The Pricing Kernel Puzzle in Forward Looking Data[*]

Horatio Cuesdeanu   Jens Carsten Jackwerth

June 29, 2016

Abstract

The pricing kernel puzzle concerns the locally increasing empirical pricing kernel, which is inconsistent with a risk-averse representative investor in a single period, single state variable setting. Some recent papers worry that the puzzle is caused simply by the mismatch of backward looking subjective and forward looking risk-neutral distributions of index returns. By using a novel test and forward looking information only, we generally confirm the existence of a u-shaped pricing kernel puzzle in the S&P 500 options data. The evidence is weaker for tests against an alternative with a risk-neutral investor and for longer horizons.

1

1 Introduction

The pricing kernel is a cornerstone of modern finance, as it tells us how to transform subjective probabilities into risk-neutral ones. As the price of any security should be the discounted expected payoff under the risk-neutral measure, the pricing kernel is indispensable for pricing. By assuming a simple single-period, representative investor economy, one can show that the pricing kernel is proportional to marginal utility.[1] Assuming risk-aversion in such a setting results in a monotonically decreasing pricing kernel.

Yet, when empirically estimating the pricing kernel, it turns out to be locally increasing, which is inconsistent with the above economy as that would imply risk-seeking behavior.[2] Ait-Sahalia and Lo (2000), Jackwerth (2000), and Rosenberg and Engle (2002) were the first to find this discrepancy and labeled it the pricing kernel puzzle. Those papers assume that a large index, e.g. the S&P 500, is a reasonable proxy for wealth at the end of a single period. The risk-neutral density is backed out from the forward looking option prices written on the index. The subjective density is inferred from past index returns using non-parametric or parametric models. Finally, the (discounted) ratio of the risk-neutral divided by the subjective density delivers the pricing kernel. Jackwerth (2004) and Cuesdeanu and Jackwerth (2016) survey the literature and details the different approaches, which tend to confirm the pricing kernel puzzle using slightly different methods and data sets. More recently, Christoffersen, Heston, and Jacobs (2013) and Beare and Schmidt (2014) confirm the pricing kernel puzzle yet again in the data.

Only a few studies dispute the puzzle, and we group them into three categories. Each of the groups throws up problems, which we want to address in our work. First, Bliss and Panigirtzoglou (2004) start out with pricing kernels based on power or exponential utility functions and thus restrict pricing kernels to monotonically decrease by construction. Instead, we allow both unrestricted pricing kernels and ones, which are restricted to monotonically decrease. Second, Barone-Adesi, Engle, and Mancini (2008) and Barone-Adesi and Dall'O (2010) argue that their

---

[1]See e.g. Cochrane (2000), pp. 50.

[2]Note that we do not literally suggest the existence of risk-seeking behavior, but rather does the evidence show a need for more encompassing models capable of generating the pricing kernel puzzle while maintaining the assumption of a risk-averse investor. We refer to such models in Section 4.2. For a more comprehensive view on the solutions to the pricing kernel puzzle, see Cuesdeanu and Jackwerth (2016).

pricing kernels are broadly decreasing even though they are locally increasing. This begs for a formal statistical test of monotonicity, which can distinguish between insignificant local fluctuation and significant non-monotonicity of the pricing kernel. We propose such tests. Third, Linn, Shive, and Shumway (2014) voice concern about the usual methodology, which mixes forward looking information for the risk-neutral density and backward looking information for the subjective density. As usual, they find the risk-neutral density from option prices based on a standard forward looking method. To avoid using backward looking subjective densities, they assume a (potentially increasing) pricing kernel based on B-splines, which allows them to compute the forward looking subjective density. They leave the density unscaled, so that the probabilities do not need to add up to one. They then work out the likelihood that a realized return over the next month is drawn from the subjective distribution. Maximizing the aggregated likelihood over many months allows them to find the pricing kernel that is most consistent with the realized future returns. According to the authors, that best pricing kernel is monotonically decreasing and hence consistent with economic theory. It is somewhat surprising that the pricing kernel does not increase even locally due to small errors in the data. Note that the authors do not formally test for monotonicity. We argue in Section 5.5 that their surprising results could be driven by the lack of scaling and their particular combination of different objective functions for the data fit.

