

Vogt, Erik

**Working Paper**

## Option-implied term structures

Staff Report, No. 706

**Provided in Cooperation with:**

Federal Reserve Bank of New York

*Suggested Citation:* Vogt, Erik (2014) : Option-implied term structures, Staff Report, No. 706, Federal Reserve Bank of New York, New York, NY

This Version is available at:

<https://hdl.handle.net/10419/120788>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

Federal Reserve Bank of New York  
Staff Reports

# Option-Implied Term Structures

Erik Vogt

Staff Report No. 706  
December 2014



This paper presents preliminary findings and is being distributed to economists and other interested readers solely to stimulate discussion and elicit comments. The views expressed in this paper are those of the author and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. Any errors or omissions are the responsibility of the author.

## **Option-Implied Term Structures**

Erik Vogt

*Federal Reserve Bank of New York Staff Reports*, no. 706

December 2014

JEL classification: C12, C14, C58, G12, G13, G17

### **Abstract**

The illiquidity of long-maturity options has made it difficult to study the term structures of option spanning portfolios. This paper proposes a new estimation and inference framework for these option-implied term structures that addresses long-maturity illiquidity. By building a sieve estimator around the risk-neutral valuation equation, the framework theoretically justifies (fat-tailed) extrapolations beyond truncated strikes and between observed maturities while remaining nonparametric. New confidence intervals quantify the term structure estimation error. The framework is applied to estimating the term structure of the variance risk premium and finds that a short-run component dominates market excess return predictability.

Key words: equity risk premium, finance, options, predictability, sieve M estimation, state-price density, term structures, variance risk premium, VIX

---

Vogt: Federal Reserve Bank of New York (e-mail: erik.vogt@ny.frb.org). An earlier version of this paper circulated as “Term Structures, Shape Constraints, and Inference for Option Portfolios” and appeared as a chapter in the author’s Ph.D. dissertation at Duke University. 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. The author 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, 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, and QMS Capital Management for valuable comments and suggestions. The views expressed in this paper are those of the author and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

# 1 Introduction

A large class of economically important option portfolios have a natural term structure, obtained by varying the maturities of the options combined. For example, the VIX and model-free variance spanning portfolios of Carr and Wu (2009), Jiang and Tian (2005), Britten-Jones and Neuberger (2000) and the skewness and kurtosis portfolios of Bakshi et al. (2003), Chang et al. (2013), and Conrad et al. (2013) each have a term structure that is indexed by option maturity. With such term structures in hand, asset pricing models that make predictions about prices and risk premia over multiple horizons can be tested by their ability to replicate the observed term structures. This intuition has been pursued extensively in the fixed income literature on the term structure of interest rates and could by analogy be carried over to the option setting.

However, unlike the term structure of interest rates, option-implied term structures are not directly observed along two dimensions (maturity and strike) and hence must often be constructed or estimated from available option prices. Any such estimates must invariably confront the realities of option data, which truncate strikes, display only a handful of maturities, and are often contaminated with microstructure noise.<sup>1</sup> These realities are exacerbated at longer maturities, where options are well known to be less liquid.<sup>2</sup> Thus, while intuition suggests that option portfolios computed from liquid short-maturity option contracts should be more precise than their illiquid counterparts, there exists, to date, no formal framework for quantifying this intuition.

This paper proposes a nonparametric estimation and inference framework for the term structures of option portfolios like the VIX and related measures. By building a sieve estimator around the risk-neutral valuation equation, the framework theoretically justifies fat-tailed extrapolations beyond truncated strikes while remaining nonparametric. This feature is particularly relevant for long-maturity options, whose strikes often do not provide sufficient tail coverage to produce reliable estimates of long-run spanning portfolios. Furthermore, the framework makes theoretically supported predictions for option portfolios even for sparsely traded maturities, sidestepping the need for atheoretical maturity interpolations. Finally, it puts new confidence intervals on estimated portfolio term structures that arise when option prices are observed with microstructure error.

---

<sup>1</sup>See, for example, Carr and Wu (2009, sec. 4) and Jiang and Tian (2005, sec. 1.2).

<sup>2</sup>The concern about longer maturity options in the context of forming model-free option spanning portfolios has been pointed out, e.g. in Ait-Sahalia et al. (2014, p. 4) and Driessen et al. (2009, p. 1384)).

The confidence intervals apply to each maturity along the portfolio term structure and facilitate comparisons of precision across maturities.

The framework is motivated by the tension that arises when trying to estimate model-free option spanning portfolios from less reliable long-maturity options: While the “model-free” requirement calls for a nonparametric approach, the reduced signal from long-maturity options calls for additional structure. Therefore, to incorporate structure in a model-free way, I build a sieve estimator around the risk-neutral valuation equation, which itself contains valuable shape information about option prices at all maturities (even unobserved ones).<sup>3</sup> Thus by forming basis function expansions to the term structure of state-price densities (SPDs) and integrating against option payoffs, I obtain candidate option surfaces that incorporate the shape information from the valuation equation. A sieve least squares regression then fits the candidate price surface to observed option prices by optimizing over the coefficients of the expansion terms. Option portfolio term structures, which are the final objects of interest, are then merely functionals of the estimated price surface, so that a simple nonlinear least squares (NLLS) sandwich-covariance matrix leads to their confidence intervals.

The main results of the paper are closed-form expressions for sieve option prices, the consistency of the nonparametric price surface, its rate of convergence, and the asymptotic distribution theory for the term structure of the option spanning portfolios. The distribution theory is the first of its kind in the context of estimating option spanning portfolios and quantifies the intuition that not all portfolios along the term structure are estimated with equal precision. Furthermore, the sieve option prices themselves have an appealing interpretation as expansions around the [Black and Scholes \(1973\)](#) formula, with higher order terms that account for non-Gaussian features of the risk-neutral return distribution. However, in contrast to existing expansions involving the Black-Scholes formula, e.g. [Jarrow and Rudd \(1982\)](#) or [Kristensen and Mele \(2011\)](#), the present sieve framework requires the number of expansion terms to grow slowly with the sample size, allowing ever-increasing flexibility to fitting prices from the unknown data-generating process. Asymptotically, the tails of the sieve SPDs are of polynomial order, which gives rise to fat-tailed extrapolations beyond truncated

---

<sup>3</sup>This shape information goes back to [Breedon and Litzenberger \(1978\)](#) and [Banz and Miller \(1978\)](#) and has been since been exploited elsewhere, for example, in [Aït-Sahalia and Lo \(1998\)](#), [Aït-Sahalia and Duarte \(2003\)](#), [Yatchew and Härdle \(2006\)](#), [Figlewski \(2008\)](#), [Rompolis and Tzavalis \(2008\)](#), and [Polkovnichenko and Zhao \(2013\)](#) in the context of estimating state-price densities.

strikes along every point on the term structure of SPDs.

I apply the framework to estimating the term structure of the VIX and its square, the synthetic variance swap, in both simulation and empirical settings. The simulation exercises show that the sieve substantially outperforms existing methods in situations where a subset of maturities displays significant strike truncation. This situation is quite common in the data and arises somewhat surprisingly in crisis periods, when the liquidity for longer maturity options appears to dry up. The sieve's performance in these situations is largely due to its structure, which, in addition to providing tail extrapolations, uses information from all option maturities simultaneously to inform VIX estimates at illiquid maturities. This differs from existing methods, which can be viewed as using information only local to the maturity of interest.

The simulations further show that the sieve can fit option prices well regardless of whether the underlying data were generated by a [Heston \(1993\)](#) stochastic volatility (SV) model, an SV model with jumps in prices, or an SV model with jumps in prices and jumps in volatility ([Duffie et al. \(2000\)](#)). In particular, a Monte Carlo exercise demonstrates that the sieve term structures display size control in finite samples when the expansion terms are chosen to minimize a data-driven and computationally convenient information criterion. Finally, the simulations also show that the number of expansion terms selected in this way increases with the complexity of the DGP.

In an empirical application, I study the term structure of the variance risk premium embedded in S&P 500 index options from 1996 to 2013. The variance risk premium measures the compensation that investors demand for bearing return variance risk over a given horizon.<sup>4</sup> Using the sieve framework proposed in this paper combined with a novel set of expectation hypothesis and return predictability regressions, I find that the compensation for bearing variance risk is dominated by a short-run component. That is, investors earn a significant premium for selling securities that pay off only if realized return variance spikes in the next one to two months. In contrast, a security that affords similar protection from realized variance over longer periods that exclude the next two months commands no measurable premium. This result is consistent with the recent evidence presented in [Dew-Becker et al. \(2014\)](#), who find a similar pattern in the unconditional Sharpe ratios of a trading strategy involving variance swaps.

---

<sup>4</sup>See, for example, [Bakshi and Madan \(2006\)](#), [Carr and Wu \(2009\)](#), [Bollerslev and Todorov \(2011\)](#), [Bollerslev et al. \(2011\)](#), [Drechsler and Yaron \(2011\)](#), [Bollerslev et al. \(2013\)](#), [Bekaert and Hoerova \(2014\)](#) and the references therein.

The paper connects with several strands of the literatures in finance and econometrics. The study of option spanning portfolios for a single maturity is well-established, and has been developed theoretically, for example, by [Britten-Jones and Neuberger \(2000\)](#), [Bakshi and Madan \(2000\)](#), [Jiang and Tian \(2005\)](#), and [Carr and Wu \(2009\)](#), and explored empirically by [Bakshi et al. \(2003\)](#), [Chang et al. \(2013\)](#), [Conrad et al. \(2013\)](#). The framework developed in this paper is for option-implied term structures and is distinct because estimation involves simultaneous use of options across all maturities. This extension is related to the burgeoning empirical literature on option-implied term structures, including exponential claims to integrated variance as in [Bakshi et al. \(2011\)](#), the dividend strip term structure of [van Binsbergen et al. \(2012\)](#), as well as the term structure of the variance risk premium, which has only recently been explored in [Ait-Sahalia et al. \(2014\)](#) and [Dew-Becker et al. \(2014\)](#). Furthermore, the development of confidence intervals for the portfolio term structure is new to this literature. Econometrically, the sieve framework I develop is an adaptation of the theory of [Shen \(1997\)](#), [Chen and Shen \(1998\)](#), [Chen \(2007\)](#), and [Chen et al. \(2014\)](#). The contribution to this literature is a sieve regression framework in which the regressor is a functional of the infinite-dimensional parameter of interest, requiring additional continuity results that are specific to option pricing. Finally, the basis functions employed for the sieve expansions are conditional analogs to those of [Gallant and Nychka \(1987\)](#) and [León et al. \(2009\)](#).

## 2 Main Idea

Standing at time 0, the object of interest is the term structure of a general class of model-free option spanning portfolios. To fix ideas consider a specific member of this class, the synthetic variance swap (SVS) ([Carr and Wu \(2009\)](#)), whose square root is commonly referred to as the VIX volatility index when the options are written on the S&P 500 index. For a given time horizon  $\tau$ , the SVS is obtained by combining European put and call options with different strikes  $\kappa$  and common maturity  $\tau$  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  $\tau$  of interest, the SVS term structure is the function

$$SVS_0(\tau) = \frac{2}{\tau} e^{r\tau} \int_0^{F(\mathbf{Z})} \frac{1}{\kappa^2} P_0(\mathbf{Z}) d\kappa + \frac{2}{\tau} e^{r\tau} \int_{F(\mathbf{Z})}^{\infty} \frac{1}{\kappa^2} C_0(\mathbf{Z}) d\kappa, \quad (2.1)$$

Table 1: Sample Averages for Monthly S&amp;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 - $\infty$	18.7	1521.4	90 235.1	4.46
0 - $\infty$	420.9	229 072.0	4 321 320.0	1.35

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  $\tau$  across an infinite continuum of  $\kappa$  in order to get at the portfolio term structure  $SV S_0(\tau)$ .

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 cross-section of S&P 500 index options contains about  $n = 420$  prices, with most of the observations concentrated at short maturities. The thinning of available option quotes for increasing  $\tau$  is also associated with smaller trading volumes, less open interest, and widening bid-ask spreads.<sup>5</sup> The widening spreads introduce varying levels of uncertainty about  $P_0(\mathbf{Z})$  for different  $\tau$ , which I allow for by letting  $P_i = P_0(\mathbf{Z}_i) + \varepsilon_i$  for  $\varepsilon_i$  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, I follow the literature and refer to the thinning of prices and increased noise as manifestations of illiquidity at longer maturities.<sup>6</sup>

To preserve the SVS's interpretation as a model-free spanning portfolio, I estimate  $P_0(\mathbf{Z})$  non-parametrically. However, instead of regressing  $P_i$  on  $\mathbf{Z}_i$  directly, the approach here is to exploit our economic knowledge of the functional form  $P_0(\mathbf{Z})$ . As shown below, the risk-neutral valuation

<sup>5</sup>Note that [Table 1](#) reports dollar bid-ask spreads because dollar values enter the integral in [\(2.1\)](#). The data set follows the CBOE data filters and is discussed further in [Section 6](#).

<sup>6</sup>See, e.g. [Aït-Sahalia et al. \(2014\)](#), [Driessen et al. \(2009\)](#).



equation states that there exists a conditional density  $f_0$  such that  $P_0(\mathbf{Z}) = P(f_0, \mathbf{Z})$  for known  $P(\cdot, \mathbf{Z})$ . The fact that  $f_0$  is a density imposes economically motivated shape restrictions on  $P_0(\mathbf{Z})$  in the  $(\kappa, \tau)$  space. A key element of the proposed framework is that this shape information can be used to guide our estimates of  $P_0$  for regions of  $(\kappa, \tau)$  that are poorly observed due to liquidity issues.

