ELSEVIER

Journal of Econometrics 102 (2001) 67–110

JOURNAL OF Econometrics

www.elsevier.nl/locate/econbase

Do option markets correctly price the probabilities of movement of the underlying asset?

Yacine Aït-Sahalia[a,*], Yubo Wang[b], Francis Yared[c]

[a]*Department of Economics, Princeton University, Princeton, NJ 08544-1021, USA*
[b]*Fixed Income Research, J.P. Morgan Securities Inc., New York, NY 10017, USA*
[c]*Lehman Brothers International (Europe), London EC 2M7HA, UK*

Received 4 January 1999; received in revised form 21 August 2000; accepted 2 October 2000

Abstract

We answer this question by comparing te risk-neutral density estimated in complete markets from cross-section of S&P 500 option prices to the risk-neutral density inferred from the time series density of the S&P 500 index. If investors are risk-averse, the latter density is different from the actual density that could be inferred from the time series of S&P 500 returns. Naturally, the observed asset returns do not follow the risk-neutral dynamics, which are therefore not directly observable. In contrast to the existing literature, we avoid making any assumptions on investors' preferences, by comparing two risk-adjusted densities, rather than a risk-adjusted density from option prices to an unadjusted density from index returns. Our only maintained hypothesis is a one-factor structure for the S&P 500 returns. We propose a new method, based on an empirical Girsanov's change of measure, to identify the risk-neutral density from the observed unadjusted index returns. We design four different tests of the null hypothesis that the S&P 500 options are efficiently priced given the S&P 500 index dynamics, and reject it. By adding a jump component to the index dynamics, we are able to partly reconcile the differences between the index and option-implied risk-neutral densities, and propose a peso-problem interpretation of this evidence.

JEL classification: C14; G13

* Corresponding author. Tel.: + 1-609-258-4015; fax: + 1-609-258-0719.
E-mail address: yacine@princeton.edu (Y. Aït-Sahalia).

0304-4076/01/$ - see front matter
PII: S0304-4076(00)00091-9

Keywords: State-price densities; Risk-neutral densities; Density comparison; Arbitrage relationships; Girsanov's Theorem; Implied volatility smile; Jump risk; Peso problem

1. Introduction

Do investors rationally forecast future stock price distributions when they price options? Provided that markets are dynamically complete, arbitrage arguments tie down option prices to the prices of primary assets. We may therefore expect that option prices will accurately reflect the risk implicit in the stochastic dynamics of their underlying assets – and nothing else. Is this the case empirically, at least within the framework of a given class of models?

An arbitrage-free option pricing model can be reduced to the specification of a density function assigning probabilities to the various possible values of the underlying asset price at the option's expiration. This density function is named the state-price density (SPD), due to its intimate relationship to the prices of Arrow–Debreu contingent claims.[1] A number of econometric methods are now available to infer SPDs from option prices. These methods deliver SPD estimates which either relax the Black–Scholes (1973) and Merton (1973) log-normal assumption in specific directions,[2] or explicitly incorporate the deviations from the Black–Scholes model when estimating the option-implied SPD and pricing other derivative securities.[3] One of the main attractions of the latter methods is that they produce direct estimates of the Arrow–Debreu state prices implicit in the market prices of options. Their main drawback is that they empirically ignore the evidence contained in the observed time series dynamics of the underlying asset – which, in theory, should indeed be redundant information.

The goal of this paper is to examine whether the information contained in the *cross-sectional* option prices and the information contained in the *time series* of underlying asset values are empirically consistent with each other, i.e., whether option prices are rationally determined, under the maintained hypothesis that the only source of risk in the economy is the stochastic nature of the asset price. More concretely, does the S&P 500 option market correctly price the S&P 500 index risk? It is certainly tempting to answer this question simply by comparing features of the SPD implied by S&P 500 option prices to features of the *observable* time series of the underlying asset price.

[1] See Ross (1976), Banz and Miller (1978), Breeden and Litzenberger (1978) and Harrison and Kreps (1979).

[2] See Cox and Ross (1976), Merton (1976), Jarrow and Rudd (1982), Bates (1991), Goldberger (1991), Longstaff (1995) and Bakshi et al. (1997).

[3] See Shimko (1993), Derman and Kani (1994), Dupire (1994), Rubinstein (1994), Stutzer (1996), Jackwerth and Rubinstein (1996), Aït-Sahalia and Lo (1998) and Dumas et al. (1998).

In fact, previous studies in the literature have always compared the observed option data to the observed underlying returns data, i.e., a risk-neutral density to an actual density. Naturally, in the case where all the Black–Scholes assumptions hold, the comparison between the option-implied and the asset-implied SPDs, both log-normal with the same mean, reduces to a comparison of implied volatilities, assumed common across the cross-section of options, to the index's time series realized volatility over the life of the option.[4] More generally, the well-documented presence in option prices of an 'implied volatility smile', whereby out-of-the-money put options are more expensive than at-the-money options, directly translate into a negatively skewed option-implied SPD. At the same time, there is evidence in the S&P 500 time series of time-varying volatility, generally negatively correlated with the index changes, as well as jumps; both of these effects can give rise to a negatively skewed SPD. We argue in this paper that no conclusions regarding the rationality of option prices should be drawn from putting these two pieces of evidence together. The distribution of the underlying asset values that can be inferred from its observed time series, and the SPD implied by option prices, are *not* comparable without assumptions on investors' preferences.

In fact, it is possible to use the exact same data to determine the representative preferences that are implicit in the *joint observations* on option and underlying asset prices.[5] In other words, the extra degree of freedom introduced by preferences can 'reconcile' any set of option prices with the observed time series dynamics of the underlying asset (up to suitable restrictions). Thus it is always possible to conclude that the options are rationally priced from a comparison of the option-implied SPD and the actual density of index returns, when preferences (albeit strange ones) can be found that legitimize the observed option prices. Yet very little is known about aggregate investors' preferences – even within the class of isoelastic utility functions, there is wide disagreement in the literature regarding what constitutes a reasonable value of the coefficient of relative risk aversion: is the coefficient of relative risk aversion equal to 5 or 250?[6] Therefore there is no reason a priori to compare the dynamics of the underlying asset that are implied by the option data to the dynamics implied by the actual time series, unless for some reason one holds a strong prior view on preferences. Consequently, the combination of the facts that the observed S&P 500 returns are negatively skewed and that there exists a persistent implied volatility smile, cannot be used as evidence either in favor or against the rationality of option prices.

---

[4] See Chiras and Manaster (1978), Schmalensee and Trippi (1978), Macbeth and Merville (1979), Rubinstein (1985), Day and Lewis (1988, 1990), Engle and Mustafa (1992), Canina and Figlewski (1993), Lamoureux and Lastrapes (1993), and for a study of the effect of institutional features on these comparisons Harvey and Whaley (1992).

[5] See Aït-Sahalia and Lo (2000) and Jackwerth (2000).

[6] See for example Mehra and Prescott (1985) and Cochrane and Hausen (1992).

In fact, the two SPDs need not even belong to the same parametric family, meaning that comparisons of their variances alone, as would be appropriate under the Black–Scholes assumptions, can lead to type-II errors (if the variances happen to be equal but other features of the densities are distinct) and/or type-I errors (if the SPDs are in fact equal but the variance estimates are inconsistent due to misspecification of the assumed parametric form). For instance, Lo and Wang (1995) give an example of misspecified variance estimates in a simpler context. The difficulty at this point is that we do not observe the data that would be needed to infer directly the time-series SPD, since we only observe the actual realized values of the S&P 500, not its risk-neutral values.

The main novelty in this paper is to show that it is nevertheless possible to use the observed asset prices to infer indirectly the time-series SPD that should be equal to the option-implied cross-sectional SPD. We rely on Girsanov's characterization of the change of measure from the actual density to the SPD: the diffusion function of the asset's dynamics is identical under both measures, only the drift needs to be adjusted. This fact is of course not surprising from a theoretical point of view, but had not been exploited empirically. Our econometric strategy is to start by estimating the diffusion function from the *observed* S&P 500 index returns. Then by Girsanov's Theorem, this is the same diffusion function as the diffusion function of the risk-neutral S&P 500 returns. What remains to be done is to adjust the drift to reflect the risk-neutral rate of return. Fortunately, it turns out that this drift can be inferred from the observed prices of market-traded instruments. The price of an S&P 500 futures contract gives the risk-neutral expected value of the S&P 500 (cash) index at the contract's expiration. From these we obtain the risk-neutral drift rate of the S&P 500 index. Having fully characterized the S&P 500 risk-neutral dynamics, it is then easy to obtain the corresponding time-series SPD. We already discussed how to extract cross-sectional SPDs from option prices. By comparing the two SPDs – cross-sectional and time-series – our methodology allows us to assess whether options are priced rationally, without making assumptions about, or having to estimate, the representative preferences that are embodied in the market prices of options.

We examine, using every option traded on the S&P 500 index between April 1986 and December 1994, whether the rationality condition that the two SPDs be equal holds empirically. Because we construct an exactly identified system with separate identification of the cross-sectional and time-series SPDs, the equality between the two SPDs becomes an *over-identifying* restriction. As such, that restriction is testable. Our approach can be contrasted to the situation where the option-implied SPD is compared to the actual distribution implied by the time-series of asset values: such a system can only be identified through assumptions on preferences. In that case, no implications for market efficiency are left to be tested since all the information contained in the data has already been exploited just to construct the estimators. To obtain over-identifying

restrictions, even more stringent restrictions must be imposed on preferences, which opens the door to a potentially severe joint hypothesis problem.

In addition to escaping the need to specify preferences, we also wish to avoid a second type of joint hypothesis problem: when we compare the two SPDs, we make as few assumptions as possible on the cross-section of options and the dynamics of the underlying asset. For that reason, we rely on nonparametric estimates (see Aït-Sahalia (1996) for a different nonparametric approach to derivative pricing). Our test for rationality is therefore robust to the biases arising from the potential misspecification of the option pricing model, the data-generating process for the asset values, and the specification of investors' preferences. In effect, our only maintained assumption is that markets are dynamically complete,[7] with the aggregate risk solely driven by the variations in the S&P 500 index.

We find empirically that the option-implied SPDs exhibit systematic excessive skewness and kurtosis with respect to the index-implied SPDs and reject the null hypothesis that the S&P 500 options can be efficiently priced within the limited context of a univariate specification with no jumps. This is not just restating that the option data exhibit an implied volatility smile, which is of course well-known by now. Instead this demonstrates that the implied volatility smile is incompatible with the dynamics of the S&P 500 returns – as captured by a univariate diffusion – independently of investors' preferences. We then design out-of-sample trading schemes to exploit the SPD differences and show that they capture superior profits, due to the irrationality of option prices (at least during the period under consideration). We also document that the high Sharpe ratios achieved by these trading schemes demand excessive variation in investors' marginal utilities, in the sense of Hansen and Jagannathan (1991).

In a sense, we have taken the univariate specification as far as it could go by being flexible in every possible dimension (nonparametric on the specification of the dynamic process, no assumptions whatsoever on representative preferences). The first main conclusion of our paper is that this model, even pushed to its most flexible limits, simply cannot explain the joint time series dynamics of S&P 500 returns and cross-sectional properties of S&P 500 option prices. Moving away from the restrictive nature of the univariate diffusion specification, we are able to partly reconcile the differences between the index and option implied SPDs by adding a jump component to the index dynamics. We propose a peso-problem interpretation of this evidence: cross-sectional option prices capture a premium as compensation for the risk of a market crash, but actual realizations of that jump are too infrequent to be consistently observed and reflected in estimates drawn from the time series of asset returns. The prediction of this model is that time series SPD estimates should be insufficiently skewed and leptokurtic

[7] Of course, it might be difficult to justify in practice why option markets enjoy such large trading volumes if they truly were redundant assets. Our approach shares this limitation with the entire literature that relies on market completeness to tie down the prices of different assets.

relative to their cross-sectional counterparts, which is exactly what we found empirically. In future work, we intend to incorporate stochastic volatility in addition to jumps and test the ability of that broader specification to fully reconcile the joint observations on option and asset returns.[8]

The paper is organized as follows. In Section 2 we present a brief review of the no-arbitrage paradigm, propose an estimation strategy for the cross-sectional and time-series SPDs and construct a statistical test of the null hypothesis that the option-implied and index-implied SPDs are identical. We apply our estimators to the S&P 500 options and index prices in Section 3. We conclude in Section 4. Technical assumptions and results are in the Appendix.

2. Cross-sectional vs. time-series SPDs

2.1. Implications of no-arbitrage

Suppose that the uncertainty in the economy is driven by the stochastic values S_T of an underlying asset at a future date T. There exists a riskless cash account which can be used to borrow and lend without restrictions between dates t and $T = t + \tau$ at the instantaneous rate of return $r_{t,\tau}$. During that period, the asset pays dividends continuously at rate $\delta_{t,\tau}$. For simplicity, we take $r_{t,\tau}$ and $\delta_{t,\tau}$ to be nonstochastic. When markets are frictionless, a path-independent derivative security with payoff $\psi(S_T)$ at T can be perfectly replicated by a dynamic trading strategy involving the asset and a riskless cash account, i.e., the derivative is a redundant asset which can be priced by arbitrage.

In order to rule out arbitrage opportunities among the asset, the derivative and the cash account, Harrison and Kreps (1979) showed that the pricing operator mapping payoffs at date T into prices at date t must be linear, continuous and strictly positive. The Riesz representation theorem then characterizes the derivative price as an integral, or expectation operator, applied to the derivative's payoff function. The SPD is the density function $f_t^*(S_t, S_T, \tau, r_{t,\tau}, \delta_{t,\tau})$ to be used in the expectation, i.e.:

$$e^{-r_{t,\tau}\cdot\tau}\int_0^{+\infty}\psi(S_T)f_t^*(S_t, S_T, \tau, r_{t,\tau}, \delta_{t,\tau})\,dS_T \tag{2.1}$$

is the price of a European-style derivative security with a single liquidating payoff $\psi(S_T)$.

If we are willing to be more specific about the nature of the uncertainty in S_T, we can further characterize the SPD. Suppose that the vector S of n_1 asset prices

[8] See the conclusion for a discussion of what the over-identifying restrictions would be in this case.

follows Itô diffusions driven by n_2 independent Brownian motions W:

$$\mathrm{d}S_t = \mu(S_t)\,\mathrm{d}t + \sigma(S_t)\,\mathrm{d}W_t \tag{2.2}$$

with $n_1 \geqslant n_2$.[9] Consider the conditional density that is generated by the dynamics

$$\mathrm{d}S_t^* = (r_{t,\tau} - \delta_{t,\tau})S_t^*\,\mathrm{d}t + \sigma(S_t^*)\,\mathrm{d}W_t^* \tag{2.3}$$

where W^* is a Brownian motion. The transformation from W to W^* and S to S^* is an application of Girsanov's Theorem [see Harrison and Kreps (1979)]. Let $g_t^*(S_t, S_T, \tau, r_{t,\tau}, \delta_{t,\tau})$ denote this conditional density.

