense of Chen, Liao, and Sun (2014), derivation of an asymptotic distribution for this class, including $\widehat{SVS}(\tau)$ or its square-root $\widehat{VIX}(\tau)$, is an application of their theory. Proposition 2 below shows how this theory can be used to establish results of the form $\sqrt{n}\widehat{V}^{-1/2}(\widehat{SVS}(\tau) - SVS_0(\tau)) \rightarrow^d N(0,1)$.

Proposition 2. *Under Assumptions A.1 and B.1–B.5,*

$$\sqrt{n}\widehat{V}_n^{-1/2}[\Gamma(\widehat{P}_n) - \Gamma(P_0)] \xrightarrow{d} N(0,1) \tag{17}$$

where

$$\widehat{V}_n = \widehat{G}'_{K_n} \widehat{R}_{K_n}^{-1} \widehat{\Sigma}_{K_n} \widehat{R}_{K_n}^{-1} \widehat{G}_{K_n} \tag{18}$$

[9] The function g is added for convenience to allow for transformations of the option spanning portfolio of interest, e.g. $VIX = 100\sqrt{SVS}$, in which case $g(x) = 100\sqrt{x}$. In many applications, $g(x) = x$.

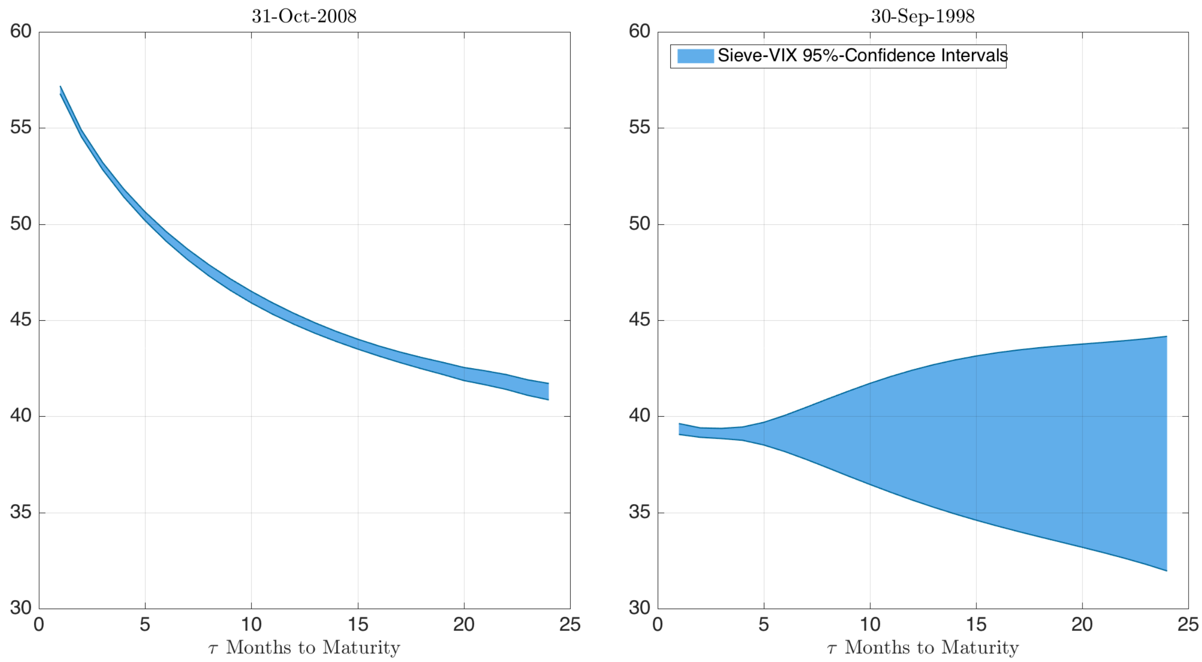

Figure 1: This figure plots sieve-estimated $VIX(\tau) = 100\sqrt{SVS(\tau)}$ term structures along with their 95-% confidence intervals, obtained from Proposition 2. The underlying data are the full set of S&P 500 index options available on the indicated dates.

and where

$$\widehat{G}_{K_n} = \frac{\partial \Gamma(P(\widehat{\beta}_n, \mathbf{Z}))}{\partial \beta}, \quad \widehat{R}_{K_n} = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \ell(\widehat{\beta}_n, \Xi_i)}{\partial \beta \partial \beta'}, \quad \widehat{\Sigma}_{K_n} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \ell(\widehat{\beta}_n, \Xi_i)}{\partial \beta} \frac{\partial \ell(\widehat{\beta}_n, \Xi_i)}{\partial \beta}'.$$

Proof. Appendix B. □

Proposition 2 says that inference on risk-neutral moments follows from nonlinear least squares standard errors, once the number of state-price density expansions K_n is chosen. Monte Carlo simulations below show that selecting a K_n that minimizes an intuitive information criterion yields confidence intervals with good coverage properties, regardless of whether the underlying data were generated by complicated stochastic volatility double-jump processes as modeled in Duffie, Pan, and Singleton (2000). The conditional Hermite polynomial basis functions therefore represent a flexible family of state-price density expansions for estimating the term structures of option prices without requiring knowledge of the exact underlying data-generating process.

Figure 1 illustrates the resulting confidence intervals for the SVS term structure by using S&P 500 index options, converted to standard deviations in order to be directly comparable to the familiar CBOE $VIX = 100\sqrt{SVS}$. The left panel plots the sieve-estimated VIX term structure's confidence intervals during the height of the 2008 financial crisis, whereas the right panel shows the corresponding term structure confidence intervals during the height of the 1998 Russian financial crisis and Long-Term Capital Management bailout. The figure shows that precise inferences about risk-neutral expectations of long-run realized variance are possible, though not always a given: during the Russian financial crisis, for example, sampling uncertainty around long-run options made it difficult to draw firm conclusions about the shape of the implied volatility term structure at longer maturities. The inference result outlined in Proposition 2 therefore presents a useful robustness check for empirical work that seeks to extract information from long-maturity options, a subject which is undertaken below.

2.4 Interpretation: Sieve Prices as Nonparametric Black-Scholes

The sieve put option price in (12) has an intuitive interpretation. Rearranging equation (12), one obtains

$$\begin{aligned} P(f_K, \mathbf{Z}) &= \kappa e^{-r\tau}\Phi(d(\mathbf{Z})) - S_0 e^{-r\tau+\mu(\mathbf{Z})} e^{\sigma(\mathbf{Z})^2/2}\Phi(d(\mathbf{Z}) - \sigma(\mathbf{Z})) \\ &\quad - \sum_{k=1}^{K_y} \gamma_k(B,\tau)\left[\frac{1}{\sqrt{k}} H_{k-1}(d(\mathbf{Z}))\phi(d(\mathbf{Z})) + S_0 e^{-r\tau+\mu(\mathbf{Z})} I_k^*(d(\mathbf{Z}))\right]. \end{aligned} \tag{19}$$

Inspection of equation (19) shows that choosing

$$\sigma(\mathbf{Z}) \equiv \sigma\sqrt{\tau}, \quad \mu(\mathbf{Z}) \equiv (r - q - \sigma^2/2)\tau \tag{20}$$

will cause the leading term in equation (19) 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 (6) is now $d(\mathbf{Z}) = (\log(\kappa/S_0) - (r - q - \sigma^2/2)\tau)/(\sigma\sqrt{\tau})$. The value σ is a constant in the sieve framework and can be chosen to equal the implied volatility of an at-the-money option.

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,

$$P(f_K, \mathbf{Z}) = P_{BS}(\sigma, \mathbf{Z}) - \sum_{k=1}^{2K_y} \gamma_k(B, \tau) \left[\frac{\kappa e^{-r\tau}}{\sqrt{k}} H_{k-1}(d(\mathbf{Z}))\phi(d(\mathbf{Z})) + S_0 e^{-q\tau - \sigma^2\tau/2} I_k^*(d(\mathbf{Z})) \right]. \quad (21)$$

This formula can be interpreted as centering the sieve at Black-Scholes, and then supplementing it with higher-order correction terms.[10] As the sample size n increases, the number of correction terms, K_y and K_τ, also increase, albeit at a slower rate than n.[11] Thus, the more data one has, the more complex the sieve option pricer is permitted to be relative to Black-Scholes. This intuition also carries over to hedging, since sieve-implied Greeks will be the standard Black-Scholes Greeks augmented with higher order correction terms that can be derived in closed-form. For brevity, this application is illustrated in the Online Appendix. In addition to their theoretical appeal, the option prices (19) have computational advantages for studying option-implied term structures: Since the integrals for risk-neutral moments in (16) require a continuum of option prices, it is convenient to have closed-form expressions for the prices in order to improve upon the speed and accuracy of numerical integration routines.

Finally, note that if the $\gamma_k(B, \tau)$ terms for $k \geq 1$ above are significantly different from zero 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, Cao, and Chen (1997), Duffie, Pan, and Singleton (2000)). The simulation study in Section 3 explores how these continuous time parametric features feed into the coefficients of the Hermite expansion and shows that an empirically tractable number of expansion terms is quite capable of fitting the conditional distributions implied by complicated stochastic volatility and jump specifications.

[10] Recently, Kristensen and Mele (2011), Xiu (2011), and León, Mencía, and Sentana (2009) have employed Hermite polynomials in a parametric option pricing setting.

[11] Recall that the $\gamma_k(B, \tau)$ terms also contain expansions in the maturity dimension.

2.5 Extension to Higher Order Moments

As is clear from (16), the theory of inference in Proposition 2 generalizes to different portfolio weights. This enables the study of risk-neutral expectations of higher order moments, as proposed for example in Bakshi, Kapadia, and Madan (2003). To this end, we compare the SVS prices to the prices of the cubic and quartic portfolios of Bakshi, Kapadia, and Madan (2003), defined as the risk-neutral expectations

$$\begin{aligned}
MOM3(\tau) \equiv \tau^{-1} E_0^{\mathbb{Q}}[e^{-r\tau} R(0,\tau)^3] &= \tau^{-1} \int_{S_0}^{\infty} \frac{6 \log(\kappa/S_0) - 3\left[\log(\kappa/S_0)\right]^2}{\kappa^2} C(\mathbf{Z}) d\kappa \\
&\quad - \tau^{-1} \int_{0}^{S_0} \frac{6 \log(S_0/\kappa) + 3\left[\log(S_0/\kappa)\right]^2}{\kappa^2} P(\mathbf{Z}) d\kappa \\
MOM4(\tau) \equiv \tau^{-1} E_0^{\mathbb{Q}}[e^{-r\tau} R(0,\tau)^4] &= \tau^{-1} \int_{S_0}^{\infty} \frac{12\left[\log(\kappa/S_0)\right]^2 - 4\left[\log(\kappa/S_0)\right]^3}{\kappa^2} C(\mathbf{Z}) d\kappa \\
&\quad + \tau^{-1} \int_{0}^{S_0} \frac{12\left[\log(S_0/\kappa)\right]^2 + 4\left[\log(S_0/\kappa)\right]^3}{\kappa^2} P(\mathbf{Z}) d\kappa
\end{aligned} \tag{22}$$

where $R(0,\tau) \equiv \log(S_\tau/S_0)$. Note that these moments clearly fall into the class of option spanning portfolios covered by (16).

The VIX, cubic, and quartic portfolio term structures are plotted in Figure 2, which shows how the they evolve over time, with red shading indicating term structures with short maturities of $\{1, 2, 3, \dots\}$ months, and blue shading indicating term structures with long maturities of $\{\dots, 22, 23, 24\}$ months. The top panel illustrates that the volatility term structure embedded in options is highly time varying in level and slope. While the level of the term structure moves strongly with the familiar 1-month VIX, the slope shows occasional signs of inversion: In times when the overall term structure level is high, short-run volatility peaks above long-run volatility, showing that index options price in a volatility mean reversion. In most periods, however, the volatility term structure appears upward sloping. In contrast, the cubic and quartic portfolios do not reveal such term structure inversions, suggesting that in all periods, long-run options are pricing in more skewness and kurtosis than short-run options.[12]

[12]Note that this effect is not mechanically due to a lengthening of maturity, as the moments in (22) are scaled by maturity.

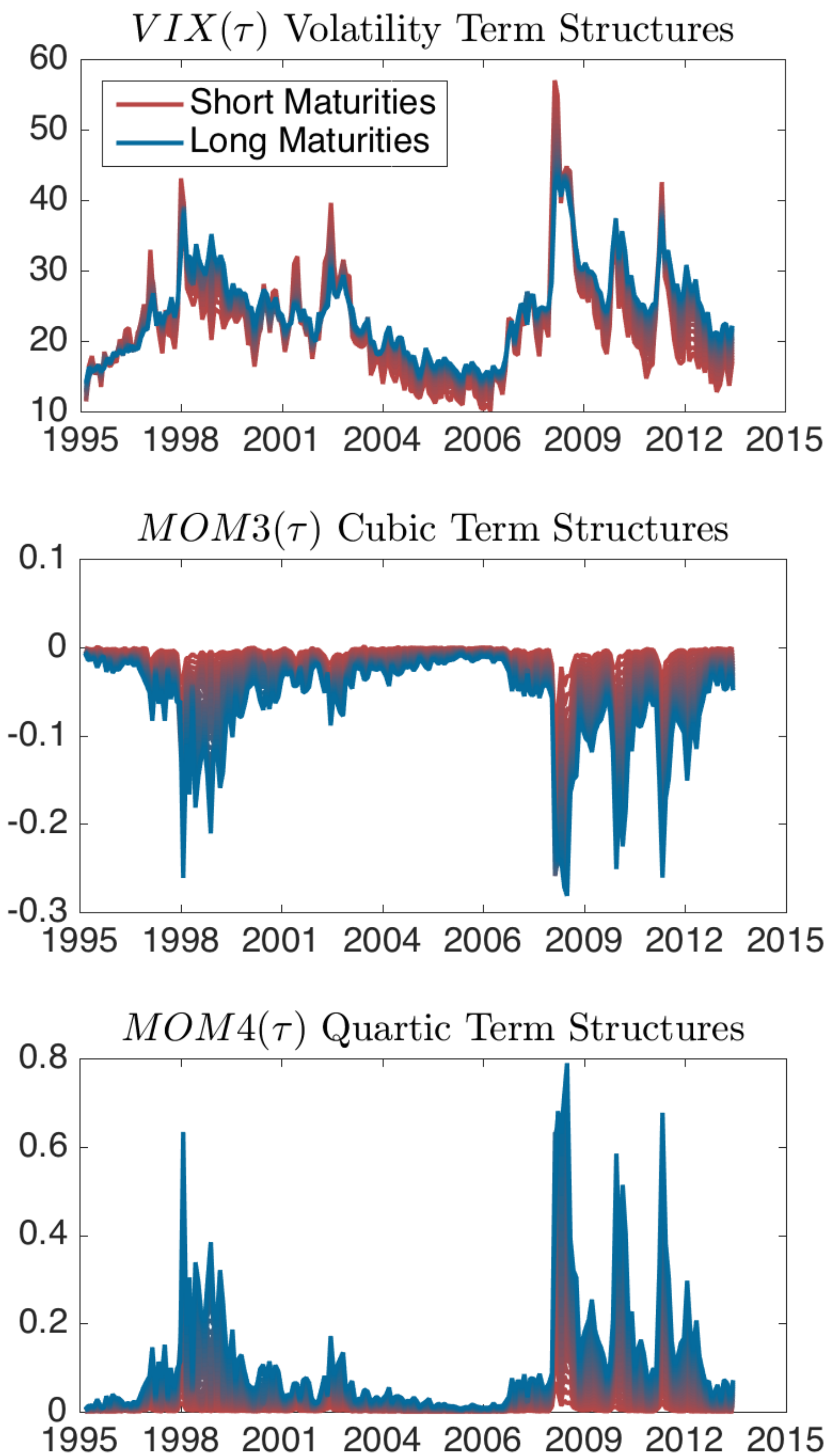

Figure 2: This figure plots the time series of estimated risk-neutral volatility, cubic, and quartic term structures. Each figure plots 24 time series, representing term structures that cover 1 to 24 months. The sample period is 1996 to 2013.

In level terms, the cubic and quartic portfolios in Figure 2 reveal that the strongest negative skewness and positive kurtosis effects by far occur in crisis periods, which is also when the volatility term structure is at its highest level. This suggests a factor structure across all three portfolios and maturities. In fact, the first principal component of the volatility, cubic, and quartic 24-month term structures explains 95% of their combined time series variation (Table 2). For reference, the one-month VIX has a correlation of 80% to this principal component, whereas the one-month cubic and quartic portfolios have correlations of -84% and 80%. But while the one-month VIX, cubic, and quartic portfolios are all strongly related to the first principal component across all portfolios and maturities, the second principal component is most closely related to the VIX's term spread, defined as the difference between 24-month VIX and 1-month VIX, $VIX(24) - VIX(1)$. The VIX term spread's correlation to the second principal component is 70%, whereas the cubic and quartic term spreads are only weakly related to the second principal component.

Table 2: Principal Components of Volatility, Cubic, and Quartic Term Structures

	Variation Explained	PC Corr to VIX Level and Term Spread	PC Corr to MOM3 Level and Term Spread	PC Corr to MOM4 Level and Term Spread
PC1	0.95	0.80	-0.84	0.80
PC2	0.03	0.70	-0.29	0.22

Notes: The principal components of the combined VIX, cubic (MOM3), and quartic (MOM4) 24-month portfolio term structures are extracted. The first column reports the variation explained by the first two principal components PC1 and PC2. The remaining columns report the correlation of PC1 with level, given by VIX(1), MOM3(1), and MOM4(1), and correlations of PC2 with term spreads, given by $[VIX(24) - VIX(1)]$, $[MOM3(24) - MOM3(1)]$, and $[MOM4(24) - MOM4(1)]$.

While Figure 2 shows the point estimates of the term structures over time, Figure 3 examines how precisely they are estimated according to the inference theory in Proposition 2. The term structures are averaged according to a pre-financial crisis subperiod (1996-2006), when the index options were less liquid, to a crisis subperiod (2007-2009) when the volatility term structure was often inverted, and a post-crisis period (2010-2013), when options in the sample were at their most liquid. The figure shows that in general, higher moment spanning portfolios are less precisely estimated when considering the confidence interval width relative to the estimated level. For example, over the 1996-2006 period, the cubic portfolio is clearly downward sloping and negative as a point estimate.

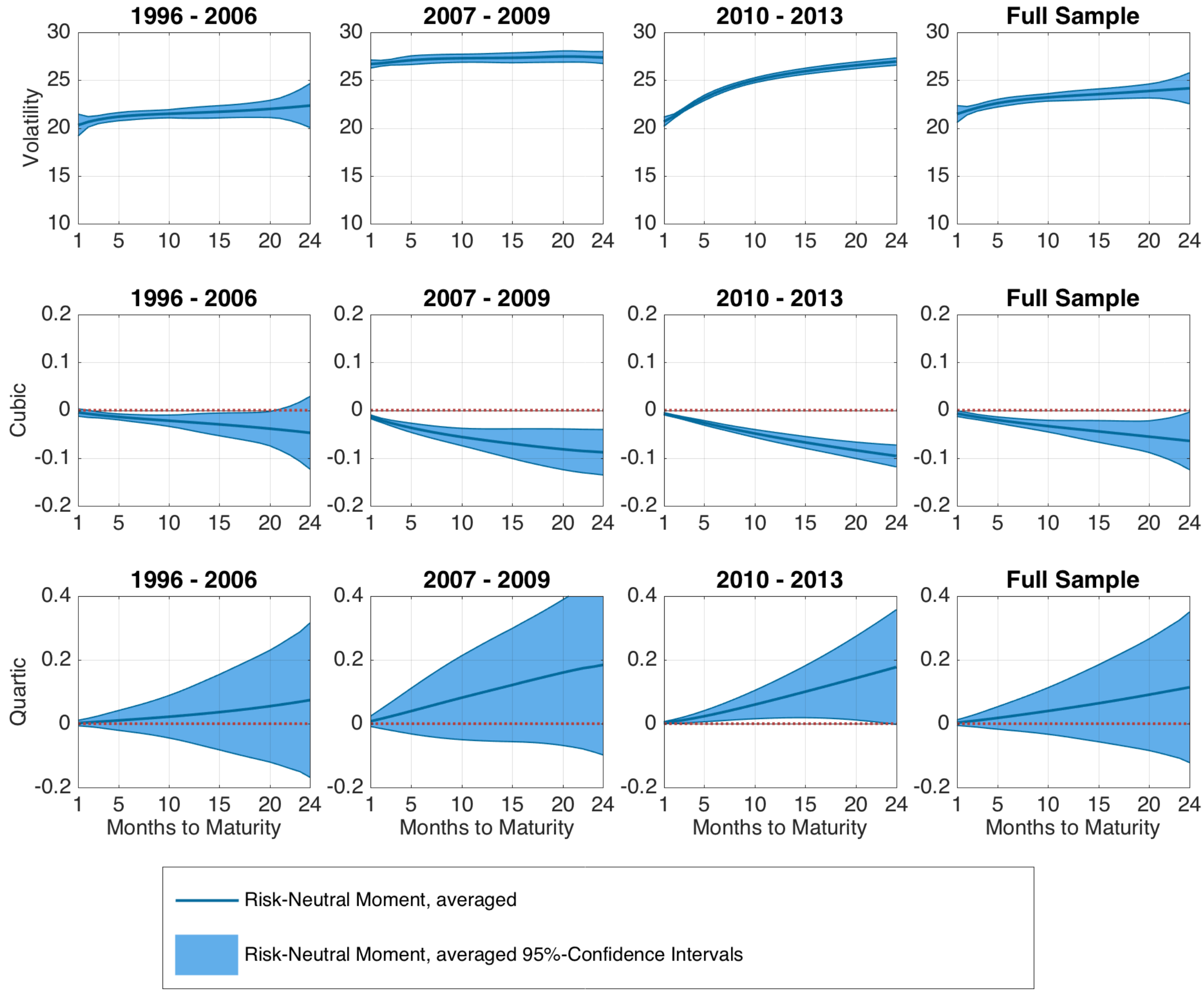

Figure 3: This figure plots average risk-neutral term structures and 95% confidence intervals as introduced in Proposition 2.