The goal, therefore, is to only consider  $P(f, \mathbf{Z})$  for candidates  $f$  that are valid densities.<sup>7</sup> Denote the collection of these candidate densities by  $\mathcal{F}$ . One possible approach might then involve solving

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

thereby obtaining estimates  $\hat{P}(\mathbf{Z}) \equiv P(\hat{f}, \mathbf{Z})$  of  $P_0(\mathbf{Z})$ . However, in the present nonparametric framework, the space  $\mathcal{F}$  is infinite-dimensional. As is well known, in general optimizing over an infinite-dimensional function space may not be feasible or could even be ill-posed. In this case, it is typical to proceed by the method of sieves, which involves approximating  $\mathcal{F}$  by a sequence of finite-dimensional function spaces (the "sieve" spaces)  $\mathcal{F}_K \subset \mathcal{F}_{K+1} \subset \dots \subset \mathcal{F}$  that are compact and arbitrarily dense in  $\mathcal{F}$  for large  $K$  [see [Chen \(2007\)](#), [Chen and Shen \(1998\)](#), and [Shen \(1997\)](#)].

Thus, given such a sequence of approximating spaces, the feasible optimization problem to be solved will be of the form

$$\hat{f}_{K_n} = \arg \min_{f \in \mathcal{F}_{K_n}} \left\{ \frac{1}{n} \sum_{i=1}^n [P_i - P(f, \mathbf{Z}_i)]^2 W(\mathbf{Z}_i) \right\}, \quad (2.2)$$

where  $K_n \rightarrow \infty$  slowly as  $n \rightarrow \infty$ . Informally, as the sample size grows, the approximating spaces  $\mathcal{F}_{K_n}$  increasingly resemble the parent space  $\mathcal{F}$ , so that optima on  $\mathcal{F}_{K_n}$  should indeed converge to  $f_0$ . [Proposition 2](#) below makes this notion precise and further shows that the option price estimates  $\hat{P}(\mathbf{Z}) \equiv P(\hat{f}_{K_n}, \mathbf{Z})$  also converge to the true option price  $P(f_0, \mathbf{Z})$ .

Computationally, [\(2.2\)](#) is equivalent to non-linear least squares. This is because every sieve density  $f \in \mathcal{F}_{K_n}$  can be represented as nonlinear combinations of certain basis functions. The coefficients of these basis functions,  $\beta_n \in \mathbb{R}^{K_n}$  then become the variables of optimization. Moreover, [Proposition 1](#) shows that sieve densities, when integrated against the option's payoff, yield closed-

---

<sup>7</sup>If one were to instead evaluate  $P(g, \mathbf{Z})$  for  $g$  taking negative values on a set of non-zero Lebesgue measure, then the prices  $P(g, \cdot)$  would admit an arbitrage opportunity.

form option prices  $P(f, \mathbf{Z}_i)$  for all  $f \in \mathcal{F}_{K_n}$ . Thus, (2.2) amounts to a nonlinear least-squares problem with closed-form regressors and is therefore accessible with standard statistical software.

Estimates of the SVS term structure are then obtained by evaluating

$$\widehat{SVS}(\tau) = \frac{2}{\tau} e^{r\tau} \int_0^{F(\mathbf{Z})} \frac{1}{\kappa^2} \widehat{P}(\mathbf{Z}) d\kappa + \frac{2}{\tau} e^{r\tau} \int_{F(\mathbf{Z})}^{\infty} \frac{1}{\kappa^2} \widehat{C}(\mathbf{Z}) d\kappa, \quad (2.3)$$

point-wise in  $\tau$ , 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  $\tau$ .

Finally, 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

$$\Gamma(\widehat{P}) = g \left( \int_{\mathcal{Z}_1} a(\mathbf{Z}_1, \mathbf{Z}_2) \widehat{P}(\mathbf{Z}_1, \mathbf{Z}_2) d\mathbf{Z}_1 + \int_{\mathcal{Z}_1^c} b(\mathbf{Z}_1, \mathbf{Z}_2) \widehat{C}(\mathbf{Z}_1, \mathbf{Z}_2) d\mathbf{Z}_1 \right), \quad (2.4)$$

where  $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)'$  and  $\mathcal{Z}_1$  is a subset of the domain of  $\mathbf{Z}_1$ .<sup>8</sup> This class of portfolios encompasses many objects of interest beyond the SVS, and can include e.g. the skewness and kurtosis portfolios of Bakshi et al. (2003), in which case  $\mathbf{Z}_1 = \kappa$ . Because portfolios of this form represent regular functionals in the sense of Chen et al. (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 4 below shows how this theory can be used to establish results of the form  $\sqrt{n} \widehat{V}^{-1/2} (\widehat{SVS}(\tau) - SVS_0(\tau)) \rightarrow^d N(0, 1)$ .

Figure 1 illustrates the resulting confidence intervals for the SVS term structure, converted to standard deviations in order to be directly comparable to the CBOE's  $VIX = 100\sqrt{SVS}$ . The figure illustrates that the object of interest is a curve indexed by option maturity, and that sampling error around this curve of as much as 500bp can emerge at longer maturities. For comparison, I plotted an alternative VIX term structure by applying the CBOE's discrete approximation to the same option data. The figure shows that the two curves can deviate substantially, in particular during the two crisis periods 23-Sep-1998 (bailout of Long-Term Capital Management) and 31-Oct-2008 (financial crisis after Lehman bankruptcy). Below, we will see in the data and in a simulation environment that these deviations are attributed to severe strike truncation at long maturities, a

<sup>8</sup>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,  $g(x) = x$ .

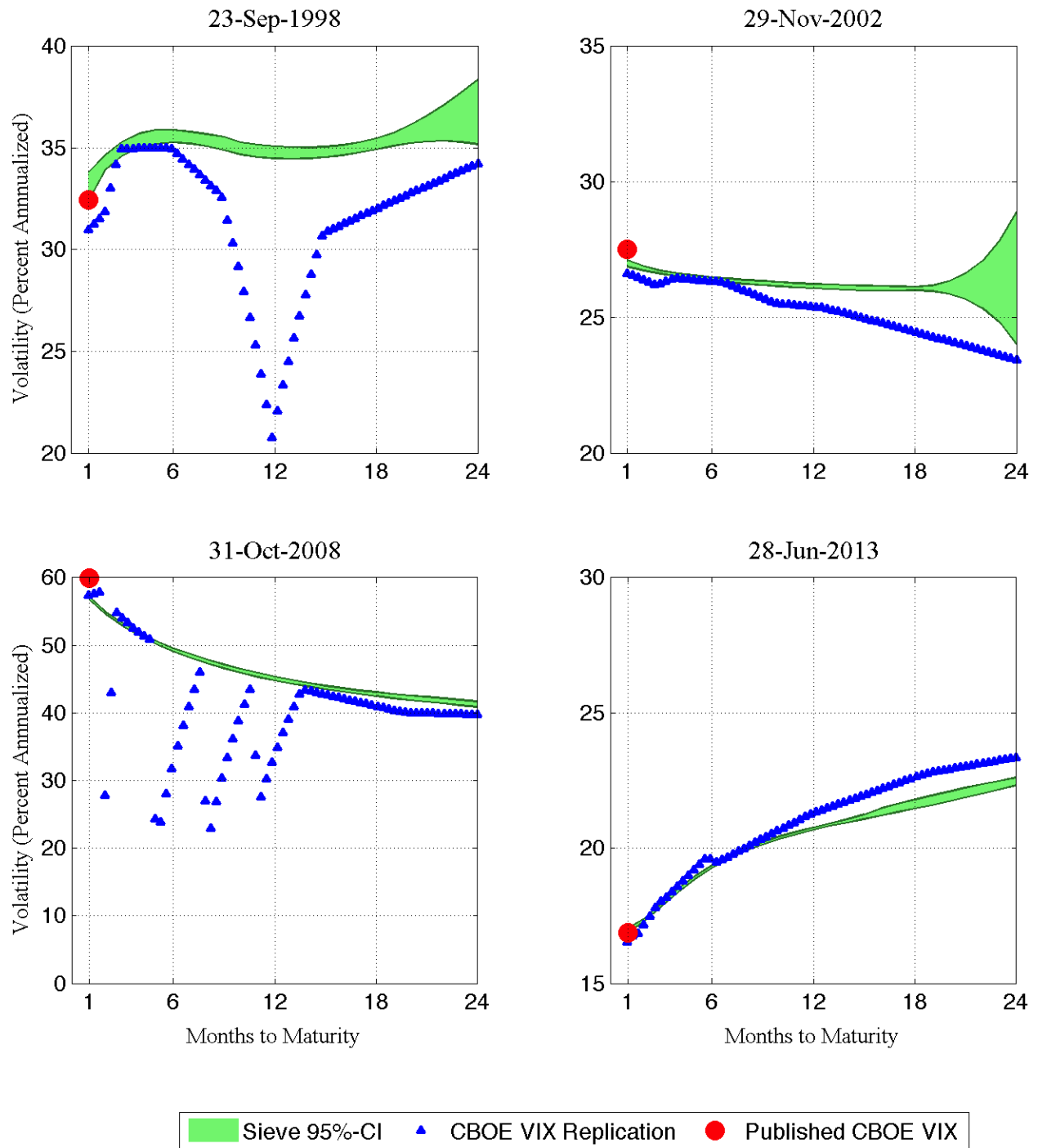


Figure 1: Sieve Estimated  $VIX = 100\sqrt{SVS}$  Term Structure. Using the full option cross-sections on the indicated days, the sieve regression (2.2) is estimated to obtain an option surface  $\hat{P}(\mathbf{Z})$ . The functional (2.3) is then computed, with  $\hat{C}(\mathbf{Z})$  computed by put-call parity. The shaded (green) region corresponds to the 95% - confidence intervals of the VIX term structure following the result in Proposition 4 in the text. Blue triangles correspond to the CBOE's VIX calculation on the same option data, and the red dot is the actual published 30-day CBOE VIX.

feature that persists at long maturities for a number of days during crises.

In sum, the framework can be described as consisting of two steps: In the first step, a shape-conforming price surface is estimated via the sieve least squares regression (2.2). In the second step, the portfolio term structures and confidence intervals of (2.4) are obtained using the sieve estimates as inputs. The next section is concerned with deriving each ingredient for the regression (2.2): it outlines the pricing functional  $P(f, \mathbf{Z})$ , followed by the function spaces  $\mathcal{F}_K$  and  $\mathcal{F}$ , and the closed-form regressors  $P(f_K, \mathbf{Z})$  for any  $f_K \in \mathcal{F}_K$ .

### 3 Estimation Framework

#### 3.1 Constructing the Regressors from Theory

The goal is to exploit the structure afforded by the well-known risk-neutral valuation equation.<sup>9</sup> Thus, for a vector of characteristics  $\mathbf{Z} = (\kappa, \tau, r, q)$ , the true option price is modeled as

$$P_0(\mathbf{Z}) \equiv e^{-r\tau} E_0^{\mathbb{Q}} \left[ [\kappa - S]_+ \middle| \tau, \mathbf{V} = \mathbf{v}_0 \right] = e^{-r\tau} \int_0^{\kappa} [\kappa - S] f_0^{\mathbb{Q}}(S | \tau, \mathbf{V} = \mathbf{v}_0) dS, \quad (3.1)$$

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 et al. (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, I therefore define  $f_0^{\mathbb{Q}}(S | \tau) \equiv f_0^{\mathbb{Q}}(S | \tau, \mathbf{V} = \mathbf{v}_0)$ , since  $\mathbf{v}_0$  is static across the option surface. On the other hand,  $\tau$  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

$$P_0(\mathbf{Z}) \equiv e^{-r\tau} \int_0^{\kappa} [\kappa - S] f_0^{\mathbb{Q}}(S | \tau) dS. \quad (3.2)$$

---

<sup>9</sup>See, for example, Chapters 6 and 8 in Duffie (2001) for a discussion of risk-neutral pricing.

The dependence of the option price on the SPD  $f_0^{\mathbb{Q}}$  and the characteristics  $\mathbf{Z}$  can be expressed as

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

The no-arbitrage pricing equation (3.2) implies shape restrictions on the option prices. Differentiating  $P_S(f_0^{\mathbb{Q}}, \mathbf{Z})$  repeatedly with respect to the strike price  $\kappa$  yields the conditions

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

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  $\kappa$  for any  $\tau$ , and additionally has slope  $e^{-r\tau}$  as  $\kappa \rightarrow \infty$  and slope 0 as  $\kappa \rightarrow 0$ . 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  $\tau$ .<sup>10</sup>

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

To this end, let  $Y$  be the random variable that satisfies

$$\log\left(\frac{S}{S_0}\right) = \mu(\mathbf{Z}) + \sigma(\mathbf{Z})Y, \tag{3.3}$$

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 (3.3) hold.

---

<sup>10</sup>These shape constraints have been exploited elsewhere in the nonparametric option pricing literature for a single  $\tau$ . See, for example, [Aït-Sahalia and Duarte \(2003\)](#), [Bondarenko \(2003\)](#), [Yatchew and Härdle \(2006\)](#), and [Figlewski \(2008\)](#).

Under this change of variables, the valuation equation (3.2) becomes

$$P_S(f_0^{\mathbb{Q}}, \mathbf{Z}) = e^{-r\tau} \int_0^{\kappa} (\kappa - S) f_0^{\mathbb{Q}}(S|\tau) dS = e^{-r\tau} \int_0^{d(\mathbf{Z})} (\kappa - S_0 e^{\mu(\mathbf{Z}) + \sigma(\mathbf{Z})Y}) f_0(Y|\tau) dY \equiv P(f_0, \mathbf{Z}), \quad (3.4)$$

where

$$d(\mathbf{Z}) = \frac{\log(\kappa/S) - \mu(\mathbf{Z})}{\sigma(\mathbf{Z})} \quad (3.5)$$

and

$$f_0^{\mathbb{Q}}(s|\tau) = (s\sigma(\mathbf{Z}))^{-1} f_0(s|\tau) \quad (3.6)$$

follow from a Jacobian transformation.

Since (3.4) says  $P_S(f_0^{\mathbb{Q}}, \mathbf{Z}) = P(f_0, \mathbf{Z})$ , one can focus on option pricing equations of the form

$$P(f, \mathbf{Z}) = e^{-r\tau} \int_0^{d(\mathbf{Z})} (\kappa - S_0 e^{\mu(\mathbf{Z}) + \sigma(\mathbf{Z})Y}) f(Y|\tau) dY. \quad (3.7)$$

This is the functional that appears as the regressor in (2.2). Furthermore, it is easy to verify that (3.7) satisfies the same shape restrictions as (3.2) for any  $f$  with  $f(y|\tau) \geq 0$  and  $\int f(y|\tau) dy = 1$ .

