Sieve Estimation of Option-Implied State Price Density*

Junwen Lu† Boston University

Zhongjun Qu‡ Boston University

March 4, 2021

Abstract

This paper proposes a nonparametric estimator for the state price density implied by a cross-section of European options with different strikes and the same maturity. The proposed estimator has two distinctive features. First, it extracts information from both call and put options, as opposed to only call options. Second, it does not require estimating any second-order derivative. Instead, it is obtained as the solution to a constrained and penalized linear regression. The technical analysis faces two challenges because the state price density is defined by a Fredholm integral equation of the first kind with an unbounded support, and the kernel function is unbounded and non-differentiable. We address these challenges by exploiting the structure of the option pricing problem. After establishing the consistency and the convergence rate of the estimator, we apply it to estimate the state price densities implied by the S&P500 index options and those by the VIX options. The sample period includes the recent financial crisis and the Great Recession, during which the turbulent market conditions imposed substantial challenges on the estimation. We show that the procedure can work with both daily and high-frequency observations. We also study whether the various features of this density can predict future asset returns and obtain positive findings. Finally, we apply the method to examine the causal effects of monetary policy announcements on the market, using high-frequency data.

Keywords: state price density, high-frequency data, inverse problem, Tikhonov regularization.
JEL codes: C14, C22, G13.

*We thank the co-editor (Serena Ng), two anonymous referees, and the seminar and conference participants at CIREQ-Concordia, Syracuse, CEMMAP UCL/Vanderbilt Joint Conference (2019), and NBER-NSF Time Series Conference (2017) for their helpful comments and suggestions.

†Department of Economics, Boston University, Boston, MA, 02215 (junwenl@bu.edu).

‡Department of Economics, Boston University, Boston, MA, 02215 (qu@bu.edu).

1 Introduction

The state price density (SPD) corresponds to a probability distribution under which expected returns on relevant financial assets are equal to the risk-free rate. The SPD exists if the market is arbitrage-free, and it is unique if the market is also complete (Harrison and Kreps, 1979). All else equal, its value in a density region increases if (i) the market assigns a higher probability for the relevant states to occur, or (ii) the market assigns a higher price to the realized gains in these states. Consequently, the change in this density can reveal how the financial market reacts dynamically to policy interventions and economic fundamentals. The recent financial crisis calls for a better understanding of these reactions. It also generates valuable data to study them.

The options market provides a natural laboratory for studying the dynamics of the SPD. Suppose there is a continuum of strikes over a sufficiently wide range in the absence of arbitrage. Then, it is possible to recover this density from a cross-section of European option prices and the risk-free rate only. This identification result, attributed to Cox and Ross (1976) and Breeden and Litzenberger (1978), has inspired a sizable literature to study SPD estimation. The current paper makes three contributions. First, it proposes a nonparametric estimator for the SPD. Second, it applies this estimator to recover the SPDs implied by the S&P500 index options and those by the VIX options. Third, it examines the predictive power of these densities for future asset prices and presents positive findings. For the last two contributions, the sample period includes the recent financial crisis and the Great Recession, during which the turbulent market environments imposed substantial challenges for the estimation.

The proposed estimator has four features: (a) It is based on the method of sieves of Grenander (1981). The SPD is approximated by a sequence of models, where the model complexity increases progressively with sample size. (b) The estimator extracts information from both put and call options, as opposed to only call options. This feature is essential for estimating both tails of the density. (c) The method does not apply smoothing (i.e., averaging) over time. As a result, the estimator is advantageous in situations where the density can jump abruptly by large magnitudes, for example, during an economic crisis, after a significant market event, or following a monetary policy announcement. We document such abrupt changes in our empirical application. The disadvantage is that the estimate no longer directly quantifies the relation between the SPD and relevant state variables. It sacrifices efficiency if the density evolved smoothly as a function of the correctly identified state variables. (d) The estimator is the solution to a penalized constrained linear regression, where the optimization problem is convex and easy to solve.

For comparison, it is informative to separate the existing estimation methods into parametric and nonparametric methods. The parametric methods assume that the SPD, or the stochastic process of the underlying asset, belongs to a parametric family. For example, Jarrow and Rudd

(1982), Longstaff (1995), Abken, Madan and Ramamurtie (1996), Backus, Foresi, and Wu (2004), and Beber and Brandt (2006) assume that the SPD has a four-parameter Gram-Charlier series representation, with the four parameters capturing the mean, variance, skewness, and kurtosis of the distribution. Melick and Thomas (1997) assume that the future price of the underlying asset at the option's expiration is drawn from a mixture of three lognormal distributions. Bates (2000) proposes a two-factor stochastic volatility model with jumps for the underlying asset price, showing that the model fits the S&P 500 futures options better than without jumps. The parametric methods are often parsimonious, with some of them well equipped at identifying relevant state variables. However, constrained by the parametric assumptions, they may lack flexibility and suffer from misspecification, especially when market conditions are rapidly changing, such as during a financial crisis or after a major policy intervention. Therefore, to obtain an unbiased view of the SPD, it is critical to confront such procedures with nonparametric estimates that do not require imposing tight parametric structures.

The existing nonparametric methods typically estimate the density in two steps. First, a call price function (or an implied volatility function) is estimated by kernel, polynomial, spline, or nonparametric least square methods. Then, the second-order derivative of the call price function with respect to the strike price is computed and used as an estimate for the SPD. Related methodological contributions include Aït-Sahalia and Lo (1998), based on a kernel estimator; Aït-Sahalia and Duarte (2003), based on constrained local polynomial estimators; Yatchew and Härdle (2006), using nonparametric least squares; and Shimko (1993) and Figlewski (2010), using global polynomial and spline methods, respectively. Some of these methods are applied or extended in, among others, Li and Zhao (2009), Birru and Figlewski (2012), Kitsul and Wright (2013), and Song and Xiu (2016).

Unlike the aforementioned nonparametric methods, our proposed SPD estimator is obtained by running a constrained linear regression without computing any derivatives. Hermite functions generate the regressors, and the constraints help regulate the estimate. In practice, the user can easily include both call and put options in the estimation or discard some observations based on transaction volume or moneyness. Also, because the estimate is over the real line, there is no need to separately estimate the tails, as in Figlewski (2010). The binomial tree method proposed by Rubinstein (1994) and Jackwerth and Rubinstein (1996) also does not require computing derivatives; its differences from the proposed method are in terms of (a), (b), and (d), described above.

Our conceptual framework treats the SPD estimation problem as an inverse problem involving linear integral equations. The general literature on linear integral equations is quite rich (see Kress, 2014 for an introductory analysis); however, specific applications to option pricing are limited. Du, Wang, and Du (2012) is an exception, which explicitly formulates the SPD estimation as a linear inverse problem. However, unlike our approach, they do not introduce any basis function for the

approximation. Instead, they discretize the integral and penalize the roughness (i.e., the second-order derivative) of the estimator, arriving at a procedure that resembles the smoothing spline estimation. They do not study the theoretical properties of the resulting estimator.

We illustrate the proposed method using both daily and high-frequency (e.g., minute-by-minute) observations. Daily data are usually available for an extended period, which makes them ideal for examining the SPD dynamics at business cycle frequencies. High-frequency data tend to be more limited in their span; however, they contain valuable information about the immediate impacts of monetary policies and other influential events on the financial market. We show that the method is suitable for both types of applications.

The SPD is forward-looking. To explore this, we run predictive regressions for asset returns, using various quantiles of the density as the predictor. The sample comprises daily observations from January 2007 to April 2016, spanning the recent financial crisis and recovery. We also consider a subsample excluding January 2007 – May 2009 to evaluate the impact of the financial crisis on the predictive analysis. The predictive horizon is equal to one month (similar results hold for the two-month horizon; see the online appendix). We find that the upper quantiles of the SPD have significant predictive power for the corresponding quantiles of the S&P 500 returns in both samples. For lower quantiles, the results are mixed, with the predictive power being mostly absent. This finding of predictability is somewhat expected because the return volatility is strongly predictable from the implied volatility. Still, the asymmetry between upper and lower quantiles suggests a more nuanced picture that merits further consideration. We also find that the quantiles of the SPD can predict the mean of the realized return in the subsample, but not the full sample. Since the predictors exhibit only mild serial dependence, these regressions do not suffer from the Stambaugh bias as when the dividend yield is the predictor. To our knowledge, the above results are new to the predictive regression literature.

Motivated by the finding of predictive power, we suggest presenting a sequence of SPDs as a way to reflect the market's perception (its prediction and the associated risk premium) of the future. This proposal is inspired by the "River of Blood" measure of the Bank of England, which is displayed as a fan chart to convey the central bank's perception about future inflation.

Finally, to illustrate the method's potential for causal analysis, we consider two FOMC announcements made on December 18, 2013, and January 27, 2016. We estimate the SPDs within a three-hour window surrounding the announcement using high-frequency data to trace out the dynamic causal effects. In each case, we find that the SPD remained mostly unchanged until the announcement, reacted immediately at the news, and then continued to evolve over the next 90-minute interval. Beber and Brandt (2006) documented, using the parametric estimator of Jarrow and Rudd (1982), that regularly scheduled macroeconomic announcements had effects on the

skewness and kurtosis of the SPD implied by Treasury bond futures options. The current paper illustrates that similar analyses can be carried out using our nonparametric estimator. We conjecture that the proposed method can be useful for identifying dynamic causal effects in other contexts, owing to its suitability for high-frequency data.

The remainder of the paper is organized as follows. Section 2 presents the estimator focusing on its implementation. Section 3 studies the identification and the asymptotic properties of the estimator, while Section 4 examines its finite sample properties. Section 5 presents empirical applications related to the S&P500 and the VIX options. Section 6 offers the conclusion. All proofs are included in the online appendix.

2 The estimator

We consider European options on a common underlying asset with the same maturity date denoted by T. As shown by Harrison and Kreps (1979), in the absence of arbitrage and market imperfection, there exists a state price density $f_t^*(\cdot)$ for the payoff at time T, such that the prices of the call and put options at time t satisfy

$$C_t(K) = e^{-r_t\tau} \int_0^\infty (S_T - K)^+ f_t^*(S_T) dS_T \tag{1}$$

and

$$P_t(K) = e^{-r_t\tau} \int_0^\infty (K - S_T)^+ f_t^*(S_T) dS_T, \tag{2}$$

respectively, where K is the strike price, $\tau = T - t$, $(S_T - K)^+ = \max(S_T - K, 0)$, r_t is the risk-free rate at time t, and S_T denotes the potential price of the underlying asset at time T. Our framework can allow for a time-varying risk-free rate r_t; all subsequent analysis goes through after replacing $r_t\tau$ with $\int_t^T r_s ds$. The SPD leads to Arrow-Debreu prices when the number of states is finite.

The proposed estimator of $f_t^*(\cdot)$ does not apply smoothing over t. As a result, we do not specify the state variables affecting $f_t^*(S_T)$. To reflect this feature, and to simplify the notation, we will suppress the subscript t in subsequent analysis. In particular, we will write $C_t(K)$, $P_t(K)$, r_t, and $f_t^*(S_T)$ as $C(K)$, $P(K)$, r, and $f^*(S_T)$, respectively. Accordingly, (1) and (2) will be expressed as $C(K) = e^{-r\tau} \int_0^\infty (S_T - K)^+ f^*(S_T) dS_T$ and $P(K) = e^{-r\tau} \int_0^\infty (K - S_T)^+ f^*(S_T) dS_T$, respectively.

2.1 Two issues in achieving effective approximations

Before introducing the estimation procedure, we consider two issues that are important for a sieve estimator of $f^*(\cdot)$ to perform well in practice.

The first issue arises because the SPD is a dynamic object. Its dispersion differs substantially between asset classes, and it varies significantly depending on, among other factors, market

volatility, economic outlook, and time to maturity. Its dispersion eventually decreases to zero as t approaches T. Because we aim to develop an estimator for general applications, we need a method to standardize the density or the data before approximating the SPD. The usual standardization method (i.e., dividing the data by their standard deviation) does not work here because we do not observe direct realizations from the distribution associated with $f^*(\cdot)$.

To address this issue, we apply the following change of variables:

$$x = \frac{\log(S_T/S) - r\tau}{\sigma\sqrt{\tau}} \tag{3}$$

and

$$z = \frac{\log(K/S) - r\tau}{\sigma\sqrt{\tau}}. \tag{4}$$

where S is the spot price of the underlying asset at time t, r and τ are defined in (1) and (2), and σ is equal to the Black-Scholes implied volatility for at-the-money call option of the same underlying asset and maturity. Because the implied volatility is always reported as a summary statistic in options data, these two transformations are straightforward to compute. We denote the density after the change of variables by $f(x)$. Then, (1) and (2) can be rewritten as

$$C(z) = \int_{-\infty}^{\infty} S\left(e^{\sqrt{\tau}\sigma x} - e^{\sqrt{\tau}\sigma z}\right)^+ f(x)dx \tag{5}$$

and

$$P(z) = \int_{-\infty}^{\infty} S\left(e^{\sqrt{\tau}\sigma z} - e^{\sqrt{\tau}\sigma x}\right)^+ f(x)dx. \tag{6}$$

We propose to approximate $f(\cdot)$ using a suitable set of basis functions. After estimating $f(\cdot)$, we recover $f^*(\cdot)$ by reversing the monotonic transformation (3), that is, by computing

$$f^*(S_T) = \frac{1}{\sigma\sqrt{\tau}S_T} f\left(\frac{\log(S_T/S) - r\tau}{\sigma\sqrt{\tau}}\right). \tag{7}$$

The SPD of the return $R_T = \log(S_T/S)$ is also straightforward to obtain, given by

$$\frac{1}{\sigma\sqrt{\tau}} f\left(\frac{R_T - r\tau}{\sigma\sqrt{\tau}}\right).$$

The transformations (3) and (4) are not new. For example, (4) is often used as a measure of moneyness to indicate how many standard deviations the option is in- or out-of-the-money; see, for example, Carr and Wu (2003) and Beber and Brandt (2006). However, their role in the current context is different, achieving the following two goals at once. First, the division by$\sqrt{\tau}\sigma$ serves as a variance-stabilizing transformation to ensure that the dispersion of $f(x)$ is insensitive to changes in market conditions and the time to maturity. In particular, if the dispersion of $f^*(x)$ increases abruptly due to a large negative shock, the Black-Scholes volatility will rise instantly, and the dispersion of $f(x)$ will remain relatively unchanged. Because σ is computed by inverting the

Black-Scholes formula, it is not treated as an unknown parameter in the subsequent estimation. Consequently, the estimation problem remains linear in parameters after the transformation. Second, the logarithmic transformation $\log(S_T/S)$ shifts the density's support from the positive axis to the real line. If $f^*(x)$ is close to the log-normal density, $f(x)$ will be close to the normal density. This feature is useful in addressing the second issue, described below.

The second issue is about the choice of basis functions for the approximation. This choice is critical because the sample size is usually small, sometimes having less than a hundred observations in total. The basis functions need to deliver an adequate approximation with a few terms only; otherwise, the resulting method will not be useful in practice.

To address this issue, we observe that $f(x)$ has three features. First, its support has no fixed boundary. There is no simple rescaling of the data that can reduce the support of $f(x)$ to, say, $[0, 1]$. Second, this density is closely related to the normal density. When the stochastic process for the underlying asset is a geometric Brownian motion with drift r and volatility σ, this density is exactly $N(0, 1)$. Third, the density can have thicker tails than the normal distribution. These three features suggest that Hermite functions are suitable basis functions for approximating $f(x)$. Recall that the Hermite functions $\{h_j\}$ are the complete orthonormal system in $L^2(-\infty, \infty)$ given by

$$h_j(x) = \frac{H_j(x)}{(2^j j! \pi^{1/2})^{1/2}} e^{-x^2/2} \quad (j = 0, 1, 2, ...), \tag{8}$$

where $\int_{-\infty}^{\infty} h_j^2(x)dx = 1$ for all $j = 0, 1, 2, ...$, $\int_{-\infty}^{\infty} h_i(x)h_j(x)dx = 0$ for all $i \neq j$, and $\{H_j\}$ are the standard physicist's Hermite polynomials given by

$$H_j(x) = (-1)^j e^{x^2} \frac{d^j}{dx^j} e^{-x^2} \quad (j = 0, 1, 2, ...). \tag{9}$$

By expressing $f(x)$ in terms of $h_j(x)$, we obtain

$$f(x) \sim \sum_{j=0}^{\infty} \beta_j h_j(x), \tag{10}$$

where

$$\beta_j = \int_{-\infty}^{\infty} f(x)h_j(x)dx. \tag{11}$$

The right-hand side of (10) is the Gauss-Hermite expansion of $f(x)$. This expansion is well defined under mild conditions. For example, by Theorem 9.1.6 in Szegő (2003, p.247), for any function $f(x)$ that is Lebesgue-measurable in $[-\infty, +\infty]$, its Gauss-Hermite expansion converges to $f(x)$ uniformly on any finite interval of $\mathbb{R}$ if the following four conditions hold: (i) $\int_{-a}^{a} |f(x)|\, dx$ exists for every $a > 0$, (ii) there exists $M < \infty$ such that $|f'(x)| \leq M$ for all $x \in \mathbb{R}$, (iii) $\int_{-\infty}^{\infty} |f'(x)|\, dx < \infty$, and (iv) $f(x) = O(e^{x^2/2} |x|^{-\delta})$ for some $\delta > 0$ as $|x| \to \infty$. Condition (i) is satisfied here

because $f(x)$ is a density. Condition (iv) permits thicker tails than that of the normal distribution. Throughout the paper, we assume that the Gauss-Hermite series expansion of $f(x)$ is well defined.

Longstaff (1995) and Abken, Madan, and Ramamurtie (1996) considered the SPD estimation under the assumption that it has a finite-order Gram-Charlier series representation. Their estimator is parametric and is nonlinear in parameters (see Displays (8)-(9) and (20)-(22) in the two papers, respectively). If we replace $H_j(x)$ by the probabilist's Hermite polynomial, $H_{ej}(x) = (-1)^j e^{(x^2/2)} \frac{d^j}{dx^j} e^{-(x^2/2)}$, and let $\alpha_j = (1/j!) \int_{-\infty}^{\infty} f(x) H_{ej}(x) dx$, we obtain the Gram-Charlier series approximation to $f(x)$:

$$(2\pi)^{-1/2} \sum_{j=0}^{J} \alpha_j H_{ej}(x) e^{-x^2/2}. \tag{12}$$

However, as shown in Cramér (1946, p.223), for the Gram-Charlier series to converge, the tails of $f(x)$ in general need to approach zero faster than $\exp(-x^2/4)$. This condition is restrictive given the third feature of $f(x)$ described above. As a result, the Gram-Charlier series is unsuitable for directly approximating $f(x)$ in (5) and (6).

2.2 The estimation procedure

Let $\{y_i, z_i\}$ $(i = 1, ..., n)$ denote the sample of observed option prices and transformed strike prices at time t (see (4)). Options with zero open interests or zero transaction volumes can be excluded from the sample to avoid stale information. We assume that the data are ordered such that the first n_c observations are call options, and the remaining $n - n_c$ observations are put options. Then,

$$y_i = \begin{cases} \int_{-\infty}^{\infty} S \left(e^{\sqrt{\tau}\sigma x} - e^{\sqrt{\tau}\sigma z_i} \right)^+ f(x) dx + \varepsilon_i & \text{for } i = 1, ..., n_c, \\ \int_{-\infty}^{\infty} S \left(e^{\sqrt{\tau}\sigma z_i} - e^{\sqrt{\tau}\sigma x} \right)^+ f(x) dx + \varepsilon_i & \text{for } i = n_c + 1, ..., n, \end{cases} \tag{13}$$

where S denotes the spot price of the underlying asset at time t, and ε_i represents deviations from the theoretical prices due to market frictions and other factors.

The estimation procedure is based on a Gauss-Hermite series approximation to $f(x)$:

$$f(x) \approx \sum_{j=0}^{J} \beta_j h_j(x), \tag{14}$$

where β_j is defined in (11) and J is the truncation order. It comprises the following steps:
STEP 1. For $j = 0, ..., J$, compute

$$x_{i,j} = \begin{cases} \int_{-\infty}^{\infty} S \left(e^{\sqrt{\tau}\sigma x} - e^{\sqrt{\tau}\sigma z_i} \right)^+ h_j(x) dx & \text{for } i = 1, ..., n_c, \\ \int_{-\infty}^{\infty} S \left(e^{\sqrt{\tau}\sigma z_i} - e^{\sqrt{\tau}\sigma x} \right)^+ h_j(x) dx & \text{for } i = n_c + 1, ..., n. \end{cases} \tag{15}$$

STEP 2. Let

$$x_i = (x_{i,0}, ..., x_{i,J})' \text{ and } \beta = (\beta_0, ..., \beta_J)'. \tag{16}$$

Solve the following optimization problem

$$\min_{\beta \in \mathcal{H}_J} \sum_{i=1}^{n} \left(y_i - x_i'\beta\right)^2 + \beta' Q_\alpha \beta, \tag{17}$$

where

$$\mathcal{H}_J = \left\{ \beta \in R^{J+1} \colon \inf_{x \in \mathbb{R}} \sum_{j=0}^{J} \beta_j h_j(x) \geq \eta \right\}, \tag{18}$$

Q_α is a $(J+1)$-dimensional regularization matrix, and η is a small negative constant, whose values are specified in the next subsection. Let $\hat{\beta}_0, ..., \hat{\beta}_J$ be the solutions to (17). Compute

$$\hat{f}(x) = \sum_{j=0}^{J} \hat{\beta}_j h_j(x).$$

STEP 3 (Optional). Compute the final estimate of $f(x)$ as

$$\widetilde{f}(x) = \max(0, \hat{f}(x) - c), \tag{19}$$

where c is a constant is such that

$$\int_{-\infty}^{\infty} \widetilde{f}(x) dx = 1. \tag{20}$$

2.3 Discussion

In this subsection, we discuss several issues related to the method's implementation, including the choice of the tuning parameters. These issues are denoted as (I1) to (I6).

(I1) In $\mathcal{H}_J$, the parameter η needs to be slightly negative to account for the effect of the truncation. This follows because, by (10), the truncation error is equal to $f(x) - \sum_{j=0}^{J} \beta_j h_j(x) = \sum_{j=J+1}^{\infty} \beta_j h_j(x)$, which can be negative for some $x \in \mathbb{R}$. We suggest setting η to a value between $-$1E-4 and $-$1E-3 in practice. In the R code we develop, η is specified as an input parameter, so that the user can apply different values, and compare the results.

(I2) The term $\beta' Q_\alpha \beta$ is used to stabilize the estimate. We consider two specifications of Q_α. The first is a standard specification:

$$Q_\alpha = \alpha I \tag{21}$$

where I is an identity matrix. This is used as the benchmark. The second specification follows Fuhry and Reichel (2012):

$$Q_\alpha = V D_\alpha V' \tag{22}$$

with

$$D_\alpha = \text{diag}\left[\max\left\{\alpha - n\lambda_1, 0\right\}, \max\left\{\alpha - n\lambda_2, 0\right\}, ..., \max\left\{\alpha - n\lambda_{J+1}, 0\right\}\right], \tag{23}$$

where $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_{J+1}$ are the ordered eigenvalues of $n^{-1}\sum_{i=1}^n x_i x_i'$, and V contains the corresponding orthonormal eigenvectors as columns. Clearly, $\sum_{i=1}^n x_i x_i' + Q_\alpha = V\,\text{diag}[n\lambda_1 + \max\{\alpha - n\lambda_1, 0\},\ n\lambda_2 + \max\{\alpha - n\lambda_2, 0\},\ ...,\ n\lambda_{J+1} + \max\{\alpha - n\lambda_{J+1}, 0\}]V'$. Being different from (21), the specification (22) can be used to regularize a subset of eigenvalues of $\sum_{i=1}^n x_i x_i'$. For example, when $J = 4$ and $n\lambda_3 > \alpha > n\lambda_4$, we have $D_\alpha = \text{diag}\,[0, 0, 0, \alpha - n\lambda_4, \alpha - n\lambda_5]$ and $\sum_{i=1}^n x_i x_i' + Q_\alpha = V\,\text{diag}\,[n\lambda_1, n\lambda_2, n\lambda_3, \alpha, \alpha]\,V'$. Fuhry and Reichel (2012, Theorem 2.1) show that $\sum_{i=1}^n x_i x_i' + V D_\alpha V'$ is the closest matrix to $\sum_{i=1}^n x_i x_i'$ in the Frobenius norm among all real symmetric matrices with the smallest eigenvalue no less than α. Having a second regularization method allows the user to check the robustness of the estimation results. The theory in the next section holds under both methods.

(I3) The optimization problem in Step 2 is strictly convex. This follows because $\mathcal{H}_J$ is a convex set, and the criterion function (17) is strictly convex. In simulations and empirical applications, we implement this step using the routine `solve.QP` in R.

(I4) The parameter J controls the approximation error in (14), while α controls the degree of regularization for a given J. Based on the theoretical and simulation results obtained in the next two sections, we propose choosing J and α using the following procedure. Basically, this procedure specifies a sufficiently large J, and then uses α to regulate the estimates in case this J is larger than necessary. First, set J =ceiling$(2(n/\log n)^{1/5})$, where the constant "2" is chosen such that $J+1$ is equal to 6 and 4 when n is around 200 and 20, respectively. Next, for this J, determine α using cross validation. In our experimentations, the leave-one-out and the ten-fold cross validation produce similar results. To implement the leave-one-out cross validation, let $\hat{c}$ denote the (1,1)-*th* element of $n^{-1}\sum_{i=1}^n x_i x_i'$. Write α as $\alpha = \xi \hat{c} n^{1/3}$, where the scaling by $\hat{c}$ ensures that α will adapt to the unit of measurement. Generate a grid of points between 0 and 0.1, and search ξ over this grid to minimize $\sum_{i=1}^n (y_i - x_i'\hat{\beta}_{(-i)}(\xi))^2$, where $\hat{\beta}_{(-i)}(\xi)$ is obtained from (17) while leaving out the ith observation. The ten-fold cross validation is implemented in the same way, except that the ith fold of the partition replaces the ith observation. The upper bound for the grid can be further increased, although such values are never selected in our simulations and applications.

(I5) Step 3 of the procedure corresponds to the L_2-projection of $\hat{f}(x)$ on the class of nonnegative densities on the real line by Lemma 4 in Gajek (1986). That is, $\widetilde{f}(x)$ minimizes $\int_{-\infty}^{\infty} |\hat{f}(x) - g(x)|^2 dx$, where $g(x)$ can be any square-integrable function over $\mathbb{R}$, satisfying $g(x) \geq 0$ and $\int_{-\infty}^{\infty} g(x)dx = 1$. In our simulations and empirical applications, the differences between $\widetilde{f}(x)$ and $\hat{f}(x)$ are small.

(I6) When computing the integrals (15) and (20), $\int_{-\infty}^{\infty}$ needs to be replaced by $\int_{-m}^{m}$ for some $m > 0$. We suggest setting $m = \max\{10, \log n\}$. A formal justification is provided in the next section (see Lemma A.1). Our theory also allows for an asymmetric range, as long as the strike coverage expands similarly asymptotically on both sides. In our simulations and empirical applications, we vary m between 10 and 30 and do not find any meaningful difference in the results.

3 Properties of the estimator

We first study the identification of the state space density based on put and call options, and then establish the convergence rate of the estimator $\hat{f}(x)$. Throughout the analysis, the results are conditional on the spot price S, the riskless rate r, and the scaling factor σ.

3.1 Identification based on call and put options

We consider the identification of $f^*(x)$ in (1)-(2) based on a continuum of call and put options. Let $[\mathcal{K}_{c,L}, \mathcal{K}_{c,U}]$ and $[\mathcal{K}_{p,L}, \mathcal{K}_{p,U}]$ denote the spans of the call and put strikes, respectively. These two spans can be identical, or only partially overlap. For example, for S&P 500 options, $\mathcal{K}_{p,L}$ is often smaller than $\mathcal{K}_{c,L}$ after excluding options with zero open interest. More generally, for index options, there is more liquid coverage on the downside via out-of-the-money put options, implying a lower $\mathcal{K}_{p,L}$ than $\mathcal{K}_{c,L}$. Define

$$[\mathcal{K}_L, \mathcal{K}_U] = [\mathcal{K}_{c,L}, \mathcal{K}_{c,U}] \cup [\mathcal{K}_{p,L}, \mathcal{K}_{p,U}].$$

Assumption 1 *There exists a Lipschitz continuous density over $x \in \mathbb{R}$ such that (1) and (2) hold for $K \in [\mathcal{K}_{c,L}, \mathcal{K}_{c,U}]$ and $K \in [\mathcal{K}_{p,L}, \mathcal{K}_{p,U}]$, respectively, where $[\mathcal{K}_{c,L}, \mathcal{K}_{c,U}] \cap [\mathcal{K}_{p,L}, \mathcal{K}_{p,U}]$ is nonempty.*

The next lemma characterizes the identification of $f^*(x)$ within and outside $[\mathcal{K}_L, \mathcal{K}_U]$.

Lemma 1 *Suppose $f_0^*(x)$ satisfies Assumption 1. Let $f_1^*(x)$ be any density that is Lipschitz continuous over $x \in \mathbb{R}$. Then, the following two statements are equivalent:*

(a) $f_1^*(x)$ satisfies (1) and (2) for $K \in [\mathcal{K}_{c,L}, \mathcal{K}_{c,U}]$ and $K \in [\mathcal{K}_{p,L}, \mathcal{K}_{p,U}]$, respectively.

(b) $f_1^*(x)$ is equal to $f_0^*(x)$ for any $x \in [\mathcal{K}_L, \mathcal{K}_U]$. Furthermore, $\int_{-\infty}^{\mathcal{K}_L} w(x) f_1^*(x) dx = \int_{-\infty}^{\mathcal{K}_L} w(x) f_0^*(x) dx$ and $\int_{\mathcal{K}_U}^{\infty} w(x) f_1^*(x) dx = \int_{\mathcal{K}_U}^{\infty} w(x) f_0^*(x) dx$, where $w(x) = [1 \;\; x]$ and $x \in \mathbb{R}$.

Lemma 1 generalizes the identification results in Cox and Ross (1976, p.154) and Breeden and Litzenberger (1978) in two ways. First, the put options are included in the identification analysis. Second, the support of the density and that of the strikes are allowed to be different. The result shows that $f^*(x)$ is point identified only within $[\mathcal{K}_L, \mathcal{K}_U]$. Outside this interval, it is not identified,

but it satisfies the integral restrictions stated in (b). Consequently, if $[\mathcal{K}_{c,L}, \mathcal{K}_{c,U}]$ and $[\mathcal{K}_{p,L}, \mathcal{K}_{p,U}]$ are different, pooling the options can lead to better identification than using the call options only. The lemma can also be formulated equivalently in terms of $f(x)$, after applying (3) and (4).

The next assumption imposes requirements on the data such that $f^*(x)$ is point identified in probability as $n \to \infty$. Intuitively, we need the option strikes to be dense asymptotically, and their range to cover all relevant positive out-of-the-money call and put option prices, such that we can estimate the full SPD consistently. Let $z_{(1)}, ..., z_{(n)}$ be the order statistic of $z_1, ..., z_n$. Let $F(x)$ denote the CDF of $f^*(x)$.

Assumption 2 *As* $n \to \infty$*:* (i) $\max_{i \leq n-1} F(z_{(i+1)}) - F(z_{(i)}) \xrightarrow{p} 0$;$(ii)$ $F(z_{(1)}) \xrightarrow{p} 0$ *and* $F(z_{(n)}) \xrightarrow{p} 1$.

