An empirical test of pricing kernel monotonicity

Brendan K. Beare and Lawrence D. W. Schmidt*

Department of Economics, University of California, San Diego

June 30, 2014

Abstract

A large class of asset pricing models predicts that securities which have high payoffs when market returns are low tend to be more valuable than those with high payoffs when market returns are high. More generally, we expect the projection of the stochastic discount factor on the market portfolio – that is, the discounted pricing kernel evaluated at the market portfolio – to be a monotonically decreasing function of the market portfolio. Numerous recent empirical studies appear to contradict this prediction. The nonmonotonicity of empirical pricing kernel estimates has become known as the pricing kernel puzzle. In this paper we propose and apply a formal statistical test of pricing kernel monotonicity. We apply the test using seventeen years of data from the market for European put and call options written on the S&P 500 index. Statistically significant violations of pricing kernel monotonicity occur in a substantial proportion of months, suggesting that observed nonmonotonicities are unlikely to be the product of statistical noise.

*We thank seminar participants at Johns Hopkins University, Pennsylvania State University, Princeton University, UC Irvine, UC Riverside, UC San Diego and the University of Chicago for helpful comments. We also thank Peter Hansen for providing the realized volatility data and related advice, Ana Monteiro for providing MatLab code for the cubic spline estimator of the risk neutral density, and Yacine Aït-Sahalia, Tim Bollerslev, Jong-Myun Moon, Allan Timmermann and Rossen Valkanov for helpful discussions.

1 Introduction

A ubiquitous intuition in asset pricing theory is that securities which covary positively with a broad market portfolio tend to be less valuable than those which covary negatively. When the prices of contingent claims written on a market portfolio satisfy this principle, we expect the projection of the stochastic discount factor on the market portfolio – that is, the discounted pricing kernel evaluated at the market portfolio – to be a monotonically decreasing function of the market portfolio. In this paper we investigate whether this general principle is in fact consistent with observed option prices and market returns.

Let us take a moment to explain what we mean here by stochastic discount factor, pricing kernel, and projection. Many authors take the terms stochastic discount factor and pricing kernel to be synonymous, but they have distinct meanings in much of the recent literature on pricing kernel monotonicity, and this is sometimes the source of confusion. The *stochastic discount factor* at time t is a random variable M_t, known at time $t+1$, such that the current price Y_t of any asset satisfies $Y_t = E_t(M_t Y_{t+1})$, where E_t denotes the (objective) expectation conditional on information available at time t. This is a standard concept in arbitrage-free models of asset pricing; see e.g. Cochrane (2001). Let S_t denote the price of the market portfolio at time t. The *projection of the stochastic discount factor on the market portfolio* is $M_t^* = E_t(M_t|S_{t+1})$, the expected value of the stochastic discount factor at time t, conditional on the price of the market portfolio at time $t+1$. Letting r_t denote the one-period risk-free interest rate at time t, we may write $M_t^* = (1+r_t)^{-1}\pi_t(S_{t+1})$ for some uniquely determined function π_t dependent on information available at time t. We shall refer to this function π_t as the *pricing kernel*.

Consider a contingent claim written at time t on the market portfolio whose price at time $t+1$ is given by $Y_{t+1} = f_t(S_{t+1})$ for some payoff function f_t. The price of such an asset at time t satisfies

$$Y_t = E_t(M_t f_t(S_{t+1})) = \frac{1}{1+r_t}\int_0^\infty f_t(x)q_t(x)\mathrm{d}x, \tag{1.1}$$

where q_t is the one-period *risk neutral density* of the market portfolio at time t. Alterna-

tively, applying the law of iterated expectations we may write

$$Y_t = \frac{1}{1+r_t} E_t(\pi_t(S_{t+1}) f_t(S_{t+1})) = \frac{1}{1+r_t} \int_0^\infty f_t(x)\pi_t(x)p_t(x)\mathrm{d}x, \tag{1.2}$$

where p_t is the one-period *physical density* of the market portfolio at time t; that is, the (objective) conditional probability density function of S_{t+1} given information available at time t. Since (1.1) and (1.2) hold for any payoff function f_t, it must be the case that $\pi_t = q_t/p_t$. The pricing kernel is therefore the ratio of the risk neutral and physical densities of the market portfolio.

If the ubiquitous intuition referred to in our opening paragraph is correct, the pricing kernel π_t should be a nonincreasing function. This was made clear in an important contribution by Dybvig (1988), who showed that if π_t is not nonincreasing, then it is possible to obtain the payoff distribution of the market portfolio at a reduced price by using contingent claims to reallocate payoffs from good states of the world to bad states of the world. More precisely, if π_t is not nonincreasing then we can find a payoff function f_t – necessarily decreasing over some region – such that $f_t(S_{t+1})$ and S_{t+1} have the same (objective) probability distribution conditional on information available at time t, and yet $E_t(M_t f_t(S_{t+1})) < E_t(M_t S_{t+1})$. Further discussion of this result, and related results, are given by Beare (2011).

Though intuitively appealing, pricing kernel monotonicity was drawn into question by three empirical studies published in the early 2000s: Aït-Sahalia and Lo (2000), Jackwerth (2000) and Rosenberg and Engle (2002). Each of these studies produced a flexible empirical estimate of the pricing kernel for the S&P 500 index at one or more dates between 1988 and 1995. All three studies obtained estimated pricing kernels that were monotonically decreasing at high and low levels of markets returns, but increasing over an intermediate range of returns.[1] Jackwerth (2000) drew particular attention to the curious shape of the estimated pricing kernel, emphasizing that it was inconsistent with the existence of a risk averse representative agent. Even without a representative agent, in the context of a one-period (myopic) portfolio choice problem, pricing kernel nonmonotonic-

[1] In the paper by Rosenberg and Engle (2002), the estimates obtained using the orthogonal polynomial specification are nonmonotone, while the estimates obtained using the power specification are monotonically decreasing by construction.

ity remains puzzling, as there cannot exist even a single individual agent who rationally chooses to invest all their wealth in the market portfolio, or in any security or combination of securities whose payoff is a monotonically increasing function of the market return. Brown and Jackwerth (2012) coined the term *pricing kernel puzzle* in reference to the apparent nonmonotonicity of pricing kernel estimates.[2] Chabi-Yo et al. (2008) use the term *risk aversion puzzle* to refer to the same phenomenon. Hens and Reichlin (2012) provide a general review of the literature on the pricing kernel puzzle, and we refer the reader to their paper for additional discussion and references.

There is an important distinction between the empirical study of Rosenberg and Engle (2002) and those of Aït-Sahalia and Lo (2000) and Jackwerth (2000). In the latter two papers, the authors estimate p_t by applying a kernel density estimator to historical market returns. Since the estimator of p_t does not take into account measurable changes in market volatility, it is perhaps better viewed as an estimate of the *unconditional* probability density function of the market portfolio S_{t+1}. Rosenberg and Engle (2002) instead use an estimator of p_t that incorporates current information on market volatility. In general we should only expect the ratio $\pi_t = q_t/p_t$ to be nonincreasing when the physical density p_t properly conditions on information available at time t. This point was made forcefully by Chabi-Yo et al. (2008), who argued that the nonmonotonicity of empirical pricing kernel estimates might be explained by a failure to condition on relevant state variables in the estimation of p_t.

In this paper we seek to determine whether the observed departures from monotonicity in empirical pricing kernel estimates are statistically significant. The pricing kernel estimates appearing in Aït-Sahalia and Lo (2000), Jackwerth (2000) and Rosenberg and Engle (2002) are not accompanied by uniform confidence bands, so it is not clear whether their nonmonotonicity can be attributed to statistical noise. The same is true for more recent pricing kernel estimates appearing in Barone-Adesi et al. (2008, 2012). In these two papers, the estimated pricing kernels are not in fact monotone, yet the authors argue that they provide evidence in favor of pricing kernel monotonicity because their general shape appears to be roughly monotone. A formal statistical test of monotonicity should be able to provide a rigorous basis for statements of this kind. Indirect evidence in favor of pricing

[2] An early version of Brown and Jackwerth (2012) was circulated in 2000. The term *pricing kernel puzzle* appears in a number of studies published during the following decade.

kernel nonmonotonicity has been documented by Constantinides et al. (2009), who found that option prices appear to violate bounds implied by pricing kernel monotonicity, and by Bakshi et al. (2010), who found that the average returns on call options written on the S&P 500 index tend to be decreasing in the strike price – a pattern suggestive of a U-shaped pricing kernel. On the other hand, Chaudhuri and Schroder (2010) found that options written on individual stocks exhibited average returns consistent with a monotone pricing kernel.

The contribution of this paper is a statistical test of pricing kernel monotonicity that can account for uncertainty about the risk neutral and physical densities associated with the market portfolio. Our test is an extension of a procedure developed by Carolan and Tebbs (2005) and Beare and Moon (2014). We consider two test statistics based on the L_1 and L_2 distances between an estimated ordinal dominance curve (defined below) and its least concave majorant, and apply our procedure using data on 180 cross sections of European put and call options written on the S&P 500 index in distinct months between 1997 and 2013, with times to maturity of between 18 and 22 trading days. Using the L^1 (L^2) test statistic, we reject the null hypothesis of monotonicity 37% (46%) of the time at the 5% significance level, and 22% (26%) of the time at the 1% significance level.

Our paper is similar in spirit to recent papers by Golubev et al. (2014) and Härdle et al. (2014) that also attempt to formally assess the statistical significance of apparent violations of pricing kernel monotonicity. Using methods different to ours, those papers identify statistically significant violations of pricing kernel monotonicity using German options data during the years 2000–2006. Here we identify statistically significant violations using US options data during the years 1997–2013. One major advantage of the approach taken here is that, similar to Rosenberg and Engle (2002), we incorporate current information on market volatility in our estimation of p_t. For dates between 1975 and 1996, we obtain volatilities using the MiDaS estimator of Ghysels et al. (2005, 2006). For dates between 1997 and 2013, when high-frequency intraday data are available, we obtain volatilities using the Real-GARCH model of Hansen et al. (2012). Golubev et al. (2014) and Härdle et al. (2014) do not account for market volatility or other state variables in their estimation of p_t, so that, similar to Aït-Sahalia and Lo (2000) and Jackwerth (2000), their estimates of p_t are more properly interpreted as estimates of the unconditional physical density of the market portfolio. This means that, in periods of high or low volatility,

there may be a mismatch between the scales of their estimated risk neutral and physical densities, leading to spurious rejections of pricing kernel monotonicity.

The remainder of this paper is structured as follows. Section 2 provides a relatively nontechnical discussion of the statistical framework we adopt, showing how the approach of Carolan and Tebbs (2005) and Beare and Moon (2014) may be applied in the present context. Details of how we estimate the risk neutral and physical distributions, and compute critical values for our test statistics, are given in Section 3. Section 4 contains our empirical results. Detailed results are provided for two dates at which our tests are implemented, and a summary of the results obtained at all 180 dates in the sample period. We also subject our results to a number of robustness checks. In Section 5 we give some final thoughts and conclude. In a supplemental appendix we state and prove technical results concerning the asymptotic properties of our statistical procedure, and provide additional details about the dataset used to obtain our empirical results.

2 Statistical framework

In this section we provide a basic outline of the statistical framework used to obtain our empirical results. Details are given in the supplemental appendix, and we refer to results given there when appropriate. A reader who is primarily concerned with our empirical results should be able to understand the gist of our approach from the discussion in this section and the following section.