### 3.2 Characterizing $\mathcal{F}$ and its Sieve Spaces $\mathcal{F}_K$

To arrive at consistency, a convergence rate, and inference results,  $\mathcal{F}$  and its sieve spaces  $\mathcal{F}_K$  are assumed to belong to Sobolev spaces. Restricting  $\mathcal{F}$  and  $\mathcal{F}_K$  in this way effectively rules out the possibility that the true option prices are generated by a very rough or oscillatory state-price density. However, because a full description of the relevant Sobolev spaces is not needed for implementation, I relegate the formal presentation of  $\mathcal{F}$  and  $\mathcal{F}_K$  to [Appendix A](#) and focus instead on intuition.

For an option surface to conform to the theoretical shape restrictions of (3.7) for any  $\tau$ ,  $\mathcal{F}$  must be a function space consisting of conditional densities  $f(Y|\tau)$  in the sense that  $f(y|\tau) \geq 0$  and  $\int f(y|\tau) dy = 1$  for all  $\tau$ . I construct such functions by first defining a collection of joint densities  $\mathcal{F}^{Y,\tau}$  with elements  $f^{Y,\tau}(y, \tau)$ , and then defining  $\mathcal{F}$  to consist of those functions  $f(y|\tau)$  such that  $f(y|\tau) = f^{Y,\tau}(y, \tau) / \int f^{Y,\tau}(y, \tau) dy$  for some  $f^{Y,\tau} \in \mathcal{F}^{Y,\tau}$ .

[Gallant and Nychka \(1987\)](#) show that if  $\mathcal{F}^{Y,\tau}$  is a certain Sobolev subspace (see [Appendix A](#)

below) and  $\{\mathcal{F}_K^{Y,\tau}\}_{K=0}^\infty$  is a collection of squared and scaled Hermite functions, then  $\{\mathcal{F}_K^{Y,\tau}\}_{K=0}^\infty$  is a valid sieve for  $\mathcal{F}^{Y,\tau}$ . To extend their results to an option pricing setting, [Appendix A](#) below shows that the conditional approximating spaces  $\{\mathcal{F}_K\}_{K=0}^\infty$  consisting of those functions  $f_K$  for which  $f_K(y|\tau) = f_K^{Y,\tau}(y, \tau) / \int f_K^{Y,\tau}(y, \tau) dy$  for some  $f_K^{Y,\tau} \in \mathcal{F}_K^{Y,\tau}$  is also a valid sieve for the conditional parent space  $\mathcal{F}$ , although the topologies differ.

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

$$f_K^{Y,\tau}(y, \tau) = \left[ \sum_{k=0}^{K_y} \left( \sum_{j=0}^{K_\tau} \beta_{kj} H_j(\tau) \right) H_k(y) \right]^2 e^{-\tau^2/2} e^{-y^2/2} = \left[ \sum_{k=0}^{K_y} \alpha_k(B, \tau) H_k(y) \right]^2 e^{-\tau^2/2} e^{-y^2/2}, \quad (3.8)$$

where  $H_k$  are Hermite polynomials of degree  $k$ , and where  $B$  is a matrix of coefficients with  $kj$ -entry  $\beta_{kj}$  and  $K = (K_y + 1)(K_\tau + 1)$ .<sup>11</sup> This function is clearly non-negative. Then, using orthogonality properties of Hermite polynomials, it can be shown that in order for  $\iint f_K^{Y,\tau}(y, \tau) dy d\tau = 1$  for any  $K$ , it suffices to impose  $\sum_{k=0}^{K_y} \sum_{j=0}^{K_\tau} \beta_{kj}^2 = 1$ .<sup>12</sup>

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

$$f_K(y|\tau) = f_K^{Y,\tau}(y, \tau) \Big/ \int f_K^{Y,\tau}(y, \tau) dy \quad (3.9)$$

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

$$\hat{\beta}_n = \arg \min_{\beta \in \mathbb{R}^{K_n}} \left\{ \frac{1}{n} \sum_{i=1}^n [P_i - P(\beta, \mathbf{Z}_i)]^2 W(\mathbf{Z}_i) \right\} \quad \text{s.t.} \quad \sum_{k=0}^{K_y(n)} \sum_{j=0}^{K_\tau(n)} \beta_{kj}^2 = 1, \quad (3.10)$$

<sup>11</sup>The Hermite polynomials are orthogonalized polynomials. They are defined, for scalars  $x$ , by

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

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

<sup>12</sup>Strictly speaking, the joint densities of interest are of the form  $f^{Y,\tau} = [P_K \phi^{1/2}]^2 + \varepsilon h_0$  given in [Appendix A](#), where  $P_K = \sum_{k=0}^{K_y} (\sum_{j=0}^{K_\tau} \beta_{kj} H_j(\tau)) H_k(y)$  and  $\phi$  is the standard normal density. The term  $\varepsilon h_0$ , where  $h_0$  can also be Gaussian, was introduced in [Gallant and Nychka \(1987\)](#) for technical reasons and is also helpful in numerical implementation to prevent division by zero when forming conditional densities. See [Appendix A](#).

which is numerically equivalent to nonlinear least squares estimation for fixed  $K_n$ . As written,  $P(\beta, \mathbf{Z}_i)$  is identical to  $P(f_K, \mathbf{Z}_i)$  from (3.7), which still requires an integration to obtain a candidate option price. Section 3.3 shows that in fact,  $P(f_K, \mathbf{Z}_i)$  is available in closed-form for any  $f_K \in \mathcal{F}_K$ , which considerably facilitates implementation.

*Remark 3.1.* As written, the objective function in (3.10) 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  $[P_i - P(\beta, \mathbf{Z}_i)]^2 = [\{P_i - I(\mathbf{Z}_i)\} - \{P(\beta, \mathbf{Z}_i) - I(\mathbf{Z}_i)\}]^2$ . Because  $\{P_i - I(\mathbf{Z}_i)\}$  and  $\{P(\beta, \mathbf{Z}_i) - I(\mathbf{Z}_i)\}$  assume their largest values at-the-money, the objective function in (3.10) 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 et al. (2013)). However, for inference problems related to option-implied term structures, the Monte Carlo simulations below suggest that simply setting  $W(\mathbf{Z}_i) = 1$  yields superior coverage properties. Additionally, closed-form expressions for the gradients and Hessians of (3.10) are provided in the Online Appendix to this paper.

Finally, note that the sieve densities preserve the required economic shape restrictions. To see this, one differentiates with respect to  $\kappa$  to obtain

$$e^{r\tau} \frac{\partial P(f_K, \mathbf{Z})}{\partial \kappa} = \int_0^{d(\mathbf{Z})} f_K(Y|\tau) dY \tau = F_K \left( \frac{\log(\kappa/S_0) - \mu(\mathbf{Z})}{\sigma(\mathbf{Z})} \middle| \tau \right) \quad (3.11)$$

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

$$\lim_{\kappa \nearrow +\infty} e^{r\tau} \frac{\partial P(f_K, \mathbf{Z})}{\partial \kappa} = 1, \quad \lim_{\kappa \searrow 0} e^{r\tau} \frac{\partial P(f_K, \mathbf{Z})}{\partial \kappa} = 0. \quad (3.12)$$

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

I now provide closed-form expressions for the sieve option prices  $P(f_K, \mathbf{Z})$  to be used in the regression (2.2) and show that  $\mu(\mathbf{Z})$  and  $\sigma(\mathbf{Z})$  can be chosen so that the sieve option prices have a natural interpretation as expansions around the Black-Scholes model.



### 3.3 The Option Price Regressors in Closed-Form

#### 3.3.1 A Building Block: Closed-Form Sieve Densities

To obtain closed-form option prices, it is convenient to first establish a linear representation for the conditional sieve densities  $f_K$  of (3.9). This is done by expanding the squared polynomial term in the joint densities of (3.8) using results from León et al. (2009).

**Lemma 3.2.** *Any  $f_K \in \mathcal{F}_K$  can be expressed in the form*

$$f_K(y|\tau) = \sum_{k=0}^{2K_y} \gamma_k(B, \tau) H_k(y) \phi(y), \quad (3.13)$$

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 (3.8).

*Proof.* [Appendix A](#). □

The use of densities  $f_K(y|\tau)$  that are linear combinations of functions in  $y$  helps with the derivation of closed-form option prices.

#### 3.3.2 Closed-Form Sieve Option Prices

The following result is an extension of Proposition 9 in León et al. (2009) to the case allowing for conditioning on  $\tau$ .

**Proposition 1.** *For a candidate SPD  $f_K(y|\tau) \in \mathcal{F}_K$  of the form given in equation (3.13), the put option price  $P(f_K, \mathbf{Z})$  from equation (3.7) is given by*

$$\begin{aligned} P(f_K, \mathbf{Z}) = & \kappa e^{-r\tau} \left[ \Phi(d(\mathbf{Z})) - \sum_{k=1}^{2K_y} \frac{\gamma_k(B, \tau)}{\sqrt{k}} H_{k-1}(d(\mathbf{Z})) \phi(d(\mathbf{Z})) \right] \\ & - S_0 e^{-r\tau + \mu(\mathbf{Z})} \left[ e^{\sigma(\mathbf{Z})^2/2} \Phi(d(\mathbf{Z}) - \sigma(\mathbf{Z})) + \sum_{k=1}^{2K_y} \gamma_k(B, \tau) I_k^*(d(\mathbf{Z})) \right] \end{aligned} \quad (3.14)$$

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

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

$$I_0^*(d(\mathbf{Z})) = e^{\sigma(\mathbf{Z})^2/2} \Phi(d(\mathbf{Z}) - \sigma(\mathbf{Z})),$$

and  $\gamma_k(B, \tau)$  is the coefficient function given in equation (3.13).

The price of a call option is given by

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

$$- \kappa e^{-r\tau} \left[ [1 - \Phi(d(\mathbf{Z}))] - \sum_{k=1}^{2K_y} \frac{\gamma_k(B, \tau)}{\sqrt{k}} H_{k-1}(d(\mathbf{Z})) \phi(d(\mathbf{Z})) \right]. \quad (3.15)$$

*Proof.* [Appendix B](#). □

*Remark 3.3.* A consequence of this result is that it makes the sieve regressor function of (3.10) available in closed-form. Indeed, in sharp contrast to the large class of parametric option pricing models of [Heston \(1993\)](#) and [Duffie et al. \(2000\)](#), no numerical integrations are required to compute an option price, which significantly facilitates the optimization problem (3.10). Moreover, the Online Appendix to this paper also provides closed-form gradients and second derivatives of the prices  $P(f_K, \mathbf{Z})$ . And finally, having closed-form price estimates drastically simplifies the ultimate objective of computing integrated portfolios of  $P(f_K, \mathbf{Z})$ .

### 3.3.3 Theoretical Interpretation of Option Pricing Formulas

The sieve put option price in (3.14) has an intuitive interpretation. Rearranging equation (3.14), one obtains

$$P(f_K, \mathbf{Z}) = \kappa e^{-r\tau} \Phi(d(\mathbf{Z})) - S_0 e^{-r\tau + \mu(\mathbf{Z})} e^{\sigma(\mathbf{Z})^2/2} \Phi(d(\mathbf{Z}) - \sigma(\mathbf{Z}))$$

$$- \sum_{k=1}^{K_y} \gamma_k(B, \tau) \left[ \frac{1}{\sqrt{k}} H_{k-1}(d(\mathbf{Z})) \phi(d(\mathbf{Z})) + S_0 e^{-r\tau + \mu(\mathbf{Z})} I_k^*(d(\mathbf{Z})) \right]. \quad (3.16)$$

Inspection of equation (3.16) shows that choosing

$$\sigma(\mathbf{Z}) \equiv \sigma\sqrt{\tau}, \quad \mu(\mathbf{Z}) \equiv (r - q - \sigma^2/2)\tau \quad (3.17)$$

will cause the leading term in equation (3.16) 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 (3.5) is now  $d(\mathbf{Z}) = (\log(\kappa/S_0) - (r - q - \sigma^2/2)\tau)/(\sigma\sqrt{\tau})$ . The value  $\sigma$  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,

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

This formula can be interpreted as ‘‘centering’’ the sieve at Black-Scholes, and then supplementing it with higher-order ‘‘correction’’ terms.<sup>13</sup> As the sample size  $n$  increases, the number of correction terms,  $K_y$  and  $K_\tau$ , also increase, albeit at a slower rate than  $n$ .<sup>14</sup> Thus, the more data one has, the more complex the sieve option pricer is permitted to be relative to Black-Scholes.

If the  $\gamma_k(B, \tau)$  terms for  $k \geq 1$  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 et al. \(1997\)](#), [Duffie et al. \(2000\)](#)). The simulation study in [Section 5](#) 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

<sup>13</sup>Recently, [Kristensen and Mele \(2011\)](#), [Xiu \(2011\)](#), and [León et al. \(2009\)](#) have employed Hermite polynomials in a parametric option pricing setting.

<sup>14</sup>Recall that the  $\gamma_k(B, \tau)$  terms also contain expansions in the maturity dimension.

quite capable of fitting the conditional distributions implied by complicated stochastic volatility and jump specifications.

## 4 Asymptotic Results

The critical feature of  $P(f, \mathbf{Z})$  in (3.7) is that it generates shape-conforming option prices for any  $\tau$ . It does so by indexing state-price densities with maturity  $\tau$ , which appears as a conditioning variable. A straightforward extension of (3.7) is allowing a state-price density with arbitrary conditioning information,  $f(Y|\mathbf{X})$ , where  $\mathbf{X} \subseteq \mathbf{Z}$  and contains  $\tau$ . For example, one could have  $\mathbf{X} = (\tau, r)$  to accommodate a risk-free rate term structure that does not match the maturities of observed option prices. Allowing for general  $\mathbf{X}$  is instructive in order to see how the rate of convergence is slowed by the dimension of the conditioning variable  $\mathbf{X}$ . Therefore, this section establishes the asymptotic theory for the extension that allows for arbitrary conditioning information in the SPD. The results below also refer to

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

the set of option prices obtained by integrating option payoffs against SPDs in  $\mathcal{F}$ .