This Assumption is used in place of the more standard assumption of $\{z_1, ..., z_n\}$ being drawn from a continuous distribution with support $(-\infty, \infty)$. Bassett (1997) shows that the approximate slope of the option curve implies bounds on the underlying cumulative distribution $F(z)$. This property allows us to evaluate Assumption 2's empirical relevance using inequalities; see Online Appendix B.

3.2 Asymptotic properties

We first present and discuss the assumptions. Then, we establish the convergence rate of the estimator $\hat{f}(x)$.

The estimation problem has two nonstandard features. First, $f(x)$ is defined by a Fredholm integral equation of the first kind with an unbounded support, where the two kernel functions $S(e^{\sqrt{\tau}\sigma x} - e^{\sqrt{\tau}\sigma z})^+$ and $S(e^{\sqrt{\tau}\sigma z} - e^{\sqrt{\tau}\sigma x})^+$ are unbounded and are non-differentiable in x or z. In other words, we can write the call pricing function $C(z) = \int_{-\infty}^{\infty} S\left(e^{\sqrt{\tau}\sigma x} - e^{\sqrt{\tau}\sigma z}\right)^+ f(x)dx$ as $C(z) = \int_{-\infty}^{\infty} K(x,z) f(x) dx$ with $K(x,z) = S\left(e^{\sqrt{\tau}\sigma x} - e^{\sqrt{\tau}\sigma z}\right)^+$, where $K(x,z)$ diverges to infinity for any fixed z as $x \to \infty$. Second, as J and n increase, the $(J+1)$-th diagonal element of $n^{-1}\sum_{i=1}^{n} x_i x_i'$ diverges to infinity, while its smallest eigenvalue approaches zero, the orders of which are unknown. These features, particularly the first one, imply that the results in the recent econometric literature on linear inverse problems, such as those in Blundell, Chen, and Kristensen (2007), Carrasco, Florens, and Renalt (2007), Horowitz (2011, 2014), and Horowitz and Lee (2012), can not be directly applied to our analysis. More generally, to the best of our knowledge, most asymptotic results for linear integral equations are obtained under settings where the kernel function is smooth and bounded, and the support is compact. Unfortunately, both conditions are violated with the current problem. We are unaware of any existing results that would imply our estimator's statistical properties, in particular, its rate of convergence. Fortunately, the option pricing problem has specific structures that we can exploit. To this end, we begin by stating the

following assumption, which concerns the real axis decay of the SPD and the asymptotic behavior of its Hermite coefficients.

Assumption 3 *The density $f(x)$ and its Hermite coefficients satisfy: (i) there exist $a > 0$ and $\kappa > 1$, such that $f(x) = O(e^{-a|x|^{\kappa}})$ as $|x| \to \infty$ along the real axis; and (ii) there exist $r > 0$ and $p > 0$, such that, for β_j in (11),*

$$\beta_j = O(\exp(-pj^r)) \text{ as } j \to \infty. \tag{24}$$

Assumption 3(i) allows heavier tails than that of a normal distribution. It is important to note that this assumption is imposed on $f(x)$, not on $f^*(x)$. The latter density does not satisfy this assumption even when the underlying asset price follows a geometric Brownian motion. Sufficient conditions for Assumption 3(ii) can be found in the literature. For example, Hille (1940, Theorem 1) shows that (24) holds with $p = \sqrt{2}\tau$ and $r = 1/2$ if (a) $f(x)$ is analytic and $f(z)$ is differentiable over a strip $-\tau \leq y \leq \tau$ for some $\tau > 0$, where $z = x + iy$; and (b) for every given c with $0 \leq c < \tau$, there exists a finite $B(c)$ such that $|f(z)| \leq B(c)\exp[-|x|\left(c^2 - y^2\right)^{1/2}]$ for $-\infty < x < \infty$ and $-c \leq y \leq c$. Building on Hille's result, Boyd (1984) shows that (24) holds with $r > 1/2$ if (a) $f(z)$ is differentiable over the complex plane; (b) and $f(x) \sim \exp(-a\,|x|^{\kappa})$ for some $\kappa > 1$ as $|x| \to \infty$ along the real axis, where a is some finite constant, and '$\sim$' denotes that the two sides can differ by an algebraic factor. Assumption 3(ii) requires $r > 0$, substantially weaker than requiring $r \geq 1/2$.

The results of Hille (1940) and Boyd (1984) imply that Assumption 3(ii) holds in the following four settings: (S1) $f(x)$ is equal to the product of a normal density and a finite order polynomial in x; (S2) $f(x)$ is a finite mixture, with each mixing component being a normal distribution multiplied by a finite order polynomial in x; (S3) $f(x)$ is a finite mixture, where the mixing components satisfy the conditions of Hille (1940), possibly with different τ, or the conditions of Boyd (1984), possibly with different a and κ; and (S4) $f(x)$ is any density that has a finite order Hermite series representation. In (S2) and (S3), Assumption 3(ii) holds because the Hermite coefficients of a finite mixture are equal to the sum of the individual coefficients. Finally, (S2) and (S4) imply that Assumption 3(ii) is compatible with the parametric models considered in Jarrow and Rudd (1982), Longstaff (1995), Abken, Madan and Ramamurtie (1996), and Melick and Thomas (1997), because (S2) and (S4) include the densities of these models as special cases.

Assumption 3(i) rules out some heavy-tailed distributions. For example, if S_T has a Lévy distribution (i.e., $f^*(x) \sim x^{-3/2}$ as $x \to \infty$), then the SPD satisfies $f(x) \sim \exp(-\sigma\sqrt{\tau}x/2)$ as $x \to \infty$, violating the requirement $\kappa > 1$ in Assumption 3(i). However, because the mean of a Lévy distribution is infinite, the price of any call option with a finite strike is also infinite, casting doubt on whether such a distribution is directly useful for option pricing without any truncation.

Assumption 3 is the key assumption of the paper. Given its importance, we now examine it in the context of several commonly used parametric models for option pricing. The first model, which is also the most general one, is a stochastic volatility model with contemporaneous jumps in return and volatility (SVCJ, see Duffie, Pan, and Singleton, 2000 and Eraker, 2004). Its dynamics under the risk neutral measure is given by

$$\begin{aligned} \frac{dS_t}{S_t} &= r_t dt + \sqrt{V_t} dW_t + (e^{Z_t^s} - 1) dN_t - \lambda \bar{\mu} dt \\ dV_t &= k(\theta - V_t) dt + \sigma_v \sqrt{V_t} dB_t + Z_t^v dN_t, \end{aligned} \tag{25}$$

where S_t is the spot price of the underlying asset; r_t is the risk free rate; W_t and B_t are two Brownian motions such that $E(dW_t dB_t) = \rho dt$; $N_t \sim Poi(\lambda)$ is a Poisson counting process with a constant intensity λ; Z_t^s denotes the jump in the return with $Z_t^s \sim N(\mu_s, \sigma_s^2)$; and Z_t^v denotes the jump in the volatility which follows an exponential distribution: $Z_t^v \sim Exp(\mu_v)$. Amongst the parameters, θ determines the long run level of the volatility, k affects the speed of the mean-reversion, σ_v is the volatility-of-volatility parameter, and $-\lambda\bar{\mu}$, with $\bar{\mu} = exp(\mu_s + \sigma_s^2/2) - 1$, compensates for the instantaneous change in the expected return due to the presence of Z_t^s. When restricting $\mu_v = 0$, the model in (25) reduces to a stochastic volatility model with jumps in return only. When further restricting $\mu_s = \sigma_s = \lambda = 0$, the model reduces to the baseline stochastic volatility model. Finally, by further restricting V_t to be constant, we obtain the Black-Scholes model. We denote these three models by SVJ, SV, and BS, respectively. Duffie, Pan, and Singleton (2000) show that the SPDs implied by the SVCJ, SVJ, and SV models have closed-form expressions (up to numerical integration). This feature allows us to examine Assumption 3 in close detail.

To make this analysis empirically relevant, we estimate the parameters of the four models using daily observations on the S&P500 index and its options. To reflect different market conditions, we choose three representative time periods and estimate the model parameters (except for ρ and λ) separately. These three periods are 2008.10 - 2009.4 (Crisis Period), 2009.6 - 2009.12 (Recovery Period), and 2013.6 - 2013.12 (Expansion Period), during which the average VIX levels are equal to 50, 25, and 15, respectively. After obtaining twelve sets of parameter estimates (i.e., four models and three periods), we apply them to compute the implied SPDs, and then run regressions to estimate the values of κ and r defined in Assumption 3. The results are summarized in Tables 1 and 2; see their footnotes for additional details on computation. As the results show, the estimated values of κ and r are well above one and zero, respectively. The estimated values of κ are close to two in the BS case, consistent with the normality of the resulting distributions. In the remaining cases, the estimated values of κ for the left tail are always less than two, while those for the right tail are greater than two, indicating thicker left and thinner right tails than those of a normal distribution. Figures 1 and 2 provide additional information by displaying the tails of the densities,

the associated Hermite coefficients, and the fitted values from the regressions for the crisis period under the BS and SVCJ models. The remaining cases are similar and are displayed in Figures S1-S10 of the online appendix. The goodness of fit is seen to be adequate in all cases. Together, these results support that Assumption 3 is compatible with the four parametric models for option pricing.

We now examine how the approximation error $f(x) - \sum_{j=0}^{J} \beta_j h_j(x)$ depends on J for the SVCJ model. The results for the crisis and expansion periods (Figures 3 and 4) show that the errors are large when $J = 1$, they substantially decrease when J increases to 3, and become negligible when $J = 5$. The recovery period results are similar; we report them in the online appendix (Figure S11) to save space. Further, in Figures 5-6 and S12, we display the same approximations but in terms of the index level's densities (i.e., $f^*(S_T)$ in (7)) to reflect the effect of the transformation (3). The conclusions are the same. Finally, because the findings for the SV and SVJ models are similar, we omit them to save space. These results are encouraging, showing that it is possible to achieve parsimonious and adequate approximations across different data generating processes and market conditions.

The next two assumptions concern the sampling process, the tuning parameters J, α, and η, and the ill-posedness of the inverse problem.

Assumption 4 *(i)* $\{z_i, \varepsilon_i\}$ *are independent of* $\{z_j, \varepsilon_j\}$ *for* $i \neq j$*; (ii)* $E(\varepsilon_i) = 0$*,* $E(\varepsilon_i^2) = \sigma_i^2$ *with* $0 < \sigma_i^2 < \infty$*, and* ε_i *is independent of* z_i *for any* i*. (iii)* $\max_{i \in \{n_c+1,\ldots,n\}} z_i = O_p(\log n)$*.*

Assumption 5 *As* $n \to \infty$*: (i)* $\log n / J^r + J/n^{1/5} \to 0$*; (ii)* $\gamma = O_p(J^\nu)$ *for some* $\nu < \infty$*, where* $\gamma = [\beta'(n^{-1}\sum_{i=1}^{n} x_i x_i')^{-1/2}\beta]^{1/2}$*; (iii)* $\alpha/n \to 0$*; and (iv)* $\eta < 0$ *and* η *is independent of* n*.*

Assumption 4(ii) allows for heteroskedasticity. Assumption 4(iii) ensures that the integral $\int_{-\infty}^{\infty}$ in (13) can be adequately approximated by $\int_{-\log n}^{\log n}$. Assumption 4(iii) is for put options only. It allows $\min_{i \in \{1,\ldots,n\}} z_i$ and $\max_{i \in \{1,\ldots,n_c\}} z_i$ to approach $-\infty$ and $+\infty$ at any rate. For call options, the tail integrals $\int_{-\infty}^{-\log n} S\left(e^{\sqrt{\tau}\sigma x} - e^{\sqrt{\tau}\sigma z}\right)^+ f(x)dx$ and $\int_{\log n}^{\infty} S\left(e^{\sqrt{\tau}\sigma x} - e^{\sqrt{\tau}\sigma z}\right)^+ f(x)dx$ are positive and bounded from above by $\int_{-\infty}^{-\log n} S e^{\sqrt{\tau}\sigma x} f(x)dx$ and $\int_{\log n}^{\infty} S e^{\sqrt{\tau}\sigma x} f(x)dx$, respectively. Because these two upper bounds do not involve z, we do not need any restriction on z to approximate them. In other words, Assumption 4(iii) is not needed for call options because the call price $C(z)$ is a non-negative, decreasing function of z.

Assumption 5(i) requires the truncation order J to increase faster than $(\log n)^{1/r}$ to ensure that the bias from the Hermite series approximation is asymptotically negligible. Assumption 5 (ii) imposes an upper bound on the decay rate of the Hermite coefficients relative to the eigenspace of $n^{-1}\sum_{i=1}^{n} x_i x_i'$, where x_i and β are defined in Step 2 of the estimation procedure; see (16).

The parameter γ is related to, but different from, the sieve measure of ill-posedness τ_n defined in Blundell, Chen, and Kristensen (2007). To see this, let $\lambda_{\min}$ denote the smallest eigenvalue of $n^{-1}\sum_{i=1}^{n} x_i x_i'$ and let u_{jk} denote the (j,k)-th element of $(n^{-1}\sum_{i=1}^{n} x_i x_i')^{-1/2}$. Then, $\gamma = (\sum_{j,k=0}^{J} \beta_j \beta_k u_{(j+1)(k+1)})^{1/2}$ and $\tau_n = \lambda_{\min}^{1/2}$. Because β_j decays to zero at the rate $\exp(-pj^r)$, the coefficients in the summation $\sum_{j,k=0}^{J} \beta_j \beta_k u_{(j+1)(k+1)}$ are close to zero except when j and k are both small, which implies that the diverging rate of γ can be much slower than that of $\lambda_{\min}^{-1/4}$ as $J \to \infty$. By imposing a restriction on γ, instead of on τ_n, this assumption allows the inverse problem to be severely ill-posed, as long as γ decays at the rate $O_p(J^\nu)$ for some $\nu < \infty$.

The parameter α in Assumption 5(iii) appears in the regularization matrix Q_α. Note that Q_α is not a generic matrix. Instead, it is defined explicitly in (21) and (22). In Assumptions 5(ii) and 5(iii), we impose regularity conditions on α and $n^{-1}\sum_{i=1}^{n} x_i x_i'$. These conditions determine the properties of Q_α, allowing us to derive the convergence rate. Finally, Assumption 5 (iv) implies that the effect of η on the estimator $\hat{f}(x)$ is negligible asymptotically.

Theorem 1 *Let Assumptions 1-5 hold. Then, for any $x \in \mathbb{R}$:*

$$\hat{f}(x) - f(x) = O_p\left(\left(\frac{\alpha}{n}\right)^{1/4} J^{1/2+\nu} + \sqrt{\frac{J}{\alpha}}\right),$$

where $(\alpha/n)^{1/4} J^{1/2+\nu}$ reflects the bias caused by the Tikhonov regularization:

$$B(x) = -E\left[h(x)'\left(\sum_{i=1}^{n} x_i x_i' + Q_\alpha\right)^{-1} Q_\alpha \beta\right].$$

Theorem 1 holds for both regularization methods described in Step 2 of the estimation procedure. By letting $J \sim (\log n)^{1/r+c_1}$ for some finite c_1 and $\alpha \sim n^{c_2}$ with $2/7 < c_2 < 1/3$, we obtain $\hat{f}(x) - f(x) = O_p(n^{-1/7})$. This rate corresponds to the optimal convergence rate when estimating the second-order derivative of a smooth function using a second-order polynomial and a second-order kernel. Note that the terms $(\alpha/n)^{1/4} J^{1/2+\nu}$ and $\sqrt{J/\alpha}$ are both increasing in J. This reflects Assumptions 3 and 5(i), under which the bias from the Hermite series approximation is asymptotically negligible.

4 Finite sample properties

We first study the effects of the truncation order and the regularization method on the precision of the proposed estimator. Then, we compare this estimator with several notable estimators in the literature: the constrained local linear estimator of Aït-Sahalia and Duarte (2003), the positive convolution approximation (PCA) approach of Bondarenko (2003), and the four-parameter Gram-Charlier series estimator of Jarrow and Rudd (1982) and Longstaff (1995).

4.1 The effects of tuning parameters on the estimator

To study this issue, we generate option prices using the SVCJ model with the parameter values set to those in Table 1. The error ε_i is set to $m_i\%$ of the theoretical price level with a cap of 5 dollars, where m_i is independently and uniformly distributed as $U(-10, 10)$. To mimic the S&P500 options market, we let $S_t = 1000$, $n_c = 68$, and $n_p = 72$, so that the total number of options is equal to 140. The spans of the call and put strikes are [660,1330] and [540, 1290], respectively. The time-to-maturity is equal to 31 days, and the scaling factor σ is equal to the Black-Scholes implied volatility for the at-the-money call option. For each regularization method and truncation order, the parameter α is determined by the ten-fold cross validation. The outcomes with $\alpha = 0$ are also reported, which serve as a benchmark for comparison. The reported mean integrated squared errors (MISE) are based on 5000 simulation replications.

Table 3 presents the results. In all cases, the smallest MISE is achieved when $3 \leq J \leq 7$, and the decrease in the MISE (if there is any) is always small when J increases from 4 to 7. This finding shows that a small number of sieve terms can deliver adequate approximations across different market conditions. A comparison with $\alpha = 0$ shows that the Tikhonov regularization is useful in stabilizing the performance of the estimator. In particular, the regularization has little effect on the estimator when J is small, whereas it leads to a significant reduction in the MISE when J is large. Between the two regularization methods, using $Q_\alpha = VD_\alpha V'$ produces smaller MISEs overall than using $Q_\alpha = \alpha I$, although the difference is mild and occurs mainly when $J \geq 7$. Note that if we let $J =$ceiling$(2(n/\log n)^{1/5})$, we obtain $J = 4$.

4.2 Comparison with other estimators

We use the DGP of Aït-Sahalia and Duarte (2003) for this comparison. The spot price, interest rate, and dividend yield are equal to $S_t = 1365$, $r_t = 4.5\%$, and $\delta_t = 2.5\%$, respectively. The time-to-maturity is equal to 30 days, and the spans of the put and call strikes are both $[1000, 1700]$. The implied volatility is a linear function of the strike price: it equals 40% when $K = 1000$ and 20% when $K = 1700$. At each volatility level, the option price is generated using the Black-Scholes formula, which is then contaminated by an additive noise with a uniform distribution. In addition to considering $n_c = n_p = 25$, we also consider a larger sample size $n_c = n_p = 50$. We report the MISEs in Table 4 and display the empirical distributions of the density estimates for $n_c = 25$ in Figures 7-10. The figures reveal how the methods perform for different parts of the distribution.

For the proposed sieve method, by letting $J =$ceiling$(2(n/\log n)^{1/5})$, we obtain $J = 4$. We also report the MISEs for $J = 3, 5$, and 6 to examine result sensitivity.

The first method we compare with is the constrained local linear estimator of Aït-Sahalia and Duarte (2003). However, their paper does not provide a bandwidth selection rule for this

estimator. To this end, we apply the optimal bandwidth for estimating the conditional mean in an unconstrained local linear regression (i.e., their expression (3.21)), and adjust the constant $C_{0,1}$ such that the bandwidth minimizes the *ex-post* MISE under the current simulation design with $n_c = 25$. Because this bandwidth choice requires the knowledge of the true DGP, it favors their method. The implementation uses call options only, as in their work.

The second method is that of Bondarenko (2003). The estimator solve the optimization problem

$$\min_{\hat{f}\epsilon\mathcal{W}_h^{\Delta z}([v,w])} \sum_{i=1}^{n_p} \left(P_i - e^{-r\tau} \int_{-\infty}^{K_i} \left(\int_{-\infty}^{y} \hat{f}(x)dx \right) dy \right),$$

where P_i is the price of a put option, K_i is a strike price, $\hat{f}(x)$ represents the SPD of S_T, $e^{-r\tau}$ provides the discounting by the risk-free rate, and $[v, w] = [1000, 1700]$ for the current setting. The set $\mathcal{W}_h^{\Delta z}([v, w])$ is determined by

$$\mathcal{W}_h^{\Delta z}([v,w]) = \left\{ g(x) = \sum_{j=1}^{M} a_j \phi_h(x - z_j) \middle| a_j \geq 0, \sum_{j=1}^{M} a_j = 1 \right\},$$

where a_j $(j = 1, ..., M)$ are unknown coefficients and $\phi_h(x - z_j) = \phi((x - z_j)/h)/h$, i.e., a Gaussian kernel centered at z_j with bandwidth (i.e., standard deviation) equal to h. As in Bondarenko (2003), we distribute z_j evenly between the strikes 1000 and 1700, and set $\Delta z = z_j - z_{j-1} = 0.5h$. Bondarenko (2003) suggested that a reasonable range for M is from 21 to 25. We follow this suggestion, and then determine the value of M within this range by 10-fold cross validation. Finally, as in Bondarenko (2003), we require that the estimator satisfy the following two constraints: $\int_{-\infty}^{\infty} \hat{f}(x)dx = 1$ and $e^{-(r-\delta)\tau} \int_0^{\infty} x\hat{f}(x)dx = S_t$ with δ being the daily dividend rate. The implementation uses put options only, as in the original study.

The third estimator is that of Jarrow and Rudd (1982) and Longstaff (1995), which assumes that SPD of S_T belongs to a four-parameter family:

$$f(x) = \frac{\exp\left(-x^2/2\right)}{\sqrt{2\pi\tau}\sigma} \left(1 + \frac{\beta}{6}\left(x^3 - 3x\right) + \frac{\gamma}{24}\left(x^4 - 6x + 3\right) \right),$$

where

$$x = \frac{\ln(S_T/S_t) - \left(\mu - \sigma^2/2\right)\tau}{\sigma\sqrt{\tau}}$$

and μ, σ, β, and γ are unknown parameters. This method is parametric because the model's complexity does not vary with the sample size. The estimate for the SPD is determined by the value of $(\mu, \sigma, \beta, \gamma)$ that minimizes the sum of squared pricing errors. The estimation can use a subset of options or all options. We use all options because it delivers the best performance for the current simulation design.

Table 4 shows that the proposed sieve estimator produces smaller MISEs than all competing methods when using the default truncation order J =ceiling$(2(n/\log n)^{1/5})$ and the default penalization method. Some potential explanations for this difference in performance are as follows. (i) The constrained local linear estimator uses call options only. Because the noises in call options are higher at lower strikes by the simulation design, the left tail's estimation is imprecise. This feature is clear from Figure 8, which shows that the left tail's estimates are much more volatile than the right tail. The proposed estimator uses all options; therefore, it can overcome this limitation. (ii) The positive convolution approximation estimator requires estimating about 20 parameters. Although the constraints embedded in $\mathcal{W}_h^{\Delta z}([v,w])$ help stabilize the estimates as explained in Bondarenko (2003), the remaining degrees of freedom are still high for the sample sizes considered here. Consequently, the estimates exhibit high variability, especially in the right tail, where the put option prices are noisier. See Figure 9. The proposed sieve estimator estimates fewer parameters, equal to 5 when J =ceiling$(2(n/\log n)^{1/5})$. Consequently, the estimates are more tightly distributed around the true value. (iii) The parametric estimator of Jarrow and Rudd (1982) and Longstaff (1995) solves a non-convex optimization problem. It is not straightforward to impose restrictions to prevent large negative density estimates. As a result, the estimates take on unreasonable values in a substantial portion of our simulation replications. For example, after sorting the estimates $\hat{f}(x)$ at each x, we find that they reach -0.52, -0.14, -0.13, -0.10, -0.08, and -0.06 at the following quantile levels across x values: 0, 0.05, 0.10, 0.20, 0.30, and 0.40. See Figure 10. In contrast, the proposed sieve estimator solves a convex optimization problem, for which imposing parameter restrictions is straightforward. In particular, our implementation guarantees that the density estimates are bounded from below by -0.001. Overall, the difficulty in borrowing information from constraints explains why the parametric estimator exhibits higher variability than the proposed sieve estimator. We also repeated the above simulation using a smaller sample size with $n_c = n_p = 20$. The conclusions are the same; see Supplemental Table S2 and Figure S13 for details.

The above results also show that discarding call or put options can result in substantial efficiency losses. It is interesting to study how to generalize the methods of Aït-Sahalia and Duarte (2003) and Bondarenko (2003) to incorporate more data into the estimation, potentially by using the put-call parity. However, such an investigation is beyond the scope of the current paper.

5 Applications

We first estimate the state price densities implied by the S&P 500 index options and by the VIX options using daily observations from January 18, 2007, to April 27, 2016. Then, we study the predictive power of these densities using quantile and least square regressions. Next, we discuss a proposal to use the state price densities to measure the market's perception of the near future.

Finally, to illustrate the application of the method to high-frequency data, we consider two FOMC announcements, and, for each one, we trace out the state price densities within a short time window to examine its causal effect on the financial market. The data used for the first three analyses are obtained from Optionmetrics and the last analysis from Ivolatility. All results reported are obtained using $J = \text{ceiling}(2(n/\log n)^{1/5})$, under the standard Tikhonov regularization with α determined by the 10-fold cross-validation.

5.1 State price densities during the recent financial crisis and recovery

To study the state price densities implied by the S&P 500 index options, we first delete the options with zero open interest or zero transaction volume to avoid stale information. We then identify the trading days with at least 40 different option strikes with $\tau = 30$ in the data. In the sample, there are 279 trading days with $\tau = 30$. Among them, 232 days have at least 40 different option strikes. The remaining 47 days are excluded from the analysis. Note that the expiration date was the third Friday of each month at the beginning of the sample. Later, the CBOE offered more settlement dates as the demand for options increased over time. For prices, Optionmetrics provides the best closing bid and ask quotes for the records dated March 4th 2008 and older, and the best 15:59 EST bid and ask quotes for the records dated March 5th 2008 till present. The midpoint of these quotes and the closing price of the index value are used for our estimation. Because we do not use the restriction on the spot price (i.e., $S_t = e^{-(r-\delta)\tau} \int_0^\infty x f^*(x) dx$), we do not require the option prices and the index value to be exactly synchronized. Instead, the index value induces a monotonic transformation (see (3) and (4)), which can be inverted after the estimation (see (4)).

After the data processing and the estimation, we obtain 232 state price densities with $\tau = 30$. The 0.1, 0.25, 0.5, 0.75, and 0.9 quantiles of these densities are reported in Figure 11. The realized monthly returns are also included in the same figure for ease of comparison.

The results reveal interesting connections between the realized return and the state price density. In particular, they show that 44.3% of the realized returns landed between the 0.25 and 0.75 quantiles of the estimated densities, and 85.3% of the realized returns landed between the 0.1 and 0.9 quantiles. To further examine their connections under the financial crisis and other major market disruptions, we consider four adverse events that each triggered a more than 3% loss in the index value: September 15, 2008 (-4.7%), when Lehman Brothers filed bankruptcy; September 29, 2008 (-8.7%), when the House rejected a $700 billion bailout plan; May 6, 2010 (-3.24%), when automated trading triggered the "flash crash"; and August 8, 2011 (-6.66%), when Standard & Poors downgraded the US credit rating. The findings are summarized below.

First, the densities' left tails indicate that some fear was already building up in the financial market before September 2008. In particular, in the first three months of 2007, the 10th percentiles

of the densities corresponded to market returns of −3.6%, −3.5%, and −4.8%, respectively. From June to August of 2008, these values worsened substantially to −8.6%, −10.0%, and −7.6%, respectively. Second, the densities reveal that, in the beginning, the market greatly underestimated the severity of the market outcome. For example, on September 18, 2008, three days after the bankruptcy of Lehman Brothers, the 10th percentile of the density implied a market loss of 14.6%. Although this magnitude is unusually high from a historical perspective, it is well below the subsequent loss of 24.9%. Except for a thickened left tail, this density did not indicate any clear tendency, such as a large increase in the dispersion, implying that the market was still perceiving a severe crisis as unlikely. Third, a comparison of the timing shows that a subsequent significant decline in the index value triggered a jump in the density's dispersion. In particular, from September 18 to October 23, the 10th and 90th percentiles of the density jumped from −14.6% and 9.3% to −35.8% and 18.9%, respectively. Clearly, at that point, the market was no longer perceiving a financial crisis as an unlikely event but a reality whose consequence was very uncertain. For the remaining three adverse events, we observe similar jumps in the dispersion of the densities after large negative realized returns, which implies that the timings and magnitudes of these events were unexpected and that, when they occurred, they significantly impacted the market's perception of the future.

To further illustrate the dynamics of the state price density over this period, in Figure 12, we display three densities obtained using the data on October 23, 2008 (during the financial crisis), June 18, 2009 (the tipping point between the recession and the recovery), and June 20, 2013 (during the expansion). The figures show that, as the underlying economic conditions improved, not only the dispersion of the density declined substantially, its left tail also continued to decrease towards the center of the distribution. This finding suggests that changing-in-variance does not fully describe the rich dynamics in the data.

In theory, because the state price density is equal to the product of the physical likelihood and the risk attitude (i.e., marginal utility), bimodality can occur if the decrease in the former is more than offset by the increase in the latter as we move away from the center of the distribution. However, in practice, except for a few studies, including the current one and Jackwerth and Rubinstein (1996, Figure 4), most studies document densities with tails decaying monotonically to zero. Therefore, it is essential to examine whether the bimodality feature reported here is due to irregularities in option prices (e.g., caused by illiquid quotes) or the Gauss-Hermite approximation. For the first possibility, we display the option prices for the three dates in Supplemental Figure S14. No visible irregularity is detected. Next, we exclude the three lowest and three highest strikes for both put and call options (i.e., excluding 12 observations in each case) and repeat the estimation. The results do not change; see Supplemental Figure S15. Based on these considerations, we conclude that the bimodality is unlikely the result of any data irregularity. For the second possibility, we

repeat the estimation with the J's value increased by one. The estimates, shown in Figure S16, display the same pattern as in Figure 12. In addition, classifying the bimodality as an artifact does not explain why the second mode is always in the left tail and why it moved closer to the center of the distribution as the market condition improved. Together, the evidence supports that the bimodality is a genuine feature of the state price density behind the data.

For the VIX options, we consider the same sample period and process the data in the same way, except that we use the trading days that have at least 20 different option strikes with $\tau = 30$ in the data. This yields 111 state price densities with $\tau = 30$. The results, summarized in Figure 13, show that 43.2% of the realized VIX values landed between the 0.25 and 0.75 quantiles of the estimated densities, and 79.3% of the VIX values landed between the 0.1 and 0.9 quantiles. As in Figure 11, the quantile bands show that the market initially underestimated the severity of the market outcome. In particular, on the following three dates in 2008, July 21, August 18, and September 22, the medians of the state price densities were 18.2, 17.2, and 19.4, and the 90th percentiles were only 30.0, 28.9, and 34.1, whereas the realized VIX levels were 20.4, 36.2, and 69.7, respectively. After the House rejected the $700 billion bailout plan and a series of adverse events, on October 22, the 90th percentile of the density jumped to 67.5, which was now comparable to the realized VIX of 74.3. Figure 14 displays three representative densities, showing that as the economy improved, the density shifted to the left, and the right tail decreased towards the center of the distribution.

Figures 11-14 convey an important message in the asset and derivatives pricing literature. That is, by estimating a sequence of state price densities, we can uncover rich information about the market's perception of downside risks. Such information is not easy to obtain through other channels. The "perception" here is a mix of predictions and associated risk premia. One of the challenges in this literature is to disentangle the state price density into the effects of objective expectations and risk premia; however, investigating this issue is outside the scope of the current paper.

5.2 Predictive power of the state price density

We consider predictive mean and quantile regressions using quantiles of the state price density as the predictor. The prediction horizon is set as one month. The quantiles are computed based on the estimated densities in the previous subsection. Using estimated quantiles, instead of the true ones, does not bias the results towards overrejection of the null hypothesis of no predictability as long as the pricing errors are serially uncorrelated. In addition to the full sample ($2007.1 - 2016.4$), we also consider a subsample $2009.6 - 2016.4$. This allows us to evaluate the effect of the financial crisis on the parameter estimates and document the differences. Below, we highlight the empirical findings. Providing theoretical explanations for them is beyond the scope of the current paper.

