# Estimating State Price Densities Implied by American Options*

Zhongjun Qu†

Boston University

Guang Zhang‡

Hong Kong University of Science and Technology (Guangzhou)

November 9, 2023

**Abstract**

We introduce a new method to estimate state price densities implicit in American-style options, which are financial contracts granting holders the right to exercise the option at any time before expiration. The method involves estimating the parameters of a Gauss-Hermite series expansion and solving a sequence of recursive equations for the early exercise premium. The resulting estimator can capture sudden shifts in density that may occur during financial crises or in response to significant policy events. It also provides an estimate of the early exercise premium that is of independent interest. We illustrate the proposed method using calibrated simulations and empirical applications. Our findings indicate that the state price densities implied by S&P 500 ETF options can predict future returns up to a one-year horizon for the period 2009-2023.

**Keywords**: State price density, American option, early exercise premium, nonparametric estimation.

**JEL classification**: C14, C22, G13.

*We are grateful to Samuel Messer for comments and suggestions.

†Department of Economics, Boston University (qu@bu.edu).

‡Thrust of Financial Technology, Hong Kong University of Science and Technology (Guangzhou) (guangzhang@hkust-gz.edu.cn).

# 1 Introduction

Equilibrium prices of financial assets reflect both uncertainty and the market's collective preferences regarding potential payoffs. This relationship is summarized by a State Price Density (SPD) in the form of a probability distribution. Intuitively, the SPD allocates weights to various states of potential economic outcomes, with a larger value indicating a higher likelihood and/or a greater preference for positive asset returns in that state. The SPD plays a central role in asset pricing theories, particularly in derivative pricing and term structure modeling. Its importance is also well recognized by central banks, as it provides a predictive distribution that yields insights into future economic conditions. For example, the Federal Reserve Bank of Minneapolis maintains SPD estimates across diverse markets, including commodities, equities, exchange rates, inflation, and interest rates.[1]

The options market represents an ideal setting for estimating SPDs for several reasons: (1) Option contracts have a simple payoff structure; (2) their strike prices encompass a wide range of economic states; and (3) their maturities span from a few days to over a year, offering insights into various future horizons. In practice, index options are primarily European-style, which have a fixed exercise date. In contrast, American-style options, which allow exercise at any time before contract expiration, are more common and cover a broader array of assets. These assets include equities such as individual stocks, market index futures, and equity ETFs; commodities such as gold, silver, and agricultural products; energy commodities including natural gas and crude oil; and fixed-income securities such as US Treasuries and European government bonds. Researchers often work with American options when estimating SPDs for this diverse set of assets.

Many methods exist to estimate SPDs using European-style options. These include: Jackwerth and Rubinstein's (1996) binomial tree-based approach; Jarrow and Rudd's (1982) and Longstaff's (1995) Gram-Charlier based estimator; Ait-Sahalia and Lo's (1998) and Dalderop's (2020) kernel-based estimators; Ait-Sahalia and Duarte's (2003) constrained local polynomial estimator; Shimko's (1993) global polynomial method; and Figlewski's (2010) spline method. As these methods were designed for European options, they do not account for the early exercise premium inherent in American options. Therefore, applying these methods directly to American options can result in biased estimates, especially when considering horizons that extend beyond a few months. Section 4 of this paper quantifies this potential bias through calibrated

[1]Federal Reserve Bank of Minneapolis, "Current and Historical Market-Based Probabilities," https://www.minneapolisfed.org/banking/current-and-historical-market–based-probabilities.

simulations. The bias issue is further examined in the empirical section using actual data, which confirms that the bias can indeed be substantial (See Figure 7 for a visual comparison and the empirical section for more details). These findings underscore the importance of considering the early exercise feature, especially in environments of rising interest rates, as the premium for put options can be sizable in such situations.

In contrast, few methods are available for estimating SPDs using American options. The two main approaches include the method developed by Melick and Thomas (1997), which approximates the price of an American option as a weighted average of upper and lower bounds related to European option prices, and the method introduced by Tian (2011), which iteratively estimates the early exercise premium and the SPD using a binomial-tree model. In empirical studies, researchers often apply Melick and Thomas's (1997) method or assume that the early exercise premium is negligible and resort to methods designed for European options. However, the latter approach requires the exclusion of in-the-money options or limiting the analysis to short horizons to mitigate potential bias. There exists a demand for new methods capable of providing SPD estimates for both short and long horizons without discarding data and that can facilitate a better understanding of the early exercise premium in this context.

This paper introduces a nonparametric estimator to recover SPDs from American options. The proposed approach consists of two components: a Gauss-Hermite series expansion and a recursive equation characterizing the early exercise premium. The estimation entails determining the coefficients of this expansion and solving the recursive equation iteratively. For data input, we require only a single cross-section of American option prices with a common maturity, a proxy for the risk-free interest rate, and an estimate of the dividend rate. Unlike methods that apply temporal smoothing, it can capture sudden shifts in the SPD, such as those resulting from financial crises or major market events. It is suitable for both low- and high-frequency data and also provides an estimate of the early exercise premium that is of independent interest. The estimator generalizes the method of Lu and Qu (2021) from European to American options, incorporating additional steps to account for the early exercise premium.

We evaluate the method's performance using empirically calibrated simulations and compare it with the estimators of Melick and Thomas (1997) and Tian (2011). We then proceed to an empirical application with two objectives: to assess the performance of the proposed estimator using real data and to evaluate whether the recovered SPD can predict future asset returns at various horizons up to one year. The first objective presents a challenge because, for most assets,

we do not observe their European and American options simultaneously, making it difficult to assess issues related to the early exercise premium. Nevertheless, we have identified a pair of assets that provide valuable insights: the S&P 500 index (SPX) and an ETF, called SPY, designed to track the SPX. Importantly, the S&P 500 index provides European-style options, while the SPY options are American-style. This means that we can view SPY options as nearly equivalent to SPX options, with the difference in their prices primarily reflecting the early exercise premium. Therefore, the SPD derived from the SPX options can serve as a reference point for evaluating the accuracy of our method in recovering both the SPD and the early exercise premium.

The empirical results indicate that our methods effectively recover the SPDs, as the estimates based on the SPY options tend to closely mirror those based on the SPX. The results also confirm that accounting for the early exercise feature is indeed important at extended horizons. For example, for the transaction date 02/17/2023, at which the London Inter-Bank Offered Rate (LIBOR) is at 5.6%, ignoring the early exercise premium resulted in a sizable estimation bias at the one-year horizon, consistent with the findings from our simulation analysis.

For the predictive analysis, we examine both predictive mean and quantile regressions for future SPY returns, using a quantile of the SPD as the predictor. We present results across three sample periods to assess potential shifts in predictability due to extreme market conditions: a benchmark sample from 2009.06 to 2020.02, an extended sample that includes the COVID pandemic period (2009.06-2023.02), and a further extended sample which includes the 2008 financial crisis (2007.01-2023.02). For each sample period, we consider four different horizons, at approximately one month, three months, six months, and one year.

The predictive mean regressions indicate a statistically significant predictive relationship (at the 10% level) for the first two sample periods. The evidence is clear-cut at monthly to semi-annual horizons and becomes more mixed at the annual horizon. Lower quantiles consistently display stronger predictive power than upper quantiles. Intuitively, a substantial decrease in the lower quantiles is usually triggered by a negative market event (e.g., the onset of the COVID pandemic). Over the past 20 years, the market has generally recovered well from large market declines. As a result, such shifts in lower quantiles often signaled an investment opportunity, which, on average, produced higher subsequent returns over various horizons.

The predictive quantile regressions yield clear-cut results for the monthly, quarterly, and semi-annual horizons for all three sample periods. Specifically, when a quantile above the median is used to predict the same quantile of the return distribution, the slope coefficients are mostly

significant. In contrast, when using a quantile below the median as the predictor, the estimates are often insignificant. This asymmetry reflects the same intuition as in the mean regression case: given that the market has generally rebounded from significant declines or crashes, the real market outcomes often exceeded initial pessimistic expectations. In other words, the lower quantiles of the SPD captured a fear (or risk premium) that was accompanied by higher rather than lower subsequent returns. These lower quantiles did not have a strong correlation with the lower quantiles of the actual return distribution.

Our empirical results add to the extensive literature examining the relationship between the options market and subsequent stock returns. Bollerslev, Gibson, and Zhou (2011) demonstrate that the volatility risk premium predicts excess returns on the S&P500 index at monthly and quarterly horizons from 1990Q1 to 2003Q2. An, Ang, Bali, and Cakici (2014) find that stocks with large prior increases in call (put) implied volatilities subsequently exhibit high (low) future returns, a pattern most pronounced at the monthly horizon and persisting up to six months for the sample period 1996-2011. Xing, Zhang, and Zhao (2010) examine the predictability of the implied volatility skewness of individual stock options from 1996-2005. They find that the predictability weakens after one month but can extend up to half a year. Some studies have used the tails of the SPD to forecast future returns; see Andersen, Fusari, and Todorov (2015) by parametric methods and Andersen, Todorov, and Ubukata (2021) by nonparametric methods. Since our methods allow us to extract the SPD from American options, they facilitate the assessment of the predictive power of the SPD for a broader range of markets than those analyzed using European options.

The rest of the paper is organized as follows: Section 2 provides an overview of concepts related to American options in preparation for subsequent analyses. Section 3 introduces the SPD estimator. In Section 4, we evaluate the estimator's performance using simulations, with parameters calibrated to empirical estimates. Section 5 presents an empirical application related to SPDs implied by the SPX index and the SPY ETF, and Section 6 offers the conclusion. Additional results and details are available in an online appendix.

# 2 American options and the early exercise premium

This section reviews basic properties of American options needed for subsequent analyses. Our estimator utilizes a cross-section of call and put options with the same maturity date.

An American option is a financial contract granting holders the right to buy or sell an un-

derlying asset at a predetermined price (i.e., the strike price) at any time before a specified expiration date. Throughout the analysis, we denote present time by $t$, and the expiration date by $T$. We let $S_v$ be the price of the option's underlying asset at time-$v$ for any $t \leq v \leq T$. Using this notation, the arbitrage-free price of an American call option with a strike price $K$ can be expressed as

$$AmCall_t = \sup_{t \leq v \leq T} e^{-r(v-t)} E\left[(S_v - K)^+ | \mathcal{F}_t\right],$$

where $(X)^+ = \max(X, 0)$; $r$ is the risk-free rate; and $\mathcal{F}_t$ is the information set used by the market to price the option at time $t$. The conditional expectation is always taken with respect to the risk-neutral probability measure, and the supremum operator indicates that the option can be exercised at any time prior to the expiration date.

Our SPD estimator is based on the following well-known decomposition of the equilibrium price:

$$AmCall_t = EuCall_t + EEC_t, \tag{1}$$

where $EuCall_t$ is the price of a European call option for the same asset with the same strike price and expiration date, given by

$$EuCall_t = e^{-r(T-t)} E\left[(S_T - K)^+ | \mathcal{F}_t\right], \tag{2}$$

and $EEC_t$ is the early exercise premium, equal to

$$EEC_t = \int_t^T e^{-r(v-t)} E\left[(\delta S_v - rK)\, 1_{(S_v \geq B_v)} | \mathcal{F}_t\right] dv, \tag{3}$$

where $\delta$ denotes the dividend rate.

The early exercise premium reflects a trading strategy of borrowing cash of amount $K$ at the rate $r$ and holding one share of the underlying asset when its price is above the exercise boundary $B_v$ for $v \in [t, T]$; see Carr, Jarrow, and Myneni (1992). An option holder is indifferent between exercising the option or not at the exercise boundary, therefore the following recursive equation must hold when $S_v = B_v$:

$$B_v - K = EuCall_v + EEC_v \quad \text{for any } v \in [t, T). \tag{4}$$

The left hand side of the equation represents the revenue from exercising the option, and the right hand side represents the continuing value of the option if it is not exercised at time $v$. In particular, $EuCall_v$ is the European call option price in (2) at time $v$ when the underlying asset

price equals $B_v$, and $EEC_v$ is the corresponding early exercise premium as in (3), which depends on the remaining path of the early exercise boundary. Kim and Yu (1996) and Detemple and Tian (2002) verified (1) and (3) for general diffusion models. For geometric Brownian motions, more explicit expressions are available and are provided by Kim (1990), Jacka (1991), and Carr, Jarrow, and Myneni (1992). For jump diffusions, the equation (1) continues to hold; however, (3) requires modifications; see Gukhal (2001). For our analysis, we first derive the SPD estimator assuming no jumps in asset prices and then examine the estimator's performance when there are jumps via empirically calibrated simulations.

Similar decompositions as those in (1)-(4) hold for put options: The arbitrage-free price of a put option is the sum of a European put option and an early exercise premium:

$$AmPut_t = EuPut_t + EEP_t,$$

where

$$\begin{aligned} EuPut_t &= e^{-r(T-t)} E\left[(K - S_T)^+ |\mathcal{F}_t\right], \\ EEP_t &= \int_t^T e^{-r(v-t)} E\left[(rK - \delta S_v)\, 1_{(S_v \le B_v)} |\mathcal{F}_t\right] dv. \end{aligned} \tag{5}$$

The resulting early exercise boundary $B_v$ satisfies

$$K - B_v = EuPut_v + EEP_v \quad \text{for any } v \in [t, T), \tag{6}$$

where $EuPut_v$ is the European put option price at time $v$ when the underlying asset price equals $B_v$, and $EEP_v$ is the corresponding early exercise premium.

The conditional expectations in these expressions all involve the SPD. We denote its value for the time-$v$ distribution conditional on time-$s$ information as $f_s^*(S_v) = f^*(S_v|\mathcal{F}_s)$. For instance, in the case of (2), the equation now reads explicitly as

$$E\left[(S_T - K)^+ |\mathcal{F}_t\right] = \int (S_T - K)^+ f_t^*(S_T) dS_T.$$

Once $f_s^*(S_v)$ is specified for any $(s, v)$ in $t \le s \le v \le T$, all the expressions between (1) and (6) can be computed numerically or through simulations. In essence, our estimation represents an inverse problem, in which we infer the SPD from these expressions using observed option prices as input.

# 3 Proposed estimator

Our proposed estimator has two components: a Gauss-Hermite series approximation to the SPD and a recursive equation for the early exercise premium. We first describe these two components and then present the main steps of the estimation procedure.

## 3.1 Approximations to the SPD and the early exercise premium

It is well known that the shape of an SPD, especially its dispersion, systematically varies with market conditions, asset types, and time-to-maturity. Its variance can increase sharply during market stress and decreases to zero as the option approaches maturity. Consequently, utilizing a fixed set of basis functions, without any data-dependent standardization, will not capture this density accurately across diverse applications. To address this, we follow Lu and Qu (2021) to implement a change-of-variables technique before introducing any approximation.