**Proposition 2** now shows that under sufficient conditions given in [Appendix B](#), sieve-estimated option prices obtained by setting  $\hat{P}_n \equiv P(\hat{f}_n, \mathbf{Z})$ , where  $\hat{f}_n$  solves the least squares problem (2.2), are consistent for the true option prices  $P_0 = P(f_0, \mathbf{Z})$ . The consistency norm is the  $L^2(\mathbb{R}^{d_z}, W d\mathbb{P})$  norm given by

$$\|P_1 - P_2\|_2^2 = \mathbb{E}\{[P_1(\mathbf{Z}) - P_2(\mathbf{Z})]^2 W(\mathbf{Z})\} = \int [P_1(\mathbf{Z}) - P_2(\mathbf{Z})]^2 W(\mathbf{Z}) \mathbb{P}(d\mathbf{Z}) \quad (4.1)$$

on the space of option prices  $\mathcal{P}$ .

**Proposition 2.** (*Consistency*) *Assumptions A.1 and B.1 imply  $\|\hat{P}_n - P_0\|_2 \xrightarrow{p} 0$ .*

*Proof.* [Appendix B](#). □

The proof of [Proposition 2](#) establishes a link between option prices  $\mathcal{P}$  and the state-price density space  $\mathcal{F}$  from which they are generated. Next, [Proposition 3](#) establishes the rate at which  $\hat{P}_n$  converges to  $P_0$ . Informally, the rate depends on the size and complexity of the space of admissible

option prices  $\mathcal{P}$ , which is indexed by the space of conditional densities  $\mathcal{F}$ , which in turn is indexed by the space of joint densities  $\mathcal{F}^{Y,X}$ . Thus, the size and complexity of  $\mathcal{P}$  is determined by the size and complexity of the joint densities, which reside in a Sobolev space whose members have bounded derivatives up to order  $m_0 + m$  (Appendix A).

**Proposition 3.** *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.* Appendix B. □

Intuitively, the rate result says that option prices converge at a faster rate when (a) the state-price densities considered are smoother (larger  $m_0 + m$ ), (b) the sieve spaces  $\{\mathcal{F}_K\}$  fill in the parent space  $\mathcal{F}$  at a faster rate (larger  $\alpha$ ), and (c) the dimension of the conditioning variable is smaller (smaller  $d_u = 1 + d_x$ ). Expressions of  $\alpha$  in terms of the more primitive  $m_0, m$ , and  $d_u$  are, as yet, unavailable for the particular weighted Sobolev spaces that underlie this framework.

The rate result is used to establish the asymptotic distribution of a variety of option portfolios, including the synthetic variance swap of Carr and Wu (2009), the skewness and kurtosis portfolios of Bakshi et al. (2003), and the exponential claims to integrated variance given in Bakshi et al. (2011). The key observation is that these option portfolios are functionals of the estimated price  $\widehat{P}_n$ , which was introduced as  $\Gamma(\widehat{P}_n)$  in (2.4) above. Given this functional relationship, intuition then suggests that inference on the portfolio functional  $\Gamma(\cdot)$  should follow from a functional delta method on  $\Gamma(\widehat{P}_n)$ .

Fortunately, Proposition 4 now shows that things are much simpler in practice because the standard parametric sandwich covariance matrix is applicable once  $K_n$  is chosen appropriately. This feature is due to an insight of Chen et al. (2014), who show an equivalence between the parametric covariance matrix and the Riesz representer of the derivative of functionals like  $\Gamma(\cdot)$ . Specifically, let  $\Xi_i \equiv (P_i, \mathbf{Z}_i)$  denote observations on option prices and characteristics, and define  $\ell(\beta, \Xi_i) \equiv -\frac{1}{2}[P_i - P(\beta, \mathbf{Z}_i)]^2 W(\mathbf{Z}_i)$ .

**Proposition 4.** *Under Assumptions A.1 and B.1–B.5,*

$$\sqrt{n}\widehat{V}_n^{-1/2}[\Gamma(\widehat{P}_n) - \Gamma(P_0)] \xrightarrow{d} N(0, 1) \quad (4.2)$$

where

$$\widehat{V}_n = \widehat{G}'_{K_n} \widehat{R}_{K_n}^{-1} \widehat{\Sigma}_{K_n} \widehat{R}_{K_n}^{-1} \widehat{G}_{K_n} \quad (4.3)$$

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.](#) □

The objects in (B.8) are the usual quantities involved in the estimation of the variance matrix in nonlinear least squares problems. For example, if  $\Gamma$  represents the 1-month  $VIX(1)$ , i.e.  $VIX(1) = 100\sqrt{SVS(1)}$  for the synthetic variance swap of (2.1), then [Proposition 4](#) says that the  $VIX(1)$  is asymptotically normally distributed with estimated variance  $\widehat{V}_n$  computed above. Moreover, this calculation can be done to obtain  $VIX(\tau)$  for arbitrary  $\tau$ , which enables the construction of  $VIX$  term structures that quantify the estimation error involved with the construction of long-term  $VIX$ 's.

## 5 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, I simulate empirically realistic option price data from DGPs of varying complexity. Within the simulations, while I observe the DGP, the sieve does not. Instead, 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. In this section, I propose a data-driven method for choosing  $K_n$  that performs well across several tests.

The simulations in this section refer to various subcases of the following general data gener-

ating process.

$$\begin{aligned}
dX_t &= \left( r - q - \lambda\bar{\mu} - \frac{1}{2}V_t \right) dt + \rho\sqrt{V_t}dW_t + J_t dN_t \\
dV_t &= \kappa_v(\bar{V} - V_t)dt + \rho v\sqrt{V_t}dW_t + (1 - \rho^2)^{1/2} v\sqrt{V_t}dW'_t + Z_t dN_t
\end{aligned}
\tag{5.1}$$

where  $V_t$  is a stochastic volatility process,  $W_t$  and  $W'_t$  are standard Brownian motions, and  $\kappa_v, \bar{V}, \rho, v$  parametrize the volatility process' mean reversion, long-run mean, the leverage effect, and the volatility of volatility, respectively.  $N_t$  is a Poisson process with arrival intensity  $\lambda$  and compensator  $\lambda\bar{\mu}$ , where  $\bar{\mu} = \exp(\mu_J + 0.5\sigma_J^2)/(1 - \mu_v - \rho_J\mu_v) - 1$ . 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 et al. \(2000\)](#) 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 et al. \(2012\)](#) and are given in the Online Appendix to this paper.

## 5.1 Goodness of Fit: VIX Term Structures

Recall from [Figure 1](#) that the sieve-estimated VIX term structure and the corresponding term structure obtained using the CBOE's method can deviate substantially in the data on S&P 500 index options. I now show that it takes only one illiquid or poorly observed maturity to generate this deviation. To this end, I simulate a dataset that mimics the one observed on September 23, 1998, which corresponds to the top left panel of [Figure 1](#) and coincides with the failure of Long-Term Capital Management (LTCM). This date is illustrative because it displays severe strike truncation for options at the one-year horizon, but is otherwise richly observed. The bottom left panel of [Figure 1](#) shows that a similar pattern repeats during the more recent financial crisis.

In direct correspondence with the option cross-section observed on this date, I simulate options 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. Notice the particularly deficient 9 observations with 12-month maturity. 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, I simulate 1000 option datasets from the SVJJ process in (5.1), and for each dataset, I compute the true VIX term structure (free of noise, discretization, and truncation error), the sieve’s estimate of the VIX term structure, and several workhorse benchmarks frequently used in applied work.<sup>15</sup>

Table 2 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. 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 ( $\kappa/S_0$ ) of 0.86 to 1.11. To make this range comparable to other maturities, Table 2 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 2 re-estimate at each maturity. Thus, the CBOE’s discrete approximation to the integral (2.1), 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.<sup>16</sup> 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 2 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-

---

<sup>15</sup>See Dew-Becker et al. (2014), Bollerslev et al. (2011), Carr and Wu (2009), and Jiang and Tian (2005) for applications using these benchmarks.

<sup>16</sup>See CBOE (2003) for details about its VIX construction.



Table 2: 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.

neutral valuation equation, it performs a theoretically supported projection of option prices beyond the 23% and 68% implied quantiles. Second, and importantly, the sieve regression (2.2) 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 2, 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.<sup>17</sup> It also provides practical guidance as to when the extrapolation can stop for the purpose of computing the option spanning integral (2.1): one should ideally continue until sufficient tail information is incorporated. For the implementations in this paper, I set the strike integration range for spanning portfolios to

<sup>17</sup>It can be shown from Lemma 3.2 that the sieve-implied risk-neutral CDF is given in closed-form by

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

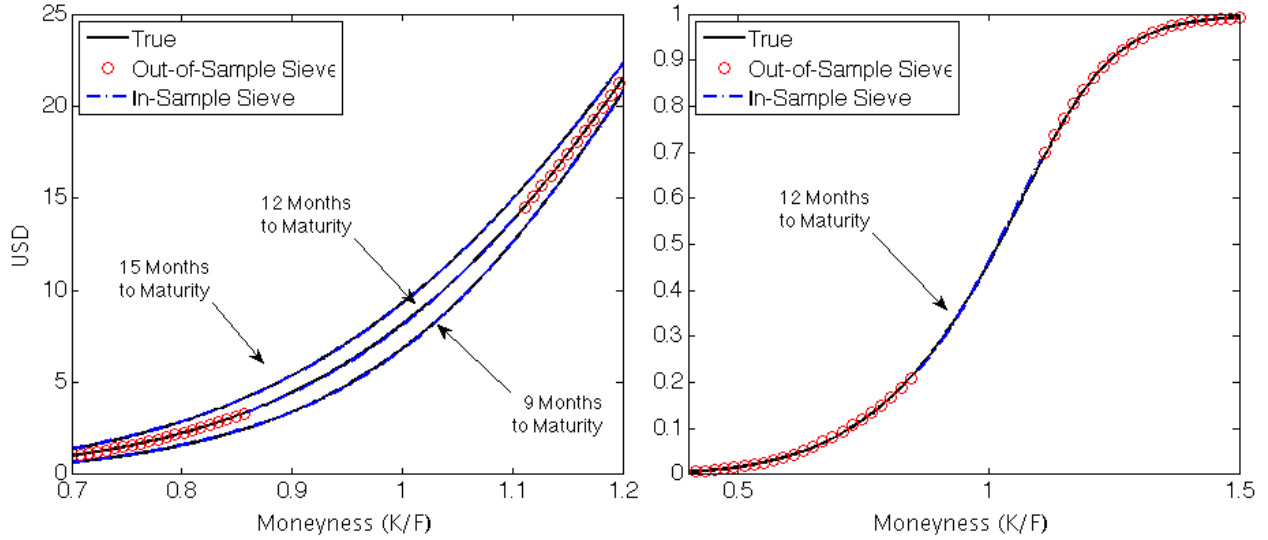


Figure 2: Simulated Out-of-Sample Sieve Price Projections and Implied Risk-Neutral CDF. An option surface with 1, 2, 3, 6, 9, 12, 15, 21 months-to-maturity and with respective number of observations 32, 20, 44, 31, 30, 9, 23, 27 is simulated. The sieve regression (3.10) is estimated, with expansion terms  $(K_y, K_\tau) = (8, 2)$  selected via the procedure in Remark 5.1. Strikes at the 12-month maturity were truncated according to the cutoffs in Table 2. 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.

consistently cover 0.5% to 99.5% of the risk-neutral distribution.

*Remark 5.1.* (Selection of Expansion Terms). 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 (3.10)  $(n - 1)$  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 the well-known Mallows (1973) criterion, which is asymptotically equivalent to leave-one-out cross-validation in certain settings, are natural candidates.<sup>18</sup> Since Coppejans and Gallant (2002) also show that minimizing the BIC tends to under-select the optimal expansion terms, I use the BIC to set a lower bound on  $K_y(n)$  and  $K_\tau(n)$ , and subsequently minimize

<sup>18</sup>Mallows (1973) criterion involves solving

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

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

Table 3: Monte Carlo Rejection Frequencies.

DGP	Maturity (months)								Expansions ( $K_y, K_\tau$ )
	1	2	3	6	9	12	15	21	
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 under the moneyness truncation ranges of [Table 2](#). 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 4](#), 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 [Mallows \(1973\)](#) criterion on a grid of  $(K_y(n), K_\tau(n))$  that is bounded below by the BIC choice.

## 5.2 Monte Carlo: Coverage

In addition to RMSE fit, I examine the finite-sample performance of the inference procedure in [Proposition 4](#) in a Monte Carlo experiment. Using the same maturity, strike range, observation, and bid-ask error configuration just described, I simulate 1000 datasets each from the SVJJ, SVJ, Heston, and Black-Scholes processes nested in [\(5.1\)](#). For each dataset, I estimate the sieve least squares regression [\(3.10\)](#), form estimated option prices  $\widehat{P}(\mathbf{Z}) \equiv P(\widehat{\beta}, \mathbf{Z})$  as in [Proposition 1](#) with arbitrarily dense strikes, and compute the integral in [\(2.1\)](#) at each observed maturity. I also compute the corresponding true VIX term structure, which is observed within the simulation study. The integration range is set to ensure that prices representing the 0.5% to 99.5% quantiles of the implied risk-neutral distribution are included. Finally, following [Proposition 4](#), 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 3](#) 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 data-driven procedure in [Remark 5.1](#), 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  $(0, 0)$ .