We first consider the S&P 500 case. The results for the predictive mean regressions are summarized in Table 5, where each column represents a separate least-square regression. For example, in the first column, the one-month-ahead realized return on the S&P 500 index is regressed on the 10th percentile of the estimated state price density. The p-values that are below 10% are in bold. When the financial crisis is excluded from the sample (i.e.. for $2009.6 - 2016.4$), the estimates are all statistically significant. Because the first-order correlations of the predictors are below 0.80, this evidence for predictability is unlikely to be spurious. However, when the financial crisis is included in the sample (i.e., for $2007.1 - 2016.4$), none of the estimates is statistically significant. Is this break-down due to the unusual nature of the recent crisis, or is it commonly associated with major market disruptions? This question requires an extended sample and is left for future study. Note that the estimates for the lower quantiles are negative, while those for the upper quantiles are positive, implying that lower (higher) quantiles of the density and the future return tend to move in opposite (same) directions. This pattern is consistent with a risk-return trade-off; that is, a higher dispersion of the state price density predicts a higher return on the S&P 500 index. The above evidence for predictability is based on in-sample estimates. To examine out-of-sample predictability, we would need to generate estimates of the regression coefficients and then compare the forecasting performance with a benchmark model. We leave this investigation for future work.

The results for the predictive quantile regressions are summarized in Table 6, where, in each column, a quantile of the state price density is used to predict the corresponding quantile of the return distribution. Interestingly, for quantiles that are above the median, the estimates are statistically significant, irrespective of whether the financial crisis is included in the sample. Meanwhile, for quantiles that are below the median, the estimates are insignificant in most cases. This asymmetry merits further consideration.

We also repeat the analysis using the trading days with at least 50, instead of 40, different option strikes. The results (see Tables S3 and S4) are essentially the same as those in Tables 5 and 6. In addition, we consider a longer forecasting horizon of 60 days. The findings (see Tables S5 and S6) are qualitatively similar to those in Tables 5 and 6, showing that they are not unique to the monthly horizon.

The results for the VIX case are reported in Supplemental Tables S7 and S8. The estimates are significant in all cases. This finding is unsurprising because VIX is strongly serially correlated. At the same time, it provides favorable evidence for the performance of the proposed estimator.

The above analysis contributes to the growing literature that uses some aspects of the state price density to predict future returns or risk premia. For example, Bollerslev, Tauchen, and Zhou (2009) show that the variance risk premium (the gap between the VIX and realized variance) can predict the equity risk premium. Some studies have used the tail of the state price density to forecast future

returns; see Andersen, Fusari, and Todorov (2015) by parametric methods and Andersen, Todorov, and Ubukata (2020) by nonparametric methods. Interestingly, most of these procedures tend to find predictability at intermediate horizons, e.g., three to eight months. Results obtained using our methods suggest that such predictability exists at shorter horizons, of one and two months.

5.3 A measure for market perception

Every state price density represents the market's valuation about a future day. Given this property, we propose to use a sequence of densities corresponding to different option maturities to measure the market's perception about the near future. We illustrate this proposal in Figure 15, where we display the quantiles of nine state price densities estimated using VIX options on May 16, 2016, for $\tau = 9, 16, 23, 30, 37, 65, 93, 128, 156$. These densities are asymmetric with a long right tail, implying that the market was concerned about downside risk, i.e., the risk associated with a higher VIX value than its current level. This concern is more pronounced at longer horizons. For example, when $\tau = 9$, 65, and 156, the density's medians are $15.9, 16.6$, and 17.5, respectively, while their 0.9 quantiles are 19.9, 28.4, and 33.1. This measure is model-free because it does not apply any parametric model for the VIX index. Its value reflects not only expectations of future volatility levels, but also the associated risk premia.

5.4 Identifying causal effects using high-frequency data

The proposed estimator can be used to identify and estimate the impact of a monetary policy shock on the financial market. We illustrate this point by considering two FOMC announcements made on December 18, 2013 and January 27, 2016. In each case, we estimate the state price densities implied by the S&P500 options over five-minute intervals within a three-hour window of the policy announcement. The quantiles of the estimates are displayed in Figure 16.

On December 18, 2013, the Committee announced that it would be tapering back on QE3, by slowing down its purchase of mortgage-backed securities at a pace of $35 billion per month rather than $40 billion per month, and the longer-term Treasury securities at a pace of $40 billion per month rather than $45 billion per month. It also reaffirmed that the exceptionally low target range for the federal funds rate of 0 to 0.25% would be in place at least as long as the unemployment rate remained above 6.5%. Figure 16(a) reveals that the state price density remained approximately unchanged until the announcement, reacted immediately at the news, and then continued to evolve over the next 90-minute interval. The most significant change in the density is a reduction in its left tail. The changes in the quantiles above the median are much smaller. Therefore, the main causal effect of this announcement is on the perception of tail probability and/or the associated risk premium, not the overall growth potential.

On January 27, 2016, the Committee announced that it would maintain the target range for the federal funds rate at 0.25% to 0.5%. It also expressed a view that the economic conditions would warrant only gradual increases in the federal funds rate and that the federal funds rate would likely remain, for some time, at a low level. Figure 16(b) shows that the market reacted negatively to the announcement. Different from the previous case, this announcement did not impact the lower or upper tails only. Instead, it caused a small downward change in the overall distribution.

In the above analysis, we have controlled for factors that can affect the market outside a small window by using high-frequency data. In practice, similar analyses can be carried out to study the causal effects of other events on the market. Prior to our work, Hattori, Schrimpf, and Sushko (2016) examined the impact of unconventional monetary policy announcements on investors' expectations of tails risks. Their analysis is at the daily frequency and is based on computing the second-order derivatives of the option pricing function.

6 Conclusion

This paper has proposed a sieve estimator for the state price density implied by a cross-section of European options with different strikes and the same maturity. A variance-stabilizing transformation enables the estimator to perform well across asset classes under different market conditions. The empirical applications demonstrate that the estimator is suitable for both daily and high-frequency data. The results from predictive regressions suggest that the state price density has predictive power for future returns. Finally, the analysis of two FOMC announcements shows that the method can be used to identify and estimate causal effects using high-frequency data.

References

Abken, P. A., D. B. Madan, and B. S. Ramamurtie (1996): "Estimation of risk-neutral and statistical densities by Hermite polynomial approximation: with an application to Eurodollar futures options," FRB Atlanta Working Paper 96-5, Federal Reserve Bank of Atlanta.

Aït-Sahalia, Y., and J. Duarte (2003): "Nonparametric option pricing under shape restrictions," *Journal of Econometrics*, 116(1-2), 9–47.

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

Andersen, T. G., N. Fusari, and V. Todorov (2015): "The risk premia embedded in index options," *Journal of Financial Economics*, 117(3), 558 – 584.

Andersen, T. G., V. Todorov, and M. Ubukata (2020): "Tail risk and return predictability for the Japanese equity market," *Journal of Econometrics*.

Backus, D. K., S. Foresi, and L. Wu (2004): "Accounting for biases in Black-Scholes," Discussion paper, New York University.

Bassett, G. (1997): "Nonparametric bounds for the probability of future prices based on option values," *L1-Statistical Procedures and Related Topics*, 31, 288–300.

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

Beber, A., and M. W. Brandt (2006): "The effect of macroeconomic news on beliefs and preferences: Evidence from the options market," *Journal of Monetary Economics*, 53(8), 1997 – 2039.

Birru, J., and S. Figlewski (2012): "Anatomy of a meltdown: the risk neutral density for the S&P 500 in the fall of 2008," *Journal of Financial Markets*, 15(2), 151–180.

Blundell, R., X. Chen, and D. Kristensen (2007): "Semi-nonparametric IV esstimation of shape-invariant Engel Curves," *Econometrica*, 75(6), 1613–1669.

Bollerslev, T., G. Tauchen, and H. Zhou (2009): "Expected Stock Returns and Variance Risk Premia," *The Review of Financial Studies*, 22(11), 4463–4492.

Bondarenko, O. (2003): "Estimation of risk-neutral densities using positive convolution approximation," *Journal of Econometrics*, 116(1), 85 – 112, Frontiers of financial econometrics and financial engineering.

Boyd, J. P. (1984): "Asymptotic coefficients of Hermite function series," *Journal of Computational Physics*, 54, 382–410.

Breeden, D. T., and R. H. Litzenberger (1978): "Prices of state-contingent claims implicit in option prices," *Journal of business*, pp. 621–651.

Broadie, M., M. Chernov, and M. Johannes (2007): "Model specification and risk premia: evidence from futures options," *Journal of Finance*, 62, 1453–1490.

Carr, P., and L. Wu (2003): "The finite moment log stable process and option pricing," *The Journal of Finance*, 58(2), 753–777.

Carrasco, M., J.-P. Florens, and E. Renault (2007): "Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization," *Handbook of econometrics*, 6, 5633–5751.

Cox, J. C., and S. A. Ross (1976): "The valuation of options for alternative stochastic processes," *Journal of Financial Economics*, 3(1-2), 145–166.

Cramér, H. (1946): *Mathematical methods of statistics*. Princeton university press.

Du, Y., C. Wang, and Y. Du (2012): "Inversion of option prices for implied risk-neutral probability density functions: general theory and its applications to the natural gas market," *Quantitative Finance*, 12(12), 1877–1891.

Duffie, D., J. Pan, and K. Singleton (2000): "Transform analysis and asset pricing for affine jump-diffusions," *Econometrica*, 68(6), 1343–1376.

Eraker, B. (2004): "Do stock prices and volatility jump? Reconciling evidence from spot and option prices," *Journal of Finance*, 59(3), 1367–1404.

Figlewski, S. (2010): "Estimating the implied risk neutral density for the U.S. market portfolio," *Volatility and TimeSeries Econometrics: Essays in Honor of Robert F. Engle*.

Fuhry, M., and L. Reichel (2012): "A new Tikhonov regularization method," *Numerical Algorithms*, 59(3), 433–445.

Gajek, L. (1986): "On improving density estimators which are not bona fide functions," *The Annals of Statistics*, 14(4), 1612–1618.

Grenander, U. (1981): *Abstract inference*. Wiley New York.

Harrison, J. M., and D. M. Kreps (1979): "Martingales and arbitrage in multiperiod securities markets," *Journal of Economic Theory*, 20(3), 381–408.

Hattori, M., A. Schrimpf, and V. Sushko (2016): "The response of tail risk perceptions to unconventional monetary policy," *American Economic Journal: Macroeconomics*, 8(2), 111–36.

Hille, E. (1940): "Contributions to the theory of Hermitian series II. the representation problem," *Transactions of the American Mathematical Society*, pp. 80–94.

Horowitz, J., and S. Lee (2012): "Uniform confidence bands for functions estimated nonparametrically with instrumental variables," *Journal of Econometrics*, 168(2), 175–188.

Horowitz, J. L. (2011): "Applied nonparametric instrumental variables estimation," *Econometrica*, 79(2), 347–394.

——— (2014): "Adaptive nonparametric instrumental variables estimation: Empirical choice of the regularization parameter," *Journal of Econometrics*, 180(2), 158–173.

Jackwerth, J. C., and M. Rubinstein (1996): "Recovering probability distributions from option prices," *Journal of Finance*, 51(5), 1611–32.

Jarrow, R., and A. Rudd (1982): "Approximate option valuation for arbitrary stochastic processes," *Journal of Financial Economics*, 10(3), 347–369.

Kitsul, Y., and J. H. Wright (2013): "The economics of options-implied inflation probability density functions," *Journal of Financial Economics*, 110(3), 696–711.

Kress, R. (2014): *Linear integral equations*. Springer-verlag, third edition edn.

Li, H., and F. Zhao (2009): "Nonparametric estimation of state-price densities implicit in interest rate cap prices," *Review of Financial Studies*, 22(11), 4335–4376.

Longstaff, F. A. (1995): "Option pricing and the martingale restriction," *The Review of Financial Studies*, 8(4), 1091–1124.

Melick, W. R., and C. P. Thomas (1997): "Recovering an asset's implied PDF from option prices: an application to crude oil during the Gulf crisis," *Journal of Financial and Quantitative Analysis*, 32(01), 91–115.

Rubinstein, M. (1994): "Implied binomial trees," *The Journal of Finance*, 49(3), 771–818.

Shimko, D. (1993): "Bounds of probability," *Risk*, 6, 33–37.

Song, Z., and D. Xiu (2016): "A tale of two option markets: pricing kernels and volatility risk," *Journal of Econometrics*, 190(1), 176–196.

Szegö, G. (2003): *Orthogonal polynomials*, Colloquium Publications, Volume 23. American Mathematical Society.

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

Table 1: Parameter Estimates for the Four Parametric Models in Three Volatility Regimes

	BS			**SV**			**SVJ**			**SVCJ**		
	crisis	recovery	expansion	crisis	recovery	expansion	crisis	recovery	expansion	crisis	recovery	expansion
k				0.0128	0.0070	0.0116	0.0100	0.0043	0.0094	0.0107	0.0047	0.0090
θ	3.6283	1.2453	0.4300	3.7053	2.0417	0.6644	3.0485	1.7498	0.6375	2.5483	1.5187	0.6113
σ_v				0.4936	0.1792	0.1403	0.4267	0.1207	0.1078	0.4440	0.1195	0.1119
ρ				-0.5268	-0.5268	-0.5268	-0.6655	-0.6655	-0.6655	-0.6108	-0.6108	-0.6108
λ							0.0044	0.0044	0.0044	0.0035	0.0035	0.0035
μ_s							-10.1538	-4.7521	-1.8591	-12.3802	-6.4082	-3.0804
σ_s							6.5910	7.2807	2.5553	6.0028	6.6369	2.4780
μ_v										1.4579	0.3005	0.0501

Note. The estimates correspond to daily returns in percentages. They are obtained in two steps. First, we estimate each model under the physical measure by MCMC using S&P 500 daily returns from January 2, 1980 to December 30, 2016. Then, we follow Broadie et al. (2007) and estimate the models under the risk neutral measure using Wednesday data on the S&P 500 index and its options. We report the estimates for three representative periods: crisis (2008.10-2009.04), recovery (2009.06-2009.12) and expansion (2013.06-2013.12), where the average VIX levels are 50, 25 and 15, respectively. Because ρ and λ are challenging to identify and are invariant between the two measures, we fix their values at the estimates from the first step. The variance parameter of Black-Scholes (BS) model is reported in the θ row.

Table 2: Tail Decay Rates and Hermite Coefficients

	Crisis			Recovery			Expansion		
	r	κ(L)	κ(R)	r	κ(L)	κ(R)	r	κ(L)	κ(R)
BS	1.103	2.051	1.952	1.089	2.028	1.973	1.069	2.016	1.984
SV	0.461	1.287	2.142	0.534	1.460	2.358	0.552	1.329	2.298
SVJ	0.464	1.361	2.889	0.537	1.452	2.582	0.482	1.423	2.810
SVCJ	0.523	1.342	2.636	0.629	1.446	2.528	0.535	1.408	2.701

Note. The tail decay is estimated using $f(x) = b\exp(-a\left|x\right|^{\kappa}) + u(x)$ with $x \in [0.5, 5]$ for the right tail and $x \in [-5, -0.5]$ for the left tail. The estimates are denoted by κ(R) and κ(L), respectively. The Hermite coefficients are estimated using $\log(|\beta_j|) = q - pj^r + e_j$, where $j = 0, 1, ..., 8$ for the BS model, and $j = 0, 1, ..., 30$ and for the remaining three models. The logarithm transformation is used because the values of $|\beta_j|$ are small after the first few j. The upper bound of j for the BS model is smaller because $|\beta_j|$ approaches zero very rapidly to the extent that the values are dominated by rounding errors when j is outside this range. All estimates are significant at the 1% level.

Table 3: Mean Integrated Squared Errors under different J and Q_α

	Crisis Period			Recovery Period			Expansion Period		
J	$Q_\alpha = 0$	$Q_\alpha = \alpha I$	$Q_\alpha = VD_\alpha V'$	$Q_\alpha = 0$	$Q_\alpha = \alpha I$	$Q_\alpha = VD_\alpha V'$	$Q_\alpha = 0$	$Q_\alpha = \alpha I$	$Q_\alpha = VD_\alpha V'$
1	0.0884	0.0882	0.0884	0.0679	0.0678	0.0679	0.1890	0.1889	0.1890
2	0.0817	0.0817	0.0817	0.0591	0.0593	0.0591	0.1756	0.1761	0.1756
3	0.0376	0.0375	0.0376	0.0152	0.0153	0.0152	0.0589	0.0590	0.0598
4	0.0126	0.0123	0.0126	0.0104	0.0094	0.0098	0.0549	0.0467	0.0494
5	0.0119	0.0113	0.0121	0.0173	0.0142	0.0142	0.0697	0.0558	0.0706
6	0.0111	0.0104	0.0105	0.0214	0.0170	0.0129	0.0779	0.0615	0.0692
7	0.0112	0.0096	0.0082	0.0361	0.0274	0.0179	0.1058	0.0839	0.0841
8	0.0150	0.0123	0.0099	0.0439	0.0328	0.0220	0.1229	0.0948	0.0870
9	0.0195	0.0155	0.0113	0.0526	0.0374	0.0269	0.1541	0.1094	0.1015
10	0.0236	0.0180	0.0124	0.0650	0.0426	0.0320	0.1813	0.1204	0.1097

The DGP is the SVCJ model and the parameter values are given in Table 1. The integrated squared error is computed as $\int_{-\infty}^{\infty} |\hat{f}(x) - f(x)|^2 dx$, where the integration is approximated by a summation over a grid. The averages over 5000 replications are reported. In each case, the three columns display the MSEs with no regularization ($Q_\alpha = 0$), the standard Tikhonov regularization ($Q_\alpha = \alpha I$), and the modified Tikhonov of Fuhry and Reichel ($Q_\alpha = VD_\alpha V'$), respectively.

Table 4: Mean Integrated Squared Errors of the Estimators

Estimator		$n_c = n_p = 25$			$n_c = n_p = 50$		
	J	$Q_\alpha = 0$	$Q_\alpha = \alpha I$	$Q_\alpha = VD_\alpha V'$	$Q_\alpha = 0$	$Q_\alpha = \alpha I$	$Q_\alpha = VD_\alpha V'$
Sieve	3	0.0373	0.0374	0.0373	0.0365	0.0364	0.0365
	4	0.0501	0.0397	0.0445	0.0315	0.0267	0.0286
	5	0.0679	0.0513	0.0579	0.0409	0.0327	0.0365
	6	0.0881	0.0637	0.0571	0.0522	0.0386	0.0351
Local linear			0.1011			0.0870	
PCA			0.2675			0.1883	
Parametric			0.0477			0.0280	

Note. For the sieve estimator, the three columns display the MISEs over 5000 replication with no regularization ($Q_\alpha = 0$), the standard Tikhonov regularization ($Q_\alpha = \alpha I$), and the modified Tikhonov of Fuhry and Reichel ($Q_\alpha = VD_\alpha V'$), respectively. Setting J =ceiling$(2(n/\log n)^{1/5})$, gives J=4 for both sample sizes. The parameter α is determined by 10-fold cross validation. The last three rows of the table correspond to the constrained local linear estimator of Ait-Sahalia and Duarte (2003), the positive convolution approximation estimator of Bondarenko (2003), and the four-parameter Gram-Charlier series estimator of Jarrow and Rudd (1982) and Longstaff (1995).

Table 5: Predictive Mean Regressions for S&P500 Returns Using a Quantile as the Predictor

Quantile	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
01/18/2007 - 04/27/2016									
Estimate	-0.12	-0.36	-0.72	-1.84	0.63	0.57	0.42	0.30	0.21
s.d.	0.14	0.34	0.62	1.37	1.91	0.86	0.50	0.34	0.24
p-value	0.42	0.30	0.25	0.16	0.73	0.51	0.41	0.39	0.39
R-square	0.01	0.01	0.02	0.02	0.00	0.00	0.01	0.01	0.01
06/18/2009 - 04/27/2016									
Estimate	-0.36	-1.27	-2.45	-5.34	3.12	2.07	1.39	0.98	0.69
s.d.	0.09	0.31	0.57	1.40	1.17	0.56	0.36	0.25	0.18
p-value	**1e-4**	**1e-5**	**2e-5**	**2e-4**	**8e-3**	**3e-4**	**2e-4**	**1e-4**	**1e-4**
R-square	0.06	0.10	0.10	0.07	0.03	0.07	0.08	0.08	0.08

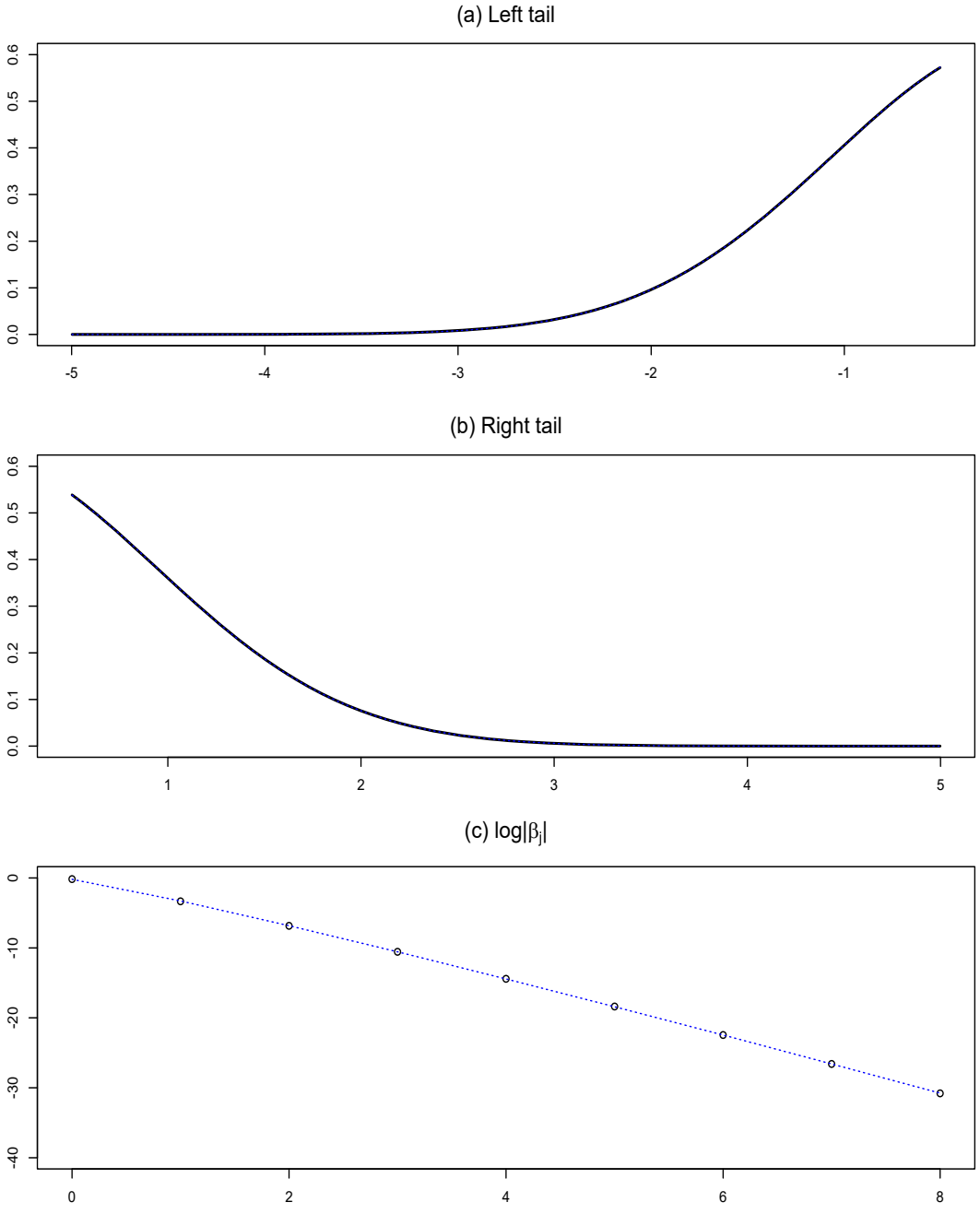
Note. Each column represents a least square regression. For example, the first column displays the results from a least square regression of the one-month-ahead realized return on the 10th percentile of the state price density. The standard errors allow for heteroscedasticity and autocorrelation. The estimates that are significant at the 10% level are in bold.

Table 6: Predictive Quantile Regressions for S&P500 Returns

Quantile	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
01/18/2007 - 04/27/2016									
Estimate	0.25	0.28	-0.60	-1.58	1.60	2.15	1.61	1.01	0.79
s.d.	0.06	0.97	1.06	1.61	1.78	0.55	0.19	0.14	0.21
p-value	**5e-5**	0.77	0.57	0.33	0.37	**1e-4**	**1e-16**	**2e-13**	**2e-4**
06/18/2009 - 04/27/2016									
Estimate	-0.28	-0.72	-1.72	-5.98	2.93	2.14	1.86	1.33	1.22
s.d.	0.24	0.61	1.16	2.70	1.93	0.62	0.53	0.35	0.21
p-value	0.24	0.24	0.14	**0.03**	0.13	**4e-5**	**5e-4**	**1e-4**	**1e-8**

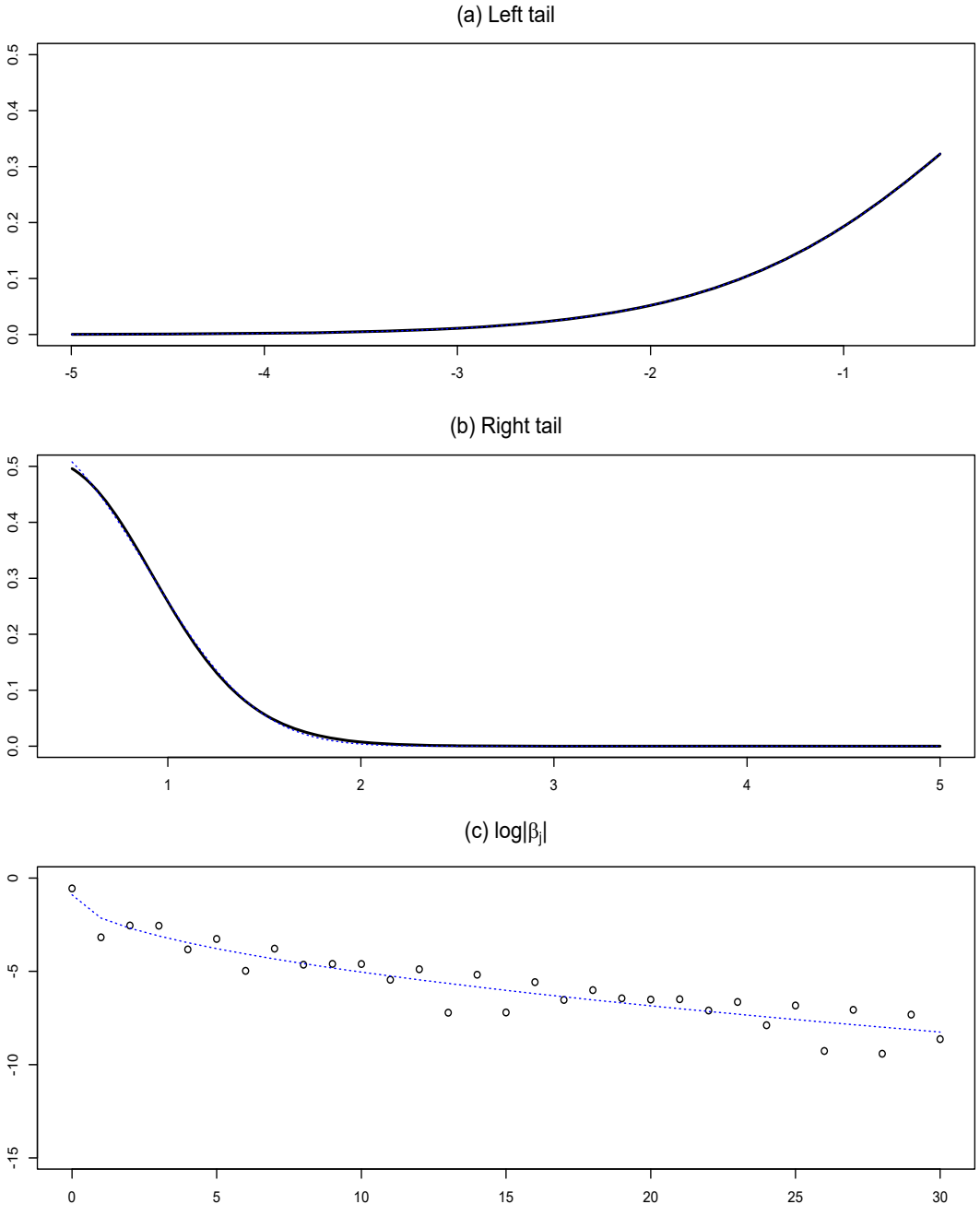
Note. Each column represents a quantile regression. For example, the first column displays the results from a 10th percentile quantile regression of the one-month-ahead realized return on the 10th percentile of the state price density. The standard errors allow for heteroscedasticity and autocorrelation. The estimates that are significant at the 10% level are in bold.

Figure 1: Density tails and Hermite coefficients implied by the Black-Scholes model (the crisis period)



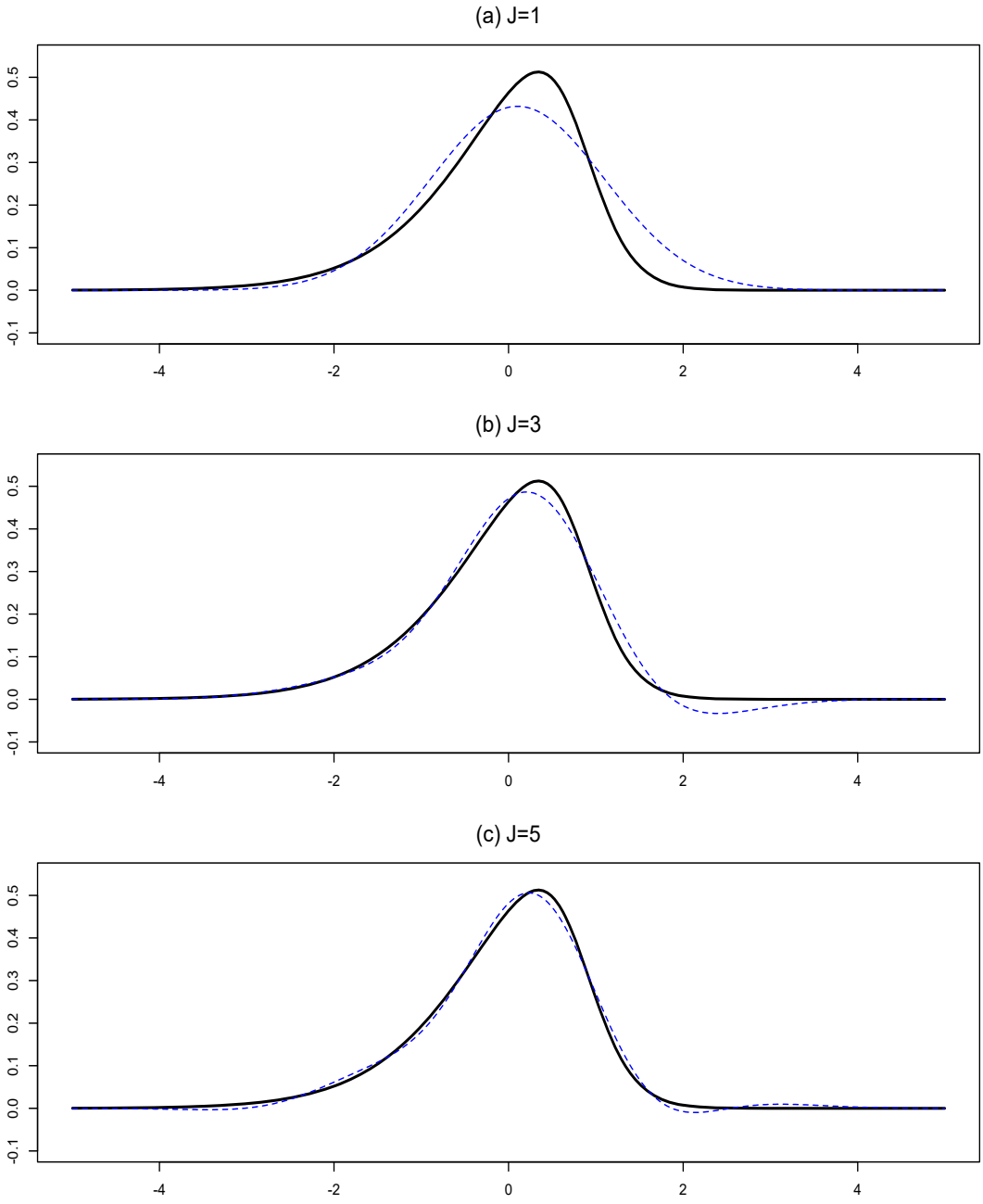
Note. In (a) and (b), the solid line represents the true values, and the dotted line (indistinguishable from the solid line) represents the fitted values from the regressions. In (c), the circles denote the actual Hermite coefficients, and the dotted line the fitted regression line.