Specifically, let $S_t$ be the spot price of the underlying asset at time $t$. We define the following change-of-variables operations:

$$x = \frac{\log(S_T/S_t) - r\tau}{\sigma\sqrt{\tau}} \tag{7}$$

and

$$z = \frac{\log(K/S_t) - r\tau}{\sigma\sqrt{\tau}}. \tag{8}$$

Here, $r$ is the risk-free rate, $\tau = T - t$ represents time to maturity, $S_T$ is the potential price of the underlying asset at maturity, $K$ is the strike price, and $\sigma$ is the Black-Scholes implied volatility, computed using the at-the-money call option price. We let $f_t(x)$ denote the SPD of the transformed variable $x$. Using $f_t(x)$, the European call and put option prices in (2) and (5) can be expressed as, respectively,

$$EuCall_t = \int_{-\infty}^{\infty} S_t \left[e^{\sqrt{\tau}\sigma x} - e^{\sqrt{\tau}\sigma z}\right]^+ f_t(x)dx \tag{9}$$

and

$$EuPut_t = \int_{-\infty}^{\infty} S_t \left[e^{\sqrt{\tau}\sigma z} - e^{\sqrt{\tau}\sigma x}\right]^+ f_t(x)dx.$$

The transformation in (7) achieves two goals simultaneously. First, the division by $\sqrt{\tau}\sigma$ acts as a variance-stabilizing transformation. For instance, if the variance of $f_t^*(x)$ increases abruptly during a market crash, then $\sigma$ will rise immediately, ensuring the variance of $f_t(x)$ remains relatively stable and close to one. Second, the logarithmic transformation, $\log(S_T/S_t)$, shifts the support of the density from the positive axis to the real line. Thus, if $f_t^*(x)$ is close to

a log-normal density, $f_t(x)$ will be close to a normal density due to this transformation. These two properties are crucial for achieving an effective approximation using the standard normal distribution as the center of the series expansion. Additionally, the $z$ in (8) is frequently used as a measure of moneyness for 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). In our estimation, we first estimate $f_t(x)$ and then we reverse the transformation in (7) to obtain $f_t^*(x)$ as follows:

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

After the change-of-variable operation, we approximate $f_t(x)$ in (9) using a Gauss-Hermite series expansion. Although we follow the same approach as in Lu and Qu (2021), we provide full details here to ensure clarity. Recall that the Hermite functions $\{h_j\}$ are the complete orthonormal system in $L^2(-\infty, \infty)$, given by

$$h_j(x) = [H_j(x)/(2^j j! \pi^{1/2})^{1/2}] e^{-x^2/2}$$

with $\int_{-\infty}^{\infty} h_j^2(x)dx = 1$ for all $j = 0, 1, 2, ...$; $\int_{-\infty}^{\infty} h_i(x)h_j(x)dx = 0$ for all $i \neq j$; and $\{H_j(x)\}$ are the standard physicist's Hermite polynomials given by $H_j(x) = (-1)^j e^{x^2}(d^j/dx^j)e^{-x^2}$. By expressing $f_t(x)$ in terms of $h_j(x)$, we obtain

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

where

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

Applying a truncation to (10), we obtain a Gauss-Hermite series approximation to $f_t(x)$:

$$f_t(x) \approx \sum_{j=0}^{J} \beta_j h_j(x), \tag{12}$$

where $\beta_j$ is defined in (11), and $J$ is the truncation order. The center of this approximation is proportional to a standard normal density, and the additional terms capture deviations from normality such as a thick left tail. Finally, after reversing the change-of-variables, we obtain:

$$f_t^*(S_T) \approx \frac{1}{\sigma\sqrt{\tau}S_T} \sum_{j=0}^{J} \beta_j h_j \left(\frac{\log(S_T/S_t) - r\tau}{\sigma\sqrt{\tau}}\right). \tag{13}$$

We now turn to the early exercise premium. Recall that the equation determining this premium is given by (c.f.(3))

$$EEC_t = \int_t^T \int_0^{\infty} e^{-r(v-t)} (\delta S_v - rK) 1_{(S_v \geq B_v)} f_s^*(S_v) dS_v dv.$$

Thus, $f_s^*(S_v)$ is required to compute $EEC_t$. For this purpose, we use the same Gauss-Hermite series approximation as in (13):

$$f_s^*(S_v) \approx \frac{1}{\sigma\sqrt{v-s}S_v} \sum_{j=0}^{J} \beta_j h_j \left( \frac{\log(S_v/S_s) - r(v-s)}{\sigma\sqrt{v-s}} \right),$$

where the truncation order $J$ and the Gauss-Hermite coefficients are as in (12). Using this approximation, we obtain the early exercise boundary by solving the nonlinear equation in (4) and computing the integration in (3) numerically. The overall approximation is internally consistent because $f_s^*(S_v)$ converges to $f_t^*(S_T)$ as $s \to t$ and $v \to T$.

## 3.2 Estimation procedure

We now present the estimation procedure. Let $y_i$ and $z_i$ $(i = 1, ..., n)$ denote the observed option prices and transformed strike prices at time $t$ (c.f. (7) and (8)). The $t$ index is omitted here and in subsequent discussions to simplify the notation. We assume the data are arranged such that the first $n_c$ observations are call options, and the remaining $n - n_c$ observations are put options. Then, we have

$$y_i = \begin{cases} EuCall_i + EEC_i + \varepsilon_i & \text{for } i = 1, ..., n_c, \\ EuPut_i + EEP_i + \varepsilon_i & \text{for } i = n_c + 1, ..., n. \end{cases}$$

In each equation, the first two terms represent the theoretical option price, and $\varepsilon_i$ accounts for pricing errors. We define

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

for $j = 0, ..., J$. Those $x_{i,j}$ terms will serve as the regressors in our regression. Let

$$x_i = (x_{i,0}, ..., x_{i,J})'.$$

The estimation procedure consists of three steps:

**STEP 1.** Obtain an initial estimate of the SPD: Assume $EEC_i = EEP_i = 0$ and solve

$$\min_{\beta \in \mathcal{H}_J} \sum_{i=1}^{n} \left(y_i - x_i'\beta\right)^2, \tag{14}$$

with

$$\mathcal{H}_J = \left\{ \beta \in R^{J+1} \colon \inf_{x \in \mathbb{R}} \sum_{j=0}^{J} \beta_j h_j(x) \geq \eta \right\}, \tag{15}$$

where $\eta$ is a small negative constant introduced to account for the Gauss-Hermite approximation error. Let $(\hat{\beta}_0, ..., \hat{\beta}_J)$ be the estimate of $\beta$, and compute

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

**STEP 2.** Compute the early exercise boundary and the early exercise premium using this SPD estimate: Let

$$\hat{f}_s^*(S_v) = \frac{1}{\sigma\sqrt{v-s}S_v} \sum_{j=0}^{J} \hat{\beta}_j h_j \left( \frac{\log(S_v/S_s) - r(v-s)}{\sigma\sqrt{v-s}} \right).$$

To determine the early exercise boundary of a call option with strike price $K_i$, solve numerically for $B_v$ using

$$\hat{B}_v - K_i = \widehat{EuCall}_v + \widehat{EEC}_v \quad \text{for any } v \in [t, T),$$

where $\widehat{EuCall}_v$ and $\widehat{EEC}_v$ correspond to $EuCall_v$ and $EEC_v$, but with $f_v^*(\cdot)$ and $B_v(\cdot)$ replaced by $\hat{f}_v^*(\cdot)$ and $\hat{B}_v(\cdot)$, respectively. The resulting estimate for the early exercise premium is

$$\widehat{EEC}_i = \int_t^T \int_{-\infty}^{\infty} e^{-r(v-t)} \left(\delta S_v - rK\right) 1_{(S_v \geq \hat{B}_v)} \hat{f}^*(S_v) dS_v dv.$$

Compute the early exercise premium for put options, denoted $\widehat{EEP}_i$, in a similar way.

**STEP 3.** Update the SPD estimate and the early exercise premium: First, use the early exercise premium from STEP 2 to re-estimate the SPD. Specifically, re-solve (14) after replacing $y_i$ with $y_i - \widehat{EEC}_i$ for call options and replacing $y_i$ with $y_i - \widehat{EEP}_i$ for put options. Next, use the updated SPD estimate to recompute the early exercise premium as in STEP 2. Repeat this process until changes in the SPD estimate are small.

**Remark 1** *STEP 1 of the procedure produces an initial estimate of the SPD by treating American options as their European counterparts. STEP 2 calculates expressions associated with the early exercise feature based on this initial estimate. STEP 3 refines these estimates in an iterative manner. This sequential approach breaks the computation into simple segments, avoiding an otherwise complex nonlinear problem.*

**Remark 2** *By (10), the Gauss-Hermite approximation error equals* $f(x) - \sum_{j=0}^{J} \beta_j h_j(x) = \sum_{j=J+1}^{\infty} \beta_j h_j(x)$, *which can be negative for some* $x \in \mathbb{R}$. *Therefore, in* $\mathcal{H}_J$, *the parameter* $\eta$ *should be slightly negative to account for the effect of the truncation. We suggest setting it to* $\eta$ *=-1E-3. The optimization problem in Step 2 is strictly convex because* $\mathcal{H}_J$ *is a convex set and the criterion function (14) is strictly convex. We use the* `solve.QP` *routine in R to implement this step.*

*The parameter $J$ controls the approximation error in (12). We suggest choosing $J$ using ten-fold cross-validation over the set $J = \{J^*, J^*+1, J^*+2\}$ with $J^* = ceiling(2 * (n/log(n))^{0.2})$. These recommendations follow Lu and Qu (2021). The estimator presented in this paper extends the methodology of Lu and Qu (2021) from European options to American options by incorporating STEPs 2 and 3 to account for the early exercise premium.*

# 4 Empirically calibrated simulations

In this section, we evaluate the performance of the proposed SPD estimator using five well-known parametric models as data generating processes. These five models span diverse markets: equity index options, equity index futures options, volatility index options, and commodity options. We generate data using empirically calibrated parameter values. A summary of these data-generating processes is provided below, with more details available in the appendix.

**Heston's stochastic volatility (SV) model**. This is a benchmark model used for pricing equity and equity index options. The model specifies the risk-neutral dynamics of the spot price of the underlying asset as

$$dS_t = (r-\delta)S_t dt + \sqrt{V_t} S_t dW_t,$$
$$dV_t = \kappa(\theta - V_t)dt + \sigma\sqrt{V_t} dU_t,$$

where $r$ represents the risk-free rate, $\delta$ represents the dividend rate, $W_t$ and $U_t$ are two Brownian motions with a correlation coefficient $\rho$. The parameter $\theta$ determines the long-run variance level, $\kappa$ affects the speed of mean-reversion, and $\sigma$ is the volatility-of-volatility parameter. For simulations, we use the parameters from the SV panel of Table 1 of Lu and Qu (2021), estimated using S&P500 index options from the period 2013.06–2013.12.

**Two-factor SV model**. As a generalization of the basic SV model, this model allows two factors in the volatility process ($i = 1, 2$):

$$dS_t/S_t = rdt + \sqrt{V_{1,t}} dW_{1,t} + \sqrt{V_{2,t}} dW_{2,t},$$
$$dV_{i,t} = (\alpha_i - \beta_i V_{i,t})dt + \sigma_i \sqrt{V_{i,t}} dU_{i,t},$$
$$Cov(dW_{i,t}, dU_{i,t}) = \rho_i dt,$$
$$Cov(dW_{1,t}, dW_{2,t}) = Cov(dU_{1,t}, dU_{2,t}) = 0.$$

For simulations, we use parameter values from Bates (2000, p.203), which are estimated using S&P500 futures options.

**DMR model**. This model is mainly used for variance swaps, currency options, and interest rates, and it appeared in Bates (2012), Mencia and Sentana (2013), and Xiu (2014). Our DGP is the same as that in Mencia and Sentana (2013):

$$dV_t = \beta(\theta_t - V_t)dt + \sigma\sqrt{V_t}dW_t,$$
$$d\theta_t = \xi(\alpha - \theta_t)dt + \kappa\sqrt{\theta_t}dU_t,$$

where $V_t$ is a variance instrument, and the correlation between the Wiener processes $W$ and $U$ is zero. The parameter values are taken from Table 4 of Mencia and Sentana (2013), where the authors conducted an empirical analysis of VIX derivative valuation models.

**CEV model**. The constant elasticity of variance (CEV) model, introduced by Cox (1996), is a generalization of the Geometric Brownian Motion, which allows the conditional variance of asset returns to depend on the price level. This model has been applied in various contexts, such as modeling short-term interest rates and pricing commodity options. We use the following DGP:

$$dV_t/V_t = (r - \delta)dt + \sigma {V_t}^{\beta} dW_t.$$

The parameter values are taken from Geman and Shih (2008), estimated based on crude oil over the period 01/01/2000 to 12/11/2007, which were also used in Carlos Dias and Pedro Vidal Nunes (2011).

**SVCJ model**. This is a stochastic volatility model with contemporaneous jumps in return and volatility, proposed by Duffie, Pan, and Singleton (2000):

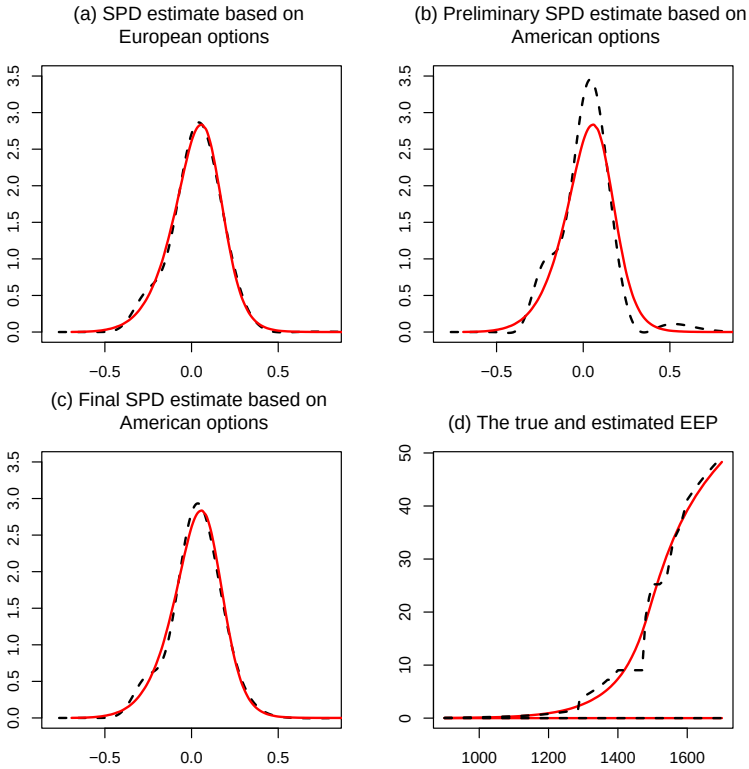
$$dS_t/S_t = rdt + \sqrt{V_t}dW_t + (e^{Z_t^s} - 1)dN_t - \lambda\mu dt$$
$$dV_t = \kappa(\theta - V_t)dt + \sigma\sqrt{V_t}dU_t + Z_t^v dN_t,$$

where the definitions of the variables are the same as in the SV case, $N_t \sim \text{Poi}(\lambda)$ is a Poisson counting process with a constant intensity $\lambda$; $Z_t^s$ denotes jump in return with $Z_t^s \sim \mathcal{N}(\mu_s, \sigma_s^2)$; and $Z_t^v$ denotes jump in volatility which follows an exponential distribution: $Z_t^v \sim \text{Exp}(\mu_v)$. Amongst the parameters, $\theta$ determines the long-run variance level, $\kappa$ affects the speed of mean reversion, $\sigma$ is the volatility-of-volatility parameter, and $-\lambda\mu$, with $\mu = \exp(\mu_s + \sigma_s^2/2) - 1$, compensates for the instantaneous change in the expected return due to the presence of $Z_t^s$. For parameter values, we use the SVCJ panel of Table 1 from Lu and Qu (2021). This DGP is included here to assess the estimation accuracy when jumps are present.