## 6 Application: The Term Structure of Variance Risk Premia

I combine the sieve framework with a novel set of expectation hypothesis and return predictability regressions to study the term structure of the variance risk premium, defined as follows. Let realized variance from month  $t$  to  $T = t + \tau$  be given by the annualized sum of squared daily returns

$$RV_{t,T} \equiv \frac{252}{n} \sum_{i=1}^n \left( \frac{sp500(t + i\Delta_n) - sp500(t + (i-1)\Delta_n)}{sp500(t + (i-1)\Delta_n)} \right)^2, \quad (6.1)$$

where  $n = \tau/\Delta_n$  is the number of trading days between  $t$  and  $T$ ,  $\Delta_n$  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  $RV_{t,T}$

$$VRP_t(t, T) \equiv \mathbb{E}_t^{\mathbb{P}}[RV_{t,T}] - \mathbb{E}_t^{\mathbb{Q}}[RV_{t,T}]. \quad (6.2)$$

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 (6.2) 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 (2.1):  $SVS_t(\tau) = \mathbb{E}_t^{\mathbb{Q}}[RV_{t,T}]$  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

$$RV_{t,T} = a(\tau) + b(\tau)\mathbb{E}_t^{\mathbb{Q}}[RV_{t,T}] + \varepsilon_{t+\tau}. \quad (6.3)$$

The idea is to test several hypotheses of this form and to relate them to well-established findings for the 1-month VRP. I therefore augment (6.3) with the concept of a *forward variance*, which takes advantage of the additive properties of  $RV_{t,T}$ : Note that from (6.1), one has for horizons  $\tau > 1$ ,  $RV_{t,T} = RV_{t,t+1} + RV_{t+1,T}$ , giving rise to a decomposition

$$\begin{aligned} VRP_t(t, T) &= \mathbb{E}_t^{\mathbb{P}}[RV_{t,t+1} + RV_{t+1,T}] - \mathbb{E}_t^{\mathbb{Q}}[RV_{t,t+1} + RV_{t+1,T}] \\ &= \mathbb{E}_t^{\mathbb{P}}[RV_{t,t+1}] - \mathbb{E}_t^{\mathbb{Q}}[RV_{t,t+1}] + \mathbb{E}_t^{\mathbb{P}}[RV_{t+1,T}] - \mathbb{E}_t^{\mathbb{Q}}[RV_{t+1,T}] \\ &\equiv VRP_t(t, t+1) + VRP_t(t+1, T). \end{aligned} \quad (6.4)$$

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.<sup>19</sup> Therefore, in order to relate findings on  $VRP_t(t, T)$  to our existing understanding of the 1-month variance risk premium, I also test hypotheses regarding the forward variance risk premium

$$RV_{t+1,T} = a(\tau) + b(\tau)\mathbb{E}_t^{\mathbb{Q}}[RV_{t+1,T}] + \varepsilon_{t+\tau}, \quad (6.5)$$

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 et al. (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  $SVS_t(\tau)$ , I 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  $VRP_t(t+1, T)$ . I therefore estimate both

$$\begin{aligned} Re_{t+h} &= \alpha_{h,\tau} + \beta_{h,\tau}VRP_t(t, T) + \varepsilon_{t+h} \quad \text{and} \\ Re_{t+h} &= \alpha_{h,\tau} + \beta_{h,\tau}VRP_t(t, t+1) + \gamma_{h,\tau}VRP_t(t+1, T) + \varepsilon_{t+h} \end{aligned} \quad (6.6)$$

for various forecasting horizons  $h$  and term structure maturities  $\tau$ , where  $Re_{t+h}$  denotes the  $h$ -month

---

<sup>19</sup>See, for example, Carr and Wu (2009), Bollerslev and Todorov (2011), Bollerslev et al. (2011), Drechsler and Yaron (2011), Bollerslev et al. (2013), Bekaert and Hoerova (2014).

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, I need a  $\mathbb{P}$ -measure forecast of realized variance,  $\mathbb{E}_t^{\mathbb{P}}[RV_{t,T}]$ , which I obtain from the standard heterogeneous AR model of [Corsi \(2009\)](#),

$$RV_{t,t+1} = b_0 + b_1 RV_t + b_2 \left( \frac{1}{6} \sum_{i=0}^5 RV_{t-i} \right) + b_3 \left( \frac{1}{24} \sum_{i=0}^{23} RV_{t-i} \right) + \varepsilon_{t+1}, \quad (6.7)$$

which effectively captures the long-memory dynamics of the  $RV_t$  process. To avoid look-ahead bias, (6.7) is estimated for each month  $t$  in my option sample 1996-2013, using monthly S&P 500 index RV (6.1) 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 (6.7) forward.

## 6.1 S&P 500 Index Option Data

To run the regressions in (6.3), (6.5), and (6.6), I 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 et al. \(2012\)](#), I 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, I 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, I solve the sieve least squares problem (3.10) and compute the portfolio integration in (2.1) for  $\tau = 1, 2, \dots, 24$  months-to-maturity. At each of these maturities, the integration limits in (2.1) 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, I 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 5](#). The unconditional correlation between the two series is 0.9976,

and the number of expansion terms selected via the data-driven criterion ([Remark 5.1](#)) was about  $(K_y, K_\tau) = (8, 3)$  on average.

Finally, note that inference on regressions of the form [\(6.3\)](#) and [\(6.5\)](#) 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  $b = 1$ . As a diagnostic, the bottom panel of [Figure 5](#) plots the ratio of average  $\widehat{SV\widehat{S}_t}(\tau)$  standard errors ([Proposition 4](#)) relative to their sample standard deviations, as well as  $\widehat{SV\widehat{S}_t}(\tau)$  autocorrelations for  $\tau = 1, \dots, 24$  months to maturity. The plots suggest that measurement error and serial correlation in the regressors are muted for maturities  $\tau = 1, \dots, 12$ , which motivates a focus on the first 12 months to maturity.

## 6.2 Results

**Expectation Hypothesis**      The results of the regressions [\(6.3\)](#) and [\(6.5\)](#) are surprising.  $p$ -values in the first row of [Table 4](#) 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  $\tau = 1, \dots, 12$ . 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.<sup>20</sup>

**Return Predictability**      The results of the expectation hypothesis tests are further corroborated in the return predictability regressions. [Table 5](#) reports  $t$ -statistics on the slope coefficient of the first regression in [\(6.6\)](#) using [Hodrick \(1992\)](#) standard errors.<sup>21</sup> The left-most column shows the same pattern of excess return predictability on horizons  $h = 2, \dots, 7$  found in [Bollerslev et al. \(2009\)](#) for one-month  $VRP_t(t, t + 1)$ . The pattern is noteworthy given that [Bollerslev et al. \(2009\)](#) use S&P 500 index excess returns, whereas I use CRSP value-weighted excess returns over a different sample

<sup>20</sup>For reference, the full regression output is provided in the Online Appendix.

<sup>21</sup>See the discussion in [Ang and Bekaert \(2007\)](#) in favor of using [Hodrick \(1992\)](#) standard errors in overlapping return predictability regressions.

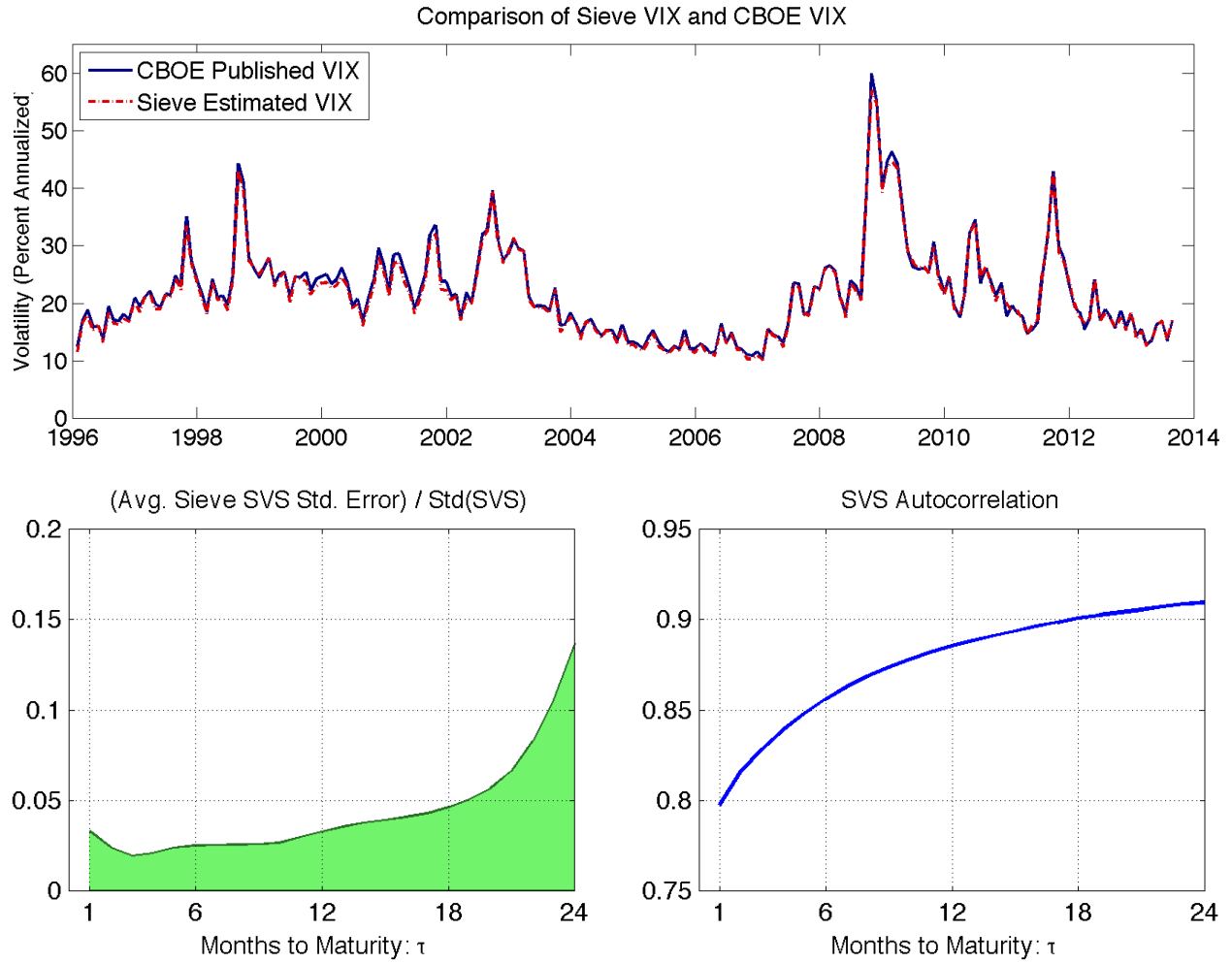


Figure 5: Pre-regression Diagnostics. The top panel plots a comparison of the sieve-estimated 30-day VIX and the CBOE published 30-day VIX. The bottom left panel plots the ratio of average sieve standard errors of  $\widehat{SVS}_t(\tau)$  to the time series standard deviation of  $\widehat{SVS}_t(\tau)$ , and the bottom right panel plots the sample first-order autocorrelation of  $\widehat{SVS}_t(\tau)$  for maturities  $\tau = 1, \dots, 24$ . 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 4](#).



Table 4:  $p$ -Values for Expectation Hypothesis Regressions.

	$\tau$											
	1	2	3	4	5	6	7	8	9	10	11	12
SVS	0.000	0.056	0.036	0.005	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Forward	-	0.035	0.066	0.107	0.133	0.186	0.236	0.296	0.359	0.408	0.440	0.459

*Notes:* The OLS regressions from (6.3) and (6.5) of realized variance on sieve synthetic variance swaps  $SVS_t(\tau)$  and forward variance swaps  $SVS_t(\tau) - SVS_t(1)$ , respectively, are estimated for each of the monthly horizons  $\tau = 1, \dots, 12$ .  $p$ -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 (6.3), and the second row shows the corresponding outcome for the forward variance swap (6.5). Newey and West (1987) standard errors for lag length 12 are used.

period as the left-hand side variable. Further out into the term structure, the remaining columns of Table 5 show that the strong predictive pattern is mirrored for  $VRP_t(t, T)$  for  $\tau = 2, \dots, 12$ . 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.

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

To dig deeper into the source of the return predictability, Table 6 presents the results of the second regression in (6.6), which decomposes  $VRP_t(t, T) = VRP_t(t, t+1) + VRP_t(t+1, T)$  into its short-run and forward VRP components. As with the Expectation Hypothesis tests, the return predictability regressions show that virtually all of the return predictability in  $VRP_t(t, T)$  is due to the one-month variance risk premium. This result strongly suggest that prices for protection against long-run variance risk are disproportionately driven by the protection afforded over the

Table 5: Excess Return Predictability.

$h \setminus \tau$	$t(\hat{\beta}_{h,\tau})$											
	1	2	3	4	5	6	7	8	9	10	11	12
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
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
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
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
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
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
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
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
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
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
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
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

*Notes:* This table reports [Hodrick \(1992\)](#)  $t$ -statistics from regressions of monthly CRSP value-weighted excess returns on the lagged term structure of variance risk premia (6.6), i.e.  $Re_{t+h} = \alpha_{h,\tau} + \beta_{h,\tau} VRP_t(t, T) + \varepsilon_{t+h}$ , where  $T = t + \tau$ .

leading month.

Taken together, these findings provide conditional evidence that the short-run variance risk premium is a significant driver of the compensation that investors receive when buying long-run variance swaps, which is consistent with the recent work of [Dew-Becker et al. \(2014\)](#), who find (unconditionally) high risk-adjusted returns for selling short-run variance swaps. A reconciliation of these results with asset pricing theory is still open and makes for interesting future work.