The pricing kernel puzzle concerns the shape of the ratio of the risk neutral and physical densities governing the payoff of some base asset – typically, a market index – at a given future date. Let[3] Q and P be two cumulative distribution functions (cdfs) on the real line with $P(x) = Q(x) = 0$ for all $x \leq 0$. The cdf P, referred to as the *physical distribution*, describes the value after one period of a \$1 investment in the base asset. It should be interpreted as being conditional on all information available at the time of investment. The cdf Q, referred to as the *risk neutral distribution*, determines the price of derivative contracts delivering a payoff after one period determined by the value of the base asset at

[3]In what follows, to simplify notation we drop the the t subscripts that were applied to the pricing kernel and risk neutral and physical distributions in the Introduction.

that time. Such contracts have price equal to their discounted expected payoff under Q.

We assume that the cdfs Q and P admit continuous probability density functions (pdfs) q and p, with common support. The pricing kernel is the density ratio $\pi(x) = q(x)/p(x)$, defined over the common support of q and p. We wish to test whether π is nonincreasing. Patton and Timmermann (2010) discuss a variety of ways to set up the null and alternative hypotheses in statistical tests of monotonicity. We shall adopt the following formulation.

$$\mathrm{H}_0 : \pi \text{ is nonincreasing}; \quad \mathrm{H}_1 : \pi \text{ is not nonincreasing}.$$

The null hypothesis H_0 is composite, meaning that it admits multiple pricing kernels π. Consistent with Carolan and Tebbs (2005), Golubev et al. (2014) and Beare and Moon (2014), we shall choose critical values for our test statistics that deliver the correct asymptotic size at a particular choice of nonincreasing pricing kernel: $\pi = 1$. This is the only choice of π that is constant, since q and p must integrate to one. When π is constant it has exactly zero derivative everywhere; we wish to reject the null hypothesis if there is some region over which π has positive derivative. Therefore, in an intuitive sense, $\pi = 1$ comes closer to violating H_0 than any other nonincreasing choice of π. With a suitably chosen test statistic, the rejection rate achieved at other nonincreasing choices of π should be no greater than the rejection rate at $\pi = 1$. Beare and Moon (2014) provide a more detailed discussion of this subject.

The approach to monotonicity testing proposed by Carolan and Tebbs (2005) is based on an equivalence between the monotonicity of π and the concavity of a function called the *ordinal dominance curve*. Given our cdfs Q and P, the corresponding ordinal dominance curve is the map $\phi : [0,1] \to [0,1]$ given by

$$\phi(u) = Q\left(P^{-1}(u)\right), \quad u \in [0,1].$$

Here, $P^{-1}(u) = \inf\{x : P(x) \geq u\}$, the quantile function for P. When q and p have common support, ϕ is continuous and nondecreasing with $\phi(0) = 0$ and $\phi(1) = 1$. It turns out that π is nonincreasing if and only if ϕ is concave. To see this, differentiate both sides of the equality $Q(x) = \phi(P(x)))$ to obtain $\pi(x) = \phi'(P(x))$. Since P is nondecreasing and continuous, we find that π is nonincreasing if and only if ϕ' is nonincreasing. One may therefore consider testing H_0 against H_1 by constructing a test statistic that is in some

sense an empirical measure of the nonconcavity of ϕ.

In Figure 2.1 we provide three examples of pairs of pdfs, their ratios, and the corresponding ordinal dominance curves. It may be helpful to think of the light pdfs as risk neutral densities and the dark pdfs as physical densities. In the first row, the two pdfs are normal with different means and the same variance, with the physical density shifted to the right of the risk neutral density. In this case, the pricing kernel is monotone decreasing, and the ordinal dominance curve is concave. In the second row the two pdfs are normal with equal means, but the variance of the risk neutral density is greater than the variance of the physical density. In this case the pricing kernel is U-shaped, broadly consistent with the empirical findings of Bakshi et al. (2010), and the ordinal dominance curve is not concave. In the third row the two pdfs are the risk neutral and physical pdfs estimated by Jackwerth (2000) for monthly S&P 500 returns on April 15, 1992. In this case the pricing kernel is decreasing at the extremes but nondecreasing around the center of the return distribution, and the ordinal dominance curve again fails to be concave.

In our empirical application we shall estimate Q using a cross-section of current option prices, and P using a time series of market returns, interest rates and volatilities. Let m denote the number of observed option prices, let n denote the length of the time series, and let Q_m and P_n denote our estimates of Q and P. Carolan and Tebbs (2005) and Beare and Moon (2014) take Q_m and P_n to be the empirical distribution functions of independent samples of size m and n drawn from Q and P. Here we employ more complicated estimators. Details about the estimators of Q and P used in our application are provided in Section 3. For the purposes of developing an asymptotic approximation to the statistical behavior of Q_m and P_n, and of the test statistics defined below, we will assume that $m \to \infty$ and $n \to \infty$ simultaneously, with $n/(m+n) \to \lambda$ for some $\lambda \in (0,1)$.

Given our estimated cdfs Q_m and P_n, we construct an estimated ordinal dominance curve $\phi_{m,n} : [0,1] \to [0,1]$ as follows:

$$\phi_{m,n}(u) = Q_m(P_n^{-1}(u)), \quad u \in [0,1].$$

The *least concave majorant* of $\phi_{m,n}$, denoted $\mathcal{M}\phi_{m,n}$, is the smallest concave function on $[0,1]$ that is everywhere equal to or greater than $\phi_{m,n}$. Figure 2.2 provides an example of an estimated ordinal dominance curve $\phi_{m,n}$ and its least concave majorant $\mathcal{M}\phi_{m,n}$.

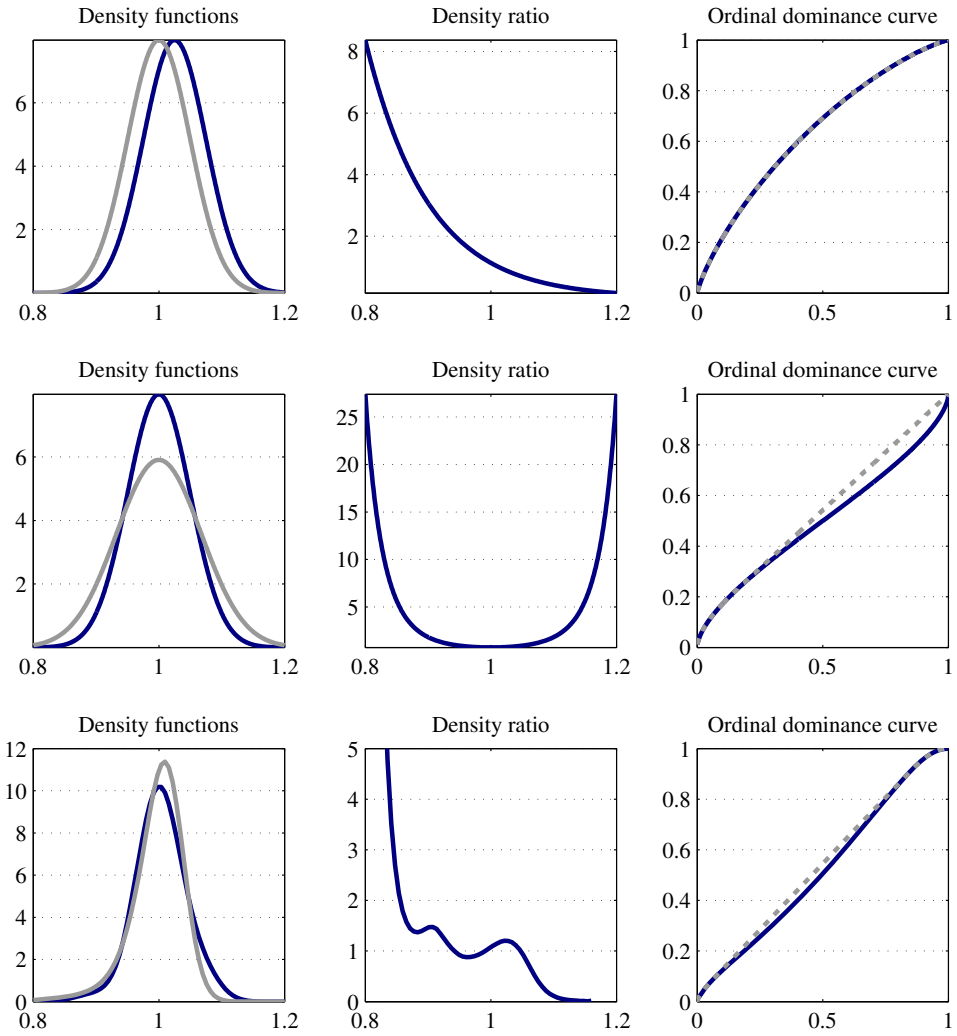

Figure 2.1: Density ratios and ordinal dominance curves.

The densities used in the first row are Gaussian with different means and equal variances. The densities used in the second row are Gaussian with equal means and different variances. The densities used in the third row are the estimated risk neutral and physical densities reported by Jackwerth (2000) for monthly S&P 500 returns on April 15, 1992. In the third column of panels, ordinal dominance curves are given by dark solid lines, while lighter dashed lines indicate their least concave majorants.

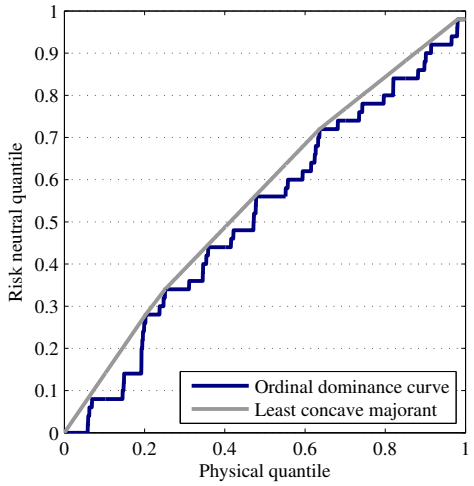

Figure 2.2: The least concave majorant of an estimated ordinal dominance curve.

The estimated ordinal dominance curve $\phi_{m,n}$ displayed here was constructed using the empirical distribution functions of two independent samples from the standard normal distribution, each with 50 observations. In this case the true ordinal dominance curve is $\phi(u) = u$, and the implied density ratio is $\pi(x) = 1$.

The least concave majorant may be efficiently computed using the `convhull` command in MATLAB, or comparable functions in other numerical software packages.

We will use the following statistics $M^1_{m,n}$ and $M^2_{m,n}$ for testing H_0 against H_1:

$$M^1_{m,n} = \sqrt{\frac{mn}{m+n}} \int_0^1 (\mathcal{M}\phi_{m,n}(u) - \phi_{m,n}(u))\,\mathrm{d}u, \tag{2.1}$$

$$M^2_{m,n} = \sqrt{\frac{mn}{m+n}} \left(\int_0^1 (\mathcal{M}\phi_{m,n}(u) - \phi_{m,n}(u))^2\,\mathrm{d}u \right)^{1/2}. \tag{2.2}$$

$M^1_{m,n}$ is proportional to the L^1-distance between $\phi_{m,n}$ and $\mathcal{M}\phi_{m,n}$, while $M^2_{m,n}$ is proportional to the L^2-distance between $\phi_{m,n}$ and $\mathcal{M}\phi_{m,n}$. The statistic $M^1_{m,n}$ was proposed by Carolan and Tebbs (2005) for testing whether ϕ is concave. Those authors also proposed a second statistic based on the maximum distance between $\phi_{m,n}$ and $\mathcal{M}\phi_{m,n}$. Beare and Moon (2014) showed that this second statistic is poorly behaved and instead recommended that suitable statistics should be formed by considering the L^ν-distance between

$\phi_{m,n}$ and $\mathcal{M}\phi_{m,n}$, with $1 \leq \nu \leq 2$. In this paper we confine ourselves to the cases $\nu = 1$ and $\nu = 2$. Numerical simulations reported by Beare and Moon (2014) suggest that the choice $\nu = 2$ may yield higher power than $\nu = 1$ against some relevant alternatives, and in the empirical results reported in Section 4.3 we do indeed obtain more frequent rejections of pricing kernel monotonicity using the test statistic $M^2_{m,n}$ than we do with $M^1_{m,n}$.