We divide the analysis into two parts. In the first part, we evaluate the estimation methods in an environment when the pricing relationship is exact – meaning there are no pricing

errors or noise in option prices. This allows us to examine the adequacy of the Gauss-Hermite approximation and the early exercise premium approximation, without the interference of estimation uncertainty. It also enables us to assess the magnitude of the estimation bias when the early exercise feature is disregarded, i.e., treating American options as if they were European options. Then, in the second part of the analysis, we introduce noises to option prices and evaluate the finite sample performance of the proposed methods. In all cases, the truncation order of the Hermite approximation is determined via a ten-fold cross-validation over the set $J = \{J^*, J^* + 1, J^* + 2\}$ with $J^* = ceiling(2 * (n/log(n))^{0.2})$. The option prices are generated using Longstaff and Schwartz's (2001) method. The maturity is set to one year, and the results for a six-month horizon are reported in the appendix.

Figure 1: SPD and EEP estimates for the SV model




Note. SPD: state price density for the return. EEP: early exercise premium. Horizon: 1-year. Solid curve: true value. Dashed line: estimates. In (d), the increasing part is the put option EEP. See Appendix for more details about the DGP.

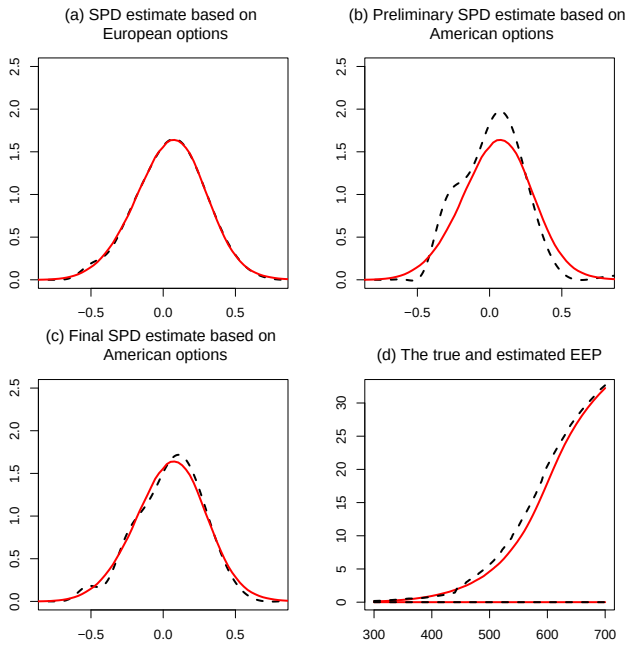
## 4.1 Adequacy of approximations

We present the results in five figures, Figures 1-5, for the five models respectively. Each of these figures consists of four panels. Panel (a) displays the SPD estimate using European option prices,

which serves as a benchmark for comparison. Panel (b) displays the conventional approach that treats American options as European options, resulting in biased estimates. Panel (c) reports the proposed estimator, while panel (d) displays the estimated and the true early exercise premiums. In each case, the true value is shown by the red solid curve, and the estimate is represented by the black dashed line. For panels (a) and (b), the estimates are obtained using the sieve method of Lu and Qu (2021).

Figure 1 shows that large estimation biases are present when ignoring the early exercise feature (Panel b of Figure 1). It also shows that accounting for the early exercise premium using the proposed methods improves the results substantially (Panel c). The resulting estimates are close to the infeasible estimates (Panel a), for which the unknown early exercise premium is replaced by its true value. The recovered early exercise premium sometimes deviates from the true value for in-the-money put options. Nevertheless, this difference has no visible detrimental effect on the SPD estimation; it appears that for the estimation of the SPD, what matters is the approximate value of the early exercise premium rather than its exact value.

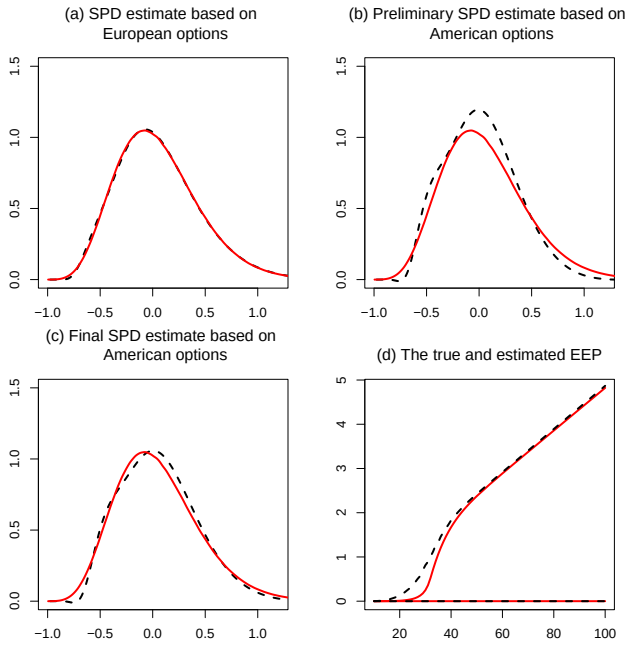
Figure 2: SPD and EEP estimates for the two-factor SV model




Note. Horizon: 1-year. Solid curve: true value. Dashed line: estimates. In (d), the increasing part is the put option EEP. See Appendix for more details about the DGP.

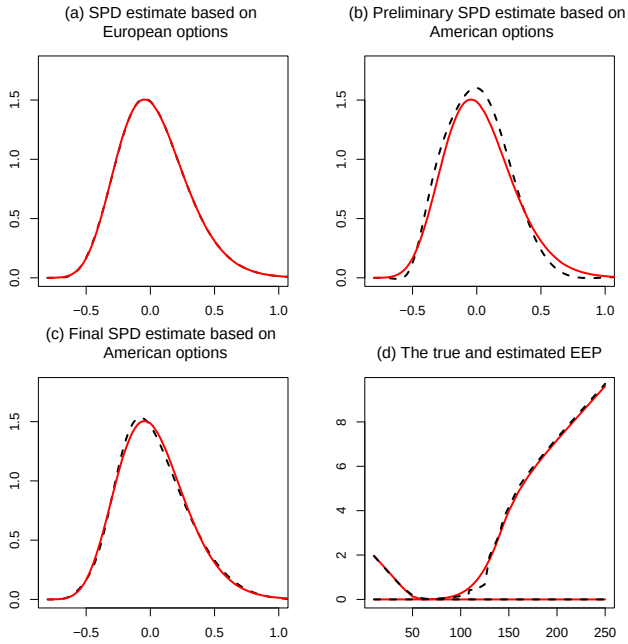
Figures 2-5 indicate similar conclusions as Figure 1. In particular, ignoring the early exercise premium can lead to large biases, and accounting for it using the proposed methods can improve the results substantially. The methods perform well in the SVCJ model case, despite the fact

Figure 3: SPD and EEP estimates for the DMR model




Note. 1-year horizon. Solid curve: true value. Dashed line: estimates. In (d), the increasing part is the put option EEP. See Appendix for more details about the DGP.

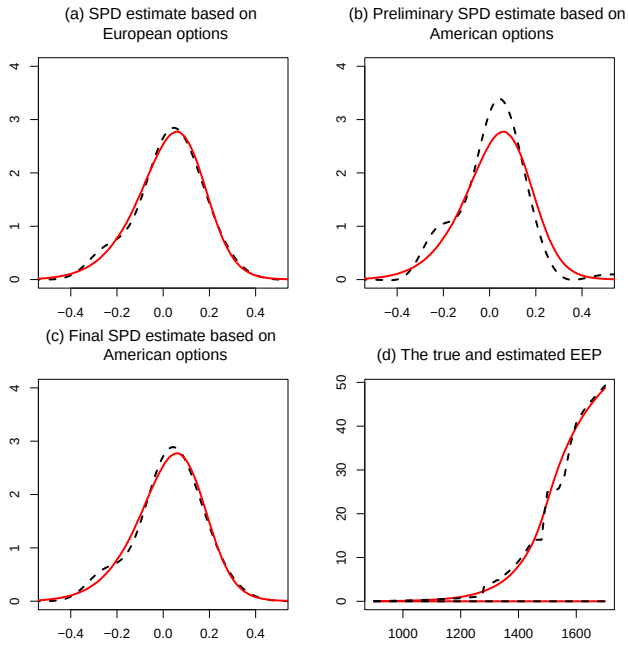
Figure 4: SPD and EEP estimates for the CEV model




Note. 1-year horizon. Solid curve: true value. Dashed line: estimates. In (d), the increasing part is the put option EEP. See Appendix for more details about the DGP.

that the early exercise boundary formula does not explicitly account for jumps. Figures A.1-A.5 in the appendix provide further evidence that incorporating the early exercise premium can be

Figure 5: SPD and EEP estimates for the SVCJ model




Note. 1-year horizon. Solid curve: true value. Dashed line: estimates. In (d), the increasing part is the put option EEP. See Appendix for more details about the DGP.

important even at the six-month horizon. In particular, for the SV, two-factor SV, and SVCJ models, neglecting the early exercise premium results in significant biases and the proposed methods are able to reduce them effectively.

## 4.2 Finite sample performance

We now introduce noise to option prices. For each of the five models described above, we first generate option prices as in Section 4.1 and then add an independent noise term that is uniformly distributed between -10% and 10% of the price level. We further assume that the noise is capped by an upper bound. Specifically, for Heston's SV model, the two-factor SV model, and the SVCJ model, we set an upper bound of either five dollars or ten dollars. For the DMR model and the CEV model, we set a cap of either fifty cents or one dollar because the option prices are lower (with maximum option prices being approximately 149 and 77 in these two cases, respectively). This simulation design is similar to that in Lu and Qu (2021).

We compare our methods with those of Melick and Thomas (1997) and Tian (2011). Melick and Thomas (1997) specified the SPD as a mixture of three lognormal distributions (the MLN

estimator):

$$f_t^{MLN}(S_T) = \sum_{i=1}^{3} \frac{\pi_i}{\sqrt{2\pi}\sigma_i S_T} \exp\left[-\frac{1}{2}\left(\frac{\log(S_T) - \mu_i}{\sigma_i}\right)^2\right],$$

where $\sum_{i=1}^{3} \pi_i = 1$, and $\pi_i > 0$ for $i = 1, 2, 3$. To address the issue of the early exercise premium, Melick and Thomas (1997) derived bounds of American option prices relative to their European counterparts and subsequently estimated the nine parameters $(\pi_i, \mu_i, \sigma_i), i = 1, 2, 3$, by minimizing the sum of squared option pricing errors. Tian's (2011) method is based on estimating a binomial tree (hereafter, the *i*IB estimator). Starting with an initial value for the SPD, this method calculates the implied early exercise premium and subtracts this value from the American option prices to get an estimate of the European option prices. Afterward, it fits the resulting European option prices to a binomial tree. Tian (2011) did not explicitly provide the number of iterations in the paper but did state that convergence is fast. In our comparison, we implement this method for four iterations, the same as our proposed method.

Table 1 presents the mean integrated squared errors (MISEs) based on 5,000 simulation replications for the five models, for a total of ten cases. The lowest MISE in each case is highlighted in bold. The results show that the proposed method produces the smallest MISE in eight out of ten cases. The reduction in MISE is often substantial, with a decrease of at least 30% relative to the second-best method in seven of these eight cases. For the remaining two cases (DMR model and CEV model with an error bound of \$1), the sieve method produces the second-lowest MISE. These MISE values are only marginally higher than the best values in these instances.

# 5 Empirical application

We carry out this empirical application with two objectives. First, we use an actual dataset to evaluate whether the method can adequately recover the SPD and the early exercise premium. Second, we examine whether the recovered SPD is useful for predictive purposes, specifically for forecasting future realized asset returns at various horizons.

The first objective presents a challenge because most financial assets do not have both European and American options issued on them simultaneously, making it difficult to assess issues related to the early exercise premium. Nevertheless, we have identified a pair of assets that can offer valuable insights: the S&P500 index (SPX), and an ETF, called SPY, designed to track the SPX. These two assets co-move closely and are barely distinguishable from each other, as illus-

Table 1: Mean Integrated Squared Errors

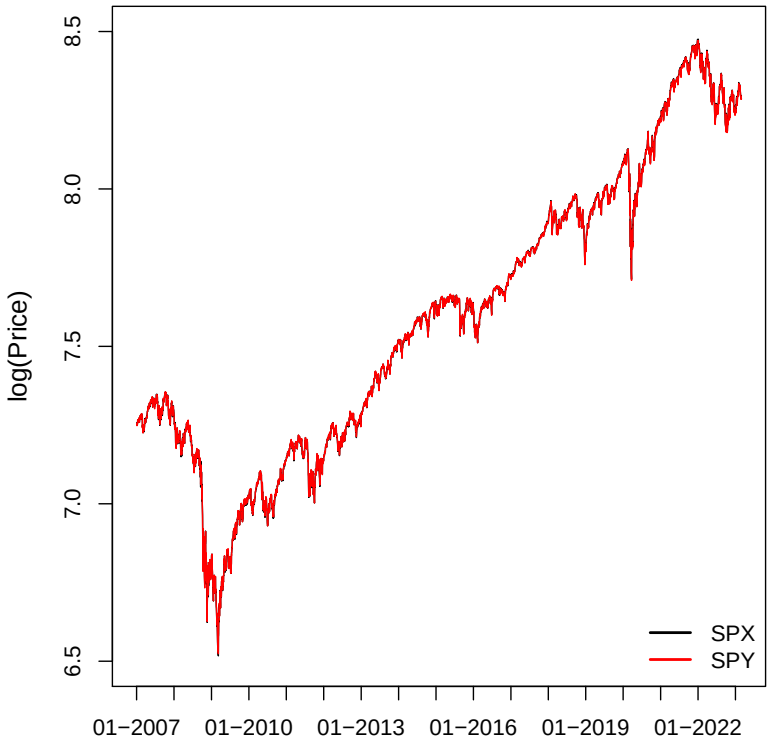
| *A. SV model* | | |
|---|---|---|
| Method | Error bound: \$5 | Error bound: \$10 |
| Proposed | **0.0132** | **0.0179** |
| *i*IB | 0.0503 | 0.0503 |
| MLN | 0.0287 | 0.0325 |
| *B. Two-factor SV model* | | |
| Method | Error bound: \$5 | Error bound: \$10 |
| Proposed | **0.0129** | **0.0166** |
| *i*IB | 0.0239 | 0.0239 |
| MLN | 0.0440 | 0.0445 |
| *C. DMR model* | | |
| Method | Error bound: \$0.5 | Error bound: \$1 |
| Proposed | **0.0107** | 0.0138 |
| *i*IB | 0.0112 | **0.0113** |
| MLN | 0.0381 | 0.0420 |
| *D. CEV model* | | |
| Method | Error bound: \$0.5 | Error bound: \$1 |
| Proposed | **0.0039** | 0.0066 |
| *i*IB | 0.0052 | **0.0052** |
| MLN | 0.0200 | 0.0204 |
| *E. SVCJ model* | | |
| Method | Error bound: \$5 | Error bound: \$10 |
| Proposed | **0.0139** | **0.0176** |
| *i*IB | 0.0535 | 0.0535 |
| MLN | 0.0325 | 0.0366 |

Note. In each panel, the first row corresponds to the proposed estimator, the second row to the *i*IB estimator of Tian (2011), and the third row to the MLN estimator of Melick and Thomas (1997). The lowest MISE is highlighted in bold.

trated in Figure 6, which displays these two series from 01/01/2007 to 02/28/2023. Importantly, the S&P500 index carries European-style options, while the SPY options are American-style. Given the strong linkage between these two series, we can view SPY options as approximately equivalent to SPX options, with the difference in their prices primarily reflecting the early exercise premium. This implies that we can use the SPD computed from the SPX options as a reference point to assess the accuracy of our method in recovering the SPD from American options based on SPY.