Given the contradicting evidence, is the pricing kernel puzzle really a feature of the data? To answer this question, we suggest a number of methodological improvements and a novel test for pricing kernel monotonicity. For one, we use forward looking data only as we share the concern of Bliss and Panigirtzoglou (2004) and Linn, Shive, and Shumway (2014) that historical index return data might not be informative about the forward looking subjective distribution in the presence of regime shifts or non-stationary distributions. Second, we use flexible pricing kernels, which allow for increasing segments, in line with Linn, Shive, and Shumway (2014) and extending the methodology of Bliss and Panigirtzoglou (2004). Picking a tentative pricing kernel, we can obtain the forward looking subjective density by dividing the forward looking risk-neutral density by the assumed pricing kernel. We can now find the log score of the observed future return on the market. Maximizing the average log scores gives

us the best *unrestricted* pricing kernel. Maximizing the average log scores, while requiring the pricing kernel to monotonically decrease, gives us the best *restricted* pricing kernel. Third, we provide a formal test of pricing kernel monotonicity by simulating the difference in average log scores between restricted and unrestricted pricing kernels. We extend our test to a number of alternative test statistics measuring the quality of the pricing kernel to explain the data, namely, the moment fit criterion presented in Knüppel (2015), the Cramér van Mises statistic, and the Berkowitz (2001) test.

Existing pricing kernel monotonicity tests confirmed the existence of the pricing kernel puzzle but used different assumptions. Examining options written on the German market index on three dates at the end of June 2000, 2002, and 2004, Golubev, Haerdle, and Timofeev (2014) reject pricing kernel monotonicity. Their approach is similar to ours, in that they avoid the explicit estimation of subjective densities inferred from historical return data. However, this comes at the cost of assuming *iid* returns and, hence, unconditional subjective and risk-neutral distributions at each of the three dates, which have been picked in order to represent different market regimes: a bull market, a stable market, and a bear market. Beare and Schmidt (2014) allow for conditional distributions and still find monotonicity violations in the S&P 500 data with two tests based on ordinal dominance curves. Unfortunately, their use of historical returns for obtaining the subjective distribution exposes their results to the criticism of mixing forward and backward looking information.

Using our new test and only forward looking data, we confirm the pricing kernel puzzle in the data. Our findings are stable for different samples and goodness of fit criteria, including the ones used in Bliss and Panigirtzoglou (2004) and Linn, Shive, and Shumway (2014). Consistent with the recent literature (see Christoffersen, Heston, and Jacobs (2013)), the estimated pricing kernel displays a u-shape. In our most conservative test, which tests against an alternative assuming a risk-neutral investor, two out of four goodness of fit criteria still reject pricing kernel monotonicity at the monthly horizon. Considering longer horizons, pricing kernel monotonicity cannot be rejected.

The remainder of the paper proceeds as follows. Section 2 outlines our empirical approach and the construction of potential pricing kernels. Section 3 describes the data set. In Section

4, we present the results when the pricing kernel is modeled as a piece-wise linear function. Results from various robustness checks are provided in Section 5, and Section 6 concludes.

2 Methodology

We identified the desirability of a forward looking assessment of the subjective probability distribution, which is typically unknown. What we have readily available instead is the option implied risk-neutral density, which, in a single period representative agent economy as in e.g. Cochrane (2000), pp. 50, is linked to the subjective density via the pricing kernel. Formally, the subjective density is given by:

$$p_t(R_t) = \frac{q_t(R_t)}{m(R_t)} \tag{1}$$

where $p_t(R_t)$ denotes the time t conditional subjective density for the rate of return realized over the following month $R_t = S_{t+\Delta t}/S_t$ with S_t being the stock price at time t. The time t conditional risk-neutral density is denoted by $q_t(R_t)$ and the pricing kernel is $m(R_t)$. We assume that the pricing kernel, and hence preferences, are constant through time. This assumption does not imply stationarity for the subjective or risk-neutral distribution. Both distributions are allowed to vary over time; only the ratio of the densities has to be the same. The constant preferences assumption is essential for the approach used in this study and was also applied by Bliss and Panigirtzoglou (2004) and Linn, Shive, and Shumway (2014). Being worried that the pricing kernel might change somewhat over time, we show in Section 5 that our results survive for different subsamples and the shape of the pricing kernel does not change extensively.