The asymptotic statistical properties of $M^1_{m,n}$ and $M^2_{m,n}$ depend on the behavior of Q_m and P_n, viewed as sequences of random cdfs on the real line. Naturally, the behavior of Q_m and P_n depends on how they are constructed from the observed data, and details of this construction are deferred to Section 3. For now, it suffices for us to assume that, as $m, n \to \infty$ with $n/(m+n) \to \lambda$, the normalized estimation errors $\sqrt{m}(Q_m(x)-Q(x))$ and $\sqrt{n}(P_n(x)-P(x))$ converge weakly to continuous random functions of x. More specifically, we assume that

$$\begin{pmatrix} \sqrt{m}(Q_m(x) - Q(x)) \\ \sqrt{n}(P_n(x) - P(x)) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \xi(Q(x)) \\ \zeta(P(x)) \end{pmatrix}, \tag{2.3}$$

where $\rightsquigarrow$ denotes weak convergence, and ξ and ζ are independent continuous random functions on $[0, 1]$. A more formal statement of this condition is given in Theorem A.1 in the supplemental appendix. If Q_m and P_n are the empirical distribution functions of independent samples of size m and n drawn from Q and P, as they are in the case considered by Carolan and Tebbs (2005) and Beare and Moon (2014), then Donsker's theorem ensures that (2.3) holds with ξ and ζ independent standard Brownian bridges on $[0, 1]$. Other choices of ξ and ζ may accommodate other estimators of Q and P. In the supplemental appendix we identify the processes ξ and ζ corresponding to the risk neutral and physical estimators proposed in Section 3; alternative derivations may be required for other estimators.

The assumed independence of ξ and ζ means that *uncertainty about* Q is asymptotically independent of *uncertainty about* P. This condition appears reasonable in our application, because uncertainty about Q derives from random pricing errors present in *current* option price data, whereas uncertainty about P derives primarily from random variation in *historical* asset returns, the bulk of which become increasingly distant as the time series length n increases. Historical asset returns are used to estimate the *shape* of P. A secondary source of uncertainty about P is uncertainty about the *scale* of P; that is, the present level of physical volatility. We assume that the physical volatility can be accu-

rately measured using the RealGARCH model of Hansen et al. (2012), which augments a standard GARCH model using high frequency intra-day data. If uncertainty about the scale of P is small relative to uncertainty about the shape of P, our assumption that ξ and ζ are independent should not be too unreasonable.[4]

We reject H_0 in favor of H_1 when our test statistic $M^1_{m,n}$ or $M^2_{m,n}$ exceeds a suitably chosen critical value. Theorem A.1 in the supplemental appendix establishes that, when π is constant and $m, n \to \infty$ with $n/(m+n) \to \lambda$, we have

$$M^1_{m,n} \xrightarrow{d} \int_0^1 (\mathcal{M}\mathscr{G}(u) - \mathscr{G}(u))\,\mathrm{d}u \quad \text{and} \tag{2.4}$$

$$M^2_{m,n} \xrightarrow{d} \left(\int_0^1 (\mathcal{M}\mathscr{G}(u) - \mathscr{G}(u))^2\,\mathrm{d}u \right)^{1/2}, \tag{2.5}$$

where $\mathscr{G}(u) = \lambda^{1/2}\xi(u) - (1-\lambda)^{1/2}\zeta(u)$. If the random behavior of $\mathscr{G}$ is known,[5] critical values for our test statistics may be extracted from the $(1-\alpha)$-quantiles of the limit distributions given in (2.4) and (2.5). This will deliver an asymptotic rejection rate of α when π is constant, and of no greater than α when π is nonincreasing. If the random behavior of $\mathscr{G}$ is not known, it may still be possible to obtain asymptotically valid critical values for our test statistics by consistently estimating the $(1-\alpha)$-quantiles of the limit distributions in (2.4) and (2.5). We discuss this further in Section 3.3, after our estimators Q_m and P_n have been introduced.

3 Construction of test statistics and critical values

In Section 2 we outlined the statistical framework we shall use for testing pricing kernel monotonicity. We assumed that our risk neutral and physical estimators Q_m and P_n satisfied the weak convergence condition (2.3), but did not explain precisely how they are computed in our application. In Section 3.1 we provide details about the risk neutral

[4]There is an easier way to justify the independence of ξ and ζ: we may simply assume that uncertainty about Q is asymptotically negligible relative to uncertainty about P. This assumption was used by Golubev et al. (2014) to justify their test of pricing kernel monotonicity, and in our framework amounts to assuming that $\xi = 0$, thereby trivially assuring the independence of ξ and ζ.

[5]For instance, when ξ and ζ are independent standard Brownian bridges, as in Carolan and Tebbs (2005) and Beare and Moon (2014), $\mathscr{G}$ is itself a standard Brownian bridge.

estimator used in our application, and in Section 3.2 we provide details about our physical estimator. We identify continuous random functions ξ and ζ such that, under suitable conditions, our estimators satisfy (2.3). In Section 3.3 we explain how critical values are calculated for our test statistics.

3.1 Risk neutral distribution estimation

Nonparametric methods for estimating the risk neutral density were first introduced by Jackwerth and Rubinstein (1996) and Aït-Sahalia and Lo (1998). A large number of alternative estimators have been proposed in subsequent literature. In our application we employ the estimator of Monteiro et al. (2008), which involves approximating the risk neutral density with a cubic spline. Two properties of this estimator make it ideally suited to our statistical framework: it is *flexible*, but *parametric*. Flexibility is crucial because, to identify potential departures from pricing kernel monotonicity, we need an estimator that is sufficiently flexible to accommodate irregularities in the shape of the risk neutral distribution. We can achieve flexibility with the cubic spline by employing a modest number of spline knots. At the same time, the parametric nature of the cubic spline makes it possible for us to obtain regularity conditions under which the weak convergence postulated in (2.3) is satisfied. This is typically not possible with fully nonparametric estimators of the risk neutral distribution. For instance, the estimator proposed by Aït-Sahalia and Lo (1998) delivers an asymptotically normal estimate of the risk neutral cdf at any given point in its support, but we do not have weak convergence in the functional sense, as required in (2.3). Technically, the sequence of risk neutral cdf estimates fails to satisfy a tightness condition, ruling out weak convergence. The parametric nature of the cubic spline allows us to circumvent this difficulty.

The data used for estimating the risk neutral distribution at a given date consist of m quadruplets (d_i, s_i, w_i, v_i), $i = 1, \ldots, m$, describing the observed characteristics of European put and call options written on our base asset, expiring after one period. d_i is equal to zero if the ith option is a call and one if the ith option is a put, s_i is the strike at which the ith option is written, w_i is the price of the ith option, and v_i is the trading volume of options with strike s_i and of type d_i on the date in question. We also observe the one period risk-free interest rate r that applies at the present date. Further details about how

our data are constructed are provided in Section 4.1.

Following Monteiro et al. (2008), we assume that the risk neutral density q is a cubic spline (i.e., a smooth piecewise cubic polynomial) with fixed knots $x_0 < \cdots < x_k$. The choice of knots is arbitrary; in our application we employ an ad hoc data dependent method to choose the knot locations. Given the knots, q is specified up to a vector of parameters $\theta \in \mathbb{R}^{4k}$:

$$q(x;\theta) = \begin{cases} \theta_{4j-3} + \theta_{4j-2}x + \theta_{4j-1}x^2 + \theta_{4j}x^3 & \text{if } x \in (x_{j-1}, x_j] \\ 0 & \text{if } x \notin (x_0, x_k]\,. \end{cases} \tag{3.1}$$

The parameter vector θ is restricted in such a way that q integrates to one, is twice continuously differentiable on (x_0, x_k), and is continuous and equal to zero at x_0 and x_k. These requirements amount to $3k$ linear equality restrictions on θ, which we write as $R\theta = \gamma$ using a suitably chosen $3k \times 4k$ matrix R and $3k \times 1$ vector γ. In addition to these linear equality restrictions, we require that q is nonnegative. This condition imposes a family of linear inequality restrictions on θ. We require

$$\theta_{4j-3} + \theta_{4j-2}x + \theta_{4j-1}x^2 + \theta_{4j}x^3 \geq 0 \tag{3.2}$$

for all $x \in (x_{j-1}, x_j]$ and all $j = 1, \ldots, k$. Let Θ denote the collection of $\theta \in \mathbb{R}^{4k}$ satisfying $R\theta = \gamma$ and the inequality restrictions in (3.2).

Monteiro et al. (2008) propose choosing $\theta \in \Theta$ to minimize a weighted sum of squared differences between the observed prices w_i and the prices implied by $q(\cdot;\theta)$. The trading volumes v_i are used to weight the different squared pricing errors. This approach is consistent with the idea that highly traded assets are more likely to be accurately priced. The estimator for the "true" value of θ, which we denote θ^*, can be written as

$$\theta_m = \operatorname*{argmin}_{\theta \in \Theta} \sum_{i=1}^{m} v_i \left(w_i - h(d_i, s_i; \theta)\right)^2, \tag{3.3}$$

where

$$h(d, s; \theta) = \frac{1}{1+r} \int_0^\infty \max\{(-1)^d (x - s), 0\} q(x;\theta) \mathrm{d}x.$$

Note that, in view of the spline model (3.1), $h(d, s; \theta)$ is a linear function of θ for each (d, s).

For each i, we may therefore write $h(d_i, s_i; \theta) = z_i'\theta$ for all θ, with z_i an element of $\mathbb{R}^{4k}$ determined by s_i, d_i, r, and the spline knots $x_0, \ldots, x_k$. Estimation of θ^* thus amounts to weighted linear regression subject to linear equality and inequality restrictions. Monteiro et al. (2008) provide a fast algorithm to compute θ_m based on the method of semidefinite programming. We shall not repeat the details here.

Given the estimated risk neutral density $q(\cdot; \theta_m)$, our estimate for the risk neutral distribution $Q(\cdot) = Q(\cdot; \theta^*)$ is given by

$$Q_m(x) = Q(x; \theta_m) = \int_0^x q(y; \theta_m) \mathrm{d}y.$$

To implement our test for pricing kernel monotonicity, we need to characterize the asymptotic statistical behavior of Q_m. Specifically, we need to identify a continuous random function ξ on $[0, 1]$ such that $\sqrt{m}(Q_m(x) - Q(x)) \rightsquigarrow \xi(Q(x))$, consistent with (2.3). Theorem A.2 in the supplemental appendix establishes that, under suitable conditions, we may take

$$\xi(u) = \left(\frac{\partial}{\partial \theta} Q(Q^{-1}(u; \theta^*); \theta) \bigg|_{\theta = \theta^*} \right)' \Psi^{1/2} N, \tag{3.4}$$

where N is a vector of $4k$ independent standard normal random variables, and Ψ is the symmetric positive semidefinite $4k \times 4k$ matrix defined in equation (A9) in the supplemental appendix.