We collect daily observations from OptionMetrics and process the data as follows: (1) Remove

Figure 6: SPX and SPY from 01/01/2007 to 02/28/2023




Note. Daily log price levels. SPY: the SPY ETF series; SPX: the S&P500 index. The SPY series is multipled by 10 before taking the log.

all options with zero open interest to exclude outdated information. (2) Retain only option contracts that expire on the third Friday of any month. (3) For these Fridays, keep option contracts with maturities of roughly 1, 3, 6, and 12 months, or equivalently, with $\tau = 30, 91, 182,$ or 365 days. Specifically, when $\tau$ equals 30 or 91 days, retain contracts expiring in [$\tau$-2, $\tau$+2] days, and when $\tau$ is 182 or 365 days, retain those expiring in [$\tau$-5, $\tau$+5] days. The usage of a window ensures that actual return horizons do not deviate significantly from their targets. If two contracts are equidistant from maturity (e.g., 28 days and 32 days when $\tau = 30$), select the contract with more available strikes. (4) Exclude trading days with fewer than 40 available contracts. The resulting sample sizes are detailed in Table 2.

Table 2: Sample sizes for different maturities

| | $\tau = 30$ | $\tau = 91$ | $\tau = 182$ | $\tau = 365$ |
|---|---|---|---|---|
| SPX | 192 | 182 | 90 | 68 |
| SPY | 192 | 164 | 89 | 73 |

## 5.1 SPDs implied by SPX and SPY options

In this subsection, we compare the SPDs estimated from SPX (European-style) and SPY (American-style) options. We first examine an example date closely and then present a summary of the estimates for the entire sample period.

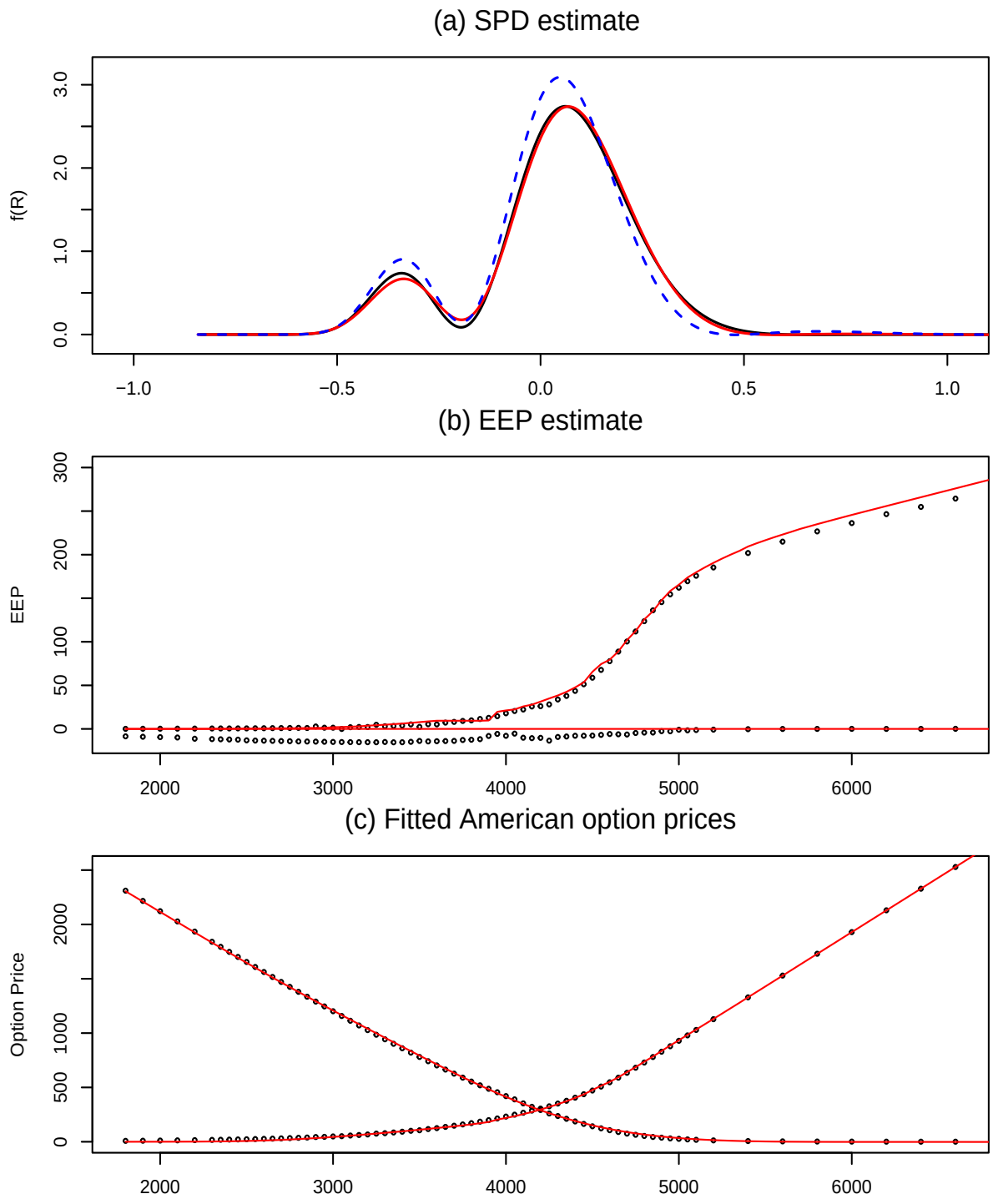
The example date is 02/17/2023, which is the most recent Friday available in the data at the time of writing this paper. On that date, the short-term interest rate stood at 5.6%, following the Fed's efforts to control inflation after COVID. The option maturity is 336 days.

Panel (a) of Figure 7 displays three sets of SPD estimates: the solid black curve represents the SPD estimate based on SPX (European-style) options; the solid red curve is produced by the proposed estimator using SPY ETF (American-style) options; and the dashed blue curve represents the estimate using SPY ETF options but ignoring the early exercise feature. The results show that the proposed estimator yields an SPD close to that obtained from the European options, while ignoring the early exercise feature leads to significant estimation bias. Panel (b) displays the estimated early exercise premiums (represented by the solid curve) relative to their true values computed as differences between SPY and SPX option prices (denoted by circles). The estimates adequately capture the put option early exercise premiums, and they also correctly indicate that the call premiums are nearly zero. In Panel (c), the actual SPY ETF option values (circles) are plotted alongside the fitted values produced by the proposed method (solid line). The fitted values are close to the actual option values across the entire range of strike prices.

Next, we summarize the entire set of SPD estimates for this sample period. Specifically, we plot the 0.1, 0.25, 0.5, 0.75, and 0.9 quantiles of these SPDs, alongside the actual realized future returns for comparison. We report the results for $\tau$ equal to 1 month and 1 year in the main text and include the cases for three and six months in the appendix (Figures A.6 and A.7). Note that the option expiration dates for SPX and SPY available in the data do not always align perfectly with one another, especially for longer horizons; for example, in the annual horizon case, SPX has 68 available transaction dates while SPY has 73. This leads to some variations in the return series and the SPD estimates, but it does not affect the overall patterns observed in the estimates.

Figure 8 confirms that the SPDs implied by SPX and SPY options resemble each other at the one-month horizon, with both distributions displaying clear leftward skewness. The 10th percentile tends to show more extreme negative values than the 90th percentile does positive values, particularly during periods of market stress, such as the recent financial crisis and the

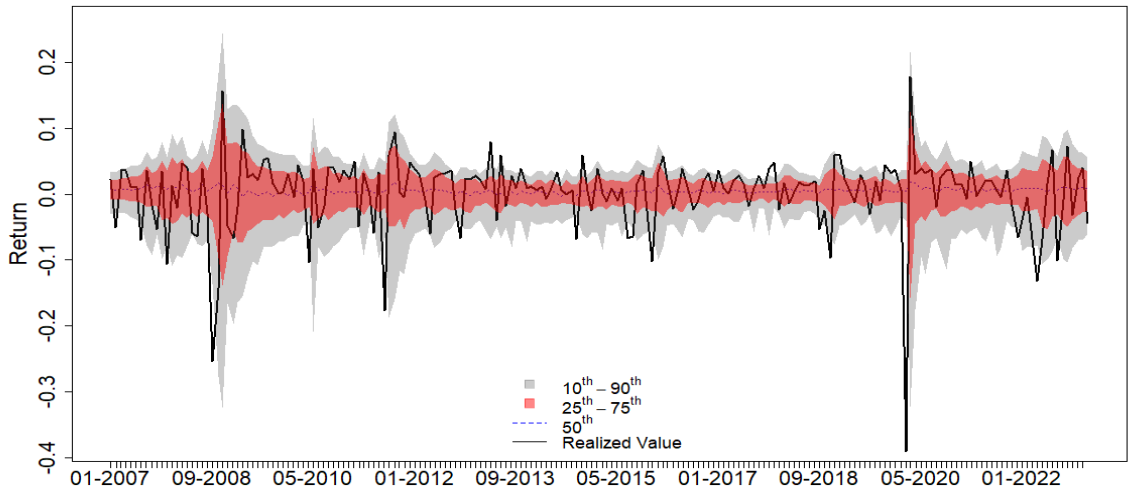
Figure 7: SPD and EEP estimates 02/17/2023



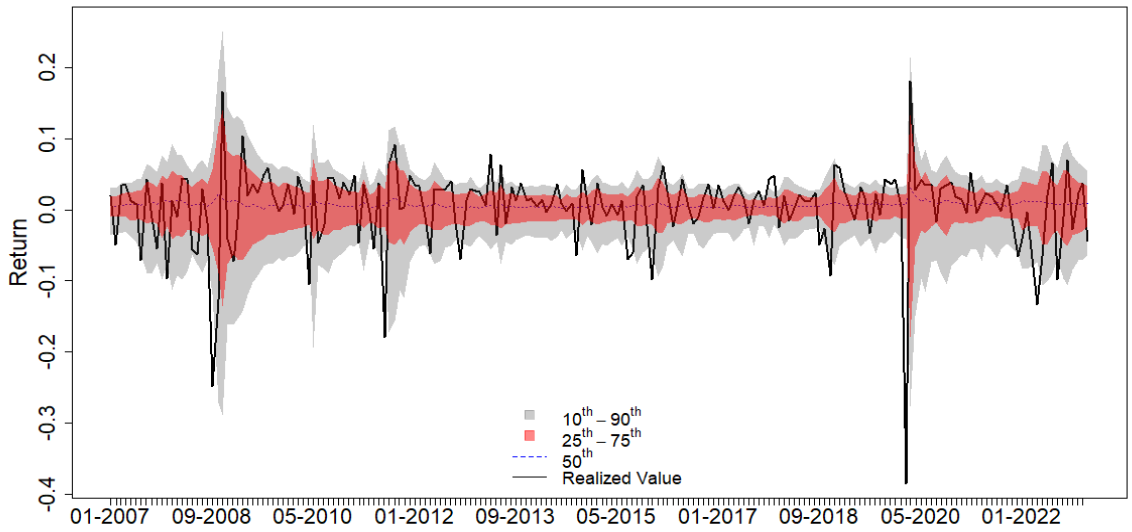

Note. In panel (a), the solid black curve represents the SPD estimate derived from the SPX index options; the solid red curve is the estimate using SPY ETF options; and the dashed blue curve shows the estimate using SPY ETF options that ignores the early exercise feature. In panel (b), the red curves denote the estimated early exercise premiums, while the black circles represent the empirical EEP values, computed as differences between ETF and index option prices. In panel (c), the red solid curves display the fitted option values based on the estimated SPD and EEP using our methodology. The black circles are actual American option prices. The expiration spans 336 days, with the SPX and SPY spot prices being 4079.1 and 4072.6, respectively.

Figure 8: SPD estimates for $\tau = 30$ days

(a) SPY ETF implied densities




(b) SPX implied densities

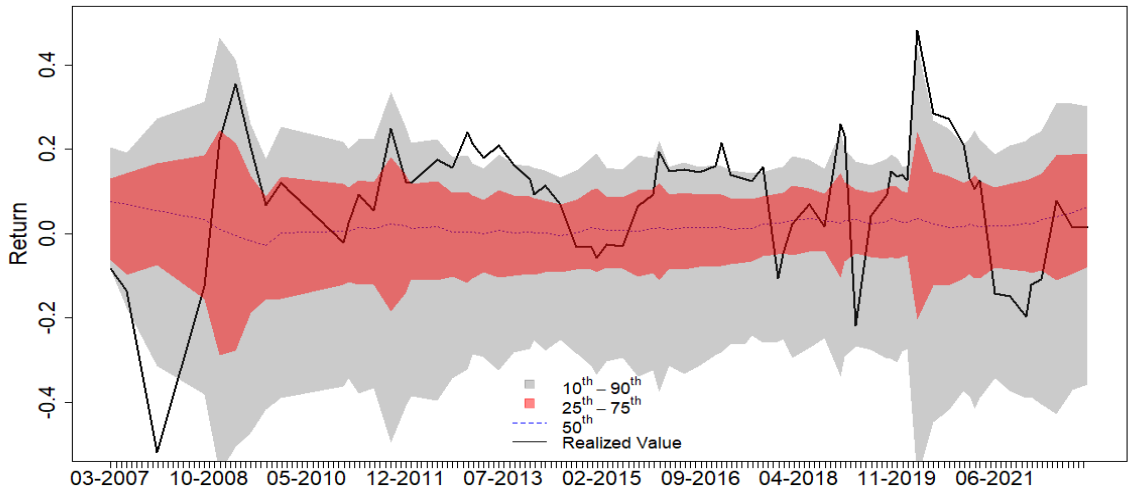



COVID pandemic. The results also reveal that the distribution's upper and lower tails undergo sharp fluctuations during times of economic turmoil, while the median remains stable. This pattern suggests that the tails may offer more predictive information about future returns than the distribution's center can provide. These findings are in line with those reported by Lu and Qu (2021), who analyzed SPDs implied by S&P 500 index options over the 2007-2016 period.