However, at long maturities, this measure of skewness is no longer statistically distinguishable from zero. In contrast, the volatility term structure appears the most precisely estimated as a fraction of the estimated level, and that its shape in general cannot be attributed to estimation error. Finally, of particular note is that the quartic portfolios are very imprecisely estimated, which calls into question their use in empirical work. For example, the width of these confidence intervals may explain why the asset pricing implications of the kurtosis portfolios in Conrad, Dittmar, and Ghysels (2013) were relatively weaker than those obtained from the variance and skewness portfolios.

Taken together, these results suggest that the risk-neutral variance term structure is both more precisely estimated and also captures a substantial portion of the time-series variation in higher-order moments, motivating in part a focus on variance term structures below.

3 Simulations

Despite its parametric appearance, the sieve is still model-free in the sense that it can fit option prices from a variety of unknown data generating processes (DGPs). To illustrate, this section presents simulations of empirically realistic option price data from DGPs of varying complexity, from which VIX term structures and confidence intervals are computed. Within the simulations, while the researcher observes the DGP and consequently the true VIX term structure, the sieve does not. Instead, the sieve must estimate the VIX from a finite sample of noisy option prices. In doing so, the sieve is only permitted to vary the number of expansion terms K_n in a data-dependent manner, making the choice of K_n as important as the choice of a bandwidth in a kernel regression. A data-driven method for choosing K_n is proposed below and is shown to perform well across several DGPs.

The simulations in this section refer to various subcases of the following general data generating process,

$$\begin{aligned} dX_t &= \left(r - q - \lambda\overline{\mu} - \frac{1}{2}V_t\right)dt + \rho\sqrt{V_t}dW_t + J_t dN_t \\ dV_t &= \kappa_v(\overline{V} - V_t)dt + \rho\upsilon\sqrt{V_t}dW_t + \left(1-\rho^2\right)^{1/2}\upsilon\sqrt{V_t}dW_t' + Z_t dN_t \end{aligned} \tag{23}$$

where V_t is a stochastic volatility process, X_t is the underlying's log price, W_t and W_t' are standard Brownian motions, and κ_v, $\overline{V}$, ρ, υ 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 stochastic volatility double-jump process (SVJJ), which is a special case of the general affine-jump diffusion processes treated in Duffie, Pan, and Singleton (2000) that is nonetheless general enough to nest the seminal 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, Fusari, and Todorov (2012) and are given in

	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

Table 3: Parameter values used in the simulation exercises in Section 3.

Table 3.

Given the parameter values in Table 3, a dataset with empirically realistic options is simulated by mimicking features of September 23, 1998, the bailout date of LTCM. This date is chosen to represent crisis conditions while keeping the option dataset computationally manageable, due to the many optimizations that need to be solved across all Monte Carlo datasets. That is, options are simulated with 1, 2, 3, 6, 9, 12, 15, 21 months-to-maturity and with respective number of observations 32, 20, 44, 31, 30, 9, 23, 27. The range of strikes simulated at each maturity corresponds to the same moneyness of options observed in the data. Finally, each drawn option price is perturbed with uniformly distributed noise corresponding to the width of the bid-ask spread observed in the actual data. In this way, 1000 option datasets are simulated from the SVJJ process in (23), and for each dataset, the true VIX term structure (free of noise, discretization, and truncation error) is computed.

Furthermore, for each finite simulated sample observed with noise, the sieve least squares regression (14) is estimated. The number of sieve expansion terms K_n are chosen with both theoretical and computational considerations in mind. While Coppejans and Gallant (2002) have shown that leave-one-out and hold-out cross-validations perform well for univariate Hermite series in the context of density estimations, these cross-validations typically involve heavy computation. The curse of dimensionality compounds the problem for the two-dimensional Hermite polynomials studied here. For example, for a sample of size n, leave-one-out cross validation requires computation of the nonlinear regression (14) $(n-1)$ times for many configurations of $K_n = (K_y(n)+1)(K_\tau(n)+1)$. Among computationally feasible selection criteria, minimizing the Bayesian Information Criterion (BIC) or

Table 4: Monte Carlo Rejection Frequencies.

DGP	1	2	3	6	9	12	15	21	Expansions (K_y, K_τ)
	Maturity (months)								
SVJJ	0.051	0.020	0.030	0.020	0.030	0.030	0.020	0.010	(7,2)
SVJ	0.003	0.027	0.038	0.034	0.044	0.022	0.041	0.050	(6,2)
Heston	0.034	0.018	0.018	0.020	0.030	0.049	0.087	0.078	(4,2)
Black-Scholes	0.035	0.036	0.026	0.018	0.031	0.043	0.052	0.046	(0,0)

Notes: A dense surface of true option prices was simulated under an SVJJ specification for each of the maturities shown, from which a true VIX was computed without moneyness truncation error. Then, 1000 random subsamples were drawn from this surface. These sample prices were perturbed with a uniformly distributed error corresponding to the width of observed bid-ask spreads on S&P 500 index options. $\widehat{VIX}(\tau)$ estimates were computed and studentized according to Proposition 2, and corresponding 95% confidence intervals were constructed by inverting nominal level 5% tests. Rejection frequencies report the proportion of simulated draws for which the true $VIX(\tau)$ was outside the 95% confidence intervals.

the well-known Mallows (1973) criterion, which is asymptotically equivalent to leave-one-out cross-validation in certain settings, are natural candidates that perform equally well in the simulations studied in this paper.[13]

At each Monte Carlo iteration, the result of the sieve least squares regression are closed-form option prices $\widehat{P}(\mathbf{Z}) \equiv P(\widehat{\beta}, \mathbf{Z})$ (derived in Proposition 1), which can be computed with arbitrarily dense strikes for any given maturity. These prices are then fed into the integral in (1), which is computed at each observed maturity. Because the sieve estimates allow extrapolations into the strike tails, it is possible to set the strike integration range in a manner that ensures that prices representing the 0.5% to 99.5% quantiles of the implied risk-neutral distribution are included, allowing the sieve to substantially reduce strike truncation errors. Finally, following Proposition 2, the studentized $VIX(\tau)$ curve is computed for each simulated dataset, and the corresponding 95% confidence intervals are formed using standard normal critical values.

Table 4 shows the rejection frequencies of the inference procedure, i.e. the proportion of datasets for which the the 95% confidence intervals do not cover the true VIX at each maturity along the term structure. The results show that for the nominal level 5% test considered, the confidence intervals display good, though often slightly conservative, size control. The right-most column, which shows the modal number of expansion terms that were selected by the aforementioned data-

[13] Mallows (1973) criterion involves solving

$$K_n = \arg\min_K \frac{1}{n} \sum_{i=1}^{n} \Big[P_i - P(\beta_K, \mathbf{Z}_i)\Big]^2 W(\mathbf{Z}_i) + 2\hat{\sigma}^2 (K/n).$$

See also Li and Racine (2007, p. 451).

driven procedure, suggests a clear relationship between the complexity of the underlying DGP and the number of expansions $K = (K_y + 1)(K_\tau + 1)$ selected. Importantly, when the underlying DGP is in fact Black-Scholes, the modal number of expansions chosen was the correct $(0, 0)$.

Finally, notice that while the sieve state-price densities $f(Y|\tau)$ do not explicitly depend on stochastic volatility, the sieve nonetheless performs well in capturing the option-implied term structures of stochastic processes that do allow for stochastic volatility. This is because the effect of stochastic volatility (as well as jumps) is to create option prices whose underlying state-price densities are different by maturity. Hence, stochastic volatility and jumps can be viewed as modeling devices that uncouple long-run option prices from short-run option prices. The sieve captures this effect directly by allowing long-maturity state-price densities to differ from their short-maturity counterparts. The sieve's effectiveness in doing so is explored in further simulations in the Online Appendix C.

4 Return Predictability in the Term Structure of Variance Risk Premia

This section combines the sieve framework with a novel set of expectation hypothesis and return predictability regressions to study the term structure of the variance risk premium and its link to equity risk premia. Following Bollerslev, Tauchen, and Zhou (2009), the variance risk premium is defined as follows. Let realized variance from month t to $T = t + \tau$ be given by the annualized sum of squared daily returns

$$RV_{t,T} \equiv \frac{252}{n} \sum_{i=1}^{n} \left(\frac{sp500(t + i\Delta_n) - sp500(t + (i-1)\Delta_n)}{sp500(t + (i-1)\Delta_n)} \right)^2, \tag{24}$$

where $n = \tau/\Delta_n$ is the number of trading days between t and T, Δ_n is the daily increment, and $sp500(t)$ represents the level of the S&P 500 index at time t. The variance risk premium is the difference between objective ($\mathbb{P}$-measure) and risk-neutral ($\mathbb{Q}$-measure) conditional expectations of $RV_{t,T}$

$$VRP_t(t, T) \equiv \mathbb{E}_t^{\mathbb{P}}[RV_{t,T}] - \mathbb{E}_t^{\mathbb{Q}}[RV_{t,T}]. \tag{25}$$

Note that for a pricing kernel $M_{t,T}$ and $m_{t,T} \equiv M_{t,T}/\mathbb{E}_t^{\mathbb{P}}[M_{t,T}]$,

$$\mathbb{E}_t^{\mathbb{Q}}[RV_{t,T}] = E_t^{\mathbb{P}}[m_{t,T}RV_{t,T}] = E_t^{\mathbb{P}}[RV_{t,T}] + Cov_t^{\mathbb{P}}[m_{t,T}, RV_{t,T}],$$

so that the difference in (25) measures covariation of realized variance with the pricing kernel, or in other words, a risk premium. Following Carr and Wu (2009), the quantity $\mathbb{E}_t^{\mathbb{Q}}[RV_{t,T}]$ is well replicated by the synthetic variance swap, i.e. the integrated option portfolio (1): $SVS_t(\tau) = \mathbb{E}_t^{\mathbb{Q}}[RV_{t,T}]$ up to a third-order approximation error.

Expectation Hypothesis Under a null hypothesis $H_0 : Cov_t^{\mathbb{P}}[m_{t,T}, RV_{t,T}] = 0$ of no variance risk premium, one has $\mathbb{E}_t^{\mathbb{Q}}[RV_{t,T}] = \mathbb{E}_t^{\mathbb{P}}[RV_{t,T}]$, so that for $\varepsilon_{t+\tau}$ with $\mathbb{E}_t^{\mathbb{P}}[\varepsilon_{t+\tau}] = 0$, $RV_{t,T} = \mathbb{E}_t^{\mathbb{Q}}[RV_{t,T}] + \varepsilon_{t+\tau}$. Therefore, H_0 is equivalent to the joint null hypothesis $a = 0$ and $b = 1$ in the regressions

$$RV_{t,T} = a(\tau) + b(\tau)\mathbb{E}_t^{\mathbb{Q}}[RV_{t,T}] + \varepsilon_{t+\tau}. \tag{26}$$

The idea is to test several hypotheses of this form and to relate them to well-established findings for the 1-month VRP. We therefore augment (26) with the concept of a *forward variance*, which takes advantage of the additive properties of $RV_{t,T}$: Note that from (24), one has for horizons $\tau > 1$, $RV_{t,T} = RV_{t,t+1} + RV_{t+1,T}$, giving rise to a decomposition

$$\begin{aligned} VRP_t(t,T) &= \mathbb{E}_t^{\mathbb{P}}[RV_{t,t+1} + RV_{t+1,T}] - \mathbb{E}_t^{\mathbb{Q}}[RV_{t,t+1} + RV_{t+1,T}] \\ &= \mathbb{E}_t^{\mathbb{P}}[RV_{t,t+1}] - \mathbb{E}_t^{\mathbb{Q}}[RV_{t,t+1}] + E_t^{\mathbb{P}}[RV_{t+1,T}] - \mathbb{E}_t^{\mathbb{Q}}[RV_{t+1,T}] \\ &\equiv VRP_t(t,t+1) + VRP_t(t+1,T). \end{aligned} \tag{27}$$

Notice that the first component on the right-hand side is the familiar one-month variance risk premium that has been extensively studied in the literature using published (one-month) VIX data.[14] Therefore, in order to relate findings on $VRP_t(t,T)$ to our existing understanding of the 1-month variance risk premium, we also test hypotheses regarding the forward variance risk premium

$$RV_{t+1,T} = a(\tau) + b(\tau)\mathbb{E}_t^{\mathbb{Q}}[RV_{t+1,T}] + \varepsilon_{t+\tau}, \tag{28}$$

[14] See, for example, Carr and Wu (2009), Bollerslev and Todorov (2011), Bollerslev, Gibson, and Zhou (2011), Drechsler and Yaron (2011), Bollerslev, Osterrieder, Sizova, and Tauchen (2013), Bekaert and Hoerova (2014).

where $\mathbb{E}_t^{\mathbb{Q}}[RV_{t+1,T}] = [SVS_t(\tau) - SVS_t(1)]$ captures the steepness of the synthetic variance swap curve in maturity. A test of $H_0 : a = 0 \cap b = 1$ is a test of the forward variance risk premium $VRP_t(t+1, T)$.

Return Predictability Bollerslev, Tauchen, and Zhou (2009) and Bekaert and Hoerova (2014), among others, provide evidence that $VRP_t(t, t+1)$ predicts excess stock market returns. Using the sieve-estimated term structure of $SVS_t(\tau)$, one can test whether their predictability result extends to long-run variance risk premia as well as forward variance risk premia, i.e. $VRP_t(t, T)$ and $VRP_t(t+1, T)$. We therefore estimate both

$$\begin{aligned} Re_{t+h} &= \alpha_{h,\tau} + \beta_{h,\tau} VRP_t(t, T) + \varepsilon_{t+h} \qquad \text{and} \\ Re_{t+h} &= \alpha_{h,\tau} + \beta_{h,\tau} VRP_t(t, t+1) + \gamma_{h,\tau} VRP_t(t+1, T) + \varepsilon_{t+h} \end{aligned} \tag{29}$$

for various forecasting horizons h and term structure maturities τ, where Re_{t+h} denotes the h-month ahead CRSP value-weighted return (including dividends) in excess of the risk-free rate. To ensure that $VRP_t(t, T) = \mathbb{E}_t^{\mathbb{P}}[RV_{t,T}] - \mathbb{E}_t^{\mathbb{Q}}[RV_{t,T}]$ lies in the time t information set, we need a $\mathbb{P}$-measure forecast of realized variance, $\mathbb{E}_t^{\mathbb{P}}[RV_{t,T}]$, which we obtain from the standard heterogeneous AR model of Corsi (2009),

$$RV_{t,t+1} = b_0 + b_1 RV_t + b_2 \left(\frac{1}{6} \sum_{i=0}^{5} RV_{t-i} \right) + b_3 \left(\frac{1}{24} \sum_{i=0}^{23} RV_{t-i} \right) + \varepsilon_{t+1}, \tag{30}$$

which effectively captures the long-memory dynamics of the RV_t process. To avoid look-ahead bias, (30) is estimated for each month t in our option sample 1996-2013, using monthly S&P 500 index RV (24) from 1950 to t. The results and conclusions below do not materially depend on the exact lag structure of this RV forecasting regression, since they are upheld under various specifications. Long-run forecasts of $RV_{t,T}$ can be obtained by iterating (30) forward.

4.1 S&P 500 Index Option Data

To run the regressions in (26), (28), and (29), we use the proposed sieve framework to estimate a balanced monthly time series of $SVS_t(\tau) = \mathbb{E}_t^{\mathbb{Q}}[RV_{t,T}]$ term structures from data on S&P 500 index options (SPX) spanning January, 1996 to August, 2013. Following the data filtering procedure of

Andersen, Fusari, and Todorov (2012), we use the average of closing bid and ask quotes, discard all in-the-money options, and options with maturities of less than 7 days. Call option information is incorporated by converting out-of-the-money calls to in-the-money puts by put-call parity. Furthermore, we follow the CBOE (2003) VIX White Paper procedure of excluding options with strikes beyond the first pair of zero-bid option prices. Table 1 presents summary statistics of the resulting dataset, which includes option surfaces observed at the end of the month, for a total of 212 months.

To be specific, for each month t of these 212 option cross-sections, we solve the sieve least squares problem (14) and compute the portfolio integration in (1) for $\tau = 1, 2, \ldots, 24$ months-to-maturity. At each of these maturities, the integration limits in (1) were set to cover the 0.5% to 99.5% quantiles of the implied risk-neutral CDF, yielding a balanced monthly term structure $\widehat{SVS_t}(\tau)$. To check that the resulting $\widehat{SVS_t}(\tau)$ produces coherent estimates of implied volatility at the one-month horizon, we plot $100 \cdot \widehat{SVS_t}(1)^{1/2} = \widehat{VIX_t}(1)$ against the CBOE's published VIX in the top panel of Figure 4. The unconditional correlation between the two series is 0.9976, and the number of expansion terms selected via the data-driven criterion (Section 3) was about $(K_y, K_\tau) = (8, 3)$ on average.

Finally, note that inference on regressions of the form (26) and (28) can be affected by measurement error and persistence in the regressors. Measurement error in the regressors is known to cause attenuation bias in the slope coefficient, which is especially problematic when testing hypotheses of the form $b = 1$. That is, by replacing $\mathbb{E}_t^{\mathbb{Q}}[RV_{t,T}]$ in equation (26) with its estimate $\widehat{SVS_t}(\tau) = \mathbb{E}_t^{\mathbb{Q}}[RV_{t,T}] + \eta_{t,T}$, one obtains $RV_{t,T} = \alpha(\tau) + \beta(\tau)\widehat{SVS_t}(\tau) + u_{t+\tau} = \alpha(\tau) + \beta(\tau)[\mathbb{E}_t^{\mathbb{Q}}[RV_{t,T}] + \eta_{t,T}] + u_{t+\tau}$. Assuming for simplicity that each $\eta_{t,T}$ is independent of other variables, we see from standard arguments that $\beta(\tau) = Cov(RV_{t,T}, \widehat{SVS_t}(\tau))/Var[\widehat{SVS_t}(\tau)] = b(\tau)Var[\mathbb{E}_t^{\mathbb{Q}}[RV_{t,T}]]/\{Var[\mathbb{E}_t^{\mathbb{Q}}[RV_{t,T}]] + \mathbb{E}[\sigma_{\eta,t}^2]\} \equiv b(\tau)\phi(\tau)$, which is the classical measurement error attenuation bias. Note that Proposition 2 ensures that for fixed t, $\sqrt{n_t}(\widehat{SVS_t}(\tau) - \mathbb{E}_t^{\mathbb{Q}}[RV_{t,T}]) \to^d N(0, \sigma_{\eta,t}^2)$, and hence one can use the sieve standard errors to perform a simple bias correction by multiplying the estimated $\hat{\beta}(\tau)$ by an estimate of $\phi(\tau)^{-1}$, which is greater than one. The bottom left panel of Figure 4 plots the estimated bias correction $\phi(\tau)^{-1}$ (in excess of one) and shows that the bias can play a significant role (up to 14%) for long-horizon $\tau = 24$. On average, term structure estimation error ranges from 2% to 14% of time-series variation in the synthetic variance swap curve (Figure 4, middle right panel). Finally, note that serial correlation in the regressors is increasing in τ, but all cases well below 1.