Figure 2: Density tails and Hermite coefficients implied by the SVCJ model (the crisis period)



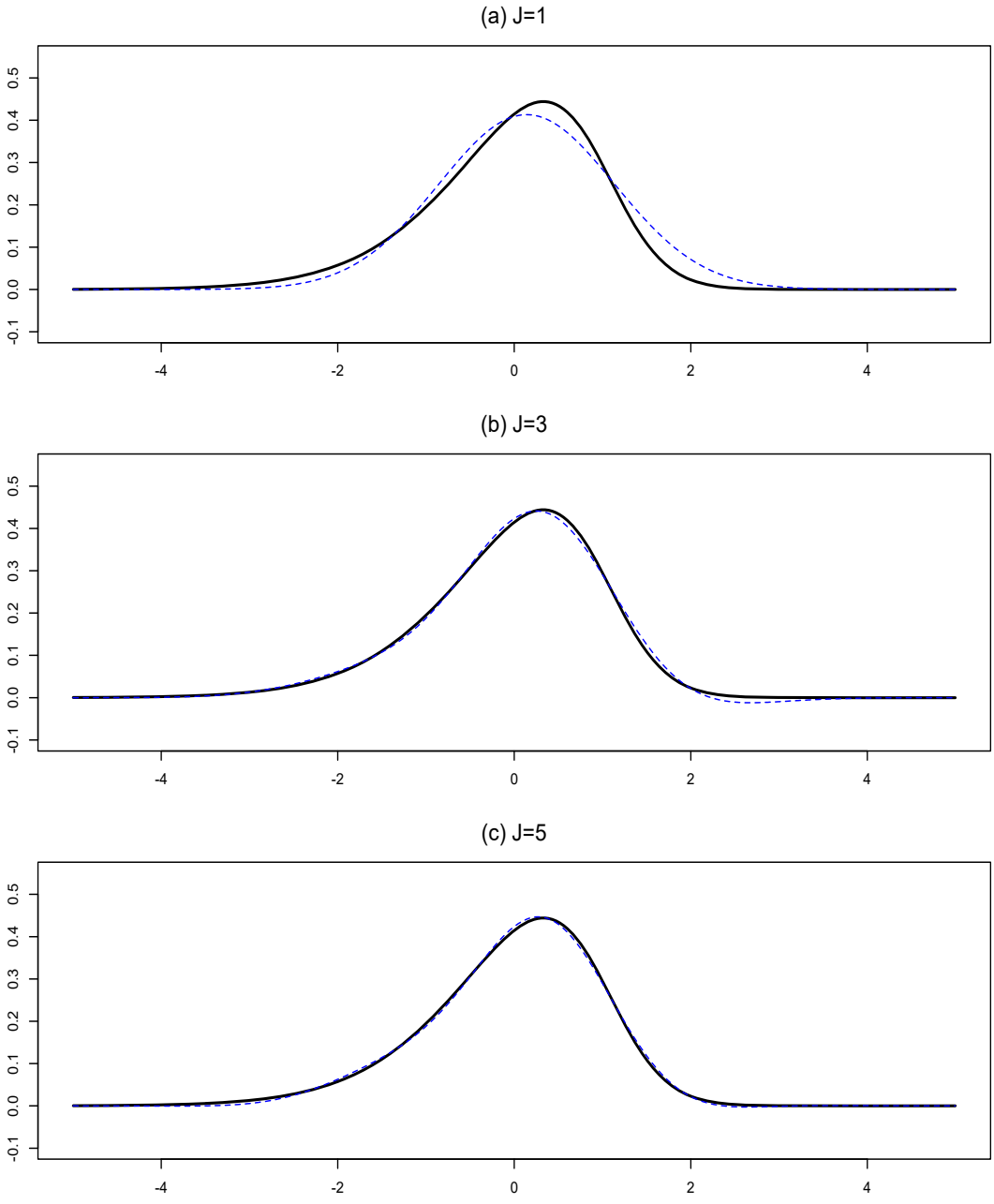
Note. In (a) and (b), the solid line represents the true values, and the dotted line (barely distinguishable from the solid line) represents the fitted values from the regressions. In (c), the circles denote the actual Hermite coefficients, and the dotted line the fitted regression line.

Figure 3: Hermite Series approximations with different truncation orders for the SVCJ model (the crisis period)



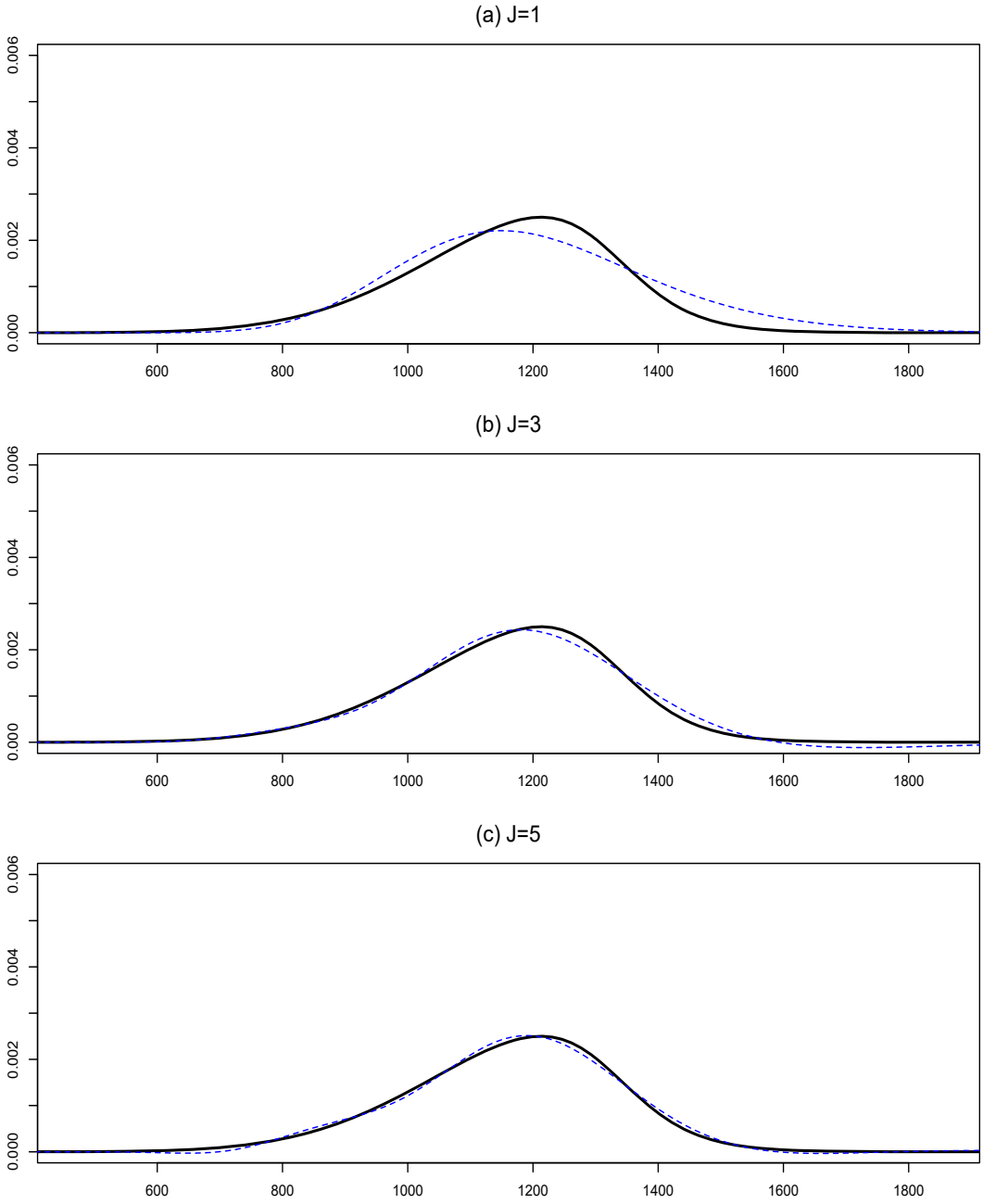
Note. The density is for the transformed variable x. The solid and dashed lines represent the true density and the approximation, respectively.

Figure 4: Hermite Series approximations with different truncation orders for the SVCJ model (the expansion period)



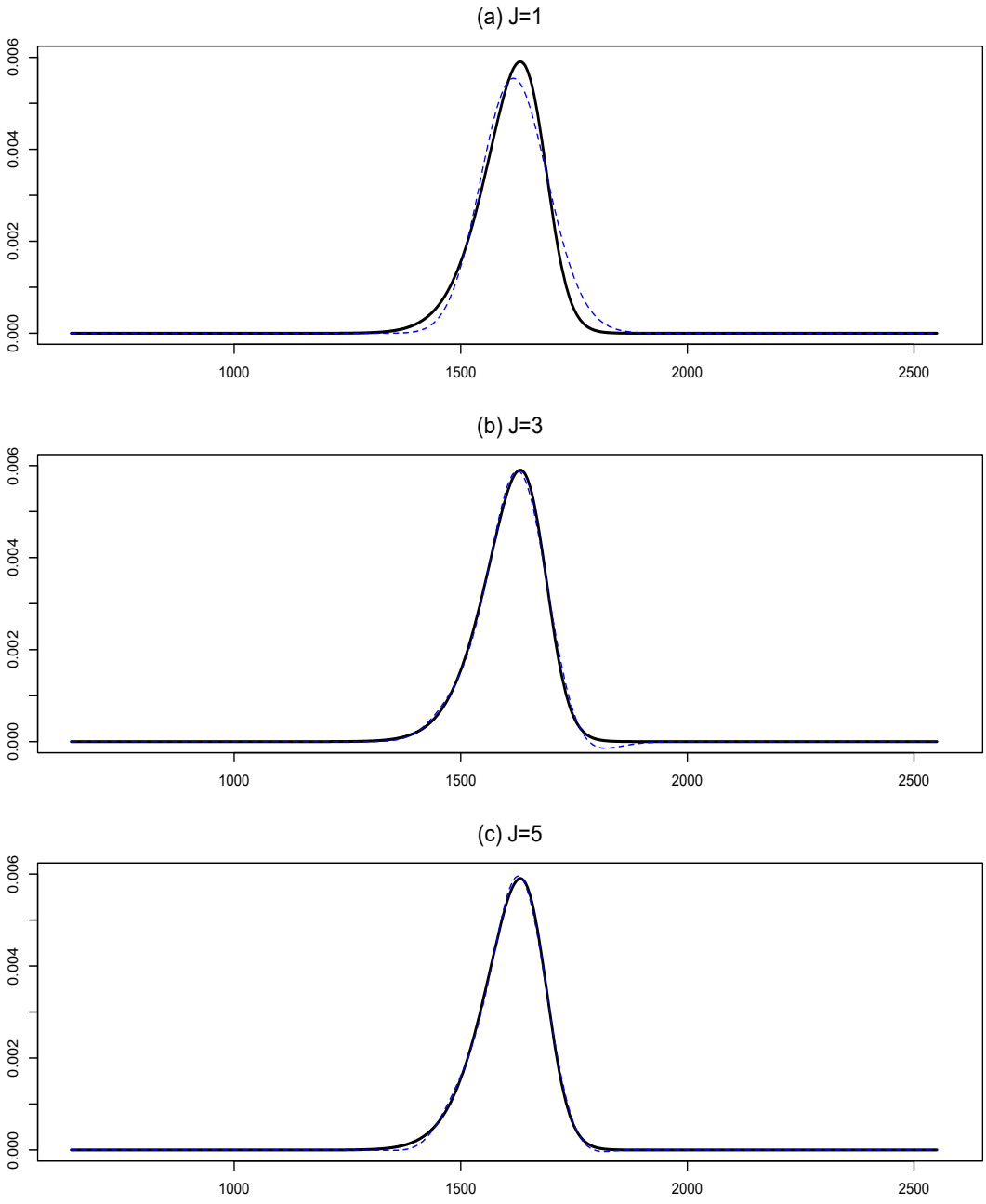
Note. The density is for the transformed variable x. The solid and dashed lines represent the true density and the approximation, respectively.

Figure 5: Hermite series approximations and the original density for the SVCJ model (the crisis period)



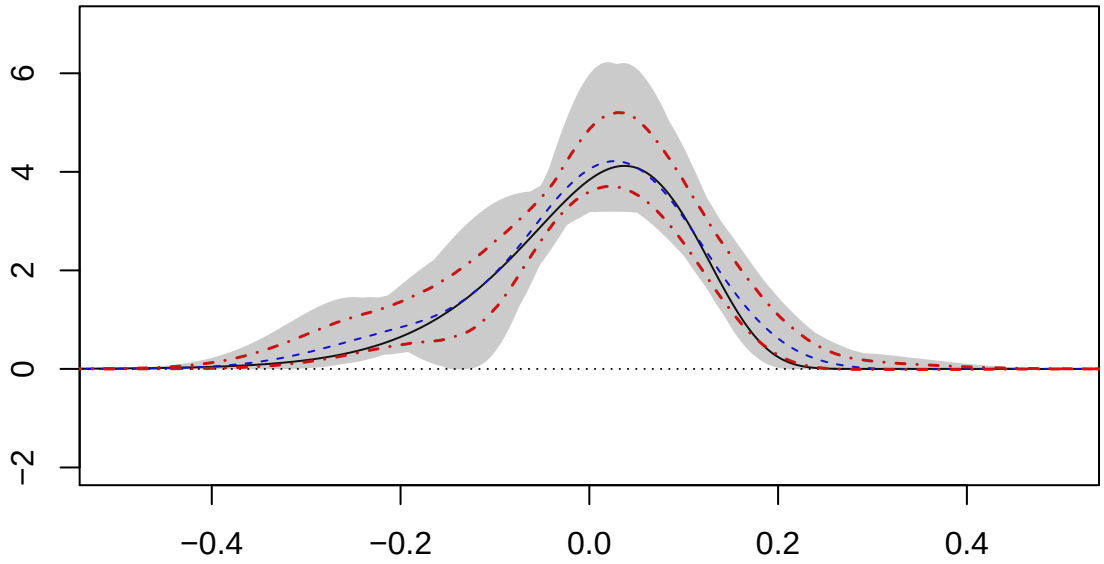
Note. The density is for the original variable S. The solid and dashed lines represent the true density and the approximation, respectively.

Figure 6: Hermite series approximations and the original density for the SVCJ model (the expansion period)



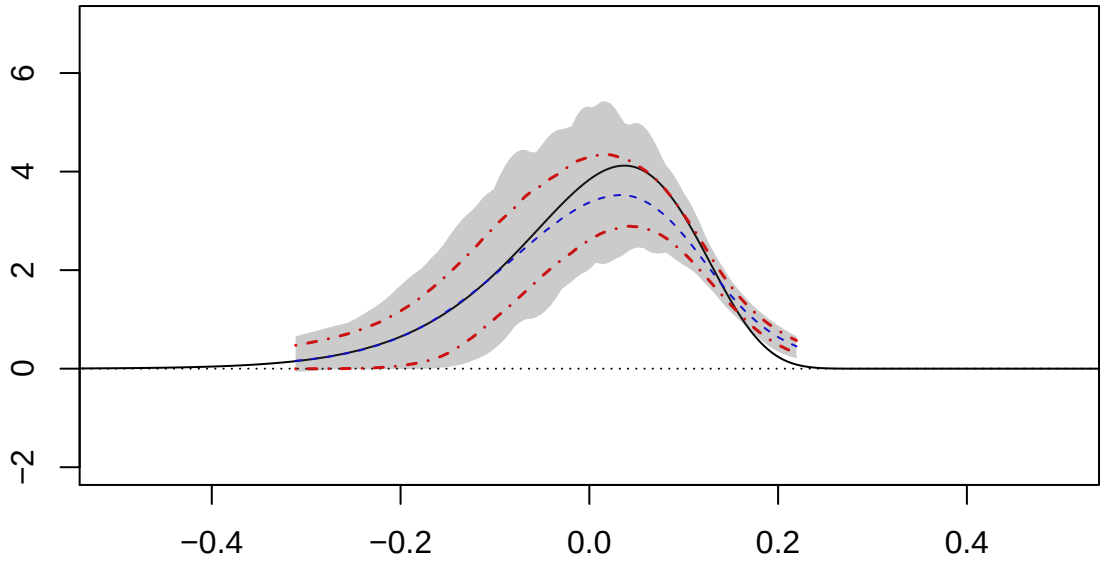
Note. The density is for the original variable S. The solid and dashed lines represent the true density and the approximation, respectively.

Figure 7: The empirical distribution of the Sieve estimator



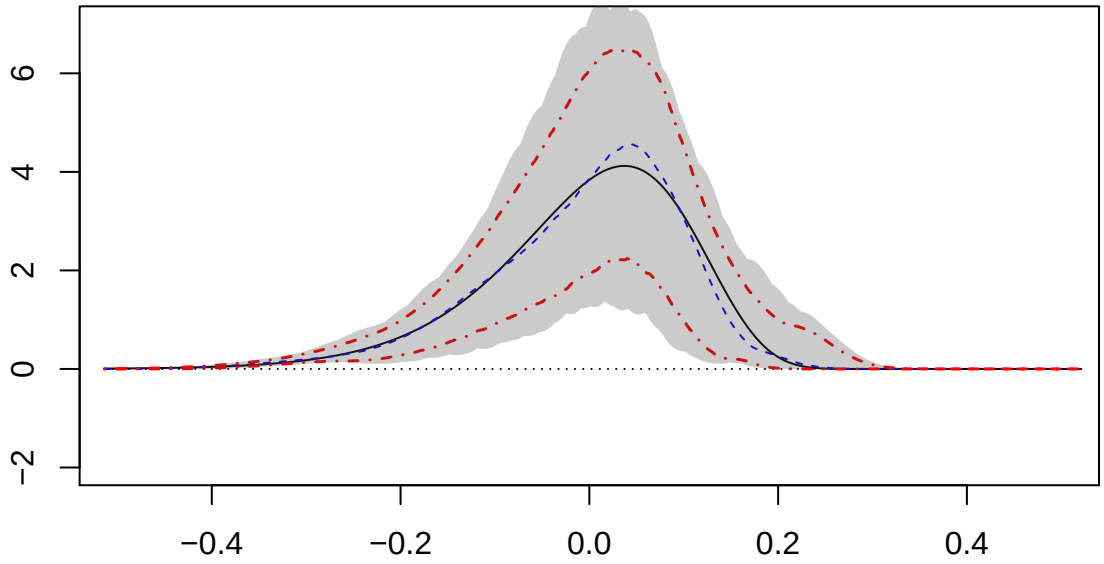
Note. The solid line: the true density; the dashed line: the median; the two dot-dash lines: the 0.025 and 0.975 quantiles; the shaded area: the entire set of estimates.

Figure 8: The empirical distribution of the constrained local linear estimator



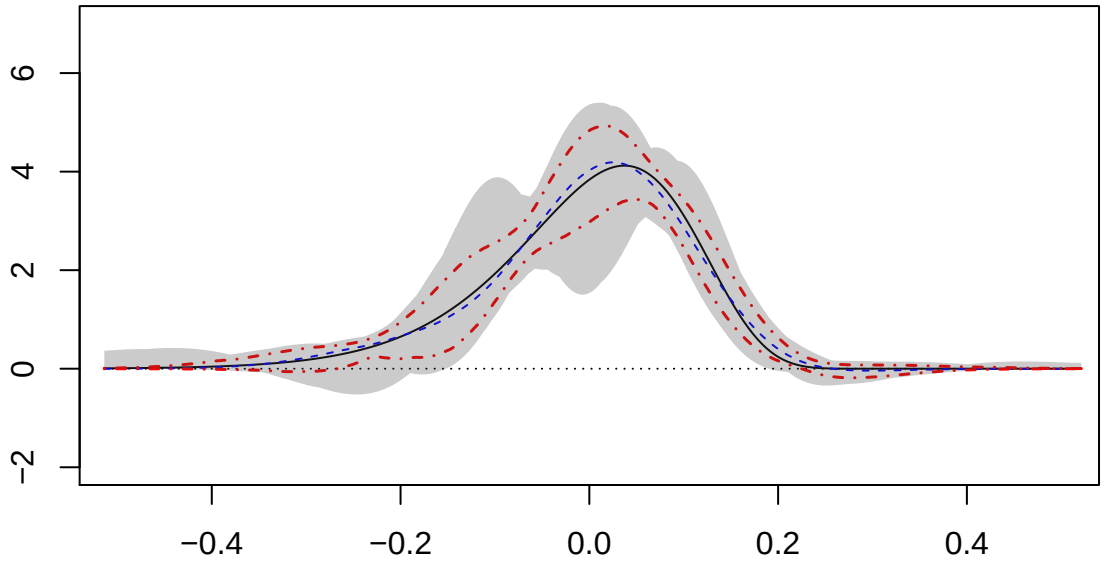
Note. See Figure 7. This estimator is local, therefore it does not estimate the SPD beyond the strike range.

Figure 9: The empirical distribution of the positive convolution approximation estimator



Note. See Figure 7.

Figure 10: The empirical distribution of the parametric estimator



Note. See Figure 7.