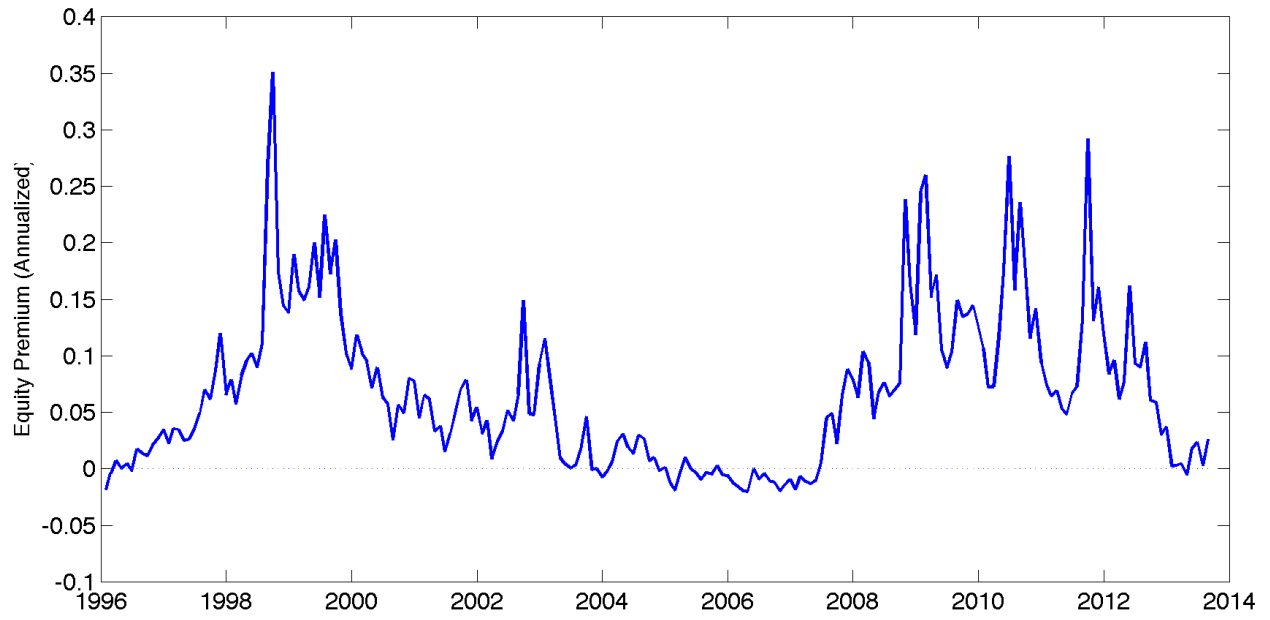


Figure 6: The Equity Risk Premium  $\hat{E}_t[Re_{t+6}] = \hat{\alpha}_{6,12} + \hat{\beta}_{6,12}VRP_t(t, t + 12)$ .

Table 6: Excess Return Predictability of Forward Variance Risk Premia.

$h \setminus \tau$	$t(\hat{\beta}_{h,\tau})$											
	1	2	3	4	5	6	7	8	9	10	11	12
1	-1.28	-1.40	-1.49	-1.55	-1.56	-1.54	-1.51	-1.48	-1.47	-1.45	-1.44	
2	-2.02	-2.12	-2.22	-2.28	-2.29	-2.27	-2.24	-2.22	-2.20	-2.19	-2.18	
3	-2.15	-2.23	-2.34	-2.43	-2.46	-2.45	-2.44	-2.42	-2.41	-2.40	-2.40	
4	-1.94	-2.04	-2.17	-2.26	-2.31	-2.30	-2.29	-2.28	-2.28	-2.27	-2.26	
5	-1.97	-2.10	-2.22	-2.30	-2.33	-2.32	-2.30	-2.29	-2.27	-2.26	-2.25	
6	-2.08	-2.17	-2.27	-2.33	-2.36	-2.35	-2.34	-2.33	-2.32	-2.31	-2.30	
7	-2.03	-2.11	-2.19	-2.25	-2.29	-2.29	-2.28	-2.27	-2.26	-2.25	-2.24	
8	-1.78	-1.85	-1.92	-1.96	-1.99	-1.99	-1.98	-1.98	-1.97	-1.96	-1.95	
9	-1.59	-1.65	-1.70	-1.74	-1.76	-1.76	-1.76	-1.76	-1.76	-1.76	-1.76	
10	-1.40	-1.46	-1.52	-1.56	-1.58	-1.58	-1.58	-1.58	-1.58	-1.58	-1.58	
11	-1.28	-1.35	-1.41	-1.45	-1.46	-1.47	-1.47	-1.47	-1.46	-1.46	-1.46	
12	-1.18	-1.25	-1.30	-1.32	-1.34	-1.34	-1.34	-1.34	-1.34	-1.33	-1.33	

$h \setminus \tau$	$t(\hat{\gamma}_{h,\tau})$											
	1	2	3	4	5	6	7	8	9	10	11	12
1	-0.06	-0.33	-0.52	-0.62	-0.67	-0.64	-0.61	-0.58	-0.55	-0.53	-0.52	
2	-0.27	-0.63	-0.88	-0.98	-0.98	-0.91	-0.85	-0.80	-0.77	-0.74	-0.72	
3	0.14	-0.33	-0.73	-0.95	-1.03	-1.01	-0.97	-0.94	-0.92	-0.90	-0.88	
4	-0.26	-0.70	-1.04	-1.24	-1.32	-1.31	-1.28	-1.25	-1.23	-1.21	-1.19	
5	-1.03	-1.40	-1.64	-1.75	-1.78	-1.73	-1.67	-1.63	-1.59	-1.55	-1.52	
6	-1.01	-1.31	-1.52	-1.63	-1.66	-1.63	-1.60	-1.57	-1.54	-1.51	-1.48	
7	-1.06	-1.32	-1.51	-1.61	-1.65	-1.63	-1.60	-1.56	-1.54	-1.50	-1.47	
8	-1.10	-1.30	-1.43	-1.50	-1.52	-1.50	-1.47	-1.44	-1.41	-1.38	-1.36	
9	-1.06	-1.21	-1.30	-1.34	-1.36	-1.34	-1.32	-1.31	-1.30	-1.28	-1.27	
10	-1.15	-1.27	-1.34	-1.36	-1.36	-1.34	-1.32	-1.30	-1.28	-1.27	-1.25	
11	-1.19	-1.29	-1.34	-1.35	-1.34	-1.31	-1.29	-1.27	-1.25	-1.23	-1.21	
12	-1.01	-1.12	-1.15	-1.16	-1.15	-1.12	-1.10	-1.08	-1.06	-1.04	-1.03	

*Notes:* This table reports [Hodrick \(1992\)](#)  $t$ -statistics from regressions of monthly CRSP value-weighted excess returns on the lagged components of the term structure of variance risk premia (6.6), i.e.  $Re_{t+h} = \alpha_{h,\tau} + \beta_{h,\tau}VRP_t(t, t+1) + \gamma_{h,\tau}VRP_t(t+1, T) + \varepsilon_{t+h}$ , where  $T = t + \tau$ .

## 7 Conclusion

This paper presented a new estimation and inference framework for the term structures of model-free option spanning portfolios, including the VIX and related measures. By constructing sieve approximations to the term structure of state-price densities, the framework inherits the structural shape information contained in the risk-neutral valuation equation at all maturities. The shape information is used to inform nonparametric estimates of the option price surface at long maturities, where options are less liquid and therefore increasingly subject to measurement error, strike truncation, and maturity sparseness. To quantify the intuition that long-run option spanning portfolios are less precisely estimated than their liquid short-run counterparts, the paper also develops an asymptotic distribution theory for option spanning portfolios.

The framework is tested both empirically and in Monte Carlo simulations. The simulations show that the sieve-estimated VIX term structure considerably outperforms existing benchmarks in situations where a subset of maturities display significant strike truncation. Furthermore, the distribution theory for spanning portfolios is used to construct confidence intervals for the VIX term structure, which are shown to display size control in empirically realistic finite samples. In an empirical application, the sieve was used to estimate the term structure of the synthetic variance swap. Expectation hypothesis and return predictability regressions involving forward variance risk premia found that the term structure of the variance risk premium is dominated by compensation for bearing short-run variance risk.

In future work, the framework could be used to explore option-implied term structures in the cross-section. To the extent that options written on individual stocks, industry ETFs, and international stock indices 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.

## References

- Adams RA, Fournier JJ. 2003. *Sobolev spaces*, volume 140. Access Online via Elsevier.
- Aït-Sahalia Y, Duarte J. 2003. Nonparametric option pricing under shape restrictions. *Journal of Econometrics* **116**: 9–47.
- Aït-Sahalia Y, Karaman M, Mancini L. 2014. The term structure of variance swaps, risk premia, and the expectation hypothesis. *Available at SSRN 2136820* .
- Aït-Sahalia Y, Lo AW. 1998. Nonparametric estimation of state-price densities implicit in financial asset prices. *The Journal of Finance* **53**: 499–547.
- Andersen TG, Fusari N, Todorov V. 2012. Parametric inference and dynamic state recovery from option panels. Technical report, National Bureau of Economic Research.
- Ang A, Bekaert G. 2007. Stock return predictability: Is it there? *Review of Financial Studies* **20**: 651–707.
- Bakshi G, Cao C, Chen Z. 1997. Empirical performance of alternative option pricing models. *The Journal of Finance* **52**: 2003–2049.
- Bakshi G, Kapadia N, Madan D. 2003. Stock return characteristics, skew laws, and the differential pricing of individual equity options. *Review of Financial Studies* **16**: 101–143.
- Bakshi G, Madan D. 2000. Spanning and derivative-security valuation. *Journal of Financial Economics* **55**: 205–238.
- Bakshi G, Madan D. 2006. A theory of volatility spreads. *Management Science* **52**: 1945–1956.
- Bakshi G, Panayotov G, Skoulakis G. 2011. Improving the predictability of real economic activity and asset returns with forward variances inferred from option portfolios. *Journal of Financial Economics* **100**: 475–495.
- Banz RW, Miller MH. 1978. Prices for state-contingent claims: Some estimates and applications. *Journal of Business* : 653–672.
- Bates D. 1996. Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Review of financial studies* **9**: 69–107.
- Bates D. 2000. Post-'87 crash fears in the s&p 500 futures option market. *Journal of Econometrics* **94**: 181–238.
- Bekaert G, Hoerova M. 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, Scholes M. 1973. The pricing of options and corporate liabilities. *The Journal of Political Economy* : 637–654.
- Bollerslev T, Gibson M, Zhou H. 2011. Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *Journal of Econometrics* **160**: 235–245.

- Bollerslev T, Osterrieder D, Sizova N, Tauchen G. 2013. Risk and return: Long-run relations, fractional cointegration, and return predictability. *Journal of Financial Economics* **108**: 409–424.
- Bollerslev T, Tauchen G, Zhou H. 2009. Expected stock returns and variance risk premia. *Review of Financial Studies* **22**: 4463–4492.
- Bollerslev T, Todorov V. 2011. Tails, fears, and risk premia. *The Journal of Finance* **66**: 2165–2211.
- Bondarenko O. 2003. Estimation of risk-neutral densities using positive convolution approximation. *Journal of Econometrics* **116**: 85–112.
- Breedon DT, Litzenberger RH. 1978. Prices of state-contingent claims implicit in option prices. *Journal of business* : 621–651.
- Britten-Jones M, Neuberger A. 2000. Option prices, implied price processes, and stochastic volatility. *The Journal of Finance* **55**: 839–866.
- Carr P, Wu L. 2009. Variance risk premiums. *Review of Financial Studies* **22**: 1311–1341.
- CBOE. 2003. The cboe volatility index - vix white paper. URL: <http://www.cboe.com/micro/vix/vixwhite.pdf> .
- Chang BY, Christoffersen P, Jacobs K. 2013. Market skewness risk and the cross section of stock returns. *Journal of Financial Economics* **107**: 46–68.
- Chen X. 2007. Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics* **6**: 5549–5632.
- Chen X, Liao Z, Sun Y. 2014. Sieve inference on possibly misspecified semi-nonparametric time series models. *Journal of Econometrics* **178**: 639–658.
- Chen X, Shen X. 1998. Sieve extremum estimates for weakly dependent data. *Econometrica* : 289–314.
- Christoffersen P, Fournier M, Jacobs K. 2013. The factor structure in equity option prices .
- Conrad J, Dittmar RF, Ghysels E. 2013. Ex ante skewness and expected stock returns. *The Journal of Finance* **68**: 85–124.
- Coppejans M, Gallant AR. 2002. Cross-validated snp density estimates. *Journal of Econometrics* **110**: 27–65.
- Corsi F. 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* **7**: 174–196.
- Dew-Becker I, Giglio S, Le A, Rodriguez M. 2014. The price of variance risk. Available at SSRN 2471079 .
- Drechsler I, Yaron A. 2011. What’s vol got to do with it. *Review of Financial Studies* **24**: 1–45.
- Driessen J, Maenhout PJ, Vilkov G. 2009. The price of correlation risk: Evidence from equity options. *The Journal of Finance* **64**: 1377–1406.
- Duffie D. 2001. *Dynamic asset pricing theory*. Princeton Univ Pr.

- Duffie D, Pan J, Singleton K. 2000. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* **68**: 1343–1376.
- 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 AR, Nychka DW. 1987. Semi-nonparametric maximum likelihood estimation. *Econometrica: Journal of the Econometric Society* : 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**: 327–343.
- Hodrick RJ. 1992. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial studies* **5**: 357–386.
- Jarrow R, Rudd A. 1982. Approximate option valuation for arbitrary stochastic processes. *Journal of financial Economics* **10**: 347–369.
- Jiang G, Tian Y. 2005. The model-free implied volatility and its information content. *Review of Financial Studies* **18**: 1305–1342.
- Kristensen D, Mele A. 2011. Adding and subtracting black-scholes: a new approach to approximating derivative prices in continuous-time models. *Journal of financial economics* **102**: 390–415.
- León Á, Mencía J, Sentana E. 2009. Parametric properties of semi-nonparametric distributions, with applications to option valuation. *Journal of Business & Economic Statistics* **27**: 176–192.
- Leslie RA. 1991. How not to repeatedly differentiate a reciprocal. *American Mathematical Monthly* : 732–735.
- Li Q, Racine JS. 2007. *Nonparametric econometrics: Theory and practice*. Princeton University Press.
- Mallows CL. 1973. Some comments on c p. *Technometrics* **15**: 661–675.
- Newey W, West K. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica: Journal of the Econometric Society* : 703–708.
- Newey WK. 1991. Uniform convergence in probability and stochastic equicontinuity. *Econometrica* **59**: 1161–1167.
- Nickl R, Pötscher BM. 2007. Bracketing metric entropy rates and empirical central limit theorems for function classes of besov-and sobolev-type. *Journal of Theoretical Probability* **20**: 177–199.
- Polkovnichenko V, Zhao F. 2013. Probability weighting functions implied in options prices. *Journal of Financial Economics* **107**: 580–609.
- Rompolis LS, Tzavalis E. 2008. Recovering risk neutral densities from option prices: A new approach. *Journal of Financial and Quantitative Analysis* **43**: 1037–1053.
- Shen X. 1997. On methods of sieves and penalization. *The Annals of Statistics* **25**: 2555–2591.
- van Binsbergen J, Brandt M, Koijen R. 2012. On the timing and pricing of dividends. *The American Economic Review* **102**: 1596–1618.