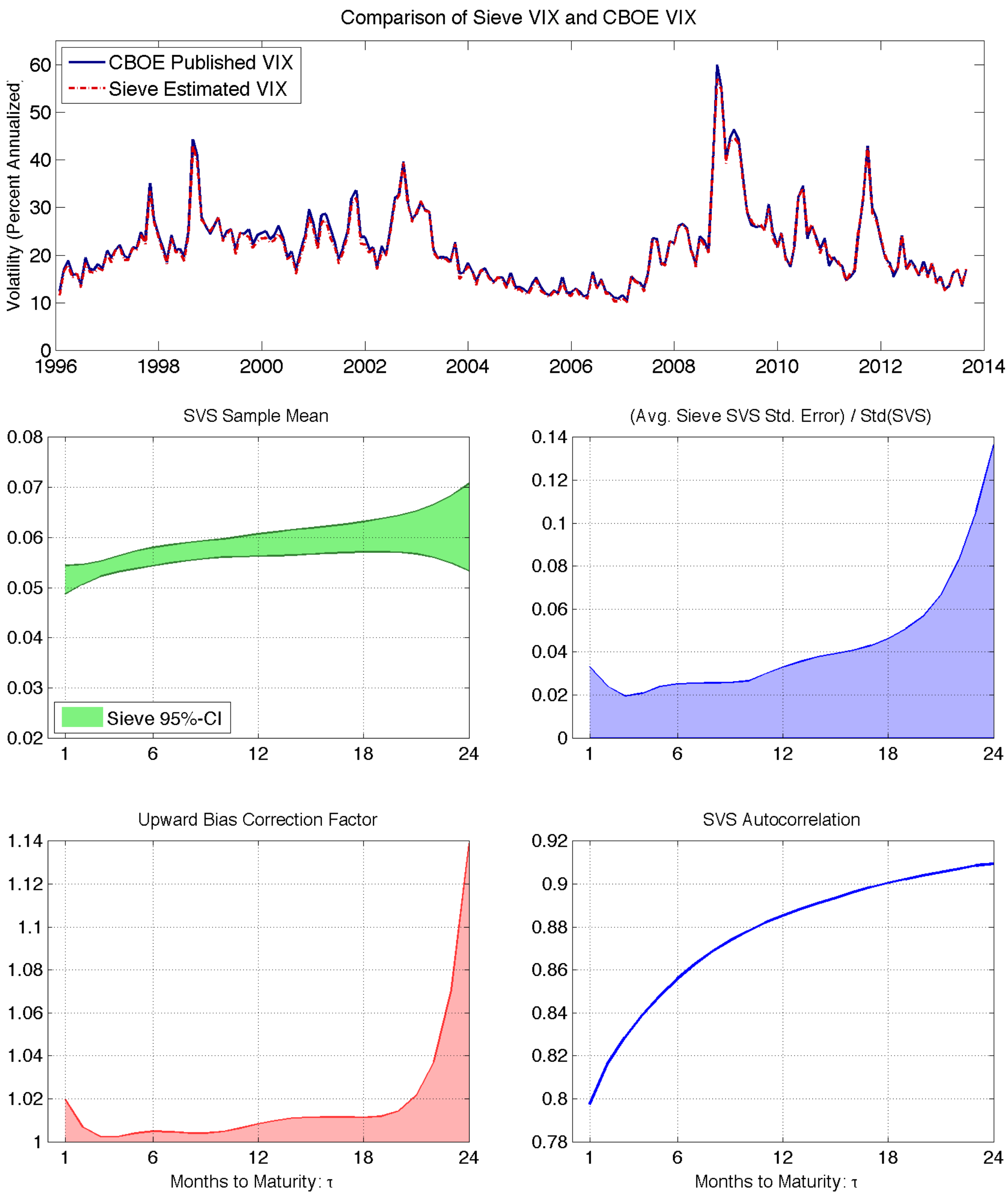

Figure 4: Pre-regression Diagnostics. The top panel plots a comparison of the sieve-estimated 30-day VIX and the CBOE published 30-day VIX. The middle left panel plots the average 95% confidence intervals of the synthetic variance swap term structure. The middle right panel shows the ratio of average sieve standard errors of $\widehat{SVS}_t(\tau)$ to the time series standard deviation of $\widehat{SVS}_t(\tau)$. The bottom left panel shows the bias correction factor (in excess of one) to be applied to the slope coefficient in the expectation hypothesis regression, and the bottom right panel plots the sample first-order autocorrelation of $\widehat{SVS}_t(\tau)$ for maturities $\tau = 1, \ldots, 24$. Sieve standard errors for $\widehat{SVS}_t(\tau)$ are computed for 212 months from January, 1996, to August, 2013, using S&P 500 index options and the inference procedure in Proposition 2.

4.2 Results

Expectation Hypothesis The results of the bias-corrected regressions (26) and (28) are surprising. p-values in the first row of Table 5 show strong evidence against the null hypothesis $H_0 : a = 0 \cap b = 1$ of no variance risk premium $VRP_t(t,T)$ across all maturities $\tau = 1, \ldots, 12, 18, 24$. In contrast, the forward variance tests reported in the bottom panel are unable to reject the null hypothesis of no forward variance risk premium $VRP_t(t+1,T)$. A notable exception is at the $\tau = 2$ horizon, whose p-value in the forward regression is smaller than in the full regression, suggesting that investors earn a premium for being exposed to variance risk between $t+1$ and $t+2$ as well. In sum, the strong rejections in the first row and the lack of rejection in the second row suggest that compensation for variance risk is concentrated on the first one or two maturities.[15]

Table 5: p-Values for Expectation Hypothesis Regressions.

	τ													
	1	2	3	4	5	6	7	8	9	10	11	12	18	24
SVS	0.000	0.074	0.019	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Forward	-	0.032	0.072	0.111	0.141	0.189	0.227	0.271	0.324	0.371	0.401	0.421	0.529	0.537

Notes: The bias-corrected OLS regressions from (26) and (28) of realized variance on sieve synthetic variance swaps $\widehat{SVS}_t(\tau)$ and forward variance swaps $\widehat{SVS}_t(\tau) - \widehat{SVS}_t(1)$, respectively, are estimated for each of the monthly horizons $\tau = 1, \ldots, 12, 18, 24$. p-values in the first row report the outcome of the joint tests $a(\tau) = 0 \cap b(\tau) = 1$ for the regression on the full variance swap regression (26), and the second row shows the corresponding outcome for the forward variance swap (28). Newey and West (1987) standard errors for lag length 24 are used.

Return Predictability The results of the expectation hypothesis tests are further corroborated in the return predictability regressions. Table 6 reports t-statistics on the slope coefficient of the first regression in (29) using Hodrick (1992) standard errors.[16] The left-most column shows the same pattern of excess return predictability on horizons $h = 2, \ldots, 7$ found in Bollerslev, Tauchen, and Zhou (2009) for one-month $VRP_t(t, t+1)$. The pattern is noteworthy given that Bollerslev, Tauchen, and Zhou (2009) use S&P 500 index excess returns, whereas we use CRSP value-weighted excess returns over a different sample period as the left-hand side variable. Further out into the

[15] For reference, the full regression output is provided in the Online Appendix.

[16] See the discussion in Ang and Bekaert (2007) in favor of using Hodrick (1992) standard errors in overlapping return predictability regressions.

Table 6: Excess Return Predictability of the VRP Term Structure.

	t-stat on $VRP_t(t,T)$													
$h \setminus \tau$	1	2	3	4	5	6	7	8	9	10	11	12	18	24
h=1	-1.21	-1.28	-1.41	-1.53	-1.61	-1.65	-1.63	-1.59	-1.55	-1.52	-1.49	-1.47	-1.32	-1.27
h=2	-2.00	-2.02	-2.12	-2.24	-2.30	-2.30	-2.24	-2.18	-2.11	-2.06	-2.01	-1.97	-1.77	-1.65
h=3	-2.28	-2.17	-2.22	-2.32	-2.40	-2.41	-2.36	-2.30	-2.23	-2.18	-2.14	-2.09	-1.89	-1.71
h=4	-2.03	-1.96	-2.04	-2.16	-2.26	-2.30	-2.27	-2.23	-2.19	-2.15	-2.12	-2.09	-1.93	-1.77
h=5	-1.92	-1.97	-2.10	-2.24	-2.34	-2.38	-2.36	-2.32	-2.28	-2.25	-2.22	-2.18	-1.99	-1.90
h=6	-2.09	-2.09	-2.17	-2.27	-2.33	-2.35	-2.33	-2.30	-2.26	-2.23	-2.20	-2.16	-1.97	-1.87
h=7	-2.04	-2.03	-2.10	-2.19	-2.26	-2.28	-2.27	-2.24	-2.21	-2.18	-2.14	-2.11	-1.91	-1.76
h=8	-1.75	-1.78	-1.85	-1.92	-1.97	-1.99	-1.98	-1.96	-1.94	-1.91	-1.89	-1.86	-1.67	-1.58
h=9	-1.54	-1.59	-1.65	-1.71	-1.75	-1.77	-1.76	-1.75	-1.74	-1.73	-1.71	-1.70	-1.53	-1.46
h=10	-1.30	-1.39	-1.47	-1.54	-1.59	-1.61	-1.62	-1.61	-1.61	-1.60	-1.59	-1.58	-1.40	-1.32
h=11	-1.17	-1.27	-1.36	-1.43	-1.48	-1.51	-1.52	-1.52	-1.51	-1.50	-1.49	-1.48	-1.31	-1.21
h=12	-1.10	-1.18	-1.25	-1.31	-1.35	-1.37	-1.37	-1.36	-1.35	-1.34	-1.33	-1.32	-1.15	-1.08

	t-stat on forward $VRP_t(t+1,T)$													
$h \setminus \tau$	1	2	3	4	5	6	7	8	9	10	11	12	18	24
h=1		0.18	0.01	-0.10	-0.14	-0.15	-0.11	-0.08	-0.04	-0.01	0.01	0.02	0.13	0.19
h=2		0.23	0.01	-0.15	-0.20	-0.19	-0.13	-0.09	-0.05	-0.02	-0.00	0.02	0.16	0.24
h=3		0.66	0.35	0.04	-0.14	-0.22	-0.22	-0.20	-0.18	-0.16	-0.14	-0.13	-0.00	0.15
h=4		0.17	-0.13	-0.40	-0.57	-0.65	-0.65	-0.63	-0.61	-0.60	-0.58	-0.57	-0.45	-0.31
h=5		-0.66	-0.93	-1.10	-1.18	-1.20	-1.16	-1.12	-1.08	-1.04	-1.01	-0.97	-0.78	-0.70
h=6		-0.65	-0.85	-1.00	-1.07	-1.10	-1.08	-1.05	-1.02	-1.00	-0.97	-0.94	-0.74	-0.65
h=7		-0.72	-0.88	-1.01	-1.08	-1.11	-1.10	-1.08	-1.05	-1.02	-0.99	-0.96	-0.74	-0.60
h=8		-0.81	-0.93	-1.02	-1.06	-1.07	-1.06	-1.04	-1.01	-0.98	-0.96	-0.93	-0.71	-0.62
h=9		-0.83	-0.90	-0.95	-0.97	-0.98	-0.97	-0.96	-0.94	-0.93	-0.92	-0.90	-0.71	-0.65
h=10		-0.96	-1.02	-1.05	-1.05	-1.04	-1.02	-1.01	-0.99	-0.97	-0.95	-0.94	-0.72	-0.64
h=11		-1.03	-1.09	-1.10	-1.09	-1.07	-1.04	-1.02	-0.99	-0.97	-0.95	-0.94	-0.72	-0.61
h=12		-0.86	-0.91	-0.92	-0.90	-0.88	-0.86	-0.84	-0.81	-0.79	-0.78	-0.76	-0.56	-0.50

Notes: The top panel reports Hodrick (1992) t-statistics from regressions of monthly CRSP value-weighted excess returns on the lagged term structure of variance risk premia (29), i.e. $Re_{t+h} = \alpha_{h,\tau} + \beta_{h,\tau} VRP_t(t,T) + \varepsilon_{t+h}$, where $T = t+\tau$.The bottom panel reports analogous t-statistics for regressions of excess returns on lagged forward variance risk premia $Re_{t+h} = \alpha_{h,\tau} + \gamma_{h,\tau} VRP_t(t+1,T) + \varepsilon_{t+h}$. The forward variance risk premia reflect compensation for volatility occurring over month $t+1$ to T. Gray shading denotes significance at the 5% level..

term structure, the remaining columns of Table 6 show that the strong predictive pattern is mirrored for $VRP_t(t,T)$ for $\tau = 2,\ldots,12$. Note that the negative sign on the t-statistics indicates that declines in the variance risk premium (e.g. $\mathbb{P}$-measure forecasts of volatility exceed $\mathbb{Q}$-measure implied volatility) tend to predict positive excess returns. Heuristically, the sign is consistent with the intuition that long positions in variance swaps are often used as hedges against high-marginal utility states, since they pay out when realized variance exceeds implied variance.

One implication of excess return predictability is that it provides a measure for the time-variation of the equity risk premium, $\hat{\mathbb{E}}_t[Re_{t+h}] = \hat{\alpha}_{h,\tau} + \hat{\beta}_{h,\tau} VRP_t(t,t+\tau)$. Figure 5 illustrates examples for the 6-month ahead equity risk premium using the slower-moving 6,12,18, and 24-month variance risk

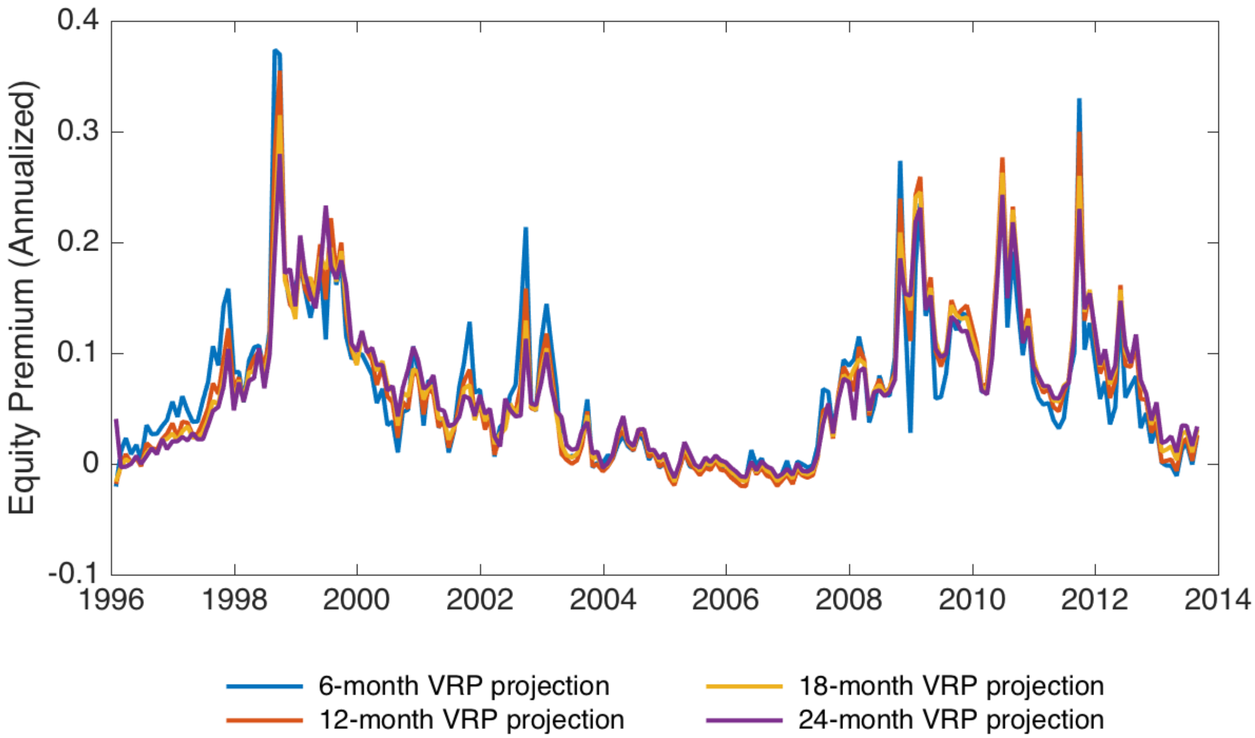

Figure 5: The Equity Risk Premium as projected onto various points on the variance risk premium term structure, $\hat{E}_t[Re_{t+6}] = \hat{\alpha}_{6,\tau} + \hat{\beta}_{6,\tau} VRP_t(t, t+\tau)$, for $\tau = 6, 12, 18, 24$.

premia. The figure shows that the equity premium is significantly varying over time and assumes its largest values in crisis periods, with peaks occurring during the Asian financial crisis and LTCM bankruptcy, the 2002-03 Iraq invasion, the 2008-09 Great Recession, and the subsequent European sovereign debt crises.

4.3 Economics of Return Predictability Results

The results of the return predictability regressions in Table 6 can be interpreted within the models of Bollerslev, Tauchen, and Zhou (2009) (BTZ) and Drechsler and Yaron (2011) (DY), which provide theoretical foundations for a link between the variance risk premium and the predictability of excess returns.

The BTZ Model Bollerslev, Tauchen, and Zhou (2009) specify an equilibrium consumption model that links the one-month variance risk premium to expected excess returns. In their model, consumption growth $g_{t+1} = \log(C_{t+1}/C_t)$ follows

$$g_{t+1} = \mu_g + \sigma_{g,t} z_{g,t+1}, \tag{31}$$

where $z_{g,t+1}$ is i.i.d. $N(0,1)$. The consumption volatility dynamics are governed by mean-reverting square root processes

$$\sigma_{g,t+1} = a_\sigma + \rho_\sigma \sigma^2_{g,t} + \sqrt{q_t} z_{\sigma,t+1}, \tag{32}$$

$$q_{t+1} = a_q + \rho_q q_t + \phi_q \sqrt{q_t} z_{q,t+1}, \tag{33}$$

where $z_{\sigma,t+1}$ and $z_{q,t+1}$ are i.i.d. $N(0,1)$ innovations, and where the parameters satisfy certain regularity conditions. The process q_t represents the volatility of consumption volatility, which is itself stochastic. The main implication of the Bollerslev, Tauchen, and Zhou (2009) is that under Epstein and Zin (1991) recursive preferences, the variance risk premium for returns on the economy's consumption asset is given by

$$\mathbb{E}^{\mathbb{P}}_t \left[\sigma^2_{r,t+1}\right] - \mathbb{E}^{\mathbb{Q}}_t \left[\sigma^2_{r,t+1}\right] = const. \times q_t, \tag{34}$$

where *const.* is a constant scalar consisting of sums and products of the model's fundamental parameters. Equation (34) says that in the Bollerslev, Tauchen, and Zhou (2009) economy, time-variation in the variance risk premium is driven entirely by the single consumption vol-of-vol process q_t.

It is straightforward to show that the entire *term structure* of variance risk premia in this economy, $\mathbb{E}^{\mathbb{P}}_t \left[\sigma^2_{r,t+j}\right] - \mathbb{E}^{\mathbb{Q}}_t \left[\sigma^2_{r,t+j}\right]$ for $j > 1$, is also linearly driven by the same consumption vol-of-vol process q_t, as are the economy's *forward* VRPs. Thus an implication of the Bollerslev, Tauchen, and Zhou (2009) is that forward variance risk premia must also predict returns, a result that is strongly at odds with the bottom panel of Table 6. Instead, the results of Table 6 suggest a model in which at least two factors drive time-variation in the term structure of the variance risk premium, with the first factor driving short-run VRP having the ability to predict excess returns, and the second factor driving forward VRP, which does not have predictive content. Furthermore, note that the second factor may not simply be white noise, since the top panel of Table 6 suggests that the sum of short-run VRP and forward VRP still predict excess returns.

The DY Model Similar implications are readily obtained from the Drechsler and Yaron (2011) model. The state vector in their long-run risk economy with recursive preferences is given by $Y_t \in \mathbb{R}_n$, which follows a VAR

$$Y_{t+1} = \mu + FY_t + G_t z_{t+1} + J_{t+1}, \tag{35}$$

whose shocks are driven by both Gaussian innovations $z_{t+1} \sim N(0, I_n)$ and a Poisson jump process J_{t+1}. The state vector of main interest is $Y_{t+1} = (g_{t+1}, x_{t+1}, \overline{\sigma}^2_{t+1}, \sigma^2_{t+1}, \Delta d_{t+1})$, where x_{t+1} is the persistent conditional mean of consumption growth g_{t+1}, where σ^2_{t+1} drives stochastic volatility across the state vector ($G_t G_t' = h + H_\sigma \sigma^2_t$), $\overline{\sigma}^2_{t+1}$ is the long-run mean of volatility, and Δd_{t+1} is dividend growth.

The key state variable that emerges from the Drechsler and Yaron (2011) model is σ^2_t, which drives time variation in both the variance risk premium as well as the equity risk premium. However, Drechsler and Yaron (2011) show that *forward* VRP (the *drift difference* of VRP in their terminology) is also driven by an affine function of σ^2_t whose intercept and coefficients depend on model fundamentals. Hence in the Drechsler and Yaron (2011) economy, forward VRP also predicts excess returns, which is at odds with the findings in the bottom panel of Table 6.

Broader Implications The results of this exercise suggests a link between the term structure of variance risk premia (as considered in Dew-Becker, Giglio, Le, and Rodriguez (2015), Andries, Eisenbach, Schmalz, and Wang (2015), and Aït-Sahalia, Karaman, and Mancini (2015) and the equity term structure (van Binsbergen, Brandt, and Koijen (2012)). In particular, van Binsbergen, Brandt, and Koijen (2012) show that the equity premium is closely tied to discount rates of short-term dividend cash flows. Analogously, the above return predictability exercise shows that only discount rates of *short-run* variance cash flows have implications for the equity premium. Taken together, these results appear to connect the discount rates of short-term variance cash flows to those of short-run equity cash flows, which in turn has implications for the joint modeling of equity and variance risk premia.