We next discuss the estimation and construction of the two ingredients of the subjective density as given in equation (1): the risk-neutral density q_t and the pricing kernel m.

2.1 The Risk-Neutral Density

Empirically, we need to estimate the risk-neutral density $\widehat{q}_t$, which we obtain from the fast and stable method of Jackwerth (2004), which uses average implied volatilities based on bid

and ask option prices. The method then finds implied volatilities $\{\sigma_j\}_{j=0}^{J}$ on a fine grid of 2,500 moneyness levels from 0.2 to 1.8 and trades off between fitting a smooth implied volatility function while matching the implied volatilities observed in the market. The optimal implied volatilities are given by the closed form solution (given the trade-off parameter λ) of the following minimization problem:

$$\min_{\sigma_j} \frac{\delta^4}{2(J+1)} \sum_{j=0}^{J} (\sigma_j'')^2 + \frac{\lambda}{2I} \sum_{i=1}^{I} (\sigma_i - \bar{\sigma}_i)^2 \tag{2}$$

Thus, we obtain $J+1$ implied volatilities σ_j corresponding to a grid consisting of $J+1$ strike prices K_j. The fineness of the grid is given by δ, which is the difference between two successive strikes in the grid. The index i in the second term of equation (2) represents those implied volatilities σ_i, at which an option with implied volatility $\bar{\sigma}_i$ for the strike price K_i is observed in the market. The trade-off between smoothness and fit is controlled by the parameter λ. The smooth implied volatility curve is translated into the estimated risk-neutral density $\hat{q}_t$ by means of the Breeden and Litzenberger (1978) result. Our results are robust to using instead the fully parametric method of Gatheral (2004) for the risk-neutral density, see Section 5.3.

2.2 Construction of Potential Pricing Kernels

We model a tentative pricing kernel $\widetilde{m}$ as a piece-wise linear function and, in a robustness test, as a cubic spline. For specifying a tentative pricing kernel, we fix nine equally spaced values on the return axis: 0.8, 0.85, 0.9, 0.95, 1, 1.05, 1.1, 1.15, and 1.2. We fix the pricing kernel at a return of 0.8 at a value of 5, allow each of the other eight points to take pricing kernel values between zero and five, and interpolate linearly. Changing the fixed value at 0.8 away from 5 and changing the upper limit has virtually no effect on our results since the subjective densities are rescaled, see Section 2.3. As we only use option quotes with moneyness ($= K/S_t$) between 0.8 and 1.2, we follow Rosenberg and Engle (2002) and set the pricing kernel below 0.8 to its value at 0.8 and above 1.2 to its value at 1.2.[3] Hence, in our base setting, a tentative pricing kernel

---

[3]Our method is robust to extending the pricing kernel below 0.8 and above 1.2 at the slopes of the last interior segments instead of horizontally. Moreover, the results do not change when fixing different basis points on the

$\widetilde{m}(R_t; \Theta)$ is a piece-wise linear function of eight parameters $\Theta = \{\theta_1, ..., \theta_8\}$, which correspond to the values of the pricing kernel at the returns 0.85, 0.9, 0.95, 1, 1.05, 1.1, 1.15, and 1.2.