3.2 Physical distribution estimation

We seek to estimate P, the conditional distribution of the payoff at time $t + 1$ of a \$1 investment in the S&P 500 index at time t. The option prices used to estimate Q are quoted at time t, with the options expiring at time $t + 1$. To estimate P we employ a historical time series of n triples (X_j, r_j, σ_j), $j = 1, \ldots, n$, covering the span 1975–2013. The date t corresponds to one of these time periods.[6] An equal number of trading days fall between any consecutive dates labelled j and $j+1$. Depending on t, this is between 18 and 22 trading days; further details are given at the beginning of Section 4.3. X_j is the rate of

[6] In the robustness checks reported in Section 4.4 we also consider setting n equal to t, so that only data observed up to period t are used to estimate P.

return on an investment in the S&P 500 index from time j to time $j+1$. It is calculated as the natural logarithm of the price ratio between the two dates. r_j is the risk-free rate of interest from time j to time $j+1$. σ_j^2 is a measure of the conditional volatility of excess returns on the S&P 500 index from time j to time $j+1$, given information available at time j. It is constructed from daily (1975–1996) or high frequency intraday (1997–2013) historical return data. Additional details about the construction of volatilities, and our data more generally, are provided in Section 4.1.

P is the conditional cdf of $\exp(X_t)$ given r_t and σ_t. Our estimator of P is given by

$$P_n(x) = \frac{1}{n}\sum_{j=1}^{n} 1\left(\frac{\sigma_t}{\sigma_j}(X_j - r_j) + r_t \leq \ln x\right) \tag{3.5}$$

if $x > 0$, or $P_n(x) = 0$ if $x \leq 0$. This is the empirical cdf of $\exp(X_j^*)$, where $X_j^* = (\sigma_t/\sigma_j)(X_j - r_j) + r_t$ is a recentered, rescaled version of the rate of return X_j. To implement our test we must identify a continuous random function ζ on $[0,1]$ such that $\sqrt{n}(P_n(x) - P(x)) \rightsquigarrow \zeta(P(x))$, consistent with (2.3). In Theorem A.3 in the supplemental appendix, we identify conditions under which this weak convergence holds with ζ a standard Brownian bridge on $[0,1]$. The main requirement is that volatility adjusted excess returns are iid over time. Though an imperfect approximation to reality, this condition compares favorably to the recent empirical studies of pricing kernel monotonicity by Golubev et al. (2014) and Härdle et al. (2014), in which time variation in conditional market volatility is ignored. We impose no structure on the shape of P beyond continuity.

3.3 Calculating critical values

In (2.4) and (2.5) we stated the limit distributions of our test statistics $M_{m,n}^1$ and $M_{m,n}^2$ when $\pi = 1$ in terms of the random process $\mathscr{G}(u) = \lambda^{1/2}\xi(u) - (1-\lambda)^{1/2}\zeta(u)$. For the estimators Q_m and P_n defined in Sections 3.1 and 3.2, ξ is a finite dimensional Gaussian process given by (3.4), and ζ is a standard Brownian bridge. As discussed in Section 2, the two processes are assumed independent. To obtain critical values c_1 and c_2 for $M_{m,n}^1$ and $M_{m,n}^2$ yielding a test with asymptotic size α we employ the following procedure.

1. Randomly generate ζ, a standard Brownian bridge.

2. Randomly generate ξ in accordance with (3.4). Since the distribution of ξ depends on unknown parameters θ^* and Ψ, we substitute these with consistent estimators θ_m and Ψ_m. The former is defined in (3.3) and the latter in equation (A.9) in the supplemental appendix.

3. Calculate $\mathscr{G} = \lambda^{1/2}\xi - (1-\lambda)^{1/2}\zeta$. The parameter λ is set equal to its finite sample analogue $n/(m+n)$.

4. Evaluate

$$M^1 = \int_0^1 (\mathcal{M}\mathscr{G}(u) - \mathscr{G}(u))\,\mathrm{d}u \quad \text{and} \quad M^2 = \left(\int_0^1 (\mathcal{M}\mathscr{G}(u) - \mathscr{G}(u))^2\,\mathrm{d}u\right)^{1/2}.$$

5. Repeat steps 1 through 4 a large number of times, so as to obtain a large number of independent realizations of M^1 and M^2. Set c_1 and c_2 equal to the $(1-\alpha)$ quantiles of the simulated distributions of M^1 and M^2.

4 Empirical results

The presentation of our empirical results is divided into four subsections. In Section 4.1 we describe our dataset. In Section 4.2 we provide a detailed description of our results for two representative dates in our sample period – April 15, 2009, and April 17, 2013. In Section 4.3 we summarize our results for the full set of 180 dates. In Section 4.4 we report the outcome of a number of robustness checks.

4.1 Data

Our primary dataset consists of prices for European call and put options written on the S&P 500 from January 1997 to December 2013. Daily option prices were obtained from the data provider DeltaNeutral[7], which collects prices for options reported by the Options Price Reporting Authority[8] (OPRA). OPRA compiles information from a number

[7] `http://www.deltaneutral.com`.
[8] `http://www.opradata.com`.

of different exchanges in order to find a national option price. Our dataset consists of bid-ask averages of OPRA's reported prices for different options at the close of the market on each trading day, along with the corresponding trading volumes.

We exclude all options that do not have between 18 and 22 trading days to maturity, and nonzero trading volume. Tables A.1 and A.2, located in the supplemental appendix, summarize some key features of the remaining prices. In Table A.1 we report the maximum, minimum and average number of put and call option prices observed in the different months of each year. The average number of option prices increased from 63 in 1997 to 164 in 2013. In Table A.2 we report average implied volatilities by option moneyness, computed using the Black-Scholes formula. The figures exhibit the familiar volatility smile, where at-the-money options tend to have lower implied volatilities than their far in-the-money or (to a lesser extent) out-of-the-money counterparts.

For the benchmark results reported in Sections 4.2 and 4.3, we exclude options with moneyness outside ±15% from the risk neutral estimation procedure. In the robustness checks reported in Section 4.4, we also examine the impact of excluding calls with moneyness less than -3% and puts with moneyness greater than 3%. The effect is minimal because of the volume weighting used in our estimation procedure.

For dates between 1997–2013, our physical volatility measure is derived from a daily series of realized volatilities kindly provided by Peter R. Hansen for the years 1997–2009, and extended through 2013 using the Oxford-Man realized volatility library[9] (Heber et al., 2009). These realized volatilities were constructed using high frequency intraday data on the SPY, an exchange traded fund which tracks the S&P 500 index, using the realized kernel method described by Barndorff-Nielsen et al. (2008, 2009). Following Hansen et al. (2012), we used a log-linear RealGARCH(1,2) model to produce forward-looking volatilities from the historical SPY returns and realized volatilities.[10] In Figure 4.1 we graph the annualized realized volatility forecasts alongside the annualized volatilities from our cubic spline risk-neutral density estimators. The realized volatility forecasts are based on a 20-day volatility projection using the RealGARCH model, while the risk neutral volatilities are estimated from options with a time-to-maturity of between 18 and

[9] `http://realized.oxford-man.ox.ac.uk/data`.

[10] Code for estimating the log-linear RealGARCH(1,2) model is available at `http://qed.econ.queensu.ca/jae/datasets/hansen003`.

22 days, as discussed at the beginning of Section 4.3.

Volatility estimates based on intraday data are only available from 1997-2013. To extend the time series of standardized excess returns used in our estimator of the physical distribution, we constructed volatilities for the period 1975-1996 using the MiDaS estimator of Ghysels et al. (2006), which models the monthly variance as a parsimonious parametric function of lagged daily ranges for the S&P 500. Applying the MiDaS estimator to the period 1997-2013 yields volatility estimates that are generally very close to those obtained using the RealGARCH(1,2) model. Our baseline results reported in Sections 4.2 and 4.3 use the MiDaS volatilities for 1975-1996 and RealGARCH(1,2) volatilities for 1997-2013. In the robustness checks reported in Section 4.4 we include results obtained using the MiDaS volatilities for the full sample period 1975-2013 used to estimate the physical distribution. There is essentially no change.

Data for the risk-free interest rate were obtained from Kenneth French's website[11], which reports the daily return on the one-month Treasury bill rate (from Ibbotson Associates).

4.2 Representative examples

In Figure 4.2 we present detailed results for the dates April 15, 2009, and April 17, 2013. We chose these two dates from the 180 dates at which we apply our test because they provide nice examples of two pricing kernel shapes that recur frequently over our sample period and have been identified in prior studies. In the left column of panels in Figure 4.2 we graph the estimated risk neutral and physical densities, implied pricing kernel, and ordinal dominance curve for the date April 15, 2009, and in the right column of panels we graph the same curves for the date April 17, 2013.

For the date April 15, 2009, we see that the physical density has a sharper peak than the risk neutral density, while the risk neutral density is more disperse than the physical density. This results in a broadly U-shaped pricing kernel, similar to the findings of Bakshi et al. (2010). Care should be exercised when judging the shape of the pricing kernel, as sampling uncertainty may be large in the tails of either distribution. The corresponding

[11] `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

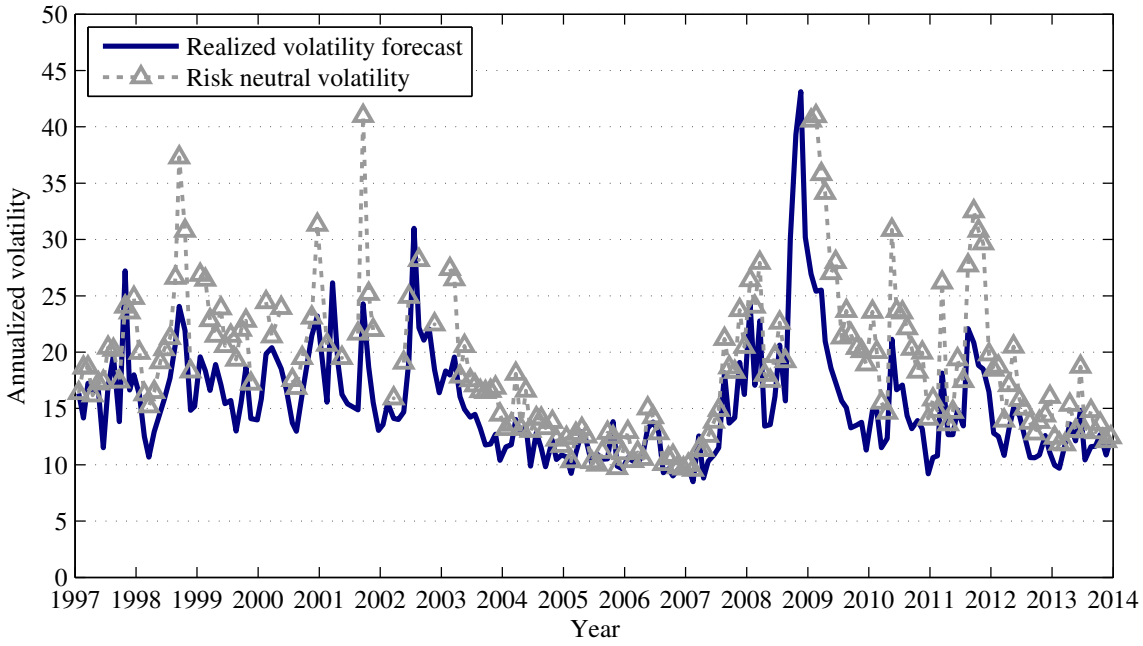

Figure 4.1: Forward looking realized and risk neutral volatilities.