5 Conclusion

This paper developed a methodological framework for measuring long-run risk-neutral expectations of variance and other moments from options. By constructing sieve approximations to the term structure of state-price densities, the paper derives new closed-form option prices that can be interpreted as nonparametric extensions of the Black-Scholes formula. The sieve approximations involve basis function expansions that grow slowly with the sample size and can fit a variety of unknown DGPs, as confirmed in a simulation exercise. The methodological framework suggests future lines of research that explore option-implied term structures in the cross-section. In particular, to the extent that options written on individual stocks, industry ETFs, and even fixed-income instruments like swaps are less liquid than the S&P 500 index options considered above, the nonparametric confidence intervals provided in this paper provide a useful metric with which to compare the precision of option spanning portfolios across assets and asset classes.

The paper's main empirical application concerns the term structure of variance risk premia. Using the sieve's estimates of risk-neutral implied variance, the paper presents new findings on the relationship between variance risk premia and equity market return predictability. In particular, a decomposition of the variance risk premium into a short-run component and a forward risk premium suggests that only the short-run component predicts excess returns. This finding is at odds with existing asset pricing models that seek to explain the variance risk premium's predictive content, since these models counterfactually imply that the forward variance risk premium should also predict excess returns. Hence information contained in the term structure of risk-neutral variance provides an alternative benchmark against which to test existing asset pricing theories. The results therefore suggest future directions for expanding existing equilibrium models of the variance risk premium and return predictability to account for term structure effects.

References

Adams, R. A., and J. J. Fournier (2003): *Sobolev spaces*, vol. 140. Access Online via Elsevier.

Aït-Sahalia, Y., and J. Duarte (2003): "Nonparametric option pricing under shape restrictions," *Journal of Econometrics*, 116(1), 9–47.

Aït-Sahalia, Y., M. Karaman, and L. Mancini (2015): "The Term Structure of Variance Swaps, Risk Premia, and the Expectation Hypothesis," *Available at SSRN 2136820*.

Aït-Sahalia, Y., and A. W. Lo (1998): "Nonparametric estimation of state-price densities implicit in financial asset prices," *The Journal of Finance*, 53(2), 499–547.

Andersen, T. G., N. Fusari, and V. Todorov (2012): "Parametric inference and dynamic state recovery from option panels," Discussion paper, National Bureau of Economic Research.

Andries, M., T. Eisenbach, M. Schmalz, and Y. Wang (2015): "The term structure of the price of variance risk," Discussion paper, Working Paper.

Ang, A., and G. Bekaert (2007): "Stock return predictability: Is it there?," *Review of Financial Studies*, 20(3), 651–707.

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

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

Bakshi, G., and D. Madan (2000): "Spanning and derivative-security valuation," *Journal of Financial Economics*, 55(2), 205–238.

Bansal, R., and A. Yaron (2004): "Risks for the long run: A potential resolution of asset pricing puzzles," *The Journal of Finance*, 59(4), 1481–1509.

Bates, D. (1996): "Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche Mark options," *Review of financial studies*, 9(1), 69–107.

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

Bekaert, G., and M. Hoerova (2014): "The VIX, the variance premium and stock market volatility," *Journal of Econometrics*, forthcoming.

Billingsley, P. (1995): "Probability and measure. 1995," *John Wiley&Sons, New York*.

Black, F., and M. Scholes (1973): "The pricing of options and corporate liabilities," *The Journal of Political Economy*, pp. 637–654.

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

Bollerslev, T., D. Osterrieder, N. Sizova, and G. Tauchen (2013): "Risk and return: Long-run relations, fractional cointegration, and return predictability," *Journal of Financial Economics*, 108(2), 409–424.

BOLLERSLEV, T., G. TAUCHEN, AND H. ZHOU (2009): "Expected stock returns and variance risk premia," *Review of Financial Studies*, 22(11), 4463–4492.

BOLLERSLEV, T., AND V. TODOROV (2011): "Tails, fears, and risk premia," *The Journal of Finance*, 66(6), 2165–2211.

BONDARENKO, O. (2003): "Estimation of risk-neutral densities using positive convolution approximation," *Journal of Econometrics*, 116(1), 85–112.

CAMPBELL, J., AND J. COCHRANE (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *The Journal of Political Economy*, 107(2), 205–251.

CARR, P., AND L. WU (2009): "Variance risk premiums," *Review of Financial Studies*, 22(3), 1311–1341.

CBOE (2003): "The CBOE Volatility Index - VIX White Paper," *URL: http://www.cboe.com/micro/vix/vixwhite.pdf*.

CHEN, X. (2007): "Large sample sieve estimation of semi-nonparametric models," *Handbook of Econometrics*, 6, 5549–5632.

CHEN, X., Z. LIAO, AND Y. SUN (2014): "Sieve inference on possibly misspecified semi-nonparametric time series models," *Journal of Econometrics*, 178, 639–658.

CHEN, X., AND X. SHEN (1998): "Sieve extremum estimates for weakly dependent data," *Econometrica*, pp. 289–314.

CHRISTOFFERSEN, P., M. FOURNIER, AND K. JACOBS (2013): "The Factor Structure in Equity Option Prices," .

CONRAD, J., R. F. DITTMAR, AND E. GHYSELS (2013): "Ex ante skewness and expected stock returns," *The Journal of Finance*, 68(1), 85–124.

COPPEJANS, M., AND A. R. GALLANT (2002): "Cross-validated SNP density estimates," *Journal of Econometrics*, 110(1), 27–65.

CORSI, F. (2009): "A simple approximate long-memory model of realized volatility," *Journal of Financial Econometrics*, 7(2), 174–196.

DEW-BECKER, I., S. GIGLIO, A. LE, AND M. RODRIGUEZ (2015): "The price of variance risk," Discussion paper, National Bureau of Economic Research.

DRECHSLER, I., AND A. YARON (2011): "What's vol got to do with it," *Review of Financial Studies*, 24(1), 1–45.

DRIESSEN, J., P. J. MAENHOUT, AND G. VILKOV (2009): "The price of correlation risk: Evidence from equity options," *The Journal of Finance*, 64(3), 1377–1406.

DUFFIE, D., J. PAN, AND K. SINGLETON (2000): "Transform analysis and asset pricing for affine jump-diffusions," *Econometrica*, 68(6), 1343–1376.

EPSTEIN, L. G., AND S. E. ZIN (1991): "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis," *Journal of political Economy*, pp. 263–286.

Fackler, P. L. (2005): "Notes on matrix calculus," *North Carolina State University.*

Figlewski, S. (2008): "Estimating the Implied Risk Neutral Density," *Volatility and Time Series Econometrics: Essay in Honor of Robert F. Engle. Editors: Tim Bollerslev, Jeffrey R. Russell, Mark Watson, Oxford, UK: Oxford University Press.*

Gallant, A. R., and D. W. Nychka (1987): "Semi-nonparametric maximum likelihood estimation," *Econometrica: Journal of the Econometric Society*, pp. 363–390.

Heston, S. (1993): "A closed-form solution for options with stochastic volatility with applications to bond and currency options," *Review of financial studies*, 6(2), 327–343.

Hodrick, R. J. (1992): "Dividend yields and expected stock returns: Alternative procedures for inference and measurement," *Review of Financial studies*, 5(3), 357–386.

Jiang, G., and Y. Tian (2005): "The model-free implied volatility and its information content," *Review of Financial Studies*, 18(4), 1305–1342.

Kristensen, D., and A. Mele (2011): "Adding and subtracting Black-Scholes: a new approach to approximating derivative prices in continuous-time models," *Journal of financial economics*, 102(2), 390–415.

León, Á., J. Mencía, and E. Sentana (2009): "Parametric properties of semi-nonparametric distributions, with applications to option valuation," *Journal of Business & Economic Statistics*, 27(2), 176–192.

Leslie, R. A. (1991): "How not to repeatedly differentiate a reciprocal," *American Mathematical Monthly*, pp. 732–735.

Li, Q., and J. S. Racine (2007): *Nonparametric econometrics: Theory and practice*. Princeton University Press.

Mallows, C. L. (1973): "Some comments on C p," *Technometrics*, 15(4), 661–675.

Newey, W., and K. West (1987): "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica: Journal of the Econometric Society*, pp. 703–708.

Newey, W. K. (1991): "Uniform convergence in probability and stochastic equicontinuity," *Econometrica*, 59(4), 1161–1167.

Nickl, R., and B. M. Pötscher (2007): "Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov-and Sobolev-type," *Journal of Theoretical Probability*, 20(2), 177–199.

van Binsbergen, J., M. Brandt, and R. Koijen (2012): "On the Timing and Pricing of Dividends," *American Economic Review*, 102(4), 1596–1618.

van Binsbergen, J., W. Hueskes, R. Koijen, and E. Vrugt (2013): "Equity yields," *Journal of Financial Economics*, 110(3), 503 – 519.

Van Binsbergen, J. H., and R. S. Koijen (2015): "The Term Structure of Returns: Facts and Theory," *Available at SSRN 2597481.*

Van De Geer, S. (2000): *Empirical processes in M-estimation*, vol. 45. Cambridge university press Cambridge.

Van Der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence*. Springer.

Xiu, D. (2011): "Dissecting and Deciphering European Option Prices using Closed-Form Series Expansion," *Chicago Booth Research Paper*, (11-40).

Yatchew, A., and W. Härdle (2006): "Nonparametric state price density estimation using constrained least squares and the bootstrap," *Journal of Econometrics*, 133(2), 579–599.

A Definitions and Preliminary Theoretical Results

A.1 Defining the Sobolev Sieve Spaces

We begin by presenting a precise definition of the sieve spaces referenced in the main paper. 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.

A.1.1 The Gallant-Nychka Joint Density Spaces

Let $\mathbf{u} = (y, \mathbf{x})' \in \mathcal{Y} \times \mathcal{X} \equiv \mathcal{U}$, where $\mathcal{Y} = \mathbb{R}$ and $\mathcal{X} \subset \mathbb{R}^{d_x}$ is a compact rectangle. Let $d_u \equiv 1 + d_x$, 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. (Sobolev norms). For $1 \le p < \infty$, define the Sobolev norm of f with respect to the nonnegative weight function $\zeta(\mathbf{u})$ by

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

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

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

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

$$W^{m,p,\zeta}(\mathcal{U}) \equiv \{ f \in L^p(\mathcal{U}) : \|f\|_{m,p,\zeta} < \infty \},$$

where $1 \le p \le \infty$.

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

Definition A.2. (The 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 $\|h_0\|_{m_0+m,2,\zeta_0} \le \mathcal{B}_0$, let $\mathcal{F}^{Y,X}$ consist of those probability density functions $f(\mathbf{u})$ that have the form

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

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

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

Let

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

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.3. (The Joint Sieve Space $\mathcal{F}_K^{Y,X}$). Let $\phi(\mathbf{u}) = \exp(-\mathbf{u}'\mathbf{u}/2)/\sqrt{2\pi}$, and let $P_K(\mathbf{u})$ denote a Hermite polynomial of degree K. $\mathcal{F}_K^{Y,X}$ consists of those probability density functions that are of the form

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

with $\left\|P_K(\mathbf{u})\phi(\mathbf{u})^{1/2}\right\|_{m_0+m,2,\zeta_0} \le \mathcal{B}_0$ and $\varepsilon > \varepsilon_0$ and $\beta'\beta = 1$, where β is the stacked vector of all Hermite polynomial coefficients in $P_K(\mathbf{u})$. Denote $\mathcal{H}_K \equiv \{h \in W^{m_0+m,2,\zeta_0} : h = P_K\phi^{1/2}\}$.

Because the ultimate object of interest is a conditional density, we put additional structure on the term $\varepsilon h_0(\mathbf{u})$ to prevent explosive tail behavior when dividing by marginal densities on $\mathbf{x}$.

Assumption A.1. *(Support and Tail Conditions)*

(i) The function h_0 satisfies

$$h_0(\mathbf{u}) = \phi(y) \cdot h_x(\mathbf{x}), \tag{36}$$

where h_x is bounded away from zero on its compact support $\mathcal{X} \subseteq \mathcal{Z}$, $\phi(y)$ is Gaussian, and $\mathcal{Y} = \mathbb{R}$.

(ii) For option characteristics $\mathbf{Z} = (\kappa, \tau, r, q)$, $\mathbf{Z} \in \mathcal{Z}$, where $\mathcal{Z}$ is a compact hyperrectangle in $\mathbb{R}^{d_z}$.

Remark A.4. The compactness of $\mathcal{X}$ and the functional form of the lower bound in (36) are not constraining in empirical implementations. Since $\mathcal{X}$ represents the support of variables related to option maturity, it can be set wide enough to encompass maturities ranging from zero to 1000 years. Similarly, the decomposition of $h_0(\mathbf{u}) = \phi(y) \cdot h_x(\mathbf{x})$ is only slightly more restrictive than the workhorse choice $\phi(y) \cdot \phi(x_1) \cdot \dots \cdot \phi(x_{d_x})$. In any case, one can argue as in Gallant and Nychka (1987) that the value of ε can be set so that the term $\varepsilon h_0(\mathbf{u})$ is smaller than machine epsilon in applications. Importantly, the return variable can have unbounded support $\mathcal{Y}$. The condition on $\mathcal{Z}$ ensures integrability of the put option payoff, and furthermore enables the invocation of Sobolev embedding theorems. This assumption can be relaxed to domains satisfying a strong local Lipschitz condition (Adams and Fournier (2003, §4.9)), which is more general than what is needed for the option pricing application.

A.1.2 The Conditional Density Spaces

The transformed 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.5. (The Sieve Spaces $\mathcal{F}$ and $\mathcal{F}_K$). Define

$$\mathcal{F} \equiv \left\{ f : f(y|\mathbf{x}) = \frac{f^{Y,X}(y,\mathbf{x})}{\int_{\mathcal{Y}} f^{Y,X}(y,\mathbf{x})dy} \text{ some } f^{Y,X} \in \mathcal{F}^{Y,X} \right\} \quad \text{and}$$
$$\mathcal{F}_K \equiv \left\{ f_K : f_K(y|\mathbf{x}) = \frac{f_K^{Y,X}(y,\mathbf{x})}{\int_{\mathcal{Y}} f_K^{Y,X}(y,\mathbf{x})dy} \text{ some } f_K^{Y,X} \in \mathcal{F}_K^{Y,X} \right\}.$$

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 a Lipschitz continuous map $\Lambda : \mathcal{F}^{Y,X} \to \mathcal{F}$ (Lemma A.9 below). Note that the densities in $\mathcal{F}$ are related to the return distribution via the change of variables formula in (4).

Definition A.6. (The Option Spaces $\mathcal{P}$ and $\mathcal{P}_K$). Define

$$\mathcal{P} \equiv \{P : \mathcal{Z} \to \mathbb{R}_+ : P(\mathbf{Z}) = P(f,\mathbf{Z}) \text{ some } f \in \mathcal{F}\}$$
$$\mathcal{P}_K \equiv \{P : \mathcal{Z} \to \mathbb{R}_+ : P(\mathbf{Z}) = P(f_K,\mathbf{Z}) \text{ some } f_K \in \mathcal{F}_K\}.$$

Definition A.7. (Hölder Spaces). Define

$$C^{j,\eta}(\mathcal{Z}) = \left\{ g \in C^m(\mathcal{Z}) : \max_{|\lambda|\le j} \sup_{\mathbf{z}\in\mathcal{Z}} \left|D^\lambda g(\mathbf{z})\right| \le L \right.$$
$$\left. \max_{|\lambda|=j} \sup_{\mathbf{z}_1\neq\mathbf{z}_2\in\mathcal{Z}} \frac{\left|D^\lambda g(\mathbf{z}_1) - D^\lambda g(\mathbf{z}_2)\right|}{|\mathbf{z}_1-\mathbf{z}_2|^\eta} \le L \right\}.$$

A.2 Preliminary Results

Lemma A.8. *The following results will be invoked later on. Under Assumption A.1,*

(i) There exists a constant M such that for all marginals $f^X(\mathbf{x}) = \int_{\mathcal{Y}} f^{Y,X}(y,\mathbf{x})dy$ with $f^{Y,X} \in \mathcal{F}^{Y,X}$, one has $\|f^X\|_{m,1} \le M$.

(ii) The conditional space $\mathcal{F} \subset W^{m,1}(\mathcal{U})$.

(iii) The option space $\mathcal{P} \subset W^{m,1}(\mathcal{Z})$.

(iv) $P(f,\mathbf{Z})$ is a bounded linear functional in f, i.e. there exists M, such that $\|P\|_{m,1} \le M\|f\|_{m,1}$. Hence $P(f,\mathbf{Z})$ is locally bounded, that is, for any $f \in \mathcal{F}$, there exists a neighborhood $U \ni f$ such that for some M_U, $\sup_{g\in U}\|P(g,\mathbf{Z})\|_{m,1} \le M_U$.

(v) Let $m = j+k$, $k = d_z+1$, $\eta = 1$, $j > 0$. Then there exists a Hölder Space embedding $\mathcal{P} \hookrightarrow C^{j,\eta}(\mathcal{Z})$.

Proof. In what follows, M and C_j refer to generic constants and, as before, $\mathbf{u} = (y,\mathbf{x})'$.

(i) Step 1: Show that $f^{Y,X} \le C_3(1+\mathbf{u}'\mathbf{u})^{-\delta}$ for $\delta \in (d_u/2, \delta_0)$. First, from Definition A.2, $\|h\|_{m_0+m,2,\zeta_0} \le$

$\mathcal{B}_0$ implies

$$
\begin{aligned}
\left|\zeta_0(\mathbf{u})^{1/2}h(\mathbf{u})\right| &\le \max_{|\lambda|\le m}\sup_{\mathbf{u}}\left|D^\lambda\zeta_0(\mathbf{u})^{1/2}h(\mathbf{u})\right| = \left\|\zeta_0(\mathbf{u})^{1/2}h(\mathbf{u})\right\|_{m,\infty} \\
&\le C_1\left\|h\right\|_{m_0+m,2,\zeta_0} \qquad \text{Gallant-Nychka Lemma A.1(b)} \\
&\le C_1\mathcal{B}_0 \\
\Rightarrow \qquad \zeta_0(\mathbf{u})h(\mathbf{u})^2 &\le (C_1\mathcal{B}_0)^2 \\
h(\mathbf{u})^2 &\le (C_1\mathcal{B}_0)^2(1+\mathbf{u}'\mathbf{u})^{-\delta_0} \le (C_1\mathcal{B}_0)^2(1+\mathbf{u}'\mathbf{u})^{-\delta} \\
\text{and} \qquad |h(\mathbf{u})| &\le (C_1\mathcal{B}_0)(1+\mathbf{u}'\mathbf{u})^{-\delta/2}.
\end{aligned}
\tag{37}
$$

Since $\left\|h_0\right\|_{m_0+m,2,\zeta_0} \le \mathcal{B}_0$ as well, one has $f^{Y,X} \le C_3(1+\mathbf{u}'\mathbf{u})^{-\delta}$.
Step 2: For readability let $f(\mathbf{x}) \equiv f^X(\mathbf{x})$ and $f(y,\mathbf{x}) \equiv f^{Y,X}(y,\mathbf{x})$. Let $\alpha = (0,\alpha_1,\ldots,\alpha_{d_x})$ denote a multi-index over $\mathbf{x}$. By Step 1, dominated convergence, triangle inequality, Hölder's inequality, and Gallant-Nychka Lemma A.1(c),

$$
\begin{aligned}
\left\|f^X\right\|_{m,1} &= \sum_{|\alpha|\le m}\int_{\mathcal{X}}\left|D^\alpha f(\mathbf{x})\right|d\mathbf{x} = \sum_{|\alpha|\le m}\int_{\mathcal{X}}\left|D^\alpha\int_{\mathcal{Y}} f(y,\mathbf{x})dy\right|\,d\mathbf{x} \\
&\le \sum_{|\alpha|\le m}\int_{\mathcal{X}}\int_{\mathcal{Y}}\left|D^\alpha[h(y,\mathbf{x})^2+\varepsilon_0 h_0(y,\mathbf{x})]\right|dy\,d\mathbf{x} \\
&\le \sum_{|\alpha|\le m}\left\{2\int_{\mathcal{X}}\int_{\mathcal{Y}}\left|h(y,\mathbf{x})D^\alpha h(y,\mathbf{x})\right|dy\,d\mathbf{x} + \varepsilon_0\int_{\mathcal{X}}\int_{\mathcal{Y}}\left|D^\alpha h_0(y,\mathbf{x})\right|dy\,d\mathbf{x}\right\} \\
&\le \sum_{|\alpha|\le m}2\left\{\sup_{y,\mathbf{x}}\left|D^\alpha h(y,\mathbf{x})\right|\right\}\int_{\mathcal{X}}\int_{\mathcal{Y}}\left|h(y,\mathbf{x})\right|dy\,d\mathbf{x} \\
&+ \sum_{|\alpha|\le m}\varepsilon_0\left\{\int_{\mathcal{X}}\int_{\mathcal{Y}}\left|\phi(y)\cdot D^\alpha h_x(\mathbf{x})\right|dy\,d\mathbf{x}\right\} \\
&\le C_1\left\{\max_{|\alpha|\le m}\sup_{y,\mathbf{x}}\left|D^\alpha h(y,\mathbf{x})\right|\right\}\int_{\mathcal{X}}\int_{\mathcal{Y}}(1+y^2+\mathbf{x}'\mathbf{x})^{-\delta/2}dy \\
&+ \sum_{|\alpha|\le m}\varepsilon_0\left\{\int_{\mathcal{X}}\left|D^\alpha h_x(\mathbf{x})\right|\ d\mathbf{x}\right\} \\
&\le C_2\left\|h\right\|_{m,\infty} + \sum_{|\alpha|\le m}\varepsilon_0\left\{\sup_{\mathbf{x}}\left|D^\alpha h_x(\mathbf{x})\right|\right\} leb(\mathcal{X}) \le C_2\left\|h\right\|_{m,\infty} + C_3\left\|h_x\right\|_{m,\infty} \\
&\le C_4\left\|h\right\|_{m_0+m,2,\zeta} \le C_4\mathcal{B}_0 < \infty,
\end{aligned}
$$

where $leb(\mathcal{X})$ is the Lebesgue measure of the compact hyperrectangle $\mathcal{X}$. Thus $f^X \in W^{m,1}(\mathcal{X})$.

