Term Structures, Shape Constraints, and Inference for Option Portfolios

Erik Vogt*
Department of Economics
Duke University

Job Market Paper
Version: January 8, 2014

Abstract

The illiquidity of options at long horizons has made it difficult to study the long-run properties of several key option portfolios, including the volatility index (VIX) and related measures. This paper proposes a new nonparametric framework to aid in the estimation of option portfolios at arbitrary maturities (even sparsely traded ones) and develops an asymptotic distribution theory for those portfolios. The distribution theory is used to quantify the estimation error induced by computing integrated option portfolios from a finite sample of noisy option data. Moreover, by relying on the method of sieves, the framework is nonparametric, adheres to economic shape restrictions for arbitrary maturities, yields closed-form option prices, and is easy to compute. The framework also permits the extraction of the entire term structure of risk-neutral distributions in closed-form. Monte Carlo simulations confirm the framework’s performance in finite samples. An application to the term structure of the synthetic variance swap portfolio finds sizeable uncertainty around the swap’s true fair value, particularly when the variance swap is synthesized from noisy long-maturity options. A nonparametric investigation into the term structure of the variance risk premium finds growing compensation for variance risk at long maturities.

JEL Classification: G12, G13, C14, C58

Keywords: Sieve M Estimation, Option Pricing, State price density, Nonparametric, Constraints.

*Email: erik.vogt@duke.edu. I am grateful to George Tauchen, Tim Bollerslev, Federico Bugni, Ron Gallant, Jia Li, Andrew Patton, Torben Andersen, Peter Christoffersen, Enrique Sentana, and Dacheng Xiu, as well as participants of the Duke Financial Econometrics lunchgroup and the 2013 Triangle Econometrics Conference for valuable comments and suggestions. All errors are my own.
Introduction

This paper is concerned with nonparametrically estimating a shape-conforming option price surface and quantifying the statistical uncertainty around associated integrated option portfolios. The use of option prices in the extraction of economically significant quantities is linked to their ability to approximate state-contingent claims. This observation is due to the fundamental insights of Ross (1976) and Breeden and Litzenberger (1978), who show that options can be combined into portfolios that replicate the role of Arrow-Debreu securities in spanning or hedging against uncertain future states. More recently, option prices and their portfolios have been used to extract state-price densities,\(^1\) to learn about the market prices of jump risk and crash fears,\(^2\) to estimate investor risk aversion and risk-neutral skewness\(^3\), to forecast returns,\(^4\) and to study how investors price and perceive volatility risk.\(^5\) The latter category, in particular, has benefitted from a collection of recently developed model-free implied volatility measures that are obtained by forming integrated portfolios of option prices, the most well-known of which is arguably the synthetic variance swap and its square-root, the VIX volatility index.\(^6\)

However, a central issue with implementing the above theory is the sparseness and noise of option data due to illiquidity. For example, in order to construct the synthetic variance swap portfolio or risk-neutral density at some horizon \(\tau\), an infinite continuum of European options expiring in \(\tau\) periods is required (Carr and Wu (2009)). In reality, option prices are discrete and truncated in strikes and maturity, since there may only be a few dozen observations available from which to infer the infinite option portfolio. The problem is even more severe when the objective is to investigate term structures implied by option prices, since the typical option panel has only a handful of maturities clustered at short horizons. To overcome this mismatch between data sparseness and theory, it has been customary to numerically interpolate observed options and then to treat the resulting estimates as though they represent actual observations on the theoretical object of interest. This approach omits at least two important considerations: first, the replacement of option prices by estimates should induce an estimation error. How quickly does the estimation error vanish as more options become available? Second, options are frequently observed with microstructure error arising from synchronization issues, bid-ask spreads, and quote staleness. Does the presence and variance of these errors affect the precision of the estimated portfolio, and can this be meaningfully quantified?

To answer these questions, I propose a new nonparametric framework to (1) overcome the discreteness and truncation of option data in both the strike and maturity dimension and (2) additionally provide a distribution theory for option portfolios. The key ingredient in this framework

---


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

\(^3\)See Bliss and Panigirtzoglou (2004) and Bakshi et al. (2003).

\(^4\)Bakshi et al. (2011) and Bollerslev et al. (2013).

\(^5\)Carr and Wu (2009), Bollerslev et al. (2011), Drechsler and Yaron (2011)

\(^6\)See Britten-Jones and Neuberger (2000), Jiang and Tian (2005), and Carr and Wu (2009).
is the nonparametric estimation of an option surface that satisfies certain shape constraints implied by economic theory. Thus, in illiquid regions of the option panel where maturities or strikes are only sparsely available, economic theory guides the estimator to maintain the proper structure. Moreover, the estimator has a number of appealing properties from an empirical perspective: first, option prices can be solved in closed-form. Second, because the option prices are shape-conforming along any maturity of interest, the estimator yields an entire term structure of valid state-price densities (SPDs) and risk-neutral CDFs indexed by maturity. That is, for any maturity of interest, the estimated SPDs always integrate to one, even in finite samples and even off the support of observed options. Third, the term structures of SPDs and risk-neutral CDFs are available in closed-form, avoiding the need to consider numerical differentiation or integration errors. Fourth, while having nonparametric properties, the estimator is easy to implement.

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

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

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

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

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

The sieve-estimated variance swap term structures are then used to estimate the term structure
of the variance risk premium. An active literature in financial economics has documented the
existence of a significant and time-varying risk premium that investors demand for bearing return
variance risk.\(^8\) Recently, Aït-Sahalia et al. (2012) and Fusari and Gonzalez-Perez (2012) have

\(^7\)This result is included in an Online Appendix for brevity.
\(^8\)See Bakshi and Madan (2006), Carr and Wu (2009), Bollerslev and Todorov (2011), Bollerslev et al. (2011),
and the related literature exploring parametric estimates of the volatility risk premium, e.g. Pan (2002), Eraker
(2004), and Broadie et al. (2007). Equilibrium models that seek to explain the existence and size of the variance risk
extended this literature by examining the variance risk premium at longer horizons using a flexible parametric model combined with data on variance swaps. The present paper complements their work from a nonparametric perspective and confirms that the variance risk premium term structure grows with maturity. Moreover, I find that the shape of the term structure depends on current volatility levels by applying a set of novel expectation hypothesis regressions.

The paper is organized as follows. Section 1 introduces the sieve least squares estimator for the shape-conforming option surface. Section 2 gives the closed-form option pricing formulas that are used in the sieve framework, and Section 3 establishes the estimator’s consistency and its rate of convergence. Section 4 derives the asymptotic distribution theory for integrated option portfolios, which are functionals of the option surface estimated in the preceding sections. The results of Monte Carlo simulations that examine the sieve estimator’s properties in finite samples is given in Section 5. Section 6 studies the term structure of the synthetic variance swap portfolio and associated variance risk premia, and Section 7 concludes.

1 A Nonparametric, Shape-Conforming Option Surface

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

1.1 Setup

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

$$P_0(\kappa, \tau) \equiv e^{-r\tau} E^Q_0 \left[ \left( \kappa - S_\tau \right)_+ \right] = e^{-r\tau} \int_0^\kappa \left[ \kappa - S \right] f^Q_0(S|\tau, \mathbf{V}) dS, \tag{1.1}$$

where $\mathbf{V}$ is a vector of state variables that generate the current information set, $f^Q_0(\cdot | \tau, \mathbf{V})$ is the unobserved transition or state-price density (SPD), and $r$ is the risk-free rate. The components of $\mathbf{V}$ are left unspecified and can contain any number of variables relevant to pricing options. The Heston model, for example, specifies $\mathbf{V} = (S_0, V_0)$, where $S_0$ is the current underlying price and $V_0$ represents spot volatility (see Heston (1993), Duffie et al. (2000)).

Since the goal is to estimate a shape-conforming option surface at a single point in time, $\mathbf{V}$ realizes to some fixed value $v_0$, so that the density in (1.1) becomes $f^Q_0(S|\tau, \mathbf{V} = v_0)$. To avoid cumbersome notation, I therefore define $f^Q_0(S|\tau) \equiv f^Q_0(S|\tau, \mathbf{V} = v_0)$, since $v_0$ is static across the option surface. On the other hand, $\tau$ is not static on the option surface because it indexes maturity. premium from a preference-based point of view are examined in Bakshi and Madan (2006), Bollerslev et al. (2009), and Drechsler and Yaron (2011).

---

9See, for example, Chapters 6 and 8 in Duffie (2001).
In this form, the risk-neutral valuation formula on a single option cross-section becomes

$$P_0(\kappa, \tau) \equiv e^{-r\tau} \int_0^\kappa [\kappa - S] f_0^Q(S|\tau) dS. \quad (1.2)$$

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

$$P_0(\kappa, \tau) \equiv P(f_0^Q, Z).$$

The no-arbitrage pricing equation (1.2) implies shape restrictions on the option prices. By differentiating $P(f_0^Q, Z)$ repeatedly with respect to the strike price $\kappa$, one has

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

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

Since the option price’s shape constraints are implied by the fact that $f_0^Q$ 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 $f_0^Q$ 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 $f_0^Q$. The results of this straightforward change-of-variables are analytically closed-form option prices that are theoretically informative and computationally convenient.

### 1.2 Change of Variables

I propose the following change of variables to obtain closed-form expressions of estimates to the option price in Eq. (1.2). Let $Y$ be the $\tau$-measurable random variable that satisfies

$$\log \left( \frac{S}{S_0} \right) = \mu(Z) + \sigma(Z)Y, \quad (1.3)$$

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

---

10 These shape constraints have been exploited elsewhere in the nonparametric option pricing literature for a single $\tau$. See, for example, Aït-Sahalia and Lo (1998), Aït-Sahalia and Duarte (2003), Bondarenko (2003), Yatchew and Härdle (2006), and Figlewski (2008).
Under this change of variables, the valuation equation (1.2) becomes

\[ P(f_0^Q, Z) = e^{-r\tau} \int_0^\kappa (\kappa - S) f_0^Q(S|\tau) dS \]
\[ = e^{-r\tau} \int_0^{d(Z)} \left( \kappa - S_0 e^{\mu(Z) + \sigma(Z)Y} \right) f_0(Y|\tau) dY \]
\[ \equiv P_Y(f_0, Z), \]  

where

\[ d(Z) = \frac{\log(\kappa/S) - \mu(Z)}{\sigma(Z)}. \]  

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

\[ f_0^Q(s|\tau) = (s\sigma(Z))^{-1} f_0(s|\tau). \]  

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

If only the option price and not \( \hat{g}_n \) is needed, then one does not have to perform the Jacobian transformation, since Eq. (1.4) says \( P(f_0^Q, Z) = P_Y(f_0, Z) \). This allows the analysis to focus on option pricing equations of the form

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

It is easy to verify that Eq. (1.7) contains the same shape restrictions as Eq. (1.2) for any \( f \) with \( \int f(y|\tau) dy = 1 \). Proposition 1 below solves this integral in closed-form when \( f \) represents a sieve approximation.

### 1.3 Sieve Least Squares Regression

The goal is to obtain a shape-conforming option surface by directly using the structure implied by Eq. (1.7). Notice that the true option price \( P_Y(f_0, Z) \) is not observed because of the presence of the unknown infinite-dimensional parameter \( f_0 \), which is assumed to reside in some general function space \( F \). The space \( F \) consists of a very large class of smooth conditional densities \( f(y|\tau) \) and will be described shortly. Thus, given a random sample \( \{P_i, Z_i\}_{i=1}^n \) on put option prices \( P_i \) and characteristics \( Z_i \), the idea is to solve problems of the form

\[ \hat{f} = \arg \inf_{f \in \mathcal{F}} \left\{ \frac{1}{n} \sum_{i=1}^n \left[ P_i - P_Y(f, Z_i) \right]^2 W_i \right\}, \]  

where \( \mathcal{F} \) is the space of functions \( f \) that can be approximated by a sieve.
where \( P_Y(\cdot) \) is the known pricing functional from equation (1.7), \( Z_i \) is the vector of observables, and \( W_i \equiv W(Z_i) \) are known weights as a function of \( Z_i \).\(^{11}\)

The main difficulty with solving the optimization problem in equation (1.8) is the infinite dimension of the function space \( \mathcal{F} \). In general, optimizing over an infinite-dimensional function space may not be feasible or could even be ill-posed. Instead, it is typical to proceed by the method of sieves, which involves approximating \( \mathcal{F} \) by a sequence of finite-dimensional function spaces (the “sieve” spaces)

\[
\mathcal{F}_K \subset \mathcal{F}_{K+1} \subset \cdots \subset \mathcal{F}
\]

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

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

\[
\hat{f}_{K_n} = \arg \min_{f \in \mathcal{F}_{K_n}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left[ P_i - P_Y(f, Z_i) \right]^2 W_i \right\}
\]

are consistent for \( f_0 \). I provide conditions in Section 3 that ensure that the convergence of \( \hat{f}_{K_n} \) to \( f_0 \) also implies the convergence of \( P_Y(\hat{f}, Z) \) to \( P_Y(f_0, Z) \) under suitable norms. Also note that the minimum in (1.10) is taken over the subspace \( \mathcal{F}_{K_n} \subset \mathcal{F} \), where \( K_n \to \infty \) slowly as \( n \to \infty \). The requirement that \( K_n \to \infty \) slowly is crucial and can be interpreted as the sieve analog of a bandwidth selection in kernel estimation and has an intuitive interpretation: as the sample size grows, the approximating spaces \( \mathcal{F}_{K_n} \) increasingly resemble the parent space \( \mathcal{F} \). The regularity conditions then ensure that optima on \( \mathcal{F}_{K_n} \) indeed converge to \( f_0 \).

### 1.4 The Definition of \( \mathcal{F} \) and its Sieve Spaces \( \mathcal{F}_K \)

For an option surface to conform to the theoretical shape restrictions of Eq. (1.7), \( \mathcal{F} \) must be a function space consisting of conditional densities \( f(Y|\tau) \) in the sense that \( \int f(y|\tau) \, dy = 1 \) for all \( \tau \). 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 Sobolev subspace and \( \{ \mathcal{F}^{Y,\tau}_K \}_{K=0}^\infty \) is a collection of squared and scaled Hermite functions, then \( \{ \mathcal{F}^{Y,\tau}_K \}_{K=0}^\infty \) is a valid sieve for \( \mathcal{F}^{Y,\tau} \). I show that the conditional approximating spaces \( \{ \mathcal{F}_K \}_{K=0}^\infty \) consisting of those functions \( f_K \) for which \( f_K(y|\tau) = f^{Y,\tau}_K(y, \tau)/ \int f^{Y,\tau}_K(y, \tau) \, dy \) for some \( f^{Y,\tau}_K \in \mathcal{F}^{Y,\tau}_K \) is also a valid sieve for the conditional parent space \( \mathcal{F} \), although the topologies differ. A formal discussion of these technical details is postponed until Section 3, when the asymptotic properties of the estimator are examined. For now, it is sufficient to note that when \( \mathcal{F}_K \) is constructed from a ratio of two Gallant-Nychka densities, then there exists

\(^{11}\)Call options can be handled analogously in what follows, but for brevity I focus on puts.
a norm under which $\mathcal{F}_K$ is a valid sieve for $\mathcal{F}$.