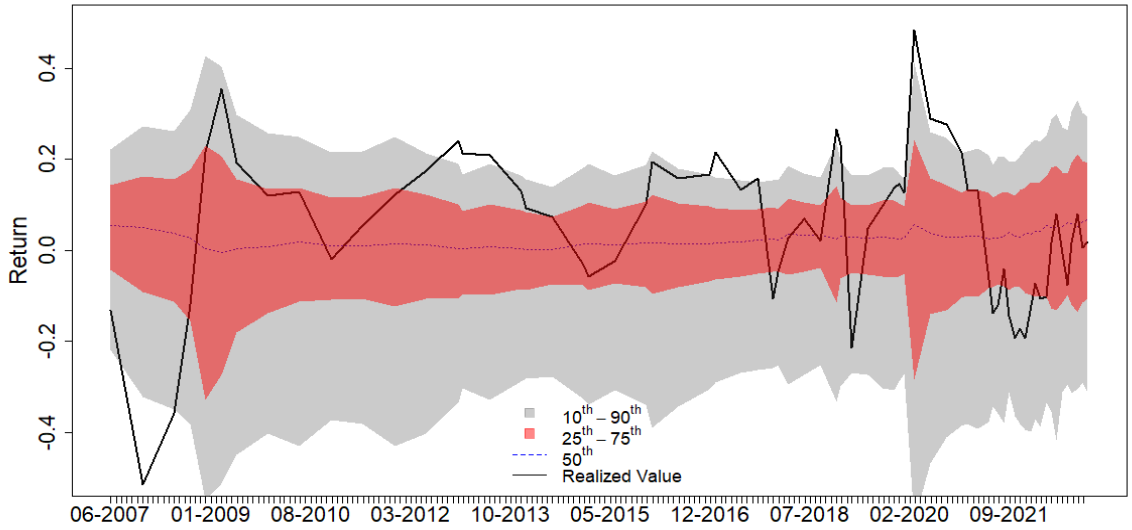
Figure 9 displays the SPD estimates for $\tau = 1$ year. As in the monthly case, the densities implied by SPX and SPY closely resemble each other. The leftward skewness is even more pronounced when contrasting the 10th percentile with the 90th percentile. The movement of

Figure 9: SPD estimates for $\tau = 365$ days

(a) SPY ETF implied densities




(b) SPX implied densities




the median of the distribution continues to remain stable throughout the sample, while the tails widen considerably during market stress, especially during the 2008 crisis and the COVID pandemic period. Appendix Figures A.6 and A.7 present the results for $\tau = 1$ quarter and $\tau = 6$ months, respectively. The observations from these figures reaffirm the conclusions drawn from the monthly and annual horizon cases.

These figures provide insights into the characteristics of the SPDs across different time horizons. Next, we turn to a formal analysis in the context of predictive regressions.

## 5.2 Predictive power of the state price density

Each SPD represents a distribution, offering the opportunity to use various features, such as the mean, variance, higher-order moments, or quantiles, as predictors of future asset returns. Given the observed comovements between the quantiles of the SPD and future realized returns in Figures 8 and 9, we focus on using the quantiles of the SPD as predictors. A similar analysis was reported by Lu and Qu (2021), focusing on a monthly horizon from 2007 to 2016 using S&P 500 index options. In contrast, the analysis here is based on American options, spans longer horizons up to one year, and incorporates the recent COVID period into the analysis.

Table 3: Predictive regression using a quantile of the SPD as predictor ($\tau$=30 days)

| Quantile | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | (a) 06/2009-02/2020 | | | | |
| Estimate | -0.42 | -0.84 | -1.32 | -1.85 | 1.49 | 1.68 | 1.16 | 0.86 | 0.62 |
| s.e. | 0.12 | 0.31 | 0.57 | 1.08 | 1.09 | 0.52 | 0.34 | 0.24 | 0.17 |
| t-value | -3.54 | -2.69 | -2.34 | -1.72 | 1.37 | 3.25 | 3.41 | 3.65 | 3.69 |
| p-value | **0.00** | **0.01** | **0.02** | **0.09** | 0.17 | **0.00** | **0.00** | **0.00** | **0.00** |
| $R^2$ | 0.05 | 0.04 | 0.03 | 0.02 | 0.01 | 0.03 | 0.04 | 0.04 | 0.04 |
| | | | | | (b) 06/2009-02/2023 | | | | |
| Estimate | -0.46 | -0.79 | -1.52 | -2.96 | 2.09 | 1.81 | 1.21 | 0.89 | 0.64 |
| s.e. | 0.12 | 0.15 | 0.44 | 1.33 | 1.21 | 0.67 | 0.39 | 0.26 | 0.19 |
| t-value | -3.98 | -5.26 | -3.43 | -2.22 | 1.73 | 2.72 | 3.10 | 3.40 | 3.46 |
| p-value | **0.00** | **0.00** | **0.00** | **0.03** | **0.09** | **0.01** | **0.00** | **0.00** | **0.00** |
| $R^2$ | 0.09 | 0.09 | 0.08 | 0.06 | 0.02 | 0.06 | 0.07 | 0.08 | 0.08 |
| | | | | | (c) 01/2007-02/2023 | | | | |
| Estimate | -0.21 | -0.43 | -0.75 | -1.63 | 0.50 | 0.79 | 0.57 | 0.41 | 0.30 |
| s.e. | 0.17 | 0.24 | 0.45 | 0.86 | 1.42 | 0.80 | 0.47 | 0.33 | 0.23 |
| t-value | -1.20 | -1.81 | -1.69 | -1.90 | 0.36 | 0.98 | 1.20 | 1.27 | 1.32 |
| p-value | 0.23 | **0.07** | **0.09** | **0.06** | 0.72 | 0.33 | 0.23 | 0.21 | 0.19 |
| $R^2$ | 0.03 | 0.04 | 0.03 | 0.03 | 0.00 | 0.01 | 0.02 | 0.02 | 0.03 |

Note. Dependent variable: return on SPY. Independent variable: a quantile of the SPD implied by SPY options. An intercept is always included. The standard errors account for heteroscedasticity and autocorrelation. The estimates significant at the 10% level are in bold.

We consider both mean and quantile predictive regressions. In each mean regression, we use a quantile of the SPD to predict the future realized return on the SPY. We present estimates from nine regressions, each utilizing one of the nine deciles as the predictor. In each quantile regression, we use a specific quantile, such as the 10th percentile of the SPD, to predict the

corresponding percentile of the future return distribution. These mean and quantile regressions are complementary as they enable us to assess the predictive power for the central tendency and the shape of the return distribution, respectively. In all regressions, the dependent variable is the return on the SPY ETF, while the predictor corresponds to the SPDs implied by SPY options. The dividend rate from the previous year is used when computing the SPD, which ensures that all conditioning information is available when making the predictions. We also conducted regressions using SPX returns and SPX options, and the results were similar.

Table 4: Predictive regression using a quantile of the SPD as predictor ($\tau$=91 days)

| Quantile | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | (a) 06/2009-02/2020 | | | | |
| Estimate | -0.45 | -1.24 | -2.25 | -4.00 | -0.28 | 1.70 | 1.40 | 1.03 | 0.77 |
| s.e. | 0.20 | 0.40 | 0.74 | 1.76 | 1.85 | 0.76 | 0.48 | 0.35 | 0.26 |
| t-value | -2.30 | -3.12 | -3.04 | -2.28 | -0.15 | 2.23 | 2.91 | 2.93 | 2.92 |
| p-value | **0.02** | **0.00** | **0.00** | **0.02** | 0.88 | **0.03** | **0.00** | **0.00** | **0.00** |
| $R^2$ | 0.09 | 0.10 | 0.10 | 0.08 | 0.00 | 0.03 | 0.06 | 0.06 | 0.07 |
| | | | | | (b) 06/2009-02/2023 | | | | |
| Estimate | -0.49 | -1.34 | -2.64 | -5.05 | 1.33 | 1.86 | 1.43 | 1.06 | 0.80 |
| s.e. | 0.21 | 0.47 | 0.93 | 1.80 | 2.20 | 1.10 | 0.67 | 0.45 | 0.32 |
| t-value | -2.31 | -2.86 | -2.83 | -2.81 | 0.60 | 1.70 | 2.15 | 2.36 | 2.49 |
| p-value | **0.02** | **0.00** | **0.01** | **0.01** | 0.55 | **0.09** | **0.03** | **0.02** | **0.01** |
| $R^2$ | 0.13 | 0.15 | 0.15 | 0.11 | 0.01 | 0.07 | 0.10 | 0.11 | 0.12 |
| | | | | | (c) 01/2007-02/2023 | | | | |
| Estimate | -0.39 | -0.69 | -1.42 | -2.99 | -0.49 | 0.81 | 0.72 | 0.55 | 0.42 |
| s.e. | 0.23 | 0.56 | 0.93 | 1.55 | 2.38 | 1.33 | 0.83 | 0.57 | 0.40 |
| t-value | -1.68 | -1.24 | -1.53 | -1.93 | -0.21 | 0.61 | 0.86 | 0.98 | 1.04 |
| p-value | **0.10** | 0.22 | 0.13 | **0.06** | 0.84 | 0.55 | 0.39 | 0.33 | 0.30 |
| $R^2$ | 0.06 | 0.04 | 0.05 | 0.04 | 0.00 | 0.01 | 0.02 | 0.03 | 0.03 |

Note. Dependent variable: return on SPY. Independent variable: a quantile of the SPY-implied SPD. An intercept is always included. The standard errors account for heteroscedasticity and autocorrelation. The estimates significant at the 10% level are in bold.

We report results for three sample periods to evaluate how predictability might change in response to extreme market conditions: a benchmark sample from 2009.6 to 2020.2, an extended sample that includes the COVID pandemic period (2009.6-2023.2), and a further extended sample which includes the 2008 financial crisis (2007.1-2023.2). This choice reflects the prior belief that predictability, even if it exists, might break down under extreme market conditions. We consider

four different maturities for each sample period as in the previous subsection: $\tau = 30, 91, 182$, and 365 days. The standard errors are computed allowing for heteroscedasticity and serial correlation.

Tables 3-6 report the results from the mean regressions, with each table representing a separate predictive horizon. Each column of the table corresponds to a distinct least-squares regression. For instance, in the first column of Table 3, the monthly realized return on the SPY is regressed on the 10th percentile of the lagged SPD. P-values significant at the 10% level are highlighted in bold.

Table 5: Predictive regression using a quantile of the SPD as predictor ($\tau$=182 days)

| Quantile | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 |
|---|---|---|---|---|---|---|---|---|---|
| (a) 6/2009-2/2020 | | | | | | | | | |
| Estimate | -0.30 | -0.99 | -1.86 | -3.02 | -1.29 | 1.46 | 1.37 | 0.99 | 0.70 |
| s.e. | 0.21 | 0.35 | 0.62 | 1.18 | 2.53 | 1.52 | 0.79 | 0.50 | 0.33 |
| t-value | -1.42 | -2.82 | -3.01 | -2.55 | -0.51 | 0.96 | 1.74 | 1.97 | 2.11 |
| p-value | 0.16 | **0.01** | **0.00** | **0.01** | 0.61 | 0.34 | **0.09** | **0.05** | **0.04** |
| $R^2$ | 0.04 | 0.13 | 0.14 | 0.11 | 0.01 | 0.03 | 0.06 | 0.07 | 0.08 |
| (b) 6/2009-2/2023 | | | | | | | | | |
| Estimate | -0.58 | -1.50 | -3.04 | -4.42 | 0.81 | 2.12 | 1.67 | 1.25 | 0.91 |
| s.e. | 0.29 | 0.39 | 1.07 | 1.64 | 2.08 | 1.53 | 0.96 | 0.65 | 0.42 |
| t-value | -1.99 | -3.87 | -2.85 | -2.69 | 0.39 | 1.39 | 1.75 | 1.93 | 2.16 |
| p-value | **0.05** | **0.00** | **0.01** | **0.01** | 0.70 | 0.17 | **0.08** | **0.06** | **0.03** |
| $R^2$ | 0.15 | 0.25 | 0.23 | 0.14 | 0.00 | 0.09 | 0.13 | 0.15 | 0.17 |
| (c) 1/2007-2/2023 | | | | | | | | | |
| Estimate | -0.55 | -1.01 | -2.02 | -4.14 | -2.63 | 0.39 | 0.73 | 0.65 | 0.52 |
| s.e. | 0.27 | 0.41 | 0.67 | 0.78 | 2.89 | 1.83 | 1.04 | 0.65 | 0.42 |
| t-value | -2.08 | -2.44 | -3.00 | -5.30 | -0.91 | 0.21 | 0.71 | 1.01 | 1.22 |
| p-value | **0.04** | **0.02** | **0.00** | **0.00** | 0.36 | 0.83 | 0.48 | 0.32 | 0.23 |
| $R^2$ | 0.10 | 0.11 | 0.12 | 0.15 | 0.04 | 0.00 | 0.02 | 0.03 | 0.05 |

Note. Dependent variable: return on SPY. Independent variable: a quantile of the SPY-implied SPD. An intercept is always included. The standard errors account for heteroscedasticity and autocorrelation. The estimates significant at the 10% level are in bold.

First, consider the benchmark sample period corresponding to panels (a) in these tables. The p-values consistently indicate a statistically significant predictive relationship at the 10% level, except for quantiles near the center of the distribution. Coefficient estimates are negative at lower quantiles and positive at upper quantiles. The regression R-squares are predictably low, but they tend to increase with a longer horizon: for the monthly horizon, the R-squares range

between 1% and 5%, while for the annual horizon, they vary from 1% to 13%. Taken together, these results indicate a significant predictive relationship between the SPD and future returns during this sample period, and the signs of the coefficients are consistent with a risk-expected return trade-off interpretation.

Next, consider panels (b) in these tables which incorporate the COVID pandemic period into the analysis. For the monthly to semi-annual horizons, the results are very similar to the 2009-2020 sample case, consistently showing a significant predictive relationship. For the annual horizon, the point estimates and R-squares are comparable to those in panel (a), but only the lower quantiles ($\tau = 0.2, 0.3, 0.4$) remain statistically significant. Therefore, the evidence of predictability is uncertain in this annual horizon case.