Under the assumptions made, the two characterizations of the SPD are identical, i.e., $f^* = g^*$. For example, in the Black–Scholes case where $n_1 = n_2 = 1$ $(r_{t,\tau} - \delta_{t,\tau})$ is constant and $\sigma(S_t^*) = \sigma \cdot S_t^*$ for a constant value of the parameter σ, the risk-neutral density is given by

$$\begin{aligned} f_{\mathrm{BS},t}^*(S_t, S_T, \tau, r_{t,\tau}, \delta_{t,\tau}) &= g_{\mathrm{BS},t}^*(S_t, S_T, \tau, r_{t,\tau}, \delta_{t,\tau}) \\ &= \frac{1}{S_T\sqrt{2\pi\sigma^2\tau}} \exp\left[-\frac{[\ln(\frac{S_T}{S_t}) - (r_{t,\tau} - \delta_{t,\tau} - \frac{\sigma^2}{2})\tau]^2}{2\sigma^2\tau} \right]. \end{aligned} \tag{2.4}$$

If we denote by H the price of a European call option with maturity date T and strike price X, i.e., (2.1) evaluated for the payoff function $\psi(S_T) = \max[S_T - X, 0]$, then Eq. (2.1) reduces to the well-known Black–Scholes formula:

$$H_{\mathrm{BS}}(F_{t,\tau}, X, \tau, r_{t,\tau}; \sigma) = \mathrm{e}^{-r_{t,\tau}\tau}\{F_{t,\tau}\Phi(d_1) - X\Phi(d_2)\} \tag{2.5}$$

where $F_{t,\tau} = S_t \exp((r_{t,\tau} - \delta_{t,\tau})\tau)$ is the forward price for delivery of the underlying asset at date T and

$$d_1 \equiv \frac{\ln(F_{t,\tau}/X) + (\sigma^2/2)\tau}{\sigma\sqrt{\tau}}, \quad d_2 \equiv d_1 - \sigma\sqrt{\tau}. \tag{2.6}$$

[9] In that case, the system of asset prices S in (2.2) supports an SPD if and only if the system of linear equations $\sigma(S_t)\cdot\lambda_t = \mu(S_t)$ admits at every instant a solution λ_t such that

$$E\left[\exp\left[\int_t^T \lambda_\tau \cdot \lambda_\tau\,\mathrm{d}\tau/2\right]\right] < \infty,$$

$$V\left[\exp\left[-\int_t^T \lambda_\tau\,\mathrm{d}W_\tau - \int_t^T \lambda_\tau \cdot \lambda_\tau\,\mathrm{d}\tau/2\right]\right] < \infty.$$

In the presence of an SPD, markets are complete if and only if rank $\sigma(S_t) = n_2$ almost everywhere.

As we discussed in the Introduction, f^* can only be compared to the conditional distribution g^* of S_T given S_t implied by the dynamics (2.3), not to the conditional distribution g implied by (2.2). This paper examines how well the equality $f^* = g^*$ holds in the data, without making the Black–Scholes (1973) assumptions – or for that matter any alternative parametric restrictions – and studies the deviations from equality that arise in the data, and their consequences. In Section 2.2, we will estimate the function f^* from the cross-section of option prices. This step essentially consists in collecting the market prices H of call options, and given that we know their payoff functions ψ, inverting or 'de-convoluting' equation (2.1) to obtain f^*.

In Section 2.3, we will then estimate the time-series SPD g^*. Of course, the risk-neutral path of the index $\{S_t^*\}$ is not observable. Our estimation method is based on the fact, which follows from Girsanov's Theorem, that the instantaneous volatility functions $\sigma(\cdot)$ are identical under both the actual and risk-neutral dynamics. That is, the *same* function $\sigma(\cdot)$ modifies the Brownian shocks in both (2.2) and (2.3). Therefore, we can use the *observable* index values $\{S_t\}$ to devise an estimator of $\sigma(\cdot)$, and then use this estimate, in conjunction with the characterization of the drift $\mu^*(S_t^*) = (r_{t,\tau} - \delta_{t,\tau})S_t^*$. The risk-neutral drift rate $(r_{t,\tau} - \delta_{t,\tau})$ is readily observable from the spot-forward parity relationship

$$r_{t,\tau} - \delta_{t,\tau} = \frac{\ln(F_{t,\tau}/S_t)}{\tau} \tag{2.7}$$

where both $F_{t,\tau}$ and S_t are date-t market prices.

Throughout, we never need to observe the process $\{S_t^*\}$, yet we are estimating its conditional density g^*, not the actual density g of the process $\{S_t\}$. We can then proceed in Section 2.4 to examine whether our nonparametric estimators of f^* and g^* are identical.

2.2. *Cross-sectional inference*: *SPD inferred from option data*

In order to estimate f^* from option prices, we use the nonparametric method of Aït-Sahalia and Lo (1998). This method exploits an insight of Banz and Miller (1978) and Breeden and Litzenberger (1978). Building on Ross's (1976) fundamental realization that options can be combined to create pure Arrow–Debreu state-contingent claims, Banz and Miller (1978) and Breeden and Litzenberger (1978) provide a strategy for obtaining an explicit expression for f^* as a function of H: f^* is the future value of the second derivative of the call option pricing formula H with respect to the option's strike price X. This can be seen either by a direct calculation of the integral in (2.1), or more intuitively by forming a butterfly portfolio with three call options, and letting the interval between the strike prices of these call options shrink to zero.

Based on this characterization, Aït-Sahalia and Lo (1998) take the option-pricing formula H to be an arbitrary nonlinear function of a pre-specified vector

of option characteristics or 'explanatory' variables, $\boldsymbol{Z} \equiv [F_{t,\tau}, X, \tau, r_{t,\tau}]'$, and use kernel regression to construct a nonparametric estimate of the function H. The estimator $\hat{H}$ can then be differentiated twice to produce an estimator of the SPD, according to $f_t^*(\cdot) = \exp(r_{t,\tau}\tau)\partial^2 H(\cdot)/\partial X^2$. In practice, the dimension of the kernel regression can be reduced by using a semiparametric approach. The d-dimensional vector of explanatory variables $\boldsymbol{Z}$ is partitioned into $[\tilde{\boldsymbol{Z}}', F_{t,\tau}, r_{t,\tau}]'$ where $\tilde{\boldsymbol{Z}} \equiv [X/F_{t,\tau}, \tau]'$ contains $\tilde{d} = 2$ regressors. Suppose that the call pricing function is given by the parametric Black–Scholes formula (2.5) except that the implied volatility parameter for that option is a nonparametric function $\sigma(\tilde{\boldsymbol{Z}})$:

$$H(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}) = H_{\mathrm{BS}}(F_{t,\tau}, X, \tau, r_{t,\tau};\ \sigma(X/F_{t,\tau}, \tau)). \tag{2.8}$$

To estimate $\sigma(X/F_{t,\tau}, \tau)$, the Nadaraya–Watson kernel estimator[10] is given by

$$\hat{\sigma}(X/F_{t,\tau}, \tau) = \frac{\sum_{i=1}^{n} k_{X/F}\left(\dfrac{X/F_{t,\tau} - X_i/F_{t_i,\tau_i}}{h_{X/F}}\right) k_\tau\left(\dfrac{\tau - \tau_i}{h_\tau}\right)\sigma_i}{\sum_{i=1}^{n} k_{X/F}\left(\frac{X/F_{t,\tau} - X_i/F_{t_i,\tau_i}}{h_{X/F}}\right) k_\tau\left(\frac{\tau - \tau_i}{h_\tau}\right)} \tag{2.9}$$

where σ_i is the volatility implied by the option price H_i, and the univariate kernel functions $k_{X/F}$ and k_τ and the bandwidth parameters $h_{X/F}$ and h_τ are chosen to optimize the asymptotic properties of the SPD estimator:

$$\hat{f}_t^*(S_t, S_T, \tau, r_{t,\tau}, \delta_{t,\tau}) = \mathrm{e}^{r_{t,\tau}\tau}\left[\frac{\partial^2 \hat{H}(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau})}{\partial X^2}\right]_{|X=S_T} \tag{2.10}$$

where $\hat{H}(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}) = H_{\mathrm{BS}}(F_{t,\tau}, X, \tau, r_{t,\tau}, \delta_{t,\tau};\ \hat{\sigma}(X/F_{t,\tau}, \tau))$.

2.3. *Time-series inference*: *SPD inferred from returns data*

Our time-series estimator of the SPD g^* is based on inferring the conditional density resulting from the risk-neutral evolution (2.3) of the asset price. The essential difficulty is that we do observe the actual index values $\{S_t\}$, rather than the risk-neutral values $\{S_t^*\}$. As we discussed above, we rely crucially on the commonality of the diffusion function under both the actual and risk neutral dynamics of the asset price. We start in Section 2.3.1 by estimating the diffusion function $\sigma^2(\cdot)$ of the asset price dynamics using the actual price series $\{S_t\}$, and then proceed in Section 2.3.2 to use this estimate, in conjunction with the known and observable drift of the risk-neutral dynamics, to estimate the time-series SPD g^*.

[10] Locally polynomial estimators (especially linear, $p = 1$) are a better choice in general than the Nadaraya–Watson estimator ($p = 0$), especially for small sample sizes. In our empirical application, the sample sizes are large enough to alleviate the need to consider locally polynomial estimators. See also Aït-Sahalia and Duarte (1999) for a local polynomial-based estimator that incorporates shape restrictions (such as the monotonicity and convexity of the price function).

2.3.1. Estimation of the asset's diffusion function

To introduce as few unnecessary restrictions as possible on the behavior of the process $\{S_t\}$, we select an estimator of the diffusion function $\sigma^2(\cdot)$ that does not place restrictions on the drift function. We use Florens–Zmirou's (1993) nonparametric version of the minimum contrast estimators (hereafter FZ). FZ provides an unbiased nonparametric estimator for the diffusion coefficient in the above model. The frequency of the data is assumed to be high enough that the drift μ can be unknown and treated as a nuisance parameter. The estimated diffusion at a point is the squared change of observations weighted by the contribution to the local time of these observations. Without loss of generality, set $t = 0$ and $T = 1$, and assume that the process is sampled at the discrete dates $t_i = i/N$. The FZ estimator is based on a discrete approximation of the local time in S of the diffusion function during $[0, t]$,

$$L_t(S) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^t 1_{|S_u - S| < \varepsilon} \, \mathrm{d}u. \tag{2.11}$$

$L_t(S)$ is estimated consistently by

$$L_t(S) = \frac{1}{nh_{\mathrm{FZ}}} \sum_{i=1}^{[Nt]} K\left(\frac{S_i - S}{h_{\mathrm{FZ}}}\right) \tag{2.12}$$

where $[\cdot]$ denotes the integer part, K is a kernel function and h_{FZ} is a bandwidth parameter such that $(Nh_{\mathrm{FZ}})^{-1} \ln(N) \to 0$ and $Nh_{\mathrm{FZ}}^4 \to 0$ as the number of observations N tends to infinity. When the trajectory of the diffusion visits the asset price level S, the natural local estimator of the diffusion function at that level is

$$\hat{\sigma}_{\mathrm{FZ}}^2(S) = \frac{\sum_{i=1}^{N-1} K(\frac{S_i - S}{h_{\mathrm{FZ}}}) N\{S_{(i+1)/N} - S_{i/N}\}^2}{\sum_{i=1}^{N} K(\frac{S_i - S}{h_{\mathrm{FZ}}})} \tag{2.13}$$

which consistently estimates $\sigma^2(S)$ as $N \to \infty$, provided that $Nh_{\mathrm{FZ}}^4 \to 0$. Moreover, if $Nh_{\mathrm{FZ}}^3 \to 0$,

$$N^{1/2} h_{\mathrm{FZ}}^{1/2} \left\{ \frac{\hat{\sigma}_{\mathrm{FZ}}^2(S)}{\sigma^2(S)} - 1 \right\} \xrightarrow{\mathrm{d}} L(S)^{-1/2} \cdot Z \tag{2.14}$$

where Z is a standard Normal variable independent of $L(S)$.

In order to assess the empirical performance of the estimator, we perform Monte-Carlo simulations in conditions that approximate the real data. We simulate sample paths of the asset price process $\{S_t\}$ according to (2.2) with a constant drift rate, $\mu(S) = \mu \cdot S$, and a diffusion function given by

$$\sigma(S) = \alpha + \beta(S - S_0) + \gamma(S - S_0)^2.$$

S_0 is set to 400. The parameter α represents the level of volatility, β the slope of the volatility function and γ its curvature at S_0. We consider various combinations of α, β and γ and obtain a range of shapes for the diffusion function which is capable of capturing the empirical features of the asset's dynamics.

To simulate the continuous-time sample paths, we generate three months of data using the Milstein scheme [see e.g., Kloeden and Platen (1992, Section 10.3)] sampled every 5 minutes, assuming 8 hours of trading per day and 21 trading days per month.[11] We check whether the estimators do capture the shape of the diffusion by considering a flat, downward sloping and upward sloping $\sigma(\cdot)$ in our path simulations. Details on the simulations can be found in Appendix C [see Table 8 and Fig. 11]. The results show that $\hat{\sigma}_{\mathrm{FZ}}(\cdot)$ captures very well the average level, slope and curvature of $\sigma(\cdot)$. Since the estimator has a local character, the diffusion function is estimated more precisely around the initial level S_0 where most of the observations are recorded. The results of our simulations also show that the estimator widely outperforms the standard maximum-likelihood estimator, which assumes that the volatility parameter is constant.

2.3.2. Estimation of the returns SPD

Once the volatility function $\sigma(\cdot)$ of the index dynamics $\{S_t\}$ is estimated, we can make use of Girsanov's characterization of the change of measure from the actual to the risk-neutral dynamics: the drift is set at the difference between the riskless rate and the dividend yield of the index, but the diffusion function is unchanged. We can then recover the time-series SPD g^* using one of two natural methods. We can either solve, for all possible values of S_T, the Fokker–Planck partial differential equation

$$\frac{\partial g^*}{\partial \tau} + (r_{t,\tau} - \delta_{t,\tau})\frac{\partial g^*}{\partial S_T} + \frac{1}{2}\hat{\sigma}^2(S_T)\frac{\partial^2 g^*}{\partial S_T} = 0 \tag{2.15}$$

with the initial condition that $g^*(S, S, 0, r, \delta)$ is a Dirac mass at S, or compute g^* by Monte-Carlo integration.

In our empirical implementation below, we adopt the second method. Given the starting value of the index at the beginning of the period, we simulate $M = 10{,}000$ sample paths of the estimated risk-neutral dynamics (2.3), where we have replaced the diffusion function by its nonparametric estimate:

$$\mathrm{d}S_t^* = (r_{t,\tau} - \delta_{t,\tau})S_t^*\,\mathrm{d}t + \hat{\sigma}_{\mathrm{FZ}}(S_t^*)\,\mathrm{d}W_t^* \tag{2.16}$$

[11] The Milstein scheme is a strong Taylor discrete approximation of the continuous-time sample path which converges strongly with an order 1.0. The traditionally-used Euler scheme converges substantially more slowly: its order of convergence is 0.5.

and use the Milstein scheme. The end points at T of these simulated paths are collected as $\{S_{T,m}: m = 1, \ldots, M\}$, and annualized three-month index log-returns are calculated. We obtain the SPD g^* of the future index values S_T by constructing a nonparametric kernel density estimator $\hat{p}_t^*$ for the index continuously-compounded log-returns u:

$$\hat{p}_t^*(u) = \frac{1}{Mh_{\mathrm{MC}}} \sum_{m=1}^{M} k_{\mathrm{MC}}\left(\frac{u_m - u}{h_{\mathrm{MC}}}\right) \tag{2.17}$$

where u_m is the log-return recorded at the end of the mth sample path. From the density of the continuously compounded log-returns we then have

$$\Pr(S_T \leqslant S) = \Pr(S_t e^u \leqslant S) = \Pr(u \leqslant \log(S/S_t)) = \int_{-\infty}^{\log(S/S_t)} p_t^*(u)\,\mathrm{d}u \tag{2.18}$$

and can recover the price density g^* corresponding to the return density p^* as

$$g_t^*(S) = \frac{\partial}{\partial S}\Pr(S_T \leqslant S) = \frac{p_t^*(\ln(S/S_t))}{S}. \tag{2.19}$$

As we show in Appendix A, both methods result in a nonparametric estimator $\hat{g}^*$ which is $\sqrt{N}$-consistent for g^*, even though the estimator $\hat{\sigma}_{\mathrm{FZ}}(\cdot)$ converges at a speed that is slower than $\sqrt{N}$.