The Gallant-Nychka sieve spaces $\{\mathcal{F}^Y_\tau\}_K$ consist of functions of the form

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

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

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

$$f_K(y|\tau) = f^Y_\tau(y, \tau) \left( \int f^Y_\tau(y, \tau) dy \right)^{-1}$$

for some joint density $f^Y_\tau \in \mathcal{F}^Y_\tau$. Notice that because the sieve joint densities $f^Y_\tau(y, \tau)$ are completely determined by the parameter matrix of coefficients $B$, then so are the conditional densities in $\mathcal{F}_K$. Therefore, for $\beta \equiv vec(B)$, the least squares problem in (1.10) becomes

$$\hat{\beta}_n = \arg \min_{\beta \in \mathbb{R}^K} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left[ P_i - P_Y(\beta, \mathbf{Z}_i) \right]^2 W_i \right\}$$

s.t. $\sum_{k=0}^{K_y} \sum_{j=0}^{K_\tau} \beta_{kj}^2 = 1,$

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

As written, $P_Y(\beta, \mathbf{Z}_i)$ is identical to $P_Y(f_K, \mathbf{Z}_i)$ from Eq. (1.7), which still requires an integration to obtain a candidate option price. Section 2 shows that in fact, $P_Y(f_K, \mathbf{Z}_i)$ is available in closed-form for any $f_K \in \mathcal{F}_K$, which considerably facilitates implementation.

---

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

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

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 $K$. 

---
1.5 The Sieve Satisfies the Required Shape Constraints

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

\[
e^{r\tau} \frac{\partial P_{Y}(f_{K}, Z)}{\partial \kappa} = \int_{0}^{d(Z)} f_{K}(Y|\tau)dY\tau = F_{K}\left(\frac{\log(S_{0}/\kappa) - \mu(Z)}{\sigma(Z)}|\tau\right) \tag{1.14}
\]

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) \( P_{Y}(f_{K}, Z) \) is increasing in the \( \kappa \) dimension (since \( F_{K} > 0 \) as a CDF), (b) \( P_{Y}(f_{K}, Z) \) is convex (since \( \partial F_{K}/\partial \kappa \) is \( f_{K}/(\kappa\sigma(Z)) \) and \( f_{K} \geq 0 \), and (c)

\[
\lim_{\kappa \to \tau, \infty} e^{r\tau} \frac{\partial P_{Y}(f_{K}, Z)}{\partial \kappa} = 1, \quad \lim_{\kappa \to 0} e^{r\tau} \frac{\partial P_{Y}(f_{K}, Z)}{\partial \kappa} = 0. \tag{1.15}
\]

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

2 Closed-form Option Prices

I now provide closed-form expressions for the sieve option prices \( P_{Y}(f_{K}, Z) \) to be used in the regression (1.10) and show that \( \mu(Z) \) and \( \sigma(Z) \) can be chosen so that the sieve option prices have a natural interpretation as expansions around the Black-Scholes model.

2.1 Closed-Form Option Prices

To obtain closed-form option prices, it is convenient to first obtain a closed-form expression for the conditional sieve densities \( f_{K} \) of Eq. (1.12). This is done by expanding the squared polynomial term in the joint densities of Eq. (1.11) using techniques similar to those in León et al. (2009).

**Lemma 2.1.** Any \( f_{K} \in \mathcal{F}_{K} \) can be expressed in the form

\[
f_{K}(y|\tau) = \sum_{k=0}^{2K_{y}} \gamma_{k}(B, \tau)H_{k}(y)\phi(y), \tag{2.1}
\]

where

\[
\gamma_{k}(B, \tau) = \frac{\alpha(B, \tau)'A_{k}\alpha(B, \tau)}{\alpha(B, \tau)'\alpha(B, \tau)'^{2}},
\]

\( A_{k} \) is a known matrix of constants, and \( \alpha(B, \tau) \) is a \((K_{y}+1) \times 1\) column vector obtained by stacking the \( \alpha_{k}(B, \tau) \) in Eq. (1.11).

The use of densities \( f_{K}(y|\tau) \) that are linear combinations of functions in \( y \) helps with the derivation of closed-form option prices. The following result is an extension of Proposition 9 in León et al. (2009) to the case allowing for conditioning on \( \tau \).
Proposition 1. For a candidate SPD $f_K(x|\tau) \in \mathcal{F}_K$ of the form given in equation (2.1), the put option price $P_Y(f_K, Z)$ from equation (1.13) is given by

$$
P_Y(f_K, Z) = \kappa e^{-r\tau} \left[ \Phi(d(Z)) - \sum_{k=1}^{2K_y} \frac{\gamma_k(B, \tau)}{\sqrt{k}} H_{k-1}(d(Z)) \phi(d(Z)) \right] - S_0 e^{-r\tau + \mu(Z)} \left[ e^{\sigma(Z)^2/2} \Phi(d(Z) - \sigma(Z)) + \sum_{k=1}^{2K_y} \gamma_k(B, \tau) I_k^*(d(Z)) \right] \tag{2.2}
$$

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

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

and $\gamma_k(B, \tau)$ is the coefficient function given in equation (2.1).

The price of a call option is given by

$$
C_Y(f_K, Z) = S_0 e^{-r\tau + \mu(Z)} \left[ e^{\sigma(Z)^2/2} [1 - \Phi(d(Z) - \sigma(Z))] + \sum_{k=1}^{2K_y} \gamma_k(B, \tau) I_k^*(d(Z)) \right] - \kappa e^{-r\tau} \left[ 1 - \Phi(d(Z)) \right] - \sum_{k=1}^{K_y} \frac{\gamma_k(B, \tau)}{\sqrt{k}} H_{k-1}(d(Z)) \phi(d(Z)) \right]. \tag{2.3}
$$

Remark 2.2. The significance of this result is that it makes the sieve regressor function of Eq. (1.13) available in closed-form. Indeed, in sharp contrast to the large class of parametric option pricing models of Heston (1993) and Duffie et al. (2000), no numerical integrations are required to compute an option price, which significantly facilitates the optimization problem Eq. (1.13). Moreover, the Online Supplement to this paper also provides closed-form gradients and second derivatives of the prices $P_Y(f_K, Z)$. And finally, having closed-form price estimates additionally simplifies the ultimate objective of computing integrated portfolios of $P_Y(f_K, Z)$.

The sieve put option price in Eq. (2.2) has an intuitive interpretation. Rearranging equation (2.2), one obtains

$$
P_Y(f_K, Z) = \kappa e^{-r\tau} \Phi(d(Z)) - S_0 e^{-r\tau + \mu(Z)} e^{\sigma(Z)^2/2} \Phi(d(Z) - \sigma(Z)) - \sum_{k=1}^{K_y} \gamma_k(B, \tau) \left[ \frac{1}{\sqrt{k}} H_{k-1}(d(Z)) \phi(d(Z)) + S_0 e^{-r\tau + \mu(Z)} I_k^*(d(Z)) \right]. \tag{2.4}
$$

Inspection of equation (2.4) shows that choosing

$$
\sigma(Z) \equiv \sigma \sqrt{\tau}, \quad \mu(Z) \equiv (\tau - q - \sigma^2/2) \tau \tag{2.5}
$$
will cause the leading term in equation (2.4) to become \( P_{BS}(\sigma, Z) = \kappa e^{-r\tau} \Phi(d(Z)) - S_0 e^{-\tau} \Phi(d(Z) - \sigma \sqrt{\tau}) \), where \( q \) is the dividend yield, and where the function \( d(Z) \) from equation (1.5) is now \( d(Z) = (\log(S_0/\kappa) - (r - q - \sigma^2/2)\tau)/(\sigma \sqrt{\tau}) \). The value \( \sigma \) is a tuning parameter in the sieve framework and is chosen to be equal to the average implied volatility of the observed option cross-section.

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

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

(2.6)

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

If the \( \gamma_k(B, \tau) \) terms for \( k \geq 1 \) above are nonzero in the data, then we can regard this as evidence against the Black-Scholes model. In particular, it has been well-documented that conditional distributions of asset prices contain substantial volatility, skewness, and kurtosis that the Black-Scholes model is unable to capture. Modeling techniques to introduce such features into the return distribution includes the addition of stochastic volatility [Heston (1993)], as well as jumps [Bates (1996), Bates (2000), Bakshi et al. (1997), Duffie et al. (2000)]. The simulation study in Section 5 explores how these continuous time parametric features feed into the coefficients of the Hermite expansion and shows that low-order expansion terms (order 4 to 6) are quite capable of fitting the conditional distributions implied by complicated stochastic volatility and jump specifications.

### 3 Consistency

The critical feature of \( P_Y(f, Z) \) in Eq. (1.7) is that it generates shape-conforming option prices for any \( \tau \). It does so by indexing state-price densities with \( \tau \), which appears as a conditioning variable. A straightforward extension of Eq. (1.7) is to permit a state-price density with arbitrary conditioning information, \( f(Y|X) \), where \( X \subseteq Z \) and contains \( \tau \). For example, one could have \( X = (\tau, r) \) to accommodate a risk-free rate term structure that does not match the maturities of observed option prices. Allowing for general \( X \) is instructive in order to see how the rate of convergence is slowed by the dimension of the conditioning variable \( X \). Therefore, this section establishes the asymptotic theory for the extension that allows for arbitrary conditioning information in the SPD.

A summary of the theoretical results developed in this section is as follows. First, I move from

---

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

\(^{14}\)Recall that the \( \gamma_k(B, \tau) \) terms also contain expansions.
Gallant-Nychka joint density spaces to conditional density spaces (the norm changes), and from conditional spaces to option price spaces (with another norm change). The theoretical contribution of this section is to show that each of these transitions corresponds to a Lipschitz map between function spaces. Thus, the complexity of the option price spaces, as measured by the $L^2(\mathbb{R}^{1+d_x}, \mathbb{P})$ metric entropy from empirical process theory, is completely determined by the complexity of the Gallant-Nychka joint density spaces, which are Sobolev subspaces with known covering numbers. Hence, one can apply existing general theorems from the sieve estimation and inference literature to obtain convergence and asymptotic distribution results.

### 3.1 Consistency of Sieve Option Prices and State-Price Densities

With $\tilde{\beta}_n$ in hand, estimated option prices are simply given by $P_Y(\tilde{\beta}_n; Z) \equiv \tilde{P}^K_Y$, which has the closed-form expression stated in Proposition 1. This subsection establishes that $\|\tilde{P}^K_Y - P^0_Y\|_2 \to 0$ as $n \to 0$, where the consistency norm is the $L^2(\mathbb{R}^{d_x}, \mathbb{P})$ norm defined below in Eq. (3.1).

The asymptotic results developed in the remainder of this section make use of Sobolev spaces and associated norms. Detailed definitions of these spaces are given in Appendix A.1. Under those definitions, the sieve spaces of conditional densities from Section 1.4 are assumed to be subspaces of $W^{m,1}(\mathbb{R}^{d_u})$ stated in Definition A.1.

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

$$\|P_{Y,1} - P_{Y,2}\|^2_2 = \mathbb{E}\{[P_{Y,1}(Z) - P_{Y,2}(Z)]^2 W(Z)\} = \int [P_{Y,1}(Z) - P_{Y,2}(Z)]^2 W(Z) \mathbb{P}(dZ) \tag{3.1}$$

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

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

#### 3.1.1 Bounded Stock Prices

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

**Assumption 3.1.** (Locally Uniformly Bounded Stock Prices). Given any $f_0 \in \mathcal{F}$, there exists an $\|\cdot\|_{m,1}$-open neighborhood $U$ containing $f_0$ and a constant $M$ (possibly depending on $U$) such that

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

where

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

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

Assumption 3.1 is a technical condition that is required in order for certain arguments in the
asymptotic theory to go through and has little bearing on practical applications. In particular, it is easy to check within optimization routines that this constraint is never close to being violated.

I also make the following assumption.

**Assumption 3.2.** Assume

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

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

Assumption 3.2 is standard and very mild. It says that the options are observed with conditional mean-zero errors with bounded \( 2 + \delta \) moments.

Taken together, Assumptions 3.1 and 3.2 imply a number of useful properties that are summarized in several Lemmas that I prove in Appendix A.1. These properties are used to establish the following consistency result.

**Proposition 2.** (Consistency) Under Assumptions 3.1 and 3.2, \( d(\hat{f}_n, f_0) \xrightarrow{P} 0 \) and \( \|\hat{P}_Y^{K_n} - P_Y^0\|_2 \xrightarrow{P} 0 \).

**Proof.** Appendix B. \( \square \)

### 3.2 Rate of Convergence

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

The rate of convergence of the sieve option prices depends on notions of size or complexity of the space of admissible option pricing functions as measured by the latter’s bracketing numbers. Note that each candidate option price \( P_Y(f, Z) \) is uniquely identified by the state-price density \( f \) (Lemma A.5). In turn, \( f \in \mathcal{F} \) is the target of a Lipschitz map with preimage \( f^{Y:X} = h^2 + \varepsilon_0 h_0 \), a Gallant-Nychka density (Lemma A.6). The Gallant-Nychka class of densities requires \( h \) to reside in \( \mathcal{H} \), a closed Sobolev ball of some radius \( B_0 \).\(^{15}\) The rate result obtained below hinges on the observation that the collection of possible option prices,

\[ \mathcal{P} = \{ P_Y : P_Y(Z) = P_Y(f, Z) \text{ for some } f \in \mathcal{F} \}, \]

is ultimately Lipschitz in the index parameter \( h \in \mathcal{H} \). Therefore, the size and complexity of \( \mathcal{P} \), as measured by its \( L^2(\mathbb{R}^d, \mathbb{P}) \) bracketing number, is bounded by the covering number of the Sobolev ball \( \mathcal{H} \) (see Van Der Vaart and Wellner (1996)). The following assumptions are used in the proof of the rate result below.

**Assumption 3.3.** \( \sigma(Z) = \mathbb{E}[e|Z] \) and \( W(Z) \) are bounded, where \( e = P - P_Y^0(Z) \).

\(^{15}\)For further details, see the Online Appendix as well as Gallant and Nychka (1987).
Assumption 3.4. The deterministic approximation error rate satisfies

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

for some $\alpha > 0$, where $h \in H$ and its orthogonal projection $\pi_{K_n} h \in H_{K_n}$ are defined in Definitions A.2 and A.3, and where $K_n \equiv \lfloor K_y(n) + 1 \rfloor \lfloor K_{x,1}(n) + 1 \rfloor \cdots [K_{x,d_u}(n) + 1]$ denotes the total number of series terms for functions in $H_{K_n}$.

Assumption 3.5. For state-price densities in $W_{m,1}(\mathbb{R}^{d_u})$, we have $m \geq d_u + 2$.

Assumption 3.3 is mild and commonly adopted in the literature (see Chen (2007)). Assumption 3.4 takes as given the deterministic approximation error rate, and Assumption 3.5 imposes additional smoothness in order to invoke Sobolev imbedding theorems (see Adams and Fournier (2003)).

Proposition 3. Let $\hat{P}_Y(Z) \equiv P_Y(\hat{f}_n, Z)$, where $\hat{f}_n$ solves (1.10), and let $P_Y^0 \equiv P_Y(f_0, Z)$ denote the true option price. Under Assumptions 3.1, 3.2, 3.3, and 3.4,

$$\|\hat{P}_Y - P_Y^0\|_2 = O_P(\varepsilon_n), \quad \text{where } \varepsilon_n = \max\{n^{-(m_0 + m)/(2(m_0 + m) + d_u)}, n^{-\alpha_d_u/(2(m_0 + m) + d_u)}\}.$$

Proof. Appendix B. \hfill \Box

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

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

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

4 Inference for Option Portfolios

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

Split $Z = (Z_1, Z_2)$. The prime example is $Z_1 = \kappa$, which includes the large class of functionals used in option hedging that integrate option prices over strikes. While the results in this subsection apply more generally, for concreteness this discussion will consider linear functionals of the form

$$\Gamma(P_Y) \equiv \Gamma_{Z_1}(P_Y) = c(Z) + \int_{Z_1} \omega(Z_1, Z_2) [P_Y(Z_1, Z_2) + b(Z_1, Z_2)]dZ_1. \quad (4.1)$$
This general functional includes so-called weighted integration functionals as well as evaluation functionals, or combinations of both (this terminology is borrowed from Chen et al. (2013)). The following examples serve to illustrate the flexibility of this functional.