Figure 11. State Price Densities Implied by S&P500 Index Options

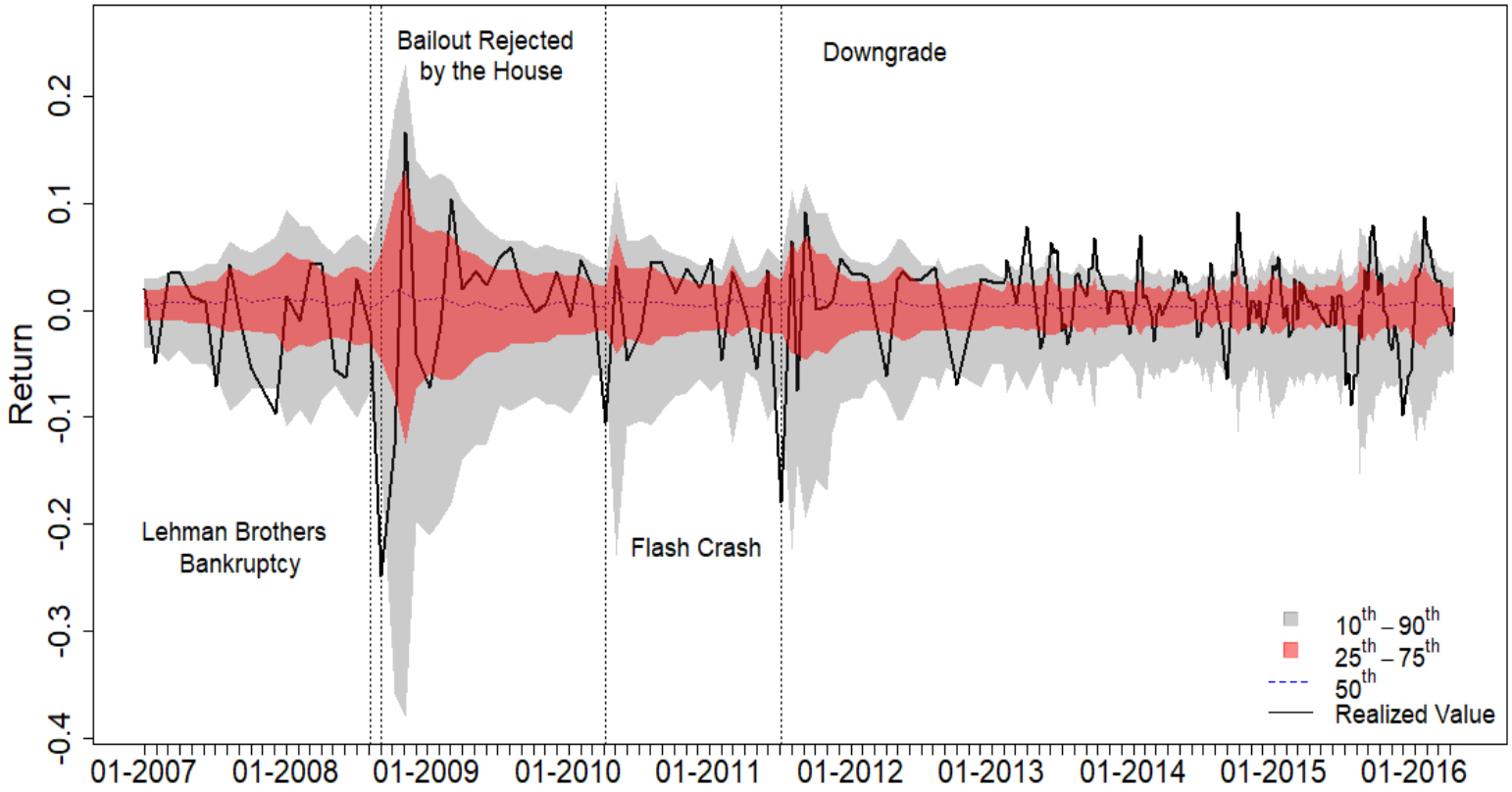


Figure 12: Three representative state price densities implied by S&P500 options

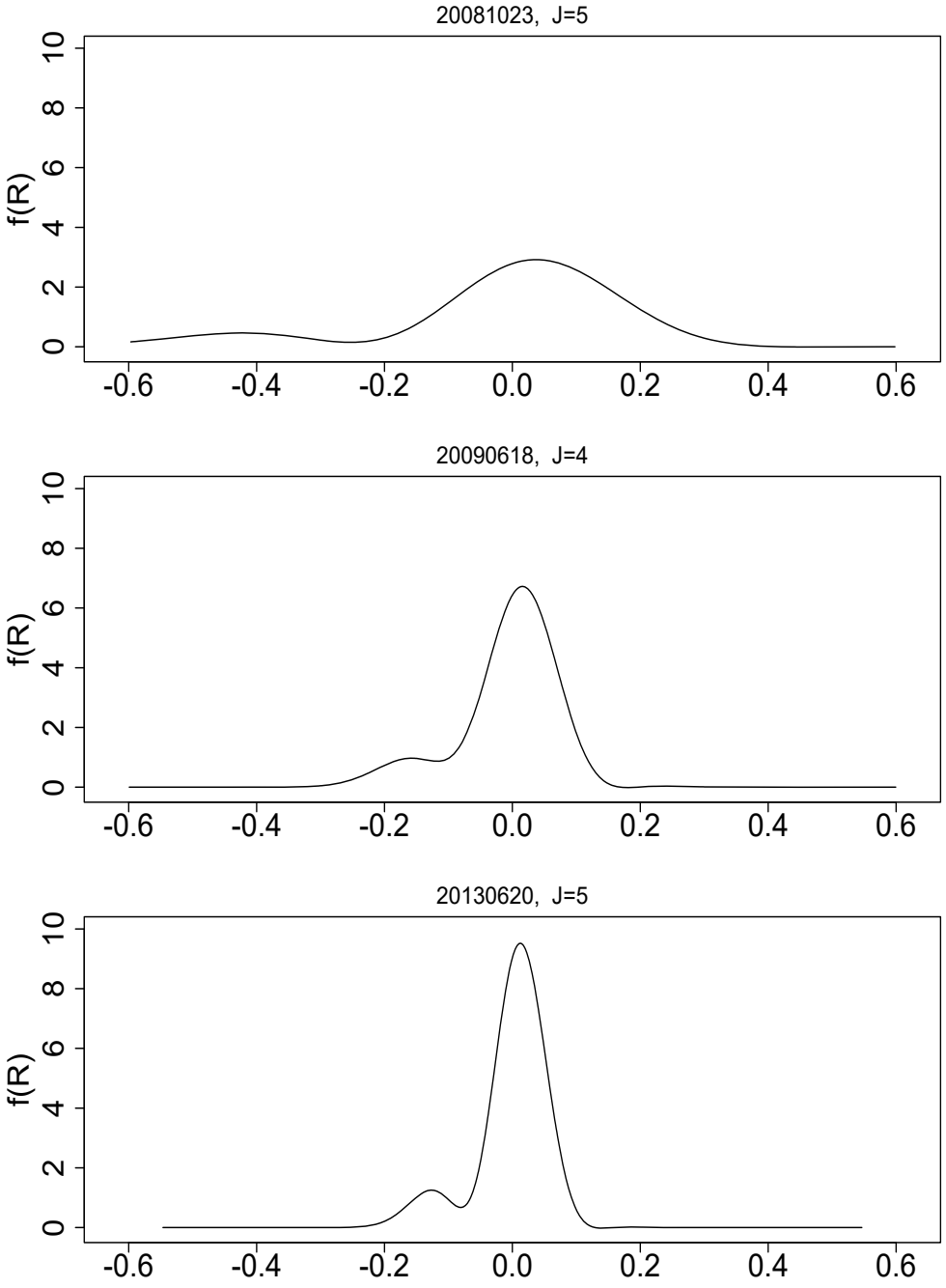


Figure 13. State Price Densities Implied by VIX options with τ=30

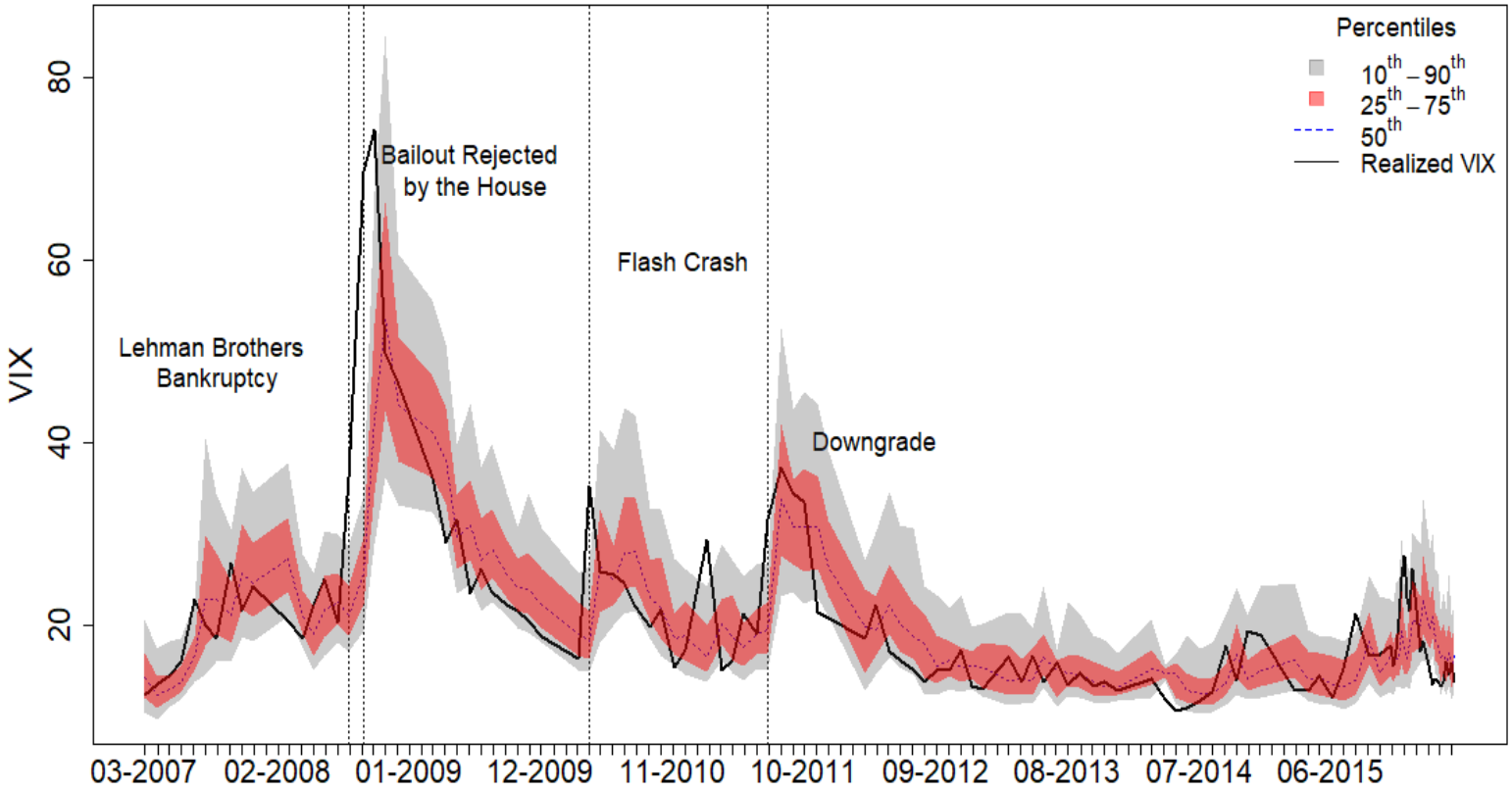


Figure 14: Three representative state price densities implied by VIX options

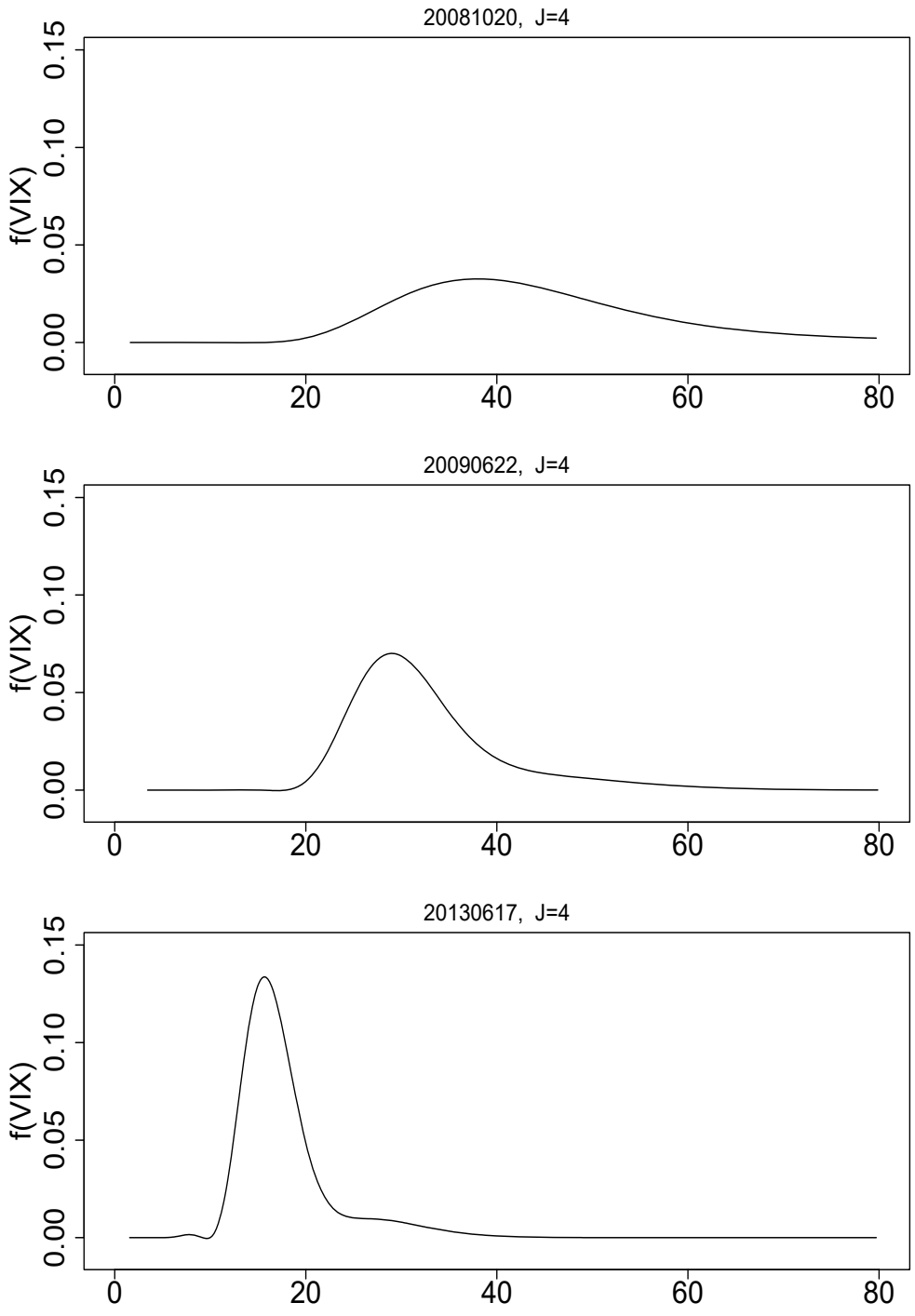


Figure 15. VIX-implied state price densities with an expanding time horizon

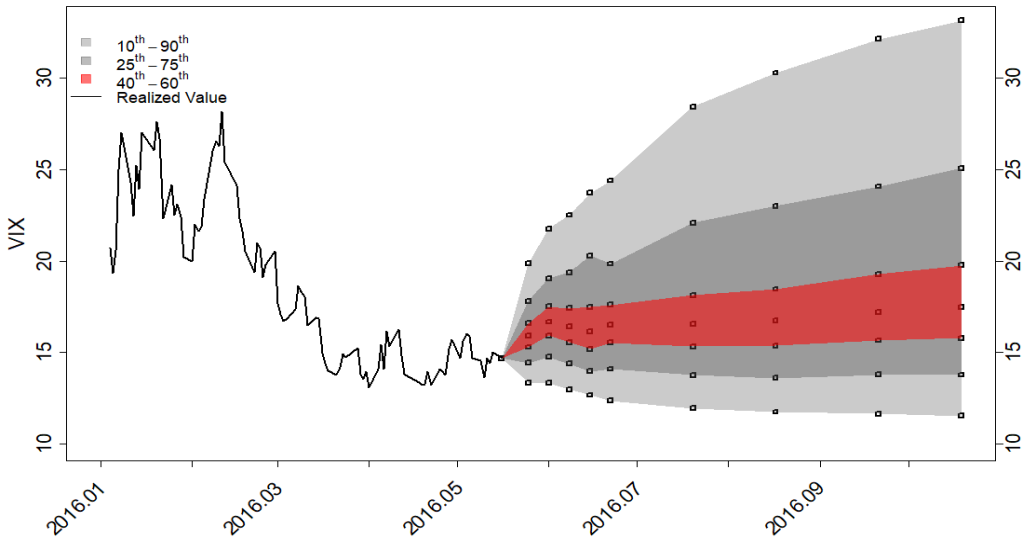
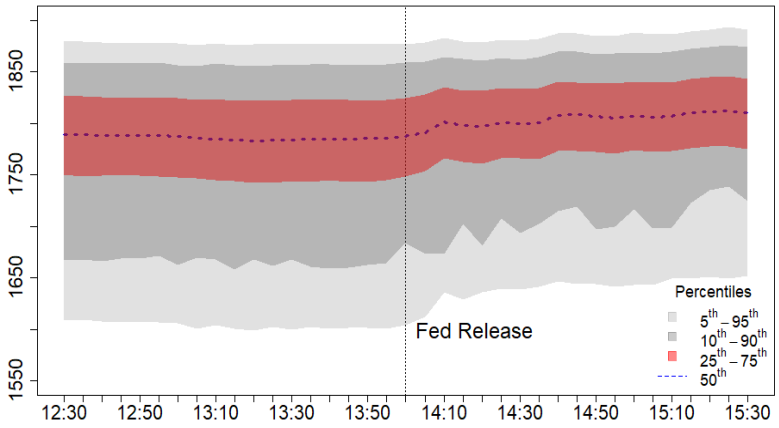
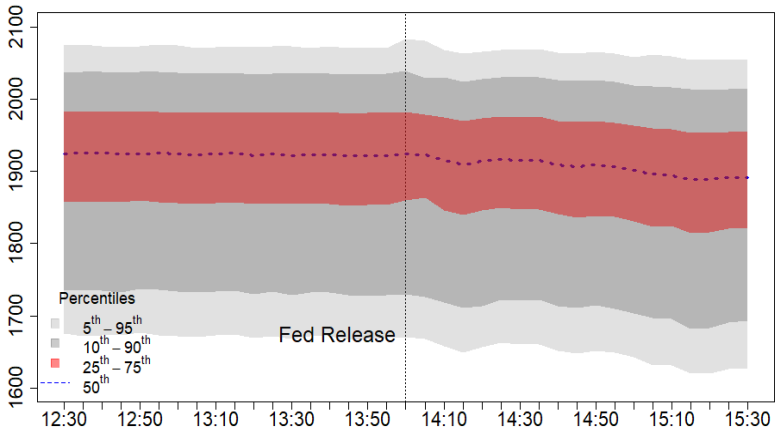


Figure 16: The Effects of Monetary Policy Announcements on State Price Densities

(a) Evolution of the state price density on 2013-12-18



(b) Evolution of the state price density on 2016-01-27



Sieve Estimation of Option-Implied State Price Density

by Junwen Lu and Zhongjun Qu
Boston University

This appendix has three parts. The first part (Online Appendix A) presents the proofs. The second part (Online Appendix B) provides some additional results related to Assumption 2 and the empirical application. The third part consists of some tables and figures.

Online Appendix A

The following notation is used throughout. $||x||$ is the Euclidean norm of a vector x. $||X||$ is the vector induced norm of a matrix X. The symbol "$\to^p$" denotes convergence in probability, and $O_p(\cdot)$ and $o_p(\cdot)$ denote the usual notation for the orders of stochastic magnitude.

Proof of Lemma 1. "(a) $\Rightarrow$ (b)": We first prove $f_1^*(x) = f_0^*(x)$ for any $x \in [\mathcal{K}_L, \mathcal{K}_U]$. Then, we show that the two integral restrictions hold when $x \notin [\mathcal{K}_L, \mathcal{K}_U]$.

The proof for $f_1^*(x) = f_0^*(x)$ for any $x \in [\mathcal{K}_L, \mathcal{K}_U]$ is by contradiction. Suppose that this equality does not hold, that is, $f_1^*(x) \neq f_0^*(x)$ for some $x \in [\mathcal{K}_L, \mathcal{K}_U]$. Without loss of generality, assume $f_1^*(x) > f_0^*(x)$ at this x. Then, we have either $x \in [\mathcal{K}_{c,L}, \mathcal{K}_{c,U}]$ or $x \in [\mathcal{K}_{p,L}, \mathcal{K}_{p,U}]$.

Suppose $x \in [\mathcal{K}_{c,L}, \mathcal{K}_{c,U}]$. By the continuity of $f_1^*(x)$ and $f_0^*(x)$, there exists an interval of positive length, $[l, u] \subseteq [\mathcal{K}_{c,L}, \mathcal{K}_{c,U}]$, such that $f_1^*(x) > f_0^*(x)$ for all $x \in [l, u]$. Define

$$d = \min_{x \in [l,u]} \left(f_1^*(x) - f_0^*(x) \right). \tag{A.1}$$

Consider three strike prices K_1, K_2, and K_3 which belong to $[l, u]$ and satisfy $K_3 = K_2 + \epsilon = K_1 + 2\epsilon$ for some $\epsilon > 0$. By Assumption 1, as in Breeden and Litzenberger (1978), we have

$$\frac{C(K_1) - 2C(K_2) + C(K_3)}{\epsilon^2} = \frac{\int_{K_1}^{K_2} (x - K_1) f_0^*(x) dx}{\epsilon^2} - \frac{\int_{K_2}^{K_3} (x - K_2) f_0^*(x) dx}{\epsilon^2} + \frac{\int_{K_2}^{K_3} f_0^*(x) dx}{\epsilon}. \tag{A.2}$$

The right hand side of (A.2) depends on $f_0^*(x)$ only through $x \in [l, u]$. Let $f_{\min} = \min_{x \in [K_1, K_3]} f_0^*(x)$ and $f_{\max} = \max_{x \in [K_1, K_3]} f_0^*(x)$. Because $(x - K_i) f_{\min} \leq (x - K_i) f_0^*(x) \leq (x - K_i) f_{\max}$ for $x \geq K_i$ and $i = 1, 2$, the first two terms on the right hand side of (A.2) satisfy

$$\frac{1}{2} f_{\min} \leq \frac{\int_{K_1}^{K_2} (x - K_1) f_0^*(x) dx}{\epsilon^2} \leq \frac{1}{2} f_{\max} \quad \text{and} \quad \frac{1}{2} f_{\min} \leq \frac{\int_{K_2}^{K_3} (x - K_2) f_0^*(x) dx}{\epsilon^2} \leq \frac{1}{2} f_{\max},$$

respectively. Therefore, the difference between the first two terms in (A.2) is bounded from below and above by $(f_{\min} - f_{\max})/2$ and $(f_{\max} - f_{\min})/2$, respectively. Because $f_0^*(x)$ is Lipschitz continuous, we have $|f_{\max} - f_{\min}|/2 \leq C(K_3 - K_1)/2 = C\epsilon$, where C is a finite constant independent of l and u. The difference between the first two right hand side terms in (A.2) is thus bounded from below and above by $-C\epsilon$ and $C\epsilon$, respectively. From this result, it follows that (A.2) satisfies

$$-C\epsilon + \frac{\int_{K_2}^{K_3} f_0^*(x) dx}{\epsilon} \leq \frac{C(K_1) - 2C(K_2) + C(K_3)}{\epsilon^2} \leq C\epsilon + \frac{\int_{K_2}^{K_3} f_0^*(x) dx}{\epsilon}. \tag{A.3}$$

Because $f_1^*(x)$ satisfies (1) for all $K \in [\mathcal{K}_{c,L}, \mathcal{K}_{c,U}]$, it must satisfy the inequality (A.3), that is,

$$-C\epsilon + \frac{\int_{K_2}^{K_3} f_1^*(x)dx}{\epsilon} \leq \frac{C(K_1) - 2C(K_2) + C(K_3)}{\epsilon^2} \leq C\epsilon + \frac{\int_{K_2}^{K_3} f_1^*(x)dx}{\epsilon}. \tag{A.4}$$

The results (A.3) and (A.4) jointly imply

$$-C\epsilon + \frac{\int_{K_2}^{K_3} f_1^*(x)dx}{\epsilon} \leq C\epsilon + \frac{\int_{K_2}^{K_3} f_0^*(x)dx}{\epsilon},$$

or, equivalently, $[\int_{K_2}^{K_3}(f_1^*(x) - f_0^*(x))dx]/\epsilon \leq 2C\epsilon$. Applying (A.1) to the left hand side of this inequality, we obtain $d \leq 2C\epsilon$. Because C and d are independent of ϵ, and ϵ can be made arbitrarily small, we arrive at a contradiction.

Now, Suppose $x \in [\mathcal{K}_{p,L}, \mathcal{K}_{p,U}]$. By the identity $(K-x)^+ = (x-K)^+ - (x-K)$, we have

$$\frac{P(K_1) - 2P(K_2) + P(K_3)}{\epsilon^2} = \frac{C(K_1) - 2C(K_2) + C(K_3)}{\epsilon^2}.$$

As a result, the same argument as the $x \in [\mathcal{K}_{c,L}, \mathcal{K}_{c,U}]$ case can be applied, which also leads to a contradiction. Therefore, we have $f_1^*(x) = f_0^*(x)$ for any $x \in [\mathcal{K}_L, \mathcal{K}_U]$.

Next, we consider the situation where $x \notin [\mathcal{K}_L, \mathcal{K}_U]$. Without loss of generality, assume $\mathcal{K}_{c,U} \geq \mathcal{K}_{p,U}$. Let K_1 and K_2 denote two generic strike prices satisfying $\mathcal{K}_{c,L} \leq K_1 < K_2 \leq \mathcal{K}_{c,U}$. By (1),

$$C(K_2) = \int_{K_2}^{\infty} x f_0^*(x)dx - K_2 \int_{K_2}^{\infty} f_0^*(x)dx \text{ and } C(K_1) = \int_{K_1}^{\infty} x f_0^*(x)dx - K_1 \int_{K_1}^{\infty} f_0^*(x)dx. \tag{A.5}$$

This implies

$$C(K_1) - C(K_2) = \int_{K_1}^{K_2} (x - K_1) f_0^*(x)dx + (K_2 - K_1) \int_{K_2}^{\mathcal{K}_U} f_0^*(x)dx + (K_2 - K_1) \int_{\mathcal{K}_U}^{\infty} f_0^*(x)dx,$$

or, equivalently,

$$\int_{\mathcal{K}_U}^{\infty} f_0^*(x)dx = \frac{C(K_1) - C(K_2) - \int_{K_1}^{K_2}(x - K_1) f_0^*(x)dx - (K_2 - K_1)\int_{K_2}^{\mathcal{K}_U} f_0^*(x)dx}{(K_2 - K_1)}.$$

Because the right hand side depends on $f_0^*(x)$ only through $x \in [\mathcal{K}_L, \mathcal{K}_U]$ and $f_0^*(x) = f_1^*(x)$ for any $x \in [\mathcal{K}_L, \mathcal{K}_U]$, we have

$$\int_{\mathcal{K}_U}^{\infty} f_0^*(x)dx = \int_{\mathcal{K}_U}^{\infty} f_1^*(x)dx \tag{A.6}$$

Furthermore, by the first equality in (A.5),

$$\int_{\mathcal{K}_U}^{\infty} x f_0^*(x)dx = C(K_2) + K_2 \int_{K_2}^{\infty} f_0^*(x)dx - \int_{K_2}^{\mathcal{K}_U} x f_0^*(x)dx.$$

The right hand side of this equation yields the same value when $f_0^*(x)$ is replaced by $f_1^*(x)$ because $f_0^*(x) = f_1^*(x)$ for $x \in [\mathcal{K}_L, \mathcal{K}_U]$ and (A.6) holds, This implies

$$\int_{\mathcal{K}_U}^{\infty} x f_0^*(x)dx = \int_{\mathcal{K}_U}^{\infty} x f_0^*(x)dx. \tag{A.7}$$

Combining (A.6) and (A.7), we obtain the integral restriction for the right tail of the density. The left tail can be studied in the same way, by considering two put options with strike prices K_3 and K_4 satisfying $\mathcal{K}_{p,L} \leq K_3 < K_4 \leq \mathcal{K}_{p,U}$. We omit the details.

The proof for "(b) $\Rightarrow$ (a)" is straightforward. For any K satisfying $\mathcal{K}_{c,L} \leq K \leq \mathcal{K}_{c,U}$, by (1),

$$C(K) = \int_K^\infty (x-K) f_0^*(x)dx = \int_K^{\mathcal{K}_U} (x-K) f_0^*(x)dx + \int_{\mathcal{K}_U}^\infty x f_0^*(x)dx - \mathcal{K}_U \int_{\mathcal{K}_U}^\infty f_0^*(x)dx.$$

The right hand side returns the same value when $f_0^*(x)$ is replaced by $f_1^*(x)$ because of the conditions in (b). Therefore, $f_0^*(x)$ and $f_1^*(x)$ yield the same call option prices for $\mathcal{K}_{c,L} \leq K \leq \mathcal{K}_{c,U}$. Similarly, by considering put options with $\mathcal{K}_{p,L} \leq K \leq \mathcal{K}_{p,U}$, we can prove that $f_0^*(x)$ and $f_1^*(x)$ yield the same put option prices for $\mathcal{K}_{c,L} \leq K \leq \mathcal{K}_{c,U}$ when the conditions in (b) hold. ■

Lemma A.1 *Under Assumptions 1-3, we have*

$$\begin{aligned}
\sup_{z \in R} \left| \int_{-\infty}^\infty S\left(e^{\sqrt{\tau}\sigma x} - e^{\sqrt{\tau}\sigma z}\right)^+ f(x)dx - \int_{-c_1 \log n}^{c_2 \log n} S\left(e^{\sqrt{\tau}\sigma x} - e^{\sqrt{\tau}\sigma z}\right)^+ f(x)dx \right| &= O_p\left(n^{-1}\right), \\
\sup_{z \leq B \log n} \left| \int_{-\infty}^\infty S\left(e^{\sqrt{\tau}\sigma z} - e^{\sqrt{\tau}\sigma x}\right)^+ f(x)dx - \int_{-c_1 \log n}^{c_2 \log n} S\left(e^{\sqrt{\tau}\sigma z} - e^{\sqrt{\tau}\sigma x}\right)^+ f(x)dx \right| &= O_p\left(n^{-1}\right),
\end{aligned}$$

where S is the spot price of the underlying asset at time t, and c_1, c_2, and B can be any finite positive constants independent of n.

Proof of Lemma A.1. Without loss of generality, let $c_1 = c_2 = 1$. Setting c_1 and c_2 to generic positive values does not affect the upper bounds that we present below.

For the first equation in the Lemma, we have

$$\begin{aligned}
&\int_{-\infty}^\infty S\left(e^{\sqrt{\tau}\sigma x} - e^{\sqrt{\tau}\sigma z}\right)^+ f(x)dx - \int_{-\log n}^{\log n} S\left(e^{\sqrt{\tau}\sigma x} - e^{\sqrt{\tau}\sigma z}\right)^+ f(x)dx \\
= &\int_{-\infty}^{-\log n} S\left(e^{\sqrt{\tau}\sigma x} - e^{\sqrt{\tau}\sigma z}\right)^+ f(x)dx + \int_{\log n}^\infty S\left(e^{\sqrt{\tau}\sigma x} - e^{\sqrt{\tau}\sigma z}\right)^+ f(x)dx. \qquad \text{(A.8)}
\end{aligned}$$

The first term on the right hand side is positive and is bounded from above by $\int_{-\infty}^{-\log n} S e^{\sqrt{\tau}\sigma x} f(x)dx$ uniformly over $z \in R$. For sufficiently large n, this upper bound satisfies

$$\begin{aligned}
\int_{-\infty}^{-\log n} S e^{\sqrt{\tau}\sigma x} f(x)dx &\leq CSe^{-\sqrt{\tau}\sigma \log n} \int_{-\infty}^{-\log n} e^{-a|x|^\kappa} dx \\
&= CSe^{-\sqrt{\tau}\sigma \log n} \int_{\log n}^\infty e^{-ax^\kappa} dx \\
&\leq CS \int_{\log n}^\infty e^{-ax^\kappa} dx \qquad \text{(A.9)} \\
&\leq \frac{CS}{a\kappa (\log n)^{\kappa-1}} \int_{\log n}^\infty a\kappa x^{\kappa-1} e^{-ax^\kappa} dx \\
&= \frac{CS}{a\kappa (\log n)^{\kappa-1}} e^{-a(\log n)^\kappa} = O_p(n^{-1})
\end{aligned}$$

where C is finite and $O_p(n^{-1})$ follows because, due to $\kappa > 1$, we have $a(\log n)^\kappa > \log n$ for large n. The second term on the right hand side of (A.8) is bounded from above by $CS\int_{\log n}^{\infty} e^{-ax^\kappa+\sqrt{\tau}\sigma x}dx$ uniformly over $z \in R$ for large n because of Assumption 3(i). This upper bound satisfies

$$CS\int_{\log n}^{\infty} e^{-ax^\kappa+\sqrt{\tau}\sigma x}dx = CS\int_{\log n}^{\infty} e^{-ax^\kappa/2}e^{(-ax^\kappa/2+\sqrt{\tau}\sigma x)}dx = CS\int_{\log n}^{\infty} e^{-ax^\kappa/2}dx * O_p(1),$$

where the second equality holds because, due to $\kappa > 1$, we have $-ax^\kappa/2 + \sqrt{\tau}\sigma x < 0$ for any $x > \log n$ for large n. By (A.9), $CS\int_{\log n}^{\infty} e^{-ax^\kappa/2}dx = O_p(n^{-1})$. Therefore, (A.8) is $O_p(n^{-1})$.

For the second equation in the Lemma, we have

$$\begin{aligned}
&\int_{-\infty}^{\infty} S\left(e^{\sqrt{\tau}\sigma z} - e^{\sqrt{\tau}\sigma x}\right)^+ f(x)dx - \int_{-\log n}^{\log n} S\left(e^{\sqrt{\tau}\sigma z} - e^{\sqrt{\tau}\sigma x}\right)^+ f(x)dx \\
= &\int_{-\infty}^{-\log n} S\left(e^{\sqrt{\tau}\sigma z} - e^{\sqrt{\tau}\sigma x}\right)^+ f(x)dx + \int_{\log n}^{\infty} S\left(e^{\sqrt{\tau}\sigma z} - e^{\sqrt{\tau}\sigma x}\right)^+ f(x)dx. \qquad \text{(A.10)}
\end{aligned}$$

The first term on the right hand side of (A.10) is bounded from above by $Se^{\sqrt{\tau}\sigma z}\int_{-\infty}^{-\log n} f(x)dx$ uniformly over $z \le B\log n$. For large n, this upper bound satisfies

$$\begin{aligned}
\sup_{z\le B\log n} Se^{\sqrt{\tau}\sigma z}\int_{-\infty}^{-\log n} f(x)dx &\le \sup_{z\le B\log n} SCe^{\sqrt{\tau}\sigma z}\int_{\log n}^{\infty} \exp\left(-ax^\kappa\right)dx \\
&\le SC\int_{\log n}^{\infty} \exp\left(-ax^\kappa + \sqrt{\tau}\sigma B\log n\right)dx \qquad \text{(A.11)} \\
&\le SC\int_{\log n}^{\infty} \exp\left(-ax^\kappa/2\right)dx = O_p(n^{-1}),
\end{aligned}$$

where the first inequality holds because of Assumption 3(i), the second holds because $z \le B\log n$, the third because $-ax^\kappa/2 + \sqrt{\tau}\sigma Bx < 0$ for sufficiently large n, and the equality holds because of (A.9). The second term on the right hand side of (A.10) is equal to $\int_{\log n}^{B\log n} S(e^{\sqrt{\tau}\sigma z} - e^{\sqrt{\tau}\sigma x})^+ f(x)dx$ because $z \le B\log n$. It satisfies

$$\begin{aligned}
\sup_{z\le B\log n}\int_{\log n}^{B\log n} S\left(e^{\sqrt{\tau}\sigma z} - e^{\sqrt{\tau}\sigma x}\right)^+ f(x)dx &\le Se^{\sqrt{\tau}\sigma B\log n}\int_{\log n}^{B\log n} f(x)dx \\
&\le Se^{\sqrt{\tau}\sigma B\log n}\int_{\log n}^{\infty} f(x)dx = O_p(n^{-1}),
\end{aligned}$$

where the equality holds because of (A.11). The Lemma follows from this result and (A.11). ■

Lemma A.2 *Under Assumptions 1-5, we have*

$$\begin{aligned}
\sup_{z\in R}\sum_{j=J+1}^{\infty}\left|\beta_j\int_{-c_1\log n}^{c_2\log n} S\left(e^{\sqrt{\tau}\sigma x} - e^{\sqrt{\tau}\sigma z}\right)^+ h_j(x)dx\right| &= O_p\left(n^{-1}\right), \\
\sup_{z\le B\log n}\sum_{j=J+1}^{\infty}\left|\beta_j\int_{-c_1\log n}^{c_2\log n} S\left(e^{\sqrt{\tau}\sigma z} - e^{\sqrt{\tau}\sigma x}\right)^+ h_j(x)dx\right| &= O_p\left(n^{-1}\right),
\end{aligned}$$

where S is the spot price of the underlying asset at time t, and c_1, c_2, and B can be any finite positive constants independent of n.

Proof of Lemma A.2. Without loss of generality, let $c_1 = c_2 = 1$. Let C be a finite constant that can be different between cases. The first equation in the Lemma satisfies, for sufficiently large n,

$$\begin{aligned}
\sum_{j=J+1}^{\infty} \left| \beta_j \int_{-\log n}^{\log n} S\left(e^{\sqrt{\tau}\sigma x} - e^{\sqrt{\tau}\sigma z}\right)^+ h_j(x) dx \right| &\leq \sum_{j=J+1}^{\infty} |\beta_j| \int_{-\log n}^{\log n} S\left(e^{\sqrt{\tau}\sigma x} - e^{\sqrt{\tau}\sigma z}\right)^+ |h_j(x)| \, dx \\
&\leq \sum_{j=J+1}^{\infty} |\beta_j| \int_{-\log n}^{\log n} S e^{\sqrt{\tau}\sigma x} |h_j(x)| \, dx \\
&\leq SC \sum_{j=J+1}^{\infty} |\beta_j| \left(\int_{-\log n}^{\log n} e^{\sqrt{\tau}\sigma x} dx \right) \\
&\leq \frac{SC}{\sqrt{\tau}\sigma} \sum_{j=J+1}^{\infty} e^{\sqrt{\tau}\sigma \log n - p j^r},
\end{aligned}$$

where the second inequality follows from the monotonicity of the function $\left(e^{\sqrt{\tau}\sigma x} - e^{\sqrt{\tau}\sigma z}\right)^+$ with respect to z, the third follows because $h_j(x)$ is bounded, and the last follows from Assumption 3 (ii). Because $J^r / \log n \to \infty$, we have $pj^r/2 > \sqrt{\tau}\sigma \log n$ for sufficiently large n, implying the last term in the preceding display satisfies

$$\begin{aligned}
SC \sum_{j=J+1}^{\infty} e^{\sqrt{\tau}\sigma \log n - p j^r} &\leq SC \sum_{j=J+1}^{\infty} \exp\left(-pj^r/2\right) \\
&= O_p(\exp\left(-pJ^r/4\right)) \\
&= O_p(\frac{1}{n} \exp(-\frac{p}{4} J^r + \log n)) \\
&= O_p(\frac{1}{n}),
\end{aligned} \tag{A.12}$$

where the first equality holds because $\sum_{j=J+1}^{\infty} \exp\left(-pj^r/4\right) < \infty$.