2.4. Testing the no-arbitrage restriction

We have now obtained consistent estimators of both the cross-sectional and time-series SPDs. In this section, we construct a test of the null hypothesis that – as specified by the no-arbitrage theory – they are identical, i.e., $f^* = g^*$. Formally, the null and alternative hypotheses that we consider are

$$\mathrm{H}_0{:}\Pr(f_t^*(S_T) = g_t^*(S_T)) = 1 \quad \text{vs.} \quad \mathrm{H}_A{:}\Pr(f_t^*(S_T) = g_t^*(S_T)) < 1. \tag{2.20}$$

A natural test statistic is the distance between the cross-sectional and time-series SPDs:

$$D(f^*, g^*) \equiv \mathrm{E}[(f_t^*(S_T) - g_t^*(S_T))^2 \omega(\tilde{\boldsymbol{Z}})] \tag{2.21}$$

where $\omega(\tilde{\boldsymbol{Z}})$ is a weighting function. Intuitively, the distance measure D will be large when f^* is far from g^*, leading to a rejection of H_0, and small when the two are sufficiently close together, and the null hypothesis cannot be rejected. The actual test statistic is D evaluated at the two nonparametric estimates $\hat{f}^*$ and $\hat{g}^*$.

How 'far' is far enough to reject H_0? We derive the asymptotic distribution of this test statistic in Appendix B. The use of the L_2 distance metric is a matter of convenience, which makes the derivation of the limiting distribution of the statistic feasible.

In practice, any numerical evaluation of the integral on the right-hand side of (2.21) can be used. We evaluate numerically the integral on a rectangle of values of the vector $\tilde{\boldsymbol{Z}}$ representing a subset of the support of its density π – so $\omega(\tilde{\boldsymbol{Z}})$ is a trimming index – and use the binning method to evaluate the kernels [see, e.g., Wand and Jones (1995) for a description of the binning method].

3. An empirical comparison of S&P 500 implied SPDs

3.1. Estimating the S&P 500 implied SPDs

We collected the entire sample of daily prices of S&P 500 call and put options between April 25, 1986, when the options first became European, and December 31, 1994. These options are traded on the Chicago Board Options Exchange (symbol SPX) calls and puts. For the underlying index, we obtained the time series of S&P 500 Index values from the Chicago Mercantile Exchange (CME) between January 2, 1986 and February 28, 1995. The CME provides a time-stamped high-frequency tick series of S&P 500 index values. The index value is recorded every few seconds.

We focus our empirical analysis on the three-month horizon. Since the goal of this paper is to assess if option markets correctly price the probabilities of movement of the S&P 500 index, we design our estimation procedure for the SPDs to avoid any look-ahead bias while including ample observations for our estimation. We use two weeks of option prices that matures at the same date T to estimate f^*. This is a reasonable compromise between having a snapshot of the cross-sectional SPD at one point in time and having enough observations for our estimation procedure to be sufficiently precise. We make sure that the estimation period for the time-series SPD (g^*) does not overlap with that for the cross-sectional SPD (f^*). To estimate g^*, we use ten weeks of index prices starting immediately after the last used option price and up to the maturity date T. Fig. 1 describes the two estimation subperiods that we use to estimate f^* and g^*, respectively.

The options dataset presents four challenges. First, it contains some implausible entries; second, dividends are not observable; third, S&P 500 futures are traded on the Chicago Mercantile Exchange, and cannot easily be time-stamped synchronously with the options; and fourth, in-the-money options are much less liquid than at and out-of-the-money options, whether calls or puts. We first drop options with implied volatility greater than 70%, or price less than 1/8. As in Aït-Sahalia and Lo (1998), we solve the second problem by relying on the

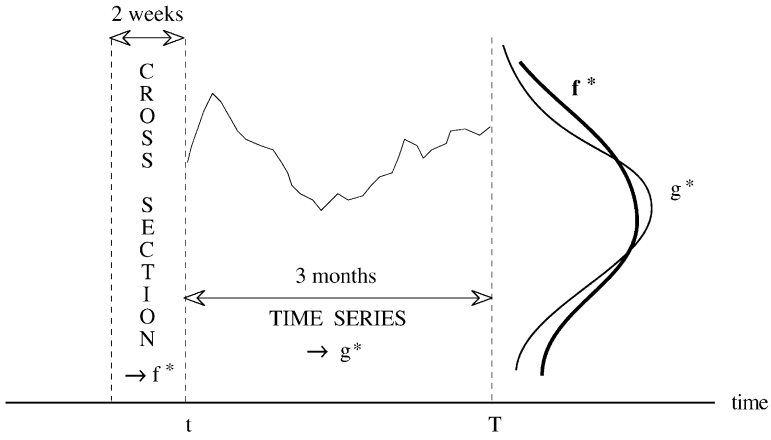


Fig. 1. Estimation periods for SPD comparison.

spot-forward parity relationship (2.7). To solve the third problem, we use prices at-the-money (where both the put and call are liquid) to infer the value of the implied futures according to put–call parity:

$$F_{t,\tau} = X + e^{r_{t,\tau}\tau}\{H(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}) - G(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau})\} \tag{3.1}$$

where G denotes the put price. Note that this equation does not require that the spot price of the index be recorded. Given the futures price $F_{t,\tau}$, we can replace the prices of all illiquid options, i.e., in-the-money options, with the price implied by put–call parity applied at each value of the strike price, using the price of the liquid out-of-money option. This solves the fourth problem. After this procedure, all the information contained in liquid put prices has been extracted and resides in corresponding call prices and we can now concentrate exclusively on call options.

The processed data contains 55 three-month periods starting from June 1986 and ending in December 1994. Until March 1992, the available maturities traded on the CBOE are at quarterly, instead of monthly, frequencies. We apply the estimation method described in Section 2.2 to these 55 periods to obtain the cross-sectional SPD estimator. For each of the 55 periods, we estimate the time-series SPD g^* as described in Section 2.3. We first estimate both the diffusion function $\sigma^2(\cdot)$ and its asymptotic variance using the FZ estimator (2.13). Then we simulate the risk-neutral sample paths to date T and non-parametrically estimate g^*.

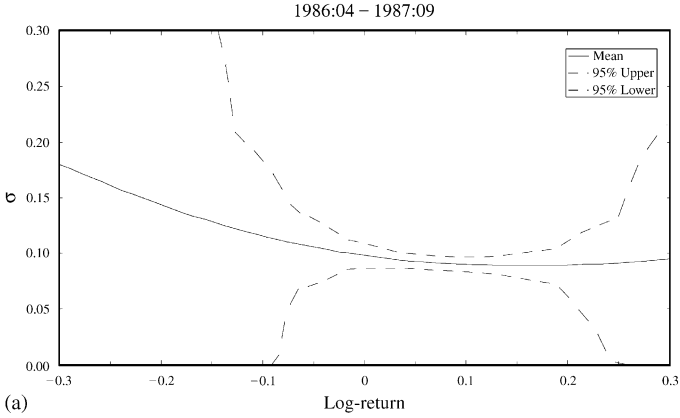
The asymptotic properties of the FZ estimator are derived as the time increment between observations goes to zero. Therefore, in our estimation procedure, we use the high-frequency S&P 500 index data described above. However, to limit the occurrence of market micro-structure effects, we sample

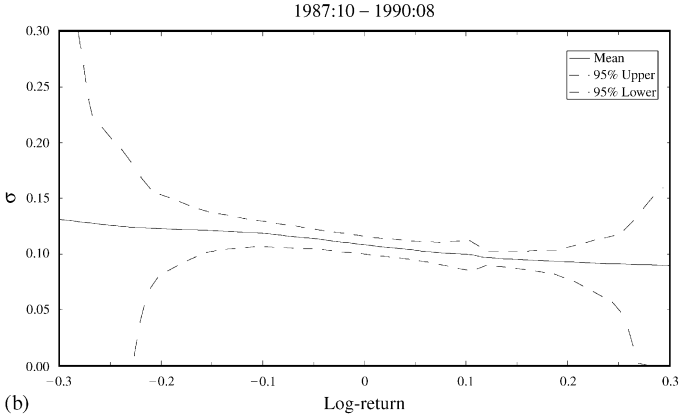
from this time series, each day, the index values every 5 minutes, starting with the first trade of the day. As the Monte Carlo results in Appendix C show, the frequency of the derived series is high enough for our estimator the diffusion function $\sigma^2(\cdot)$ to be very accurate. The Monte-Carlo simulations also provide guidance in selecting the bandwidth in our estimation of the diffusion function. In the simulation for each period, we use the same interest rate and dividend yield as those used in estimating f^*. The process is initialized at index value at the beginning of the period and we use a frequency of eight trades per day. Finally, the estimated diffusion function is linearly extrapolated at both ends of the spectrum to accommodate potential outliers.

Fig. 2 reports the average shape of the estimated function $u \mapsto \hat{\sigma}(S)$, where $u \equiv \ln(S_T/S_t)$ is the log-return corresponding to the possible values of $S_T = S$, for the three subperiods: (a) 1986:07–1987:09 (pre-crash, 5 periods); (b) 1987:10–1990:08 (12 periods); (c) 1990:09–1994:06 (post-crash, 38 periods). The average estimated diffusion function is obtained by averaging point by point each period's estimated diffusion functions. The average estimated variance is obtained in the same way. A 95% confidence interval is constructed by adding to or subtracting from the average diffusion function two times the average estimated standard deviation point by point. As is the case with local non-parametric estimation methods, the confidence interval widens rapidly as we move away from the range of observed values of the regressor. The fact in Fig. 2 that volatility becomes higher when the S&P 500 index goes down in value is compatible with Black's leverage effect explanation. Note that the overall *level* of volatility is substantially lower during the post-1990:09 period, the last substantial market downturn in our sample. Overall, these curves exhibit a significant amount of nonlinearity, which suggests that the estimated time-series SPD g^* will be different from the benchmark log-normal case given in (2.4), and that the choice of an estimator robust to departures from the standard assumptions is empirically warranted.

In the absence of arbitrage, the futures price is the expected future value of the spot price under the risk-neutral measure. Since the riskfree rate and the dividend yield are not directly observed, we make the conservative assumption that any mismatch between the means of the two SPDs is due to measurement error. We therefore translate both distributions so that their means match the implied futures price by construction. We then choose the bandwidth used in the kernel estimation of g^* to best match the variance of the two SPDs. In order to avoid over- or under-smoothing of g^*, we constrain the bandwidth to be within reasonable values of 0.5 to 5 times the asymptotic optimum derived in Appendix A, and we stop the bandwidth search whenever the variance of g^* is within 5% of the variance of f^*. This procedure allows us to focus our comparison on the conditional skewness and kurtosis of the two SPDs.

In Fig. 3, we plot for the same three subperiods (a)–(c) the average densities f^* and g^* over the subperiod. In (b), we see that the time-series SPD is actually





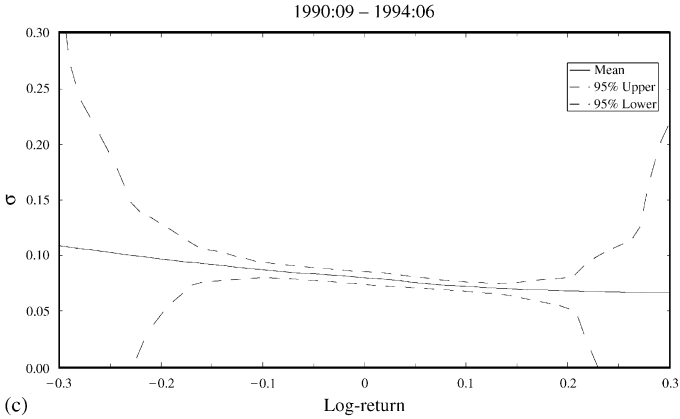
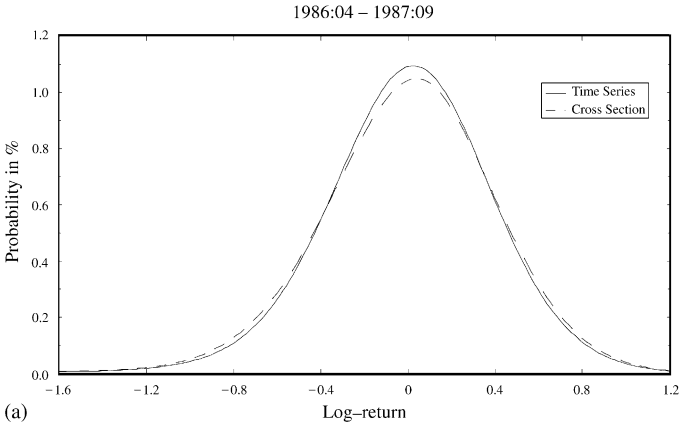
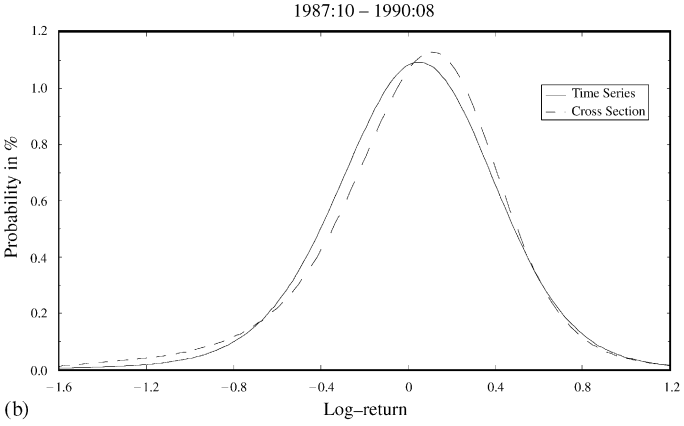


Fig. 2. Average estimated σ function.





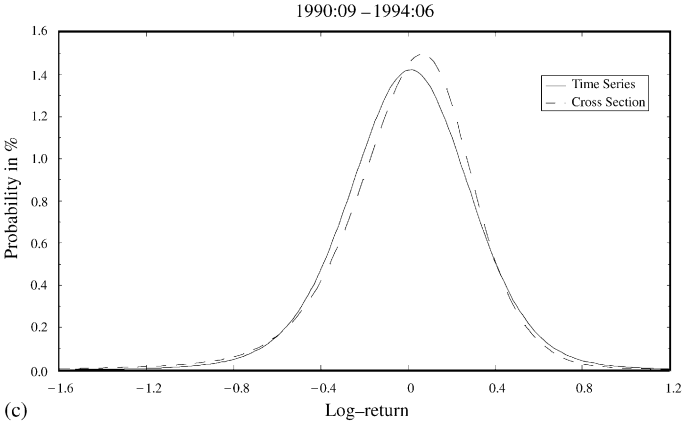


Fig. 3. Average state price densities.

more negatively skewed than the cross-sectional SPD, due to the October 1987 market crash. The fact that from that point on f^* will become more skewed (and more leptokurtic) than g^* is obviously consistent with the emergence of the implied volatility smile after the market crash. More importantly, however, the differences in skewness and kurtosis show that the options market has adjusted to pricing options according to a cross-sectional SPD that reflects negative skewness and excess kurtosis, but to a level that over-amplifies their presence in the time-series SPD. We now examine this last point in greater detail.

3.2. Comparing the cross-sectional and time-series SPDs

If investors had perfect foresight and knew the process governing the underlying asset price dynamics, then by no arbitrage, f^* and g^* should be equal. If this assumption is too unrealistic – after all, the econometrician does not know g^* before the end of the estimation period, recall Fig. 1 – then we cannot expect all differences between the two SPDs in any given period to be traded away by arbitrageurs. However, we can still expect rational nonsatiated investors to take advantage of any systematic differences between the two SPDs that keep arising over time. In much the same way, the CAPM is derived under the assumption that investors know the variance covariance matrix of the asset returns; hence in practice market efficiency conclusions can only be drawn from repeated deviations from the expected return relation. Therefore our comparisons of the two SPDs will focus only on the systematic differences between the two implied SPDs.