Forward looking realized volatilities are calculated from intraday data using the log-linear RealGARCH(1,2) model of Hansen et al. (2012). Risk neutral volatilities are extracted from our cubic spline estimates of the risk neutral density at each of the 180 dates at which we apply our monotonicity tests.

ordinal dominance curve fails to be concave, and appears roughly convex to the right of the center of the physical distribution, consistent with a U-shaped pricing kernel. The L^1- and L^2-based test statistics have p-values of 0.051 and 0.007 respectively, so we may confidently reject the null hypothesis of pricing kernel monotonicity at this date.

For the date April 17, 2013, our estimated pricing kernel resembles the shape identified early on by Aït-Sahalia and Lo (2000), Jackwerth (2000) and Rosenberg and Engle (2002): it is increasing around the center of the physical distribution, and decreasing elsewhere. This causes the corresponding ordinal dominance curve to dip below its least concave majorant at mid and lower quantiles of the physical distribution. The L^1- and L^2-based test statistics have p-values of 0.109 and 0.048, so we have borderline statistically significant evidence against pricing kernel monotonicity at this date, depending on the test statistic

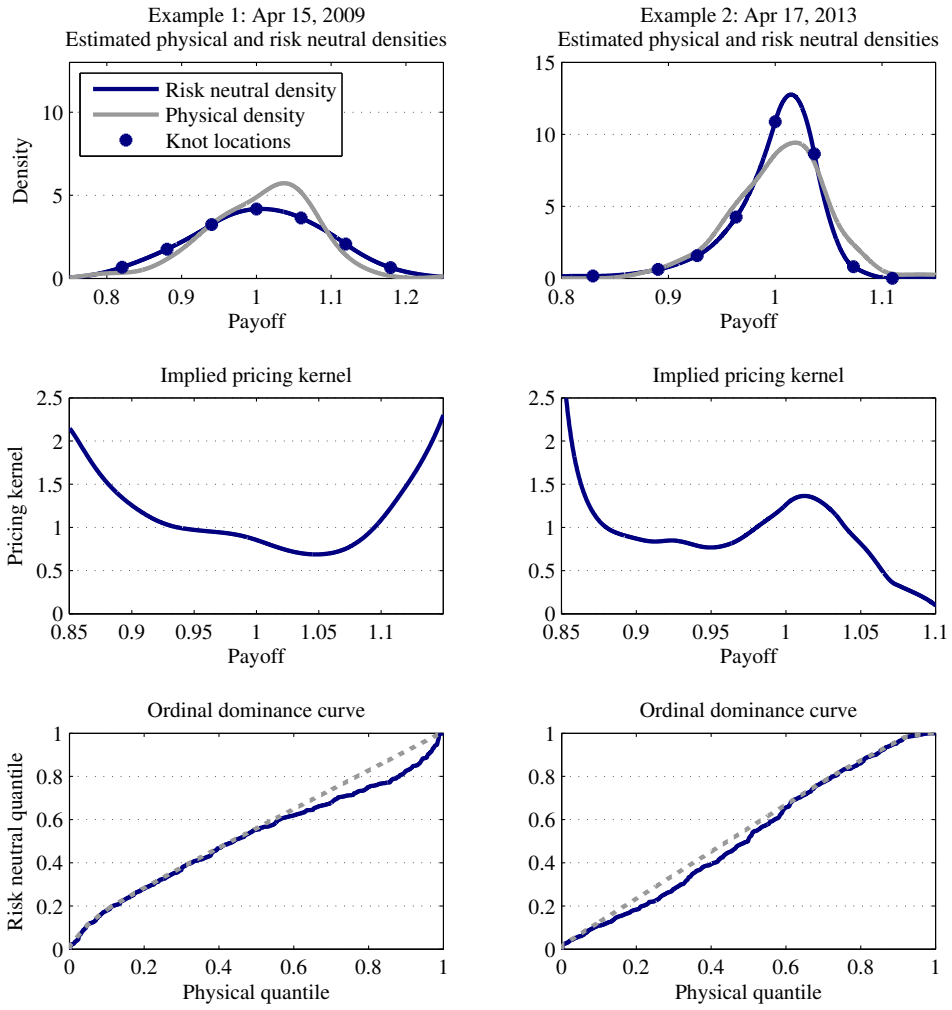

Figure 4.2: Estimated physical and risk neutral densities, pricing kernel, and ordinal dominance curve for two representative dates.

The top panel displays the estimated physical and risk-neutral density functions for two representative dates in our sample: April 15, 2009 (left panels) and April 17, 2013 (right panels). Knot locations for the cubic spline used to estimate the risk neutral density are indicated by solid dots. We used a standard kernel smoother to obtain the physical density from our estimated physical cdf, which is not differentiable. The center panel displays the estimated pricing kernel, which is the ratio of the two densities in the top panel. The bottom panel shows the corresponding ordinal dominance curve and its least concave majorant.

and desired significance level.

We used $k+1 = 11$ spline knots in the risk neutral estimation procedure. Their locations are indicated with solid dots in Figure 4.2; some of the 11 knots lie outside the range of the horizontal axes. We set the knot locations equal to $1 \pm a\sigma_t$, where σ_t^2 is the physical volatility for the date t at which we apply our test, and $a = 0, 0.75, 1.5, 2.25, 3.5, 5.5$. This procedure was used to choose knot locations for all 180 dates in our full sample.

4.3 Full sample results

For each month in the sample period, we select the date with the largest number of option prices with a single time-to-maturity of between 18 and 22 days, and test the monotonicity of the pricing kernel at that date, with that time-to-maturity. The sampling frequency of the time series used to estimate the physical distribution is chosen to match this time-to-maturity. We exclude 20 months in which there are no days with prices for at least 40 options with a common time-to-maturity of between 18 and 22 days, and the last four months of 2008 due to the extremely high realized volatility levels associated with the financial crisis. This leaves us with 180 dates at which we apply our tests.

Table 4.1 and Figure 4.3 present the results of our monotonicity tests. In Table 4.1 we report the rejection frequencies obtained with our two test statistics in each year, with 5% and 1% nominal significance levels. We find that, in nearly all years in our sample period, our test statistics exceed their critical values at a substantially higher rate than we would expect under the null hypothesis of pricing kernel monotonicity. Overall, our L^1 test rejects monotonicity 37% of the time at the 5% level, and 22% of the time at the 1% level. Our L^2 test rejects monotonicity 46% of the time at the 5% level, and 26% of the time at the 1% level.

In Figure 4.3 we plot the p-values for our test statistics over time. These p-values indicate the smallest nominal significance level at which we would reject the null hypothesis of pricing kernel monotonicity at a given date. Some p-values are extremely close to zero, indicating that we may confidently reject pricing kernel monotonicity at certain dates. Some intertemporal clustering of low p-values is apparent, indicating that there may be extended periods in which pricing kernel nonmonotonicity is more frequent, and other

periods in which it is less frequent.

The mean of the estimated risk neutral distribution, discounted by the risk-free rate, acts as a simple specification check on our cubic spline model. It should equal one if a direct investment in the S&P 500 index is correctly priced. Table A.3, located in the supplemental appendix, presents summary statistics for the moments of the estimated risk neutral densities by year. The discounted mean is in all cases very close to one. Table A.3 also summarizes the annualized volatilities associated with our estimated risk neutral distributions. These risk neutral volatilities were plotted against forward looking

	Number of	L^1 statistic		L^2 statistic	
Year	months	5%	1%	5%	1%
1997	12	33	8	42	17
1998	11	45	27	55	27
1999	11	36	36	45	36
2000	8	12	0	12	0
2001	6	50	50	50	50
2002	5	0	0	0	0
2003	11	9	9	18	9
2004	12	17	0	17	0
2005	12	33	17	42	17
2006	12	58	25	67	25
2007	12	58	33	83	50
2008	8	12	12	12	12
2009	12	50	25	67	58
2010	12	58	33	58	33
2011	12	67	50	67	50
2012	12	17	8	33	8
2013	12	42	25	58	33
All years	180	37	22	46	26

Table 4.1: Rejection rates (%) for monotonicity tests.

Figures indicate the percentage of months within a given year in which our tests yield a rejection of the null hypothesis of monotonicity. We report rejection rates for the L^1- and L^2-based test statistics, using nominal significance levels of 5% and 1%. In the final row we report rejection rates over the full sample of 180 months.

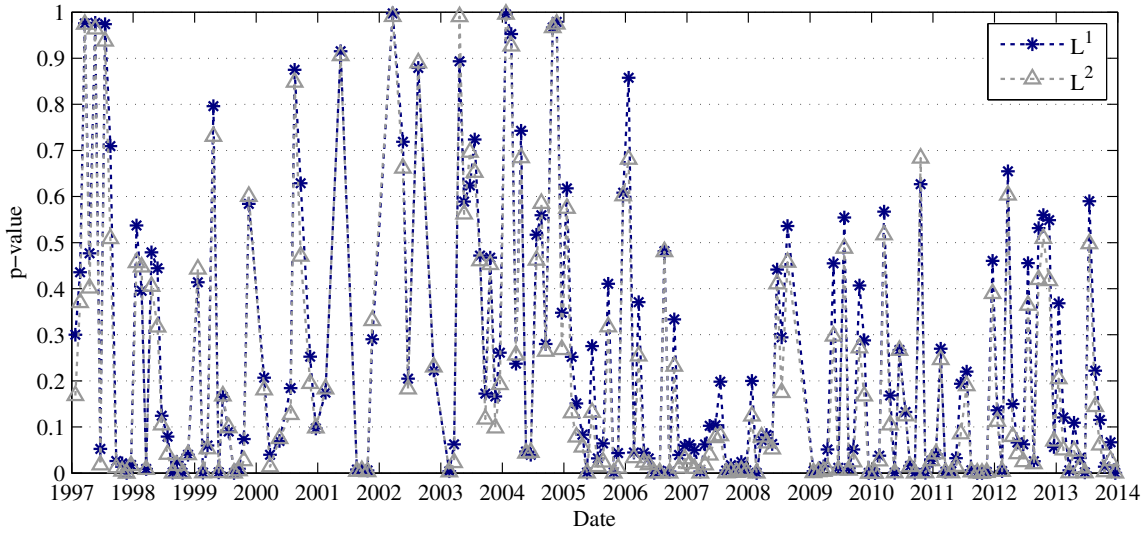

Figure 4.3: p-values for monotonicity tests.

Here we plot p-values for our L^1- and L^2-based test statistics at 180 dates during 1997-2013. p-values indicate the smallest nominal significance level at which we would reject the null hypothesis of pricing kernel monotonicity at a given date.

realized volatilities in Figure 4.1.

We also performed several specification tests on our physical distribution forecasts. Under correct specification of the physical distribution, the 180 probability-integral transformed returns should approximate an iid sequence of $U(0,1)$ random variables (Diebold et al., 1998). To test the independence of the sequence of transformed returns, we examined the first six autocorrelations of their first five moments. None were statistically significant at the 5% level. We used a Kolmogorov-Smirnov test to assess the uniformity of the transformed returns, obtaining a p-value of 0.423 and thereby failing to reject the null of uniformity at conventional significance levels.

4.4 Robustness checks

In Table 4.2 we report the overall rejection rates obtained with our two test statistics using a number of variations on our baseline testing procedure.