- Van De Geer S. 2000. *Empirical processes in M-estimation*, volume 45. Cambridge university press Cambridge.
- Van Der Vaart AW, Wellner JA. 1996. *Weak Convergence*. Springer.
- Xiu D. 2011. Dissecting and deciphering european option prices using closed-form series expansion. *Chicago Booth Research Paper* .
- Yatchew A, Härdle W. 2006. Nonparametric state price density estimation using constrained least squares and the bootstrap. *Journal of Econometrics* **133**: 579–599.

# A Definitions and Preliminary Results

## A.1 Defining the Sobolev Sieve Spaces

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

### 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  $1 \leq p < \infty$ , define the Sobolev norm of  $f$  with respect to the nonnegative weight function  $\zeta(\mathbf{u})$  by

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

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

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

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

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

where  $1 \leq p \leq \infty$ .

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  $\|h_0\|_{m_0+m,2,\zeta_0} \leq \mathcal{B}_0$ , 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  $\|h\|_{m_0+m,2,\zeta_0} \leq \mathcal{B}_0$  and  $\varepsilon > \varepsilon_0$ , where

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

Let

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

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  $\|P_K(\mathbf{u})\phi(\mathbf{u})^{1/2}\|_{m_0+m,2,\zeta_0} \leq \mathcal{B}_0$  and  $\varepsilon > \varepsilon_0$  and  $\beta'\beta = 1$ , where  $\beta$  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, I 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*

$$h_0(\mathbf{u}) = \phi(y) \cdot h_x(\mathbf{x}), \tag{A.1}$$

*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 (A.1) 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  $\phi(y) \cdot \phi(x_1) \cdots \phi(x_{d_x})$ . In any case, one can argue as in Gallant and Nychka (1987) that the value of  $\varepsilon$  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  $\Lambda : \mathcal{F}^{Y,X} \rightarrow \mathcal{F}$  ([Lemma A.9](#) below). Note that the densities in  $\mathcal{F}$  are related to the return distribution via the change of variables formula in [\(3.3\)](#).

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

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

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

**Definition A.7.** (Hölder Spaces). Define

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

## 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  $\|f^X\|_{m,1} \leq M$ .*
- (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  $\|P\|_{m,1} \leq M \|f\|_{m,1}$ . 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  $M_U$ ,  $\sup_{g \in U} \|P(g, \mathbf{Z})\|_{m,1} \leq M_U$ .*
- (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  $f^{Y,X} \leq C_3(1 + \mathbf{u}'\mathbf{u})^{-\delta}$  for  $\delta \in (d_u/2, \delta_0)$ . First, from [Definition A.2](#),  $\|h\|_{m_0+m,2,\zeta_0} \leq$

$\mathcal{B}_0$  implies

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

Since  $\|h_0\|_{m_0+m, 2, \zeta_0} \leq \mathcal{B}_0$  as well, one has  $f^{Y, X} \leq C_3 (1 + \mathbf{u}'\mathbf{u})^{-\delta}$ .

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  $\alpha = (0, \alpha_1, \dots, \alpha_{d_x})$  denote a multi-index over  $\mathbf{x}$ . By Step 1, dominated convergence, triangle inequality, Hölder's inequality, and Gallant-Nychka Lemma A.1(c),

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

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

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

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

$$\begin{aligned}
\|f\|_{m,1} &= \sum_{|\lambda|\leq m} \int_{\mathcal{X}} \int_{\mathcal{Y}} |D^\lambda f(y|\mathbf{x})| dy d\mathbf{x} \leq \sum_{|\lambda|\leq m} \sum_{\beta\leq\lambda} \begin{bmatrix} \lambda \\ \beta \end{bmatrix} \int_{\mathcal{X}} \int_{\mathcal{Y}} \left| D^\beta f(y, \mathbf{x}) D^{\lambda-\beta} \frac{1}{f(\mathbf{x})} \right| dy d\mathbf{x} \\
&= \sum_{|\lambda|\leq m} \sum_{\beta\leq\lambda} \begin{bmatrix} \lambda \\ \beta \end{bmatrix} \int_{\mathcal{X}} \int_{\mathcal{Y}} \left| D^\beta \{f(y, \mathbf{x})\} \zeta(y, \mathbf{x}) D^{\lambda-\beta} \left\{ \frac{1}{f(\mathbf{x})} \right\} \zeta(y, \mathbf{x})^{-1} \right| dy d\mathbf{x} \\
&\leq C_1 \sum_{|\lambda|\leq m} \sum_{\beta\leq\lambda} \begin{bmatrix} \lambda \\ \beta \end{bmatrix} \left\{ \sup_{y, \mathbf{x}} |D^\beta f(y, \mathbf{x})| \zeta(y, \mathbf{x}) \right\} \int_{\mathcal{X}} D^{\lambda-\beta} \left\{ \frac{1}{f(\mathbf{x})} \right\} \int_{\mathcal{Y}} (1 + y^2 + \mathbf{x}'\mathbf{x})^{-\delta} dy d\mathbf{x} \\
&\leq C_2 \left\{ \max_{|\lambda|\leq m} \sup_{y, \mathbf{x}} |D^\lambda f(y, \mathbf{x})| \zeta(y, \mathbf{x}) \right\} \|f^X\|_{m,1} = C_3 \|f^{Y,X}\|_{m,\infty,\zeta} < \infty,
\end{aligned} \tag{A.3}$$

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

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

$$\begin{aligned}
\|P\|_{m,1} &= \sum_{|\lambda|\leq m} \int_{\mathcal{Z}} |D^\lambda P(\mathbf{Z})| d\mathbf{Z} \\
&\leq \sum_{|\lambda|\leq m} \sum_{|\beta|\leq\lambda} \begin{bmatrix} \lambda \\ \beta \end{bmatrix} \int_{\mathcal{Z}-\mathcal{X}} \int_{\mathcal{X}} \int_{\mathcal{Y}} |D^\beta [\psi(y, \mathbf{Z})] D^{\lambda-\beta} f(y|\mathbf{X})| dy d\mathbf{x} d(\mathbf{z} - \mathbf{x}) \\
&\leq C_1 \sum_{|\lambda|\leq m} \sum_{|\beta|\leq\lambda} \begin{bmatrix} \lambda \\ \beta \end{bmatrix} \int_{\mathcal{Z}-\mathcal{X}} \int_{\mathcal{X}} \int_{\mathcal{Y}} |D^{\lambda-\beta} f(y|\mathbf{X})| dy d\mathbf{x} d(\mathbf{z} - \mathbf{x}) \\
&\leq C_2 \|f\|_{m,1} < \infty.
\end{aligned} \tag{A.4}$$

- (iv) Follows directly from (A.4) 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  $\Lambda : \mathcal{F}^{Y,X} \rightarrow \mathcal{F}$  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  $f_0^{Y,X}, f_j^{Y,X} \in \mathcal{F}^{Y,X}$ .

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

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

where the last inequality follows from derivations analogous to (A.3). The result follows.  $\square$

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  $K \geq 0$ .
- (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  $K \geq 0$  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  $\|h_{K,n}\|_{m_0+m, 2, \zeta_0} \leq \mathcal{B}_0$ . 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  $\lim_{j \rightarrow \infty} \|h_{K,n_j} - h\|_{m, \infty, \zeta^{1/2}} = 0$ . Then by Gallant and Nychka's Lemma A.3, one has  $\|h^2\|_{m, \infty, \zeta} < \infty$  and  $\lim_{j \rightarrow \infty} \|h_{K,n_j}^2 - h^2\|_{m, \infty, \zeta} = 0$ , whence  $\lim_{j \rightarrow \infty} \|f_{K,n_j} - f\|_{m, \infty, \zeta} = 0$ . Thus  $\mathcal{F}_K^{X,Y}$  is compact in the topology generated by  $\|\cdot\|_{m, \infty, \zeta}$ . Finally, because  $\Lambda$  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  $\lim_{K \rightarrow \infty} \|h_K - h\|_{m_0+m, 2, \zeta} = 0$ , and by their Lemmas A.1-A.3, this implies  $\lim_{K \rightarrow \infty} \|h_K^2 - h^2\|_{m, \infty, \zeta} = 0$ . One therefore has  $\lim_{K \rightarrow \infty} \|f_K^{X,Y} - f\|_{m, \infty, \zeta} = 0$ , 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  $\Lambda$  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  $\|\cdot\|_{m,1}$ .  $\square$

**Proof of Lemma 3.2.** The task is to show that conditional sieve densities have a representation  $f_K(y|\tau) = \sum_{k=0}^{2K_y} \gamma_k(B, \tau) H_k(y) \phi(y)$  that is required to get closed-form option prices. Let  $\alpha(B, \tau) = (\sum_{j=0}^{K_\tau} \beta_{0j} H_j(\tau), \dots, \sum_{j=0}^{K_\tau} \beta_{K_y j} H_j(\tau))'$ . 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,

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

where the last equality and the definition of  $A_k$  follow by applying Proposition 1 of León et al. (2009). The result follows.  $\square$

## B Derivation of Main Results

### B.1 Closed-Form Option Pricing

I begin by proving the closed-form pricing expression stated in Proposition 1.

**Proof of Proposition 1.** The proof extends Proposition 9 of León et al. (2009) to allow for conditioning on  $\tau$ . The plug-in estimator of the population option price in equation (3.4), is given by

$$\begin{aligned} P(f_K, \mathbf{Z}) &= e^{-r\tau} \int_{-\infty}^{d(\mathbf{Z})} \left( \kappa - S e^{\mu(\mathbf{Z}) + \sigma(\mathbf{Z})Y} \right) f_K(Y|\tau) dY \\ &= \kappa e^{-r\tau} \int_{-\infty}^{d(\mathbf{Z})} f_K(Y|\tau) dY - S e^{-r\tau + \mu(\mathbf{Z})} \int_{-\infty}^{d(\mathbf{Z})} e^{\sigma(\mathbf{Z})Y} f_K(Y|\tau) dY. \end{aligned} \quad (\text{B.1})$$

Using Lemma 3.2, the integral in the first term becomes

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

where the last equality follows from integration properties of the Hermite functions. The integral in the second term on the right-hand side (RHS) of equation (B.1) can further be simplified by integrating by parts. Let

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

For  $k = 0$ ,

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



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

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

Thus,

$$\begin{aligned}
\int_{-\infty}^{d(\mathbf{Z})} e^{\sigma(\mathbf{Z})Y} f_K(Y|\tau) dY &= \int_{-\infty}^{d(\mathbf{Z})} e^{\sigma(\mathbf{Z})Y} \left[ \sum_{k=0}^{2K_y} \gamma_k(B, \tau) H_k(Y) \phi(Y) \right] dY \\
&= \sum_{k=0}^{2K_y} \gamma_k(B, \tau) \int_{-\infty}^{d(\mathbf{Z})} e^{\sigma(\mathbf{Z})Y} H_k(Y) \phi(Y) dY = \sum_{k=0}^{2K_y} \gamma_k(B, \tau) I_k^*(d(\mathbf{Z})) \\
&= e^{\sigma(\mathbf{Z})^2/2} \Phi(d(\mathbf{Z}) - \sigma(\mathbf{Z})) + \sum_{k=1}^{2K_y} \gamma_k(B, \tau) I_k^*(d(\mathbf{Z})), \tag{B.3}
\end{aligned}$$

where  $\gamma_0(B, \mathbf{Z}) = 1$ . Plugging equations (B.2) and (B.3) into (B.1) obtains the desired result. The proof for call options is analogous and is therefore omitted.  $\square$

## B.2 Asymptotic Theory

I establish the consistency, rate of convergence, and asymptotic distribution results stated in [Section 4](#).

### 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 2.** *(Consistency, Restated) 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.  $\ell$  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)  $\mathcal{F}_K \subseteq \mathcal{F}_{K+1} \subseteq \dots \subseteq \mathcal{F}$ , for all  $K \geq 1$ .

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

**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  $K \geq 1$ ,  $\sup_{f \in \mathcal{F}_K} |L_n(f) - L(f)| = 0$ .

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

**Lemma B.1.** *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.  $\square$

**Lemma B.2.**  $P(f_1, \mathbf{Z}) = P(f_2, \mathbf{Z})$  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

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

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

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

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

by Lemma B.1. Thus, as  $d(f_n, f_0) \rightarrow 0$ , one has  $L(f_0) - L(f) = |L(f_0) - L(f)| \rightarrow 0$ . 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,

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

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  $\Omega_F$  with  $|P_i(\omega)| < \infty$  for all  $\omega \in \Omega_F$ ,<sup>22</sup> 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  $\Omega_F$ .

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

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

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  $n \rightarrow \infty$ , where  $L_n(f) = \frac{1}{n} \sum_{i=1}^n -\frac{1}{2} [P_i - P(f, \mathbf{Z}_i)]^2 W_i$ . 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  $\mathcal{F}_K$ .

<sup>22</sup>To see this, note by Markov's inequality that  $\mathbb{P}(|P_i| > M) \leq \text{Var}(P_i)/M^2$ . Applying the Borel-Cantelli Lemma then shows that  $P_i$  is almost surely finite. See [Billingsley \(1995\)](#).