We document these differences in four different ways. First, we focus on the first four conditional moments of the two SPDs. We start by regressing the conditional moments implied by f^* on the corresponding moments implied by g^* for the same period. If the moments estimated from option prices correctly forecast the corresponding moments of the S&P 500 index through their respective SPDs, the intercept in the regression should be 0 and the slope should be 1. Results in Table 2 indicate that the cross-sectional SPD f^* systematically

Table 1
Bandwidth values for the SPD estimators[a]

Estimator	Kernel	Sample size	q	p	m	d	h
X/F in $\hat{f}^*$	$k_{(2)}$	$n = 2{,}520$	2	5	2	2	$h_{X/F}$
τ in $\hat{f}^*$	$k_{(2)}$	$n = 2{,}520$	2	5	0	2	h_τ
$\hat{\sigma}_{FZ}$ in $\hat{g}^*$	$k_{(2)}$	$N = 63$	2	3	0	1	h_{FZ}
$\hat{g}^*$	$k_{(2)}$	$M = 10{,}000$	2	3	0	1	h_{MC}

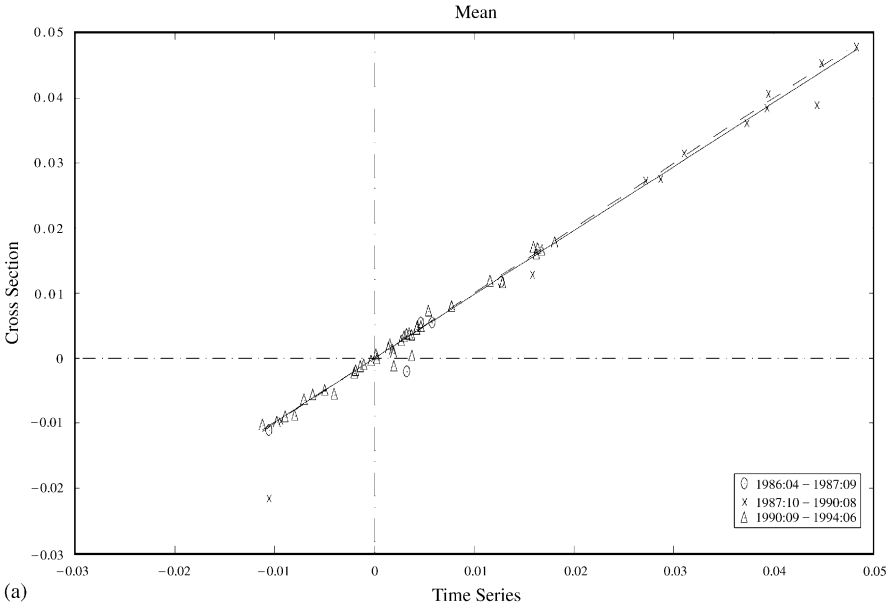
[a]Bandwidth selection for the SPD estimators, according to the selection rules given in Appendix A. All values in the table are averaged over the 52 three-month subperiods between April 1986 and June 1996.

Table 2
Predictability of cross-sectional SPD moments using the time-series SPD moments[a]

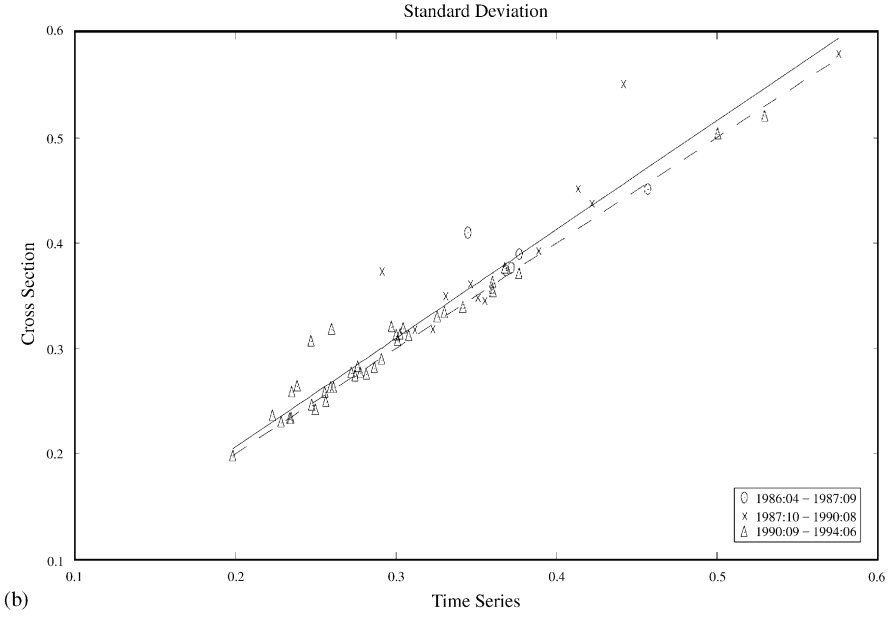
Moment	Intercept (*t*-statistic)	Slope (*t*-statistic)	R^2
Mean	− 0.0006	1.002	0.98
	(− 0.29)	(7.82)	
Volatility	0.0075	1.009	0.92
	(0.076)	(3.39)	
Skewness	− 0.51	− 0.27	0.01
	(− 0.91)	(− 0.11)	
Kurtosis	0.87	− 0.83	0.06
	(0.45)	(− 0.24)	

[a]OLS regression of the moments of the cross-sectional SPD on the corresponding moments of the time-series SPD. *t*-statistics are in parentheses. If the time-series SPD moment constitutes the true expected value of the cross-sectional SPD moment, then the intercept should be zero and the slope one. These values should hold no matter how large the relevant information set used to conduct the forecasts is, since the forecast error must be orthogonal to the forecast itself. Using a larger information set would simply result in a higher regression R^2. The sample for each regression is the 52 three-month subperiods used throughout the analysis. Five outliers (for both the skewness and kurtosis analysis) are excluded from the regressions: they correspond to the periods starting in January 1987, October 1989, August 1992, September 1992, and December 1994. The results show very clearly that the cross-sectional and time-series SPDs agree on the first two moments, but that the option-implied SPD does not reproduce the higher-order moments of the index-implied SPD. The regression is represented by the solid line on Figs. 4 and 5 with the intercept constrained to be zero, for comparison with the 45° line (dashed line). As the graph shows, the cross-sectional SPD tends to be more negatively skewed than the time-series SPD, and more leptokurtic. These results can be compared to the case where we allow for a peso-problem effect (see Fig. 6) where the fit between cross-sectional and time-series SPDs is vastly improved. Finally, these deviations are not entirely due to estimation error: they are significant on the basis of Monte-Carlo simulations (see Appendix D).

displays more skewness and kurtosis than does the time-series SPD g^*. Fig. 4 shows that the two SPDs do not exhibit significant differences in their conditional means and variances respectively given our bandwidth-selection procedure. However, in Fig. 5 the comparison of the regression (without constant) line with the 45° line (dashed line) underlines the excess skewness and kurtosis present in the options data relative to the time-series index observations. In light of the remarks above, we are interested in documenting in these figures the extent to which the data points are unevenly distributed around the 45° line, not in their particular distance to that line: how far each point is from the line only indicates the degree to which investors could forecast the particular g^* for that period, and any unexpected shock occurring in that period could have caused the mismatch of the two SPDs. However, any uneven distribution of the scatterplot signals that the conditional moments from f^* either systematically underestimate or systematically overestimate their counterparts from g^*.



(a)



(b)

Fig. 4. Moment comparison.

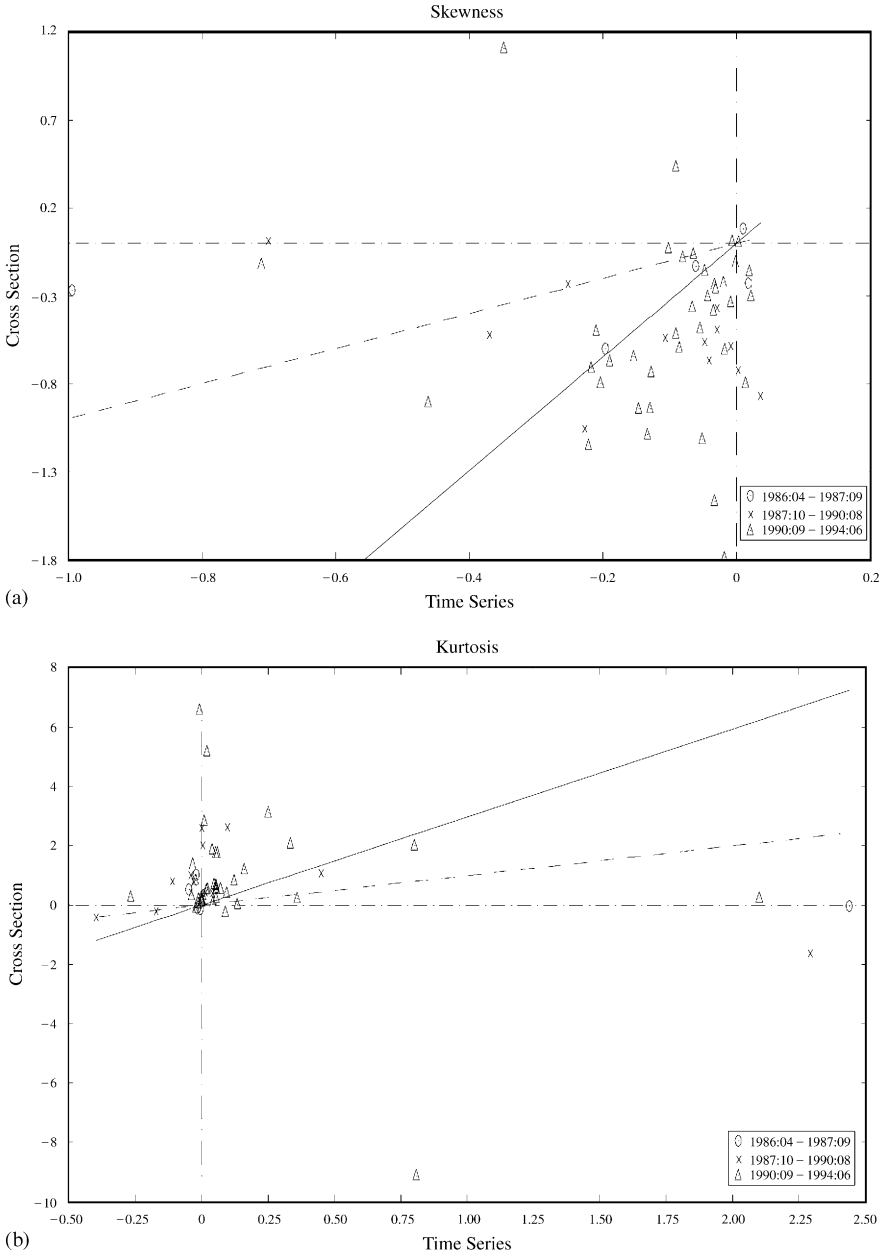


Fig. 5. Moment comparison.

We include in Appendix D Monte-Carlo simulations of these conditional moment regressions in a setup that reproduces the main features of the real data, including sample size, number of subperiods and general level of noise. The results of these simulations [see Table 9] show that the deviations between the cross-sectional and time-series conditional skewnesses and kurtoses reported in Table 2 are significant not just asymptotically but also in small samples.

As a second measure of the differences between the two SPDs, we provide a formal statistical test of the null hypothesis that the two conditional densities are identical. The test introduced in Section 2.4 focuses on whether the option-implied risk-neutral distribution of future returns, as assessed by the market, is equal to the true time-series distribution, also risk-neutral, conditional on all the information available. In other words, we test whether the market efficiently prices options according to Fama (1976)'s definition. We describe the test in Appendix B and report the results in Table 3. Because of its nonparametric nature, this test is consistent against all possible departures from the null hypothesis, but suffers from low power against any specific departure – in this case, the failure to match the third and fourth conditional moments of the densities. The average p-value for the test is 0.7, indicating that on average the test fails to reject the null hypothesis that the two SPDs are identical against an unspecified nonparametric alternative. This is largely due to the small number of observations included in each two-week period [recall Fig. 1], which limits the ability of a global nonparametric test to distinguish the two densities. Later, in Section 3.6, we propose yet another measure of the differences between the two SPDs based this time on comparing the out-of-the-money put prices implied by the two densities.

Table 3
Nonparametric test of SPD equality[a]

Integral (Average)	Test statistic (Average)	p-value (Average)
0.937	− 0.539	0.704

[a]Nonparametric tests of the null hypothesis of equality of the cross-sectional and time-series SPDs. The test is repeated for each of the 52 subperiods. The bandwidth selection rules are given in Appendix B where $\delta_{X/F}$ and $\eta_{X/F}$ are defined. The kernel function is $k_{(2)}$. The weighting function ω is a trimming index, i.e., only observations with estimated density above a certain level, and away from the boundaries of the integration space, are retained. 'Average integral' refers to the percentage of the estimated density mass on the integration space (averaged over the 52 subperiods) that is kept by the trimming index, i.e., $\int \pi(\tilde{Z})\omega(\tilde{Z})\,\mathrm{d}\tilde{Z}$, where π is the marginal density of the nonparametric regressor $\tilde{Z} = [X/F_{t,\tau}]'$. 'Average test statistic' refers to the standardized distance measure between the two SPD estimates (remove the bias term, divide by the standard deviation), again averaged over the 52 subperiods. The average p-value is similarly calculated by averaging the p-values obtained for each subperiod. The integral defining D is calculated over the interval [0.85, 1.10] in the moneyness space.

3.3. The trading profitability of SPD differences

Our third measure of the differences between the two SPDs consists in recording the trading profits that would have resulted from exploiting optimally their discrepancies. So far, we have identified statistically significant differences between the cross-sectional SPD f^* and the time-series SPD g^* when comparing their conditional moments. We now propose to measure the differences between the two SPDs using as a metric the profitability of trading from their differences. In Section 3.2, we were chiefly interested in whether the option market correctly assesses at date t the time-series SPD that will prevail over the life of the option, i.e., in the period immediately following the option price observation: recall Fig. 1. By contrast, trading decisions must obviously be made based on date t, not future, information and hence we must estimate both f^* and g^* with past data. We describe in Fig. 8 the estimation subperiods used at each point in time for forming our trading portfolio. The cross-sectional SPD f^* is estimated using option data within a window of two weeks following the latest expiration date, whereas the time-series SPD g^* is estimated for a three-month period before that date. Option quotes on the business day immediately following the cross-sectional SPD estimation subperiod are used as the input of our trading strategy. We only take positions in out-of-the-money (OTM) and at-the-money (ATM) puts and calls, since these are more competitively priced due to their liquidity. Thus our trading profits are conservative in the sense that we have restricted the investment opportunity set.

In Fig. 9, we indicate how to design such a trading strategy. Intuitively, we are buying the range of strike prices which are underpriced in the sense of $f^* < g^*$ (prices, which are determined by f^*, are cheaper than justified by the time series g^*) and selling options with payoffs in the ranges that are overpriced: $f^* > g^*$ (prices more expensive than justified by the time series). Specifically, we take long and short positions in options according to the rules described in Table 4.