First, we explore the robustness of our results to variations in the procedure used to estimate the physical distribution. As discussed earlier, the volatilities used in our baseline results were constructed using a MiDaS estimator for dates between 1975–1996, and a RealGARCH estimator for dates between 1997–2013. In row 2 of Table 4.2 we report the outcome of our tests when all volatilities are constructed using the MiDaS estimator. There is virtually no change. We also vary the length of the time series of rescaled excess returns used to estimate the physical distribution. In our baseline results we used the full 1975-2013 sample period to maximize statistical efficiency. In row 3 of Table 4.2 we report results obtained when the physical distribution is estimated using only those rescaled excess returns occurring at or prior to the date at which option prices are quoted. Again, there is virtually no change. In row 4 we report results obtained when rescaled excess returns from before 1988 are excluded from our sample. This change is motivated by the discussion in Jackwerth (2000), where it is suggested that the crash of October 1987 may have fundamentally changed investor's beliefs about the physical distribution, potentially inducing pricing kernel nonmonotonicity. We find that there is a substantial drop in the frequency of rejections when pre-1988 data are excluded from our sample. Of course, it is natural that our tests have lower power when the sample size is reduced. The rejection rates remain well above nominal size.

We also consider the sensitivity of our results to modeling choices associated with the risk neutral distribution. In order to include as many months as possible, we included in-the-money options in our baseline estimation procedure. In-the-money option prices can sometimes be unreliable. In line 5 of Table 4.2 we report results obtained after excluding calls with moneyness less than -3% and puts with moneyness greater than 3% from our sample. With fewer option prices, we are forced to eliminate 32 months from our sample, leaving us with 148 months. However, the rejection frequency remains about the same with either test statistic or significance level.

As discussed in Section 3.1, we estimate the parameters of our cubic spline by minimizing

	Specification	Number of months	L^1 statistic 5%	L^1 statistic 1%	L^2 statistic 5%	L^2 statistic 1%
1	Baseline	180	37	22	46	26
2	MiDaS throughout	180	37	21	46	28
3	Past observations only	180	39	22	48	27
4	1988-2013 sample	180	17	8	23	13
5	Exclude ITM options	148	39	21	46	28
6	Bid-ask weights	180	37	22	45	28
7	Equal weights	180	34	22	42	26
8	13 knots	156	37	23	44	27
9	9 knots	196	22	14	26	16
10	Exclude tails	180	42	22	46	27
11	Smoothed IV estimator	180	34	20	39	27
12	RND known	180	42	27	47	31
13	Constant physical volatility	180	66	52	72	60

Table 4.2: Overall rejection rates (%) under alternative specifications.

This table presents the rejection rates for our tests of pricing kernel monotonicity under a variety of alternative specifications. The first row corresponds to the baseline results given in the final row of Table 4.1. The model variations used to produce the results given in the other rows are discussed in the main text.

a weighted sum of squared pricing errors, with the weights given by observed trading volumes. In line 6 of Table 4.2 we report results obtained using bid-ask spreads in place of trading volumes, and in line 7 we report results obtained using equal weights for the pricing errors. Neither variation makes an appreciable difference to our rejection frequency. In lines 8 and 9 we vary the number of knots used in the cubic spline.[12] Increasing the number of knots from 11 to 13 has little effect. Reducing the number of knots from 11 to 9 leads to a moderate reduction in the rejection frequency, but the rejection rates remain well above nominal size.

Pricing kernel estimates can be highly unreliable in the tails of the physical density. To

[12]The number of months at which we apply our test varies when we change the number of spline knots because we only consider months including some date with at least $4k$ admissible option prices, where $k+1$ is the number of spline knots. This is done to ensure that the $4k \times 4k$ matrix Ξ_m, defined in the supplemental appendix, is invertible.

check that our rejections are not driven by nonmonotone behavior in this region, we modified our test statistics so that the integrals in (2.1) and (2.2) are over the interval $[0.02, 0.98]$, rather than the entire unit interval. A corresponding adjustment was made to critical values. Our results, displayed in line 10 of Table 4.2, differed little from the baseline results.

The risk neutral estimator proposed by Monteiro et al. (2008) is well-suited for our purposes because we are able to establish weak convergence of the normalized estimation error to a finite dimensional Gaussian process. This is Theorem A.2 in the supplemental appendix. To check the robustness of our results to the choice of risk neutral estimator, we redid our analysis using an alternative estimator. First, we used Black-Scholes inversion to transform our option prices to implied volatilities, and fit a quartic polynomial to the transformed data by weighted least squares, using bid-ask spreads of implied volatility as weights. We then transformed our fitted implied volatilities back to fitted prices using the Black-Scholes formula, and twice differentiated to obtain an estimate of the risk neutral density à la Breeden and Litzenberger (1978). Finally, we modified the tails of the estimated risk neutral density, imposing a log-normal functional form in a manner similar to Figlewski (2010). This has the effect of eliminating perverse tail behavior associated with extrapolating a fitted quartic polynomial beyond the range of the data, and ensures that the density integrates to one.

We do not have an analogue to Theorem A.2 that applies to the smoothed implied volatility estimator just described, and see no obvious way to obtain one. Therefore, when using this estimator, we are unable to account for uncertainty about the shape of the risk neutral distribution when computing our critical values. We may compute critical values under the assumption that the risk neutral distribution is known by setting $\xi = 0$ in the algorithm given in Section 3.3. Proceeding in this way, we obtain the results given in line 11 of Table 4.2.[13] We see that the rejection rates using the smoothed implied volatility estimator are similar to the baseline results. For the purposes of comparison, in line 12 we provide results obtained using the cubic spline estimator under the assumption that the risk neutral distribution is known, so that critical values are computed with $\xi = 0$. We see

[13]The robustness checks reported in lines 2,3,4,5,10 of Table 4.2 were also performed using the smoothed implied volatility estimator in place of the cubic spline. The results, which are available on request, are similar to the results obtained using the cubic spline.

that the rejection rates are only modestly higher than our baseline results. This suggests that the sampling distributions of our test statistics are dominated by uncertainty about the physical distribution, with uncertainty about the risk neutral distribution playing a smaller role in determining critical values.

In line 13 of Table 4.2 we report the rejection frequencies obtained under the assumption that the volatility of the physical distribution is constant. We achieve this by setting the ratio σ_t/σ_j in (3.5) equal to one in the construction of our estimated physical distribution. The rejection frequencies obtained with constant physical volatility are much higher than the baseline rejection rates, and also higher than the rejection rates given in all other rows of Table 4.2. This is not surprising: if we hold physical volatility constant, then in times of high (low) risk neutral volatility we are likely to observe a U-shaped (hump-shaped) pricing kernel, and reject pricing kernel monotonicity. Note that the tests of pricing kernel monotonicity reported by Golubev et al. (2014) and Härdle et al. (2014) both ignore the time variation in conditional market volatility. Since financial agents do in fact take current information about market volatility into account when valuing options, it should not be surprising that we frequently reject pricing kernel monotonicity when we do not use this information in our estimate of the physical distribution. What *is* arguably surprising is that we continue to find significant violations of pricing kernel monotonicity when we do account for current information about market volatility, as indicated by our baseline results.

5 Final remarks

We have proposed a formal statistical test of pricing kernel monotonicity, and applied it using a sample of seventeen years of options market data for the S&P 500 index. Statistically significant violations of monotonicity are detected in a substantial proportion of the dates considered. Our results provide empirical support for the growing literature on the pricing kernel puzzle, indicating that well documented nonmonotonicities in estimated pricing kernels cannot always be attributed to sampling uncertainty. This finding is inconsistent with a broad class of theoretical asset pricing models, suggesting that economic mechanisms which generate pricing kernel nonmonotonicities may play an important role

in explaining observed data.

References

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

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

Bakshi, G., Madan, D. and Panayotov, G. (2010). Returns of claims on the upside and the viability of U-shaped pricing kernels. *Journal of Financial Economics* **97** 130–154.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2008). Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* **76** 1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2009). Realized kernels in practice: trades and quotes. *Econometrics Journal* **12** C1–C33.

Barone-Adesi, G., Engle, R. F. and Mancini, L. (2008). A GARCH option pricing model with filtered historical simulation. *Review of Financial Studies* **21** 1223–1258.

Barone-Adesi, G. Dall'O, H. and Vovchak, V. (2012). Is the price kernel monotone? *ACRN Journal of Finance and Risk Perspectives* **1** 43–68.

Beare, B. K. (2011). Measure preserving derivatives and the pricing kernel puzzle. *Journal of Mathematical Economics* **47** 689–697.

Beare, B. K. and Moon, J. -M. (2014). Nonparametric tests of density ratio ordering. To appear in *Econometric Theory*.

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

Brown, D. P. and Jackwerth, J. C. (2012). The pricing kernel puzzle: reconciling index option data and economic theory. In Batten, J. A. and Wagner, N. (Eds.), *Derivative Securities Pricing and Modelling* 155–183. Emerald Group, Bingley.

Carolan, C. A. and Tebbs, J. M. (2005). Nonparametric tests for and against likelihood ratio ordering in the two sample problem. *Biometrika* **92** 159–171.

Chabi-Yo, F., Garcia, R. and Renault, E. (2008). State dependence can explain the risk aversion puzzle. *Review of Financial Studies* **21** 973–1011.

Chaudhuri, R. and Schroder, M. (2010). Monotonicity of the stochastic discount factor and expected option returns. Working paper, Eli Broad College of Business, Michigan State University.

Cochrane, J. H. (2001). *Asset Pricing.* Princeton University Press.

Constantinides, G. M., Jackwerth, J. C. and Perrakis, S. (2009). Mispricing of S&P 500 index options. *Review of Financial Studies* **22** 1247–1277.

Diebold, F. X., Gunther, T. A. and Tay, A. S. (1998) Evaluating density forecasts with applications to financial risk management. *International Economic Review* **39** 863–883.

Dybvig, P. H. (1988). Distributional analysis of portfolio choice. *Journal of Business* **63** 369–393.

Figlewski, S. (2010). Estimating the implied risk neutral density for the U.S. market portfolio. In Bollerslev, T., Russell, J. R., and Watson, M. W. (Eds.), *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle.* Oxford University Press.

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2005). There is a risk-return tradeoff after all. *Journal of Financial Economics* **76** 509–548.

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics* **131** 59–95.

Golubev, Y., Härdle, W. K. and Timonfeev, R. (2014). Testing monotonicity of pricing kernels. To appear in *Advances in Statistical Analysis*.

Hansen, P. R., Huang, Z. and Shek, H. H. (2012). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics* **27** 877–906.

Härdle, W. K., Okhrin, Y. and Wang, W. (2014). Uniform confidence bands for pricing kernels. To appear in the *Journal of Financial Econometrics*.

Heber, G., Lunde, A., Shephard, N. and Sheppard, K. (2009). Oxford-Man Institute's realized library v. 0.2. Oxford-Man Institute, University of Oxford.

Hens, T. and Reichlin, C. (2012). Three solutions to the pricing kernel puzzle. *Review of Finance* **17** 1065–1098.

Jackwerth, J. C. (2000). Recovering risk aversion from option prices and realized returns. *Review of Financial Studies* **13** 433–451.

Jackwerth, J. C. and Rubinstein, M. (1996). Recovering probability distributions from option prices. *Journal of Finance* **51** 1611–1631.

Monteiro, A. M., Tütüncü, R. H. and Vicente, L. N. (2008). Recovering risk-neutral probability density functions from options prices using cubic splines and ensuring nonnegativity. *European Journal of Operational Research* **187** 525–542.