The second equation in the Lemma satisfies, for sufficiently large n,

$$\begin{aligned}
&\sum_{j=J+1}^{\infty} \left| \beta_j \int_{-\log n}^{\log n} S\left(e^{\sqrt{\tau}\sigma z} - e^{\sqrt{\tau}\sigma x}\right)^+ h_j(x) dx \right| \\
\leq &\sum_{j=J+1}^{\infty} |\beta_j| \int_{-\log n}^{\log n} S e^{\sqrt{\tau}\sigma z} |h_j(x)| \, dx \\
\leq &S \sum_{j=J+1}^{\infty} |\beta_j| \left| \int_{-\log n}^{\log n} e^{2\sqrt{\tau}\sigma z} dx \right|^{1/2} \left| \int_{-\log n}^{\log n} h_j(x)^2 dx \right|^{1/2}.
\end{aligned}$$

The last expression in the display satisfies $\sup_{z \leq B \log n} |\int_{-\log n}^{\log n} e^{2\sqrt{\tau}\sigma z} dx|^{1/2} \leq e^{\sqrt{\tau}\sigma B \log n} \log n$ and $\int_{-\log n}^{\log n} h_j(x)^2 dx < 1$ because $h_j(x)$ $(j = 0, 1, ...)$ are orthonormal. Therefore, this expression is bounded from above by $S \sum_{j=J+1}^{\infty} e^{\sqrt{\tau}\sigma B \log n - p j^r} \log n$. This upper bound is $O_p(n^{-1})$ by the same argument as in (A.12) and the assumption $J^r / \log n \to \infty$. ■

Lemma A.3 *Under Assumptions 1-5, for any $x \in \mathbb{R}$, we have*

$$h(x)' \left(\sum_{i=1}^{n} x_i x_i' + Q_\alpha \right)^{-1} \sum_{i=1}^{n} x_i \varepsilon_i = O_p\left(\sqrt{\frac{J}{\max(\alpha, n\lambda_{\min})}} \right),$$

where $h(x) = (h_0(x), ..., h_J(x))'$, Q_α is given by (21) or (22) and $\lambda_{\min}$ is the smallest eigenvalue of $n^{-1}\sum_{i=1}^{n} x_i x_i'$. In addition, the right hand side of the above expression is $O_p(J/\sqrt{\max(\alpha, n\lambda_{\min})})$ uniformly over $x \in \mathbb{R}$.

Proof of Lemma A.3. The right hand side expression in the Lemma has mean zero. Its variance, conditional on $(x_1, ..., x_n)$ for any $x \in \mathbb{R}$, is equal to

$$\begin{aligned} & h(x)' \left(\sum_{i=1}^{n} x_i x_i' + Q_\alpha \right)^{-1} \left(\sum_{i=1}^{n} \sigma_i^2 x_i x_i' \right) \left(\sum_{i=1}^{n} x_i x_i' + Q_\alpha \right)^{-1} h(x) \\ \leq & \max_{1 \leq j \leq n} \sigma_j^2 \left\{ h(x)' \left(\sum_{i=1}^{n} x_i x_i' + Q_\alpha \right)^{-1} \left(\sum_{i=1}^{n} x_i x_i' \right) \left(\sum_{i=1}^{n} x_i x_i' + Q_\alpha \right)^{-1} h(x) \right\}. \end{aligned}$$

Below, we show that this conditional variance is $O_p(J/\max(\alpha, n\lambda_{\min}))$. The Lemma follows from this result and the Chebyshev inequality.

Denote the eigenvalue decomposition of $n^{-1}\sum_{i=1}^{n} x_i x_i'$ by $V\Lambda V'$, where $\Lambda = diag(\lambda_1, ..., \lambda_{J+1})$ with $\lambda_1 \geq ... \geq \lambda_{J+1}$. If Q_α is given by (21), the expression in the curly brackets of the preceding display is equal to

$$\begin{aligned} & h(x)' V \left[(n\Lambda + \alpha)^{-1} (n\Lambda) (n\Lambda + \alpha)^{-1} \right] V' h(x) \qquad \text{(A.13)} \\ = & ||h(x)' V||^2 \left[\frac{h(x)' V}{||h(x)' V||} (n\Lambda + \alpha)^{-1} (n\Lambda) (n\Lambda + \alpha)^{-1} \frac{V' h(x)}{||h(x)' V||} \right] \end{aligned}$$

The expression in the brackets is bounded by the maximum eigenvalue of $(n\Lambda + \alpha)^{-1} (n\Lambda) (n\Lambda + \alpha)^{-1}$, which is less than or equal to

$$\max_{1 \leq j \leq J+1} \frac{n\lambda_j}{(n\lambda_j + \alpha)^2} \leq \max\left\{ \frac{1}{4\alpha}, \frac{1}{n\lambda_{\min}} \right\},$$

where the inequality holds because the function $x/(x+\alpha)^2$ for $x > 0$ is maximized at $x = \alpha$. From this result and $||h(x)' V|| = O_p(J^{1/2})$ uniformly over $x \in \mathbb{R}$, it follows that (A.13) is $O_p(J/\max(4\alpha, n\lambda_{\min})) = O_p(J/\max(\alpha, n\lambda_{\min}))$ for any $x \in \mathbb{R}$.

If Q_α is given by (22), then, instead of (A.13), we have

$$\begin{aligned} & h(x)' V \left[(n\Lambda + D_\alpha)^{-1} (n\Lambda) (n\Lambda + D_\alpha)^{-1} \right] V' h(x) \\ = & ||h(x)' V||^2 \left[\frac{h(x)' V}{||h(x)' V||} (n\Lambda + D_\alpha)^{-1} (n\Lambda) (n\Lambda + D_\alpha)^{-1} \frac{V' h(x)}{||h(x)' V||} \right], \end{aligned}$$

where $D_\alpha = \text{diag}\left[\max\{\alpha - n\lambda_1, 0\}, \max\{\alpha - n\lambda_2, 0\}, ..., \max\{\alpha - n\lambda_{J+1}, 0\}\right]$. The expression in the brackets is bounded from above by the maximum eigenvalue of $(n\Lambda + D_\alpha)^{-1} (n\Lambda) (n\Lambda + D_\alpha)^{-1}$, which is less than or equal to

$$\max_{1 \leq j \leq J+1} \frac{n\lambda_j}{(n\lambda_j + \max\{\alpha - n\lambda_j, 0\})^2} \qquad \text{(A.14)}$$

Let j^* denote the largest j that satisfies $n\lambda_j \geq \alpha$. Then, the preceding display can be rewritten as

$$\max\left\{ \begin{array}{cccccc} \frac{1}{n\lambda_1} & ... & \frac{1}{n\lambda_{j^*}} & \frac{n\lambda_{j^*+1}}{\alpha^2} & ... & \frac{n\lambda_{J+1}}{\alpha^2} \end{array} \right\}.$$

If $j > j^*$, then

$$\frac{n\lambda_j}{\alpha^2} \leq \frac{n\lambda_{j^*+1}}{\alpha^2} \leq \frac{1}{\alpha}.$$

If $j \leq j^*$, then

$$\frac{1}{n\lambda_j} \leq \frac{1}{\alpha}. \tag{A.15}$$

Therefore, (A.14) is bounded from above by $1/\max(\alpha, n\lambda_{\min})$. From this result and $||h(x)'V|| = O_p(J^{1/2})$ uniformly over $x \in \mathbb{R}$, it follows that (A.13) is $O_p(J/\max(\alpha, n\lambda_{\min}))$ for any $x \in \mathbb{R}$.

To establish the order of $h(x)'\left(\sum_{i=1}^n x_i x_i' + Q_\alpha\right)^{-1}\sum_{i=1}^n x_i\varepsilon_i$ uniformly over $x \in \mathbb{R}$, we apply the Cauchy–Schwarz inequality:

$$\begin{aligned} & \left\| h(x)'\left(\sum_{i=1}^n x_i x_i' + Q_\alpha\right)^{-1}\sum_{i=1}^n x_i\varepsilon_i \right\|^2 \\ \leq\ & \|h(x)\|^2 \left\| \left(\sum_{i=1}^n x_i x_i' + Q_\alpha\right)^{-1}\sum_{i=1}^n x_i\varepsilon_i \right\|^2 \\ =\ & \|h(x)\|^2 \operatorname{tr}\left\{ \left(\sum_{i=1}^n x_i x_i' + Q_\alpha\right)^{-1}\left(\sum_{i=1}^n x_i\varepsilon_i\right)\left(\sum_{i=1}^n x_i'\varepsilon_i\right)\left(\sum_{i=1}^n x_i x_i' + Q_\alpha\right)^{-1} \right\}. \end{aligned}$$

Because the Hermite functions are uniformly bounded on the real line, $\|h(x)\|^2 = O_p(J)$ uniformly over $x \in \mathbb{R}$. The second term on the right hand side can be studied using the Markov inequality. In particular, it is nonnegative and its mean is bounded by $(J+1)$ times the maximum eigenvalue of $(\sum_{i=1}^n x_i x_i' + Q_\alpha)^{-1}(\sum_{i=1}^n \sigma_i^2 x_i x_i')(\sum_{i=1}^n x_i x_i' + Q_\alpha)^{-1}$. This eigenvalue can be studied using the same argument as between (A.13) and (A.15), and it is $O_p(1/\max(\alpha, n\lambda_{\min}))$. These results imply that $h(x)'\left(\sum_{i=1}^n x_i x_i' + Q_\alpha\right)^{-1}\sum_{i=1}^n x_i\varepsilon_i$ is $O_p(J/\sqrt{\max(\alpha, n\lambda_{\min})})$ uniformly over $x \in \mathbb{R}$. ■

Proof of Theorem 1. Let $\hat{\beta}^*$ denote the estimate of β from STEP 2 of the proposed estimation procedure, but with $\eta = -\infty$. To shorten the notation, define $h(x) = (h_0(x), ..., h_J(x))'$ and let

$$g_i(x) = S\left(e^{\sqrt{\tau}\sigma x} - e^{\sqrt{\tau}\sigma z_i}\right)^+ \text{ if } n \leq n_c \text{ and } g_i(x) = S\left(e^{\sqrt{\tau}\sigma z_i} - e^{\sqrt{\tau}\sigma x}\right)^+ \text{ if n} > n_c.$$

Consider the following decomposition:

$$
\begin{aligned}
& h(x)'\hat{\beta}^* - f(x) \\
= & \left\{ h(x)'\hat{\beta}^* - h(x)'\beta + h(x)'\left(\sum_{i=1}^n x_i x_i' + Q_\alpha\right)^{-1} Q_\alpha \beta \right\} \\
& + [h(x)'\beta - f(x)] - h(x)'\left(\sum_{i=1}^n x_i x_i' + Q_\alpha\right)^{-1} Q_\alpha \beta \\
= & \; h(x)'\left(\sum_{i=1}^n x_i x_i' + Q_\alpha\right)^{-1} \sum_{i=1}^n x_i \varepsilon_i - h(x)'\left(\sum_{i=1}^n x_i x_i' + Q_\alpha\right)^{-1} Q_\alpha \beta \\
& + h(x)'\left(\sum_{i=1}^n x_i x_i' + Q_\alpha\right)^{-1} \sum_{i=1}^n x_i \left(\int_{-\infty}^{-\log n} g_i(x) f(x) dx + \int_{\log n}^{\infty} g_i(x) f(x) dx\right) \quad \text{(I)} \\
& + h(x)'\left(\sum_{i=1}^n x_i x_i' + Q_\alpha\right)^{-1} \sum_{i=1}^n x_i \left(\sum_{j=J+1}^{\infty} \beta_j \int_{-\log n}^{\log n} g_i(x) h_j(x) dx\right) \qquad \text{(II)} \\
& - \sum_{s=J+1}^{\infty} h_s(x)\beta_s \qquad \text{(III)}
\end{aligned}
$$

The terms (I)-(III) represent the biases induced by truncating the Hermite series, applying the regularization with Q_α, and replacing the integral $\int_{-\infty}^{\infty}$ by $\int_{-\log n}^{\log n}$.

We examine (I)-(III) separately. Let $\lambda_{\min}$ denote the smallest eigenvalue of $n^{-1}\sum_{i=1}^n x_i x_i'$. By the Cauchy-Schwarz inequality and Lemmas A.1 and A.3, (I) is bounded from above by

$$
\begin{aligned}
& \left\{ h(x)'\left(\sum_{i=1}^n x_i x_i' + Q_\alpha\right)^{-1} \left(\sum_{i=1}^n x_i x_i'\right) \left(\sum_{i=1}^n x_i x_i' + Q_\alpha\right)^{-1} h(x) \right\}^{1/2} \\
& \times \left\{ \sum_{i=1}^n \left(\int_{-\infty}^{-\log n} g_i(x) f(x) dx + \int_{\log n}^{\infty} g_i(x) f(x) dx\right)^2 \right\}^{1/2} \\
= & \; O_p\left(\sqrt{\frac{J}{\max(\alpha, n\lambda_{\min})}} \frac{1}{\sqrt{n}}\right) = o_p\left(\sqrt{\frac{J}{\max(\alpha, n\lambda_{\min})}}\right).
\end{aligned}
$$

Similarly, (II) is bounded from above by

$$
\begin{aligned}
& \left\{ h(x)'\left(\sum_{i=1}^n x_i x_i' + Q_\alpha\right)^{-1} \left(\sum_{i=1}^n x_i x_i'\right) \left(\sum_{i=1}^n x_i x_i' + Q_\alpha\right)^{-1} h(x) \right\}^{1/2} \\
& \times \left(\sum_{i=1}^n \left(\sum_{j=J+1}^{\infty} \beta_j \int_{-\log n}^{\log n} g_i(x) h_j(x) dx \right)^2 \right).
\end{aligned}
$$

By Lemmas A.2 and A.3, this is $o_p(\sqrt{J/\max(\alpha, n\lambda_{\min})})$. Term (III) satisfies, for large n,

$$\begin{aligned}\left|\sum_{s=J+1}^{\infty} h_s(x)\beta_s\right| &\leq C\sum_{s=J+1}^{\infty}|\beta_s| \leq C\sum_{j=J+1}^{\infty}\exp(-pj^r) \leq C\exp(-pJ^r/2)\left(\sum_{s=J+1}^{\infty}\exp(-pj^r/2)\right) \\ &= O\left(\exp(-pJ^r/2)\right) = O(1/n),\end{aligned}$$

where the first inequality holds because the Hermite functions are bounded for all j and $x \in \mathbb{R}$, the second holds because of Assumption 3(ii), the first equality holds because $\sum_{s=J+1}^{\infty}\exp(-pj^r/2) < \infty$, and the last inequality holds because $pj^r/2 > \log n$ for sufficiently large n. These results for (I)-(III) hold uniformly over $x \in \mathbb{R}$.

Combining these results, we have

$$\begin{aligned}& h(x)'\hat{\beta}^* - f(x) \\ = \; & h(x)'\left(\sum_{i=1}^{n} x_i x_i' + Q_\alpha\right)^{-1}\sum_{i=1}^{n} x_i\varepsilon_i - h(x)'\left(\sum_{i=1}^{n} x_i x_i' + Q_\alpha\right)^{-1} Q_\alpha\beta \\ & + o_p\left(\sqrt{\frac{J}{\max(\alpha, n\lambda_{\min})}}\right)\end{aligned} \tag{A.16}$$

uniformly over $x \in \mathbb{R}$. We study the two leading terms on the right hand side separately. By Lemma A.3, the first term satisfies

$$h(x)'\left(\sum_{i=1}^{n} x_i x_i' + Q_\alpha\right)^{-1}\sum_{i=1}^{n} x_i\varepsilon_i = O_p\left(\sqrt{\frac{J}{\max(\alpha, n\lambda_{\min})}}\right) \tag{A.17}$$

for any $x \in \mathbb{R}$, and it is $O_p\left(J\sqrt{\max(\alpha, n\lambda_{\min})}\right)$ uniformly over $x \in \mathbb{R}$. The second term satisfies

$$\begin{aligned}& \left\|h(x)'\left(\sum_{i=1}^{n} x_i x_i' + Q_\alpha\right)^{-1} Q_\alpha\beta\right\| \\ = \; & \left\|h(x)'\left(\sum_{i=1}^{n} x_i x_i' + Q_\alpha\right)^{-1} Q_\alpha\left(n^{-1}\sum_{i=1}^{n} x_i x_i'\right)^{1/4}\left(n^{-1}\sum_{i=1}^{n} x_i x_i'\right)^{-1/4}\beta\right\| \\ \leq \; & \left\|h(x)'\left(\sum_{i=1}^{n} x_i x_i' + Q_\alpha\right)^{-1} Q_\alpha\left(n^{-1}\sum_{i=1}^{n} x_i x_i'\right)^{1/4}\right\|\left\|\left(n^{-1}\sum_{i=1}^{n} x_i x_i'\right)^{-1/4}\beta\right\| \\ = \; & \left\{h(x)'V(n\Lambda + D_\alpha)^{-1} D_\alpha\Lambda^{1/2}D_\alpha(n\Lambda + D_\alpha)^{-1}V'h(x)\right\}^{1/2}\left\|\left(n^{-1}\sum_{i=1}^{n} x_i x_i'\right)^{-1/4}\beta\right\|,\end{aligned} \tag{A.18}$$

where $D_\alpha = \alpha I$ if Q_α is given by (21) and is equal to (23) if Q_α is given by (22). In both cases, the expression in the above curly brackets can be rewritten as

$$\frac{1}{n}\left\{h(x)'V\left(\Lambda + \frac{1}{n}D_\alpha\right)^{-1} D_\alpha\Lambda^{1/2}D_\alpha\left(\Lambda + \frac{1}{n}D_\alpha\right)^{-1}V'h(x)\right\}^{1/2},$$

which is equal to

$$\begin{aligned} & \frac{1}{n}||h(x)'V|| \left\{ \frac{h(x)'V}{||h(x)'V||} \left(\Lambda + \frac{1}{n}D_\alpha\right)^{-1} D_\alpha \Lambda^{1/2} D_\alpha \left(\Lambda + \frac{1}{n}D_\alpha\right)^{-1} \frac{V'h(x)}{||h(x)'V||} \right\}^{1/2} \\ \leq \; & \frac{1}{n}||h(x)'V|| \left\| \left(\Lambda + \frac{1}{n}D_\alpha\right)^{-1} D_\alpha \Lambda^{1/2} D_\alpha \left(\Lambda + \frac{1}{n}D_\alpha\right)^{-1} \right\|^{1/2}. \end{aligned} \tag{A.19}$$

If $D_\alpha = \alpha I$, the second norm on the right hand side of (A.19) is bounded from above by

$$\max_{1\leq j\leq J+1} \frac{\lambda_j^{1/2}\alpha^2}{\left(\lambda_j + \frac{1}{n}\alpha\right)^2}.$$

To determine this value, we consider the following function of $x \geq 0$:

$$G(x) = \frac{x\alpha^4}{\left(x + \frac{1}{n}\alpha\right)^4}.$$

Its first-order condition respect to x is given by

$$\alpha^4 \frac{\left(x + \frac{1}{n}\alpha\right)^4 - 4x\left(x + \frac{1}{n}\alpha\right)^3}{\left(x + \frac{1}{n}\alpha\right)^8} = \alpha^4 \left(x + \frac{1}{n}\alpha\right)^3 \frac{\frac{\alpha}{n} - 3x}{\left(x + \frac{1}{n}\alpha\right)^5} = 0.$$

Therefore, $G(x)$ achieves its global maximum at $x = \alpha/(3n)$, which implies

$$\max_{1\leq j\leq J+1} \frac{\lambda_j^{1/2}\alpha^2}{\left(\lambda_j + \frac{1}{n}\alpha\right)^2} \leq \frac{\alpha^2\left(\frac{\alpha}{3n}\right)^{1/2}}{\left(\frac{\alpha}{3n} + \frac{1}{n}\alpha\right)^2} = \frac{\alpha^2\left(\frac{\alpha}{3n}\right)^{1/2}}{\left(\frac{4\alpha}{3n}\right)^2} \leq \alpha^2\left(\frac{n}{\alpha}\right)^{3/2}.$$

As a result, (A.19) is bounded from above by

$$\frac{1}{n}(J+1)^{1/2}\left[\alpha^2\left(\frac{n}{\alpha}\right)^{3/2}\right]^{1/2} = \left[\left(\frac{\alpha}{n}\right)^{1/2}(J+1)\right]^{1/2} = \left(\frac{\alpha}{n}\right)^{1/4}(J+1)^{1/2}$$

uniformly over $x \in \mathbb{R}$. If D_α is equal to (23), the second norm on the right hand side of (A.19) is bounded from above by

$$\max_{1\leq j\leq J+1} \frac{n^2\lambda_j^{1/2}\max\left\{\frac{\alpha}{n} - \lambda_j, 0\right\}^2}{\left(\lambda_j + \max\left\{\frac{\alpha}{n} - \lambda_j, 0\right\}\right)^2}.$$

Let j^* denote largest j that satisfies $n\lambda_j \geq \alpha$. Then, this maximum is equal to

$$\max\left\{ \begin{array}{ccccccc} 0 & \ldots & 0 & \frac{n^2\lambda_{j^*+1}^{1/2}\left(\frac{\alpha}{n} - \lambda_{j^*+1}\right)^2}{\left(\frac{\alpha}{n}\right)^2} & \ldots & \frac{n^2\lambda_{J+1}^{1/2}\left(\frac{\alpha}{n} - \lambda_{J+1}\right)^2}{\left(\frac{\alpha}{n}\right)^2} \end{array} \right\}.$$

To determine this value, we consider the following function of $0 \leq x \leq \alpha/n$:

$$G(x) = x\left(\frac{\alpha}{n} - x\right)^4.$$

The first-order condition with respect to x implies

$$\left[\left(\frac{\alpha}{n}-x\right)^4-4x\left(\frac{\alpha}{n}-x\right)^3\right]=\left(\frac{\alpha}{n}-x\right)^3(\frac{\alpha}{n}-5x)=0.$$

This shows that $G(x)$ is an increasing function over $[0,\alpha/(5n)]$ and a decreasing function over $[\alpha/(5n),\alpha/n]$. Setting $\lambda_j=\alpha/(5n)$, we have

$$\frac{n^2\lambda_j^{1/2}\left(\frac{\alpha}{n}-\lambda_j\right)^2}{\left(\frac{\alpha}{n}\right)^2}=\frac{n^2\left(\frac{\alpha}{5n}\right)^{1/2}\left(\frac{4\alpha}{5n}\right)^2}{\left(\frac{\alpha}{n}\right)^2}\leq n^{3/2}\alpha^{1/2}.$$

This upper bound is the same as in the $D_\alpha=\alpha I$ case. Therefore, (A.19) is always bounded by $(\alpha/n)^{1/4}(J+1)^{1/2}$ uniformly over $x\in\mathbb{R}$. Consequently, the expression (A.18) is

$$O_p\left(\left(\frac{\alpha}{n}\right)^{1/4}J^{1/2+\nu}\right) \tag{A.20}$$

uniformly over $x\in\mathbb{R}$.

From (A.16), (A.17), and (A.20), it follows that

$$h(x)'\hat{\beta}^*-f(x)=O_p\left(\sqrt{\frac{J}{\max(\alpha,n\lambda_{\min})}}+\left(\frac{\alpha}{n}\right)^{1/4}J^{1/2+\nu}\right).$$

for any $x\in\mathbb{R}$. Furthermore, the left hand side is $O_p(J/\max(\alpha,n\lambda_{\min})+(\alpha/n)^{1/4}J^{1/2+\nu})$ uniformly over $x\in\mathbb{R}$, which implies $\Pr(\inf_{x\in\mathbb{R}}h(x)'\hat{\beta}^*\geq\eta)\to 1$ for any $\eta<0$ independent of n. Therefore, the constraint $\inf_{x\in\mathbb{R}}h(x)'\hat{\beta}^*\geq\eta$ is not binding asymptotically, and

$$\begin{aligned}h(x)'\hat{\beta}-f(x) &= O_p\left(\sqrt{\frac{J}{\max(\alpha,n\lambda_{\min})}}+\left(\frac{\alpha}{n}\right)^{1/4}J^{1/2+\nu}\right)\\ &= O_p\left(\sqrt{\frac{J}{\alpha}}+\left(\frac{\alpha}{n}\right)^{1/4}J^{1/2+\nu}\right)\end{aligned}$$

for any $x\in\mathbb{R}$. ■

Online Appendix B

Discussion of Assumption 2. We evaluate this assumption's empirical relevance using some inequalities derived by Bassett (1997). The main idea is that the approximate slope of the option curve implies bounds on the underlying cumulative distribution $F(z)$. Define

$$\hat{F}(z_{(i)})=e^{r\tau}\left[\frac{C(z_{(i+1)})-C(z_{(i)})}{K_{(i+1)}-K_{(i)}}\right]+1=e^{r\tau}\left[\frac{P(z_{(i+1)})-P(z_{(i)})}{K_{(i+1)}-K_{(i)}}\right] \tag{A.21}$$

for $i\in\{1,...,n-1\}$, where $C(z_{(i)})$ satisfies (5), $z_{(i)}$ and $K_{(i)}$ satisfy (4), and the second equality can be verified by applying (5) and (6). Let $\hat{F}(z_{(0)})=0$ and $\hat{F}(z_{(n)})=1$. Then, by Bassett (1997), the following inequality holds for $i=1,...,n$: $\hat{F}(z_{(i-1)})\leq F(z_{(i)})\leq\hat{F}(z_{(i)})$. This inequality implies

$$F(z_{(1)})\leq\hat{F}(z_{(1)}),\qquad F(z_{(n)})\geq\hat{F}(z_{(n-1)}), \tag{A.22}$$

and

$$0 \leq F(z_{(i)}) - F(z_{(i-1)}) \leq \hat{F}(z_{(i)}) - \hat{F}(z_{(i-2)}) \tag{A.23}$$

for $i = 2, ..., n-1$. By (A.22), Assumption 2(ii) is empirically relevant if $\hat{F}(z_{(1)}) \approx 0$ and $\hat{F}(z_{(n-1)}) \approx 1$. By (A.23), Assumption 2(i) is empirically relevant if $\hat{F}(z_{(i)}) - \hat{F}(z_{(i-2)}) \approx 0$ for $i = 2, ..., n$. In the actual data, the option prices $C(z_{(i)})$ and $P(z_{(i)})$ contain noises due to market frictions, which implies that $\hat{F}(z_{(i)})$ $(i = 1, ..., n-1)$ can only be computed approximately. Nevertheless, as seen below, their values are still informative about the empirical relevance of Assumption 2.

We now apply (A.22) and (A.23) to evaluate Assumption 2 for the following five options markets: S&P 500 (SPX), Russell 2000 (RUT), Dow Jones Industrial Average (DJX), Nasdaq-100 (NDX), and Cboe Volatility (VIX). The transaction date is set as March 1, 2016, and the expiration date is the end of the first quarter. The options with zero open interest are excluded to avoid stale information. The results, summarized in Table S1, show that $\hat{F}(z_{(1)})$ and $\hat{F}(z_{(n-1)})$ are close to 0 and 1, respectively. The largest difference is observed in the VIX case, where $\hat{F}(z_{(1)}) = 0.1$. The values of $\hat{F}(z_{(i)}) - \hat{F}(z_{(i-2)})$ tend to be small, although some sizeable values are observed in the VIX case, where the 75th and 90th percentiles of $\hat{F}(z_{(i)}) - \hat{F}(z_{(i-2)})$ are equal to 0.275 and 0.375, respectively. Overall, Assumption 2 is a mild assumption for the first four options markets, while in the VIX case, it should be viewed as a strong assumption, where we face an excessively coarse strike grid in some circumstances. Finally, in all five cases, the numbers of call options and put options are comparable, which implies that including put options is important for improving the estimation efficiency. Although the above analysis is for a particular trading day, it is straightforward to download additional data (e.g., from Yahoo Finance) to verify that the reported features are generic. Similar analyses can be applied to evaluate this assumption for other options markets when applying the proposed method in practice.

Table S1: Summary Statistics for Some Index Options

Ticker	Total	Call	Put	$\hat{F}(z_{(1)})$	$\hat{F}(z_{(n-1)})$	Quantile of $\{\hat{F}(z_{(i)}) - \hat{F}(z_{i-2})\}_{2\leq i\leq n}$				
						$10th$	$25th$	$50th$	$75th$	$90th$
SPX	283	122	161	0	1	-0.010	0.000	0.005	0.030	0.060
VIX	23	15	8	0.100	0.980	0.025	0.050	0.169	0.275	0.375
RUT	176	75	101	0	1	-0.005	0.000	0.010	0.040	0.077
DJX	105	50	55	0	1	0.000	0.000	0.010	0.060	0.148
NDX	72	21	51	0.001	0.956	0.000	0.002	0.019	0.070	0.104

Note. The values are computed using the closing prices on March 1, 2016, obtained from OptionMetrics. The contracts with zero open interest are excluded. The expiration date is the end of the first quarter: $\tau = 29$ for VIX, $\tau = 30$ for SPX, RUT, and NDX, and $\tau = 17$ for DJX. The annual risk-free set is set to 3.3%.

Table S2: Mean Integrated Squared Errors

Estimator	$n_c = n_p = 20$			
	J	$Q_\alpha = 0$	$Q_\alpha = \alpha I$	$Q_\alpha = VD_\alpha V'$
Sieve	3	0.0404	0.0405	0.0403
	4	0.0536	0.0427	0.0499
	5	0.0724	0.0551	0.0627
	6	0.0961	0.0696	0.0637
Local linear			0.1103	
PCA			0.2940	
Parametric			0.0562	

Note. For the sieve estimator, the three columns display the MISEs over 5000 replication with no regularization ($Q_\alpha = 0$), the standard Tikhonov regularization ($Q_\alpha = \alpha I$), and the modified Tikhonov of Fuhry and Reichel ($Q_\alpha = VD_\alpha V'$), respectively. Setting J =ceiling$(2(n/\log n)^{1/5})$ gives J=4. The parameter α is determined by 10-fold cross validation. The last three rows of the table correspond to the constrained local linear estimator of Ait-Sahalia and Duarte (2003), the positive convolution approximation estimator of Bondarenko (2003), and the four-parameter Gram-Charlier series estimator of Jarrow and Rudd (1982) and Longstaff (1995).

Table S3: Predictive Mean Regressions for S&P500 Returns
Using a Quantile as the Predictor (with a higher cutoff)

Quantile	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
01/18/2007 - 04/27/2016									
Estimate	-0.12	-0.36	-0.73	-1.83	0.77	0.61	0.44	0.31	0.22
s.d.	0.14	0.34	0.62	1.36	2.00	0.85	0.50	0.35	0.24
p-value	0.40	0.28	0.23	0.18	0.70	0.47	0.38	0.37	0.36
R-square	0.01	0.02	0.02	0.02	0.00	0.00	0.01	0.01	0.01
06/18/2009 - 04/27/2016									
Estimate	-0.37	-1.32	-2.51	-5.32	3.46	2.14	1.46	1.03	0.72
s.d.	0.09	0.31	0.55	1.45	1.20	0.58	0.38	0.26	0.18
p-value	**1e-4**	**3e-5**	**1e-5**	**3e-4**	**4e-3**	**3e-4**	**1e-4**	**1e-4**	**1e-4**
R-square	0.07	0.10	0.10	0.06	0.04	0.07	0.08	0.09	0.09

Note. Each column represents a least square regression. For example, the first column displays the results from a least square regression of the two-month-ahead realized return on the 10th percentile of the state price density. The standard errors allow for heteroscedasticity and autocorrelation. The estimates that are significant at the 10% level are in bold.

Table S4: Predictive Quantile Regressions for S&P500 Returns
(with a higher cutoff)

Quantile	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
01/18/2007 - 04/27/2016									
Estimate	0.25	0.26	-0.52	-1.69	2.18	2.21	1.62	1.01	0.77
s.d.	0.07	0.62	1.09	1.64	1.81	0.52	0.28	0.16	0.21
p-value	**2e-4**	0.67	0.63	0.30	0.22	**2e-5**	**7e-9**	**2e-10**	**2e-4**
06/18/2009 - 04/27/2016									
Estimate	-0.31	-0.64	-1.81	-5.05	3.69	2.62	2.14	1.65	1.28
s.d.	0.29	0.61	1.06	2.03	1.91	0.68	0.58	0.34	0.26
p-value	0.28	0.30	**0.09**	**0.01**	**0.06**	**1e-4**	**2e-4**	**1e-6**	**1e-6**

Note. Each column represents a quantile regression. For example, the first column displays the results from a 10th percentile quantile regression of the two-month-ahead realized return on the 10th percentile of the state price density. The standard errors allow for heteroscedasticity and autocorrelation. The estimates that are significant at the 10% level are in bold.