In constructing our trading portfolio at time t, we value-weight each option in the portfolio by their cash outflow. The cash outflow of a long position is the cost of the option itself; that of a short position is its margin deposit. We can then measure the performance of our trading portfolio by its return. We define the rate of return of the trading portfolio as

$$\text{return} = (\text{total inflow at expiration})/(\text{total outflow at initiation}) - 1. \quad (3.2)$$

For simplicity, we assume that the margin deposit does not earn interest and short sale proceeds cannot be used to finance purchases. Based on our prior that the index option SPD is overly skewed and leptokurtic, we call 'skewness trades' trades that are triggered by the rule (S1) only and 'kurtosis trades' trades triggered by rule (K1) only.

We report the results of these trading strategies in Table 5. It is clear from the table that both trading strategies would have provided superior Sharpe ratios

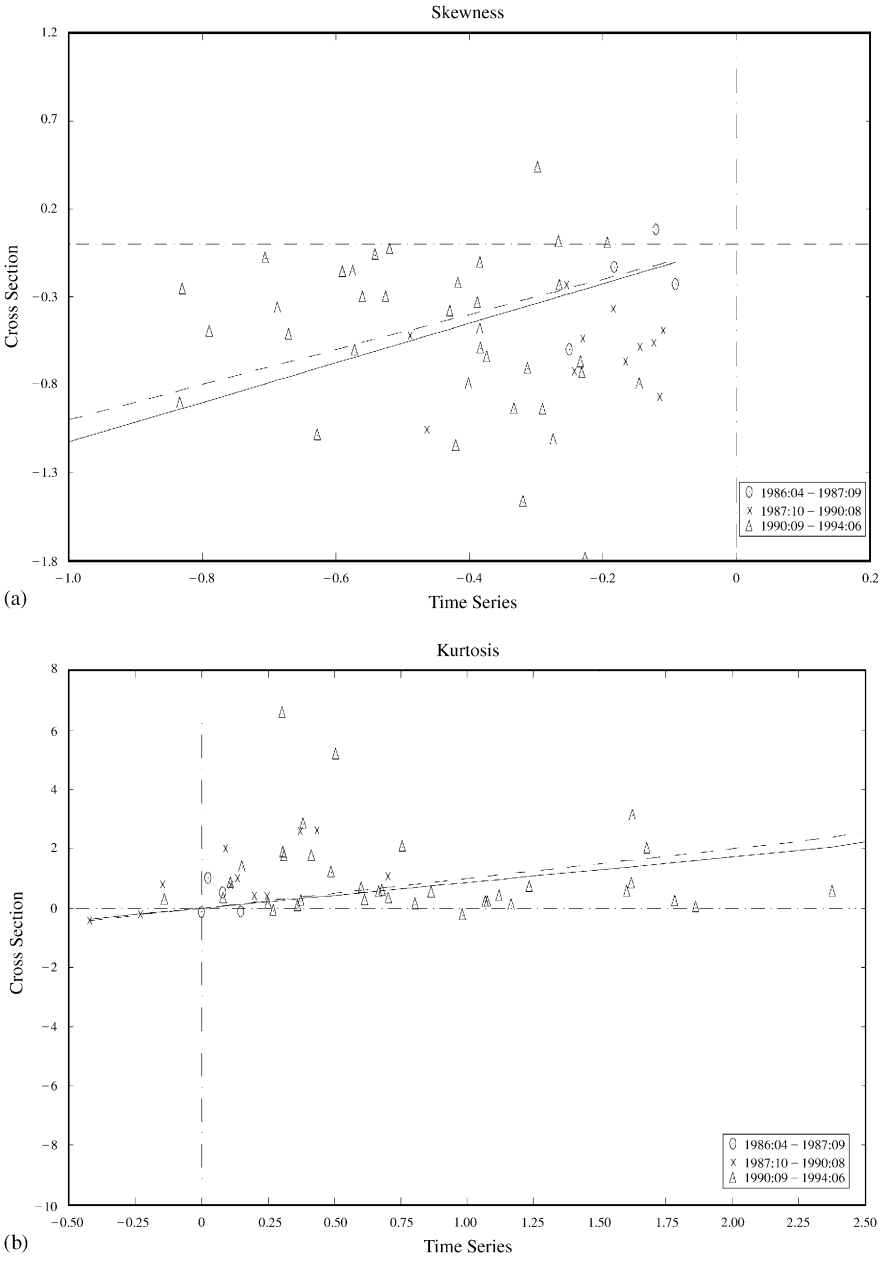


Fig. 6. Moment comparison with jumps.

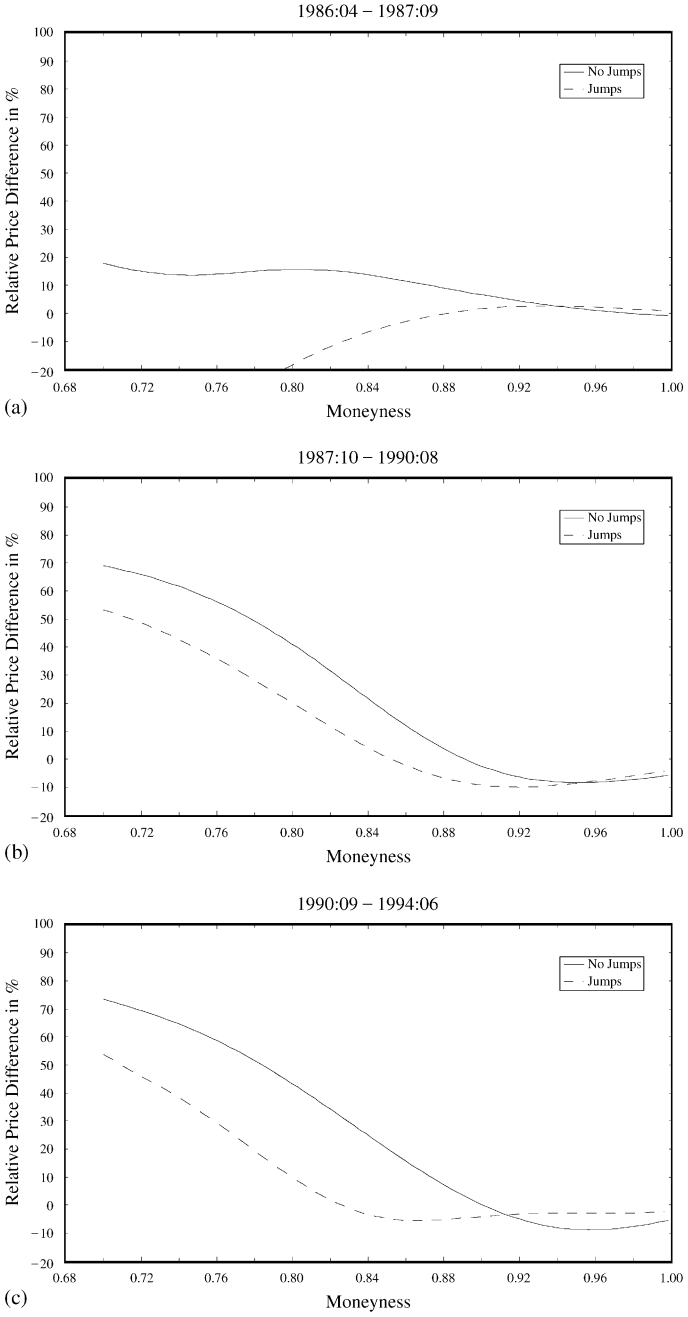


Fig. 7. Average relative price difference for out-of-the-money put.

Table 4
Trading rules to exploit SPD differences[a]

Skewness	(S1)	skew(f^*) > skew(g^*)	Sell OTM put, Buy OTM call
Trade	(S2)	skew(f^*) < skew(g^*)	Buy OTM put, Sell OTM call
Kurtosis	(K1)	kurt(f^*) > kurt(g^*)	Sell far OTM and ATM, Buy near OTM options
Trade	(K2)	kurt(f^*) < kurt(g^*)	Buy far OTM and ATM, Sell near OTM options

[a]This table describes the trading rules designed to exploit the differences between the cross-sectional and time-series SPDs. A far OTM call (put) is defined as one whose strike price is 10% higher (lower) than the futures price. A near OTM call (put) is defined as one whose strike price is 5% higher (lower) but 10% lower (higher) than the futures price. The CBOE imposes specific margin requirements on the S&P 500 index options. Uncovered writers must deposit 100% of the option proceeds plus 15% of the aggregate contract value (current index level multiplied by \$100) minus the amount by which the option is out-of-the money, if any. The minimum margin is 100% of the option proceeds plus 10% of the aggregate contract value. Long puts or calls must be paid in full.

Table 5
The profitability of trading SPD differences[a]

(S1)	Skewness		Trade		
Period	Return	Volatility	Range	Sharpe ratio	S&P 500 Sharpe ratio
86:04–87:09	7.49	0.94	[7, 8]	1.8395	− 0.0895
87:10–89:12	16.30	15.38	[− 7, 47]	0.5680	0.3478
90:01–96:06	6.18	13.16	[− 38, 21]	0.1346	0.1487
Overall	8.79	13.79	[− 38, 47]	0.2545	0.1106
(K1)	Kurtosis		Trade		
Period	Return	Volatility	Range	Sharpe ratio	S&P 500 Sharpe ratio
86:04–87:09	− 23.84	99.05	[− 138, 41]	− 0.2985	− 0.0895
87:10–89:12	17.58	21.38	[− 18, 36]	0.4755	0.3478
90:01–96:06	21.50	24.94	[− 44, 51]	0.6867	0.1487
Overall	16.51	36.39	[− 138, 51]	0.3145	0.1106

[a]This table reports the profit recorded from trading according to one of the two strategies (S1) or (K1) is measured as an annualized rate of return. The trading strategies are described in Fig. 7. The returns and their volatilities are all annualized. All numbers except Sharpe ratios are percentages.

(compared to buying and holding the S&P 500 index) over the 1986–1996 period, as well as over the same three subperiods that we considered above. Given that the trading signal is based on the current option and the lagged index observations, the persistent trading profitability suggests that the relative shapes

of the SPDs evolve very slowly over time. The option SPD is overly skewed and leptokurtic in general, not period specific.

We will provide below a peso-problem interpretation for the superior profits of the trading strategy. Indeed, by introducing a simple jump component to the index dynamics to reflect the nature of the peso problem, we are able to partially reconcile some of the differences in the time-series and cross-section SPDs. Both the 'skewness trades' and the 'kurtosis trades' sell OTM puts, which could conceivably incorporate a risk premium for the (downward) jump risk in the index. By selling these options in a time period where ex post no jumps actually occurred (but could have occurred ex ante), we are capturing this risk premium in the form of superior returns.

3.4. Implications for the aggregate pricing kernel

Our option trading simulations based on the estimated SPD differences achieves superior Sharpe ratios compared to buying and holding the S&P 500 index over our sample period: see Table 5. We now explore a different metric by which to assess the differences between the cross-sectional and time-series implied state-price densities, or equivalently pricing kernels or stochastic discount factors. In this section, we look at the volatility of the stochastic discount factor M that is consistent with our data. Hansen and Jagannathan (1991) show that the first-order condition $\mathrm{E}(MR^{\mathrm{e}}) = 0$ for the asset's excess return R^{e} implies the bound $\sigma(M) \geqslant \mathrm{E}(M)\mathrm{E}(R^{\mathrm{e}})/\sigma(R^{\mathrm{e}})$ where the expectation and the standard deviation are conditional on the information available at time t. The question we are now asking is how much tighter the bound is in the presence of options.

For a vector m of returns, the Hansen–Jagannathan bound without the positivity restriction, $M \geqslant 0$, becomes,

$$\sigma^2(M) \geqslant (1 - \mathrm{E}(M)\mathrm{E}(1 + R))'\Sigma^{-1}(1 - \mathrm{E}(M)\mathrm{E}(1 + R)) \tag{3.3}$$

where Σ is the variance–covariance matrix of the returns vector.

We present the volatility bound of the stochastic discount factor in Fig. 10 in the same format as Fig. 1 in Hansen and Jagannathan (1991, p. 228). The figure shows the feasible region for the stochastic discount factor using the S&P 500 index and T-bill (the solid line), and that using S&P 500 index, T-bill, and the option portfolios (the dashed line). The option portfolios are from our skewness and kurtosis trades reported in Table 5. For simplicity, the figure does not impose the nonnegativity restriction of the pricing kernel. Since we use our sample moments in our calculations, the short sample period and small number of assets inevitably introduces measurement error. However, our goal is to illustrate the more stringent restriction on the volatility bound when the option portfolio returns are included in our calculation.

In addition, we compare our volatility bounds obtained above to the volatility of the stochastic discount factor under the consumption-CAPM. Here, we

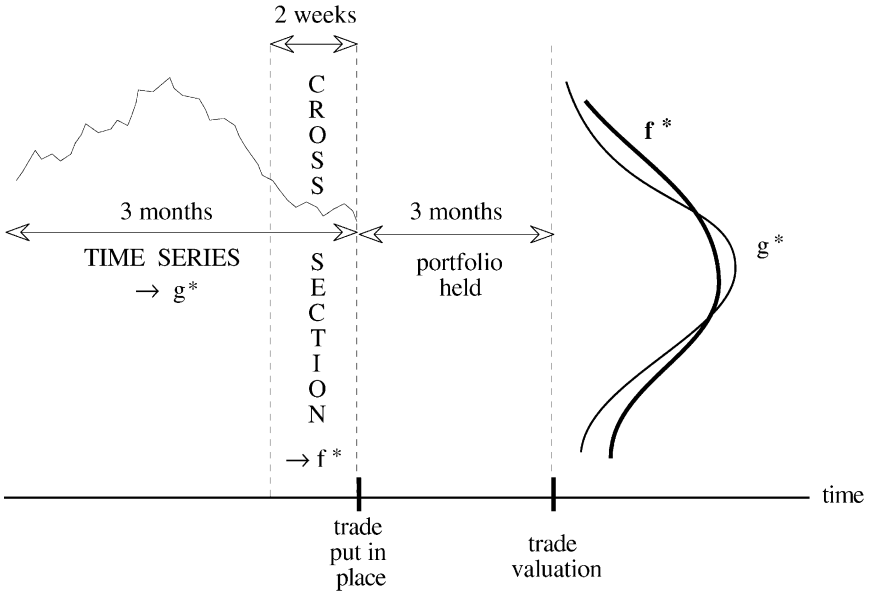


Fig. 8. Estimation periods for trading SPD differences.

take the representative agent's preferences to be given by a time-separable power utility function, $U(C) = (C^{1-\gamma} - 1)/(1-\gamma)$, where γ is the coefficient of relative risk aversion, and C represents aggregate consumption. In this case, the stochastic discount factor is given at date $t+1$ by $M_{t+1} = \delta(C_{t+1}/C_t)^{-\gamma}$, where δ is the subjective or time discount factor and C_t is the level of aggregate consumption at date t. We assume that consumption growth is an i.i.d. lognormal process. We calculate the volatility values of M corresponding to various values of the risk aversion parameter γ. The volatility of M generated by the consumption data is also plotted in Fig. 10. The first point above the horizontal axis has relative risk aversion of one; successive points have risk aversion of two, three, and so on; all values are obtained with $\delta = 1.0$. Not surprisingly, the bound for M is more restrictive in the presence of the option portfolios, although the restriction on the risk aversion parameters is essentially comparable to the benchmark case without options: the power utility pricing kernels do not enter the feasible region until the coefficient of relative risk aversion reaches the value 27, vs. 26 without options.

3.5. *Reconciling the implied SPDs: A peso-problem interpretation*

A natural interpretation for the differences in skewness and kurtosis between f^* and g^* lies in the possible existence of a peso problem for the S&P 500 index:

the option market could price the S&P 500 *options* as if the S&P 500 *index* were susceptible to large (downward) jumps of the type experienced in October 1987, even though jumps of that magnitude are absent from the time series data in the subsequent period. A number of studies have documented the presence of jumps in financial time series, and modeled their effect on option prices [see Bates (1991, 1996) and Bakshi et al. (1997)]. These studies have estimated the structural parameters of the postulated SPD from the option data. They compare the option-implied parameters to parameters estimated from the actual series of the underlying asset value – not its risk-neutralized series – which is possible through assumptions on preferences (typically log-utility or power utility). Our focus here is on the no-arbitrage restriction $f^* = g^*$, with the explicit objective of avoiding assumptions on preferences.