Patton, A. J. and Timmermann, A. (2010). Monotonicity in asset returns: new tests with applications to the term structure, the CAPM, and portfolio sorts. *Journal of Financial Economics* **98** 605–625.

Robertson, T., Wright, F. and Dykstra, R. (1988). *Order Restricted Statistical Inference.* Wiley, New York.

Rosenberg, J. V. and Engle, R. F. (2002). Empirical pricing kernels. *Journal of Financial Economics* **64** 341–372.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New York.

A Supplemental appendix

In Sections A.1-A.3 we provide a formal statement and justification of claims made in Sections 2 and 3 of the main text about the asymptotic behavior of our test statistics $M^1_{m,n}$ and $M^2_{m,n}$ and estimators Q_m and P_n. In Section A.4 we provide some tabulated summary statistics, referred to in the main text, concerning the option price data used to obtain the empirical results in Section 4 of the main text.

A.1 Limit distribution of test statistics

In Section 2 of the main text we asserted that $M^1_{m,n}$ and $M^2_{m,n}$ have limit distributions given by (2.4) and (2.5) under the weak convergence postulated in (2.3). The following result formalizes our claim. We use $\rightsquigarrow$ to denote weak convergence in some metric space (see e.g. Definition 1.3.3 in van der Vaart and Wellner, 1996), $\circ$ to denote composition, and $\ell^\infty(\mathscr{X})$ to denote the space of uniformly bounded real valued functions on $\mathscr{X}$ equipped with the uniform metric.

Theorem A.1. *Suppose there exist independent continuous random elements $\xi, \zeta \in \ell^\infty([0,1])$ such that the sequences of random cdfs Q_m and P_n satisfy the following weak convergence condition in the product metric space $\ell^\infty(\mathbb{R}) \times \ell^\infty(\mathbb{R})$ as $m, n \to \infty$ with $n/(m+n) \to \lambda$:*

$$\begin{pmatrix} \sqrt{m}(Q_m - Q) \\ \sqrt{n}(P_n - P) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \xi \circ Q \\ \zeta \circ P \end{pmatrix}.$$

If π is constant, then we must also have

$$M^1_{m,n} \xrightarrow{d} \int_0^1 (\mathcal{M}\mathscr{G}(u) - \mathscr{G}(u))\, du \quad \text{and} \quad M^2_{m,n} \xrightarrow{d} \left(\int_0^1 (\mathcal{M}\mathscr{G}(u) - \mathscr{G}(u))^2\, du \right)^{1/2},$$

where $\mathscr{G}(u) = \lambda^{1/2}\xi(u) - (1-\lambda)^{1/2}\zeta(u)$.

Proof of Theorem A.1. We begin by establishing that the following weak convergence holds in $\ell^\infty([0,1])$:

$$\sqrt{\frac{mn}{m+n}}\,(\phi_{m,n} - \phi) \rightsquigarrow \mathscr{G}. \tag{A.1}$$

Let $\mathcal{I}$ denote the identity function on $[0,1]$. Observe that

$$\sqrt{\frac{mn}{m+n}}\,(\phi_{m,n}-\phi)=\sqrt{\frac{mn}{m+n}}\,(Q_m-Q)\circ P_n^{-1}+\sqrt{\frac{mn}{m+n}}\left(Q\circ P_n^{-1}-Q\circ P^{-1}\right). \quad \text{(A.2)}$$

Focusing first on the second term on the right-hand side of (A.2), and noting that $P\circ P^{-1}=\mathcal{I}$ by virtue of the continuity of P, we use the fact that $P=Q$ when π is constant to obtain

$$\sqrt{\frac{mn}{m+n}}\left(Q\circ P_n^{-1}-Q\circ P^{-1}\right)=\sqrt{\frac{mn}{m+n}}\left((P_n\circ P^{-1})^{-1}-\mathcal{I}\right). \quad \text{(A.3)}$$

In view of the continuity of the transformation $\ell^\infty(\mathbb{R})\ni\psi\mapsto\psi\circ P^{-1}\in\ell^\infty([0,1])$, the continuous mapping theorem implies that $\sqrt{n}(P_n\circ P^{-1}-\mathcal{I})\rightsquigarrow\zeta$ in $\ell^\infty([0,1])$. Using the Hadamard differentiability of the inverse operator at $\mathcal{I}$ tangentially to $C([0,1])$ (van der Vaart and Wellner, 1996, Lemma 3.9.23(ii)) we may apply the functional delta method (van der Vaart and Wellner, 1996, Theorem 3.9.4) to obtain

$$\sqrt{n}((P_n\circ P^{-1})^{-1}-\mathcal{I})\rightsquigarrow-\zeta \quad \text{in } \ell^\infty[0,1]. \quad \text{(A.4)}$$

Since $m/(m+n)\to 1-\lambda$, we deduce from (A.3) and (A.4) that

$$\sqrt{\frac{mn}{m+n}}\left(Q\circ P_n^{-1}-Q\circ P^{-1}\right)\rightsquigarrow-(1-\lambda)^{1/2}\zeta \quad \text{in } \ell^\infty([0,1]). \quad \text{(A.5)}$$

This takes care of the second term on the right-hand side of (A.2). The first term on the right-hand side of (A.2) may be written as

$$\sqrt{\frac{mn}{m+n}}\,(Q_m-Q)\circ P_n^{-1}=\sqrt{\frac{n}{m+n}}\left(\sqrt{m}(Q_m\circ Q^{-1}-\mathcal{I})\right)\circ\left(Q\circ P_n^{-1}\right). \quad \text{(A.6)}$$

Applying the continuous mapping theorem again, we find that $\sqrt{m}(Q_m\circ Q^{-1}-\mathcal{I})\rightsquigarrow\xi$ in $\ell^\infty([0,1])$. Further, (A.5) implies that $Q\circ P_n^{-1}\rightsquigarrow Q\circ P^{-1}=\mathcal{I}$ in $\ell^\infty([0,1])$. Therefore, applying the continuous mapping theorem to the operation of composition – which is justified by the uniform continuity of ξ – we find that

$$\left(\sqrt{m}(Q_m\circ Q^{-1}-\mathcal{I})\right)\circ\left(Q\circ P_n^{-1}\right)\rightsquigarrow\xi\circ\mathcal{I}=\xi \quad \text{in } \ell^\infty([0,1]). \quad \text{(A.7)}$$

Since $n/(m+n) \to \lambda$, we deduce from (A.6) and (A.7) that

$$\sqrt{\frac{mn}{m+n}}\,(Q_m - Q) \circ P_n^{-1} \rightsquigarrow \lambda^{1/2}\xi \quad \text{in } \ell^\infty([0,1]). \tag{A.8}$$

The weak convergence posited in the statement of Theorem A.1 holds in the product metric space $\ell^\infty(\mathbb{R}) \times \ell^\infty(\mathbb{R})$. It follows that the weak convergences in (A.5) and (A.8) hold jointly in the product metric space $\ell^\infty([0,1]) \times \ell^\infty([0,1])$. In view of (A.2), this establishes the validity of (A.1).

We complete the proof with another application of the continuous mapping theorem. One may show without difficulty that the lcm operator $\mathcal{M} : \ell^\infty([0,1]) \to \ell^\infty([0,1])$ and related operator $\mathcal{D} = \mathcal{M} - \mathcal{I}$ are positive homogenous of degree one. One may also show that $\mathcal{M}(\psi_1 + \psi_2) = \mathcal{M}\psi_1 + \psi_2$ and $\mathcal{D}(\psi_1 + \psi_2) = \mathcal{D}\psi_1$ whenever $\psi_1, \psi_2 \in \ell^\infty([0,1])$ and ψ_2 is affine; see Carolan and Tebbs (2005, p. 168). Since ϕ is affine when π is constant, we may therefore write

$$M_{m,n}^1 = \int_0^1 \mathcal{D}\left(\sqrt{\frac{mn}{m+n}}(\phi_{m,n} - \phi)\right)(u)\mathrm{d}u$$

and

$$M_{m,n}^2 = \left(\int_0^1 \left(\mathcal{D}\left(\sqrt{\frac{mn}{m+n}}(\phi_{m,n} - \phi)\right)(u)\right)^2 \mathrm{d}u\right)^{1/2}.$$

In view of (A.2) and the continuity of the L^1 and L^2 norms, our desired result now follows from the continuous mapping theorem if we can show that $\mathcal{D}$ is continuous. In fact, continuity follows immediately from Marshall's Lemma (Robertson et al., 1988, ch. 7) which states that $\sup|\mathcal{M}\psi_1 - \mathcal{M}\psi_2| \leq \sup|\psi_1 - \psi_2|$ for any $\psi_1, \psi_2 \in \ell^\infty([0,1])$. □

A.2 Weak convergence of risk neutral estimator

In Section 3.1 we discussed an estimator for the risk neutral distribution proposed by Monteiro et al. (2008). We asserted that the normalized estimation error $\sqrt{m}(Q_m - Q)$ converges weakly to a finite dimensional Gaussian process. Here we provide a formal statement and justification of this claim. We employ the following technical conditions.

Assumption A.1. The following statements are true.

(a) q is a cubic spline $q(\cdot;\theta^*)$, with $\theta^* \in \Theta$.

(b) $\{(d_i, s_i, v_i) : i \in \mathbb{N}\}$ is a collection of nonrandom elements of $\{0,1\} \times (0,\infty) \times (0,\infty)$, while $\{w_i : i \in \mathbb{N}\}$ is a collection of real valued random variables.

(c) For each $i \in \mathbb{N}$, $w_i = h(d_i, s_i; \theta^*) + e_i$, with $\{e_i : i \in \mathbb{N}\}$ a mutually independent collection of real valued random variables, each with $E(e_i) = 0$.

(d) $\lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} v_i z_i z_i' = \Xi$ for some positive definite matrix Ξ.

(e) $\lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} v_i^2 z_i z_i' E(e_i^2) = \Sigma$ for some positive semidefinite matrix Σ.

(f) $\sup_{i\in\mathbb{N}} E|v_i z_{i,j} e_i|^{2+\delta} < \infty$ for each $j = 1, \ldots, 4k$ and some $\delta > 0$.

(g) q is strictly positive on (x_0, x_k), with strictly positive right-derivative at x_0 and strictly negative left-derivative at x_k.

Part (a) of Assumption A.1 ensures that our model for q is well specified. Parts (b)-(f) are standard regularity conditions for weighted linear regression with fixed regressors. Part (g) ensures that the inequality restrictions in (3.2) of the main text have an asymptotically negligible impact upon the distribution of $\sqrt{m}(\theta_m - \theta^*)$.

Under Assumption A.1 we obtain the following result, which describes the limiting behavior of $\sqrt{m}(\theta_m - \theta^*)$ and $\sqrt{m}(Q_m - Q)$. Let I denote the $4k \times 4k$ identity matrix, and let $M = I - \Xi^{-1}R' \left(R\Xi^{-1}R'\right)^{-1} R$. The matrices Σ and Ξ are defined in Assumption A.1. We define the following sample analogues to Σ, Ξ and M:

$$\Sigma_m = \frac{1}{m} \sum_{i=1}^{m} v_i^2 z_i z_i' \left(w_i - z_i'\theta_n\right)^2, \quad \Xi_m = \frac{1}{m} \sum_{i=1}^{m} v_i z_i z_i', \quad M_m = I - \Xi_m^{-1}R' \left(R\Xi_m^{-1}R'\right)^{-1} R.$$

In addition, we define

$$\Psi_m = M_m \Xi_m^{-1} \Sigma_m \Xi_m^{-1} M_m', \quad \Psi = M\Xi^{-1}\Sigma\Xi^{-1}M'. \tag{A.9}$$

Theorem A.2. *Suppose Assumption A.1 is satisfied, and let N denote a vector of $4k$ independent standard normal random variables. Then, as $m \to \infty$, the following statements are valid.*

(a) $\sqrt{m}\,(\theta_m - \theta^*) \to_d \Psi^{1/2} N$.