Example 4.1. To compute the Synthetic Variance Swap (SVS) of Carr and Wu (2009) at horizon \( \tau \), one has for \( Z_1 = \kappa, Z_2 = Z_{-\kappa} \) and by put-call parity\(^{16}\)

\[
SVS(\tau) = \Gamma_{Z_2}(P_Y) = \frac{2}{\tau} \int_{-\infty}^{\infty} e^{\tau \kappa} P_Y(Z) d\kappa + \frac{2}{\tau} \int_{F(Z)}^\infty e^{\tau \kappa} C_Y(Z) d\kappa
\]

(4.2)

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

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

and \( c(Z) = 0 \).

Example 4.2. Bakshi et al. (2011) consider the exponential claim on integrated variance proposed in Carr and Lee (2008) given by

\[
\Gamma(P_Y) \equiv e^{-rT} E^Q \left[ \exp \left( - \int_0^T \sigma_t^2 dt \right) \bigg| Z \right]
\]

(4.3)

\[
\approx e^{-rT} + \int_{-\infty}^{S_0} \omega(Z) P_Y(Z) d\kappa + \int_{S_0}^\infty \omega(Z) C_Y(Z) d\kappa
\]

(4.4)

\[
= c(Z) + \int_{F(Z)} \omega(Z) [P_Y(Z) + b(Z)] d\kappa
\]

(4.5)

where

\[
\omega(Z) = \frac{8}{\sqrt{14}} \cos \left( \arctan(1/\sqrt{7}) + \sqrt{7} \ln \left( \frac{K}{S_0} \right) \right)
\]

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

\[
c(Z) = e^{-rT}
\]

Many more functionals of option prices fall under this framework. For instance, the large class of so-called “model-free” volatility measures that are widely used in the literature, involve some type of weighted integral of option prices across strikes. See, for example, Carr and Wu (2009), Jiang and Tian (2005), Britten-Jones and Neuberger (2000), and Aït-Sahalia et al. (2012) and the...
many references therein. The Carr and Wu (2009) SVS-type portfolios will be the subject of this paper’s empirical application below. Note that because of put-call parity, the \( b(Z) \) term in Eq. (4.1) can serve to create call options from the put pricing function.

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

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

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

**Assumption 4.1.**

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

(ii) \( \lim_{K_n \to \infty} \| \frac{\partial R_n(P_{0Y})}{\partial \beta} \|_E^2 < \infty \).

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

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

Here, \( R_{K_n} \) is the population analog of (4.8) below. The main result of this section is the following proposition, which enables the construction of confidence intervals or a wide array of option portfolios that fall under the functional class in Eq. (4.1).

**Proposition 4.** Assume the conditions of Proposition 3 as well as Assumption 4.1. Then

\[
\sqrt{n} \hat{V}_n^{-1/2} [\Gamma(\hat{P}_Y) - \Gamma(P_{0Y})] \xrightarrow{d} N(0,1)
\]

(4.6)

where

\[
\hat{V}_n = \hat{G}_{K_n}' \hat{R}_{K_n}^{-1} \hat{\Sigma}_{K_n} \hat{R}_{K_n}^{-1} \hat{G}_{K_n}
\]

(4.7)
and where

\[
\begin{align*}
\hat{G}_{Kn} &= \frac{\partial \Gamma(P_Y(\hat{\beta}_n, Z))}{\partial \beta} \\
\hat{R}_{Kn} &= -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \ell(\hat{\beta}_n, \Xi_i)}{\partial \beta \partial \beta'} \\
\hat{\Sigma}_{Kn} &= \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \ell(\hat{\beta}_n, \Xi_i)}{\partial \beta} \left( \frac{\partial \ell(\hat{\beta}_n, \Xi_i)'}{\partial \beta} \right)
\end{align*}
\tag{4.8}
\]

Remark 4.3. The objects in Eq. (4.8) 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 SVS, i.e. \( SVS(1) \) of Example 4.1, then Eq. (4.6) says that the \( SVS(1) \) is asymptotically normally distributed with estimated variance \( \hat{\Sigma}_n \) computed above. Moreover, this calculation can be done for any \( SVS(\tau) \) for arbitrary \( \tau \), which enables the construction of SVS term structures that quantify the estimation error involved with the construction of long-term SVS’s. An analysis of SVS term structures and associated estimation errors is conducted in Section 6.

Remark 4.4. Proposition 4 shows \( \sqrt{n} \)-consistency of functionals of the option price, whereas Proposition 3 shows a somewhat slower rate for the convergence of \( \hat{P}_{Y}^{K_n} \) to \( P_{Y}^{0} \). This is because the functionals of interest (Eq. (4.1)) belong to the so-called regular class of functionals of Chen et al. (2013). Similar \( \sqrt{n} \)-consistency of well-behaved functionals is obtained in Newey (1997).

**Implementation Summary**

1. Choose \( \mu(Z) \) and \( \sigma(Z) \). To center expansions around Black-Scholes, use Eq. (2.5).
2. Construct \( P_Y(\beta, Z) \) from Proposition 1.
3. Choose \( K_n = (K_y(n) + 1)(K_{\tau}(n) + 1) \) to grow slowly as \( n \to \infty \), e.g. by minimizing BIC.
4. Optimize the objective function over sieve coefficients in Eq. (1.13), using all options from a given cross-section of options.
5. Form \( \hat{V}_n \) using Eq. (4.7) and use critical values from standard normal tables.

5 Simulations

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

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

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

(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 double-jump process, which is a special case of the general affine-jump diffusion processes treated in Duffie et al. (2000) that is nonetheless general enough to nest the celebrated models of Black and Scholes (1973), Heston (1993), and other jump-diffusions commonly used in the option pricing literature. The values of these parameters are set to those used in Andersen et al. (2012) and are given in the Online Appendix to this paper.

### 5.1 Shape-Constrained Fitting

The main paper contributes to an existing literature concerned with shape-constrained option price fitting. Specifically, Eqs. (1.14) and (1.15) show that the option pricing function \( P_Y \) (a.) is monotone in \( \kappa \), (b.) convex in \( \kappa \), (c.) has first derivative \( e^{\tau \bar{\kappa}} \frac{\partial P^S_Y(Z)}{\partial \kappa} \) as a CDF, yielding limits of 0 and 1 as \( \kappa \) goes to 0 and \( +\infty \), respectively, and (d.) restrictions (a.)-(c.) must hold for arbitrary time-to-maturity \( \tau \).

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

Another potential value of shape-constrained option estimators is their use in applications that require a continuum of option prices in the strike dimension that extends to infinity as inputs in empirical investigations. These include studies that use the integrated portfolios of the form Eq. (4.1) [see e.g. Britten-Jones and Neuberger (2000), Bakshi et al. (2003), Jiang and Tian]
(2005), Carr and Lee (2008), Carr and Wu (2009), or their uses in e.g. Bollerslev and Zhou (2006), Bollerslev et al. (2011), Bakshi et al. (2011) among many others]. Thus, since it is well-known that option prices are only discretely observed on a truncated interval, one often requires some type of interpolation or smoothing on the range of observed discrete options, and an extrapolation beyond the truncated range of strikes. Jiang and Tian (2005) examine the numerical properties of one such interpolation and extrapolation procedure. The sieve estimator derived above provides a complementary tool that upholds the no-arbitrage shape-constraints across all maturities \( \tau \), even for maturities for which there are few or even no observations. This is done by evaluating \( \hat{P}_\tau \) at a desired \( \tau \).

Figure 1 illustrates the estimator’s adherence to the shape constraints in Eqs. (1.14) and (1.15). The figure shows that for each observed maturity in the simulated sample, option prices satisfy the shape constraints even beyond the truncated range of observed option data and asymptote to the option’s intrinsic value. Thus, the sieve estimator performs the task of both interpolating between observed data, and extrapolating beyond observed data in a single estimation step across all maturities.

5.2 Risk-Neutral Quantiles and Densities

The purpose of shape-constrained option fitting is often to differentiate the (scaled) put pricing function once to obtain the option-implied risk neutral CDF, or differentiating it twice to obtain the state-price density. This is the subject of Jackwerth and Rubinstein (1996), A¨ıt-Sahalia and Lo (1998), Figlewski (2008), Birru and Figlewski (2012), as well as Bondarenko (2003) and the many references therein.

In contrast to methods that require numerical differentiation of the option pricing function, the sieve estimator in Eq. (1.13) delivers risk-neutral CDFs and PDFs in closed form. The closed-form expression of the risk-neutral PDF (i.e. the state-price density) is obtained by plugging Eq. (2.1) into Eq. (1.6) above. The closed-form formula for the risk-neutral CDF can be obtained by integrating the PDF and using properties of Hermite polynomials, yielding

\[
Q_K(S_T \leq \kappa | \tau) = \Phi(d(Z)) - \sum_{k=1}^{2K} \gamma_k(B, \tau) \frac{H_{k-1}(d(Z))}{\sqrt{k}} \phi(d(Z)),
\]

where \( Q_K(A) = \int_A f_K(x | \tau) dx \) is the sieve-implied risk-neutral measure obtained by integrating against the sieve state-price density. See Eq. (B.2) in the proof of Proposition 1 for a derivation of this expression.

Figures 2 and 3 show the term structures of risk-neutral CDFs and PDFs, using the estimated sieve coefficients obtained by solving the least squares problem in Eq. (1.13) on data generated by the
SVJJ process in Eq. (5.1) and the last column of Table F.1. The CDFs cannot violate the 0 and 1 bounds at all maturities by construction of the risk-neutral PDF, since it was scaled to integrate to one for expansions of any order $K$. The true CDF and true PDF are plotted as well and show remarkable fit across all maturities in the panel. However, in the more extreme quantiles of the data, and particularly in the left tail, the sieve estimator begins to oscillate. This is a consequence of the bias-variance tradeoff given in the rate result of Proposition 3. In particular, the rate shows that the estimator’s convergence is a function of bias (i.e., how quickly the sieve space fills in the parent space as $K_n$ grows), versus the variability of the approximator, which grows with $K_n$. Thus, the oscillatory behavior in the plot can be decreased by reducing expansion terms, however at the cost of introducing some bias into the estimate.\footnote{An alternative approach to reducing oscillatory behavior would be to add a penalization or “regularization” term against oscillatory solutions to the least squares problem in Eq. (1.10). An approach of this type would fall under the penalized sieve literature and would require techniques that are beyond the scope of this paper.}

[Figure 3 about here.]

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

### 5.3 Coverage

This section shows that uninformed choices of expansion terms can yield incorrect inference on portfolios of option prices. In particular, the discussion in Section 1 demonstrated that the number of sieve expansion terms $K_n = (K_y(n) + 1)(K_r(n) + 1)$ must grow slowly as the sample size $n$ tends to infinity. As mentioned above, the result in Proposition 3 shows that the estimated option price $\hat{P}_{y}^{K_n}$ converges to the true option price $P_{y}^{0}$ at a rate that trades off two criteria with bias and variance interpretations [see Chen (2007)]. That is, on the one hand, large choices of $K_n$ result in lower bias, as the sieve space has more basis functions available to approximate the parent function space. On the other hand, large $K_n$ will increase the variability of the sieve estimate, leading to oscillatory behavior. The optimal choice of $K_n$, therefore, balances these two influences. However, an inspection of the rate in Proposition 3 shows that the optimal choice of $K_n$ depends on the degree of smoothness of the true state-price density $f_{0}$. Since $f_{0}$ is unknown, the rate result does not inform us of the optimal $K_n$. A formal theory for selecting $K_n$ specifically for the least squares option pricing problem in Eq. (1.10) is therefore required, but beyond the scope of the current paper.

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

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

Then, the NLLS problem in Eq. (1.13) is solved on the sample of $n = 250$ option prices, which yields $\hat{\beta}_n$ that is then plugged into Eq. (2.2) to yield $\hat{P}^{K_s}_Y(Z)$. These estimated option prices are then used to construct the Carr-Wu Synthetic Variance Swap at each maturity [see Carr and Wu (2009)], by numerically integrating

$$SVS(\tau) = \frac{2}{\tau} e^{\tau r} \int_0^{F(Z)} \frac{1}{k^2} \hat{P}^{K_s}_Y(Z) dk + \frac{2}{\tau} e^{\tau r} \int_{F(Z)}^\infty \frac{1}{k^2} \hat{C}^{K_s}_Y(Z) dk,$$

(5.3)

where the call prices $\hat{C}^{K_s}_Y(Z)$ are obtained by put-call parity [see Example 4.1 above]. The associated sieve estimate of the $VIX(\tau)$ is given by

$$\hat{VIX}(\tau) = 100\sqrt{SVS(\tau)}.$$ 

(5.4)

Because this can be done for each $\tau$ from 1 to 24 months, this procedure yields an entire estimated VIX term structure. Because I also observe a rich set of noise-free option prices within the simulation, I can compute the true VIX term structure as well. Finally, for each point along the VIX term structure, 95%-confidence intervals are constructed using the variance matrix in Proposition 4, with $\hat{G}_{K_n}$ similar to (6.5) below with adjustment for the square root and scale factor 100. The random sampling of 250 options and sieve estimation is then repeated 1,000 times, yielding 1,000 $VIX(\tau)$ confidence intervals for each $\tau$ from 1 to 24 months.

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

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

This form of overrejection is significantly improved when the BIC is allowed to choose the expansion terms. The middle panel of Table 1 shows expansions to $(K_y, K_\tau) = (6, 2)$, which is the BIC choice for SVJ DGPs. Allowing the expansion to go from $(3, 1)$ to $(6, 2)$ improved the coverage rate from 6.9% to 97%, which is much closer to the asymptotic rejection probability of 95%.

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

6 The Term Structure of Variance Swaps and Risk Premia

An active literature in financial economics is concerned with studying the variance risk premium, i.e. the compensation that investors demand for bearing return variance risk. This literature has shown that investors are averse to return variation and have historically demanded a significant, but time-varying premium for holding securities that are exposed to such risk. Moreover, the variance risk premium, although correlated with the equity risk premium, appears to identify a source of risk that is unexplained by classic risk factors.\footnote{See, for example, Bakshi and Madan (2006), Carr and Wu (2009), Bollerslev and Todorov (2011), Bollerslev et al. (2011), Bollerslev et al. (2013), and the references therein.}

6.1 Construction of the VRP Term Structure

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

$$
dF_t = F_{t-}\sigma_t dt + \int_{\mathbb{R}} F_{t-}(e^x - 1)[\mu(dx, dt) - \nu_t(x)dxdt],$$  \hspace{1cm} (6.1)

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

$$RV_t(\tau) = \frac{1}{\tau} \int_0^{t+\tau} \sigma_{s-}^2 ds + \frac{1}{\tau} \int_0^{t+\tau} \int_{\mathbb{R}} x^2 \mu(dx, ds).$$  \hspace{1cm} (6.2)

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

$$TV_t(\tau) = \frac{1}{\tau} \int_0^{t+\tau} \sigma_{s-}^2 ds,$$  \hspace{1cm} (6.3)

which measures the continuous variation in the underlying.

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

$$VRP_t(\tau) \equiv \mathbb{E}_t^P[TV_t(\tau)] - \mathbb{E}_t^Q[TV_t(\tau)].$$  \hspace{1cm} (6.4)

Carr and Wu (2009) show that the second term in the right-hand side of Eq. (6.4) is spanned by the $SVS_t(\tau)$ portfolio given in Eq. (4.2) above. That is, $SVS_t(\tau) = \mathbb{E}_t^Q[TV_{t,\tau}]$.