2.3 The Subjective Density

Varying the parameters Θ leads to a tentative pricing kernel $\widetilde{m}$, which we can combine with our estimate of the risk-neutral density $\widehat{q}_t$, so that we obtain the subjective density as

$$\widetilde{p}_t(R_t; \Theta) = \frac{\widehat{q}_t(R_t)}{\widetilde{m}(R_t; \Theta)} \bigg/ \int_0^\infty \frac{\widehat{q}_t(R_t)}{\widetilde{m}(R_t; \Theta)} dR_t \ . \tag{3}$$

The integral in equation (3) scales the subjective density such that it integrates to unity as probabilities should sum up to one.[4] Bliss and Panigirtzoglou (2004) perform the same normalization while Linn, Shive, and Shumway (2014) ignore this step, leading them to a monotonically decreasing pricing kernel estimate, see Section 5.5.

Given some tentative pricing kernel $\widetilde{m}(R_t; \Theta)$, we can now assess the log score of the future realized return in terms of a tentative subjective density $\ln(\widetilde{p}_t(R_t; \Theta))$.[5] The sample mean of the log scores gives us the average log score and measures how likely the observed returns were drawn from the tentative density, with a high value indicating a high likelihood.[6]

$$\widetilde{LS}(\Theta) = \frac{1}{N} \sum_{t=1}^N \ln\left(\widetilde{p}_t(R_t; \Theta)\right) = \frac{1}{N} \sum_{t=1}^N \ln\left(\frac{\widehat{q}_t(R_t)}{\widetilde{m}(R_t; \Theta)}\right) \tag{4}$$

where N is the number of (monthly) observations.

By varying the tentative pricing kernel over its parameters Θ, we maximize the average log score $\widetilde{LS}(\Theta)$ subject to the no-arbitrage condition ($\widetilde{m} > 0$, i.e., $\theta_i > 0$ for $i = 1, ..., 8$) and call the result the *unrestricted* pricing kernel $\widehat{m}^{unrestr.}$.

In a second maximization, we add the condition that the pricing kernel ought to be monotonically decreasing ($\widetilde{m}' \geq 0$, i.e., $\theta_i \geq \theta_j$ for $i \leq j$), so the result is consistent with an increasing

---

return axis, see Section 5.2.

[4]Our formulation is equivalent to expressing the subjective density in terms of state prices π_t as the risk-free discount factor $e^{-r_{f,t}\Delta t}$ cancels out when normalizing the subjective densities:

$p_t = \frac{\pi_t}{m} \bigg/ \int_0^\infty \frac{\pi_t}{m} dR_t = \frac{e^{-r_{f,t}\Delta t} q_t}{m} \bigg/ \int_0^\infty \frac{e^{-r_{f,t}\Delta t} q_t}{m} dR_t = \frac{q_t}{m} \bigg/ \int_0^\infty \frac{q_t}{m} dR_t$.

[5]See Gneiting and Raftery (2007) for a comprehensive discussion of score functions.

[6]A similar approach is used in Shackleton, Taylor, and Yu (2010) for mixing forecasts from past returns and forward looking option prices.
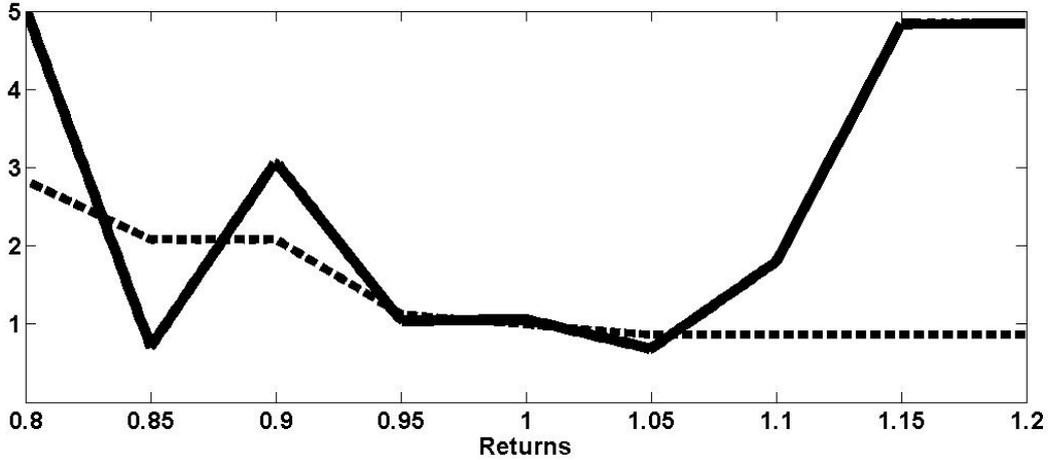
Figure 1: Linearly Interpolated Pricing Kernels. We plot an unrestricted scaled pricing kernel (solid) and a non-increasing scaled pricing kernel (dotted), constructed by our linear interpolation rule and maximizing the average log score as given in equation (4). We observe that the best fitting unrestricted pricing kernel has increasing regions. Returns are on the horizontal axis, and the values of the pricing kernel are on the vertical axis.