(ii) Step 1: Show $\left\|1/f^X\right\|_{m,1} \le M$. Apply a quotient derivative formula (e.g. Leslie (1991)) and the bound on f^X in part (i) to get $\left\|1/f^X\right\|_{m,1} \le C_1\left\|f^X\right\|_{m,1}$.

Step 2: By Leibniz' formula, Hölder's inequality, and Step 1,

$$
\begin{aligned}
\|f\|_{m,1} &= \sum_{|\lambda|\le m}\int_{\mathcal{X}}\int_{\mathcal{Y}}\left|D^{\lambda}f(y|\mathbf{x})\right| dy\, d\mathbf{x} \le \sum_{|\lambda|\le m}\sum_{\beta\le\lambda}\begin{bmatrix}\lambda\\ \beta\end{bmatrix}\int_{\mathcal{X}}\int_{\mathcal{Y}}\left|D^{\beta}f(y,\mathbf{x})D^{\lambda-\beta}\frac{1}{f(\mathbf{x})}\right| dy\, d\mathbf{x}\\
&= \sum_{|\lambda|\le m}\sum_{\beta\le\lambda}\begin{bmatrix}\lambda\\ \beta\end{bmatrix}\int_{\mathcal{X}}\int_{\mathcal{Y}}\left|D^{\beta}\left\{f(y,\mathbf{x})\right\}\zeta(y,\mathbf{x})D^{\lambda-\beta}\left\{\frac{1}{f(\mathbf{x})}\right\}\zeta(y,\mathbf{x})^{-1}\right| dy\, d\mathbf{x}\\
&\le C_1\sum_{|\lambda|\le m}\sum_{\beta\le\lambda}\begin{bmatrix}\lambda\\ \beta\end{bmatrix}\left\{\sup_{y,\mathbf{x}}\left|D^{\beta}f(y,\mathbf{x})\right|\zeta(y,\mathbf{x})\right\}\int_{\mathcal{X}}D^{\lambda-\beta}\left\{\frac{1}{f(\mathbf{x})}\right\}\int_{\mathcal{Y}}(1+y^2+\mathbf{x}'\mathbf{x})^{-\delta}dy\, d\mathbf{x}\\
&\le C_2\left\{\max_{|\lambda|\le m}\sup_{y,\mathbf{x}}\left|D^{\lambda}f(y,\mathbf{x})\right|\zeta(y,\mathbf{x})\right\}\left\|f^{X}\right\|_{m,1} = C_3\left\|f^{Y,X}\right\|_{m,\infty,\zeta}<\infty,
\end{aligned}
\tag{38}
$$

and hence conditional densities $f \in W^{m,1}(\mathcal{U})$.

(iii) Let $\psi(y,\mathbf{Z}) = e^{-r\tau}[\kappa - S_0\exp(\mu(\mathbf{Z})+\sigma(\mathbf{Z})y)]\cdot 1[y\le d(\mathbf{Z})]$ denote the discounted option payoff. Since $\mathcal{X}\subseteq\mathcal{Z}$, write $\mathcal{Z} = \mathcal{X}\bigoplus(\mathcal{Z}-\mathcal{X})$. Then by Leibniz formula and Hölder's inequality, multi-index $\lambda = (\lambda_1,\ldots,\lambda_{d_z})$, Assumption A.1, and part (ii)

$$
\begin{aligned}
\|P\|_{m,1} &= \sum_{|\lambda|\le m}\int_{\mathcal{Z}}\left|D^{\lambda}P(\mathbf{Z})\right|\; d\mathbf{Z}\\
&\le \sum_{|\lambda|\le m}\sum_{|\beta|\le\lambda}\begin{bmatrix}\lambda\\ \beta\end{bmatrix}\int_{\mathcal{Z}-\mathcal{X}}\int_{\mathcal{X}}\int_{\mathcal{Y}}\left|D^{\beta}[\psi(y,\mathbf{Z})]D^{\lambda-\beta}f(y|\mathbf{X})\right| dy\, d\mathbf{x}\, d(\mathbf{z}-\mathbf{x})\\
&\le C_1\sum_{|\lambda|\le m}\sum_{|\beta|\le\lambda}\begin{bmatrix}\lambda\\ \beta\end{bmatrix}\int_{\mathcal{Z}-\mathcal{X}}\int_{\mathcal{X}}\int_{\mathcal{Y}}\left|D^{\lambda-\beta}f(y|\mathbf{X})\right| dy\, d\mathbf{x}\, d(\mathbf{z}-\mathbf{x})\\
&\le C_2\left\|f\right\|_{m,1}<\infty.
\end{aligned}
\tag{39}
$$

(iv) Follows directly from (39) and the property that bounded linear functionals are locally bounded.

(v) This is a consequence of the Sobolev Embedding Theorem (Adams and Fournier (2003, Theorem 4.12, Part II)) and the regularity condition on $\mathcal{Z}$.

□

Lemma A.9. *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*

$$f(y|\mathbf{x}) \equiv \Lambda(f^{Y,X}(y,\mathbf{x})) = \frac{f^{Y,X}(y,\mathbf{x})}{\int_{\mathcal{Y}}f^{Y,X}(y,\mathbf{x})dy}$$

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

Proof. Let $f_0(\mathbf{x}) = \int_{\mathbf{R}}f_0^{Y,X}(y,\mathbf{x})dy$ and $f_j(\mathbf{x}) = \int_{\mathbf{R}}f_j^{Y,X}(y,\mathbf{x})dy$ denote the marginal distributions of $\mathbf{X}$ of generic $f_0^{Y,X}, f_j^{Y,X}\in\mathcal{F}^{Y,X}$.

$$\|f_j(y|\mathbf{x}) - f_0(y|\mathbf{x})\|_{m,1} = \sum_{|\lambda|\le m}\int_{\mathcal{X}}\int_{\mathcal{Y}}\left|D^{\lambda}\left\{\frac{f_j(y,\mathbf{x})}{f_j(\mathbf{x})} - \frac{f_0(y,\mathbf{x})}{f_0(\mathbf{x})}\right\}\right| dy\, d\mathbf{x}$$

$$
\begin{aligned}
&= \sum_{|\lambda|\le m} \int_{\mathcal{X}} \int_{\mathcal{Y}} \left| D^{\lambda} \left\{ f_j(y,\mathbf{x}) \left[\frac{1}{f_0(\mathbf{x})} + \frac{f_0(\mathbf{x}) - f_j(\mathbf{x})}{f_0(\mathbf{x}) f_j(\mathbf{x})} \right] - \frac{f_0(y,\mathbf{x})}{f_0(\mathbf{x})} \right\} \right| dy\, d\mathbf{x} \\
&\le \sum_{|\lambda|\le m} \int_{\mathcal{X}} \int_{\mathcal{Y}} \left| D^{\lambda} \left\{ [f_j(y,\mathbf{x}) - f_0(y,\mathbf{x})] \frac{1}{f_0(\mathbf{x})} \right\} \right| dy\, d\mathbf{x} \\
&+ \sum_{|\lambda|\le m} \int_{\mathcal{X}} \int_{\mathcal{Y}} \left| D^{\lambda} \left\{ [f_j(\mathbf{x}) - f_0(\mathbf{x})] \frac{1}{f_0(\mathbf{x}) f_j(\mathbf{x})} \right\} \right| dy\, d\mathbf{x} \\
&\le C_1 \left\| f_j(y,\mathbf{x}) - f_0(y,\mathbf{x}) \right\|_{m,\infty,\zeta},
\end{aligned}
$$

where the last inequality follows from derivations analogous to (38). The result follows. ☐

This continuity result implies that the conditional spaces inherit the topological structure from the parent joint spaces. Moreover, the strengthening to Lipschitz continuity 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.9 is also surjective by definition.

Lemma A.9 gives rise to the following two critical properties of the sieve spaces.

Lemma A.10. *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 $K \ge 0$.

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

Proof. (i) and (ii) are a result of the above Lemma A.9. To see this, fix any $K \ge 0$ and let $\{f_{K,n}\}$ denote a sequence of joint densities in $\mathcal{F}_K^{X,Y}$. By definition it can be written as $f_{K,n} = h_{K,n}^2 + \varepsilon h_0$, where $h_{K,n} = P_{K,n}\phi^{1/2}$ satisfies $\|h_{K,n}\|_{m_0+m,2,\zeta_0} \le \mathcal{B}_0$. By Lemma A.4 in Gallant and Nychka (1987), there exists h with $\|h\|_{m,\infty,\zeta^{1/2}} < \infty$ and a subsequence $\{h_{K,n_j}\}$ with $\lim_{j\to\infty} \|h_{K,n_j} - h\|_{m,\infty,\zeta^{1/2}} = 0$. Then by Gallant and Nychka's Lemma A.3, one has $\|h^2\|_{m,\infty,\zeta} < \infty$ and $\lim_{j\to\infty} \left\|h_{K,n_j}^2 - h^2\right\|_{m,\infty,\zeta} = 0$, whence $\lim_{j\to\infty} \|f_{K,n_j} - f\|_{m,\infty,\zeta} = 0$. Thus $\mathcal{F}_K^{X,Y}$ is compact in the topology generated by $\|\cdot\|_{m,\infty,\zeta}$. Finally, because Λ in the above Lemma A.9 is $\|\cdot\|_{m,\infty,\zeta} - \|\cdot\|_{m,1}$ continuous and surjective (by construction), one has that the conditional space $\mathcal{F}_K$ is compact in the topology generated by $\|\cdot\|_{m,1}$. To show (ii), note that for any joint density $f^{X,Y} \in \mathcal{F}^{X,Y}$, one has $f^{X,Y} = h^2 + \varepsilon h_0$. By Gallant and Nychka's Lemma A.5, there exists a sequence $\{h_K\}$ such that $\lim_{K\to\infty} \|h_K - h\|_{m_0+m,2,\zeta} = 0$, and by their Lemmas A.1-A.3, this implies $\lim_{K\to\infty} \|h_K^2 - h^2\|_{m,\infty,\zeta} = 0$. One therefore has $\lim_{K\to\infty} \left\|f_K^{X,Y} - f\right\|_{m,\infty,\zeta} = 0$, which implies that $\cup_{K=0}^{\infty} \mathcal{F}_K^{X,Y}$ is dense in $\mathcal{F}^{X,Y}$. Applying the above Lemma A.9 and noting that Λ is continuous and surjective shows that the conditional space $\cup_{K=0}^{\infty} \mathcal{F}_K$ is dense in $\mathcal{F}$ with the topology generated by $\|\cdot\|_{m,1}$. ☐

B Derivation of Main Results

B.1 Closed-Form Option Pricing

This section begins with a proof of the closed-form pricing expression stated in Proposition 1. To obtain these closed-form option prices, it is convenient to first establish a linear representation for the conditional sieve densities f_K of (11). This is done by expanding the squared polynomial term in the joint densities of (10) using results from León, Mencía, and Sentana (2009).

Lemma B.1. *Any $f_K \in \mathcal{F}_K$ can be expressed in the form*

$$f_K(y|\tau) = \sum_{k=0}^{2K_y} \gamma_k(B,\tau)H_k(y)\phi(y), \tag{40}$$

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 (10).

Proof. The task is to show that conditional sieve densities have the representation in (40), which is required to get closed-form option prices. 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

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

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

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

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

Proof of Proposition 1. The proof extends Proposition 9 of León, Mencía, and Sentana (2009) to allow for conditioning on τ. The plug-in estimator of the population option price in equation (5), is given by

$$\begin{aligned}P(f_K,\mathbf{Z}) &= e^{-r\tau}\int_{-\infty}^{d(\mathbf{Z})}\left(\kappa - Se^{\mu(\mathbf{Z})+\sigma(\mathbf{Z})Y}\right)f_K(Y|\tau)dY \\ &= \kappa e^{-r\tau}\int_{-\infty}^{d(\mathbf{Z})} f_K(Y|\tau)dY - Se^{-r\tau+\mu(\mathbf{Z})}\int_{-\infty}^{d(\mathbf{Z})} e^{\sigma(\mathbf{Z})Y}f_K(Y|\tau)dY.\end{aligned} \tag{41}$$

Using Lemma B.1, the integral in the first term becomes

$$\begin{aligned}&\int_{-\infty}^{d(\mathbf{Z})} f_K(Y|\tau)dY = \int_{-\infty}^{d(\mathbf{Z})}\left[\sum_{k=0}^{2K_y}\gamma_k(B,\tau)H_k(Y)\phi(Y)\right]dY \\ &= \sum_{k=0}^{2K_y}\gamma_k(B,\tau)\int_{-\infty}^{d(\mathbf{Z})}H_k(Y)\phi(Y)dY = \Phi(d(\mathbf{Z})) - \sum_{k=1}^{2K_y}\frac{\gamma_k(B,\tau)}{\sqrt{k}}H_{k-1}(d(\mathbf{Z}))\phi(d(\mathbf{Z})),\end{aligned} \tag{42}$$

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 (41) 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 $k \geq 1$,

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

Thus,

$$\begin{aligned}
&\int_{-\infty}^{d(\mathbf{Z})} e^{\sigma(\mathbf{Z})Y} f_K(Y|\tau)dY = \int_{-\infty}^{d(\mathbf{Z})} e^{\sigma(\mathbf{Z})Y}\left[\sum_{k=0}^{2K_y}\gamma_k(B,\tau)H_k(Y)\phi(Y)\right]dY \\
&= \sum_{k=0}^{2K_y}\gamma_k(B,\tau)\int_{-\infty}^{d(\mathbf{Z})} e^{\sigma(\mathbf{Z})Y} H_k(Y)\phi(Y)dY = \sum_{k=0}^{2K_y}\gamma_k(B,\tau)I_k^*(d(\mathbf{Z})) \\
&= e^{\sigma(\mathbf{Z})^2/2}\Phi(d(\mathbf{Z}) - \sigma(\mathbf{Z})) + \sum_{k=1}^{2K_y}\gamma_k(B,\tau)I_k^*(d(\mathbf{Z})),
\end{aligned} \tag{43}$$

where $\gamma_0(B,\mathbf{Z}) = 1$. Plugging equations (42) and (43) into (41) obtains the desired result. The proof for call options is analogous and is therefore omitted. □

B.2 Asymptotic Theory

We establish the consistency, rate of convergence, and asymptotic distribution that lead to the results in Section 2.

B.2.1 Consistency

Assumption B.1. *Assume*

(i) The option data and characteristics $\{\Xi_i\}_{i=1}^n \equiv \{(P_i,\mathbf{Z}_i)\}_{i=1}^n$ *are independent with* $\mathbb{E}\,|\Xi_i|^{2+\delta} < \infty$*, and the weighting function satisfies* $\mathbb{E}[W(\mathbf{Z}_i)^2] < \infty$*.*

(ii) The true state-price density $f_0 \in \mathcal{F}$ *and satisfies* $P_0 = \mathbb{E}[P(f_0,\mathbf{Z})|\mathbf{Z}]$*.*

Assumption B.1 is standard and very mild. It says that the options are observed with conditional mean-zero errors with bounded $2+\delta$ moments. Assumption B.1 and Lemma A.8 are sufficient to derive consistency:

Proposition 3. *(Consistency) Assumptions A.1 and B.1 imply* $\left\|\widehat{P}_n - P_0\right\|_2 \xrightarrow{p} 0$*.*

Proof. Let $L(f) = \mathbb{E}\{-\frac{1}{2}[P - P(f, \mathbf{Z})]^2 W\} \equiv \mathbb{E}\{\ell(f, \Xi)\}$, where $\Xi \equiv (P, \mathbf{Z})$, and $W = W(\mathbf{Z})$ is a strictly positive weighting function. ℓ is concave in f, and L is strictly concave in f. Let $d(f_1, f_2) \equiv \|f_1 - f_2\|_{m,1}$ denote the state-price density consistency norm. The goal is to estimate the unknown $P_0(\mathbf{Z}) = \mathbb{E}[P|\mathbf{Z}]$ 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 $\varepsilon > 0$, $L(f_0) > \sup_{\{f \in \mathcal{F}: d(f, f_0) \geq \varepsilon\}} L(f)$

Condition 3.2'.

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

(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 $\{\Xi_i\}_{i=1}^n$ for all $f \in \mathcal{F}_K$

(ii) For any data $\{\Xi_i\}_{i=1}^n$, $L_n(f)$ 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 $K \geq 1$, $\sup_{f \in \mathcal{F}_K} |L_n(f) - L(f)| = 0$.

We verify each of these conditions in turn but require some preliminary results that relate option prices to state-price densities:

Lemma B.2. *Assumption A.1 implies*

$$\|P(f, \mathbf{Z}) - P(f_0, \mathbf{Z})\|_2 \leq C_1 \|P(f, \mathbf{Z}) - P(f_0, \mathbf{Z})\|_{m,1} \leq C_2 d(f, f_0).$$

Proof.

$$\|P(f, \mathbf{Z}) - P(f_0, \mathbf{Z})\|_2^2 \leq C_1 \|P(f, \mathbf{Z}) - P(f_0, \mathbf{Z})\|_\infty^2 \leq C_2 \|P(f, \mathbf{Z}) - P(f_0, \mathbf{Z})\|_{m,1}^2 \leq C_3 d(f, f_0)^2,$$

where the first inequality is due to the compactness of the domain $\mathcal{Z}$ (Assumption A.1 (ii)), the second inequality follows from a Sobolev Embedding Theorem (Adams and Fournier (2003, Theorem 4.12, Part I, Case A)), and the third inequality from Lemma A.8 (iv). C_j denote generic constants. ☐

Lemma B.3. $P(f_1, \mathbf{Z}) = P(f_2, \mathbf{Z})$ *if and only if* $f_1 = f_2$ *almost everywhere.*

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

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

Since this holds for every κ, the result follows. ☐

Condition 3.1': Assumption B.1 (ii) implies $L(f_0) = 0 > -\infty$. Also,

$$\begin{aligned} L(f_0) - L(f) &= -\mathbb{E}\{\frac{1}{2}[P - P(f_0, \mathbf{Z})]^2 W(\mathbf{Z})\} + \mathbb{E}\{\frac{1}{2}[P - P(f, \mathbf{Z})]^2 W(\mathbf{Z})\} \\ &= \frac{1}{2}\mathbb{E}\{[P(f, \mathbf{Z}) - P(f_0, \mathbf{Z})][-2P + P(f, \mathbf{Z}) + P(f_0, \mathbf{Z})]W(\mathbf{Z})\} \\ &= \frac{1}{2}\mathbb{E}\{[P(f, \mathbf{Z}) - P(f_0, \mathbf{Z})]^2 W(\mathbf{Z})\} = \frac{1}{2}\left\|P(f, \mathbf{Z}) - P(f_0, \mathbf{Z})\right\|_2^2 \leq C_1\, d(f, f_0)^2, \end{aligned}$$

by Lemma B.2. Thus, as $d(f_n, f_0) \to 0$, one has $L(f_0) - L(f) = |L(f_0) - L(f)| \to 0$. 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 $\|P(f, \mathbf{Z}) - P(f_0, \mathbf{Z})\|_2 < \eta$. The contrapositive of this statement reads: Given any $\varepsilon > 0$, there exists $\eta > 0$ such that if $d(f, f_0) \geq \varepsilon$, then $\|P(f, \mathbf{Z}) - P(f_0, \mathbf{Z})\|_2 \geq \eta$. Now let $\varepsilon > 0$ be given as in Condition 3.1'(ii), and consider any $f \in \{f \in \mathcal{F} : d(f, f_0) \geq \varepsilon\}$. By the previous derivations,

$$L(f_0) - L(f) = \frac{1}{2}\left\|P(f, \mathbf{Z}) - P(f_0, \mathbf{Z})\right\|_2^2 \geq \frac{1}{2}\eta^2,$$

so

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

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.10 (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 B.1 (i) implies that P_i is almost surely finite, i.e. $\exists$ a Borel set Ω_F with $|P_i(\omega)| < \infty$ for all $\omega \in \Omega_F$,[17] and Lemma A.8 (iv) implies $P(f, \mathbf{Z}_i)$ is locally bounded $\mathbb{P} - a.s.$ on $\mathcal{F}$. Therefore $P_i - P(f, \mathbf{Z}_i)$ is finite on Ω_F.