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  $M = \max\{M_1, \dots, M_N\}$  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  $f_1, f_2 \in \mathcal{F}_K$ ,

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

This Lipschitz condition, the compactness of  $\mathcal{F}_K$ , and the pointwise convergence of  $L_n(f)$  to  $L(f)$  mean that the conditions for Corollary 2.2 in Newey (1991) are met, so that  $\sup_{f \in \mathcal{F}_K} |L_n(f) - L(f)| \xrightarrow{P} 0$ , 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.1 gives  $\|P(\hat{f}_n, \mathbf{Z}) - P(f_0, \mathbf{Z})\|_2 \xrightarrow{P} 0$ .  $\square$

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.2). 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 3.** *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,

$$\sup_{P \in \mathcal{B}_\varepsilon(P_0)} \text{Var}(\ell(P, \Xi_i) - \ell(P_0, \Xi_i)) \leq C_1 \varepsilon^2.$$

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

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

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

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

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  $\|P - P_0\|_\infty \leq \|P - P_0\|_2^{2/(2+d_z)}$ . 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

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

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

To show Condition 3.8, note that

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

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 2. Thus Lemma 2 in Chen and Shen (1998) and another appeal to the Sobolev Embedding Theorem imply that

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

for  $U(\Xi_i) = |e_i| + \text{const.}$ . 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  $\delta_n$  satisfying

$$\delta_n = \inf \left\{ \delta \in (0, 1) : \frac{1}{\sqrt{n}\delta^2} \int_{b\delta^2}^\delta \sqrt{H_{[1]}(w, \mathcal{G}_n, \|\cdot\|_2)} dw \right\},$$

for the bracketing entropy  $H_{[1]}(w, \mathcal{G}_n, \|\cdot\|_2)$ , where

$$\mathcal{G}_n = \{ \ell(P, \Xi_i) - \ell(P_0, \Xi_i) : \|P - P_0\|_2 \leq \delta, P \in \mathcal{P}_{K_n} \}. \quad (\text{B.4})$$

Consider the following chain of inequalities

$$\begin{aligned} |\ell(P, \Xi_i) - \ell(P_0, \Xi_i)| &= |[P(\mathbf{Z}_i) - P_0(\mathbf{Z}_i)] \left[ e_i + \frac{1}{2} [P_0(\mathbf{Z}_i) - P(\mathbf{Z}_i)] \right]| \\ &\leq M_1 \|f - f_0\|_{m,1} U(\Xi_i) \\ &\leq M_2 U(\Xi_i) \left\| f^{Y,X} - f_0^{Y,X} \right\|_{m,\infty,\zeta_0} \quad \text{by Lemma A.9} \\ &= M_2 U(\Xi_i) \left\| (h^{Y,X})^2 - (h_0^{Y,X})^2 \right\|_{m,\infty,\zeta_0} \quad \text{by Def. A.2} \\ &\leq M_3 U(\Xi_i) \left\| h^{Y,X} - h_0^{Y,X} \right\|_{m_0+m,2,\zeta_0} \quad (\text{B.5}) \end{aligned}$$

To see the last inequality, observe that

$$\begin{aligned}
\left\| (h^{Y,X})^2 - (h_0^{Y,X})^2 \right\|_{m,\infty,\zeta_0} &\leq C \left\| h^{Y,X} + h_0^{Y,X} \right\|_{m,\infty,\zeta_0^{1/2}} \left\| h^{Y,X} - h_0^{Y,X} \right\|_{m,\infty,\zeta_0^{1/2}} \\
&\leq C_1 \left\| \zeta_0^{1/2} (h^{Y,X} + h_0^{Y,X}) \right\|_{m,\infty} C_2 \left\| \zeta_0^{1/2} (h^{Y,X} - h_0^{Y,X}) \right\|_{m,\infty} \\
&\leq C_3 \left\| h^{Y,X} + h_0^{Y,X} \right\|_{m_0+m,2,\zeta_0} C_4 \left\| h^{Y,X} - h_0^{Y,X} \right\|_{m_0+m,2,\zeta_0} \\
&\leq C_3(2\mathcal{B}_0)C_4 \left\| h^{Y,X} - h_0^{Y,X} \right\|_{m_0+m,2,\zeta_0}.
\end{aligned} \tag{B.6}$$

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

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

where the RHS is the covering number of a Sobolev ball with dimension  $K_n \equiv [K_y(n) + 1][K_{x,1}(n) + 1] \dots [K_{x,d_x}(n) + 1]$ . By Lemma 2.5 in [Van De Geer \(2000\)](#), we can further bound the RHS, giving

$$N_{[\cdot]}(w, \mathcal{G}_n, \|\cdot\|_2) \leq N\left(\frac{w}{2CM_3}, \mathcal{H}_n, \|\cdot\|_{m_0+m,2,\zeta_0}\right) \leq \left(1 + \frac{8\mathcal{B}_0CM_3}{w}\right)^{K_n}.$$

Therefore,

$$\frac{1}{\sqrt{n}\delta_n^2} \int_{b\delta_n^2}^{\delta_n} \sqrt{H_{[\cdot]}(w, \mathcal{G}_n, \|\cdot\|_2)} dw \leq \frac{1}{\sqrt{n}\delta_n^2} \int_{b\delta_n^2}^{\delta_n} \sqrt{K_n \log\left(1 + \frac{8\mathcal{B}_0CM_3}{w}\right)} dw \leq C \frac{1}{\sqrt{n}\delta_n^2} \sqrt{K_n} \delta_n,$$

which is less than or equal to a constant for the choice  $\delta_n \asymp \sqrt{K_n/n}$ . Put  $K_y(n) \asymp K_{x,1}(n) \asymp \dots \asymp K_{x,d_x}(n) \asymp n^{1/(2(m_0+m)+d_u)}$ , 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

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

where the inequality follows from the ones in [\(B.5\)](#). 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  $\hat{R}_{K_n} = -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \ell(\hat{\beta}_n, \Xi_i)}{\partial \beta \partial \beta'}$ .

**Assumption B.5.**

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

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

(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 (2.4),  $g(\cdot)$  has bounded first and second derivatives over the domain of interest.

**Proposition 4.** Under Assumptions A.1 and B.1–B.5,

$$\sqrt{n} \widehat{V}_n^{-1/2} [\Gamma(\widehat{P}_n) - \Gamma(P_0)] \xrightarrow{d} N(0, 1) \quad (\text{B.7})$$

where

$$\widehat{V}_n = \widehat{G}'_{K_n} \widehat{R}_{K_n}^{-1} \widehat{\Sigma}_{K_n} \widehat{R}_{K_n}^{-1} \widehat{G}_{K_n} \quad (\text{B.8})$$

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.* The aim is to connect the sieve asymptotic theory with simple non-linear least squares implementations by using techniques from Chen et al. (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 et al. (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 2 and Proposition 3, one can confine the analysis to the local setting of Chen et al. (2014). That is, the convergence rate  $\varepsilon_n$  in Proposition 3 implies that  $\widehat{P} \in \mathcal{B}_n$  with probability approaching one, where

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

Let  $\mathcal{V} \equiv \text{clsp}(\mathcal{B}_0) - \{P_0\}$  and  $\mathcal{V}_n \equiv \text{clsp}(\mathcal{B}_n) - \{P_{0,n}\}$ , where  $\text{clsp}(\cdot)$  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 (2.4) 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

$$\frac{\partial \Gamma(P_0)}{\partial P}[v_n^*] = \|v_n^*\|^2 = \sup_{v \in \mathcal{V}_n, v \neq 0} \left| \frac{\partial \Gamma(P_0)}{\partial P}[v] \right|^2 / \|v\|^2. \quad (\text{B.9})$$

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

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

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

In this notation, the problem in (B.9) 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,

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

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  $v_n^*$ )

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

is required, with so-called standard deviation norm

$$\begin{aligned} \|v_n^*\|_{sd}^2 &= \text{Var} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \ell'(P_0, \Xi_i)[v_n^*] \right) \\ &= \gamma_n^{*'} \left( \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \frac{\partial \ell(\bar{\beta}^*, \Xi_i)}{\partial \beta} \frac{\partial \ell(\bar{\beta}^*, \Xi_i)'}{\partial \beta} \right] \right) \gamma_n^* \\ &= G_{K_n}' R_{K_n}^{-1} \Sigma_{K_n} R_{K_n}^{-1} G_{K_n}. \end{aligned} \quad (\text{B.10})$$

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

$$\begin{aligned} \hat{R}_{K_n} &= -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \ell(\hat{\beta}_n, \Xi_i)}{\partial \beta \partial \beta'} & \hat{\Sigma}_{K_n} &= \frac{1}{n} \sum_{i=1}^n \frac{\partial \ell(\hat{\beta}_n, \Xi_i)}{\partial \beta} \frac{\partial \ell(\hat{\beta}_n, \Xi_i)'}{\partial \beta} \\ \hat{G}_{K_n} &= \int_{\mathbf{Z}_1} \omega(\mathbf{Z}) \frac{\partial P(\hat{\beta}_n, \mathbf{Z})}{\partial \beta} d\mathbf{Z}_1 + \int_{\mathbf{Z}_1^c} \omega(\mathbf{Z}) \frac{\partial C(\hat{\beta}_n, \mathbf{Z})}{\partial \beta} d\mathbf{Z}_1 & \hat{v}_n^* &= \frac{\partial P(\bar{\beta}^*, \mathbf{Z})}{\partial \beta} \hat{R}_{K_n}^{-1} \hat{G}_{K_n}. \end{aligned}$$

Then

$$\|\widehat{v}_n^*\|_{sd,n}^2 = \widehat{G}'_{K_n} \widehat{R}_{K_n}^{-1} \widehat{\Sigma}_{K_n} \widehat{R}_{K_n}^{-1} \widehat{G}_{K_n} \equiv \widehat{V}_n \quad (\text{B.11})$$

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

**Infeasible Asymptotic Distribution** I show here that

$$\sqrt{n} \frac{\Gamma(\widehat{P}_n) - \Gamma(P_0)}{\|v_n^*\|_{sd}} \xrightarrow{d} N(0, 1). \quad (\text{B.12})$$

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 (B.12) 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 (B.12) into two parts,

$$\sqrt{n} \frac{\Gamma(\widehat{P}_n) - \Gamma(P_0)}{\|v_n^*\|_{sd}} = \underbrace{\sqrt{n} \frac{\Gamma(\widehat{P}_n) - \Gamma(P_{0,n})}{\|v_n^*\|_{sd}}}_{\text{Part A}} + \underbrace{\sqrt{n} \frac{\Gamma(P_{0,n}) - \Gamma(P_0)}{\|v_n^*\|_{sd}}}_{\text{Part B}} \quad (\text{B.13})$$

I start with Part B and prove it in steps.

Step 1: Show  $\frac{\|v_n^*\|}{\|v_n^*\|_{sd}} = O(1)$ .

$$\frac{\|v_n^*\|}{\|v_n^*\|_{sd}} = \frac{\gamma_n^{*\prime} R_{K_n} \gamma_n^*}{\gamma_n^{*\prime} \Sigma_{K_n} \gamma_n^*} \leq \frac{\lambda_{\max}(R_{K_n})}{\lambda_{\min}(\Sigma_{K_n})} = O(1), \quad (\text{B.14})$$

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 the bound in Assumption B.5 (ii) (see the discussion in Chen et al. (2014, Remark 3.2)).  
The result follows from continuity (linearity in  $v$ ) of  $\Gamma(P_0)[v]$  and denseness 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  $\langle v_n^*, P_{0,n} - P_0 \rangle = 0$ ,

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

by definition of the Riesz representer, Cauchy-Schwarz inequality, Step 3, Step 4, and Assumption B.5 (iii). Conclude that Part B in (B.13) is  $o(1)$ .

To show Part A, let  $u_n^* = v_n^* / \|v_n^*\|_{sd}$ .

Step 1: By linearity of  $\Gamma'(\cdot)[v]$  in  $v$ ,

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

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

$$\sup_{\|P_1 - P_2\| \leq \epsilon_n} \mu_n \{ \ell(P_1, \Xi) - \ell(P_2, \Xi) - \ell'(P_0, \Xi)[P_1 - P_2] \} = o_p(\epsilon_n). \quad (\text{B.15})$$

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 (B.15) 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 (B.15) 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, I follow Chen et al. (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  $\widehat{P}_n$ ,

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

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

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

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 (B.13), yields

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

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

**Feasible Asymptotic Distribution** I show that replacing  $\|v_n^*\|_{sd}$  with the estimate  $\|\widehat{v}_n^*\|_{sd,n}$  in (B.11) 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, I use the following lemma:

**Lemma B.3.** *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.1,

$$\begin{aligned} \|P_1 - P_2\|_2 &\leq M_1 \|f_1 - f_2\|_{m,1} \\ &\leq M_2 \left\| f_1^{Y,X} - f_2^{Y,X} \right\|_{m,\infty,\zeta_0} \quad \text{by Lemma A.9} \\ &= M_2 \left\| (h_1^{Y,X})^2 - (h_2^{Y,X})^2 \right\|_{m,\infty,\zeta_0} \quad \text{by Def. A.2} \\ &\leq M_3 \left\| h_1^{Y,X} - h_2^{Y,X} \right\|_{m_0+m,2,\zeta_0} \quad \text{by inequality (B.6)} \\ &\leq 2M_3 \mathcal{B}_0. \end{aligned} \tag{B.16}$$

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.

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

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

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

by Corollary 4 of [Nickl and Pötscher \(2007\)](#). Because  $m > d_u/2$  by assumption on the Gallant-Nychka spaces, we have that

$$\int_0^\infty H_{[\cdot]}^{1/2}(w, \mathcal{G}(\delta), \|\cdot\|_2) dw < \infty,$$

which is a sufficient condition for  $\mathcal{G}(\delta)$  to be  $\mathbb{P}$ -Donsker (see [Van Der Vaart and Wellner \(1996, p. 129\)](#)).  $\square$

Next, note that for  $\mathcal{W}_n \equiv \{v \in \mathcal{V}_n : \|v\| = 1\}$ , [Chen et al. \(2014\)](#) (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.3](#) 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

$$\left| \frac{\|\widehat{v}_n^*\|}{\|v_n^*\|} - 1 \right| = O_p(\epsilon_n^*), \quad \frac{\|\widehat{v}_n^* - v_n^*\|}{\|v_n^*\|} = O_p(\epsilon_n^*), \quad (\text{B.17})$$

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.3](#), the functional CLT, and [\(B.17\)](#).

To address the first term on the RHS, consider

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

by the Donsker property of [Lemma B.3](#), the functional CLT, and [\(B.17\)](#).

Combining results from the previous three displays and using (B.14), 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

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

as required. □