Table S5: Predictive Mean Regressions for S&P500 Returns
Using a Quantile as the Predictor (60 days ahead)

Quantile	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
01/18/2007 - 04/27/2016									
Estimate	-0.08	-0.13	-0.26	-0.64	-1.12	0.02	0.06	0.07	0.06
s.d.	0.18	0.44	0.81	1.66	4.08	1.53	0.84	0.56	0.39
p-value	0.67	0.77	0.75	0.69	0.78	0.98	0.95	0.89	0.88
R-square	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
06/18/2009 - 04/27/2016									
Estimate	-0.57	-1.66	-3.12	-6.57	6.04	3.69	2.43	1.68	1.18
s.d.	0.21	0.45	0.79	1.56	1.94	1.17	0.73	0.50	0.35
p-value	**7e-3**	**4e-4**	**1e-4**	**5e-5**	**2e-3**	**2e-3**	**1e-3**	**1e-3**	**1e-3**
R-square	0.12	0.18	0.17	0.12	0.06	0.13	0.16	0.17	0.17

Note. Each column represents a least square regression. For example, the first column displays the results from a least square regression of the two-month-ahead realized return on the 10th percentile of the state price density. The standard errors allow for heteroscedasticity and autocorrelation. The estimates that are significant at the 10% level are in bold.

Table S6: Predictive Quantile Regressions for S&P500 Returns
(60 days ahead)

Quantile	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
01/18/2007 - 04/27/2016									
Estimate	0.03	0.21	-0.06	0.26	0.82	2.55	1.61	1.17	1.01
s.d.	0.22	0.86	1.15	1.56	3.22	1.72	0.77	0.38	0.17
p-value	0.89	0.96	0.96	0.87	0.80	0.14	**0.04**	**2e-3**	**1e-9**
06/18/2009 - 04/27/2016									
Estimate	-0.22	-1.29	-1.85	-3.88	5.34	3.39	3.18	2.35	1.31
s.d.	0.23	1.04	1.36	2.74	1.91	1.32	0.87	0.43	0.20
p-value	0.34	0.22	0.17	0.12	**5e-3**	**1e-2**	**3e-4**	**4e-8**	**7e-11**

Note. Each column represents a quantile regression. For example, the first column displays the results from a 10th percentile quantile regression of the two-month-ahead realized return on the 10th percentile of the state price density. The standard errors allow for heteroscedasticity and autocorrelation. The estimates that are significant at the 10% level are in bold.

Table S7: Predictive Mean Regressions for VIX Using a Quantile as the Predictor

Quantile	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
					01/18/2007 - 04/27/2016				
Estimate	1.38	1.25	1.16	1.08	1.01	0.94	0.87	0.79	0.66
s.d.	0.23	0.20	0.18	0.17	0.15	0.14	0.12	0.11	0.08
p-value	**4e-8**	**1e-8**	**4e-9**	**2e-9**	**9e-10**	**3e-10**	**1e-10**	**4e-11**	**4e-12**
R-square	48.0	50.2	51.5	52.1	52.6	53.0	53.1	53.2	52.2
					06/18/2009 - 04/27/2016				
Estimate	1.01	0.94	0.88	0.83	0.78	0.73	0.68	0.61	0.51
s.d.	0.18	0.16	0.14	0.13	0.11	0.10	0.09	0.08	0.07
p-value	**4e-7**	**5e-8**	**1e-8**	**4e-9**	**1e-9**	**4e-10**	**1e-10**	**3e-11**	**4e-11**
R-square	38.8	41.7	43.2	44.0	44.5	45.2	45.7	46.6	46.9

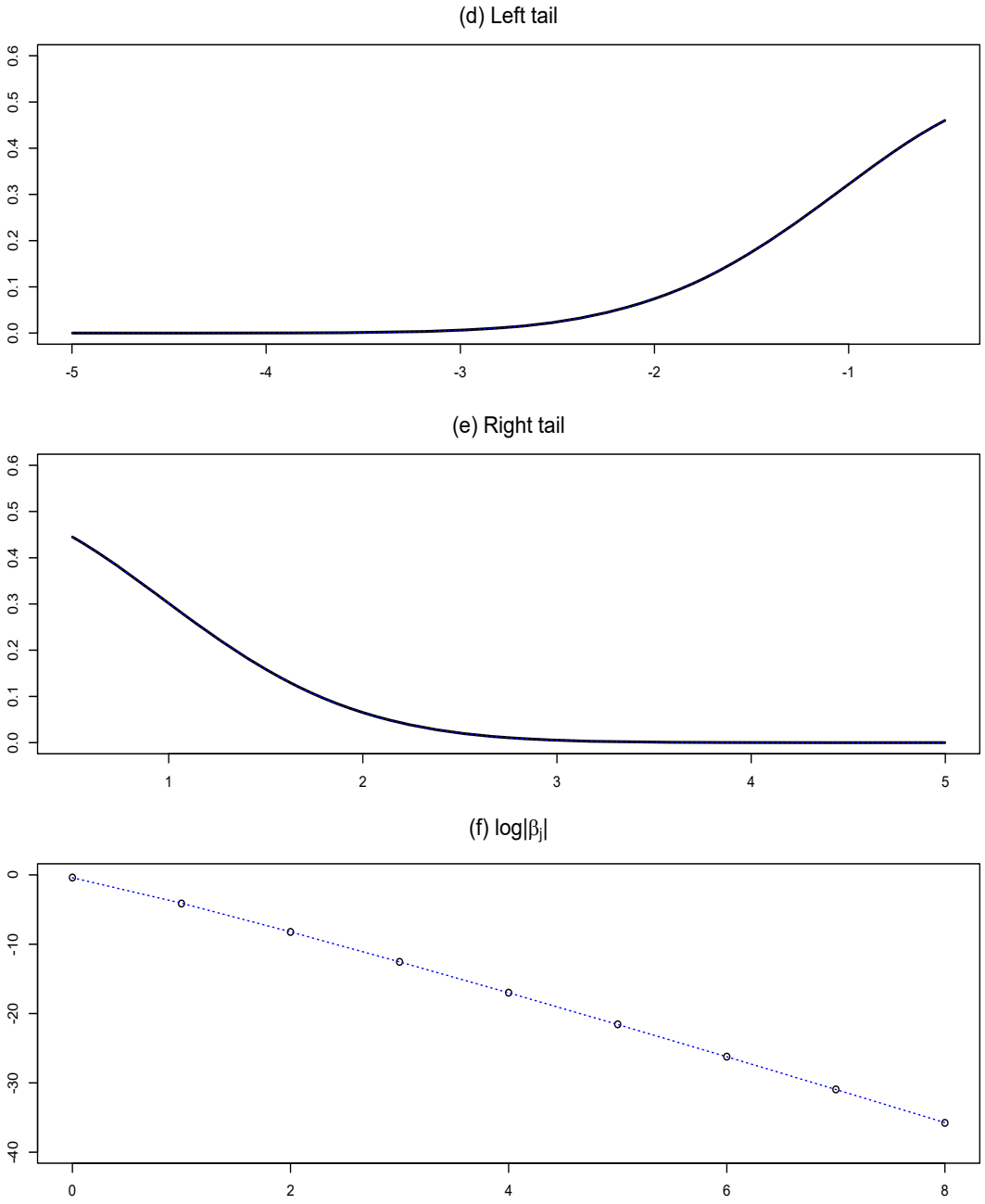
Note. Each column represents a least square regression. For example, the first column displays the results from a least square regression of the one-month-ahead realized VIX level on the 10th percentile of the state price density. The standard errors allow for heteroscedasticity and autocorrelation. The estimates that are significant at the 10% level are in bold.

Table S8: Predictive Quantile Regressions for VIX

Quantile	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
					01/18/2007 - 04/27/2016				
Estimate	0.89	0.82	0.85	0.83	0.83	0.82	0.89	0.80	0.84
s.d.	0.05	0.05	0.08	0.12	0.15	0.15	0.10	0.26	0.46
p-value	**1e-16**	**1e-16**	**1e-16**	**7e-13**	**5e-8**	**8e-8**	**1e-16**	**2e-3**	**0.07**
					06/18/2009 - 04/27/2016				
Estimate	0.84	0.78	0.77	0.71	0.73	0.69	0.73	0.75	0.67
s.d.	0.20	0.07	0.06	0.09	0.12	0.18	0.21	0.24	0.15
p-value	**2e-5**	**1e-16**	**1e-16**	**1e-15**	**1e-9**	**1e-4**	**6e-4**	**2e-3**	**1e-5**

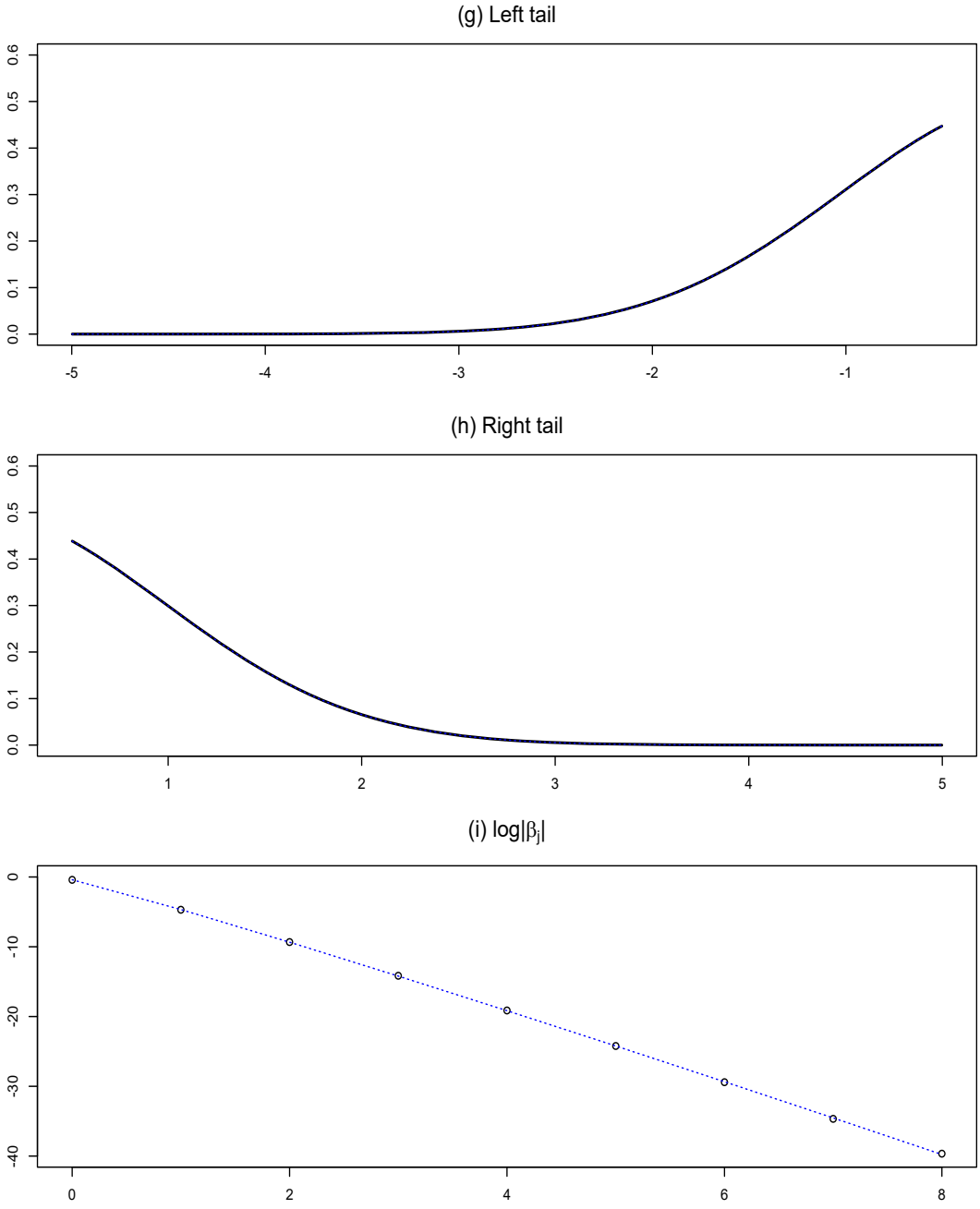
Note. Each column represents a quantile regression. For example, the first column displays the results from a 10th percentile quantile regression of the one-month-ahead realized VIX level on the 10th percentile of the state price density. The standard errors allow for heteroscedasticity and autocorrelation. The estimates that are significant at the 10% level are in bold.

Figure S1: Density tails and Hermite coefficients implied by the Black-Scholes model (the recovery period)



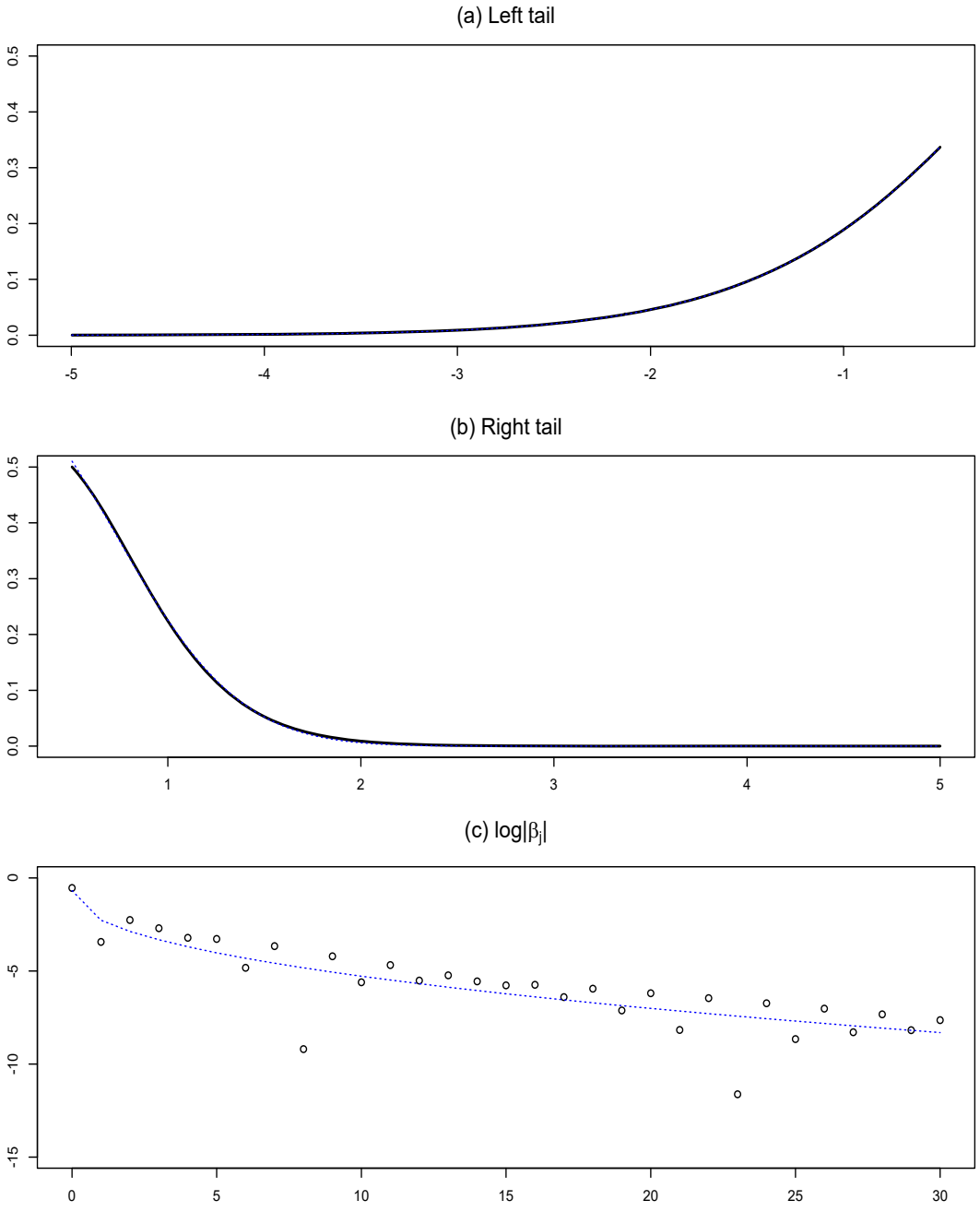
Note. In (d) and (e), the solid line represents the true values, and the dotted line represents the fitted values from the regressions. The tails look similar to those in Figure 1 because the density is for the transformed variable, as explained in the paper. In (f), the circles denote the actual Hermite coefficients, and the dotted line the fitted regression line.

Figure S2: Density tails and Hermite coefficients implied by the Black-Scholes model (the expansion period)



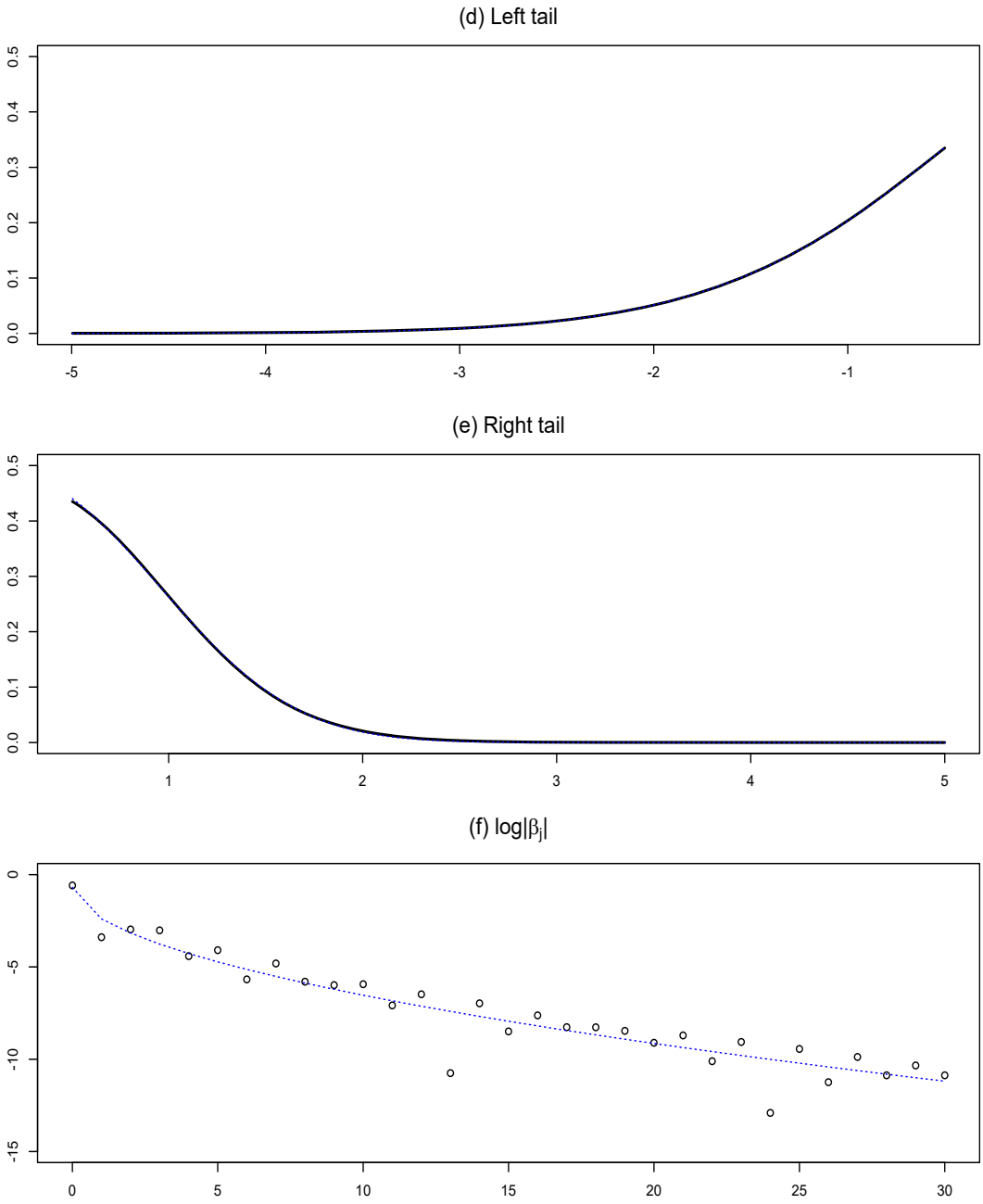
Note. In (g) and (h), the solid line represents the true values, and the dotted line represents the fitted values from the regressions. The tails look similar to those in Figure 1 because the density is for the transformed variable, as explained in the paper. In (i), the circles denote the actual Hermite coefficients, and the dotted line the fitted regression line.

Figure S3: Density tails and Hermite coefficients implied by the SV model (the crisis period)



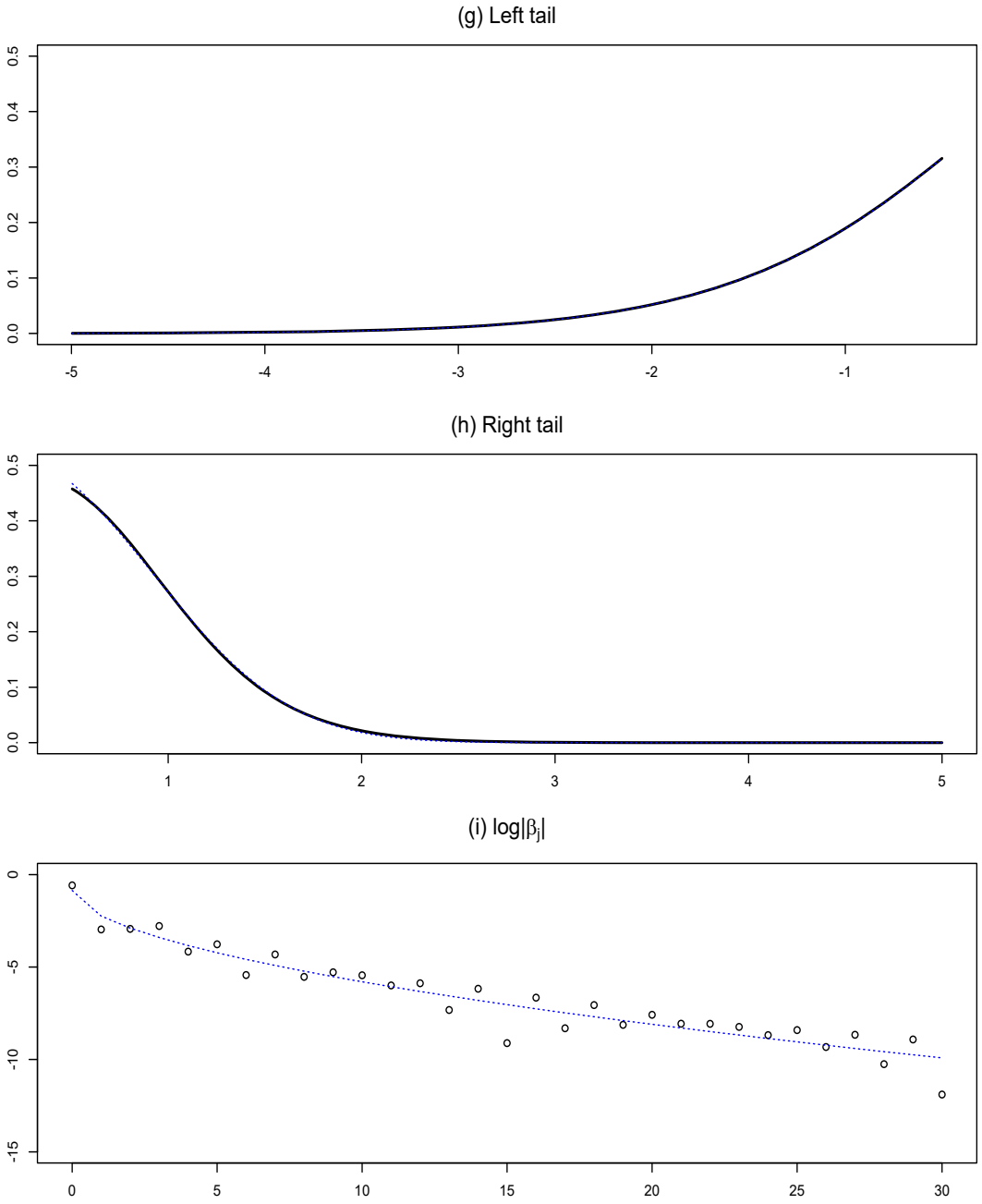
Note. In (a) and (b), the solid line represents the true values, and the dotted line represents the fitted values from the regressions. In (c), the circles denote the actual Hermite coefficients, and the dotted line the fitted regression line.

Figure S4: Density tails and Hermite coefficients implied by the SV model (the recovery period)



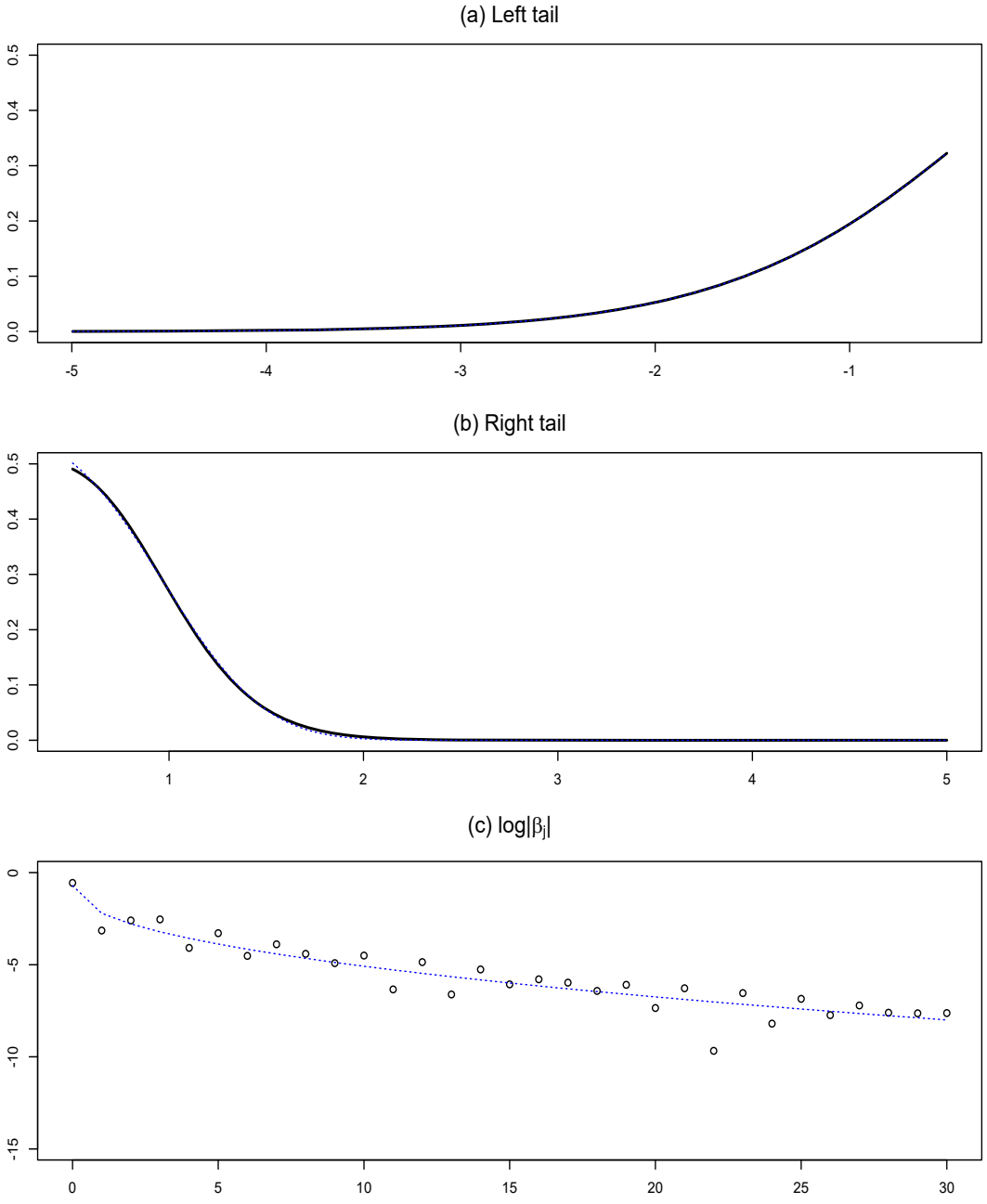
Note. In (d) and (e), the solid line represents the true values, and the dotted line represents the fitted values from the regressions. The tails look similar to those in Figure S3 because the density is for the transformed variable, as explained in the paper. In (f), the circles denote the actual Hermite coefficients, and the dotted line the fitted regression line.

Figure S5: Density tails and Hermite coefficients implied by the SV model (the expansion period)



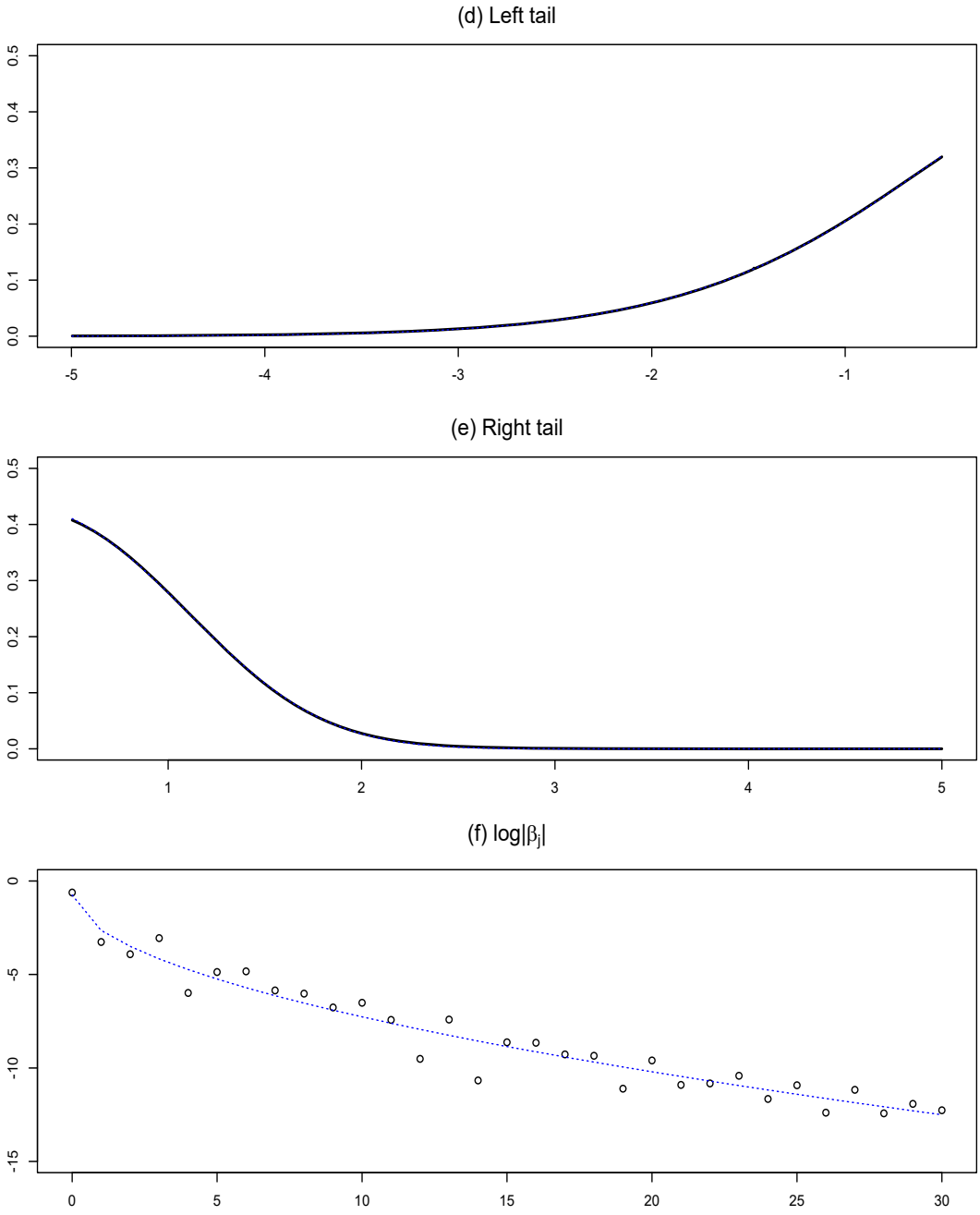
Note. In (g) and (h), the solid line represents the true values, and the dotted line represents the fitted values from the regressions. The tails look similar to those in Figure S3 because the density is for the transformed variable, as explained in the paper. In (i), the circles denote the actual Hermite coefficients, and the dotted line the fitted regression line.