In a context where realized jumps are infrequent, our estimates of the time-series SPD would not be able to show any evidence of jumps; however, the estimated cross-sectional SPD would reflect the existence of a jump risk as long as that risk is priced [Merton (1976) derives a closed-form option-pricing formula if this risk is unpriced]. To examine how the presence of jumps would affect our empirical comparison of the two SPDs, we allow the risk-neutral dynamics of the index to contain a jump term in addition to its diffusion component:

$$\mathrm{d}S_t^* = (r_{t,\tau} - \delta_{t,\tau} - q^*\kappa^*)S_t^*\,\mathrm{d}t + \sigma(S_t^*)\,\mathrm{d}W_t^* + J_t^* S_t^*\,\mathrm{d}N_t^* \tag{3.4}$$

where we specify that N_t^* is a Poisson process with constant intensity q^* ($\Pr(\mathrm{d}N_t^* = 1) = q^*\,\mathrm{d}t$), independent of W_t^*, and for simplicity we make the percentage jump size J_t^* nonrandom. The corresponding actual dynamics are

$$\mathrm{d}S_t = \mu(S_t)\,\mathrm{d}t + \sigma(S_t)\,\mathrm{d}W_t + J_t S_t\,\mathrm{d}N_t \tag{3.5}$$

where N_t is a Poisson process with intensity q and the same jump times as N_t^*, and the percentage jump size is $J_t = J_t^* = \kappa^*$. The market price of jump risk is q^*/q.

When we simulate dynamics (3.4), with the estimated $\hat{\sigma}_{\mathrm{FZ}}(\cdot)$, we will draw observations from a process that incorporates the jump term. Over the multiple sample paths simulated, we will certainly observe some jumps, even though none were present in the single sample path of the actual data – this is the essence of the peso problem. Therefore, we use the same estimator of $\sigma(\cdot)$ as is the case where no jumps have been observed during the period of interest (three months in our empirical implementation). However, our estimated time-series SPD will reflect the presence of the jumps because we are simulating a large number of sample paths from a process where jumps are present, even though none were present in a single realization – the observed sample path.

Note that the parameter q^*/q, which enters the risk-neutral dynamics of the S&P 500, is determined by investors' preferences. In other words, we can no longer fully identify the risk-neutral dynamics from the time series of actual S&P 500 values. Instead, we will utilize the excess parameter q^* to make the time-series SPD g^* as close as possible to the cross-sectional SPD f^* which we previously estimated. We still make no specific assumptions on preferences, as we do not attempt to specify separately the actual jump intensity in (3.5). Specific preferences would tell us how to go from (3.4) to (3.5), but that step is not required here since in a peso-problem context the observed Poisson process does not actually jump (i.e., the *realized* values in the sample are all $dN_t = 0$).

Specifically, when jumps are possible, the equality $f^* = g^*$ is no longer an over-identifying restriction since g^* is no longer identifiable separately from the actual observations on $\{S_t\}$. Instead, this equality allows us to restore the *exact identification* of the system when previously it was an over-identifying restriction. As a result, there are no testable implications to be drawn, but we still fully identify the system without assumptions on preferences. Of course, if we were willing to set bounds on the risk premium associated with the jump risk – or equivalently investors' risk aversion – then we could draw further conclusions regarding the plausibility of the actual jump arrival intensity q corresponding to the estimated q^*. Alternatively, we could use the additional restrictions given by the equality between the two SPDs for different option maturities. This would then restore the overidentification of the system and generate testable implications.

Empirically, using post-1962 daily returns, we estimate the standard deviation of the S&P 500 index ex-dividend returns to be 0.855%. We classify negative returns beyond five standard deviations to be downward jump events. There are seven such events out of the 8685 daily observations (roughly corresponds to once every 5 year) with an average jump size of -8%. As a realistic specification, we therefore perform our simulations with a fixed downward jump size of -10%, and moment-matching scheme for four jump frequency specifications,

Table 6
Skewness and kurtosis SPD fit in the presence of jumps[a]

	No jump	Jump frequency				
		10 yr	5 yr	3 yr	2 yr	1 yr
Skewness fit	0.5933	0.5344	0.5019	0.4806	0.5085	0.6964
Kurtosis fit	1.6053	1.5187	1.4930	1.4947	1.5642	1.8014

[a]Root-mean-squared differences between the cross-sectional and time-series SPD moments, where the time series dynamics incorporate a jump term as in (3.4). The jump frequency parameter for the jump specification is λ^*. The jump frequency is quoted in the table in terms of one jump per length of time.

corresponding to one jump every ten years, five years, two years, and one year, respectively. The root mean squared (RMS) differences in Table 6 between the cross-sectional and time-series (with jump) SPD moments exhibit a U-shaped pattern: the once-every-ten-years specification does not produce enough skewness, whereas the once-every-two-years one results in a skewness for the time-series SPD that overshoots that of the cross-sectional SPD. We find that the minimum RMS is achieved for a frequency close to once every three years. We therefore simulated (3.4) with this fifth jump frequency specification, which indeed produces a time-series SPD that best matches the cross-section SPD.

Fig. 6 is analogous to Fig. 5, but the time-series SPDs include the possibility of jumps at the once-every-three-years frequency. Note that the scatter plot is now more evenly distributed around the 45° axis, i.e., the conditional moments of the cross-sectional SPD are matched more accurately when jumps are included than they were in Fig. 5. It is apparent that the incorporation of a jump term produces an improvement towards reconciling the cross-sectional and time-series SPDs, but is not sufficient to explain the magnitude of the dispersion of the excess skewness and excess kurtosis of f^* relative to g^*. In particular, the inclusion of jumps reconciles the cross-sectional and time-series skewnesses better than it does for the corresponding kurtoses, which is to be expected since we have constrained the jumps to be exclusively of negative sign.

3.6. A comparison of out-of-the-money put prices

Our fourth and last measure of the differences between the two SPDs consists in comparing the out-of-the-money put prices implied by the two distributions. Rather than comparing conditional moments, we are now comparing integrals of the option's payoff against the respective densities. For out of the money options, we are essentially computing (weighted) tail probabilities. Indeed, excess skewness and kurtosis implies that, other things being equal, out-of-the-money put options are more expensive. We report in Fig. 7 the relative average difference between the prices implied by both f^* and g^*. We obtain put prices by integrating the left tail of the respective SPD against the respective put payoff, and discounting at the riskfree rate. For a given moneyness level (defined as the ratio of the option's strike price to the S&P 500 futures price), we average the corresponding prices over the corresponding periods used in forming the SPD estimates. We finally compute the relative difference between the prices implied by the f^* and g^* SPDs. As a result, excess skewness or kurtosis embedded in option prices should translate in a relative price difference which gets larger as their moneyness is lower. Fig. 7 shows that this is indeed the case especially after the 1987 crash. Far out of the money puts valued using f^* can be as much as 70% more expensive than their counterparts valued using g^*. Further, Fig. 7 confirms Fig. 6: it shows that the out-of-the-money put prices

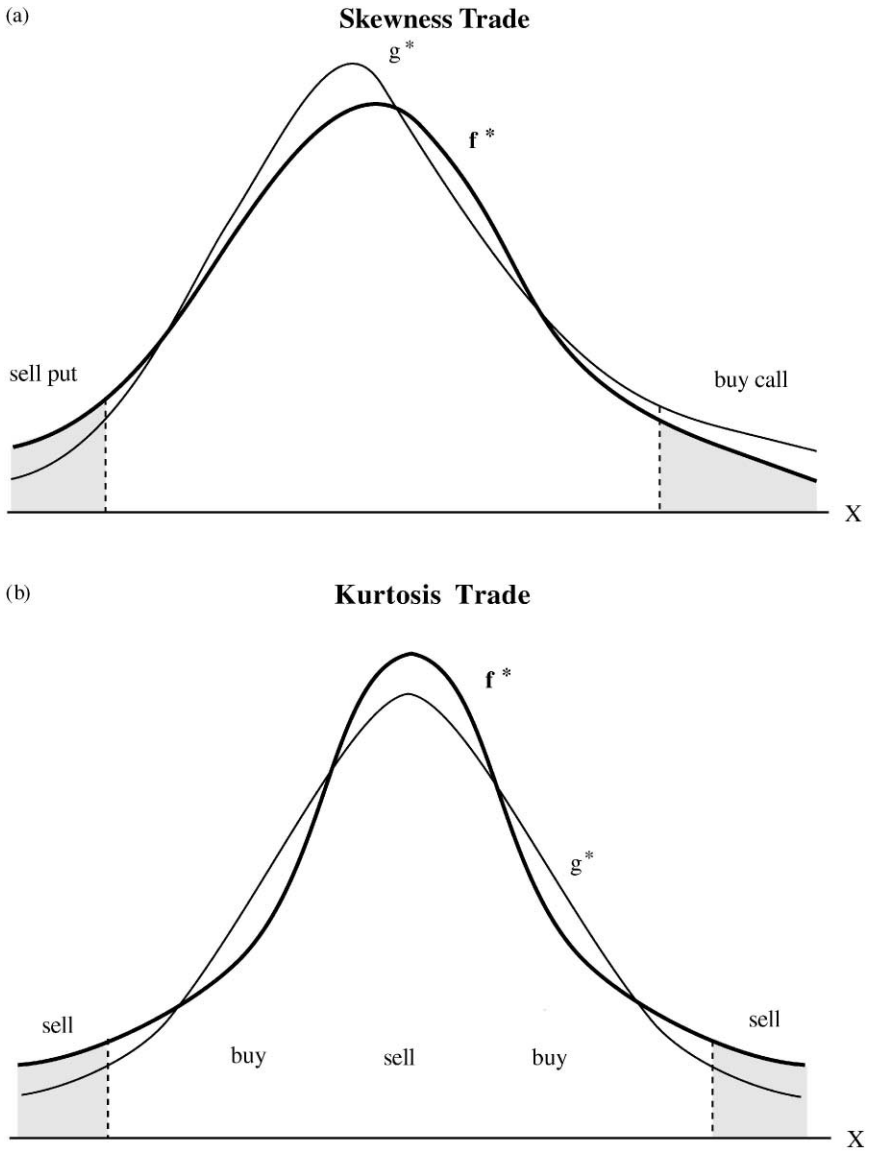


Fig. 9. (a) Skewness trade. (b) Kurtosis trade.

implied by the time-series SPD when jumps are included are closer to those implied by the cross-sectional SPD than the corresponding curve without jumps.

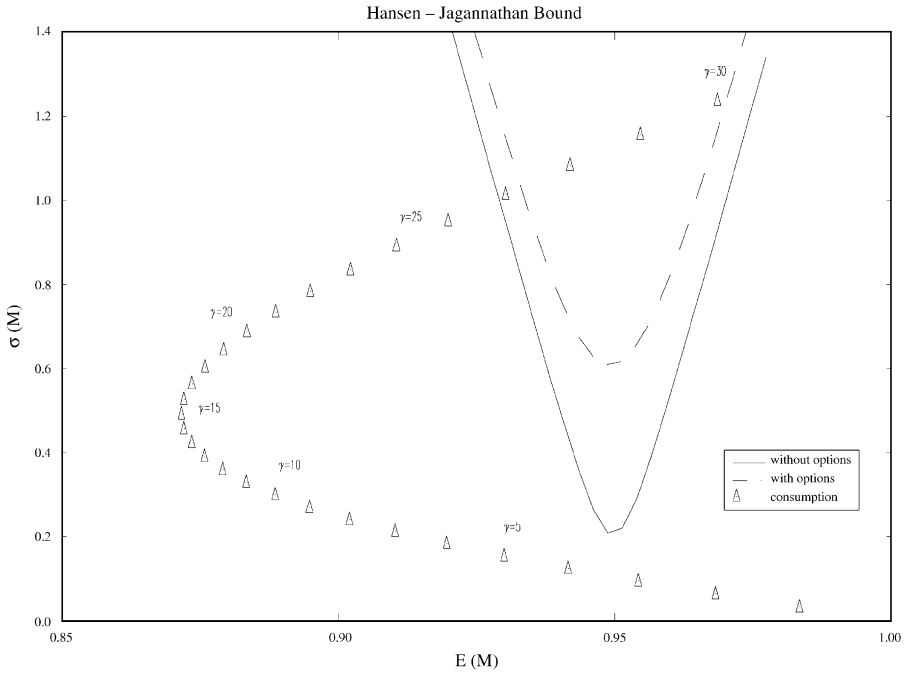


Fig. 10. Hansen–Jagannathan bound.

4. Conclusions

We have provided a method to infer from the observed time series of underlying asset values the SPD that is directly comparable to the option-implied SPD, and requires no assumptions on the representative preferences. We also showed how separate exact identification of the two SPDs from the observable data can be achieved, leaving the theoretical restriction of equality between the two SPDs as an overidentifying restriction. In that context, the equality between the cross-sectional and time-series SPDs becomes an over-identifying restriction, which is therefore testable. In addition to making no assumptions whatsoever on preferences, we also avoided parametric assumptions on the nature of the diffusion driving the observable asset's price.

The extension of this method to the case of multiple traded state variables poses no conceptual difficulties; in practice, our reliance on nonparametric estimators would result, in higher dimensions, in a loss of accuracy, and hence power to detect deviations from one SPD to the other. This 'curse of dimensionality' is a consequence of the local character of nonparametric estimators: accurate estimates can only be obtained in regions of the state space that are

visited often enough by the variables. For a given sample size, each region is revisited much less often as the dimension gets higher. With this standard caveat in mind, Monte-Carlo evidence shows that our method performs very well in the context that was relevant for our empirical application.

In the case of the S&P 500 index, our comparison of the two SPDs reveals that the market prices options with an overly skewed and leptokurtic SPD. Thus we reject the joint hypothesis that the S&P 500 options are efficiently priced and that the S&P 500 index follows a one-factor diffusion. Trading schemes designed to exploit the SPD differences are able to produce superior profits. The high Sharpe ratios achieved by these trading schemes demand excessive variation in investors' marginal utilities, which further questions market efficiency. One possible explanation for this evidence is the presence of an S&P 500 peso problem, whereby options incorporate a premium for the jump risk in the underlying index that is absent from its recorded time series. We find some evidence in favor of that interpretation.

Moving away from the restrictive nature of the univariate diffusion specification, we were able to partly reconcile the differences between the index and option implied SPDs by adding a jump component to the index dynamics. The empirical results of this paper can therefore be interpreted alternatively as evidence of the limitations of a one-factor diffusion structure for the underlying asset returns rather than an indictment on the rationality of the options markets. A further natural departure from this specification would consist in incorporating stochastic volatility to our specification. Our method can be extended to incorporate the case where the volatility of the underlying asset is stochastic as a separate process, rather than stochastic only through its dependence on the asset price. The law of motion of the volatility process under the real probability measure can be estimated using standard filtering techniques. However, given that volatility is a nontraded asset, its risk-neutral behavior cannot be solely identified by estimating the diffusion function as in Section 2.3.1. We can still estimate a three-dimensional option price function as in Section 2.2 (with volatility as the additional, third, regressor). Then by applying Itô's lemma we can derive the drift and diffusion function of any derivative price. In the absence of arbitrage, the respective market prices of S&P 500 risk and volatility risk should be identical for all derivative assets that are subject to these two sources of risk. This implies (as in the APT) that the instantaneous expected excess return of these derivative assets should be an average of the market prices of risk weighted by the size of their respective exposures, i.e., their diffusion coefficients. Therefore, given the instantaneous expected excess returns and the diffusion function for options of two different maturities, we can solve for the implied market prices of risk. We can then obtain an overidentifying restriction by checking if the implied market prices of risk are consistent with the derived expected instantaneous excess return and diffusion function for a different set of derivative prices, for instance options with a third maturity. Hence, we could

still, at least theoretically, test the validity of the model without resorting to any assumptions on preferences.