and concave utility function. We call this result the *restricted* pricing kernel $\widehat{m}^{restr.}$.

Figure 1 anticipates our main result and illustrates such tentative pricing kernels constructed by our linear interpolation rule. In particular, the unrestricted (restricted) pricing kernel in this figure is the best fitting unrestricted (restricted) pricing kernel when maximizing the average log score, as given in equation (4), see Section 4 for details. Note that the best fitting unrestricted pricing kernel has pronounced increasing regions and, hence, is not monotonically decreasing. The average log score is but one test statistic to assess the fit of the subjective densities to the data, and we detail alternatives in Section 2.5. We now turn to describing our formal test of monotonicity of the pricing kernel.

2.4 Monotonicity Test

We find that our average log score (LS) test statistic is 2.00 for the best-fitting (unrestricted) pricing kernel and 1.98 for the non-increasing (restricted) pricing kernel. We are thus left with the question, if the difference $\Delta_{LS} := LS^{unrestr.} - LS^{restr.}$ is statistically significant? As this difference is zero if and only if the true m is decreasing, we formulate our null hypothesis as:

$$H_0 : m \text{ is non-increasing} \tag{5}$$

8

As the distribution of the difference Δ_{LS} is unknown, we address this problem by simulating the distribution similarly to Bliss and Panigirtzoglou (2004). Here, we assume that the true pricing kernel $m^{power}(R_t) = R_t^{-\gamma}$ stems from a power utility function with a fixed $\gamma \geq 0$ and, hence, the true pricing kernel is non-increasing. For each month, we obtain the subjective density p_t^{power} consistent with our power utility function by dividing the observed risk-neutral density for that month by the power utility pricing kernel, as defined in Equation (3)

$$p_t^{power}(R_t) = \frac{\widehat{q}_t(R_t)}{m^{power}(R_t)} \Big/ \int_0^\infty \frac{\widehat{q}_t(R_t)}{m^{power}(R_t)} dR_t \tag{6}$$

Afterwards, we draw returns from these subjective densities, one from each monthly distribution. Finally, we employ our usual methodology and search for the best unrestricted and restricted pricing kernels which minimize the average log score. We calculate the differences Δ_{LS} and repeat the procedure 10,000 times; generating the distribution of Δ_{LS} under the null. We use this distribution to find the p-value of the empirical $\widehat{\Delta}_{LS}$, which we obtain from the data.

We the risk-aversion coefficients $\gamma = 0$, $\gamma = 2$, and $\gamma = 4$. The latter value models a risk-averse investor and is close to the estimate of $\gamma = 4.08$ in Bliss and Panigirtzoglou (2004), based on a monthly horizon. Statistically, Beare and Schmidt (2014) argue that monotonicity tests should use risk-neutrality (i.e., $\gamma = 0$ and $m^{power} = 1$). The risk-neutral density is then equal to the subjective density. This makes a rejection of non-monotonicity very difficult since a constant pricing kernel is the least decreasing pricing kernel specification and therefore most distant from the null hypothesis.[7]

2.5 Alternative Test Statistics

The average log score statistic used above is not the only way to assess if realized future returns were indeed drawn from a particular subjective density. We next present alternatives which are