Table 6: Predictive regression using a quantile of the SPD as predictor ($\tau$=365 days)

| Quantile | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 |
|---|---|---|---|---|---|---|---|---|---|
| (a) 06/2009-02/2020 | | | | | | | | | |
| Estimate | -0.49 | -0.65 | -1.23 | -1.39 | -0.96 | 0.58 | 1.41 | 1.11 | 0.77 |
| s.e. | 0.21 | 0.23 | 0.57 | 0.79 | 1.11 | 1.10 | 0.68 | 0.48 | 0.34 |
| t-value | -2.32 | -2.86 | -2.14 | -1.77 | -0.87 | 0.52 | 2.08 | 2.33 | 2.30 |
| p-value | **0.02** | **0.01** | **0.04** | **0.08** | 0.39 | 0.60 | **0.04** | **0.02** | **0.03** |
| $R^2$ | 0.08 | 0.13 | 0.11 | 0.07 | 0.02 | 0.01 | 0.06 | 0.08 | 0.09 |
| (b) 06/2009-02/2023 | | | | | | | | | |
| Estimate | -0.52 | -0.98 | -2.02 | -2.41 | -1.10 | 0.54 | 1.06 | 0.88 | 0.67 |
| s.e. | 0.38 | 0.18 | 0.69 | 0.98 | 1.19 | 1.69 | 1.27 | 0.85 | 0.52 |
| t-value | -1.36 | -5.59 | -2.93 | -2.45 | -0.92 | 0.32 | 0.83 | 1.04 | 1.28 |
| p-value | 0.18 | **0.00** | **0.00** | **0.02** | 0.36 | 0.75 | 0.41 | 0.30 | 0.20 |
| $R^2$ | 0.08 | 0.28 | 0.17 | 0.11 | 0.02 | 0.01 | 0.05 | 0.07 | 0.09 |
| (c) 01/2007-02/2023 | | | | | | | | | |
| Estimate | -0.66 | -0.95 | -1.90 | -3.06 | -3.15 | -0.92 | 0.40 | 0.54 | 0.49 |
| s.e. | 0.25 | 0.14 | 0.37 | 0.76 | 1.38 | 1.62 | 0.99 | 0.58 | 0.34 |
| t-value | -2.65 | -7.02 | -5.12 | -4.01 | -2.29 | -0.57 | 0.40 | 0.93 | 1.46 |
| p-value | **0.01** | **0.00** | **0.00** | **0.00** | **0.03** | 0.57 | 0.69 | 0.35 | 0.15 |
| $R^2$ | 0.13 | 0.24 | 0.19 | 0.19 | 0.14 | 0.02 | 0.01 | 0.03 | 0.05 |

Note. Dependent variable: return on SPY. Independent variable: a quantile of the SPY-implied SPD. An intercept is always included. The standard errors account for heteroscedasticity and autocorrelation. The estimates significant at the 10% level are in bold.

Finally, consider panels (c), which incorporate the 2008 financial crisis into the sample period. Here, the quantiles above the median are consistently insignificant, while those below the median

display varied levels of significance. The evidence of predictability is most pronounced at the semi-annual and annual horizons, where all quantiles below the median are significant, and is least evident at the quarterly horizon, where only the 10th and 40th percentiles are significant. Therefore, the results are mixed. If any conclusion is to be drawn, one might suggest that the lower quantiles appear to be overall more relevant predictors than the upper quantiles for this sample period.

In summary, these regressions suggest the existence of a predictive relationship between quantiles of the SPD and future realized returns on average when the 2008 financial crisis is not included in the sample. This evidence is clear at the monthly to semi-annual horizons and more mixed at the annual horizon. Lower quantiles exhibit stronger predictive power than upper quantiles. Intuitively, a downward shift in the lower quantiles of the SPD is often triggered by a negative market event (e.g., 2008 financial crisis or the onset of the COVID pandemic). Over the past 20 years, the market has generally recovered well from large market declines or crashes. As a result, such declines in lower quantiles signaled an investment opportunity that, on average, yielded higher subsequent returns over short to medium horizons up to one year.

We next turn to predictive quantile regressions. Appendix Tables A.1-A.4 report the results for the four predictive horizons. The results reveal clear patterns for the monthly, quarterly, and semi-annual horizons. Specifically, for quantiles above the median, the estimates are statistically significant with two exceptions only (see panels (a) and (c) in Table A.3), regardless of whether the COVID period or the 2008 financial crisis is included in the sample. For quantiles below the median, the estimates are insignificant in most cases. Lu and Qu (2021) documented this tendency for S&P500 options for a horizon of 30 days. Here, our results show that this finding is robust in the sense that it also holds for SPY options and that it holds for longer horizons up to six months. Finally, for the annual horizon case, the results are mixed: for quantiles above the median, half of the estimates are significant, while the remaining half are insignificant, therefore not leading to clear conclusions.

The results from the mean and quantile regressions have shown that the lower quantiles of the SPD predict future returns on average but may not predict the corresponding quantile of this distribution. These two findings do not contradict each other. As noted earlier, a decline in lower quantiles is typically triggered by a negative market event. Since the market has generally recovered well from significant declines or crashes, the actual market outcomes were often better than what the market had feared. In other words, the lower quantiles of the SPD captured a fear

(or risk premium) that was accompanied by higher rather than lower subsequent returns. As a result, they do not correlate closely with the lower quantiles of the actual return distribution.

Our empirical results contribute to the literature that examines the connection between the options market and subsequent stock returns. Since our methods allow us to extract the SPD from American options, they facilitate the assessment of the predictive power of the SPD for a broader range of markets than those analyzed using European options.

# 6 Conclusion

This paper introduces a new method for estimating state price densities implicit in American-style options. The method involves estimating the parameters of a Gauss-Hermite series expansion and solving recursive equations for the early exercise premium. Because the method does not involve smoothing over time, it can capture sudden shifts in density that might occur during financial crises or in response to influential policy events. It also provides an estimate of the early exercise premium that is of independent interest. We examined the predictive power of the state price density implied by S&P 500 ETF options up to a one-year horizon using both mean and quantile regressions. The mean regressions indicate a significant predictive relationship between quantiles of the SPD and future expected returns when the 2008 financial crisis is excluded from the sample. The quantile regressions show clear patterns for horizons up to six months, where the quantiles above the median consistently demonstrate significant predictive power, irrespective of whether the 2008 financial crisis period is included in the sample. We hope to conduct a more comprehensive analysis of the predictive power of the state price densities across a broad range of asset types in future work.

# References


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.

An, B.-J., A. Ang, T. G. Bali, and N. Cakici (2014): "The joint cross section of stocks and options," *The Journal of Finance*, 69(5), 2279–2337.

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 (2021): "Tail risk and return predictability for the Japanese equity market," *Journal of Econometrics*, 222(1, Part B), 344–363, Annals Issue: Financial Econometrics in the Age of the Digital Economy.

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

Bates, D. S. (2012): "US stock market crash risk, 1926–2010," *Journal of Financial Economics*, 105(2), 229–259.

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.

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

Carlos Dias, J., and J. Pedro Vidal Nunes (2011): "Pricing real options under the constant elasticity of variance diffusion," *Journal of Futures Markets*, 31(3), 230–250.

Carr, P., R. Jarrow, and R. Myneni (1992): "Alternative characterizations of American put options," *Mathematical Finance*, 2(2), 87–106.

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

Cox, J. C. (1996): "The constant elasticity of variance option pricing model," *Journal of Portfolio Management*, 22, 15–17.

Dalderop, J. (2020): "Nonparametric filtering of conditional state-price densities," *Journal of Econometrics*, 214(2), 295–325.

Detemple, J., and W. Tian (2002): "The valuation of American options for a class of diffusion processes," *Management Science*, 48(7), 917–937.

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

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*.

Geman, H., and Y. F. Shih (2008): "Modeling commodity prices under the CEV model," *The Journal of Alternative Investments*, 11(3), 65–84.

Gukhal, C. R. (2001): "Analytical valuation of American options on jump-diffusion processes," *Mathematical Finance*, 11(1), 97–115.

Jacka, S. D. (1991): "Optimal stopping and the American put," *Mathematical Finance*, 1(2), 1–14.

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.

Kim, I. J. (1990): "The analytic valuation of American options," *The Review of Financial Studies*, 3(4), 547–572.

Kim, I. J., and G. G. Yu (1996): "An alternative approach to the valuation of American options and applications," *Review of Derivatives Research*, 1, 61–85.

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

Longstaff, F. A., and E. S. Schwartz (2001): "Valuing American options by simulation: a simple least-squares approach," *The review of financial studies*, 14(1), 113–147.

Lu, J., and Z. Qu (2021): "Sieve estimation of option-implied state price density," *Journal of Econometrics*, 224(1), 88–112.

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(1), 91–115.

Mencia, J., and E. Sentana (2013): "Valuation of VIX derivatives," *Journal of Financial Economics*, 108(2), 367–391.

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

Tian, Y. S. (2011): "Extracting risk-neutral density and its moments from American option prices," *The Journal of Derivatives*, 18(3), 17–34.

Xing, Y., X. Zhang, and R. Zhao (2010): "What does the individual option volatility smirk tell us about future equity returns?," *Journal of Financial and Quantitative Analysis*, 45(3), 641–662.

Xiu, D. (2014): "Hermite polynomial based expansion of European option prices," *Journal of Econometrics*, 179(2), 158–177.

# Supplementary Appendix: Model Details and Additional Tables and Figures

## A.1 Additional details related to the simulation analysis

This subsection provides additional details related to the data generating processes used in the simulation analysis.

**SV Model.** We use the parameter values from the SV panel of Table 1 of Lu and Qu (2021), which are estimated using S&P500 options, with $r = 5\%$, $\delta = 2.5\%$, $\rho = -0.5268$, $S_t = 1300$, $\kappa = 4.2340$, $\theta = 0.0243$, $\sigma = 0.5121$, and $V_t = 0.0243$. The spans of the put and call strikes are both $[900, 1700]$, with 79 call and 79 put options.

**Two-factor SV model.** The parameters are taken from Bates (2000), with $\alpha_1 = 0.028, \alpha_2 = 0.130, \beta_1 = 0, \beta_2 = 5.58, \sigma_1 = 1.039, \sigma_2 = 0.667, \rho_1 = -0.775, \rho_2 = -0.382, V_{1,0} = 0.01, V_{2,0} = 0.01, r = 5\%$, and the initial value of the future price is set to $F_0 = 500$. The spans of the strikes for the put and call options are $[300, 800]$, generating 111 call and 111 put options.

**DMR Model.** The parameter values are taken from Table 4 in Mencia and Sentana (2013) and correspond to an empirical analysis of VIX derivative valuation models. These parameters were estimated using the sample period from March 2006 to August 2008, with $\beta = 2.575$, $\sigma = 3.732$, $\xi = 0.446$, $\alpha = 19.795$, and $\kappa = 1.360$ and the correlation between $W$ and $U$ is 0. The initial values of $V_t$ and $\theta_t$ are set to 25. The risk-free rate is 5%. Under this DGP, the future price of $V_t$ is given by (see Mencia and Sentana, 2013) $F(T) = \alpha + \delta(\tau)[\theta_t - \alpha] + exp(-\beta\tau)[V_t - \theta_t]$, where $\tau = T - t$, and $\delta(\tau) = (\beta/(\beta - \xi))exp(-\xi\tau) - (\xi/(\beta - \xi))exp(-\beta\tau)$. The implied spot price is $S(t) = exp(-r\tau)F(T)$. Using these formulas, we generate the data as follows. First, we simulate the $V_t$ process. For each simulated $V_t$, we compute its future price $F(T)$ and discount it to obtain $S(t)$. As a result, for each simulated $V_t$ process, we have an $S(t)$ process. We repeat this for 1 million times and we obtain one million independent $S(t)$ process. Finally, we use these $S(t)$ as the underlying asset and obtain European and American option prices as we have done for other models, e.g., the SV model. The spans of the strikes for the put and call options are $[10, 100]$, with 51 call options and 51 put options.

**CEV Model.** The parameter values are taken from Geman and Shih (2009) for crude oil for

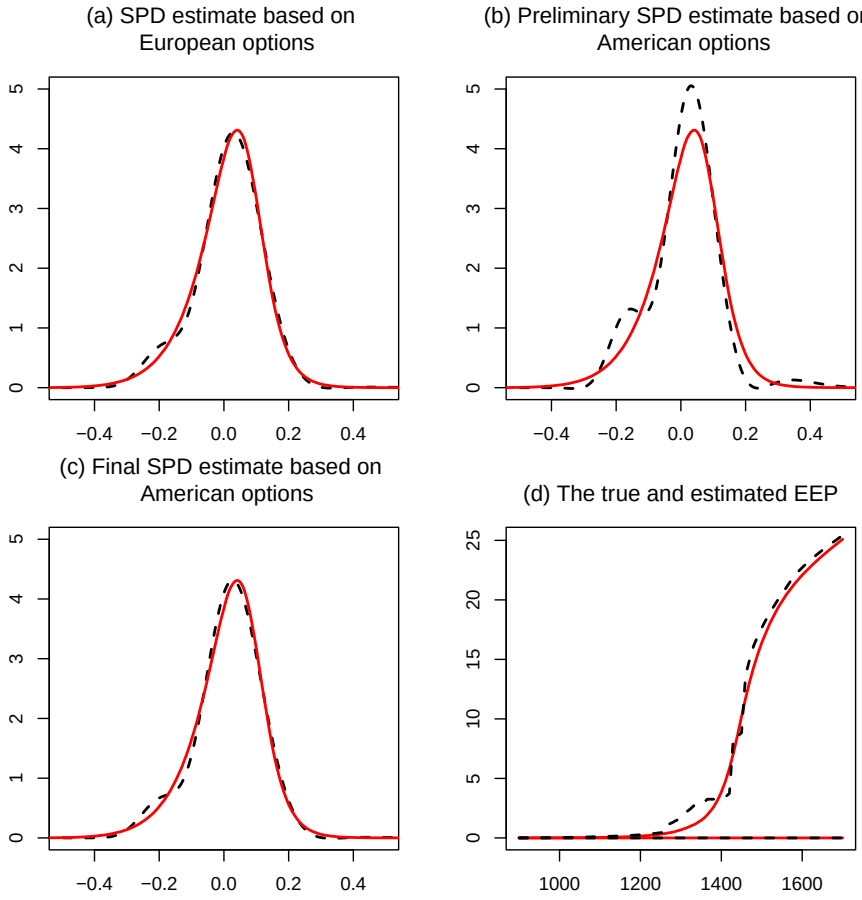
the period 01/01/2000 to 12/11/2007, with $\delta = 0.025, \sigma = 0.19, \beta = 0.68$, and $V_0 = 100$. The spans of call and put option strikes are [10,250], with 57 calls and 57 puts.