Portfolios of this type have been studied in recent years, but the focus has generally been on short (e.g. $\tau = 30$ day) horizons.\footnote{The notable exceptions are the papers by Aït-Sahalia et al. (2012) and Fusari and Gonzalez-Perez (2012).} A glance at the $SVS_t(\tau)$ portfolio given in Eq. (4.2) above, and the option data counts in Table 3 should reveal why: Beyond $\tau = 90$ days, the availability of option prices to approximate the infinite integral in the SVS portfolio (4.2) drops off significantly. Any $SVS_t(\tau)$ portfolio constructed on the sparse long-run portions of the option surface should therefore be less precise than integrated portfolios constructed from the rich short-run data. But this is exactly what Proposition 4 above contributes: it provides a formal way to quantify the precision of estimates of integrated option portfolios that are constructed from sparse and possibly noisy long-run option data.

This section therefore studies the term structure of sieve-estimated $SVS_t(\tau)$ portfolios and its implications for studying the corresponding term structure of the variance risk premium (VRP). I now turn to estimating the two quantities involved in the construction of the $VRP_t(\tau)$ in Eq. (6.4).
6.2 The Q-Measure: Estimating $\mathbb{E}_t^Q[TV_t(\tau)]$

As discussed above, the $SVS_t(\tau)$ spans $\mathbb{E}_t^Q[TV_t(\tau)]$ and is computed from actual option price data. The option data used are S&P 500 Index options obtained from OptionMetrics for the time period spanning January 1996 to January 2013. The usual filters are applied to the data, i.e. options with zero bid prices are discarded, as are in-the-money options and options with maturity less than a week. This is common practice in the literature in order to mitigate effects arising from price discreteness, liquidity effects, quote staleness, and general microstructure effects. See, for example, Andersen et al. (2012). For the analysis below, I consider the construction of 9 weekly $SVS_t(\tau)$ time series for $\tau = 1, 2, 3, 4, 6, 9, 12, 18,$ and 24 months-to-maturity that use options every Wednesday of the week. While a similar construction of daily or monthly time series is also possible, the weekly frequency strikes a balance between providing a sufficiently rich time series of variance swap term structures while avoiding observations that overlap too strongly, since the $SVS_t(\tau)$ is a forward-looking measure.

Because index options are sparse at both long-run maturities and very short-run maturities, and because the maturities vary from week to week, I use the sieve estimator derived in the previous sections to obtain a balanced time series of estimated SVS term structures. Note that the coefficient solution to Eq. (1.13) uses options across all maturities in one step and does not require second-stage $\hat{SVS}_t(\tau)$ interpolations. That is, for each Wednesday, it uses all available option prices and characteristics $\{P, Z_i\}_{i=1}^n$ and solves the NLLS problem in Eq. (1.13) for BIC-selected $K_y = 6$ and $K_\tau = 2$, yielding a parameter matrix $\hat{B}$, from which the estimated option pricing function $\hat{P}^{K_n}_Y$ is obtained. The call pricing function $\hat{C}^{K_n}_Y$ is obtained by put-call parity, and the week-t $\hat{SVS}_t(\tau)$ term structure is then numerically computed via Eq. (5.3) by evaluating $\hat{P}^{K_n}_Y$ and $\hat{C}^{K_n}_Y$ at $\tau = 1, 2, 3, 4, 6, 9, 12, 18,$ and 24 months.

To compute confidence intervals on any $\hat{SVS}_t(\tau)$, I set $\hat{G}_{K_n} = 2 \frac{e^{-\tau \tau}}{\tau^2} \int_a^b \frac{1}{\sqrt{2\pi K_n}} \partial P^{K_n}_Y(\hat{\beta}_n, Z) d\kappa$ (6.5)

and construct the estimated covariance matrix $\hat{V}_n$ in Eq. (4.7). Note that because $\hat{P}^{K_n}_Y$ is shape-conforming even for strikes that are unobserved, the integral discretization error can be made arbitrarily small. Similarly, because of the sieve-estimator’s adherence to shape-constraints, the integration limits $a$ and $b$ can be set arbitrarily wide. However, to facilitate comparisons with the CBOE’s VIX, I set the integration limits to exclude option prices that fall below 1 cent.

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

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

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

Table 2 displays the sample average of the synthetic variance swap contract in both variance units and US Dollars. Column 3 shows the 95% confidence interval width on the estimated fixed leg $\hat{SVS}_t(\tau)$. The last column of the table shows the proportion of this confidence interval width in relation to the swap’s notional value. For short and very long horizons, the 95% confidence intervals command a sizeable fraction of the swap’s fixed leg; 5.81% for 1-month swaps and up to 8.17% for two-year swaps. The corresponding Dollar amounts for these values are given in columns 4 and 5. Medium horizon swaps, in contrast, appear very well estimated and account for a relatively smaller fraction of the fixed leg payout over the sample period, compared with their short and long-maturity counterparts.

Table 2 suggests that sampling variation can account for about 8% of the observed average payoff on 2-year synthetic variance swaps. This is largely due to the lack of observations on long-maturity options for the first half of the sample, shown in Table 3. If the confidence interval width is interpreted as a gauge of precision for the fair value of the variance swap given observed option information, then this result suggests that available options at long (2-year) horizons relatively less informative hedges of the swap contract’s true value and could therefore receive a premium relative to swaps at more liquid option maturities.

### 6.4 Comparing the Sieve and CBOE VIX

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

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

(6.6)

where $\kappa_0$ is the largest observed strike below the forward price $F$. Note that the last term in (6.6) is zero when options with $\kappa_0 = F$ are observed. Because S&P 500 Index options expire on the third Friday of each month, options expiring exactly 30 days hence are not available in most instances. In such instances, the CBOE takes the two maturities that straddle 30 days, i.e. $\tau_1 < 30$ and $\tau_2 > 30$,  

\[\text{See CBOE Futures Exchange (2013).}\]
and computes the linear interpolation

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

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

It is informative to compare this volatility index with the analogous sieve estimate from Eq. (5.4). Using the above interpolation scheme, I compute a term structure of \( VIX^2_{CBOE}(\tau) \) at fixed horizons \( \tau = 1, 2, 3, 4, 6, 9, 12, 18, \) and 24 months-to-maturity for each date \( t \) in the weekly sample.

[Table 4 about here.]

An unconditional comparison of the resulting sieve and CBOE VIX term structures is given in Table 4, which shows that the \( VIX^2_{CBOE}(\tau) \) term structure is generally lower than the sieve estimator. The difference is negligible at the 1-month (=30 day) horizon (about 3bp on average), but becomes substantial at longer horizons (about 100bp on average). This difference is primarily due to truncation of available strikes, as can be seen by comparing the theoretical formula in Eq. (4.2) with the approximation in Eq. (6.6). While the theoretical formula in Eq. (4.2) extends to infinity for both call and put prices, the approximation in Eq. (6.6) sums only over observed, positive option prices. Implicitly, Eq. (6.6) has set option prices with theoretical strikes beyond observed strikes to zero, biasing down the synthetic variance swap estimate. In contrast, the sieve estimator permits extrapolation into unobserved strikes in a shape-conforming way. Figure 4 illustrates the downward bias of the CBOE VIX for long maturity options by comparing the time series of 30-day to 365-day CBOE and sieve VIX estimates.

[Figure 4 about here.]

The disagreement between the sieve estimate and the long-run CBOE VIX is further quantified in the last column of Table 4, which records the proportion of days in the sample in which the CBOE VIX lies outside the 95% confidence intervals of the sieve VIX. The proportion is clearly largest for long maturity options, suggesting clear differences between the two estimators at long maturities.

[Figure 5 about here.]

Figure 5 shows that indeed, fluctuations in the CBOE VIX from maturity to maturity are larger than the width implied by the sieve confidence intervals. This should come as no surprise: Since the CBOE VIX only uses information from two neighboring maturities, one only needs a single sparsely observed or noisy maturity to cause the CBOE VIX to lose coherence with the CBOE VIX at other maturities on the same day. In contrast, the sieve VIX estimate uses information on all maturities to construct the term structure, which has the effect of downweighting individual poorly observed maturities.
6.5 The Term Structure of Continuous Variance Risk Premia

With estimates $E_Q^t[TV_t(\tau)]$ in hand, the only object needed to compute the variance risk premium is an objective forecast of $E_P^t[TV_t(\tau)]$. I follow Andersen et al. (2003) and model the long-memory properties of realized volatilities as an ARFIMA$(5, 0.4, 0)$ process.\textsuperscript{22} The variance risk premium is then computed as in (6.4).

Table 5 provides summary statistics of the variance risk premium term structure of Eq. (6.4). The average volatility risk premium ranges from $-0.028$ at the 1-month horizon to $-0.036$ at the 2-year horizon. This finding corroborates the results of Aït-Sahalia et al. (2012) and Fusari and Gonzalez-Perez (2012). When the sample is restricted to the financial crisis period from 2007 to 2010, the variance risk premium widens significantly in magnitude and exhibits a downward-sloping term structure ranging from $-0.045$ to about $-0.059$. At the same time, the term structure of variance risk premia is itself more volatile. The decline in skewness and kurtosis and increase in persistence with $\tau$ is also consistent with Aït-Sahalia et al. (2012), who employ a different model and data set to back out a VRP term structure. First-order autocorrelations clearly show that the variance risk premium is most persistent at long horizons.

In economic terms, the magnitudes of the variance risk premium suggest that investors demand significant compensation for bearing return-variance risk and that this compensation must increase with maturity. In turbulent times, the premium is widens to about 1.6 times the average premium over the sample period.

Figure 6 shows that this premium cannot be solely accounted for by sampling variation in option prices. The top two panels show that the variance risk premium, visualized as the gap between the displayed $P$- and $Q$-measure variance term structures, widens with longer maturities. Table 5 suggests that this is standard behavior for generic variance term structures. However, the bottom right panel suggests that on certain high-volatility days, there also appears to be significant uncertainty about the long-run variance swap price itself, although it still cannot account for the entire risk premium.

6.6 Expectation Hypothesis Regressions

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

$$SVS_t(\tau) = E_Q^t[TV_t(\tau)] = E_P^t[m_t(\tau)TV_t(\tau)]$$
$$= E_P^t[TV_t(\tau)] + Cov_P^t[m_t(\tau), TV_t(\tau)],$$

\textsuperscript{22}For brevity, the details of this forecasting model are provided in the Online Appendix.
the null hypothesis of no variance risk premium is equivalent to $H_0 : \text{Cov}_P^T [m_t(\tau), TV_t(\tau)] = 0$, i.e. no covariance between the stochastic discount factor $m_t(\tau)$ and the continuous variation of the market portfolio. That is, under $H_0$, one has $SVS_t(\tau) = \mathbb{E}_P^T [TV_t(\tau)]$, so that for $\varepsilon_t(\tau)$ with $\mathbb{E}_t^P [\varepsilon_t(\tau)] = 0$,

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

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

$$TV_t(\tau) = a(\tau) + b(\tau)SVS_t(\tau) + \varepsilon_t(\tau).$$  \hspace{1cm} (6.8)

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

The results of the regression (6.8) on the weekly sample from 1996-2010 are given in Table 6. Note that the $\hat{b}(\tau)$ is monotonically declining in $\tau$ and uniformly below 1. The joint hypothesis of $a(\tau) = 0 \cap b(\tau) = 1$ is firmly rejected for $\tau = 1$ month, with $p$-values less than 0.000 at all horizons. This result corresponds to the model-based implications of Aït-Sahalia et al. (2012). The results from Table 5, however, suggest that the variance risk-premium behaves differently when conditioning on different volatility regimes. To this end, I also perform the augmented regression

$$TV_t(\tau) = a(\tau) + b(\tau)SVS_t(\tau) + c(\tau)SVS_t(\tau) \times 1_{\{t \text{ is High Volatility Day}\}} + \varepsilon_t(\tau).$$  \hspace{1cm} (6.9)

The interaction of the synthetic variance swap $SVS_t(\tau)$ with a dummy variable that is one during high-volatility periods and zero otherwise allows the slope coefficient on $SVS_t(\tau)$ to change according to volatility regimes. For this exercise, a trading day was considered “high-volatility” if the 30-day VIX exceeded its 67% sample quantile. The results of this regression do not change materially for different cutoffs ranging from 60% to 90% quantiles.\footnote{Above 90% quantiles, the long-maturity regressions samples had relatively few observations to identify $c$.}

The estimates of this augmented expectation hypothesis regressions are given in the bottom panel of Table 6. The results are quite surprising. The magnitudes of swap coefficient are uniformly higher and are significantly closer to one.\footnote{The sole exception is the 2-year maturity regression, which has zero explanatory power.} In particular, the expectation hypothesis cannot be rejected for maturities ranging from 1 to 4 months during normal times, since $\hat{b}(\tau)$ cannot be distinguished from one for these maturities. However, during high volatility periods, the slope coefficient is given by $\hat{b}(\tau) + \hat{c}(\tau)$. The significantly negative sign on $\hat{c}(\tau)$ is evidence of a sizeable risk premium on high-volatility days, which drives the wedge between $P$- and $Q$-measure expected variation. This wedge is not detected for the shorter horizon, medium- to low-volatility days.
The compensation for variance risk in the long-run, however, does not appear as sensitive. The $t$-statistics report that the $\hat{b}(\tau)$ coefficients are significantly different from one, suggesting that the expectation hypothesis is rejected for those horizons even in medium- to low-volatility days. The risk premium widens even for longer maturities on high-volatility days.

7 Conclusion

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

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

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

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


A Appendix: Technical Results and Definitions

A.1 Sobolev Sieve Spaces

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

A.1.1 The Gallant-Nychka Joint Density Spaces

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

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

with $\lambda = (\lambda_1, \ldots, \lambda_{d_u})'$ 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 $\zeta(u)$ by

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

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

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

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

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

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

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 Gallant-Nychka Joint Density Space $\mathcal{F}^{Y,X}$). Let $m$ denote the number of derivatives that characterize the degree of smoothness of the true joint SPD. Then for some integer $m_0 > d_u/2$, some bound $B_0$, some small $\varepsilon_0 > 0$, some $\delta_0 > d_u/2$, and some probability density function $h_0(u)$ with zero mean and $\|h_0\|_{m_0+m,2,\zeta} \leq B_0$, let $\mathcal{F}^{Y,X}$ consist of those probability density functions $f(u)$ with zero mean that have the form

$$f^{Y,X}(u) = h(u)^2 + \varepsilon h_0(u).$$
with \( \|h\|_{m_0+m,2,\zeta_0} \leq \mathcal{B}_0 \) and \( \varepsilon > \varepsilon_0 \), where

\[
\zeta_0(u) = (1 + u'u)^{\delta_0}.
\]

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

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

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

### A.1.2 The Conditional Density Spaces

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

**Definition A.4.** (The Sieve Spaces \( \mathcal{F} \) and \( \mathcal{F}_K \)). Define

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

and

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

This definition says that to each joint density in \( \mathcal{F}^{Y,X} \), one can associate its corresponding conditional density. This association naturally gives rise to map \( \Lambda : \mathcal{F}^{Y,X} \to \mathcal{F} \) with the following continuity property. Note that the densities in \( \mathcal{F} \) are related to the return distribution via the change of variables formula in Eq. (1.3)

### A.2 Intermediate Results

**Lemma A.5.** \( P_X(f_1, Z) = P_X(f_2, Z) \) if and only if \( f_1 = f_2 \) almost everywhere.

**Lemma A.6.** The map \( \Lambda : \mathcal{F}^{Y,X} \to \mathcal{F} \) taking joint densities to their conditional counterparts in \( \mathcal{F} \), i.e. \( \Lambda(f^{Y,X}) = f \), is \( \|f\|_{m,\zeta} - \|\cdot\|_{m,1} \) Lipschitz continuous, where \( f \) is defined pointwise by

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

and where \( \zeta(u) = (1 + u'u)^\delta \) and \( \delta \in (d_u/2, \delta_0) \).
This result formalizes an intuitive notion: when two joint densities in \( F^{Y,X} \) are close, then so are the conditional densities in \( F \). The Lemma provides the Sobolev norms for which this intuition is correct. Furthermore, this property will be used below to regulate the complexity of the space of option pricing functions that are obtained by integrating the option payoff against a candidate from \( F \). Note that the map in Lemma A.6 is also surjective by definition.