Acknowledgements

We are grateful to David Bates, George Constantinides, Robert Engle and Robert MacDonald for useful discussions. Seminar participants also provided helpful comments. This research was conducted in part during the first author's tenure as an Alfred P. Sloan Research Fellow. Financial support from the NSF (Grant SBR-9996023) and the University of Chicago's Center for Research in Security Prices is gratefully acknowledged.

Appendix A. Assumptions and asymptotic distributions

For each three-month subperiod, option prices form a panel data, consisting of N observation periods and J options per period. The sample size relevant for the computation of the cross-sectional SPD f^* is $n = NJ$, and N for the computation of the time-series SPD g^*. We make the following assumptions on the data used to construct the nonparametric regression (2.9), i.e., $(\sigma, \tilde{\boldsymbol{Z}})$ where $\tilde{\boldsymbol{Z}} \equiv [X/F_{t,\tau}, \tau]'$. The nonparametric regression function is $\hat{\sigma}(\tilde{\boldsymbol{Z}})$, and we wish to estimate its second partial derivative with respect to the first component $X/F_{t,\tau}$ of the vector $\tilde{\boldsymbol{Z}}$.

Assumption A.1.

1. The process $\{Y_i \equiv (\sigma_i, \tilde{\boldsymbol{Z}}_i)\colon i = 1, \ldots, n\}$ is strictly stationary with

$$\mathrm{E}[\sigma_i^4] < \infty, \quad \mathrm{E}[\|\tilde{\boldsymbol{Z}}_i\|^2] < \infty$$

   and is β-mixing with mixing coefficients β_j that decay as $j \to \infty$ at a rate at least as fast as j^{-b}, $b > 19/2$. The joint density of (Y_1, Y_{1+j}) exists for all j and is continuous.
2. The density $\pi(\sigma, \tilde{\boldsymbol{Z}})$ is p-times continuously differentiable with respect to $\tilde{\boldsymbol{Z}}$, with $p > m$, and π and its derivatives are bounded and in $L_2(\mathbf{R}^{1+d})$. The marginal density of the nonparametric regressors, $\pi(\tilde{\boldsymbol{Z}})$, is bounded away from zero on every compact set in $\mathbf{R}^d$.
3. $\sigma(\tilde{\boldsymbol{Z}})\pi(\tilde{\boldsymbol{Z}})$ and its derivatives are bounded. The conditional variance

$$s^2(\tilde{\boldsymbol{Z}}) \equiv \mathrm{E}[(\sigma - \sigma(\tilde{\boldsymbol{Z}}))^2 \mid \tilde{\boldsymbol{Z}}] \tag{A.1}$$

   is bounded and in $L_2(\mathbf{R}^d)$. The conditional fourth moment $\mathrm{E}[(\sigma - \sigma(\tilde{\boldsymbol{Z}}))^4 \mid \tilde{\boldsymbol{Z}}]$ is bounded.

Definition A.1. A kernel function k is of order q if

$$\int_{-\infty}^{+\infty} z^l k(z)\,\mathrm{d}z = \begin{cases} 1 & \text{if } l = 0 \\ 0 & \text{if } 0 < l < q \\ (-1)^q q!\chi_q & \text{if } l = q \end{cases}$$

where l is an integer and $\int_{-\infty}^{+\infty} |z|^l |k(z)|\,\mathrm{d}z < \infty$ for all $0 \leqslant l \leqslant q$.

Assumption A.2. The kernel functions $k_{X/F}$ and k_τ are bounded, three-times continuously differentiable, and have derivatives which are bounded and in $L_2(\mathbf{R})$. $k_{X/F}$ is of order $q_{X/F}$ and k_τ is of order q_τ. The bandwidths are given by

$$h_{X/F} = c_{X/F}\mathrm{s}(\boldsymbol{X/F})n^{-1/(\tilde{d}+2(q_{X/F}+m))}, \quad h_\tau = c_\tau \mathrm{s}(\boldsymbol{\tau})n^{-1/(\tilde{d}+2q_\tau)} \tag{A.2}$$

where $\mathrm{s}(\boldsymbol{X/F})$ and $\mathrm{s}(\boldsymbol{\tau})$ are the unconditional standard deviations of the non-parametric regressors, $c_{X/F} \equiv \gamma_{X/F}/\ln(n)$, with $\gamma_{X/F}$ constant, and $c_\tau \equiv \gamma_\tau/\ln(n)$, with γ_τ constant.

In practice, we use the kernel functions

$$k_{(2)}(z) \equiv \mathrm{e}^{-z^2/2}/\sqrt{2\pi}, \quad k_{(4)}(z) \equiv (3 - z^2)\mathrm{e}^{-z^2/2}/\sqrt{8\pi} \tag{A.3}$$

which are of order $q = 2$ and $q = 4$ respectively. We then obtain, as in Aït-Sahalia and Lo (1998), with $\tilde{\boldsymbol{Z}} = [\boldsymbol{X/F}_{t,\tau}, \tau]$:

Proposition A.1. Under Assumptions A.1 and A.2

$$n^{1/2}h_{X/F}^{5/2}h_\tau^{1/2}[\hat{f}^*(S_T) - f^*(S_T)] \xrightarrow{\mathrm{d}} \mathrm{N}(0, \sigma_{f^*}^2) \tag{A.4}$$

where $\partial H_{\mathrm{BS}}/\partial\sigma$ is the option's gamma evaluated at $\hat{\sigma}(\tilde{\boldsymbol{Z}})$ and

$$\sigma_{f^*}^2 \equiv \left[\mathrm{e}^{r_{t,\tau}\tau}\frac{\partial H_{\mathrm{BS}}}{\partial\sigma}(\hat{\sigma}(\tilde{\boldsymbol{Z}}), \boldsymbol{Z})\right]^2 \frac{s^2(\tilde{\boldsymbol{Z}})(\int_{-\infty}^{+\infty}(k_{X/F}^{(2)})^2(\omega)\,\mathrm{d}\omega)(\int_{-\infty}^{+\infty}k_\tau^2(\omega)\,\mathrm{d}\omega)}{\pi(\tilde{\boldsymbol{Z}})F_{t,\tau}^4}. \tag{A.5}$$

To estimate the time-series SPD g^*, we make the following assumptions:

Assumption A.3. The drift function $\mu(\cdot)$ in (2.2) is a bounded function, twice continuously differentiable with bounded derivatives. The instantaneous volatility function $\sigma(\cdot)$ in (2.2) and (2.3) has three continuous and bounded derivatives, and there exist two constants c and C such that for all S.

Assumption A.4. The kernel function k_{FZ} is in $L_2(\mathbf{R})$ and is of order $q_{FZ} = 2$. The bandwidth h_{FZ} to form $\hat{\sigma}_{FZ}(\cdot)$, is given by

$$h_{FZ} = c_{FZ} N^{-1/4} \tag{A.6}$$

where $c_{FZ} \equiv \gamma_{FZ}/\ln(N)$, and γ_{FZ} is a constant.

Assumption A.5. The kernel function k_{MC} is in $L_2(\mathbf{R})$ and is of order q_{MC}. The bandwidth h_{MC} is given by

$$h_{MC} = c_{MC} M^{-1/(1+2q_{MC})} \tag{A.7}$$

where M is the number of Monte-Carlo simulations, $c_{MC} \equiv \gamma_{MC}/\ln(M)$, and γ_{MC} is a constant.

We discuss in Section 3 how we select the constants γ_{FZ} and γ_{MC}. Letting the number of simulated sample paths of (2.3) go to infinity, we obtain

Proposition A.2. Under Assumptions A.3 and A.4:

$$[\hat{g}_t^*(S_T) - g_t^*(S_T)] = \mathrm{O_p}(N^{-1/2}). \tag{A.8}$$

Appendix B. A nonparametric comparison test of $f^* = g^*$

To construct the estimators involved in the test statistic of Section 2.4, we use the following bandwidth and kernel function, for the one-dimensional semiparametric model corresponding to $\tilde{\boldsymbol{Z}} = [\boldsymbol{X}/\boldsymbol{F}_{t,\tau}]$:

Assumption B.1. The bandwidth $h_{X/F}$ to estimate $\hat{f}^*$ in (2.21) is given by

$$h_{X/F} = \eta_{X/F} n^{-1/\delta_{X/F}}$$

where $\eta_{X/F}$ is constant, and $\delta_{X/F}$ satisfies

$$0 < \delta_{X/F} < 2q_{X/F} + \frac{9}{2} \tag{B.1}$$

where $q_{X/F}$ is the order of the kernel function $k_{X/F}$.

We then obtain

Proposition B.1. Under Assumptions A.1, A.3–A.5 *and under* H_0

$$nh_{X/F}^{9/2} D(\hat{f}^*, \hat{g}^*) - h_{X/F}^{-1/2} B_D \xrightarrow{\mathrm{d}} \mathrm{N}(0, \boldsymbol{\Sigma}_D^2) \tag{B.2}$$

where

$$B_D = \left(\int_{-\infty}^{+\infty} k''^2(w)\,\mathrm{d}w\right)\int_{\tilde{\boldsymbol{Z}}} \sigma^2(\tilde{\boldsymbol{Z}})\tilde{\omega}(\tilde{\boldsymbol{Z}})\,\mathrm{d}\tilde{\boldsymbol{Z}} \tag{B.3}$$

$$\boldsymbol{\Sigma}_D^2 = 2\left[\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} k''(w)k''(w+v)\,\mathrm{d}w\right)^2 \mathrm{d}v\right]\int_{\tilde{\boldsymbol{Z}}} \sigma^4(\tilde{\boldsymbol{Z}})\tilde{\omega}^2(\tilde{\boldsymbol{Z}})\,\mathrm{d}\tilde{\boldsymbol{Z}} \tag{B.4}$$

where

$$\tilde{\omega}(\tilde{\boldsymbol{Z}}) \equiv \left(\mathrm{e}^{r_{t,\tau}\tau}\frac{\partial H_{\mathrm{BS}}(\sigma(\tilde{\boldsymbol{Z}}), \boldsymbol{Z})}{\partial\sigma}\right)^2 \omega(\tilde{\boldsymbol{Z}}). \tag{B.5}$$

Note that the sampling variation of $\hat{g}^*$ does not enter the asymptotic distribution of the test statistic under H_0, due to the faster rate of convergence of $\hat{g}^*$ compared to $\hat{f}^*$ (compare (A.4) to (A.8)).

To estimate consistently the conditional variance of the regression, $s^2(\tilde{\boldsymbol{Z}})$ we calculate the difference between the kernel estimate of the regression of the squared dependent variable σ^2 on $\tilde{\boldsymbol{Z}}$ and the squared of the regression $\sigma(\tilde{\boldsymbol{Z}})$ of the dependent variable σ on $\tilde{\boldsymbol{Z}}$. The regression $\mathrm{E}[\sigma^2\,|\,\tilde{\boldsymbol{Z}}]$ is estimated with bandwidth $h_{cv} = \eta_{cv} n^{-1/\delta_{cv}}$), $\delta_{cv} = \delta_\tau$ and η_{cv} constant. The test statistic is formed by standardizing the asymptotically normal distance measure D: we estimate consistently B_D and $\boldsymbol{\Sigma}_D^2$ by plugging-in the estimate of the conditional variance $\sigma^2(\tilde{\boldsymbol{Z}})$, then subtract the asymptotic mean and divide by the asymptotic standard deviation. The test statistic then has the asymptotic N(0, 1) distribution. Since the test is one-sided (we only reject when $D(\hat{f}^*, \hat{g}^*)$, hence the test statistic, is large and positive), the 10% critical value is 1.28, while the 5% value is 1.64. We fix the variables in $\boldsymbol{Z}$ that are excluded from $\tilde{\boldsymbol{Z}}$ at their sample means. Kernel constants appearing in Propositions A.1 and A.2 are given in Table 7.

Table 7
Kernel constants in asymptotic distributions[a]

Differentiation order m	0	1	2
$\int_{-\infty}^{+\infty}(k_{(2)}^{(m)})^2(w)\,\mathrm{d}w$	$\frac{1}{2\sqrt{\pi}}$	$\frac{1}{4\sqrt{\pi}}$	$\frac{3}{8\sqrt{\pi}}$
$\int_{-\infty}^{+\infty}(k_{(4)}^{(m)})^2(w)\,\mathrm{d}w$	$\frac{27}{32\sqrt{\pi}}$	$\frac{175}{64\sqrt{\pi}}$	$\frac{273}{128\sqrt{\pi}}$
$\int_{-\infty}^{+\infty}(\int_{-\infty}^{+\infty} k_{(2)}^{(m)}(w)k_{(2)}^{(m)}(w+v)\,\mathrm{d}w)^2\,\mathrm{d}v$	$\frac{1}{2\sqrt{2\pi}}$	$\frac{3}{32\sqrt{2\pi}}$	$\frac{105}{512\sqrt{2\pi}}$

[a]Kernel constants that characterize the asymptotic variances of the nonparametric estimators, and the distribution of the test statistic. The kernel functions $k_{(2)}$ and $k_{(4)}$ are defined in (A.3). $k^{(m)}$ denotes the mth derivative of k.

Appendix C. Performance of the diffusion estimators in Monte-Carlo simulations

In our simulations, we set the drift of the log-return process to be constant at 9% per year, which corresponds to the long-run historical ex-dividend average of the S&P 500 index. We calculate the volatility estimator with a given bandwidth h_{FZ} to each path and obtain the estimated volatility as a function of the price level. To investigate the sensitivity of the estimators to the choice of bandwidths, we evaluate the estimators over a range of possible bandwidth values. Two performance measures of the estimators are used for each choice of bandwidth.

We adopt as an evaluation criterion the mean integrated squared error (MISE) of the estimator. For each path, point-by-point squared estimation errors (SE) are calculated. Then the SEs are weighted by their density (of S over all paths) and are summed up, producing the integrated squared error (ISE). The average ISE over the 1000 paths becomes the MISE. However, the MISE does not explicitly tell how well the estimator captures the shape of the diffusion function. In fact it tends to favor the smoothness of the estimate over how accurately the shape is captured. In practice, when estimating the diffusion function, we are not only interested in its level, but also its shape which determines the skewness and the kurtosis of the time-series SPD g^*.

Thus we propose a second performance measure that emphasizes how well the shape is captured (since the diffusion function is parametrically specified, with parameters that determines the shape). The second measure describes how well the estimated diffusion function can reproduce these parameters. To obtain the estimated value $\hat{\theta} = (a, b, c)$ of $\theta = (\alpha, \beta, \gamma)$ for each path, we run an OLS regression. The average of (a, b, c) over the 1000 paths are calculated. We then proceed to test the separate hypotheses $\alpha = \alpha_0$, $\beta = \beta_0$, $\gamma = \gamma_0$ (the subscript 0 denotes true values) using the t-test, and the joint hypothesis $\theta = \theta_0$, using the χ^2-test. We then select a bandwidth that balances MISE-minimization and shape preservation.