---

[7]See Beare and Moon (2015) for a detailed discussion. Intuitively, when simulating with a constant pricing kernel, the differences between the simulated restricted and unrestricted log scores are larger than the differences when simulating with a decreasing pricing kernel. Hence, the difference we obtain from the data has to be much larger in order to reject the null hypothesis, making the risk-neutral version of the test the most conservative way to test pricing kernel monotonicity.

based on the probability integral transformation: if we knew the true subjective distribution, we could, for each realized return, look up its percentile, i.e., the cumulative probability of a return being lower or equal to the realized return. This cumulative probability will always lie between zero and one and is *iid* uniformly distributed. More formally, given a sequence of realized returns r_t stemming from the true subjective density $p_t(R_t)$, the following sequence of random variables u_t is *iid* uniform on the unit interval:

$$u_t = \int_0^{r_t} p_t(x)dx \ , \quad t = 1, ..., N \tag{7}$$

Using equation (3), we can substitute the subjective density p_t with the ratio of the estimated risk-neutral density $\widehat{q}_t$ and the tentative pricing kernel $\widetilde{m}$ and obtain:

$$\widetilde{u}_t = \int_0^{r_t} \left\{ \frac{\widehat{q}_t(x)}{\widetilde{m}(x;\Theta)} \Big/ \int_0^{\infty} \frac{\widehat{q}_t(R_t)}{\widetilde{m}(R_t;\Theta)} dR_t \right\} dx \tag{8}$$

Hence, as an alternative to maximizing the average log score, we can minimize the value of any test statistic having uniformity as the null. A number of tests are available to test for uniformity of $\widetilde{u}_t$ but only Berkowitz (2001) simultaneously tests for independence, and so we start there.

- The Berkowitz (2001) test

The first step is yet another transformation of the *iid* uniformly distributed $\widetilde{u}_t$ to normality:

$$\widetilde{z}_t = \Phi^{-1}(\widetilde{u}_t), \tag{9}$$

where Φ denotes the standard normal cumulative distribution. $\widetilde{z}_t$ is distributed *iid* standard normal if $\widetilde{u}_t$ is *iid* uniform. The second step is the estimation of the following AR(1) regression:

$$\widetilde{z}_t - \mu = \rho(\widetilde{z}_{t-1} - \mu) + \epsilon_t, \tag{10}$$

where $\epsilon_t \sim \mathcal{N}(0,1)$. If $\widetilde{z}_t$ is distributed *iid* standard normal, then $\mu = 0$, $\sigma = 1$, $\rho = 0$, and ϵ_t would coincide with $\widetilde{z}_t$. Berkowitz (2001) proposes the following likelihood ratio based test which accounts for both, *iid* and normality:

$$LR_3 = -2(LL(0,1,0) - LL(\widehat{\mu}, \widehat{\sigma}, \widehat{\rho})), \tag{11}$$

where $LL(\mu, \sigma, \rho)$ denotes the log likelihood of equation (10).

- The Knüppel (2015) test

The Knüppel (2015) test fits the empirical moments of the uniform variables $\widetilde{u}_t$ to their theoretical counterpart in a generalized method of moments (GMM) fashion. When considering L moments, the test statistic is given by

$$\widetilde{\alpha}_L = N \cdot \widetilde{D}'_L \widetilde{\Omega}_L^{-1} \widetilde{D}_L \; . \tag{12}$$

Here, $\widetilde{D}_L$ is a vector of length L consisting of the differences between sample moments and their theoretical values under the null of uniformity

$$\widetilde{D}_L = \begin{pmatrix} \frac{1}{N}\sum_{t=1}^{N} \widetilde{u}_t^1 - E[u^1] \\ \frac{1}{N}\sum_{t=1}^{N} \widetilde{u}_t^2 - E[u^2] \\ \vdots \\ \frac{1}{N}\sum_{t=1}^{N} \widetilde{u}_t^L - E[u^L] \end{pmatrix} . \tag{13}$$