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

**Definition A.7.** (Sieve Spaces). The sieve spaces of interest are denoted \( F \equiv \overline{\mathrm{cl}(F^{Y,X})} \) and \( F_Y^X \equiv \overline{\mathrm{cl}(F_Y^X)} \), where \( \mathrm{cl}(\cdot) \) denotes the closure.

The following is a consequence of Lemma A.6.

**Corollary A.8.** There exists a continuous extension of \( \Lambda \) to a mapping \( \Lambda : \mathrm{cl}(F^{Y,Y}) \to \mathrm{cl}(F^{Y|Y}) \), where \( \mathrm{cl}(\cdot) \) denotes the closure.

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

**Lemma A.9.** The sieve spaces \( F_K \) satisfy the following conditions:

(i) \( F_K \) is compact in the topology generated by \( \|\cdot\|_{m,1} \) for all \( K \geq 0 \).

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

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

**Lemma A.10.** Under Assumption 3.1, the option pricing functional \( P_Y(f, Z) \) is

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

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

**B Appendix: Proofs**

**Proof of Lemma 2.1** Let \( \alpha(B, \tau) = (\sum_{j=0}^{K_0} \beta_{0j} H_j(\tau), \ldots, \sum_{j=0}^{K_r} \beta_{K_{0j}} H_j(\tau))' \). Then

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

\[
= \phi(\tau) \int \sum_{k=0}^{K_y} \alpha_k(B, \tau)^2 H_k(y)^2 \phi(y) dy = \phi(\tau) \sum_{k=0}^{K_y} \alpha_k(B, \tau)^2 = \alpha(B, \tau)' \alpha(B, \tau) \phi(\tau),
\]

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

\[
f_K(y|\tau) = \frac{f_K^{X,Z}(y, \tau)}{\int f_K^{X,Z}(y, \tau) dy} = \frac{\sum_{k=0}^{K_y} \alpha_k(B, \tau) H_k(y)}{\alpha(B, \tau)' \alpha(B, \tau) \phi(\tau)} \phi(y)
\]

\[
= \frac{\sum_{k=0}^{2K_y} \alpha(B, \tau)' A_k \alpha(B, \tau) H_k(y) \phi(y)}{\alpha(B, \tau)' \alpha(B, \tau)}
\]

37
where the last equality and the definition of $A_k$ follow by applying Proposition 1 of Leon, Mencia, and Sentana (2009). The result follows.

**Proof of Proposition 1** I follow the derivation of León et al. (2009), which differs due to the conditioning on $\tau$. The plug-in estimator of the population option price in equation (1.4), is given by

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

$$= \kappa e^{-\tau \tau} \int_{-\infty}^{d(Z)} f_K(Y|\tau) dY - S e^{-\tau \tau + \mu(Z)} \int_{-\infty}^{d(Z)} e^{\sigma(Z)Y} f_K(Y|\tau) dY.$$  \hspace{1cm} (B.1)

The integral in the first term becomes

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

$$= \sum_{k=0}^{2K_y} \gamma_k(B, \tau) \int_{-\infty}^{d(Z)} H_k(Y) \phi(Y) dY = \Phi(d(Z)) - \sum_{k=1}^{2K_y} \frac{\gamma_k(B, \tau)}{\sqrt{k}} H_{k-1}(d(Z)) \phi(d(Z)),$$  \hspace{1cm} (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

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

For $k = 0$,

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

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

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

$$= \left[ \frac{1}{\sqrt{k}} e^{\sigma(Z)Y} H_{k-1}(Y) \phi(Y) \right]_{-\infty}^{d(Z)} + \frac{\sigma(Z)}{\sqrt{k}} \int_{-\infty}^{d(Z)} e^{\sigma(Z)Y} H_{k-1}(Y) \phi(Y) dY$$

$$= \frac{1}{\sqrt{k}} e^{\sigma(Z)d(Z)} H_{k-1}(d(Z)) \phi(d(Z)) + \frac{\sigma(Z)}{\sqrt{k}} I_{k-1}^*(d(Z)).$$
Thus,

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

\[
= \sum_{k=0}^{2K_y} \gamma_k(B, \tau) \int_{-\infty}^{d(Z)} e^{\sigma(Z)Y} H_k(Y) \phi(Y) dY = \sum_{k=0}^{2K_y} \gamma_k(B, \tau) I_k^*(d(Z))
\]

\[
= \gamma_0(B, Z) e^{\sigma(Z)^2/2} \Phi(d(z) - \sigma(z)) + \sum_{k=1}^{2K_y} \gamma_k(B, \tau) I_k^*(d(Z))
\]

\[
= 1 \cdot e^{\sigma(Z)^2/2} \Phi(d(z) - \sigma(z)) + \sum_{k=1}^{2K_y} \gamma_k(B, \tau) I_k^*(d(Z))
\]

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.

**Proof of Proposition 2**

Let \( L(f) = \mathbb{E}\{-\frac{1}{2}[P - P_Y(f, Z)]^2 W\} \equiv \mathbb{E}\{\ell(f, Y)\} \), where \( Y \equiv (P, Z) \), and \( W = W(Z) \) is a strictly positive weighting function. \( \ell \) is concave in \( f \), and \( L \) is strictly concave in \( f \). The goal is to estimate the unknown \( P_0^Y(Z) = \mathbb{E}[P|Z] \) by invoking the general sieve consistency theorem in Chen (2007) (i.e. her Theorem 3.1). This requires verification of her Conditions 3.1' - 3.3', 3.4, and 3.5(i), which adapts to the present notation as follows:

**Condition 3.1'.**

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

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

**Condition 3.2'.**

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

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

**Condition 3.3'.**

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

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

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

**Condition 3.5.**

(i) For all \( K \geq 1 \), \( \sup_{f \in \mathcal{F}_K} |L_n(f) - L(f)| = 0 \).
I verify each of these conditions in turn. Condition 3.1': Assumption 3.2 (ii) implies $L(f_0) = 0 > -\infty$. Also, 

$$L(f_0) - L(f) = -\mathbb{E}\left\{\frac{1}{2}[P - P_Y(f_0, Z)]^2W(Z)\right\} + \mathbb{E}\left\{\frac{1}{2}[P - P_Y(f, Z)]^2W(Z)\right\}$$

$$= \frac{1}{2}\mathbb{E}\{[P^2 - 2PP_Y(f, Z) + P_Y(f, Z)^2 - P^2 + 2PP_Y(f_0, Z) - P_Y(f_0, Z)]W(Z)\}$$

$$= \frac{1}{2}\mathbb{E}\{[P_Y(f, Z) - P_Y(f_0, Z)][-2P + P_Y(f, Z) + P_Y(f_0, Z)]W(Z)\}$$

$$= -\mathbb{E}\{[P_Y(f, Z) - P_Y(f_0, Z)][(P - P_Y(f_0, Z)) - \frac{1}{2}(P_Y(f, Z) - P_Y(f_0, Z))]W(Z)\}$$

$$= \frac{1}{2}\mathbb{E}\{[P_Y(f, Z) - P_Y(f_0, Z)]^2W(Z)\}$$

$$= \frac{1}{2}\|P_Y(f, Z) - P_Y(f_0, Z)\|_2^2.$$

As $d(f_n, f_0) \to 0$, the local Lipschitz continuity condition derived in Lemma A.10 implies that the RHS tends to zero, i.e. $L(f_0) - L(f) = |L(f_0) - L(f)| \to 0$. This establishes Condition 3.1'(i). As for Condition 3.1'(ii), note that continuity of $L(f)$ at $f_0$ implies that for any $\eta > 0$, there exists a $\varepsilon > 0$ such that for all $f$ satisfying $d(f, f_0) < \varepsilon$, we have $\|P_Y(f, Z) - P_Y(f_0, Z)\|_2 < \eta$. The contrapositive of this statement reads: Given any $\varepsilon > 0$, there exists $\eta > 0$ such that if $d(f, f_0) \geq \varepsilon$, then $\|P_Y(f, Z) - P_Y(f_0, Z)\|_2 \geq \eta$. 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,

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

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.9 (ii).

Condition 3.3': First note that Chen’s Theorem 3.1 still goes through if we only require $L_n(f)$’s upper semi-continuity to hold almost surely. To this end, observe that Assumption 3.2 (i) implies that $P_t$ is almost surely finite, i.e. $\exists$ a Borel set $\Omega_F$ with $P_t(\omega) < \infty$ for all $\omega \in \Omega_F$; and Assumption 3.1 with no arbitrage imposed implies $P_Y(f, Z_i)$ is locally bounded $\mathbb{P}$ – a.s. on $\mathcal{F}$. Therefore $P_t - P_Y(f, Z_i)$ is finite on $\Omega_F$.

\[\text{To see this, note by Markov’s inequality that } \mathbb{P}(|P_t| > M) \leq \text{Var}(P_t)/M^2. \text{ Applying the Borel-Cantelli Lemma then shows that } P_t \text{ is almost surely finite. See Billingsley (1995).}\]
Next, fix $\omega \in \Omega_F$. Given any sequence $f_j \in \mathcal{F}_K$ with $\|f_j - f\|_{m,1} \to 0$, 

$$|L_n(f_j) - L_n(f)| \leq \frac{1}{n} \sum_{i=1}^{n} \left|[P_Y(f_j, Z_i(\omega)) - P_Y(f, Z_i(\omega))] - \frac{1}{2} (P_Y(f_j, Z_i(\omega)) - P_Y(f, Z_i(\omega))) W(Z_i(\omega))\right|$$

$$\leq \text{const.} \frac{1}{n} \sum_{i=1}^{n} \left\{ \left|P_Y(f_j, Z_i(\omega)) - P_Y(f, Z_i(\omega))\right|^2 W(Z_i(\omega)) \right\}$$

$$+ \left\{ \left|P_Y(f, Z_i(\omega))\right| \left\| f_j - f \right\|_{m,1} \right\}$$

$$\to 0$$

where the last inequality follows from the mean value theorem, and Assumption 3.1 implies that the suprema are bounded for sufficiently large $j$. Hence $L_n(f)$ is almost surely continuous and therefore upper semi-continuous. On the other hand, $L_n(f) = \frac{1}{n} \sum_{i=1}^{n} \left| P_i - P_Y(f, Z_i) \right|^2 W_i$ is continuous in $Z_i$ 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.9 (i).

Condition 3.5(i): Finally, we require the uniform convergence of the empirical criterion $L_n(f) = \frac{1}{n} \sum_{i=1}^{n} \left| P_i - P_Y(f, Z_i) \right|^2 W_i$ over sieves, i.e. for all $K \geq 1$, $\sup_{f \in \mathcal{F}_K} |L_n(f) - L(f)| \overset{L}{\to} 0$ as $n \to \infty$. First, note that by Assumption 3.2 (i) and the law of large numbers, $L_n(f) - L(f) = o_p(1)$ pointwise in $f$ on $\mathcal{F}_K$. Second, standard arguments show

$$\sup_{f \in \mathcal{F}_K} |L_n(f)| \leq \sup_{f \in \mathcal{F}_K} \frac{1}{n} \sum_{i=1}^{n} \left| P_i - P_Y(f, Z_i) \right|^2 W(Z_i)$$

$$\leq \frac{1}{n} \sum_{i=1}^{n} \left| P_i W(Z_i) \right| + \sup_{g \in \mathcal{F}_K} \left| P_Y(g, Z_i) \right| \left( \frac{1}{n} \sum_{i=1}^{n} \left| W(Z_i) \right| \right)$$

$$\leq \left( \frac{1}{n} \sum_{i=1}^{n} \left| P_i \right|^2 \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^{n} \left| W(Z_i) \right|^2 \right)^{1/2} + \sup_{g \in \mathcal{F}_K} \left| P_Y(g, Z_i) \right| \left( \frac{1}{n} \sum_{i=1}^{n} \left| W(Z_i) \right| \right).$$

The first term is $O_p(1)$ by Assumption 3.2 (i). The second term is also $O_p(1)$ by the following arguments. By Lemma A.9 (i), the $\mathcal{F}_K$ are compact. Next, cover each point in $\mathcal{F}_K$ with balls of radius small enough to make the local boundedness Assumption 3.1 hold. By compactness of $\mathcal{F}_K$, there exists a finite subcover $\{U_i\}_{i=1}^{N}$ of $\mathcal{F}_K$ where for each set $U_i$ in the subcover, $\sup_{f \in U_i} S(f, Z) \leq M_i \ P - a.s$. Then $M = \max\{M_1, \ldots, M_N\}$ is a bound on $\sup_{g \in \mathcal{F}_K} \left| P_Y(g, Z_i) \right|$, so the second term in the above display is $O_p(1)$ under Assumption 3.2 (i). Hence, by the mean value theorem, for $f_1, f_2 \in \mathcal{F}_K$, 

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

This Lipschitz condition, the compactness of $\mathcal{F}_K$, and the pointwise convergence of $L_n(f)$ to $L(f)$ mean that the conditions for Corollary 2.2 in Newey (1991) are met, so that $\sup_{h \in \mathcal{H}_K} \left| L_n(h) - L(h) \right| \overset{p}{\to} 0$, as required. Since the conditions for Chen’s Theorem 3.1 are met, we conclude that
Proof of Proposition 3. Recall that the option prices \( P_Y(Z) \) are generated by a conditional
density, i.e. \( P_Y(Z) \equiv P_Y(f, Z) \), 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 \( B_0 \). The complexity of the
space of possible option prices \( \mathcal{P} \) is then firmly linked to the complexity of the Sobolev ball \( \mathcal{H} \). The proof strategy is therefore to establish this link, and then to apply Theorem 3.2 in Chen (2007)
once we have a handle on the complexity of \( \mathcal{P} \).

Application of Theorem 3.2 in Chen (2007) requires verification of her Conditions 3.6, 3.7,
and 3.8, reproduced here for the current notation. It also requires the computation of a certain
bracketing entropy integral, which is undertaken below. Condition 3.6 requires an i.i.d. sample,
which we have already assumed in Assumption 3.2. It remains to check Conditions 3.7 and 3.8 and
to compute the bracketing entropy integral.

Condition 3.7. There exists \( C_1 > 0 \) such that \( \forall \varepsilon > 0 \) small,
\[
\sup_{P_Y \in B_2(P^0_Y)} \text{Var}(\ell(P_Y, Y_i) - \ell(P^0_Y, Y_i)) \leq C_1 \varepsilon^2.
\]

Condition 3.8. For all \( \delta > 0 \), there exists a constant \( s \in (0, 2) \) such that
\[
\sup_{P_Y \in B_2(P^0_Y)} |\ell(P_Y, Y_i) - \ell(P^0_Y, Y_i)| \leq \delta^s U(Y_i),
\]

with \( E[U(Y_i)^\gamma] \leq C_2 \) for some \( \gamma \geq 2 \).