**SVCJ Model.** Parameter values are taken from Lu and Qu (2021), estimated using S&P500 index options: $S_0 = 1300, V_0 = 0.0223, \theta = 0.0223, \kappa = 3.285, \sigma = 0.4084, \mu = 5.01e^{-4}, \rho = -0.6108, \lambda = 1.2775, \mu_s = -0.0308$, and $\sigma_s = 0.0248$. The spans of call and put option strikes are [900,1700], with 79 calls and 79 puts.

## A.2 Figures and Tables

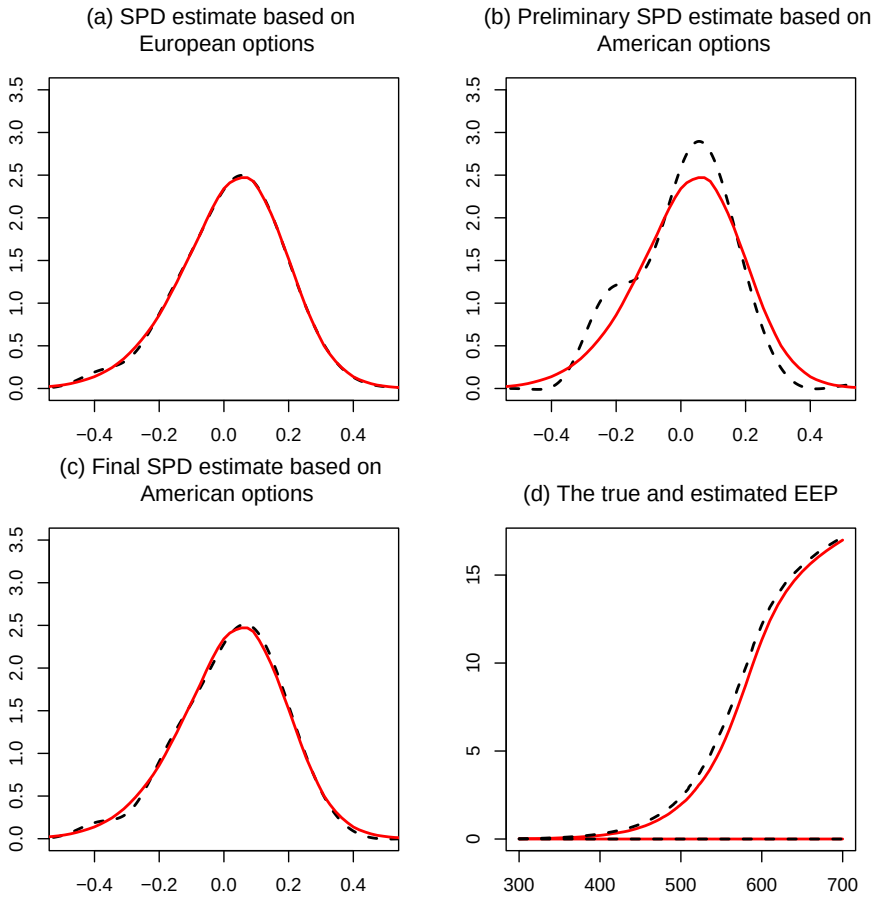
This subsection provides additional tables and figures for the simulation and empirical analyses in the paper.

Figure A.1: SPD and EEP estimates for the SV model



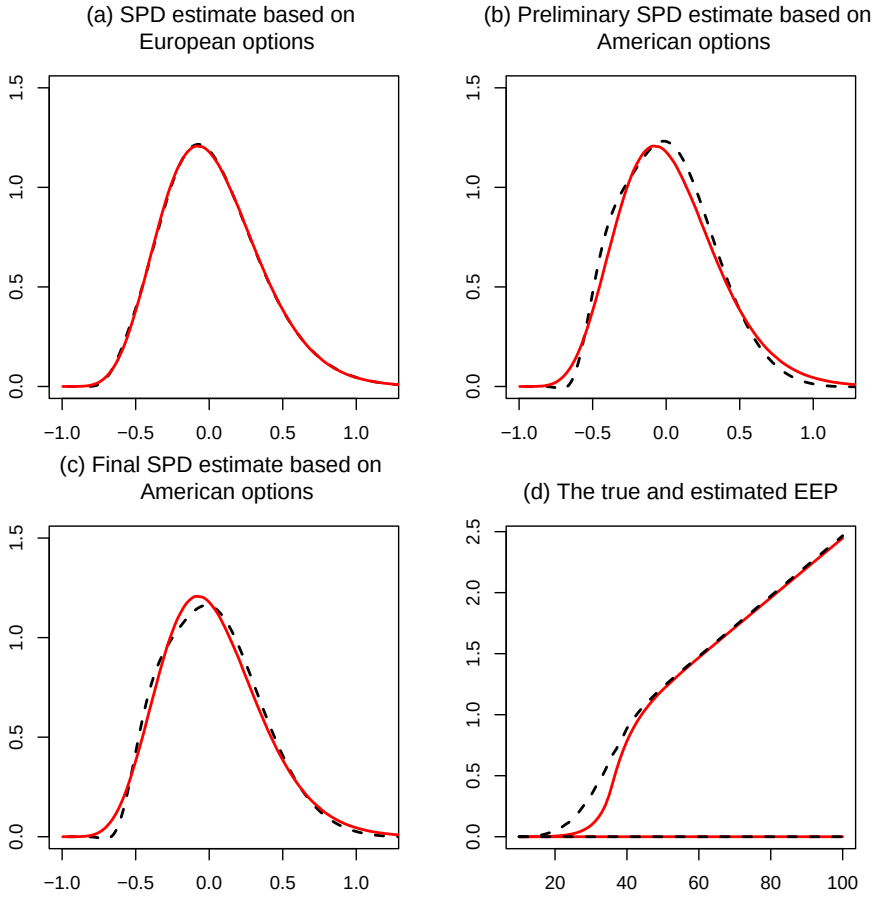

Note. Six-month horizon. Red curve represents the true value, and the dashed line corresponds to the estimate. In (d), the increasing function denotes the put option EEP. See the Appendix for more details about the DGP.

Figure A.2: SPD and EEP estimates for the two-factor SV model



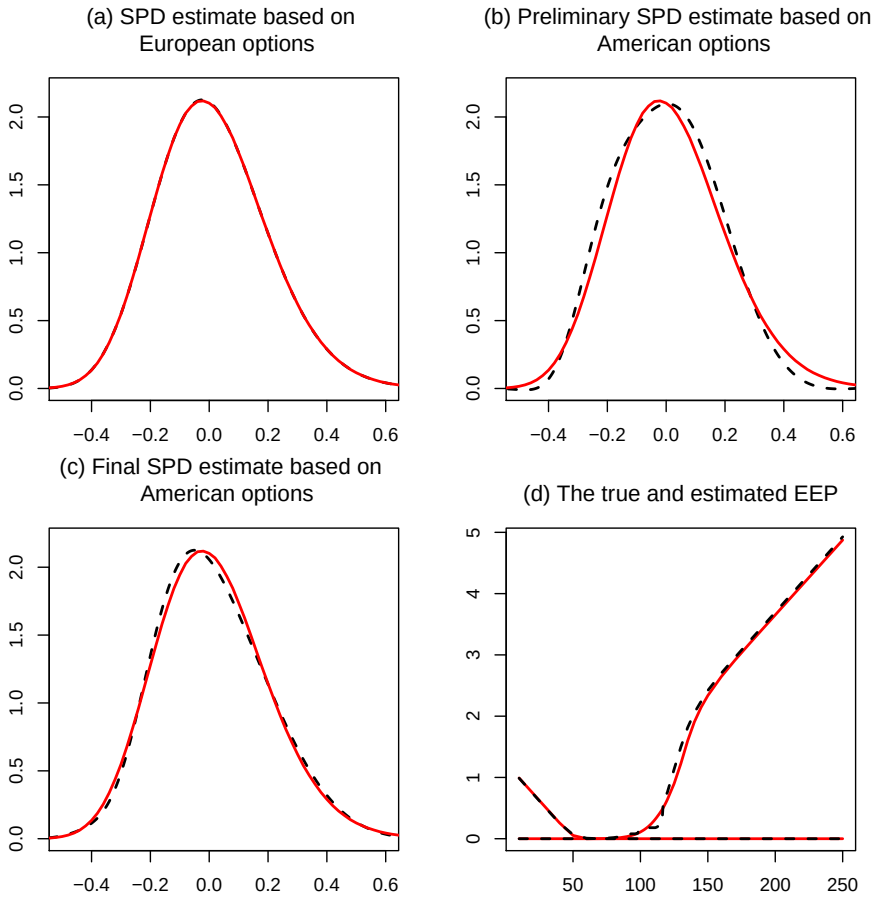

Note. Six-month horizon. Red curve represents the true value, and the dashed line corresponds to the estimate. In (d), the increasing function denotes the put option EEP. See the Appendix for more details about the DGP.

Figure A.3: SPD and EEP estimates for the DMR model



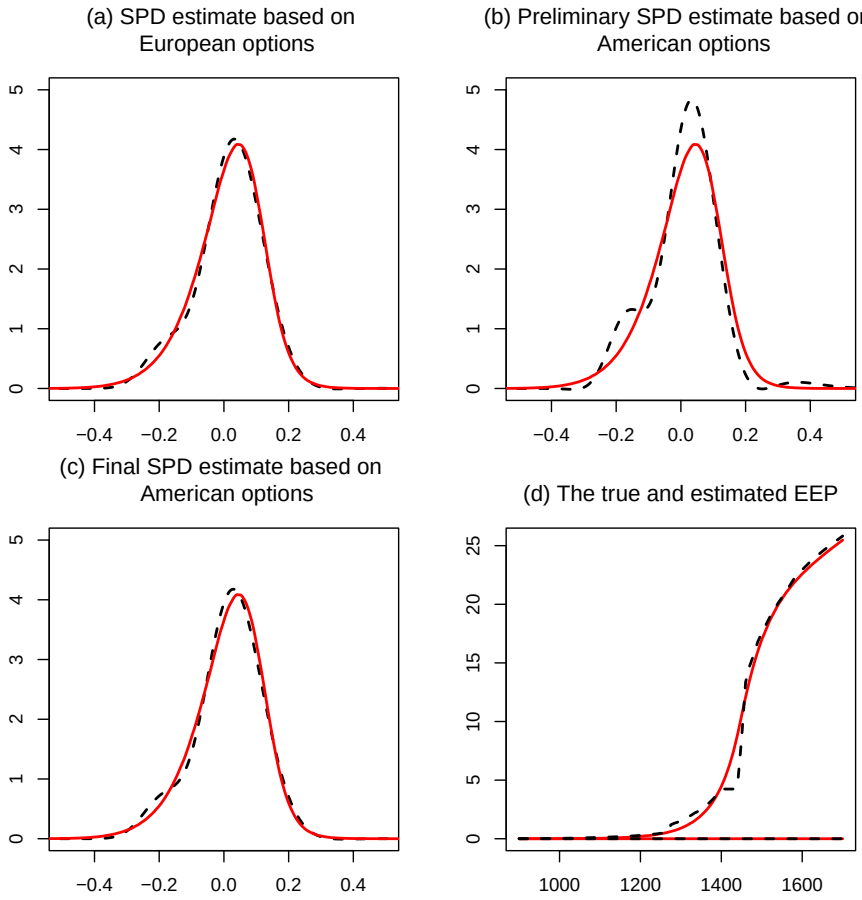

Note. Six-month horizon. Red curve represents the true value, and the dashed line corresponds to the estimate. In (d), the increasing function denotes the put option EEP. See the Appendix for more details about the DGP.

Figure A.4: SPD and EEP estimates for the CEV model




Note. Six-month horizon. Red curve represents the true value, and the dashed line corresponds to the estimate. In (d), the increasing function denotes the put option EEP. See the Appendix for more details about the DGP.

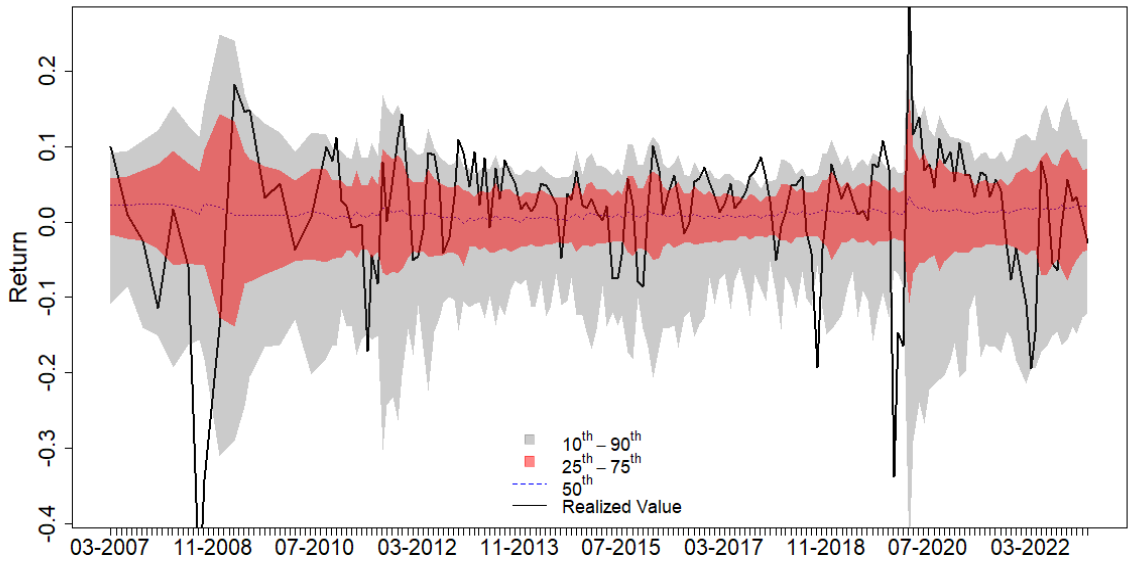
Figure A.5: SPD and EEP estimates for the SVCJ model




Note. Six-month horizon. Red curve represents the true value, and the dashed line corresponds to the estimate. In (d), the increasing function denotes the put option EEP. See the Appendix for more details about the DGP.