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

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

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

where the last inequality follows from the mean value theorem, and Lemma A.8 (iv) 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, $L_n(f) = \frac{1}{n}\sum_{i=1}^n -\frac{1}{2}[P_i - P(f, \mathbf{Z}_i)]^2 W(\mathbf{Z}_i)$ 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.10 (i).
Condition 3.5(i): Finally, we require the uniform convergence of the empirical criterion over sieves, i.e. for all $K \geq 1$, $\sup_{f \in \mathcal{F}_K} |L_n(f) - L(f)| \xrightarrow{p} 0$ as $n \to \infty$, where $L_n(f) = \frac{1}{n}\sum_{i=1}^n -\frac{1}{2}[P_i - P(f, \mathbf{Z})]^2 W_i$. First, note that by Assumption B.1 (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

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

The first term is $O_p(1)$ by Assumption B.1 (i). The second term is also $O_p(1)$ by the following arguments. By Lemma A.10 (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 Lemma A.8 (iv) 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, $\sup_{f \in U_i} P(f, \mathbf{Z}) \leq M_i$ $\mathbb{P} - a.s.$ Then $M = \max\{M_1, \ldots, M_N\}$ is a bound on $\sup_{g \in \mathcal{F}_K} |P(g, \mathbf{Z}_i)|$, so the second term in the above display is $O_p(1)$ under Assumption B.1 (i). Hence, by the mean value theorem, for $f_1, f_2 \in \mathcal{F}_K$,

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

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 $\sup_{f \in \mathcal{F}_K} |L_n(f) - L(f)| \xrightarrow{p} 0$, 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 B.2 gives $\left\| P(\widehat{f}_n, \mathbf{Z}) - P(f_0, \mathbf{Z}) \right\|_2 \xrightarrow{p} 0$. □

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 $P(f, \mathbf{Z})$ is uniquely identified by the state-price density f (Lemma B.3). 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 (Appendix A and Lemma A.9). The Gallant-Nychka class of densities requires h to reside in $\mathcal{H}$, a closed Sobolev ball of some radius $\mathcal{B}_0$. The rate result obtained below hinges on the observation that the collection of possible option prices $\mathcal{P}$ is ultimately Lipschitz in the index parameter $h \in \mathcal{H}$. Therefore, the size and complexity of $\mathcal{P}$, as measured by its $L^2(\mathbb{R}^{d_z}, \mathbb{P})$ bracketing number, is bounded by the covering number of the Sobolev ball $\mathcal{H}$ (see Van Der Vaart and Wellner (1996)).

B.2.2 Rate of Convergence

Assumption B.2. *$\sigma_e^2(\mathbf{Z}) \equiv \mathbb{E}[e^2|\mathbf{Z}]$ and $W(\mathbf{Z})$ are bounded above and away from zero, where $e = P - P_0(\mathbf{Z})$.*

Assumption B.3. *The deterministic approximation error rate satisfies*

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

for some $\alpha > 0$, where $h \in \mathcal{H}$ and its orthogonal projection $\pi_K h \in \mathcal{H}_K$ are defined in Definitions A.2 and A.3, and where $K \equiv [K_y + 1][K_{x,1} + 1] \dots [K_{x,d_x} + 1]$ denotes the total number of series terms for functions in H_K.

Assumption B.4. *For state-price densities in $W^{m,1}(\mathcal{Y} \times \mathcal{X})$, we have $m \geq d_u + 2$.*

Assumption B.2 is mild and commonly adopted in the literature (see Chen (2007)). Assumption B.3 takes as given the deterministic approximation error rate, and Assumption B.4 imposes additional smoothness in order to invoke Sobolev embedding theorems (see Adams and Fournier (2003)).

Proposition 4. *Under Assumptions A.1 and B.1–B.4,*

$$\begin{aligned}\left\|\widehat{P}_n - P_0\right\|_2 &= O_P(\varepsilon_n),\\ \varepsilon_n &= \max\{n^{-(m_0+m)/(2(m_0+m)+d_u)}, n^{-\alpha d_u/(2(m_0+m)+d_u)}\}.\end{aligned}$$

and $K_n \asymp n^{d_u/(2(m_0+m)+d_u)}$.

Proof. Recall that the option prices $\mathcal{P}(\mathbf{Z})$ are generated by a conditional density, i.e. $\mathcal{P}(\mathbf{Z}) \equiv P(f, \mathbf{Z})$, 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 independent sample, which is already assumed in Assumption B.1. 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,

$$\sup_{P \in B_\varepsilon(P_0)} Var\big(\ell(P, \Xi_i) - \ell(P_0, \Xi_i)\big) \leq C_1 \varepsilon^2.$$

Condition 3.8. For all $\delta > 0$, there exists a constant $s \in (0, 2)$ such that

$$\sup_{P \in B_\delta(P_0)} \big|\ell(P, \Xi_i) - \ell(P_0, \Xi_i)\big| \leq \delta^s U(\Xi_i),$$

with $\mathbb{E}[U(\Xi_i)^\gamma] \leq C_2$ for some $\gamma \geq 2$.

First, note that $\ell(P, \Xi_i) - \ell(P_0, \Xi_i) = W(\mathbf{Z}_i)[P(\mathbf{Z}_i) - P_0(\mathbf{Z}_i)]\{e_i + \frac{1}{2}[P(\mathbf{Z}_i) - P_0(\mathbf{Z}_i)]\}$. Then

$$\begin{aligned}
\mathbb{E}\{[\ell(P, \Xi_i) - \ell(P_0, \Xi_i)]^2\} &= \mathbb{E}\Big\{W(\mathbf{Z}_i)^2[P(\mathbf{Z}_i) - P_0(\mathbf{Z}_i)]^2\{e_i + \frac{1}{2}[P(\mathbf{Z}_i) - P_0(\mathbf{Z}_i)]\}^2\Big\} \\
&= \mathbb{E}\{W(\mathbf{Z}_i)^2[P(\mathbf{Z}_i) - P_0(\mathbf{Z}_i)]^2 e_i^2\} + \mathbb{E}\{\frac{1}{4}W(\mathbf{Z}_i)^2[P(\mathbf{Z}_i) - P_0(\mathbf{Z}_i)]^4\} \\
&= \mathbb{E}\{W(\mathbf{Z}_i)^2[P(\mathbf{Z}_i) - P_0(\mathbf{Z}_i)]^2\sigma_e^2(\mathbf{Z}_i)\} + \frac{1}{4}\mathbb{E}\{W(\mathbf{Z}_i)^2[P(\mathbf{Z}_i) - P_0(\mathbf{Z}_i)]^4\} \\
&\leq const.\, \|P - P_0\|_2^2 + \frac{1}{4}\mathbb{E}\{W(\mathbf{Z}_i)^2[P(\mathbf{Z}_i) - P_0(\mathbf{Z}_i)]^4\}
\end{aligned}$$

where the last inequality uses the bound from Assumption B.2. The second term on the RHS can be further bounded,

$$\begin{aligned}
\mathbb{E}\{W(\mathbf{Z}_i)^2[P(\mathbf{Z}_i) - P_0(\mathbf{Z}_i)]^4\} &\leq C \sup_{\mathbf{Z}\in\mathcal{Z}}[P(\mathbf{Z}) - P_0(\mathbf{Z})]^2\mathbb{E}\{[P(\mathbf{Z}_i) - P_0(\mathbf{Z}_i)]^2 W(\mathbf{Z}_i)\} \\
&= C\|P - P_0\|_\infty^2 \|P - P_0\|_2^2
\end{aligned}$$

The smoothness of P and P_0 can be used to bound $\|P - P_0\|_\infty^2$ as follows. Let $\eta = 1$ and let $j = m - (d_z + 1)$, which is greater than or equal to 1 by Assumption B.4. Thus, by Lemma A.8 (v), there exists a Hölder Space embedding $\mathcal{P} \hookrightarrow C^{j,\eta}(\mathcal{Z})$. Applying Lemma 2 in Chen and Shen (1998), one then has $\|P - P_0\|_\infty \leq \|P - P_0\|^{2/(2+d_z)}$. Therefore

$$\mathbb{E}\{W(\mathbf{Z}_i)^2[P(\mathbf{Z}_i) - P_0(\mathbf{Z}_i)]^4\} \leq C\|P - P_0\|_2^{2+4/(2+d_z)},$$

and one has

$$\mathbb{E}\{[\ell(P, \Xi_i) - \ell(P_0, \Xi_i)]^2\} \leq const.\, \|P - P_0\|_2^2 + \frac{C}{4}\|P - P_0\|_2^{2+4/(2+d_z)}.$$

This implies that Condition 3.7 is satisfied for all $\varepsilon \leq 1$.

To show Condition 3.8, note that

$$\begin{aligned}
|\ell(P, \Xi_i) - \ell(P_0, \Xi_i)| &= \left|[P(\mathbf{Z}_i) - P_0(\mathbf{Z}_i)]\big[e_i + \frac{1}{2}[P_0(\mathbf{Z}_i) - P(\mathbf{Z}_i)]\big]\right| \\
&\leq const.\, \|P - P_0\|_\infty \{|e_i| + \frac{1}{2}\|P_0\|_\infty + \frac{1}{2}\|P\|_\infty\}.
\end{aligned}$$

The terms involving $\|P_0\|_\infty$ and $\|P\|_\infty$ are bounded by Assumption B.1 as well as the arguments in the proof of Proposition 3. Thus Lemma 2 in Chen and Shen (1998) and another appeal to the Sobolev Embedding Theorem imply that

$$\begin{aligned}
|\ell(P, \Xi_i) - \ell(P_0, \Xi_i)| &\leq const.\, \|P - P_0\|_\infty U(\Xi_i) \\
&\leq const. U(\Xi_i)\|P - P_0\|_2^{2/(2+d_z)}
\end{aligned}$$

for $U(\Xi_i) = |e_i| + const.$ Thus $s = 2/(2 + d_z)$ is the required modulus of continuity, and $\gamma = 2$ by Assumption B.1. 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

$$\mathcal{G}_n = \{\ell(P, \Xi_i) - \ell(P_0, \Xi_i) : \|P - P_0\|_2 \leq \delta,\ P \in \mathcal{P}_{K_n}\}. \tag{44}$$

Consider the following chain of inequalities

$$\begin{aligned} |\ell(P, \Xi_i) - \ell(P_0, \Xi_i)| &= |[P(\mathbf{Z}_i) - P_0(\mathbf{Z}_i)| \left| e_i + \frac{1}{2}[P_0(\mathbf{Z}_i) - P(\mathbf{Z}_i)] \right| \\ &\leq M_1 \|f - f_0\|_{m,1} U(\Xi_i) \\ &\leq M_2 U(\Xi_i) \left\| f^{Y,X} - f_0^{Y,X} \right\|_{m,\infty,\zeta_0} \qquad \text{by Lemma A.9} \\ &= M_2 U(\Xi_i) \left\| (h^{Y,X})^2 - (h_0^{Y,X})^2 \right\|_{m,\infty,\zeta_0} \qquad \text{by Def. A.2} \\ &\leq M_3 U(\Xi_i) \left\| h^{Y,X} - h_0^{Y,X} \right\|_{m_0+m,2,\zeta_0} \end{aligned} \tag{45}$$

To see the last inequality, observe that

$$\begin{aligned} \left\| (h^{Y,X})^2 - (h_0^{Y,X})^2 \right\|_{m,\infty,\zeta_0} &\leq C \left\| h^{Y,X} + h_0^{Y,X} \right\|_{m,\infty,\zeta_0^{1/2}} \left\| h^{Y,X} - h_0^{Y,X} \right\|_{m,\infty,\zeta_0^{1/2}} \\ &\leq C_1 \left\| \zeta_0^{1/2}(h^{Y,X} + h_0^{Y,X}) \right\|_{m,\infty} C_2 \left\| \zeta_0^{1/2}(h^{Y,X} - h_0^{Y,X}) \right\|_{m,\infty} \\ &\leq C_3 \left\| h^{Y,X} + h_0^{Y,X} \right\|_{m_0+m,2,\zeta_0} C_4 \left\| h^{Y,X} - h_0^{Y,X} \right\|_{m_0+m,2,\zeta_0} \\ &\leq C_3(2\mathcal{B}_0) C_4 \left\| h^{Y,X} - h_0^{Y,X} \right\|_{m_0+m,2,\zeta_0}. \end{aligned} \tag{46}$$

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 $\mathcal{H}_n \equiv \mathcal{H}_{K_n}$ 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

$$N_{[\,]}(w, \mathcal{G}_n, \|\cdot\|_2) \leq N\left(\frac{w}{2CM_3}, \mathcal{H}_n, \|\cdot\|_{m_0+m,2,\zeta_0}\right),$$

where the RHS is 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

$$N_{[\,]}(w, \mathcal{G}_n, \|\cdot\|_2) \leq N\left(\frac{w}{2CM_3}, \mathcal{H}_n, \|\cdot\|_{m_0+m,2,\zeta_0}\right) \leq \left(1 + \frac{8\mathcal{B}_0 CM_3}{w}\right)^{K_n}.$$

Therefore,

$$\frac{1}{\sqrt{n}\delta_n^2} \int_{b\delta_n^2}^{\delta_n} \sqrt{H_{[\,]}(w, \mathcal{G}_n, \|\cdot\|_2)}\, dw \leq \frac{1}{\sqrt{n}\delta_n^2} \int_{b\delta_n^2}^{\delta_n} \sqrt{K_n \log\left(1 + \frac{8\mathcal{B}_0 CM_3}{w}\right)}\, dw \leq C \frac{1}{\sqrt{n}\delta_n^2} \sqrt{K_n}\delta_n,$$

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 B.3 yields the approximation error rate

$$\|[P(\mathbf{Z}_i) - P_0(\mathbf{Z}_i)\|_2 \leq const. \left\| h^{Y,X} - h_0^{Y,X} \right\|_{m_0+m,2,\zeta_0} = O(K_n^{-\alpha}) = O\left(n^{\frac{-\alpha d_u}{2(m_0+m)+d_u}} \right),$$

where the inequality follows from the ones in (45). Applying Chen (2007)'s Theorem 3.2 yields the stated result. $\square$

B.2.3 Asymptotic Distribution of Option Portfolios

Let $\Xi_i \equiv (P_i, \mathbf{Z}_i)$ denote observations on option prices and characteristics, and define $\ell(\beta, \Xi_i) \equiv -\frac{1}{2}[P_i - P(\beta, \mathbf{Z}_i)]^2 W_i$ and $\widehat{R}_{K_n} = -\frac{1}{n}\sum_{i=1}^n \frac{\partial^2 \ell(\widehat{\beta}_n, \Xi_i)}{\partial\beta\partial\beta'}$.

Assumption B.5.

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

(ii) $\lim_{K_n\to\infty} \left\| \frac{\partial\Gamma(P_0)}{\partial P}[\frac{\partial P_{K_n}}{\partial\beta}] \right\|_2^2 < \infty$.

(iii) The deterministic sieve approximation rate (Assumption B.3) satisfies $\alpha > \frac{2(m_0+m)+d_u}{4d_u}$.

(iv) For the functional $\Gamma(\cdot)$ in (16), $g(\cdot)$ has bounded first and second derivatives over the domain of interest.

Proof of Proposition 2. The aim is to connect the sieve asymptotic theory with simple non-linear least squares implementations by using techniques from Chen, Liao, and Sun (2014).

Riesz Representers Let $P_{K_n}(\mathbf{Z}) = P(\beta_n, \mathbf{Z})$ and $\ell(P_{K_n}, \mathbf{Z}) = \ell(\beta_n, \mathbf{Z})$ for the purposes of this proof. Then following Chen, Liao, and Sun (2014), one can define the inner product

$$\langle P_1 - P_0, P_2 - P_0 \rangle \equiv -\mathbb{E}\{r(P_0, \Xi)[P_1 - P_0, P_2 - P_0]\},$$

where

$$r(P_0, \Xi)[P_1 - P_0, P_2 - P_0] \equiv \left. \frac{\partial \ell'(P_0 + \eta(P_2 - P_0), \Xi)[P_1 - P_0]}{\partial\eta} \right|_{\eta=0}$$

can be interpreted as a second-order Gateaux derivative in the directions $P_1 - P_0$ and $P_2 - P_0$. The associated norm is given by

$$\|P - P_0\|^2 = -\mathbb{E}\{r(P_0, \Xi)[P - P_0, P - P_0]\}.$$

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 Proposition 3 and Proposition 4, one can confine the analysis to the local setting of Chen, Liao, and Sun (2014). That is, the convergence rate ε_n in Proposition

4 implies that $\widehat{P} \in \mathcal{B}_n$ with probability approaching one, where

$$\mathcal{B}_n \equiv \mathcal{B}_0 \cap \mathcal{P}_{K_n}, \qquad \text{where } \mathcal{B}_0 \equiv \{P \in \mathcal{P}_{K_n} : \|P - P_0\|_2 \leq \varepsilon_n \log\log n\}.$$

Let $\mathcal{V} \equiv clsp(\mathcal{B}_0) - \{P_0\}$ and $\mathcal{V}_n \equiv clsp(\mathcal{B}_n) - \{P_{0,n}\}$, where $clsp(\cdot)$ denotes the closed linear span and where $P_{0,n} = \pi_{K_n} P_0$ denotes the orthogonal projection of P_0 onto the sieve space $\mathcal{P}_{K_n}$.

$\mathcal{V}_n$ is a finite-dimensional Hilbert space, which implies that the functional $\Gamma(P)$ in (16) 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

$$\frac{\partial \Gamma(P_0)}{\partial P}[v] \equiv \left.\frac{\partial \Gamma(P_0 + \eta v)}{\partial \eta}\right|_{\eta=0} = \langle v_n^*, v \rangle$$

and

$$\frac{\partial \Gamma(P_0)}{\partial P}[v_n^*] = \|v_n^*\|^2 = \sup_{v \in \mathcal{V}_n, v \neq 0} \left| \frac{\partial \Gamma(P_0)}{\partial P}[v] \right|^2 / \|v\|^2 . \tag{47}$$

To get a step closer to familiar expressions from non-linear least squares asymptotic theory, we linearize the option pricing functions $P \in \mathcal{P}_{K_n}$ w.r.t. its coefficient vector β_n. Since any $v \in \mathcal{V}_n$ has the form $v = P - P_{0,n}$ for $P \in \mathcal{P}_{K_n}$, one has by mean value theorem $v = \frac{\partial P(\overline{\beta}, \mathbf{Z})}{\partial \beta}(\beta_n - \beta_{0,n})$ for $\overline{\beta}$ between β_n and the coefficients of the projection $P_{0,n}$. Thus $v_n^* = \frac{\partial P(\overline{\beta}^*, \mathbf{Z})}{\partial \beta}(\beta_n^* - \beta_{0,n})$ for some β_n^* that depends on the functional $\Gamma(P)$.

Now, let $\gamma_n = (\beta_n - \beta_{0,n})$, and define the directional derivative

$$G_{K_n} \equiv \frac{\partial \Gamma(P_0)}{\partial P} \left[\frac{\partial P(\overline{\beta}, \mathbf{Z})}{\partial \beta} \right], \quad \text{and} \quad R_{K_n} \equiv \mathbb{E} \left\{ -\frac{\partial^2 \ell(\overline{\beta}, \Xi)}{\partial \beta \partial \beta'} \right\}.$$

In this notation, the problem in (47) 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 is given by $\gamma_n^* = R_{K_n}^{-1} G_{K_n}$. Therefore,

$$v_n^* = \frac{\partial P(\overline{\beta}^*, \mathbf{Z})}{\partial \beta}(\beta_n^* - \beta_{0,n}) = \frac{\partial P(\overline{\beta}^*, \mathbf{Z})}{\partial \beta}\gamma_n^* = \frac{\partial P(\overline{\beta}^*, \mathbf{Z})}{\partial \beta} R_{K_n}^{-1} G_{K_n},$$

which by definition implies the norm $\|v_n^*\|^2 = G_{K_n}' R_{K_n}^{-1} G_{K_n}$. Finally, the score process (in the direction v_n^*)

$$\ell'(P_0, \Xi_i)[v_n^*] = [P_i - P_0(\mathbf{Z}_i)] W(\mathbf{Z}_i) v_n^* = e_i W(\mathbf{Z}_i) \frac{\partial P(\overline{\beta}^*, \mathbf{Z}_i)}{\partial \beta} \gamma_n^*$$

is required, with so-called standard deviation norm

$$\begin{aligned}
\|v_n^*\|_{sd}^2 &= Var\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \ell'(P_0, \Xi_i)[v_n^*] \right) \\
&= \gamma_n^{'*} \left(\frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\frac{\partial \ell(\overline{\beta}^*, \Xi_i)}{\partial \beta} \frac{\partial \ell(\overline{\beta}^*, \Xi_i)}{\partial \beta}' \right] \right) \gamma_n^* \\
&= G_{K_n}' R_{K_n}^{-1} \Sigma_{K_n} R_{K_n}^{-1} G_{K_n}. \qquad (48)
\end{aligned}$$

This object can be estimated by replacing the Riesz representer v_n^* with an estimate $\widehat{v}_n^*$. Define

$$\widehat{R}_{K_n} = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \ell(\widehat{\beta}_n, \Xi_i)}{\partial\beta\partial\beta'} \qquad \widehat{\Sigma}_{K_n} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \ell(\widehat{\beta}_n, \Xi_i)}{\partial\beta}\frac{\partial \ell(\widehat{\beta}_n, \Xi_i)}{\partial\beta}'$$

$$\widehat{G}_{K_n} = \int_{\mathcal{Z}_1}\omega(\mathbf{Z})\frac{\partial P(\widehat{\beta}_n, \mathbf{Z})}{\partial\beta}d\mathbf{Z}_1 + \int_{\mathcal{Z}_1^c}\omega(\mathbf{Z})\frac{\partial C(\widehat{\beta}_n, \mathbf{Z})}{\partial\beta}d\mathbf{Z}_1 \qquad \widehat{v}_n^* = \frac{\partial P(\overline{\beta}^*, \mathbf{Z})}{\partial\beta}\widehat{R}_{K_n}^{-1}\widehat{G}_{K_n}.$$

Then

$$\|\widehat{v}_n^*\|_{sd,n}^2 = \widehat{G}_{K_n}'\widehat{R}_{K_n}^{-1}\widehat{\Sigma}_{K_n}\widehat{R}_{K_n}^{-1}\widehat{G}_{K_n} \equiv \widehat{V}_n \tag{49}$$

corresponds to the usual variance estimator using the familiar parametric Delta method.