In general, the FZ estimators capture the level and slope very well over almost all shapes of volatility considered. The only exception is that the estimator tends to under-estimate the slope as the volatility level gets lower. When the level and the spread is large, FZ captures the curvature well, but not when the spread is small and the level is low. However, one can argue that the ability to capture the curvature is more crucial in the former case. To be closer to the theoretical assumptions of FZ, we also generated paths using a small drift (3%) and found that it resulted in a slight improvement in the curvature estimation. Finally, when we increased the number of draws per day the performance of the estimator increased substantially, providing support for the asymptotic convergence results that we rely on. Table 8 summarizes quantitatively the estimator's ability to capture the shape of the $\sigma(\cdot)$ for a variety of possible configurations. In Fig. 11, we select four typical shapes of the true function $\sigma(\cdot)$ and plot the average $\hat{\sigma}_{FZ}(\cdot)$ estimated over all the simulated sample paths, as well as its 95% Monte-Carlo confidence interval.

Table 8
Monte-Carlo simulations of the diffusion estimator[a]

Shape of $\sigma^2(\cdot)$				Estimation $\hat{\sigma}^2_{FZ}(\cdot)$	Error $\hat{\sigma}^2_{ML}$	Relative error
Spread	Slope	Curvature	Figure			
Flat	Flat	Flat	11a	0.000233	0.000045	5.2
Large	Up	Concave		0.000512	0.0229	0.02
Large	Up	Flat		0.00106	0.00962	0.11
Large	Down	Flat	11b	0.000871	0.0253	0.03
Large	Down	Convex		0.00911	0.0323	0.28
Small	Up	Flat		0.000699	0.0112	0.06
Small	Up	Concave	11c	0.000808	0.0216	0.04
Small	Down	Flat		0.000617	0.0129	0.05
Small	Down	Convex	11d	0.000955	0.0294	0.03

[a]This table reports the results of Monte-Carlo simulations of the diffusion estimator for a variety of shapes of the true diffusion function $\sigma^2(\cdot)$, which is parameterized as $\sigma(S) = \alpha + \beta(S - S_0) + \gamma(S - S_0)^2$. The parameter α is set at 0.20. A 'large' spread represents the cases where $\beta = \pm 0.002$ and $\gamma = \pm 0.00004$, whereas a 'small' spread corresponds to the cases where $\beta = \pm 0.001$ and $\gamma = \pm 0.00002$. The sign of β and γ determines the qualifiers up, down, concave and convex in the usual sense. 'Flat' corresponds to a parameter value of 0. The estimation error reported in the table is the integrated mean squared error of the estimator (squared bias plus variance), averaged over all the simulations. For each shape of the diffusion function, we have used a bandwidth that balances MISE-minimization and shape preservation. In the case where the true function is constant (and we know it), the maximum-likelihood estimator is naturally the most efficient choice (see the first row of the table). However, in general the estimation error of $\hat{\sigma}^2_{FZ}$ is less than 5% of that of $\hat{\sigma}^2_{ML}$.

Appendix D. Monte-Carlo simulations of the conditional moment regressions

In Table 9, we report the results of Monte-Carlo simulations of the conditional moments regressions used to compare the time-series and cross-sectional conditional SPD moments in Section 3.2, specifically Table 2. These simulations are useful to assess the degree of small sample bias and variance inherent in these conditional moments comparisons in conditions that approximate those of the real data.

We replicate 500 economies with 11 year histories consisting of 55 subperiods each. Each subperiod consists of 10,000 simulated sample paths for the underlying S&P 500 index assumed to follow Cox's CEV process

$$dS_t^* = (r - \delta)S_t^* \, dt + \sigma(S_t^*)^{\alpha} \, dW_t^*.$$

We set the initial value of the index at 450, the risk free rate r at 5% and the dividend payout rate δ at 0%. For each period, we draw the volatility parameter from a normal distribution with mean of 35% (the real data average) and a standard deviation of 15%, with a cap and floor of 70% and 5%, respectively.

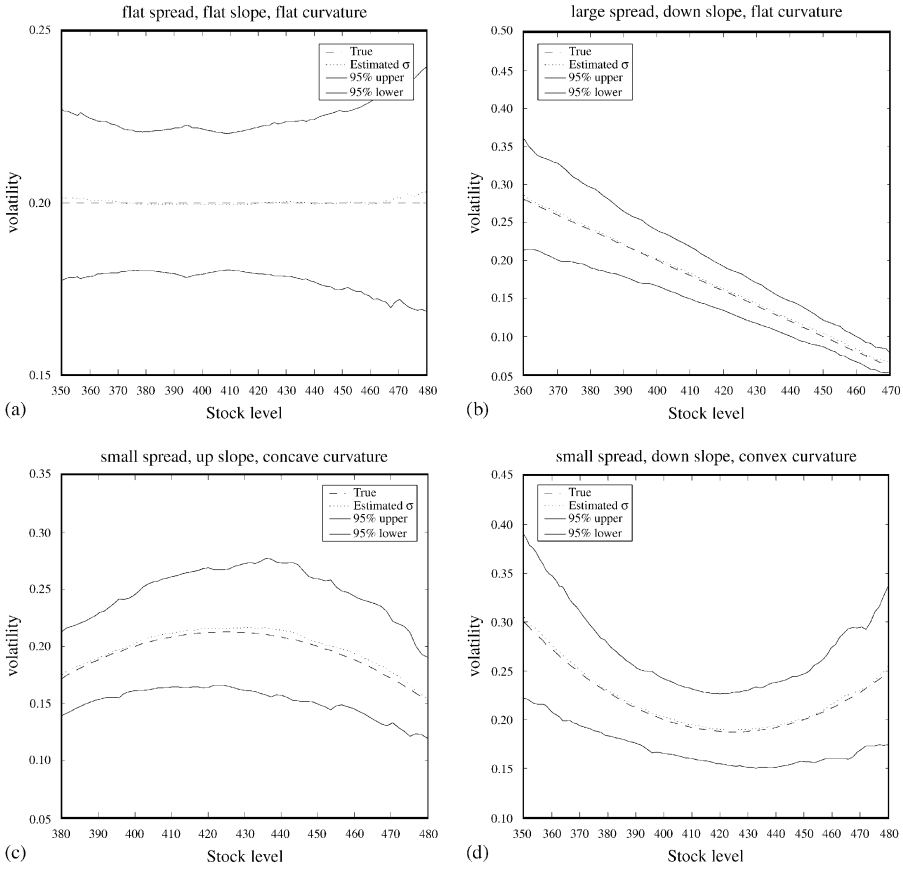


Fig. 11. Monte-Carlo confidence interval for the estimated σ. (a) flat spread, flat slope, flat curvature; (b) large spread, down slope, flat curvature; (c) small spread, up slope, concave curvature; (d) small spread, down slope, convex curvature.

We then calculate returns from the simulated paths and estimate the time-series SPD using the same estimator as with the real data. On the option side, we analytically price a cross-section of call options for each of the 55 periods of the 500 sets using the closed-form CEV formula. The strikes range from 25% in the money to 35% out of the money. The time-to-maturity is three months, as in the real data. We use the same volatility parameter as that in the matching index simulation subperiod so that in the absence of noise and estimation error the two SPDs would by construction be identical. As with the real data, we then compute the cross-sectional SPD for each of the 55 subperiods from the associated cross-section of option prices. Since the cross-sectional moments are more precisely estimated than their time-series counterparts, the regressions are

Table 9
Monte-Carlo simulations of the conditional moments regressions[a]

Moment	Intercept MC average (MC standard deviation)	Slope MC average (MC standard deviation)	R^2 MC average (MC standard deviation)
Mean	− 0.0004 (0.001)	1.019 (0.09)	0.73 (0.06)
Volatility	− 0.002 (0.006)	0.988 (0.009)	0.99 (0.001)
Skewness	0.012 (0.03)	1.08 (0.08)	0.75 (0.06)
Kurtosis	− 0.14 (0.11)	1.22 (0.20)	0.42 (0.10)

[a]This table reports the results of Monte-Carlo simulations of the conditional moment regressions. For each simulation, we compute the intercept and slope coefficients and R^2 from a regression of the conditional moments using the simulated 55 subperiods. The table reports the mean and standard deviations obtained from 500 overall simulations. The setup for the simulations is detailed in Appendix D and reproduces the main features of the real data, including sample size, number of subperiods and general level of noise. The results of these simulations show that the deviations between the cross-sectional and time-series conditional skewnesses and kurtoses reported in Table 2 are significant in the sense of being outside their respective Monte-Carlo confidence intervals.

corrected for the potential errors in variables problem (where y is regressed on a constant and x, with x measured with error) by inferring the coefficients from the better-estimated reverse regression.

The closed-form expression of the CEV SPD is

$$f_t^*(S_t, S_T, \tau, r, \delta) = g_t^*(S_t, S_T, \tau, r, \delta)$$
$$= \frac{2S_T^{-1+2(1-\alpha)}\left(\dfrac{e^{2\tau(1-\alpha)\kappa}S_t^{2(1-\alpha)}}{S_T^{2(1-\alpha)}}\right)^{1/4(1-\alpha)}\kappa I_{1/2(1-\alpha)}\left(\dfrac{2\sqrt{e^{2\tau(1-\alpha)\kappa}S_T^{2(1-\alpha)}S_t^{2(1-\alpha)}}\kappa}{(-1+e^{2\tau(1-\alpha)\kappa})(1-\alpha)\sigma^2}\right)}{\exp\left\{\dfrac{(S_T^{2(1-\alpha)}+e^{2\tau(1-\alpha)\kappa}S_t^{2(1-\alpha)})\kappa}{(-1+e^{2\tau(1-\alpha)\kappa})(1-\alpha)\sigma^2}\right\}(-1+e^{2\tau(1-\alpha)\kappa})\sigma^2}$$

where $\kappa = r - \delta$ (see Cox, 1996). This expression reduces to

$$f_t^*(S_t, S_T, \tau, r, \delta) = g_t^*(S_t, S_T, \tau, r, \delta) = \frac{2\sqrt{\kappa}\sinh\left(\dfrac{2e^{\tau\kappa}S_T S_t \kappa}{(-1+e^{2\tau\kappa})\sigma^2}\right)}{\exp\left\{\dfrac{(S_T^2+e^{2\tau\kappa}S_t^2)\kappa}{(-1+e^{2\tau\kappa})\sigma^2}\right\}\sqrt{-1+e^{2\tau\kappa}}\sqrt{\pi}\sigma}$$

for the absolute model ($\alpha = 0$).

References

Aït-Sahalia, Y., 1996. Nonparametric pricing of interest rate derivative securities. Econometrica 64, 527–560.

Aït-Sahalia, Y., Duarte, J., 1999. Nonparametric option pricing under shape restrictions. Working paper, Princeton University.

Aït-Sahalia, Y., Lo, A.W., 1998. Nonparametric estimation of state-price densities implicit in financial asset prices. Journal of Finance 53, 499–547.

Aït-Sahalia, Y., Lo, A.W., 2000. Nonparametric risk management and implied risk aversion. Journal of Econometrics 94, 9–51.

Bakshi, G., Cao, C., Chen, Z., 1997. Empirical performance of alternative option pricing models. Journal of Finance 52, 2003–2049.

Banz, R., Miller, M., 1978. Prices for state-contingent claims: some estimates and applications. Journal of Business 51, 653–672.

Bates, D.S., 1991. The crash of '87: was it expected? The Evidence from Options Markets. Journal of Finance 46, 1009–1044.

Bates, D.S., 1996. Jumps and stochastic volatility: exchange rate processes implicit in Deutsche Mark options. Review of Financial Studies 9, 69–107.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–659.

Breeden, D., Litzenberger, R., 1978. Prices of state-contingent claims implicit in option prices. Journal of Business 51, 621–651.

Canina, L., Figlewski, S., 1993. The informational content of implied volatility. Review of Financial Studies 6, 659–681.

Chiras, D., Manaster, S., 1978. The informational content of option prices and a test of market efficiency. Journal of Financial Economics 6, 213–234.

Cochrane, J.H., Hausen, L.P., 1992. Asset pricing explorations for macroeconomics, in NBER Macroeconomics Annual, ed. by O. Blanchard and S. Fisher, MIT Press, Cambridge, MA, pp. 115–165.

Cox, J., 1996. The constant elasticity of variance option pricing model. The Journal of Portfolio Management, Special Issue, 15–17.

Cox, J., Ross, S., 1976. The valuation of options for alternative stochastic processes. Journal of Financial Economics 3, 145–166.

Day, T., Lewis, C., 1988. The behavior of volatility implicit in the prices of stock index options. Journal of Financial Economics 22, 103–122.

Day, T., Lewis, C., 1990. Stock market volatility and the information content of stock index options. Journal of Econometrics 52, 267–287.

Derman, E., Kani, I., 1994. Riding on the smile. RISK 7, 32–39.

Dumas, B., Flemming, J., Whaley, R.E., 1998. Implied volatility functions: empirical tests. Journal of Finance 53, 2059–2106.

Dupire, B., 1994. Pricing with a smile. RISK 7, 18–20.

Engle, R.F., Mustafa, C., 1992. Implied ARCH models from option prices. Journal of Econometrics 52, 289–311.

Fama, E.F., 1976. Foundations of Finance. Basil Books, New York, NY.

Florens–Zmirou, D., 1993. On estimating the diffusion coefficient from discrete observations. Journal of Applied Probability 30, 790–804.

Goldberger, D.H., 1991. A unified method for pricing options on diffusion processes. Journal of Financial Economics 29, 3–34.

Hansen, L.P., Jagannathan, R., 1991. Implications of security market data for models of dynamic economies. Journal of Political Economy 99, 225–262.

Harrison, M., Kreps, D., 1979. Martingales and arbitrage in multiperiod securities markets. Journal of Economic Theory 20, 381–408.

Harvey, C., Whaley, R., 1992. Market volatility prediction and the efficiency of the S&P 100 index option market. Journal of Financial Economics 31, 43–73.
Jackwerth, J.C., 2000. Recovering risk aversion from option prices and realized returns. Working paper, London Business School.
Jackwerth, J.C., Rubinstein, M., 1996. Recovering probability distributions from contemporary security prices. Journal of Finance 51, 1611–1631.
Jarrow, R., Rudd, A., 1982. Approximate option valuation for arbitrary stochastic processes. Journal of Financial Economics 10, 347–369.
Kloeden, P.E., Platen, E., 1992. Numerical Solution of Stochastic Differential Equations. Springer, Berlin, Germany.
Lo, A.W., Wang, J., 1995. Option pricing models when asset returns are predictable. Journal of Finance 50, 87–129.
Longstaff, F., 1995. Option pricing and the martingale restriction. Review of Financial Studies 8, 1091–1124.
Lamoureux, C.G., Lastrapes, W.D., 1993. Forecasting stock-return variance: toward an understanding of stochastic implied volatilities. Review of Financial Studies 6, 293–326.
Macbeth, J.D., Merville, L.J., 1979. An empirical examination of the Black–Scholes call option pricing model. Journal of Finance 34, 1172–1186.
Mehra, R., Prescott, E.C., 1985. The equity premium: A puzzle. Journal of Monetary Economics 15, 145–161.
Merton, R., 1973. Rational theory of option pricing. Bell Journal of Economics and Management Science 4, 141–183.
Merton, R., 1976. Option Pricing when Underlying Stock Returns are Discontinuous. Journal of Financial Economics 3, 125–144.
Ross, S., 1976. Options and efficiency. Quarterly Journal of Economics 90, 75–89.
Rubinstein, M., 1985. Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active CBOE option classes from August 23, 1976 through August 31, 1978. Journal of Finance 40, 455–480.
Rubinstein, M., 1994. Implied binomial trees. Journal of Finance 49, 771–818.
Schmalensee, R., Trippi, R.R., 1978. Common stock volatility expectations implied by option premia. Journal of Finance 33, 129–147.
Shimko, D., 1993. Bounds of probability. RISK 6, 33–37.
Stutzer, M., 1996. A simple nonparametric approach to derivative security valuation. Journal of Finance 51, 1633–1652.
Wand, M.P., Jones, M.C., 1995. Kernel Smoothing. Chapman & Hall, London, UK.