Figure A.6: SPD estimates for $\tau = 91$ days

(a) SPY ETF implied densities




(b) SPX implied densities

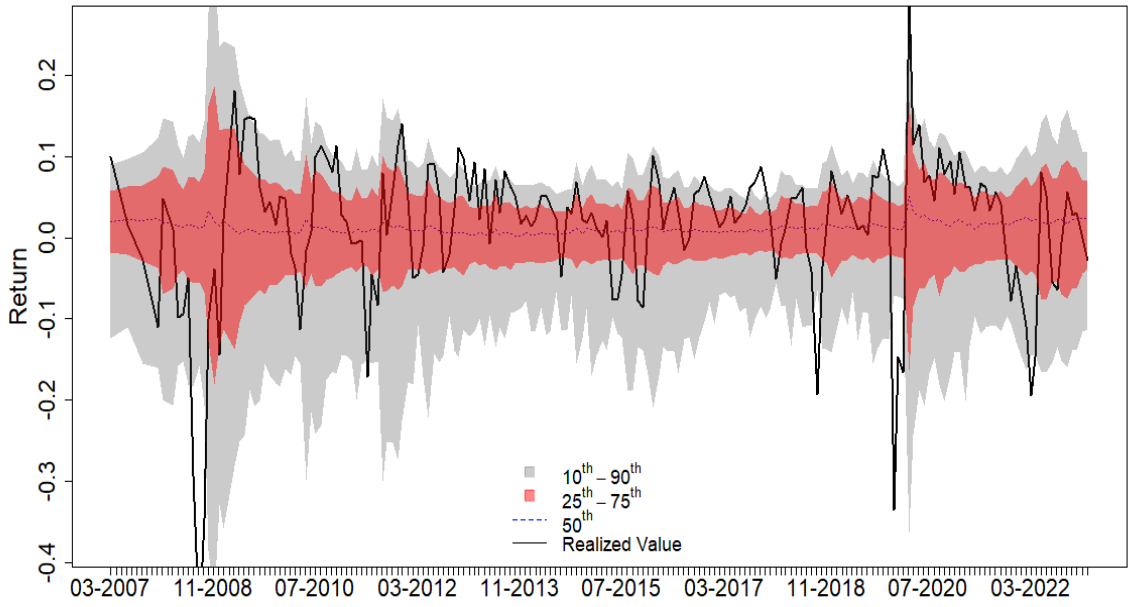
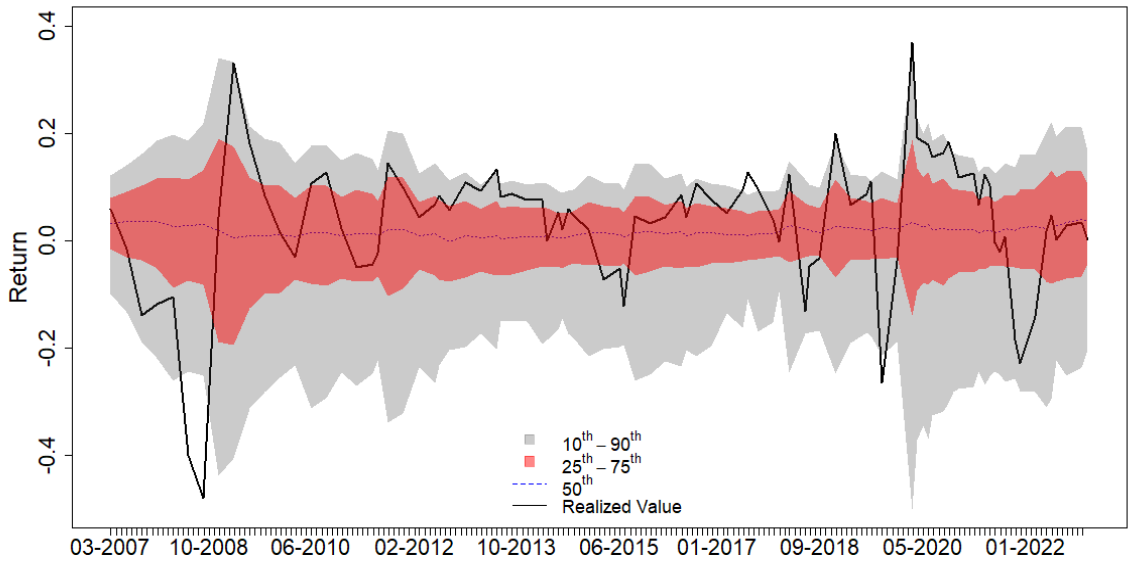


Figure A.7: SPD estimates for $\tau = 182$ days

(a) SPY ETF implied densities




(b) SPX implied densities

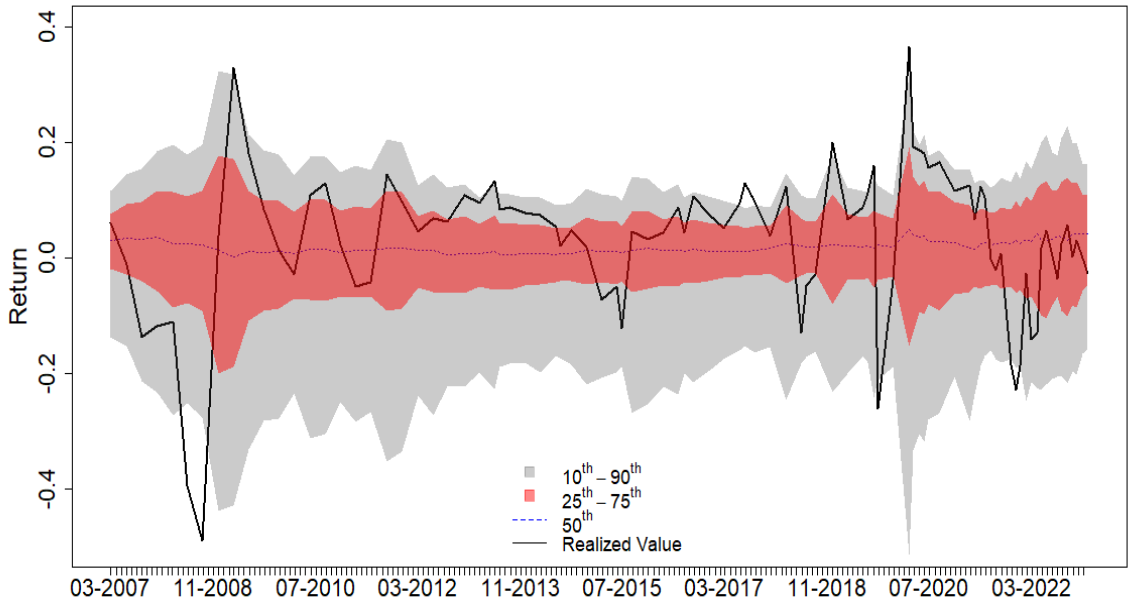


Table A.1: Quantile regression using a quantile of the SPD as the predictor ($\tau$=30 days)

| Quantile | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | (a) 06/2009-02/2020 | | | | |
| Estimate | -0.56 | -0.28 | -0.73 | -0.63 | 1.69 | 1.77 | 1.68 | 1.10 | 0.65 |
| s.e. | 0.20 | 0.67 | 0.76 | 0.79 | 1.02 | 0.66 | 0.35 | 0.06 | 0.11 |
| t-value | -2.83 | -0.41 | -0.96 | -0.80 | 1.66 | 2.70 | 4.83 | 18.82 | 5.94 |
| p-value | **0.00** | 0.68 | 0.34 | 0.43 | **0.10** | **0.01** | **0.00** | **0.00** | **0.00** |
| | | | | | (b) 06/2009-02/2023 | | | | |
| Estimate | -0.61 | -0.36 | -0.84 | -1.06 | 1.43 | 1.77 | 1.57 | 1.13 | 0.78 |
| s.e. | 0.29 | 0.53 | 0.59 | 0.52 | 0.89 | 0.53 | 0.42 | 0.28 | 0.26 |
| t-value | -2.13 | -0.67 | -1.42 | -2.04 | 1.61 | 3.34 | 3.75 | 4.01 | 2.99 |
| p-value | **0.03** | 0.50 | 0.15 | **0.04** | 0.11 | **0.00** | **0.00** | **0.00** | **0.00** |
| | | | | | (c) 01/2007-02/2023 | | | | |
| Estimate | 0.39 | 0.29 | 0.06 | -0.76 | 1.36 | 1.21 | 1.22 | 0.97 | 0.70 |
| s.e. | 0.26 | 0.39 | 0.85 | 0.91 | 0.82 | 0.41 | 0.31 | 0.18 | 0.10 |
| t-value | 1.47 | 0.73 | 0.07 | -0.84 | 1.66 | 2.93 | 3.97 | 5.30 | 7.00 |
| p-value | 0.16 | 0.46 | 0.94 | 0.40 | **0.10** | **0.00** | **0.00** | **0.00** | **0.00** |

Note. The standard errors allow for heteroscedasticity and autocorrelation. The estimates that are significant at the 10% level are in bold.

Table A.2: Quantile regression using a quantile of the SPD as the predictor ($\tau$=91 days)

| Quantile | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | (a) 06/2009-02/2020 | | | | |
| Estimate | -0.64 | -1.03 | -1.32 | -1.55 | 0.68 | 1.66 | 1.58 | 1.12 | 0.96 |
| s.e. | 0.45 | 0.68 | 0.91 | 1.52 | 1.14 | 0.63 | 0.49 | 0.41 | 0.30 |
| t-value | -1.40 | -1.52 | -1.45 | -1.02 | 0.60 | 2.62 | 3.25 | 2.70 | 3.23 |
| p-value | 0.16 | 0.13 | 0.15 | 0.31 | 0.55 | **0.01** | **0.00** | **0.01** | **0.00** |
| | | | | | (b) 06/2009-02/2023 | | | | |
| Estimate | -0.61 | -1.13 | -1.49 | -2.71 | 1.44 | 1.87 | 1.39 | 1.11 | 1.08 |
| s.e. | 0.66 | 0.48 | 0.82 | 1.57 | 1.54 | 0.82 | 0.67 | 0.47 | 0.48 |
| t-value | -0.93 | -2.36 | -1.81 | -1.73 | 0.94 | 2.29 | 2.08 | 2.37 | 2.25 |
| p-value | 0.35 | **0.02** | **0.07** | **0.08** | 0.35 | **0.02** | **0.04** | **0.02** | **0.02** |
| | | | | | (c) 01/2007-02/2023 | | | | |
| Estimate | 0.36 | -0.48 | -1.14 | -2.83 | 0.06 | 1.63 | 1.42 | 1.10 | 0.98 |
| s.e. | 0.45 | 0.92 | 0.86 | 1.00 | 1.32 | 0.78 | 0.44 | 0.42 | 0.30 |
| t-value | 0.80 | -0.52 | -1.33 | -2.83 | 0.05 | 2.10 | 3.21 | 2.62 | 3.33 |
| p-value | 0.42 | 0.60 | 0.18 | **0.00** | 0.96 | **0.04** | **0.00** | **0.01** | **0.00** |

Note. The standard errors allow for heteroscedasticity and autocorrelation. The estimates that are significant at the 10% level are in bold.

Table A.3: Quantile regression using a quantile of the SPD as the predictor ($\tau$=182 days)

| Quantile | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 |
|---|---|---|---|---|---|---|---|---|---|
| | | | | (a) 06/2009-02/2020 | | | | | |
| Estimate | -0.03 | -0.59 | -2.21 | -3.90 | 0.40 | 1.96 | 1.53 | 1.20 | 0.72 |
| s.e. | 0.60 | 0.67 | 1.23 | 1.67 | 2.16 | 1.09 | 0.70 | 0.54 | 0.61 |
| t-value | -0.06 | -0.88 | -1.80 | -2.34 | 0.19 | 1.80 | 2.20 | 2.22 | 1.18 |
| p-value | 0.95 | 0.38 | **0.07** | **0.02** | 0.85 | **0.07** | **0.03** | **0.03** | 0.24 |
| | | | | (b) 06/2009-02/2023 | | | | | |
| Estimate | -0.67 | -1.33 | -2.21 | -4.13 | -0.48 | 2.69 | 1.97 | 1.41 | 1.17 |
| s.e. | 0.73 | 0.78 | 0.73 | 1.31 | 2.28 | 1.62 | 0.63 | 0.38 | 0.30 |
| t-value | -0.93 | -1.70 | -3.04 | -3.14 | -0.21 | 1.66 | 3.13 | 3.75 | 3.91 |
| p-value | 0.35 | **0.09** | **0.00** | **0.00** | 0.83 | **0.10** | **0.00** | **0.00** | **0.00** |
| | | | | (c) 01/2007-02/2023 | | | | | |
| Estimate | -0.31 | -0.43 | -1.99 | -4.27 | -1.14 | 1.73 | 1.96 | 1.57 | 1.05 |
| s.e. | 1.37 | 0.56 | 1.10 | 0.84 | 1.27 | 1.60 | 0.76 | 0.27 | 0.20 |
| t-value | -0.23 | -0.76 | -1.81 | -5.10 | -0.90 | 1.08 | 2.58 | 5.80 | 5.14 |
| p-value | 0.82 | 0.45 | **0.07** | **0.00** | 0.37 | 0.28 | **0.01** | **0.00** | **0.00** |

Note. The standard errors allow for heteroscedasticity and autocorrelation. The estimates that are significant at the 10% level are in bold.

Table A.4: Quantile regression using a quantile of the SPD as the predictor ($\tau$=365 days)

| Quantile | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 |
|---|---|---|---|---|---|---|---|---|---|
| | | | | (a) 06/2009-02/2020 | | | | | |
| Estimate | -0.96 | -1.10 | -1.19 | -0.90 | 0.27 | 0.67 | 1.41 | 0.80 | 0.91 |
| s.e. | 0.44 | 0.34 | 1.11 | 1.46 | 1.76 | 1.18 | 0.90 | 0.64 | 0.61 |
| t-value | -2.16 | -3.25 | -1.07 | -0.62 | 0.15 | 0.57 | 1.57 | 1.24 | 1.50 |
| p-value | **0.03** | **0.00** | 0.28 | 0.54 | 0.88 | 0.57 | 0.12 | 0.21 | 0.13 |
| | | | | (b) 06/2009-02/2023 | | | | | |
| Estimate | -0.80 | -1.17 | -1.43 | -1.18 | 0.67 | 2.38 | 2.75 | 1.83 | 1.07 |
| s.e. | 0.87 | 0.60 | 0.89 | 1.48 | 1.58 | 1.34 | 1.17 | 0.62 | 0.62 |
| t-value | -0.92 | -1.95 | -1.60 | -0.80 | 0.42 | 1.78 | 2.36 | 2.94 | 1.72 |
| p-value | 0.36 | **0.05** | 0.11 | 0.43 | 0.67 | **0.07** | **0.02** | **0.00** | **0.09** |
| | | | | (c) 01/2007-02/2023 | | | | | |
| Estimate | -0.07 | -1.18 | -1.94 | -2.22 | -2.57 | -1.34 | 1.42 | 1.31 | 1.07 |
| s.e. | 0.46 | 0.19 | 0.38 | 0.74 | 1.61 | 1.58 | 0.91 | 0.65 | 0.18 |
| t-value | -0.14 | -6.17 | -5.09 | -2.99 | -1.60 | -0.85 | 1.56 | 2.01 | 5.83 |
| p-value | 0.89 | **0.00** | **0.00** | **0.00** | 0.11 | 0.40 | 0.12 | **0.04** | **0.00** |

Note. The standard errors allow for heteroscedasticity and autocorrelation. The estimates that are significant at the 10% level are in bold.