First, note that \( \ell(P_Y, Y_i) - \ell(P^0_Y, Y_i) = W(Z_i)[P_Y(Z_i) - P^0_Y(Z_i)]\{e_i + \frac{1}{2}[P_Y(Z_i) - P^0_Y(Z_i)]\}. \)

Then
\[
E\{|\ell(P_Y, Y_i) - \ell(P^0_Y, Y_i)|^2\} = E\{W(Z_i)^2[P_Y(Z_i) - P^0_Y(Z_i)]^2\{e_i + \frac{1}{2}[P_Y(Z_i) - P^0_Y(Z_i)]\}^2
\]
\[
= E\{W(Z_i)^2[P_Y(Z_i) - P^0_Y(Z_i)]^2 e_i^2\} + E\{\frac{1}{4}W(Z_i)^2[P_Y(Z_i) - P^0_Y(Z_i)]^4\}
\]
\[
= E\{W(Z_i)^2[P_Y(Z_i) - P^0_Y(Z_i)]^2\sigma(Z_i)\} + \frac{1}{4}E\{W(Z_i)^2[P_Y(Z_i) - P^0_Y(Z_i)]^4\}
\]
\[
\leq \text{const.}\|P_Y - P^0_Y\|_2^2 + \frac{1}{4}E\{W(Z_i)^2[P_Y(Z_i) - P^0_Y(Z_i)]^4\}
\]

where the last inequality uses the bound from Assumption 3.3. The second term on the RHS can be
further bounded,
\[
E\{W(Z_i)^2[P_Y(Z_i) - P^0_Y(Z_i)]^4\} \leq C \sup_{Z \in Z} \|P_Y(Z) - P^0_Y(Z)\|^2 W(Z_i)
\]
\[
= C\|P_Y - P^0_Y\|_\infty^4\|P_Y - P^0_Y\|_2^4
\]
The smoothness of \( P_Y \) and \( P^0_Y \) can be used to bound \( \|P_Y - P^0_Y\|_\infty^4 \) as follows. First, let
\[
C^{j,q}(cl(\mathbb{R}^d_u)) = \left\{ f \in C^m(cl(\mathbb{R}^d_u)) : \max_{|\lambda| \leq j} \sup_{u \in \mathbb{R}^d_u} |D^\lambda f(u)| \leq L \right\}
\]
\[
\max_{|\lambda| = j, u_1 \neq u_2} \sup_{u \in \mathbb{R}^d_u} \frac{|D^\lambda f(u_1) - D^\lambda f(u_2)|}{|u_1 - u_2|^q} \leq L
\]
denote a Hölder space. Let \( m = j + k, \eta = 1, k = d_u + 1 \). If the domain \( Z \) satisfies some mild regularity conditions, the Sobolev Embedding Theorem (Theorem 4.12 Adams and Fournier (2003) Part II) implies that \( W^{m,1}(\mathbb{R}^{d_u}) \hookrightarrow C^{\frac{\eta}{n}}(d(\mathbb{R}^{d_u})) \), where \( j = m - k = m - d_u - 1 \geq 1 \) by Assumption 3.5. Thus Assumption 3.5 ensures that \( W^{m,1}(\mathbb{R}^{d_u}) \) can be embedded in a Hölder space consisting of functions that are at least once continuously differentiable and therefore Lipschitz.

Thus Lemma 2 in Chen and Shen (1998) and another appeal to the Sobolev Embedding Theorem of the stock price in Assumption 3.1 as well as the compactness of the sieve space (Lemma A.9).

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

To show Condition 3.8, note that

\[
|\ell(P_Y, Y_i) - \ell(P_Y, Y_i)| = \|P_Y(Z_i) - P_Y^0(Z_i)|\| |e_i + \frac{1}{2}[P_Y^0(Z_i) - P_Y(Z_i)]| \\
\leq \text{const.} \|P_Y - P_Y^0\|_\infty \{ |e_i| + \frac{1}{2} \|P_Y\|_\infty + \frac{1}{2} \|P_Y\|_\infty \}.
\]

The terms involving \( \|P_Y^0\|_\infty \) and \( \|P_Y\|_\infty \) are bounded as a consequence of the local boundedness of the stock price in Assumption 3.1 as well as the compactness of the sieve space (Lemma A.9).\(^{26}\)

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

\[
|\ell(P_Y, Y_i) - \ell(P_Y, Y_i)| \leq \text{const.} \|P_Y - P_Y^0\|_\infty U(Y_i) \\
\leq \text{const.} U(Y_i) \|P_Y - P_Y^0\|_2^{2/(2 + d_u)}
\]

for \( U(Y_i) = |e_i| + \text{const.} \). Thus \( s = 2/(2 + d_u) \) is the required modulus of continuity, and \( \gamma = 2 \) by Assumption 3.2. This establishes Condition 3.8.

An appeal to Chen (2007)'s Theorem 3.2 requires the computation of \( \delta_n \) satisfying

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

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

\[
\mathcal{G}_n = \{ \ell(P_Y, Y_i) - \ell(P_Y^0, Y_i) : \|P_Y - P_Y^0\|_2 \leq \delta, P_Y \in \mathcal{P}_{K_n} \}.
\]

\(^{26}\text{See also the argument in the proof of Proposition 2.}\)
Consider the following chain of inequalities
\[
\|\ell(P_Y, Y_i) - \ell(P^0_Y, Y_i)\| = \|P_Y(Z_i) - P^0_Y(Z_i)\| e_i + \frac{1}{2} [P^0_Y(Z_i) - P_Y(Z_i)] \\
\leq M_1\|f - f_0\|_{m,1} U(Y_i) \\
\leq M_2 U(Y_i) \|f^{Y,X} - f^0_{Y,X}\|_{m,\infty,\zeta_0} \quad \text{by Lemma A.6} \\
= M_2 U(Y_i) \|(h^{Y,X})^2 - (h^0_{Y,X})^2\|_{m,\infty,\zeta_0} \quad \text{by Def. A.2} \\
\leq M_3 U(Y_i) \|(h^{Y,X} - h^0_{Y,X})\|_{m_0 + m_2, \zeta_0} (B.4)
\]

To see the last inequality, observe that
\[
\|(h^{Y,X})^2 - (h^0_{Y,X})^2\|_{m,\infty,\zeta_0} \leq C \|h^{Y,X} + h^0_{Y,X}\|_{m,\infty,\zeta_0} \|h^{Y,X} - h^0_{Y,X}\|_{m,\infty,\zeta_0}^{1/2}
\leq C_1 \|\zeta_0^{1/2}(h^{Y,X} + h^0_{Y,X})\|_{m,\infty} C_2 \|\zeta_0^{1/2}(h^{Y,X} - h^0_{Y,X})\|_{m,\infty}
\leq C_3 \|h^{Y,X} + h^0_{Y,X}\|_{m_0 + m_2, \zeta_0} C_4 \|h^{Y,X} - h^0_{Y,X}\|_{m_0 + m_2, \zeta_0}
\leq C_3 (2B_0) C_4 \|h^{Y,X} - h^0_{Y,X}\|_{m_0 + m_2, \zeta_0}.
\]

for some constants $M_j$ and $C_j$, and where the first inequality follows from Gallant and Nychka (1987) Lemma A.3, the second from Gallant and Nychka (1987) Lemma A.1(d), the third from Gallant and Nychka (1987) Lemma A.1(b), and the fourth by the definition of $\mathcal{H}_n$ as a bounded Sobolev ball.

Theorem 2.7.11 in Van Der Vaart and Wellner (1996) implies that the bracketing number for $\mathcal{G}_n$ can be bounded
\[
N[1](w, \mathcal{G}_n, \|\cdot\|_2) \leq N \left( \frac{w}{2CM_3}, \mathcal{H}_n, \|\cdot\|_{m_0 + m_2, \zeta_0} \right),
\]
where the RHS is by the covering number of a Sobolev ball with dimension $K_n \equiv [K_y(n) + 1][K_x,1(n) + 1] \ldots [K_{x,d_x}(n) + 1]$. By Lemma 2.5 in Van De Geer (2000), we can further bound the RHS, giving
\[
N[1](w, \mathcal{G}_n, \|\cdot\|_2) \leq N \left( \frac{w}{2CM_3}, \mathcal{H}_n, \|\cdot\|_{m_0 + m_2, \zeta_0} \right) \leq \left( 1 + \frac{8B_0 CM_3}{w} \right)^{K_n}.
\]

Therefore,
\[
\frac{1}{\sqrt{n} \delta_n^2} \int_{b_2}^{\delta_n} \sqrt{H[1](w, \mathcal{G}_n, \|\cdot\|_2)} dw \leq \frac{1}{\sqrt{n} \delta_n^2} \int_{b_2}^{\delta_n} \sqrt{K_n} \log \left( 1 + \frac{8B_0 CM_3}{w} \right) dw \leq C \frac{1}{\sqrt{n} \delta_n^2} \sqrt{K_n} \delta_n,
\]

which is less than or equal to a constant for the choice $\delta_n \propto \sqrt{K_n/n}$. Put $K_y(n) \propto K_x,1(n) \propto \cdots \propto K_{x,d_x}(n) \propto n^{1/(2(m_0 + m) + du)}$, so that $K_n \propto n^{du/(2(m_0 + m) + du)}$, yielding
\[
\delta_n \propto \frac{\sqrt{K_n}}{\sqrt{n}} \propto n^{du/[2(2(m_0 + m) + du)]} n^{-1/2} = n^{-(m_0 + m)/(2(2(m_0 + m) + du))}.
\]

On the other hand, this choice of $K_n$ combined with Assumption 3.4 yields the approximation error
rate
\[ \| P_{\gamma}(Z_i) - P_{\gamma}^0(Z_i) \|_2 \leq \text{const.} \| h_{\gamma}^{Y,X} - h_0^{Y,X} \|_{m_0+m,2,\zeta_0} = O(K_n^{-\alpha}) = O\left(n^{2/(m_0+m)+d_k} \right), \]

where the inequality follows from the ones in Eq. (B.4). Applying Chen (2007)’s Theorem 3.2 yields the stated result.

**Proof of Proposition 4**

I verify Assumptions 3.1 - 3.4 of Chen et al. (2013) (CLS). Linearity of \( v \mapsto \frac{\partial \Gamma(P^0)}{\partial P^Y} [v] \) is satisfied for the linear functional \( \Gamma \) in Eq. (4.1), since \( \frac{\partial \Gamma(P^0)}{\partial P^Y}[av_1 + bv_2] = \int \omega(Z)[av_1 + bv_2]dZ_1 = a \int \omega(Z)v_1(Z)dZ_1 + b \int \omega(Z)v_2(Z)dZ_1 = a \frac{\partial \Gamma(P^0)}{\partial P^Y}[v_1] + b \frac{\partial \Gamma(P^0)}{\partial P^Y}[v_2] \). CLS Assumption 3.1(ii) is also trivially satisfied, since \( \Gamma \) is a linear functional. By Assumption 4.1 (iii) and the rate in Proposition 3, we also have \( \|v_n - v^*\| \times \|P_{\gamma}^{0,n} - P_{\gamma}^0\| = o(n^{-1/2}) \). Thus CLS Assumption 3.1 is satisfied.

CLS Assumption 3.2 is directly assumed under our Assumption 4.1 (iv).

CLS Assumption 3.3(i) follows from the linearity of \( \ell'(P^0, \Xi)[av_1 + bv_2] = [P - P^0(Z)]W(Z)(av_1(Z) + bv_2(Z)) = a\ell'(P^0, \Xi)[v_1] + b\ell'(P^0, \Xi)[v_2] \). To show CLS Assumption 3.3(ii), we invoke Lemma 4.2 of Chen (2007). Take

\[
\sup_{\|P_{\gamma} - P_{\gamma}^0\|_2 \leq \delta} \|P_{\gamma}^0(Z) - P_{\gamma}(Z)\| \frac{\partial P_{\gamma}(\beta)}{\partial \beta} R_{K_n}^{-1} G_{K_n} \leq M \delta.
\]

by Assumption 4.1 (ii). Lemma D.1 implies that the entropy integral of Chen (2007) (4.2.2) is satisfied. CLS Assumption 3.3(ii) therefore follows after invoking Lemma 4.2 of Chen (2007). Next, note that CLS Assumption 3.3.(iii) follows by definition of the least squares objective function (see e.g. Shen (1997) Example 1).

Finally, define the empirical process \( \mu \{ g(\Xi) \} = \frac{1}{n} \sum_{i=1}^{n} g(\Xi_i) - \mathbb{E} g(\Xi_i) \), and let \( u_n^* = v_n^*/\|v_n^*\|_{sd} \).

Then
\[
\sqrt{n} \mu \{ \ell'(P^0, \Xi)[u_n^*] \} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell'(P^0, \Xi)[v_n^*] \frac{\mathbb{V}ar(\ell'(P^0, \Xi)[v_n^*])}{\mathbb{V}ar(\ell'(P^0, \Xi)[v_n^*])} = \left( G_{K_n}^t R_{K_n}^{-1} \Sigma_{K_n} R_{K_n}^{-1} G_{K_n} \right) \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell'(P^0, \Xi)[v_n^*] \right) \]
\[
= \left( G_{K_n}^t R_{K_n}^{-1} \Sigma_{K_n} R_{K_n}^{-1} G_{K_n} \right) \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i \frac{\partial P_{\gamma}(\beta)}{\partial \beta} R_{K_n}^{-1} G_{K_n} \right) \]
\[
d \rightarrow N(0, 1)
\]

by continuous mapping theorem and a standard central limit theorem for i.i.d. samples. CLS Assumption 3.4 then follows.