Figure S6: Density tails and Hermite coefficients implied by the SVJ model (the crisis period)



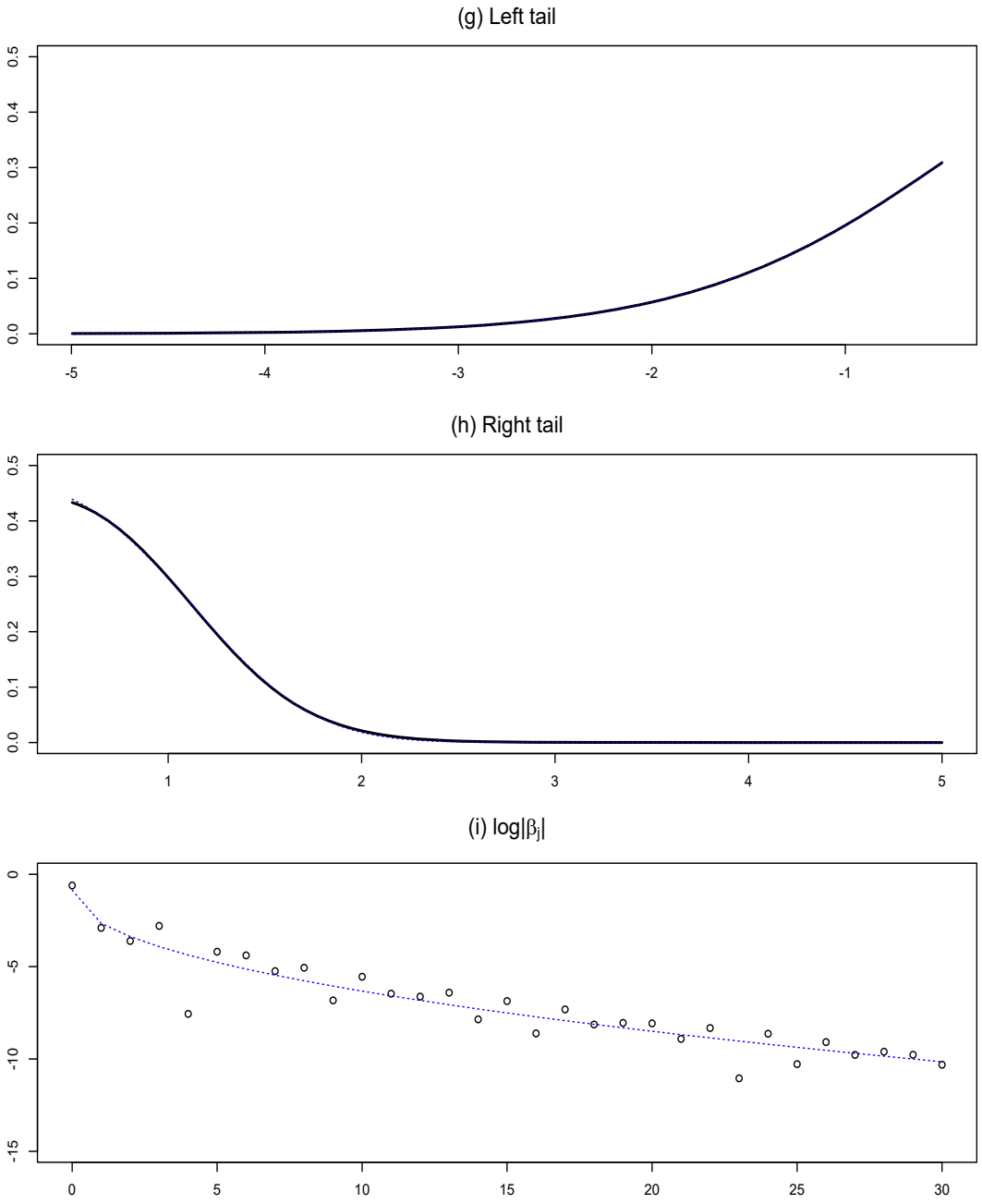
Note. In (a) and (b), the solid line represents the true values, and the dotted line represents the fitted values from the regressions. In (c), the circles denote the actual Hermite coefficients, and the dotted line the fitted regression line.

Figure S7: Density tails and Hermite coefficients implied by the SVJ model (the recovery period)



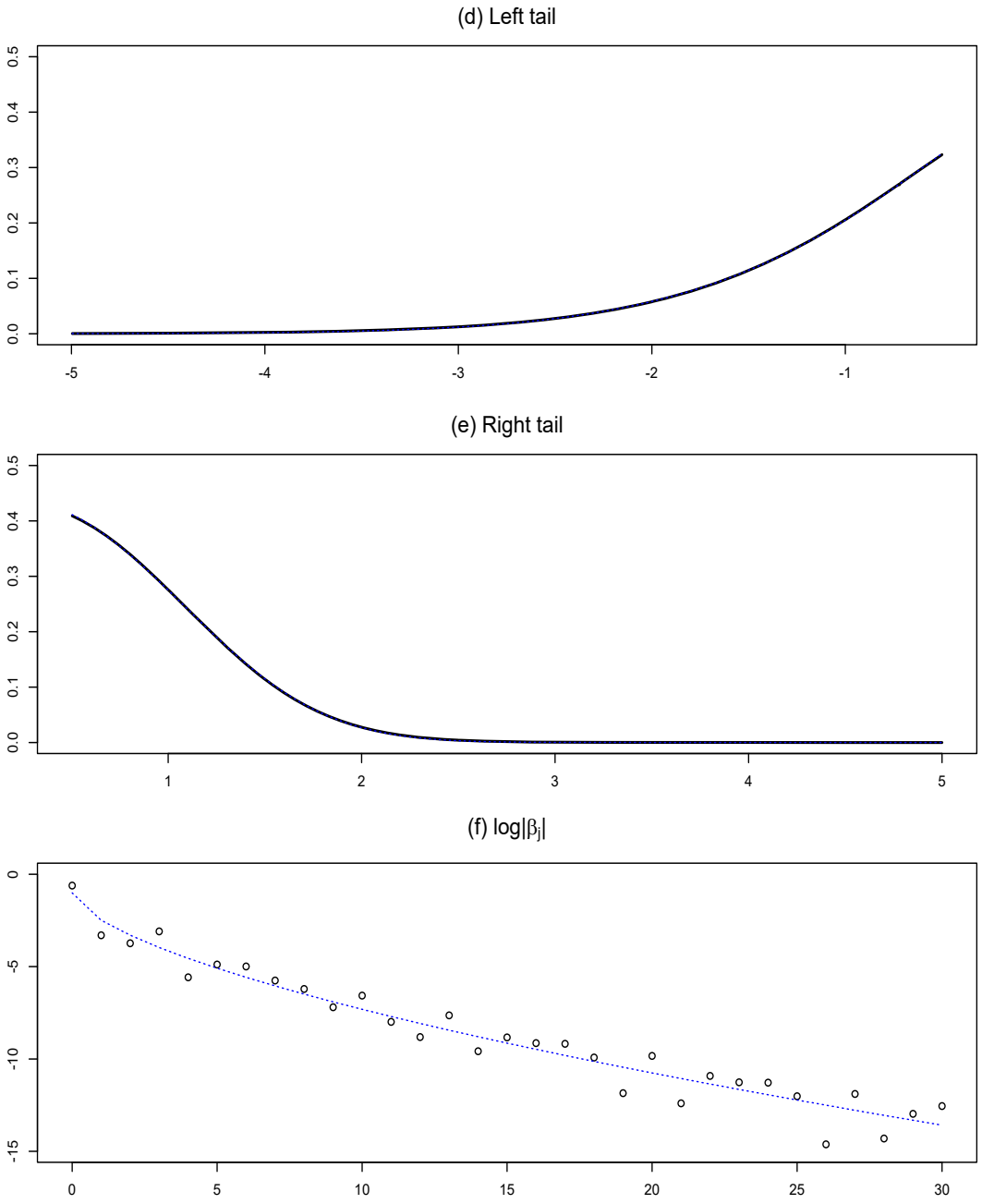
Note. In (d) and (e), the solid line represents the true values, and the dotted line represents the fitted values from the regressions. The tails look similar to those in Figure S3 because the density is for the transformed variable, as explained in the paper. In (f), the circles denote the actual Hermite coefficients, and the dotted line the fitted regression line.

Figure S8: Density tails and Hermite coefficients implied by the SVJ model (the expansion period)



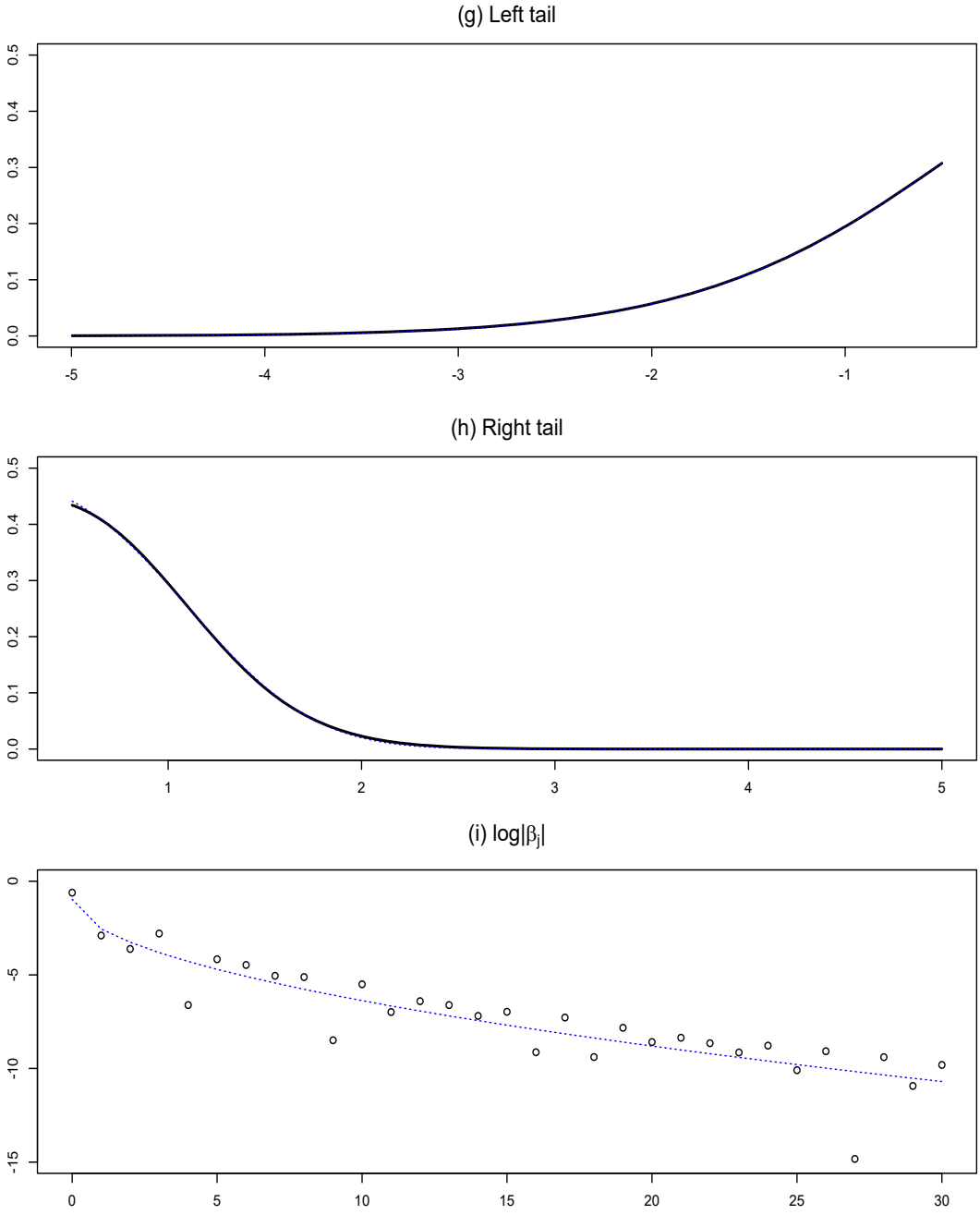
Note. In (g) and (h), the solid line represents the true values, and the dotted line represents the fitted values from the regressions. The tails look similar to those in Figure S3 because the density is for the transformed variable, as explained in the paper. In (i), the circles denote the actual Hermite coefficients, and the dotted line the fitted regression line.

Figure S9: Density tails and Hermite coefficients implied by the SVCJ model (the recovery period)



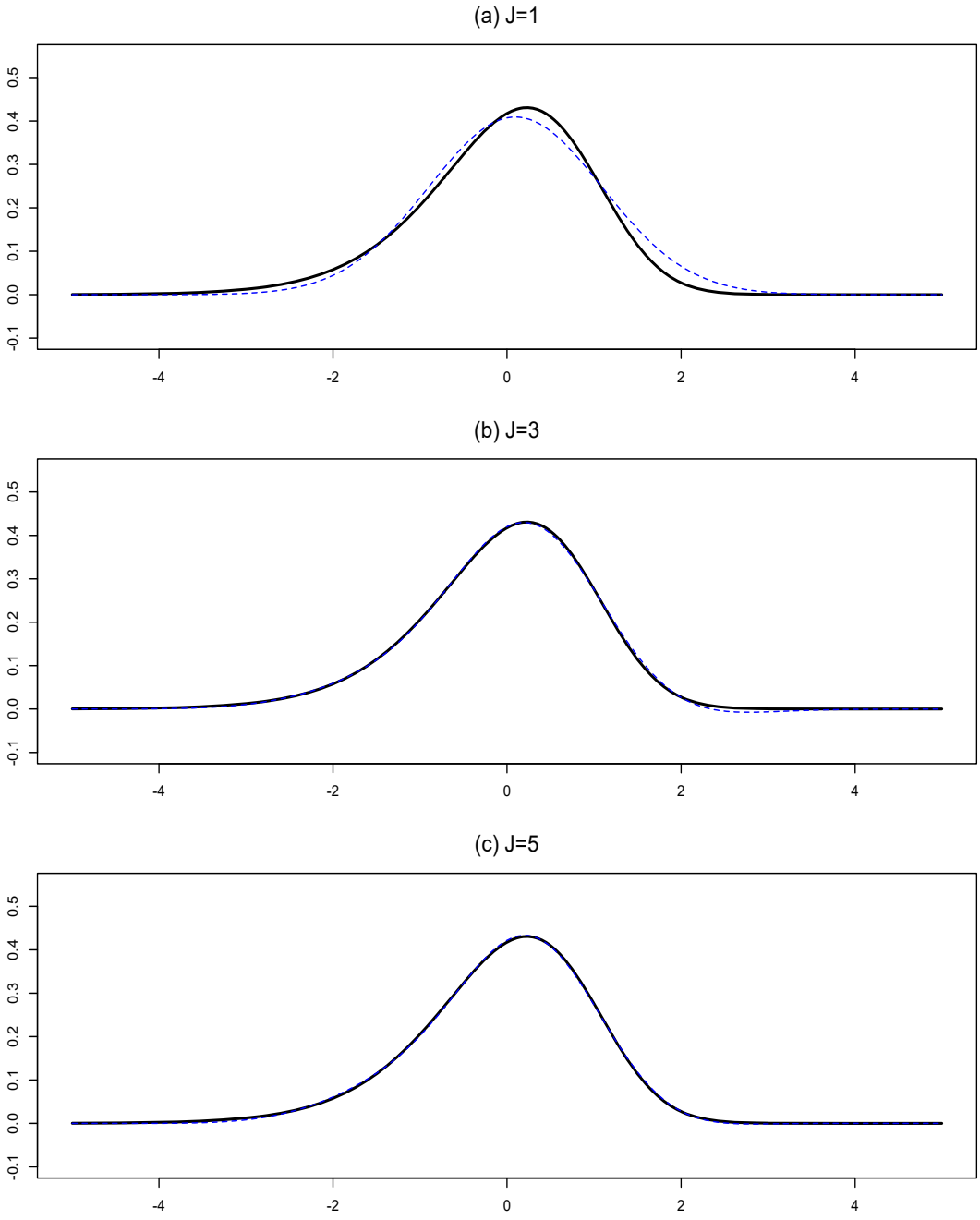
Note. In (d) and (e), the solid line represents the true values, and the dotted line represents the fitted values from the regressions. The tails look similar to those in Figure S3 because the density is for the transformed variable, as explained in the paper. In (f), the circles denote the actual Hermite coefficients, and the dotted line the fitted regression line.

Figure S10: Density tails and Hermite coefficients implied by the SVCJ model (the expansion period)



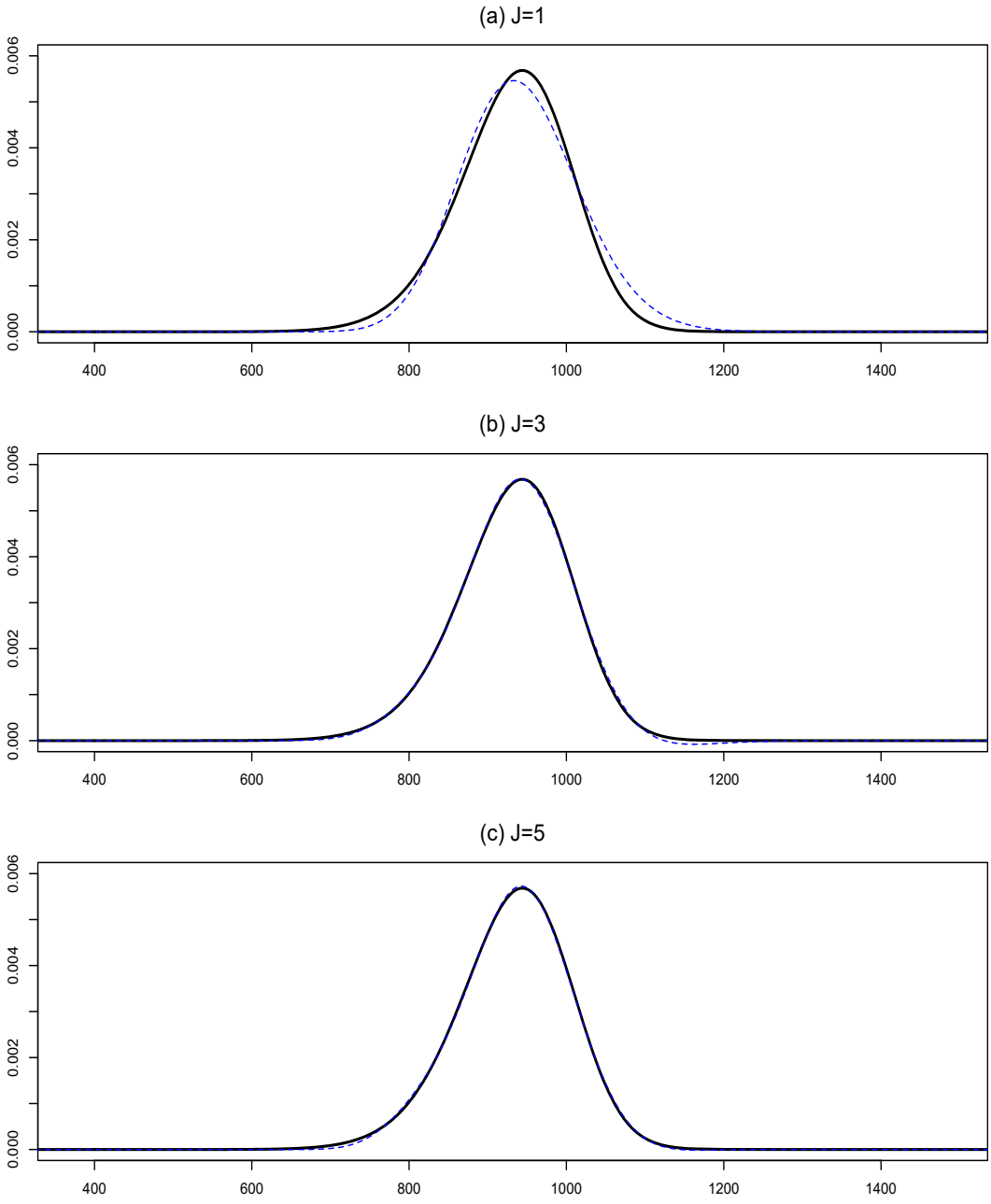
Note. In (g) and (h), the solid line represents the true values, and the dotted line represents the fitted values from the regressions. The tails look similar to those in Figure S3 because the density is for the transformed variable, as explained in the paper. In (i), the circles denote the actual Hermite coefficients, and the dotted line the fitted regression line.

Figure S11: Hermite Series approximations with different truncation orders for the SVCJ model (the recovery period)



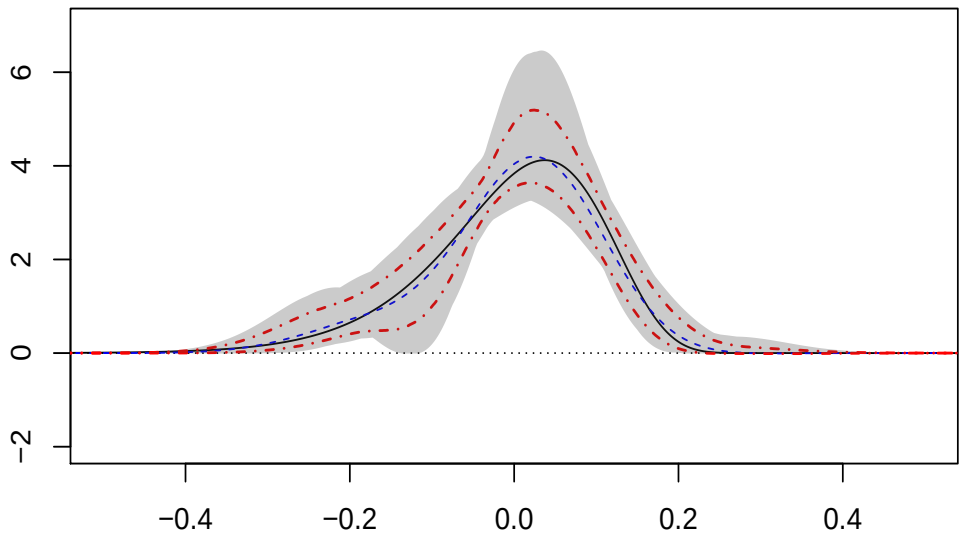
Note. The solid and dashed lines represent the true density and the approximation, respectively.

Figure S12: Hermite series approximations and the original density for the SVCJ model (the recovery period)



Note. The density is for the original variable S. The solid and dashed lines represent the true density and the approximation, respectively.

Figure S13: The empirical distribution of the Sieve estimator



Note. Sample size: n=40 with 20 call and put options, respectively. The solid line: the true density; the dashed line: the median; the two dot-dash lines: the 0.025 and 0.975 quantiles; the shaded area: the entire set of estimates.

Figure S14: Call and put option prices on the three dates

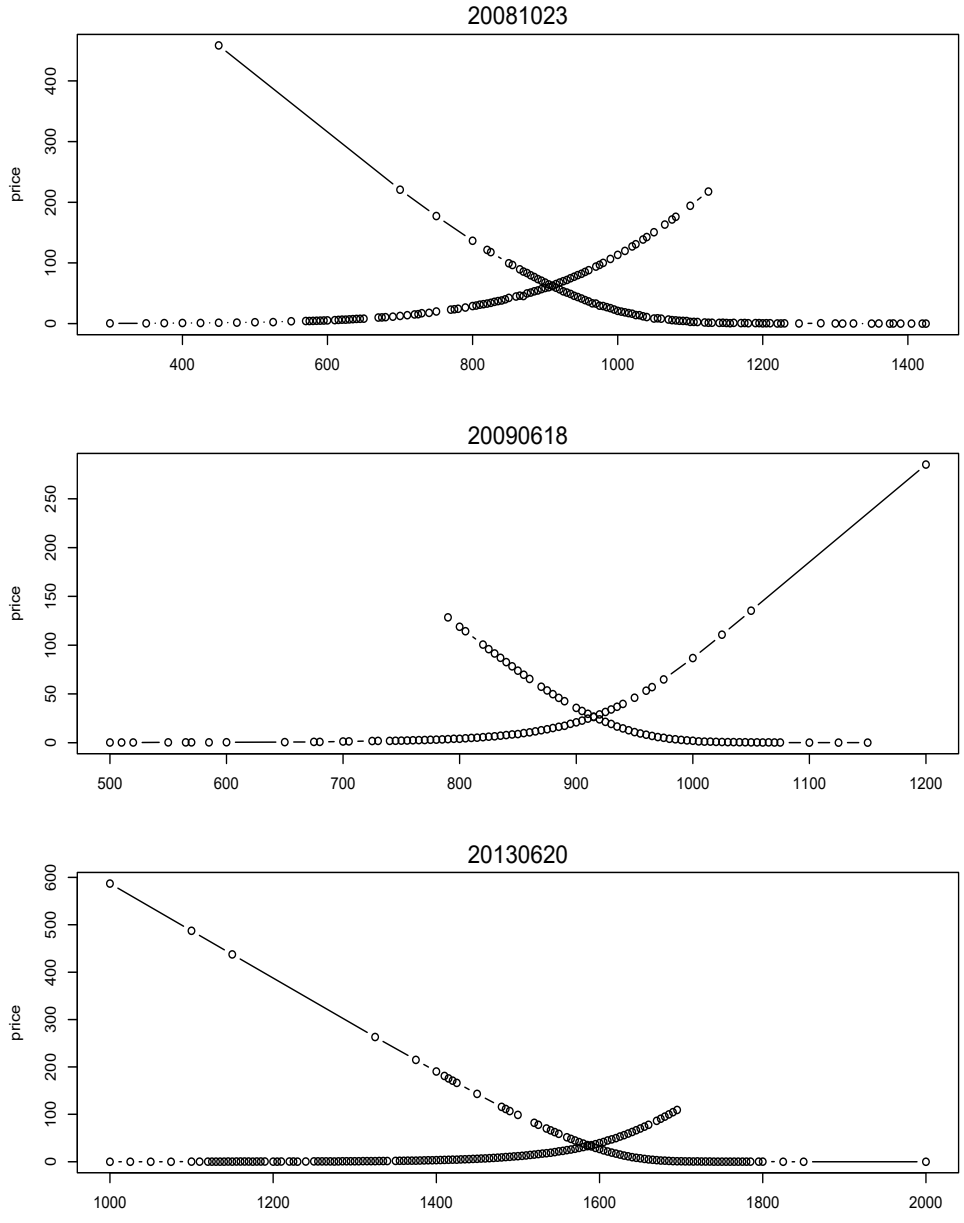


Figure S15: The estimated state price densities after excluding some extreme stikes

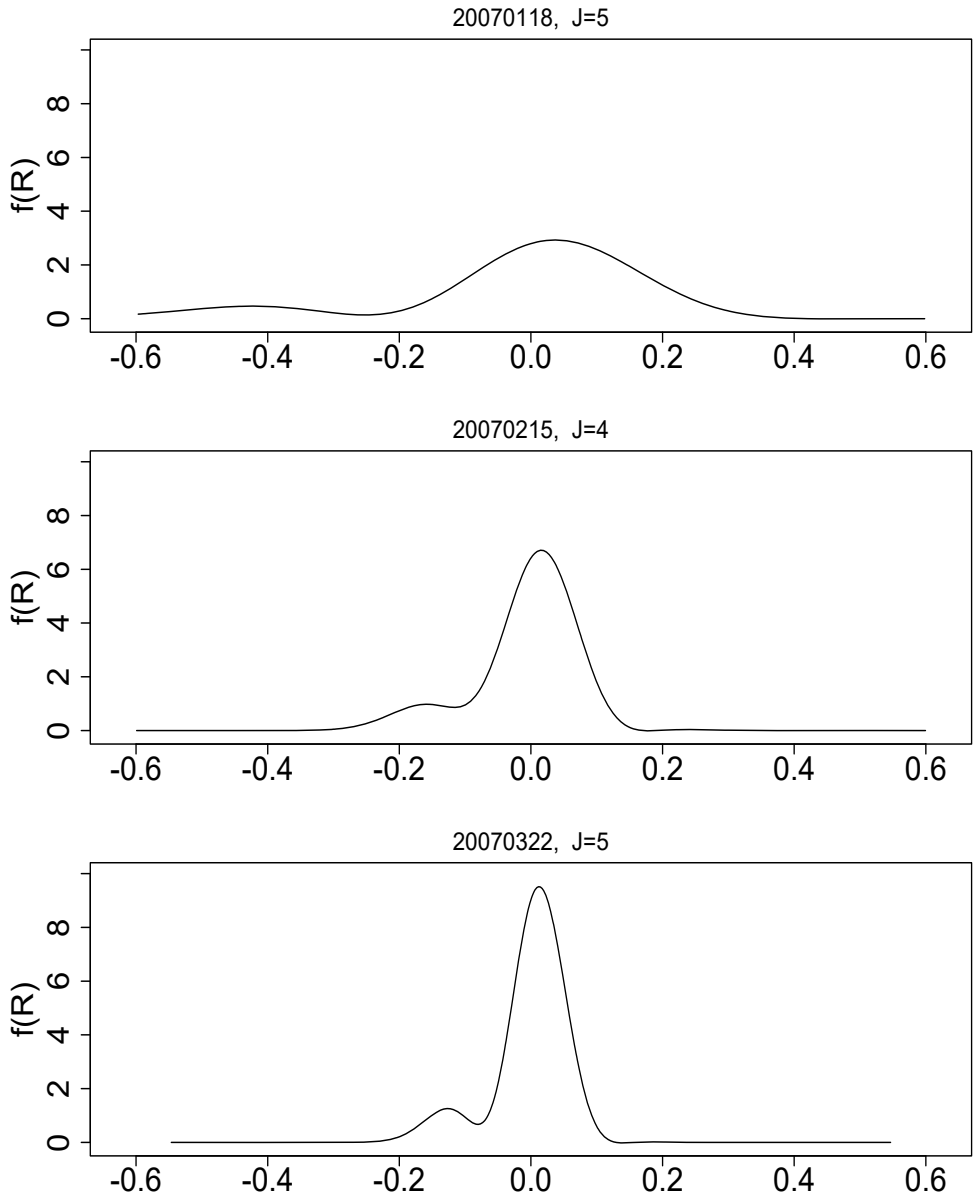


Figure S16: The estimated state price densities with J increased by 1