Furthermore, $\widetilde{\Omega}_L$ is a consistent estimator of the covariance matrix of

$$\widetilde{d}_t = \begin{pmatrix} \widetilde{u}_t^1 - E[u^1] \\ \widetilde{u}_t^2 - E[u^2] \\ \vdots \\ \widetilde{u}_t^L - E[u^L] \end{pmatrix} . \tag{14}$$

11

In our base procedure, following Knüppel (2015), we use the first four moments ($L = 4$) and set those elements of the covariance matrix to zero, which correspond to covariances between odd and even moments. We account for possible time dependence in the $\widetilde{u}_t$ by using the Newey-West covariance matrix with the automatic lag length selection proposed by Andrews (1991). Our findings do not change when using predefined lag lengths or a quadratic spectral covariance matrix.

- The Cramér van Mises test[8]

In contrast, the Cramér van Mises statistic cannot account for dependence in $\widetilde{u}_t$. It tests the entire distribution instead of the first central moments by comparing the empirical cumulative distribution of a sample to its theoretical counterpart. Being non-parametric, the test has less power than a parametric test such as Berkowitz (2001). The test statistic is:

$$CvM = \int_0^1 (\widehat{F}(u) - u)^2 du, \tag{15}$$

where $\widehat{F}$ denotes the empirical cumulative density function of the $\widehat{u}_t$ series.

We choose these statistics for a number of reasons. For one, we like to use the same criteria as the earlier studies: The Berkowitz (2001) test is the main test in Bliss and Panigirtzoglou (2004). Linn, Shive, and Shumway (2014) fit the raw moments of the uniform distribution, similarly to the Knüppel (2015) test, in a first model selection step and apply the Cramér van Mises statistic afterwards to decide upon the maximal order of moments. Results based on the Kolmogorov-Smirnov test or relative entropy are very similar to the ones from the Cramér van Mises statistic and are not reported. The Knüppel (2015) test is chosen as it provides a reasonable trade-off between power and precision: While the Berkowitz (2001) test is only able to detect deviations from non-normality through the first and second moments, the Cramér van Mises test, examining the entire distribution, has far less power than the Berkowitz (2001) test. By using the Knüppel (2015) test with four moments, we are able to detect higher moment

---

[8]See Cramer (1928).

differences and still have enough power to test our hypothesis.

3   Data

As the power of the previously mentioned test statistics benefits from large sample sizes, we choose the maximal reasonable time span: data on the European options on the S&P 500 index from January 1988 to August 2015. The data from 1988 to 1995 stem from the Berkeley Options Database, while the data from 1996 to 2015 are provided by OptionMetrics. Although the Berkeley database provides option quotes prior to 1988, we drop these data as there is evidence that the pricing kernel puzzle did not exist in the S&P 500 data before the crash of 1987, see Jackwerth (2000). Both databases also contain the corresponding return series for the underlying as well as dividends and the risk-free rate. We apply the usual filters to the option data insofar as options with moneyness exceeding 1.2 and below 0.8 are excluded from the sample, as well as option prices violating general no-arbitrage constraints, and quotes without volume. Consistent with the literature, we only use out-of-the-money calls and puts and consider a monthly horizon in order to obtain non-overlapping returns, which typically leads to a maturity of 23 calendar days. We end up using $N = 333$ non-overlapping monthly risk-neutral densities $\widehat{q}_t$ and realized returns R_t. In robustness tests, we alternatively use non-overlapping two- and three-month returns.

For the risk-neutral densities, we report in Table 1, Panel A, the annualized average expected return (1.29%) and the annualized average volatility (18.57%) for the full sample. In Panel B, we report the annualized sample average (11.15%) and the sample standard deviation (15.49%) of the realized returns. These values are consistent with a positive equity risk-premium and a negative variance risk-premium. The annualized average risk-free rate (3.68%) is slightly above the average dividend yield (2.28%), which is in line with the low average expected return of the risk-neutral densities.

13

Table 1:
Descriptive Statistics

We report descriptive statistics for risk neutral densities, realized returns, risk-free rates, and dividend yields. Panel A shows the annualized average of the expected returns and the average volatility of the risk-neutral densities. Panel B displays sample means and standard deviations of the realized returns, the risk-free rates, and the dividend yields.