Infeasible Asymptotic Distribution We show here that

$$\sqrt{n}\frac{\Gamma(\widehat{P}_n) - \Gamma(P_0)}{\|v_n^*\|_{sd}} \xrightarrow{d} N(0,1). \tag{50}$$

Define the empirical process $\mu\{f(\Xi)\} = \frac{1}{n}\sum_{i=1}^n f(\Xi_i) - \mathbb{E}f(\Xi_i)$, and let $u_n^* = v_n^*/\|v_n^*\|_{sd}$. The result (50) follows by showing that $\sqrt{n}\frac{\Gamma(\widehat{P}_n)-\Gamma(P_0)}{\|v_n^*\|_{sd}} = \sqrt{n}\,\mu_n\{\ell'(P_0,\Xi)[u_n^*]\} + o_p(1)$, since $\sqrt{n}\,\mu_n\{\ell'(P_0,\Xi)[u_n^*]\} \xrightarrow{d} N(0,1)$ by Assumption B.1 (i) Assumption B.2, and Liapunov's CLT.

Break the LHS of (50) into two parts,

$$\sqrt{n}\frac{\Gamma(\widehat{P}_n) - \Gamma(P_0)}{\|v_n^*\|_{sd}} = \underbrace{\sqrt{n}\frac{\Gamma(\widehat{P}_n) - \Gamma(P_{0,n})}{\|v_n^*\|_{sd}}}_{\text{Part A}} + \underbrace{\sqrt{n}\frac{\Gamma(P_{0,n}) - \Gamma(P_0)}{\|v_n^*\|_{sd}}}_{\text{Part B}} \tag{51}$$

We start with Part B and prove it in steps.
Step 1: Show $\frac{\|v_n^*\|}{\|v_n^*\|_{sd}} = O(1)$.

$$\frac{\|v_n^*\|}{\|v_n^*\|_{sd}} = \frac{\gamma_n^{*'} R_{K_n}\gamma_n^*}{\gamma_n^{*'}\Sigma_{K_n}\gamma_n^*} \leq \frac{\lambda_{\max}(R_{K_n})}{\lambda_{\min}(\Sigma_{K_n})} = O(1), \tag{52}$$

where $\lambda_{\max}(S)$ and $\lambda_{\min}(S)$ are the largest and smallest eigenvalues of a matrix S. The last equality follows from Assumption B.5 (i) and Assumption B.2.
Step 2: Show $\|P - P_0\| \asymp \|P - P_0\|_2$. Note that $\|P - P_0\| = \sqrt{\mathbb{E}[-r(P_0,\Xi)[P-P_0, P-P_0]]}$ $= \sqrt{\mathbb{E}[\frac{1}{2}[P(\mathbf{Z}) - P_0(\mathbf{Z})]^2 W(\mathbf{Z})]} = \frac{1}{\sqrt{2}}\|P - P_0\|_2$.
Step 3: $\|v_n^*\| \nearrow \|v^*\| < \infty$. By definition, since $\mathcal{V}_n \subset \mathcal{V}$ and $\|v_n^*\|^2 = \langle v_n^*, v_n^*\rangle$ $= \sup_{v\in\mathcal{V}_n, v\neq 0}|\Gamma(P_0)[v]|^2/\|v\|^2 \leq \sup_{v\in\mathcal{V}, v\neq 0}|\Gamma(P_0)[v]|^2/\|v\|^2 = \|v^*\|^2 < \infty$, where the last inequality follows from from the bound in Assumption B.5 (ii) (see the discussion in Chen, Liao, and Sun (2014, Remark 3.2)). The result follows from continuity (linearity in v) of $\Gamma(P_0)[v]$ and deneseness of $\mathcal{V}_n$ in $\mathcal{V}$ (Lemma A.10).
Step 4: Show $\|v^* - v_n^*\| = O(K_n^{-\alpha})$. By definition, $\|v^* - v_n^*\| = \|(P^* - P_0) - (P_n^* - P_{0,n})\|$ $= \|(P^* - P_n^*) - (P_0 - P_{0,n})\| \leq \|P^* - P_n^*\| + \|P_0 - P_{0,n}\| = O(K_n^{-\alpha})$ by Assumption B.3 and Step 2.

Step 5: By definition, since $\langle v_n^*, P_{0,n} - P_0\rangle = 0$,

$$\frac{|\Gamma(P_{0,n}) - \Gamma(P_0)|}{\|v_n^*\|_{sd}} = \frac{|\Gamma(P_{0,n}) - \Gamma(P_0)|}{\|v_n^*\|} \underbrace{\frac{\|v_n^*\|}{\|v_n^*\|_{sd}}}_{O(1) \text{ by Step 1}} = O(1)\frac{|\Gamma'(P_0)[P_{0,n} - P_0]|}{\|v_n^*\|}$$
$$= O(1)\frac{|\langle v^*, P_{0,n} - P_0\rangle|}{\|v_n^*\|} = O(1)\frac{|\langle v^* - v_n^*, P_{0,n} - P_0\rangle|}{\|v_n^*\|}$$
$$\leq O(1)\,\|v^* - v_n^*\|\,\|P_{0,n} - P_0\|\,\frac{\|v^*\|}{\|v_n^*\|} = O(1)\,\|v^* - v_n^*\|\,\|P_{0,n} - P_0\|$$
$$\leq O(1)O(K_n^{-\alpha})O(K_n^{-\alpha}) = O(n^{\frac{-2\alpha d_u}{2(m_0+m)+d_u}}) = o(n^{-1/2})$$

by definition of the Riesz representer, Cauchy-Schwarz inequality, Step 3, Step 4, and Assumption B.5 (iii). Conclude that Part B in (51) is $o(1)$.

To show Part A, let $u_n^* = v_n^* / \|v_n^*\|_{sd}$.
Step 1: By linearity of $\Gamma'(\cdot)[v]$ in v,

$$\sqrt{n}\frac{\Gamma(\widehat{P}_n) - \Gamma(P_{0,n})}{\|v_n^*\|_{sd}} = \sqrt{n}\frac{\Gamma(P_0) + \Gamma'(P_0)[\widehat{P}_n - P_0] - \{\Gamma(P_0) + \Gamma'(P_0)[P_{0,n} - P_0]\}}{\|v_n^*\|_{sd}}$$
$$= \sqrt{n}\frac{\Gamma'(P_0)[\widehat{P}_n - P_{0,n}]}{\|v_n^*\|_{sd}} = \frac{\langle \widehat{P}_n - P_{0,n}, v_n^*\rangle}{\|v_n^*\|_{sd}} = \langle \widehat{P}_n - P_{0,n}, u_n^*\rangle.$$

Step 2: For $\epsilon_n = o(n^{-1/2})$, show that

$$\sup_{\|P_1 - P_2\| \leq \epsilon_n} \mu_n\{\ell(P_1, \Xi) - \ell(P_2, \Xi) - \ell'(P_0, \Xi)[P_1 - P_2]\} = o_p(\epsilon_n). \tag{53}$$

Let $Q_n(P) \equiv \mu_n\{\ell(P, \Xi_i) - \ell(P_0, \Xi_i) - \ell'(P_0, \Xi_i)[P - P_0]\}$. In this notation, the LHS in (53) becomes $\mu_n\{\ell(P_1, \Xi) - \ell(P_2, \Xi) - \ell'(P_0, \Xi)[P_1 - P_2]\} = Q_n(P_1) - Q_n(P_2)$. Then by the functional mean value theorem,

$$|Q_n(P_1) - Q_n(P_2)| \leq \sup_{P \in \mathcal{P}} |Q_n'(P)|\,\|P_1 - P_2\| = O_p(1)\,\|P_1 - P_2\|\,,$$

since by definition of the least squares objective function,

$$Q_n'(P) = \mu_n\{\ell'(P, \Xi_i) - \ell'(P_0, \Xi_i) - \ell''(P_0, \Xi_i)[P - P_0]\} = -\mu_n\{P_0(\mathbf{Z})W(\mathbf{Z})\} = O_p(1).$$

The result in (53) follows.
Step 3: Consider how the optimized sample objective function behaves in response to small changes in the direction of the Riesz representer v_n^*. To this end, we follow Chen, Liao, and Sun (2014) and set $\widehat{P}_{u,n}^* = \widehat{P}_n \pm \epsilon_n u_n^*$, where $\epsilon_n = o(n^{-1/2})$. Note that since $\widehat{P}_n \in \mathcal{B}_n$ with probability approaching one, one has that $\widehat{P}_{u,n}^* \in \mathcal{B}_n$ with probability approaching one. Then by definition of $\widehat{P}_n$,

$$-O_p(\epsilon_n^2) \leq \frac{1}{n}\sum_{i=1}^{T} \ell(\widehat{P}_n, \Xi_i) - \frac{1}{n}\sum_{i=1}^{T} \ell(\widehat{P}_{u,n}^*, \Xi_i)$$
$$= \mathbb{E}[\ell(\widehat{P}_n, \Xi_i) - \ell(\widehat{P}_{u,n}^*, \Xi_i)] + \mu_n\{\ell'(P_0, \Xi_i)[\widehat{P}_n - \widehat{P}_{u,n}^*]\}$$
$$+ \mu_n\{\ell(\widehat{P}_n, \Xi_i) - \ell(\widehat{P}_{u,n}^*, \Xi_i) - \ell'(P_0, \Xi_i)[\widehat{P}_n - \widehat{P}_{u,n}^*]\}$$
$$= \mathbb{E}[\ell(\widehat{P}_n, \Xi_i) - \ell(\widehat{P}_{u,n}^*, \Xi_i)] \mp \mu_n\{\ell'(P_0, \Xi_i)[\epsilon_n u_n^*]\} + o_p(\epsilon_n)$$

by Step 2. Next, note that by definition of $r(P_0, \Xi)$,

$$\ell(P, \Xi) = \ell(P_0, \Xi) + \ell'(P_0, \Xi)[P - P_0] + \frac{1}{2} r(P_0, \Xi)[P - P_0, P - P_0],$$

so that

$$\begin{aligned}
\mathbb{E}[\ell(\widehat{P}_n, \Xi_i) - \ell(\widehat{P}^*_{u,n}, \Xi_i)] = \frac{\left\|\widehat{P}^*_{u,n} - P_0\right\|^2 - \left\|\widehat{P}_n - P_0\right\|^2}{2} &= \pm\epsilon_n \langle \widehat{P}_n - P_0, u^*_n \rangle + \frac{1}{2}\epsilon_n^2 \left\|u^*_n\right\|^2 \\
&= \pm\epsilon_n \langle \widehat{P}_n - P_0, u^*_n \rangle + O_p(\epsilon_n^2).
\end{aligned}$$

Thus $-O_p(\epsilon_n^2) \leq \pm\epsilon_n \langle \widehat{P}_n - P_0, u^*_n \rangle + O_p(\epsilon_n^2) \mp \epsilon_n \mu_n \{\ell'(P_0, \Xi_i)[u^*_n]\} + o_p(\epsilon_n)$, so that

$$\left| \langle \widehat{P}_n - P_0, u^*_n \rangle - \mu_n \{\ell'(P_0, \Xi_i)[u^*_n]\} \right| = O_p(\epsilon_n) = o_p(n^{-1/2}).$$

Finally, since the definition of $P_{0,n}$ implies $\langle P_{0,n} - P_0, v \rangle = 0$ for any $v \in \mathcal{V}_n$,

$$\left| \langle \widehat{P}_n - P_{0,n}, u^*_n \rangle - \mu_n \{\ell'(P_0, \Xi_i)[u^*_n]\} \right| = o_p(n^{-1/2}).$$

This expression, plugged into Step 1 and (51), yields

$$\begin{aligned}
\sqrt{n}\frac{\Gamma(\widehat{P}_n) - \Gamma(P_0)}{\|v^*_n\|_{sd}} &= \sqrt{n}\frac{\Gamma(\widehat{P}_n) - \Gamma(P_{0,n})}{\|v^*_n\|_{sd}} + \sqrt{n}\frac{\Gamma(P_{0,n}) - \Gamma(P_0)}{\|v^*_n\|_{sd}} \\
&= \sqrt{n}\,\mu_n \left\{\ell'(P_0, \Xi)[u^*_n]\right\} + o_p(1) \\
&\xrightarrow{d} N(0, 1).
\end{aligned}$$

by Assumption B.1 (i) Assumption B.2, and Liapunov's CLT.

Feasible Asymptotic Distribution We show that replacing $\|v^*_n\|_{sd}$ with the estimate $\|\widehat{v}^*_n\|_{sd,n}$ in (49) results in

$$\sqrt{n}\frac{\Gamma(\widehat{P}_n) - \Gamma(P_0)}{\|\widehat{v}^*_n\|_{sd,n}} \xrightarrow{d} N(0, 1).$$

To establish the requisite stochastic equicontinuity results, we use the following lemma:

Lemma B.4. *For $\delta > 0$, the subset of option pricing functions $\mathcal{G}(\delta) \equiv \{P_1, P_2 \in \mathcal{P} : \|P_1 - P_2\|_2 \leq \delta\}$ is $\mathbb{P}$-Donsker.*

Proof. By Lemma B.2,

$$\begin{aligned}
\|P_1 - P_2\|_2 &\leq M_1 \|f_1 - f_2\|_{m,1} \\
&\leq M_2 \left\|f_1^{Y,X} - f_2^{Y,X}\right\|_{m,\infty,\zeta_0} && \text{by Lemma A.9} \\
&= M_2 \left\|(h_1^{Y,X})^2 - (h_2^{Y,X})^2\right\|_{m,\infty,\zeta_0} && \text{by Def. A.2} \\
&\leq M_3 \left\|h_1^{Y,X} - h_2^{Y,X}\right\|_{m_0+m,2,\zeta_0} && \text{by inequality (46)} \\
&\leq 2M_3 \mathcal{B}_0.
\end{aligned} \tag{54}$$

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.

$$N_{[\,]}(w, \mathcal{G}(\delta), \|\cdot\|_2) \leq N\left(\frac{w}{4\mathcal{B}_0 M_3}, \mathcal{H}, \|\cdot\|_{m_0+m,2,\zeta_0}\right) \leq N\left(\frac{w}{4\mathcal{B}_0 M_3}, \mathcal{H}, \|\cdot\|_{\infty}\right),$$

where the second inequality follows from Gallant and Nychka (1987) Lemma A.1(c). Therefore,

$$H_{[\,]}(w, \mathcal{G}(\delta), \|\cdot\|_2) \leq C_2 w^{-d_u/m}$$

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)). □

Next, note that for $\mathcal{W}_n \equiv \{v \in \mathcal{V}_n : \|v\| = 1\}$, Chen, Liao, and Sun (2014) (CLS) Assumption 5.1(i) is trivially satisfied for the least squares regression function, and CLS Assumptions 5.1(ii) is satisfied by Lemma B.4 and an application of the Glivenko-Cantelli theorem. CLS Assumption 5.1(iii) can be obtained from Assumption B.5 (iv), so that CLS Lemma 5.1 of can be invoked, which states

$$\left|\frac{\|\widehat{v}_n^*\|}{\|v_n^*\|} - 1\right| = O_p(\epsilon_n^*), \qquad \frac{\|\widehat{v}_n^* - v_n^*\|}{\|v_n^*\|} = O_p(\epsilon_n^*), \tag{55}$$

for $\epsilon_n^* = o(1)$.

The object of interest is

$$\|v_n^*\|_{sd}^{-1} \|\widehat{v}_n^*\|_{sd,n}^2 \|v_n^*\|_{sd}^{-1} = \widehat{Var}\left(\|v_n^*\|_{sd}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \ell'(\widehat{P}_n, \Xi_i)[\widehat{v}_n^*]\right).$$

Focusing on the term inside the variance, linearity of the directional derivative implies

$$\begin{aligned}\frac{1}{\sqrt{n}} \sum_{i=1}^n \ell'(\widehat{P}_n, \Xi_i)[\widehat{v}_n^*] &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{\ell'(\widehat{P}_n, \Xi_i)[\widehat{v}_n^*] - \ell'(P_0, \Xi_i)[\widehat{v}_n^*]\right\} \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n \ell'(P_0, \Xi_i)[v_n^*] + \frac{1}{\sqrt{n}} \sum_{i=1}^n \ell'(P_0, \Xi_i)[\widehat{v}_n^* - v_n^*].\end{aligned}$$

The third term on the RHS is

$$\left|\frac{1}{\sqrt{n}} \sum_{i=1}^n \ell'(P_0, \Xi_i)[\widehat{v}_n^* - v_n^*]\right| \leq \|\widehat{v}_n^* - v_n^*\| \sup_{v \in \mathcal{W}_n} \left|\frac{1}{\sqrt{n}} \sum_{i=1}^n \ell'(P_0, \Xi_i)[v]\right| = O_p(\|v_n^*\| \, \epsilon_n^*)$$

by the Donsker property of Lemma B.4, the functional CLT, and (55).

To address the first term on the RHS, consider

$$\sup_{P\in\mathcal{B}_n}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\{\ell'(P,\Xi_i)[\widehat{v}_n^*]-\ell'(P_0,\Xi_i)[\widehat{v}_n^*]\}\right| \le \|\widehat{v}_n^*\| \sup_{\substack{P\in\mathcal{B}_n\\ v\in\mathcal{W}_n}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n v(\mathbf{Z}_i)W(\mathbf{Z}_i)[P_0(\mathbf{Z}_i)-P(\mathbf{Z}_i)]\right|$$
$$= O_p(\|v_n^*\|\,\epsilon_n^*),$$

by the Donsker property of Lemma B.4, the functional CLT, and (55).

Combining results from the previous three displays and using (52), one has

$$\begin{aligned}
\|v_n^*\|_{sd}^{-1}\,\|\widehat{v}_n^*\|_{sd,n}^2\,\|v_n^*\|_{sd}^{-1} &= \widehat{Var}\left(\|v_n^*\|_{sd}^{-1}\,\frac{1}{\sqrt{n}}\sum_{i=1}^n\ell'(\widehat{P}_n,\Xi_i)[\widehat{v}_n^*]\right)\\
&= \widehat{Var}\left(\|v_n^*\|_{sd}^{-1}\,\frac{1}{\sqrt{n}}\sum_{i=1}^n\ell'(P_0,\Xi_i)[v_n^*]+o_p(1)\right)\\
&\xrightarrow{p} \|v_n^*\|_{sd}^{-1}\,\|v_n^*\|_{sd}^2\,\|v_n^*\|_{sd}^{-1}=1
\end{aligned}$$

by LLN. Therefore

$$\frac{\sqrt{n}[\Gamma(\widehat{P}_n)-\Gamma(P_0)]}{\|\widehat{v}_n^*\|_{sd,n}} = \sqrt{n}\frac{\Gamma(\widehat{P}_n)-\Gamma(P_0)}{\|v_n^*\|_{sd}}\frac{\|v_n^*\|_{sd}}{\|\widehat{v}_n^*\|_{sd,n}} \quad \xrightarrow{d} \quad N(0,1),$$

as required. $\square$

Online Appendix - Not Intended for Publication

A Implementation Details

A.1 Implementation Summary

1. Choose $\mu(\mathbf{Z})$ and $\sigma(\mathbf{Z})$. To center expanions around Black-Scholes, use (20).
2. Construct $P(\beta, \mathbf{Z})$ from Proposition 1.
3. Choose $K_n = (K_y(n) + 1)(K_\tau(n) + 1)$ to grow slowly as $n \to \infty$. See Section 3.
4. Optimize the objective function over sieve coefficients in (14), using all options from a given cross-section of options.
5. Form $\widehat{V}_n$ using (18) and use critical values from standard normal tables.