(b) $\sqrt{m}\,(Q_m - Q) \rightsquigarrow \xi \circ Q$ *in* $\ell^\infty(\mathbb{R})$, *where* $\xi(u) = \left(\frac{\partial}{\partial \theta} Q(Q^{-1}(u;\theta^*);\theta)\big|_{\theta=\theta^*} \right)' \Psi^{1/2} N$.

Proof of Theorem A.2. First we prove part (a). The main inconvenience here is the presence of the inequality restrictions (3.2), which we temporarily dispose of. Let $\tilde{\Theta}$ be the collection of $\theta \in \mathbb{R}^{4k}$ for which $R\theta = \gamma$, and let

$$\tilde{\theta}_m = \operatorname*{argmin}_{\theta \in \tilde{\Theta}} \sum_{i=1}^m v_i \left(w_i - h(d_i, s_i; \theta)\right)^2 = \operatorname*{argmin}_{\theta \in \tilde{\Theta}} \sum_{i=1}^m v_i \left(w_i - z_i'\theta\right)^2.$$

$\tilde{\theta}_m$ is simply a weighted least squares estimator subject to linear equality restrictions. With some elementary algebra we may show that

$$\sqrt{m}\left(\tilde{\theta}_m - \theta^*\right) = M_m \Xi_m^{-1} \frac{1}{\sqrt{m}} \sum_{i=1}^m v_i z_i e_i,$$

where Ξ_m is invertible for m sufficiently large under Assumption A.1(d). This condition also ensures that $\Xi_m \to \Xi$ and $M_m \to M$ as $m \to \infty$. An application of the Liapounov central limit theorem gives $m^{-1/2} \sum_{i=1}^m v_i z_i e_i \to_d N(0, \Sigma)$ under Assumption A.1(b,c,e,f). We conclude that $\sqrt{n}(\tilde{\theta}_m - \theta^*) \to_d \Psi^{1/2} N$.

If the probability of the inequality restrictions (3.2) binding vanishes in the limit, so that $\text{Prob}(\tilde{\theta}_m = \theta_m) \to 1$, then the proof of (a) will be complete. Under Assumption A.1(a,g) we may choose $\epsilon > 0$ and $\delta > 0$ such that $q'(\cdot;\theta^*) \geq \delta$ on $(x_0, x_0 + \epsilon)$, $q(\cdot;\theta^*) \geq \delta$ on $[x_0 + \epsilon, x_k - \epsilon]$, and $q'(\cdot;\theta^*) \leq -\delta$ on $(x_k - \epsilon, x_k)$. If $q(\cdot;\tilde{\theta}_m) \to q(\cdot;\theta^*)$ and $q'(\cdot;\tilde{\theta}_m) \to q'(\cdot;\theta^*)$ in probability uniformly on (x_0, x_k), then the probabilities of the three events

$$\left\{ \inf_{x \in (x_0, x_0+\epsilon)} q'(x;\tilde{\theta}_m) > 0 \right\}, \left\{ \inf_{x \in [x_0+\epsilon, x_k-\epsilon]} q(x;\tilde{\theta}_m) > 0 \right\}, \left\{ \inf_{x \in (x_k-\epsilon, x_k)} q'(x;\tilde{\theta}_m) < 0 \right\}$$

will all converge to one as $m \to \infty$. But then we must also have $q(\cdot;\tilde{\theta}_m) \geq 0$ on (x_0, x_k) with probability approaching one, so that $\text{Prob}(\tilde{\theta}_m = \theta_m) \to 1$. To prove uniform convergence in probability of $q(\cdot;\tilde{\theta}_m)$ to $q(\cdot;\theta^*)$ and of $q'(\cdot;\tilde{\theta}_m)$ to $q'(\cdot;\theta^*)$, we note that $q(x;\theta)$ and $q'(x;\theta)$ are linear in θ for each $x \in \mathbb{R}$. Thus we may write $q(x;\tilde{\theta}_m) = q(x;\theta^*) + \left(\frac{\partial}{\partial\theta} q(x;\theta)|_{\theta=\theta^*}\right)'(\tilde{\theta}_m - \theta^*)$ and $q'(x;\tilde{\theta}_m) = q'(x;\theta^*) + \left(\frac{\partial}{\partial\theta} q'(x;\theta)|_{\theta=\theta^*}\right)'(\tilde{\theta}_m - \theta^*)$. Uniform

convergence in probability now follows from the uniform boundedness of $\frac{\partial}{\partial\theta}q(x;\theta)|_{\theta=\theta^*}$ and $\frac{\partial}{\partial\theta}q'(x;\theta)|_{\theta=\theta^*}$ and the fact that $\tilde{\theta}_m \to \theta^*$ in probability. This proves (a).

To prove (b) we use the fact that $Q(x;\theta)$ is linear in θ for each $x \in \mathbb{R}$ to write

$$\sqrt{m}(Q_m(x) - Q(x)) = \sqrt{m}(Q(x;\theta_m) - Q(x;\theta^*)) = \left(\frac{\partial}{\partial\theta} Q(x;\theta)\bigg|_{\theta=\theta^*} \right)' \sqrt{m}(\theta_m - \theta^*)$$

for each $x \in \mathbb{R}$ and $m \in \mathbb{N}$. Our desired result now follows from part (a) and the fact that $\frac{\partial}{\partial\theta}Q(x;\theta)|_{\theta=\theta^*}$ is uniformly bounded in x. □

A.3 Weak convergence of physical estimator

Here we provide a formal statement of the weak convergence of $\sqrt{n}(P_n - P)$ asserted in Section 3.2 of the main text. We treat $\{(r_j, \sigma_j) : j \in \mathbb{N}\}$ as a nonrandom sequence in $(0,1) \times (0,\infty)$, and $\{X_j : t \in \mathbb{N}\}$ as a sequence of real valued random variables.

Theorem A.3. *Suppose $\{(X_j - r_j)/\sigma_j : j \in \mathbb{N}\}$ is an iid sequence of random variables, and let P be the cdf of $\exp(X_t)$. Then, as $n \to \infty$, we have $\sqrt{n}\,(P_n - P) \rightsquigarrow \zeta \circ P$ in $\ell^\infty(\mathbb{R})$, where ζ is a standard Brownian bridge.*

Proof of Theorem A.3. P_n is the empirical cdf of the iid random variables $\exp(X_j^*)$, $j = 1, \ldots, n$. Thus Donsker's theorem implies that $\sqrt{n}(P_n - P^*) \rightsquigarrow \zeta \circ P^*$ in $\ell^\infty(\mathbb{R})$, where P^* is the common cdf of the $\exp(X_j^*)$'s. Since $X_t^* = X_t$, we have $P^* = P$. □

A.4 Summary statistics for option price data

Tables A.1, A.2 and A.3, referred to in Section 4 of the main text, include some summary statistics for the option price data used to obtain our empirical results. In Table A.1 we report the maximum, minimum and average number of put and call option prices observed in the different months of each year. In Table A.2 we report average implied volatilities by option moneyness, computed using the Black-Scholes formula. In Table A.3 we report the maximum, minimum and average of the discounted means and annualized volatilities implied by our estimated risk neutral distributions, by year.

	Number of calls			Number of puts		
Year	Avg	Min	Max	Avg	Min	Max
1997	29	24	35	34	26	52
1998	26	20	37	35	27	42
1999	26	19	35	33	24	46
2000	26	19	33	30	19	41
2001	22	17	27	32	20	44
2002	22	15	37	31	22	44
2003	26	19	31	32	25	40
2004	29	25	33	35	31	43
2005	32	28	36	37	31	43
2006	37	30	42	48	41	56
2007	43	33	63	67	53	88
2008	62	48	93	84	65	124
2009	62	55	76	86	75	100
2010	58	50	67	100	88	117
2011	59	48	75	111	87	143
2012	53	43	73	101	90	124
2013	60	47	72	104	89	139

Table A.1: Option count statistics.

Figures indicate the average, minimum and maximum number of distinct put and call option prices with nonzero trading volume and maturity between 18 and 22 trading days, sorted by year.

	Average implied volatility by moneyness category									
Year	<-15	-15:-10	-10:-6	-6:-3	-3:0	0:3	3:6	6:10	10:15	>15
1997	42	29	26	23	21	19	18	17	18	0
1998	50	32	28	24	21	20	17	17	20	24
1999	46	33	28	24	22	20	18	17	18	27
2000	45	30	26	23	21	20	19	19	22	28
2001	44	35	35	30	27	26	25	23	24	33
2002	93	35	29	27	24	23	22	21	24	36
2003	53	29	25	22	19	18	18	20	22	39
2004	34	24	20	17	14	13	12	14	17	0
2005	32	23	18	14	12	10	10	11	15	22
2006	36	23	18	15	12	10	9	11	14	0
2007	33	26	22	19	16	14	13	13	16	19
2008	62	40	36	32	30	29	27	26	30	52
2009	52	36	32	29	27	25	24	24	24	29
2010	45	31	26	23	20	18	17	16	18	22
2011	49	32	28	24	21	19	18	17	19	23
2012	39	26	22	18	16	14	12	13	16	18
2013	35	24	19	16	14	12	11	12	14	35

Table A.2: Option implied volatilities.

Figures indicate the average Black-Scholes implied volatility of distinct put and call option prices with nonzero trading volume and maturity between 18 and 22 trading days, sorted by moneyness (%) and year.

Year	Number of months	Discounted mean			Annualized volatility		
		Avg	Min	Max	Avg	Min	Max
1997	12	1.002	0.997	1.005	0.206	0.170	0.262
1998	11	1.002	1.000	1.005	0.219	0.160	0.364
1999	11	1.003	0.999	1.006	0.217	0.168	0.262
2000	8	1.005	1.003	1.007	0.217	0.164	0.306
2001	6	1.001	0.998	1.003	0.255	0.200	0.420
2002	5	1.002	1.000	1.003	0.227	0.163	0.289
2003	11	1.000	0.998	1.000	0.194	0.148	0.281
2004	12	1.000	0.998	1.002	0.143	0.120	0.185
2005	12	1.001	0.999	1.003	0.119	0.100	0.135
2006	12	1.003	1.001	1.004	0.118	0.099	0.154
2007	12	1.003	1.002	1.005	0.159	0.098	0.243
2008	8	0.999	0.995	1.005	0.225	0.179	0.287
2009	12	0.998	0.996	1.000	0.284	0.194	0.420
2010	12	0.998	0.995	1.000	0.211	0.145	0.317
2011	12	0.999	0.997	1.000	0.224	0.140	0.333
2012	12	0.999	0.998	1.000	0.163	0.131	0.210
2013	12	0.998	0.997	1.001	0.139	0.121	0.192

Table A.3: Moments of estimated risk neutral distributions.

This table presents summary statistics for the (standardized) moments of the estimated risk neutral densities by year. If the estimated risk neutral distribution prices the market portfolio perfectly, the normalized mean should be equal to one.