Chen et al. (2013) Theorem 3.1 then implies
\[
\sqrt{n} \Gamma(\hat{P}_{\gamma}) - \Gamma(P^0_Y) \xrightarrow{d} N(0, 1).
\]

45
Next, we want to replace \( \|v_n^*\|_{sd} \) with its estimate

\[
\hat{\lambda} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell'(\hat{P}_Y, \Xi_i) [\hat{v}_n^*] \\
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \ell'(\hat{P}_Y, \Xi_i) [\hat{v}_n^*] - \mathbb{E}[\ell'(\hat{P}_Y, \Xi_i) [\hat{v}_n^*]] - \ell'(P_0^Y, \Xi_i) [v_n^*] + \mathbb{E}[\ell'(P_0^Y, \Xi_i) [\hat{v}_n^*]] \right\} \\
+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \mathbb{E}[\ell'(P_0^Y, \Xi_i) [\hat{v}_n^*]] + \mathbb{E}[\ell'(P_0^Y, \Xi_i) [v_n^*]] - \mathbb{E}[r(P_0^Y, \Xi_i) [\hat{v}_n^*, \hat{P}_Y - P_0^Y]] \right\} \\
+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell'(P_0^Y, \Xi_i) [\hat{v}_n^*] \\
= \hat{I}_1 + \hat{I}_2 + \hat{I}_3 + \hat{I}_4.
\]  

Corollary D.2 (iii) and (iv) implies that \( \hat{I}_1 = o_p(||\hat{v}_n^*||) \) and \( \hat{I}_2 = O_p(\sqrt{n}e_n \varepsilon_n ||\hat{v}_n^*||) \), which makes

\[
\hat{\lambda} = o_p(||\hat{v}_n^*||) + O_p(\sqrt{n}e_n \varepsilon_n ||\hat{v}_n^*||)
+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ (\hat{v}_n^* - v_n^*, \hat{P}_Y - P_0^Y) + (v_n^*, \hat{P}_Y - P_0^Y) \right\} \\
+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell'(P_0^Y, \Xi_i) [v_n^*] - v_n^* ||P_0^Y| v_n^* |.\]

By arguments similar to the proof of CLS Theorem 5.1, we have

\[
\sqrt{n} \|v_n^*\|_{sd}^{-1} (v_n^*, \hat{P}_Y - P_0^Y) = \|v_n^*\|_{sd}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell'(P_0^Y, \Xi_i) [v_n^*] + o_p(1)
\]

and

\[
|\sqrt{n}(\hat{v}_n^* - v_n^*, \hat{P}_Y - P_0^Y)| \leq \sqrt{n}||\hat{v}_n^* - v_n^*|| \|\hat{P}_Y - P_0^Y\| = O_p(\sqrt{n} \varepsilon_n \varepsilon_{n})
\]

and

\[
|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell'(P_0^Y, \Xi_i) [\hat{v}_n^* - v_n^*]| \leq ||\hat{v}_n^* - v_n^*|| \sup_{v \in W_n} |\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell'(P_0^Y, \Xi_i) [v]| = O_p(||v_n^*\| \varepsilon_n^*).\]

46
Since \( \sqrt{n} \epsilon_n \varepsilon_n = o(1) \), we combine these results to obtain

\[
\| \hat{v}_n^* \|_{sd,n}^{sd} = \| v_n^* \|_{sd}^{-1} Var(\hat{\lambda})
\]

\[
= \| v_n^* \|_{sd}^{-1} Var \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell'(P_0^y, \Xi_i)[v_n^*] + o_p(1) \right)
\]

\[
\overset{p}{\Rightarrow} \| v_n^* \|_{sd}^{-1} \| v_n^* \|_{sd} = 1.
\]

Therefore

\[
\frac{\sqrt{n}[\Gamma(\hat{P}_Y) - \Gamma(P_Y)]}{\| \hat{v}_n^* \|_{sd,n}} = \frac{\sqrt{n}[\Gamma(\hat{P}_Y) - \Gamma(P_Y)]}{\| v_n^* \|_{sd,n} \| \hat{v}_n^* \|_{sd,n}} \overset{d}{\rightarrow} N(0,1)
\]

\[ \square \]

**Proof of Lemma A.5** If \( f_1 = f_2 \) a.e., then by definition \( P_X(f_1, Z) = P_X(f_2, Z) \). Conversely, suppose \( P_X(f_1, Z) = P_X(f_2, Z) \). Then differentiating the option price with respect to strike twice yields

\[
e^{rt} \frac{\partial^2 P_X(f_1, Z)}{\partial \kappa^2} |_{\kappa} = e^{rt} \frac{\partial^2 P_X(f_2, Z)}{\partial \kappa^2} |_{\kappa} \implies f_1(\kappa| Z) = f_2(\kappa| Z).
\]

Since this holds for every \( \kappa \), the result follows.

\[ \square \]

**Proof of Lemma A.6** Let \( f_0(x) = \int_{\mathcal{Z}} f^Y_X(y, x) dx \) and \( f_K(x) = \int_{\mathcal{Z}} f^Y_K(y, x) dx \) denote the marginal distributions of \( X \) of generic \( f^Y_X \in \mathcal{F}^Y_X \) and \( f^Y_K \in \mathcal{F}^Y_K \), and let \( g_0(x) = 1/f_0(x) \) and \( g_K(x) = 1/f_K(x) \) denote their reciprocals. In this notation, the conditional densities become \( f_0(y|x) = f^Y_0(y, x)g_0(x) \) and \( f_K(y|x) = f^Y_K(y, x)g_K(x) \). Let \( \mathcal{X} = \mathbb{R}^d \) and \( \mathcal{Y} = \mathbb{R} \).

The goal of the proof is to show that \( \|f_K(y|x) - f_0(y|x)\|_{m,1} \) is small whenever the corresponding joint distribution error \( \|f^Y_K(y, x) - f^Y_0(y, x)\|_{m_0 + m, \infty, \psi} \) is small.

Note first that by definition,

\[
\|f_K(y|x) - f_0(y|x)\|_{m,1} = \sum_{|\lambda| \leq m} \int_{\mathcal{X}} \int_{\mathcal{Y}} |D^\lambda f_K(y|x) - D^\lambda f_0(y|x)| dx dy
\]

\[
= \sum_{|\lambda| \leq m} \|D^\lambda f_K(y|x) - D^\lambda f_0(y|x)\|_{0,1}, \quad (B.6)
\]

so we can focus on the \( \|D^\lambda f_K(y|x) - D^\lambda f_0(y|x)\|_{0,1} \) terms on the RHS.
where as follows. By assumption, the dominated differentiation operator on the third line is due to the dominated between the integration and differentiation, respectively. The interchange

\[
\|D^\lambda f_K(y|x) - D^\lambda f_0(y|x)\|_{0,1} = \int_x \int_y |D^\lambda \{f_K(y|x)\} - D^\lambda \{f_0(y|x)\}| \, dx \, dy \\
= \int_x \int_y |D^\lambda \{f_K(y,x)g_K(x)\} - D^\lambda \{f_0(y,x)g_0(x)\}| \, dx \, dy \\
= \int_x \int_y |D^\lambda \{[f_K(y,x) - f_0(y,x)]g_K(x)\} + D^\lambda \{[g_K(x) - g_0(x)]f_0(y,x)\}| \, dx \, dy \\
\leq \int_x \int_y |D^\lambda \{[f_K(y,x) - f_0(y,x)]g_K(x)\}| \, dx \, dy \\
+ \int_x \int_y |D^\lambda \{[g_K(x) - g_0(x)]f_0(y,x)\}| \, dx \, dy \\
= \|D^\lambda \{[f_K(y,x) - f_0(y,x)]g_K(x)\}\|_{0,1} \\
+ \|D^\lambda \{[g_K(x) - g_0(x)]f_0(y,x)\}\|_{0,1}. \tag{B.7}
\]

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

\[
\|D^\alpha f_K(x) - D^\alpha f_0(x)\|_{0,1} = \int_x |D^\alpha f_K(x) - D^\alpha f_0(x)| \, dx \\
= \int_x |D^\alpha \int_y f^Y_X(x,y) \, dx - D^\alpha \int_y f^Y_X(y,x) \, dx| \, dx \\
= \int_x |D^\alpha \int_y f^Y_X(x,y) \, dx - \int_y D^\alpha f^Y_X(y,x) \, dx| \, dx \\
\leq \int_x |D^\alpha f^Y_X(x,y) \, dx - \int_y D^\alpha f^Y_X(y,x) \, dx| \, dx \\
\leq \|f^Y_X - f^Y_X\|_{m,\infty} \int_{\mathbb{R}^d} (1 + u'\mathbf{u})^{-\delta} \, du,
\]

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

\[
f^Y_X = h^2 + \varepsilon_0 h_0,
\]

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

\[
|\zeta_0(u)^{1/2}h(u)| \leq \max_{|\lambda| \leq m} \sup_{u \in \mathbb{R}^d} |D^\lambda \zeta_0(u)^{1/2}h(u)| \\
= \|\zeta_0^{1/2}h\|_{m,\infty} \\
\leq M_2\|h\|_{m_0+m,2,\zeta_0} \quad \text{by Gallant and Nychka (1987) Lemma A.1(b)} \\
< M_2\mathcal{B}_0.
\]
Therefore,
\[
\zeta_0(u) h(u)^2 \leq (M_2 B_0)^2
\]
\[
h(u)^2 \leq (M_2 B_0)^2 (1 + u'u)^{-\delta} \leq (M_2 B_0)^2 (1 + u'u)^{-\delta}.
\]

Similar reasoning establishes a bound on \( h_0 \), so we have
\[
f^{Y,X} \leq \text{const.}(1 + u'u)^{-\delta},
\]
where the RHS is integrable. By dominated convergence, this establishes the validity of interchanging the differentiation and integration operator in Eq. (B.8). Therefore,
\[
k f_K(x) - f_0(x) \|_{m,1} = \sum_{|\alpha| \leq m} \int_{\mathcal{X}} |D^\alpha \{ f_K(x) \} - D^\alpha \{ f_0(x) \}| dx
\]
\[
= \sum_{|\alpha| \leq m} \| D^\alpha f_K(x) - D^\alpha f_0(x) \|_{0,1}
\]
\[
\leq \text{const.} \| f^{Y,X}_K - f^{Y,X}_0 \|_{m,\infty,\zeta} \int_{\mathbb{R}^{d_u}} (1 + u'u)^{-\delta} du,
\]
which implies that if \( \| f^{Y,X}_K - f^{Y,X}_0 \|_{m,\infty,\zeta} \to 0 \), then \( \| f_K(x) - f_0(x) \|_{m,1} \to 0 \). Next, observe that this type of convergence holds for the reciprocal marginals \( g = 1/f \) too, due to the continuity of the operator \( f \mapsto 1/f \) (since \( f \) has a lower density bound of order \( \varepsilon_0 h_0 \)). Thus, \( \| g_K(x) - g_0(x) \|_{m,1} \to 0 \) as well.

We are now ready to examine the two terms on the RHS of Eq. (B.7). Apply Leibniz’ formula (see Adams and Fournier (2003)) to get
\[
D^\lambda \{ [g_K(x) - g_0(x)] f_0(y, x) \} = \sum_{\beta \leq \lambda} \left[ \frac{\lambda}{\beta} \right] D^{\lambda - \beta} \{ g_K(x) - g_0(x) \} D^\beta f_0(y, x),
\]
so that by the triangle inequality, Hölder’s inequality, and the definitions of Sobolev norms,
\[
\| D^\lambda \{ [g_K(x) - g_0(x)] f_0(y, x) \} \|_{0,1} \leq \sum_{\beta \leq \lambda} \left[ \frac{\lambda}{\beta} \right] \| D^{\lambda - \beta} \{ g_K(x) - g_0(x) \} D^\beta f_0(y, x) \|_{0,1}
\]
\[
\leq \sum_{\beta \leq \lambda} \left[ \frac{\lambda}{\beta} \right] \| D^{\lambda - \beta} \{ g_K(x) - g_0(x) \} \|_{0,1} \| D^\beta f_0(y, x) \|_{0,\infty}
\]
\[
\leq \text{const.} \| g_K(x) - g_0(x) \|_{m,1} \| f_0(y, x) \|_{m,\infty}
\]
\[
\leq \text{const.} \| g_K(x) - g_0(x) \|_{m,1}.
\]

(B.9)
Similarly,

$$
\|D^\lambda \{ [f_K(y,x) - f_0(y,x)]g_K(x) \}\|_{0,1} \leq \sum_{\beta \leq \lambda} \left[ \frac{\lambda}{\beta} \right] \|D^{\lambda-\beta} \{ f_K(y,x) - f_0(y,x) \} D^{\beta} g_K(x) \|_{0,1}
$$

$$
\leq \sum_{\beta \leq \lambda} \left[ \frac{\lambda}{\beta} \right] \|D^{\lambda-\beta} \{ f_K(y,x) - f_0(y,x) \}\|_{0,\infty} \|D^{\beta} g_K(x)\|_{0,1}
$$

$$
\leq \text{const.} \|f_K(y,x) - f_0(y,x)\|_{m,\infty,\xi_0} \|g_K(x)\|_{m,1}
$$

$$
\leq \text{const.} \|f_K(y,x) - f_0(y,x)\|_{m,\infty,\xi_0} \|g_K(x) - g_0(x)\|_{m,1}
$$

$$
\leq \text{const.} \|f_K(y,x) - f_0(y,x)\|_{m,\infty,\xi_0} \|g_0(x)\|_{m,1}
$$

$$
+ \text{const.} \|f_K(y,x) - f_0(y,x)\|_{m,\infty,\xi_0} \|g_0(x)\|_{m,1}.
$$

\tag{B.10}

Plugging Eqs. (B.9) and (B.10) into the RHS of Eq. (B.7), we see that \( D^\lambda f_K(y|x) - D^\lambda f_0(y|x) \|_{0,1} \to 0 \) whenever \( f_K(y,x) - f_0(y,x) \|_{m,\infty,\xi_0} \to 0 \). By Eq. (B.6), this means \( f_K(y,x) - f_0(y|x) \|_{m,1} \to 0 \) whenever \( D^\lambda f_K(y|x) - D^\lambda f_0(y|x) \|_{0,1} \to 0 \), which is the statement of the lemma.

\[ \square \]

**Proof of Corollary A.8** Note that \( cl(\mathcal{F}^{Y|X}) \) is a closed subset of a (complete) Sobolev space and is therefore complete (p. 194 Royden and Fitzpatrick (2010)). In addition, Lemma A.6 shows that \( \Lambda : \mathcal{F}^{Y,X} \to \mathcal{F}^{Y|X} \) is Lipschitz continuous and is therefore uniformly continuous. Therefore, this map has a unique uniformly continuous extension \( \overline{\Lambda} \) from \( \mathcal{F}^{Y,X} \) to \( cl(\mathcal{F}^{Y|X}) \) (p. 196 Royden and Fitzpatrick (2010)). This extension sends \( cl(\mathcal{F}^{Y|X}) \) into \( cl(\mathcal{F}^{Y|X}) \).

\[ \square \]

**Proof of Lemma A.9** Continuity of the \( \overline{\Lambda} \) map above means that the spaces \( \mathcal{F}_K \) inherit the topological properties of their preimages under \( \overline{\Lambda} \). Since Theorem 1 of Gallant and Nychka (1987) says that \( cl(\mathcal{F}^{Y,X}) \) is compact in \( \| \cdot \|_{m,\infty,\xi} \), we have that \( cl(\mathcal{F}^{Y,X}_K) \) are compact as well. By continuity of \( \overline{\Lambda} \), this means that \( \mathcal{F}_K \) is compact in \( \| \cdot \|_{m,1} \), which shows (i). Similarly, Theorem 2 of Gallant and Nychka (1987) shows that \( \bigcup_{K=0}^{\infty} \mathcal{F}^{Y,X}_K \) is a dense subset of \( cl(\mathcal{F}^{Y,X}) \), so \( \bigcup_{K=0}^{\infty} cl(\mathcal{F}^{Y,X}_K) \) is as well. Next, note that the definition of \( \overline{\Lambda} \) says that \( \overline{\Lambda} \) is surjective. Because the image of a dense set is again dense under a continuous surjective map, we have that \( \overline{\Lambda}(\bigcup_{K=0}^{\infty} cl(\mathcal{F}^{Y,X}_K)) = \bigcup_{K=0}^{\infty} \Lambda(cl(\mathcal{F}^{Y,X}_K)) = \bigcup_{K=0}^{\infty} \mathcal{F}_K \) is dense in \( \mathcal{F} \) under \( \| \cdot \|_{m,1} \), showing (ii).

\[ \square \]

**Proof of Lemma A.10** (i) Let \( \varepsilon > 0 \) be given, and fix an \( f_0 \in \mathcal{F} \). Under Assumption 3.1, there exists an \( \| \cdot \|_{m,1} \)-open ball \( B_\delta(f_0) \) of radius \( \delta \) such that

$$
\sup_{f \in B_\delta(f_0)} |S(f, Z)| \leq M \quad \mathbb{P} - a.s.
$$
Then choose \( \eta = \min\{\varepsilon, \delta\}/(2M) \), and consider any \( f \in B_\eta(f_0) \). Using put-call parity,

\[
|P_Y(f, Z) - P_Y(f_0, Z)| = |C_Y(f, Z) - S_0e^{-q\tau} + \kappa e^{-r\tau} - C_Y(f_0, Z) + S_0e^{-q\tau} - \kappa e^{-r\tau}| \\
= |C_Y(f, Z) - C_Y(f_0, Z)| \\
\leq \sup_{g \in (f, f_0)} |\frac{\partial C_Y(g, Z)}{\partial f}| \|f - f_0\|_{m,1} \\
= \sup_{g \in (f, f_0)} |C_Y(g, Z)| \|f - f_0\|_{m,1} \\
\leq \sup_{g \in (f, f_0)} |S_T(g, Z)| \|f - f_0\|_{m,1} \\
\leq M \|f - f_0\|_{m,1} \quad \text{a.s.,} \\
\leq \varepsilon/2
\]