A.2 Gradients and Hessian

The objective function to be minimized is

$$L_n(\beta) \equiv \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\Big[p_i - P(\beta, \mathbf{z}_i)\Big]^2 W_i \equiv \frac{1}{n}\sum_{i=1}^{n}\ell(\beta, \Xi_i). \tag{56}$$

The gradient and Hessian of the objective function are

$$\frac{\partial L_n(\widehat{\beta})}{\partial \beta} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \ell(\widehat{\beta}, \Xi_i)}{\partial \beta} = \frac{1}{n}\sum_{i=1}^{n} -\Big[p_i - P(\widehat{\beta}, \mathbf{z}_i)\Big] W_i \frac{\partial P(\widehat{\beta}, \mathbf{z}_i)}{\partial \beta} \tag{57}$$

$$\begin{aligned} \frac{\partial^2 L_n(\widehat{\beta})}{\partial \beta \partial \beta'} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \ell(\widehat{\beta}, \Xi_i)}{\partial \beta \partial \beta'} \\ &= \frac{1}{n}\sum_{i=1}^{n}\left\{ -\Big[p_i - P(\widehat{\beta}, \mathbf{z}_i)\Big] W_i \frac{\partial^2 P(\widehat{\beta}, \mathbf{z}_i)}{\partial \beta \partial \beta'} + W_i \frac{\partial P(\widehat{\beta}, \mathbf{z}_i)}{\partial \beta}\frac{\partial P(\widehat{\beta}, \mathbf{z}_i)}{\partial \beta}' \right\}. \end{aligned} \tag{58}$$

The pricing function $P(\beta, \mathbf{z}_i)$ is found in Proposition 1,

$$\begin{aligned} P(\beta, \mathbf{z}_i) = \kappa e^{-r_i \tau_i} &\left[\Phi(d(\mathbf{z}_i)) - \sum_{k=1}^{2K_x(n)} \frac{\gamma_k(\beta, \tau_i)}{\sqrt{k}} H_{k-1}(d(\mathbf{z}_i))\phi(d(\mathbf{z}_i)) \right] \\ &- S_i e^{-r_i \tau_i + \mu(\mathbf{z}_i)} \left[e^{\sigma(\mathbf{z}_i)^2/2}\Phi(d(\mathbf{z}_i) - \sigma(\mathbf{z}_i)) + \sum_{k=1}^{2K_x(n)} \gamma_k(\beta, \tau_i) I_k^*(d(\mathbf{z}_i)) \right] \end{aligned} \tag{59}$$

where $\Phi(\cdot)$ is the standard normal CDF, $K_n \equiv (K_x(n), K_\tau(n))$, and where

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

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

$$\alpha(B,\tau) = \begin{bmatrix} \sum_{j=0}^{Kz} \beta_{0j} H_j(\tau) \\ \vdots \\ \sum_{j=0}^{Kz} \beta_{K_x j} H_j(\tau) \end{bmatrix} = B \cdot H$$

for

$$B \equiv \begin{bmatrix} \beta_{00} & \beta_{01} & \dots & \beta_{0K_\tau} \\ \beta_{10} & \beta_{11} & \dots & \beta_{1K_\tau} \\ \vdots & & & \\ \beta_{K_x 0} & \beta_{K_x 1} & \dots & \beta_{K_x K_\tau} \end{bmatrix}, \quad H \equiv \begin{bmatrix} H_0(\tau) \\ \vdots \\ H_{K_\tau}(\tau) \end{bmatrix}.$$

The only place where β shows up in $P(\beta, \mathbf{z})$ is through each of the $\gamma_k(\beta,\tau)$, $k = 1, \dots, K_x$. Hence to find first and second derivatives of $P(\cdot, \mathbf{z})$ we must find them for $\gamma_k(\cdot, \tau)$.

Note that

$$\gamma_k(B,\tau) = [H'B'B'H]^{-1}H'B'A_k BH.$$

We suppress the k subscript on A in subsequent derivations. Using the matrix differential conventions in Fackler (2005), it can be shown that

$$\begin{aligned} \frac{\partial \gamma_k(B,\tau)}{\partial B} &\equiv \frac{\partial vec\{\gamma_k(B,\tau)\}}{\partial vec\{B\}} = \frac{\partial \gamma_k(\beta,\tau)}{\partial \beta} \\ &= [H'B'BH]^{-1}\Big\{(H' \otimes H'B'A) + (H'B'A' \otimes H')T_{(K_x+1),(K_\tau+1)}\Big\} \\ &- [H'B'BH]^{-2}[H'B'ABH]\Big\{(H' \otimes H'B') + (H'B' \otimes H')T_{(K_x+1),(K_\tau+1)}\Big\}, \end{aligned} \tag{60}$$

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,

$$\begin{aligned} \frac{\partial P(\widehat{\beta}, \mathbf{z}_i)}{\partial \beta} &= \kappa e^{-r_i\tau_i}\left[\Phi(d(\mathbf{z}_i)) - \sum_{k=1}^{2K_x(n)} \frac{1}{\sqrt{k}} \frac{\partial \gamma_k(\beta,\tau)}{\partial \beta} H_{k-1}(d(\mathbf{z}_i))\phi(d(\mathbf{z}_i))\right] \\ &- S_i e^{-r_i\tau_i + \mu(\mathbf{z}_i)}\left[e^{\sigma(\mathbf{z}_i)^2/2}\Phi(d(\mathbf{z}_i) - \sigma(\mathbf{z}_i)) + \sum_{k=1}^{2K_x(n)} \frac{\partial \gamma_k(\beta,\tau)}{\partial \beta} I_k^*(d(\mathbf{z}_i))\right]. \end{aligned}$$

To obtain the Hessian, decompose the expression in Eq. (60) into its four terms,

$$
\begin{aligned}
f_1(B) &= [H'B'BH]^{-1}(H' \otimes H'B'A) \\
f_2(B) &= [H'B'BH]^{-1}(H'B'A' \otimes H')T_{(K_x+1),(K_\tau+1)} \\
f_3(B) &= [H'B'BH]^{-2}[H'B'ABH](H' \otimes H'B') \\
f_4(B) &= [H'B'BH]^{-2}[H'B'ABH](H'B' \otimes H')T_{(K_x+1),(K_\tau+1)}
\end{aligned}
$$

To avoid clutter, let $u_1(B) = [H'B'BH]^{-1}$, $v_1(B) = (H' \otimes H'B'A)$. Then

$$
\begin{aligned}
u_1'(B) &= -[H'H'BH]^{-2}\Big\{(H' \otimes H'B') + (H'B' \otimes H')T_{(K_x+1),(K_\tau+1)}\Big\} \\
v_1'(B) &= H' \otimes (H \otimes A')
\end{aligned}
$$

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 we make use of the product rule for $i = 3, 4$:

$$
\begin{aligned}
f_i'(B) = Du(B)_{m\times p}v(B)_{p\times q}w(B)_{q\times n} &= [w(B)'v(B)' \otimes I_m]u'(B) \\
&+ (I_n \otimes u(B))\Big\{(w(B)' \otimes I_p)v'(B) + (I_n \otimes v(B)w'(B))\Big\}.
\end{aligned}
$$

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

$$
\begin{aligned}
u'(B) &= -2[H'B'BH]^{-3}\Big\{(H' \otimes H'B') + (H'B' \otimes H')T_{(K_x+1),(K_\tau+1)}\Big\} \\
v'(B) &= (H' \otimes H'B'A) + (H'B'A' \otimes H')T_{(K_x+1),(K_\tau+1)} \\
w'(B) &= H' \otimes (H \otimes I_{K_x+1}).
\end{aligned}
$$

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

$$
\begin{aligned}
w(B) &= (H'B' \otimes H')T_{(K_x+1),(K_\tau+1)} \\
w'(B) &= H' \otimes T'_{(K_x+1),(K_\tau+1)}(I_{K_x+1} \otimes H).
\end{aligned}
$$

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,

$$\begin{aligned}\frac{\partial^2 P(\widehat{\beta},\mathbf{z}_i)}{\partial\beta\partial\beta'} &= \kappa e^{-r_i\tau_i}\left[\Phi(d(\mathbf{z}_i)) - \sum_{k=1}^{2K_x(n)} \frac{1}{\sqrt{k}}\frac{\partial^2\gamma_k(\beta,\tau_i)}{\partial\beta\partial\beta'} H_{k-1}(d(\mathbf{z}_i))\phi(d(\mathbf{z}_i))\right] \\ &- S_i e^{-r_i\tau_i+\mu(\mathbf{z}_i)}\left[e^{\sigma(\mathbf{z}_i)^2/2}\Phi(d(\mathbf{z}_i)-\sigma(\mathbf{z}_i)) + \sum_{k=1}^{2K_x(n)} \frac{\partial^2\gamma_k(\beta,\tau_i)}{\partial\beta\partial\beta'} I_k^*(d(\mathbf{z}_i))\right].\end{aligned}$$

B Sieve Hedging

Recalling the pricing formula (21) and the definitions in Proposition 1 in the main text, we have

$$P(f_K,\mathbf{Z}) = P_{BS}(\sigma,\mathbf{Z}) + A(\sigma,\mathbf{Z}),$$

where

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

Thus sieve option Greeks are standard Black-Scholes Greeks plus higher order adjustment terms, obtained by differentiating the adjustment term $A(\sigma,\mathbf{Z})$. The process is illustrated for the option delta, yielding

$$\frac{\partial P(f_K,\mathbf{Z})}{\partial S_0} = \Delta_{BS} + \frac{\partial A(\sigma,\mathbf{Z})}{\partial S_0},$$

where $\Delta_{BS} = -e^{-q\tau}\Phi(d(\mathbf{Z})-\sigma\sqrt{\tau})$ is the standard Black-Scholes delta and

$$\frac{\partial A(\sigma,\mathbf{Z})}{\partial S_0} = -\sum_{k=1}^{2K_y}\gamma_k(B,\tau)\left[\underbrace{\frac{\partial}{\partial S_0}\left\{\frac{\kappa e^{-r\tau}}{\sqrt{k}}H_{k-1}(d(\mathbf{Z}))\phi(d(\mathbf{Z}))\right\}}_{(*)} + \underbrace{\frac{\partial}{\partial S_0}\left\{S_0 e^{-q\tau-\sigma^2\tau/2}I_k^*(d(\mathbf{Z}))\right\}}_{(**)}\right].$$

Examining each component on the right-hand side and adopting the convention $H_{-1}(x) = 0$, we have

$$\begin{aligned}(*) = \frac{\partial}{\partial S_0}\left\{\frac{\kappa e^{-r\tau}}{\sqrt{k}}H_{k-1}(d(\mathbf{Z}))\phi(d(\mathbf{Z}))\right\} &= \kappa e^{-r\tau}H_{k-2}(d(\mathbf{Z}))\phi(d(\mathbf{Z}))\frac{\partial d(\mathbf{Z})}{\partial S_0} \\ &- \frac{\kappa e^{-r\tau}}{\sqrt{k}}H_{k-1}(d(\mathbf{Z}))d(\mathbf{Z})\phi(d(\mathbf{Z}))\frac{\partial d(\mathbf{Z})}{\partial S_0},\end{aligned}$$

where $\partial d(\mathbf{Z})/\partial S_0 = -1/[S_0\sigma\sqrt{\tau}]$, and

$$(**) = \frac{\partial}{\partial S_0}\left\{S_0 e^{-q\tau-\sigma^2\tau/2}I_k^*(d(\mathbf{Z}))\right\} = e^{-q\tau-\sigma^2\tau/2}I_k^*(d(\mathbf{Z})) + S_0 e^{-q\tau-\sigma^2\tau/2}\frac{\partial I_k^*(d(\mathbf{Z}))}{\partial S_0},$$

where

$$
\begin{aligned}
\frac{\partial I_k^*(d(\mathbf{Z}))}{\partial S_0} = \frac{\sigma\sqrt{\tau}}{\sqrt{k}}\frac{\partial I_{k-1}^*(d(\mathbf{Z}))}{\partial S_0} - \Bigg[&\frac{\sigma\sqrt{\tau}}{\sqrt{k}}e^{\sigma\sqrt{\tau}d(\mathbf{Z})}\frac{\partial d(\mathbf{Z})}{\partial S_0}H_{k-1}(d(\mathbf{Z}))\phi(d(\mathbf{Z})) \\
&+\frac{1}{\sqrt{k}}e^{\sigma\sqrt{\tau}d(\mathbf{Z})}\frac{1}{\sqrt{k-1}}H_{k-2}(d(\mathbf{Z}))\frac{\partial d(\mathbf{Z})}{\partial S_0}\phi(d(\mathbf{Z})) \\
&-\frac{1}{\sqrt{k}}e^{\sigma\sqrt{\tau}d(\mathbf{Z})}H_{k-1}(d(\mathbf{Z}))d(\mathbf{Z})\phi(d(\mathbf{Z}))\frac{\partial d(\mathbf{Z})}{\partial S_0}\Bigg]
\end{aligned}
$$

and $\frac{\partial I_0^*(d(\mathbf{Z}))}{\partial S_0} = e^{\sigma^2\tau/2}\phi(d(\mathbf{Z}) - \sigma\sqrt{\tau})\frac{\partial d(\mathbf{Z})}{\partial S_0}$.

C Further Simulations: RMSE Comparisons

This section extends the simulation results of Section 3 in the main paper by exploring the sieve's ability to fit option prices across maturities and under a variety of DGPs.

C.1 Goodness of Fit: VIX Term Structures

Table 7 shows a root-mean squared error (RMSE) comparison of the sieve and the benchmarks relative to the true VIX term structure, where the units are in VIX index points.[18] The sieve outperforms the benchmarks in most instances, with significant improvements occurring at the 12-month maturity, where strike truncation is the most severe: options maturing in 12-months only span a moneyness range (κ/S_0) of 0.86 to 1.11. To make this range comparable to other maturities, Table 7 also converts moneyness values to quantiles of the implied risk-neutral CDF of the underlying SVJJ process, which shows that the 12-month truncation cuts off option price information below the 23% quantile and above the 68% quantile. Compared with the other maturities, this means that the 12-month maturity is significantly less informative about the tails of the implied risk-neutral distribution.

In contrast to the sieve, the benchmarks in Table 7 re-estimate at each maturity. Thus, the CBOE's discrete approximation to the integral (1), which makes no tail predictions and uses no information from neighboring maturities, performs poorly at the 12-month horizon but works well at the liquid 1-month horizon, where strikes cover almost the entire risk-neutral distribution.[19] The lognormal extrapolation (which is equivalent to setting extrapolated option prices to have the same implied volatility as the most extreme observed options) used in Carr and Wu (2009) and Jiang and Tian (2005) does much better at the 12-month horizon, but then deteriorates at the 1-month horizon. This is essentially due to the fact that their procedure implicitly assumes that option data are observed without error and that prices can therefore be interpolated. Finally, Table 7 shows that Black-Scholes implied volatility measures suffer most, which suggests that tail information (not captured by at-the-money Black-Scholes implied volatility and incorrectly weighted by averaging Black-Scholes volatilities) is critically important.

The sieve's strong performance at the truncated 12-month maturity comes from two structural improvements relative to the benchmarks: first, by using the shape information from the risk-neutral valuation equation, it performs a theoretically supported projection of option prices beyond the 23% and 68% implied

[18] These benchmarks refer to workhorses frequently used in applied work. See Dew-Becker, Giglio, Le, and Rodriguez (2015), Bollerslev, Gibson, and Zhou (2011), Carr and Wu (2009), and Jiang and Tian (2005) for applications using these benchmarks.

[19] See CBOE (2003) for details about its VIX construction.

Table 7: RMSE Comparison of VIX Term Structures.

VIX estimator	Maturity (months) 1	2	3	6	9	12	15	21
Sieve	1.10	0.44	0.45	0.50	0.41	0.43	0.45	0.48
CBOE	1.16	0.73	0.58	0.52	0.89	5.73	1.28	0.81
Lognormal Extrapolation	4.50	0.53	1.04	0.49	0.53	1.46	0.56	0.45
ATM Black-Scholes	3.96	3.38	3.65	3.41	3.58	3.01	3.26	2.84
Average Black-Scholes	16.25	1.97	7.61	1.37	0.58	1.69	0.72	0.68
Moneyness Range	0.65	0.75	0.49	0.55	0.62	0.86	0.63	0.44
	1.12	1.21	1.30	1.38	1.37	1.11	1.30	1.32
RN CDF Quantile Range	0.01	0.01	0.01	0.01	0.03	0.23	0.07	0.02
	1.00	1.00	1.00	1.00	0.99	0.68	0.91	0.88
Number of Obs.	32	20	44	31	30	9	23	27

Notes: A dense surface of true option prices was simulated under an SVJJ specification for each of the maturities shown, from which a true VIX was computed without moneyness truncation error. Then, 1000 random subsamples were drawn under the various shown moneyness truncation ranges. These sample prices were perturbed with a uniformly distributed error corresponding to the width of observed bid-ask spreads on S&P 500 index options. VIX estimates using each of the displayed methods were computed to arrive at their RMSE relative to the true VIX and are in percent annualized standard deviation units (e.g. RMSE of 0.50 means half a VIX point). RN CDF Quantile refers to the quantiles of the implied risk-neutral CDF corresponding to the most extreme observed strikes at the given maturity.

quantiles. Second, and importantly, the sieve regression (9) is using information from all maturities in a single pass. Combined with the shape restrictions, this implies that information across all maturities is informing the sieve projection at the 12-month maturity.

The situation is illustrated in Figure 6, where simulated true SVJJ prices are plotted alongside sieve estimates. The left panel shows that the sieve's 12-month out-of-sample price projections (extrapolations) clearly benefit from information at other maturities. The right panel shows the severity of the 12-month truncation in terms of implied risk-neutral quantiles.[20] It also provides practical guidance as to when the extrapolation can stop for the purpose of computing the option spanning integral (1): one should ideally continue until sufficient tail information is incorporated. For the implementations in this paper, we set the strike integration range for spanning portfolios to consistently cover 0.5% to 99.5% of the risk-neutral distribution.

[20]It can be shown from Lemma B.1 that the sieve-implied risk-neutral CDF is given in closed-form by

$$\mathbb{Q}_K(S_T \leq \kappa|\tau) \equiv \int_{-\infty}^{d(\mathbf{Z})} f_K(y|\tau)dy = \Phi(d(\mathbf{Z})) - \sum_{k=1}^{2K_y} \frac{\gamma_k(B,\tau)}{\sqrt{k}} H_{k-1}(d(\mathbf{Z}))\phi(d(\mathbf{Z})).$$

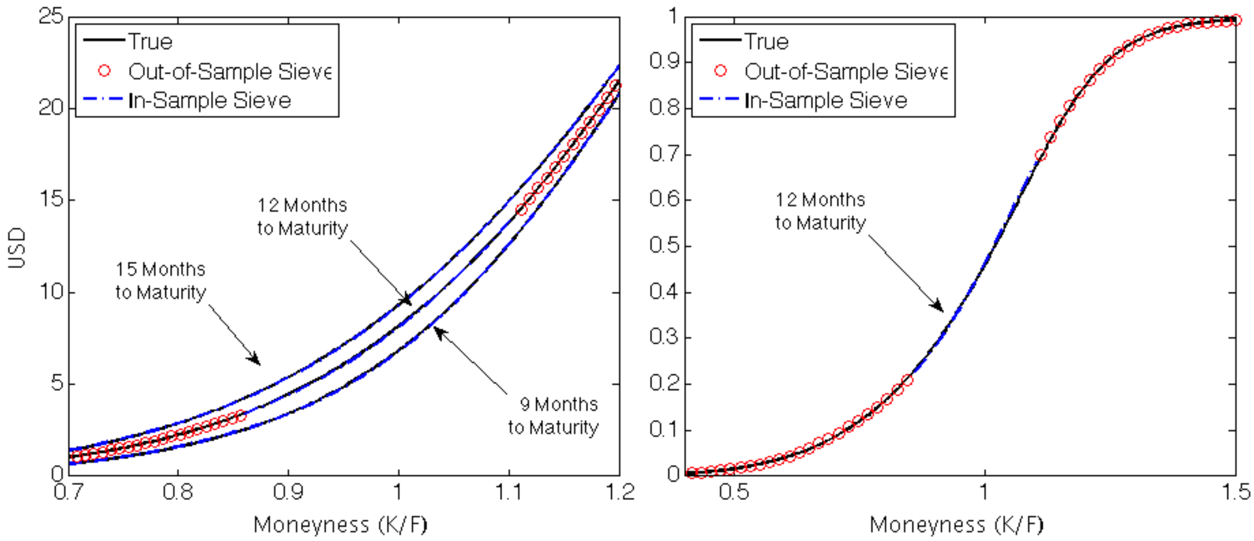

Figure 6: Simulated Out-of-Sample Sieve Price Projections and Implied Risk-Neutral CDF. An option surface with 1, 2, 3, 6, 9, 12, 15, 21 months-to-maturity and with respective number of observations 32, 20, 44, 31, 30, 9, 23, 27 is simulated. The sieve regression (14) is estimated, with expansion terms $(K_y, K_\tau) = (8, 2)$ selected via the procedure in Remark F.1. Strikes at the 12-month maturity were truncated according to the cutoffs in Table 7. The left panel shows true and sieve projected prices at the 9-month, 12-month, and 15-month maturities, and the right panel shows the risk-neutral CDF implied by the projected 12-month prices. Red circles denote extrapolations beyond the observed strike range.