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

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

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

by combining Assumption 3.1 and Assumption 3.2 (i). Choose \( \eta \) similar to (i) above, but depending on const. Then perform the derivations under (i) above, replacing \( |\cdot| \) with \( \|\cdot\|_2 \) to obtain the desired result. \( \Box \)
Figure 1: Shape-conforming option price estimates for multiple maturities. A dense set of true option prices is simulated from the double-jump process in Eq. (5.1) and are plotted in solid for eight maturities. A subset of 250 option prices is drawn from this dense set and perturbed with zero-mean measurement error (round dots). The sieve least squares problem in Eq. (1.13) is solved with BIC-selected $K_x = 6$ and $K_T = 2$ and is plotted (dash).
Figure 2: The term structure of risk-neutral CDFs. A dense set of true option prices is simulated from the double-jump process in Eq. (5.1) and are plotted in solid for eight maturities. A subset of 250 option prices is drawn from this dense set and perturbed with zero-mean measurement error (round dots). The sieve least squares problem in Eq. (1.13) is solved with BIC-selected $K_z = 6$ and $K_s = 2$. Eq. (5.2) is then evaluated at the estimated coefficient matrix $\hat{B}$ (dash).
Figure 3: The term structure of risk-neutral PDFs. A dense set of true option prices is simulated from the double-jump process in Eq. (5.1) and are plotted in solid for eight maturities. A subset of 250 option prices is drawn from this dense set and perturbed with zero-mean measurement error (round dots). The sieve least squares problem in Eq. (1.13) is solved with BIC-selected $K_x = 6$ and $K_z = 2$. Eq. (2.1) is then evaluated at the estimated coefficient matrix $\hat{B}$ (dash).
Figure 4: Short- and long-maturity time series of the sieve VIX from Eq. (5.4) and the CBOE VIX from Eq. (6.6).
Figure 5: Term Structures for the sieve VIX and the CBOE VIX. The sieve VIX 95% confidence intervals of the estimate in Eq. (5.4) are plotted alongside the CBOE VIX approximations from Eq. (6.6) for four sample days with $\tau$ ranging from 1 month to 24 months.
Figure 6: Term Structures for High-, Medium-, and Low-Volatility Days. Trading days are sorted by 1-month volatility as measured by the synthetic variance swap, SVS. A high-volatility day is chosen to exceed the 95th percentile of all 1-month SVS values, and a low-volatility trading day is chosen to lie below the 5th percentile of 1-month VIX values. Q-measure SVS 95% confidence intervals and P-measure forecasts of truncated variation, $E_t^a[TV_t(\tau)]$ are plotted for $\tau$ ranging from 1 month to 24 months.
Table 1: Simulated Coverage Probabilities Given Affine Jump-Diffusion DGP. A dense set of true option prices is simulated from the double-jump process in Eq. (5.1) with parameters from Table F.1, from which a true \( \text{VIX}(\tau) \) for each of the \( \tau \) given in the displayed horizons is computed. Then a sample of 250 option prices is drawn at random from this dense set and perturbed with zero-mean measurement error. The sieve least squares problem in Eq. (1.13) is solved with different choices of expansion terms \( K_x \) and \( K_\tau \). The estimated variance from Proposition 4 is then computed to construct 95% confidence intervals around the estimated \( \text{VIX}(\tau) \). The process of drawing 250 options prices and computing estimated \( \hat{\text{VIX}}(\tau) \) and its confidence intervals is repeated 1,000 times, and the proportion of occasions on which the true \( \text{VIX}(\tau) \) lies inside the estimated 95% confidence intervals is then recorded at each horizon \( \tau \). Shading denotes BIC selection.

<table>
<thead>
<tr>
<th>Horizon (months)</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>9</th>
<th>12</th>
<th>18</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K_y = 3, K_\tau = 1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heston</td>
<td>0.939</td>
<td>0.946</td>
<td>0.910</td>
<td>0.915</td>
<td>0.943</td>
<td>0.948</td>
<td>0.941</td>
<td>0.873</td>
</tr>
<tr>
<td>SVJ</td>
<td>0.069</td>
<td>0.040</td>
<td>0.434</td>
<td>0.886</td>
<td>0.909</td>
<td>0.802</td>
<td>0.853</td>
<td>0.908</td>
</tr>
<tr>
<td>SVJJ</td>
<td>0.076</td>
<td>0.031</td>
<td>0.039</td>
<td>0.056</td>
<td>0.317</td>
<td>0.706</td>
<td>0.861</td>
<td>0.840</td>
</tr>
<tr>
<td>( K_y = 6, K_\tau = 2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heston</td>
<td>0.930</td>
<td>0.935</td>
<td>0.960</td>
<td>0.923</td>
<td>0.929</td>
<td>0.916</td>
<td>0.919</td>
<td>0.888</td>
</tr>
<tr>
<td>SVJ</td>
<td>0.971</td>
<td>0.941</td>
<td>0.962</td>
<td>0.944</td>
<td>0.909</td>
<td>0.897</td>
<td>0.916</td>
<td>0.890</td>
</tr>
<tr>
<td>SVJJ</td>
<td>0.961</td>
<td>0.940</td>
<td>0.918</td>
<td>0.941</td>
<td>0.863</td>
<td>0.876</td>
<td>0.878</td>
<td>0.835</td>
</tr>
<tr>
<td>( K_y = 7, K_z = 2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heston</td>
<td>0.958</td>
<td>0.945</td>
<td>0.913</td>
<td>0.942</td>
<td>0.955</td>
<td>0.941</td>
<td>0.908</td>
<td>0.867</td>
</tr>
<tr>
<td>SVJ</td>
<td>0.970</td>
<td>0.968</td>
<td>0.922</td>
<td>0.931</td>
<td>0.960</td>
<td>0.934</td>
<td>0.913</td>
<td>0.883</td>
</tr>
<tr>
<td>SVJJ</td>
<td>0.934</td>
<td>0.907</td>
<td>0.915</td>
<td>0.939</td>
<td>0.909</td>
<td>0.915</td>
<td>0.921</td>
<td>0.852</td>
</tr>
</tbody>
</table>
Table 2: Variance swap confidence intervals and ex-post payoffs for the sample period 1996-2010. The synthetic variance swap term structure $\mathcal{S}V_{S_t}(\tau)$ along with 95% confidence intervals is estimated using the sieve methods derived in Sections 2-4 in the text. The $P$-measure ex-post realized analog is computed by truncating jumps from 5-minute return data according to Eq. (F.1). The payoff $(TV_t - \mathcal{S}V_{S_t})$ of a hypothetical long position in the continuous variance swap is reported. Profit and loss in US dollars are computed using the variance notional given in the text.

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Variance $SV_t$ 95%-CI Range</th>
<th>USD $SV_t$ 95%-CI Range</th>
<th>Percentage CI Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.043 0.002</td>
<td>979,692.22 59,036.48</td>
<td>5.81</td>
</tr>
<tr>
<td>2</td>
<td>0.045 0.002</td>
<td>1,009,856.95 39,873.55</td>
<td>4.00</td>
</tr>
<tr>
<td>3</td>
<td>0.045 0.002</td>
<td>1,023,791.66 33,843.16</td>
<td>3.36</td>
</tr>
<tr>
<td>4</td>
<td>0.046 0.001</td>
<td>1,030,536.66 32,237.58</td>
<td>3.19</td>
</tr>
<tr>
<td>6</td>
<td>0.046 0.001</td>
<td>1,035,921.87 26,873.87</td>
<td>2.66</td>
</tr>
<tr>
<td>9</td>
<td>0.046 0.001</td>
<td>1,038,554.44 23,017.43</td>
<td>2.23</td>
</tr>
<tr>
<td>12</td>
<td>0.046 0.001</td>
<td>1,039,254.55 23,691.63</td>
<td>2.27</td>
</tr>
<tr>
<td>18</td>
<td>0.045 0.001</td>
<td>1,037,954.92 30,035.26</td>
<td>2.82</td>
</tr>
<tr>
<td>24</td>
<td>0.045 0.004</td>
<td>1,035,319.83 87,291.28</td>
<td>8.17</td>
</tr>
</tbody>
</table>

Table 3: Option Cross-Section Average Counts. Option data spanning January 1996 to January 2013 are obtained from OptionMetrics. The number of available option prices for each displayed category is averaged over time.

<table>
<thead>
<tr>
<th>Maturity Range</th>
<th>Number of Maturities</th>
<th>Number of Options</th>
<th>Number of Maturities</th>
<th>Number of Options</th>
<th>Number of Maturities</th>
<th>Number of Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 90</td>
<td>3.3</td>
<td>170.6</td>
<td>2.7</td>
<td>82.4</td>
<td>3.8</td>
<td>266.8</td>
</tr>
<tr>
<td>90 - 180</td>
<td>1.6</td>
<td>58.3</td>
<td>1.0</td>
<td>31.6</td>
<td>2.2</td>
<td>87.5</td>
</tr>
<tr>
<td>180 - 270</td>
<td>1.3</td>
<td>41.9</td>
<td>1.0</td>
<td>27.2</td>
<td>1.7</td>
<td>57.9</td>
</tr>
<tr>
<td>270 - 360</td>
<td>1.3</td>
<td>39.1</td>
<td>1.0</td>
<td>24.0</td>
<td>1.7</td>
<td>55.6</td>
</tr>
<tr>
<td>360 - 450</td>
<td>0.5</td>
<td>17.1</td>
<td>0.5</td>
<td>12.9</td>
<td>0.6</td>
<td>21.7</td>
</tr>
<tr>
<td>450 - 540</td>
<td>0.5</td>
<td>15.1</td>
<td>0.5</td>
<td>10.9</td>
<td>0.5</td>
<td>19.7</td>
</tr>
<tr>
<td>540 - 630</td>
<td>0.5</td>
<td>13.7</td>
<td>0.5</td>
<td>10.1</td>
<td>0.5</td>
<td>17.8</td>
</tr>
<tr>
<td>630 - 720</td>
<td>0.5</td>
<td>12.4</td>
<td>0.5</td>
<td>8.1</td>
<td>0.5</td>
<td>17.1</td>
</tr>
<tr>
<td>720 - $\infty$</td>
<td>0.5</td>
<td>16.4</td>
<td>0.0</td>
<td>0.1</td>
<td>0.9</td>
<td>34.3</td>
</tr>
<tr>
<td>Totals</td>
<td>10.0</td>
<td>384.6</td>
<td>7.8</td>
<td>207.3</td>
<td>12.4</td>
<td>578.2</td>
</tr>
</tbody>
</table>
Table 4: The CBOE and Sieve VIX Term Structures from 1996 to 2013. VIX term structures and corresponding confidence intervals are obtained for each Wednesday of the sample using the sieve estimator from the main text as well as the CBOE’s discrete approximation and linear interpolation procedure.

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Mean CBOE</th>
<th>Mean Sieve</th>
<th>Difference (1) − (2)</th>
<th>Std. Dev. CBOE</th>
<th>Std. Dev. Sieve</th>
<th>Difference (4) − (5)</th>
<th>Fraction Days Significantly Different</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21.3</td>
<td>21.3</td>
<td>−0.03</td>
<td>8.4</td>
<td>8.2</td>
<td>0.18</td>
<td>0.56</td>
</tr>
<tr>
<td>2</td>
<td>21.7</td>
<td>21.9</td>
<td>−0.18</td>
<td>7.6</td>
<td>7.8</td>
<td>−0.13</td>
<td>0.70</td>
</tr>
<tr>
<td>3</td>
<td>22.0</td>
<td>22.3</td>
<td>−0.31</td>
<td>7.5</td>
<td>7.4</td>
<td>0.11</td>
<td>0.74</td>
</tr>
<tr>
<td>4</td>
<td>22.3</td>
<td>22.6</td>
<td>−0.26</td>
<td>7.2</td>
<td>7.2</td>
<td>−0.03</td>
<td>0.70</td>
</tr>
<tr>
<td>6</td>
<td>22.2</td>
<td>22.8</td>
<td>−0.58</td>
<td>6.4</td>
<td>6.9</td>
<td>−0.52</td>
<td>0.74</td>
</tr>
<tr>
<td>9</td>
<td>22.0</td>
<td>23.0</td>
<td>−0.96</td>
<td>6.1</td>
<td>6.6</td>
<td>−0.45</td>
<td>0.82</td>
</tr>
<tr>
<td>12</td>
<td>21.9</td>
<td>23.1</td>
<td>−1.16</td>
<td>6.1</td>
<td>6.4</td>
<td>−0.37</td>
<td>0.82</td>
</tr>
<tr>
<td>18</td>
<td>22.3</td>
<td>23.2</td>
<td>−0.88</td>
<td>6.3</td>
<td>6.3</td>
<td>0.05</td>
<td>0.84</td>
</tr>
<tr>
<td>24</td>
<td>22.2</td>
<td>23.3</td>
<td>−1.15</td>
<td>6.7</td>
<td>6.3</td>
<td>0.41</td>
<td>0.74</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Maturity</th>
<th>1996-2010</th>
<th></th>
<th>2007-2010</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. dev.</td>
<td>Skew</td>
<td>Kurt</td>
</tr>
<tr>
<td>1</td>
<td>-0.028</td>
<td>0.036</td>
<td>-5.10</td>
<td>39.50</td>
</tr>
<tr>
<td>2</td>
<td>-0.031</td>
<td>0.035</td>
<td>-4.19</td>
<td>28.70</td>
</tr>
<tr>
<td>3</td>
<td>-0.032</td>
<td>0.034</td>
<td>-3.72</td>
<td>23.65</td>
</tr>
<tr>
<td>4</td>
<td>-0.033</td>
<td>0.033</td>
<td>-3.42</td>
<td>20.58</td>
</tr>
<tr>
<td>6</td>
<td>-0.034</td>
<td>0.032</td>
<td>-3.02</td>
<td>16.66</td>
</tr>
<tr>
<td>9</td>
<td>-0.035</td>
<td>0.031</td>
<td>-2.61</td>
<td>13.12</td>
</tr>
<tr>
<td>12</td>
<td>-0.035</td>
<td>0.030</td>
<td>-2.35</td>
<td>11.14</td>
</tr>
<tr>
<td>18</td>
<td>-0.035</td>
<td>0.029</td>
<td>-2.04</td>
<td>8.88</td>
</tr>
<tr>
<td>24</td>
<td>-0.036</td>
<td>0.030</td>
<td>-1.85</td>
<td>7.50</td>
</tr>
</tbody>
</table>
Table 6: Expectation Hypothesis Regressions. The OLS regressions $TV_t(\tau) = a(\tau) + b(\tau)SV_t(\tau) + \varepsilon_t(\tau)$ from Eq. (6.8) of realized continuous variation on synthetic variance swaps are estimated for each of the horizons $\tau = 1, 2, 3, 4, 6, 9, 12, 18, \text{and } 24$ and are reported in the top panel. The same regressions are augmented in Eq. (6.9) to incorporate conditioning information on volatility, i.e. $TV_t(\tau) = a(\tau) + b(\tau)SV_t(\tau) + c(\tau)SV_t(\tau) \times \mathbf{1}_{\text{t is High Volatility Day}} + \varepsilon_t(\tau)$, and are reported in the bottom panel. $t$-statistics on $a(\tau)$ and $c(\tau)$ are centered at 0, whereas the $t$-statistic on $b(\tau)$ is centered at 1. $p$-values report the outcome of the joint tests $a(\tau) = 0 \cap b(\tau) = 1$.

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

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