Panel A:

	Average Expected Return	Average Volatility
Risk-Neutral	1.29%	18.57%

Panel B:

	Sample Mean	Sample Standard Deviation
Realized Returns	11.15%	15.49%
Risk-Free Rate	3.68 %	2.62%
Dividend Yield	2.28 %	1.20%

4 Empirical Findings

We next present the empirical findings leading to the conclusion that the pricing kernel puzzle exists in the S&P 500 data, even though we use forward looking information only. We first look at the statistical evidence against pricing kernel monotonicity. Afterwards, we discuss the shape of the empirical pricing kernel estimates and refer to the literature.

4.1 Statistical Evidence

In Table 2, we show the estimation results stemming from the linear pricing kernel interpolation as described in Section 2.2. The values for the optimal test statistics are collected in the columns Unrestricted and Restricted for the respective pricing kernel specification. The distributions of the differences between the two specifications are simulated by the procedure described in Section 2.4. Based on the simulated distributions, we report in the remaining columns of Table 2 p-values corresponding to the differences, assuming that the true pricing kernel is non-increasing and stems from a power utility function with $\gamma = 0$, $\gamma = 2$, and $\gamma = 4$, respectively.

Considering the average log score in the most conservative setting with $\gamma = 0$, we reject

Table 2:
Results for the Linear Pricing Kernel Optimization - Main Table

Estimation results when fitting a piecewise linear pricing kernel, with nine points in the return space and optimized values in the pricing kernel space, to the full sample. The numbers in the second and third columns show the test statistics associated with the measure mentioned in the first column. The p-values of the difference between the unrestricted and restricted test statistic are given in columns four to six. All p-values are obtained by simulation, assuming that the true pricing kernel stems from a power utility function with $\gamma = 0$, 2, and 4, respectively. The values of the Cramér van Mises test are multiplied by one million and the values of the Berkowitz (2001) test are multiplied by 10^{12}.

Statistic	Unrestricted	Restricted	$\gamma = 0$	$\gamma = 2$	$\gamma = 4$
Log Score	2.00	1.98	0.02	0.01	0.01
Knueppel	0.00	9.90	0.02	0.00	0.00
Cramér van Mises	47.31	294.10	0.34	0.06	0.01
Berkowitz	0.57	24.33	0.98	0.92	0.82

the null hypothesis of a non-increasing pricing kernel (p-value 2%). The Knüppel (2015) test confirms this finding (p-value 2%). The Cramér van Mises test achieves a p-value of 34% and cannot reject non-increasing pricing kernels, which could be due to the relatively low power of non-parametric tests compared with parametric tests. The p-value of the Berkowitz (2001) test is large at 0.98. This is not surprising since this test only takes into account the mean and variance of the transformed variables, as it is based on the likelihood function of the normal distribution. Mitchell and Hall (2005) point out that the Berkowitz (2001) test only *"has power to detect non-normality through the first two moments,"* but, e.g., excess skewness and kurtosis are not captured. We confirm this finding by running the Knüppel (2015) test with the first two moments ($L = 2$) only and find that monotonicity cannot be rejected as well. This holds true irrespective of taking care of serial correlation by a proper covariance matrix $\widetilde{\Omega}_L$ or focusing on the moments only; i.e. basing the covariance matrix on the identity matrix.

Before launching into the discussion of the results for the risk averse settings (γ being either 2 or 4), some words of caution are in order. Rejection in such setting is only a rejection of the null hypothesis that the true pricing kernel in the economy derives from a power utility function with the stated risk aversion coefficient. The tests cannot detect monotonicity violations for other utility functions or power utilities with other risk aversion coefficients. Note that this critique does not hold for the case of risk neutrality ($\gamma = 0$), which is a conservative test for

general pricing kernel monotonicity.

We still feel that the risk averse settings are interesting as they are standard reference points for economists. Recall that Bliss and Panigirtzoglou (2004) found $\gamma = 4.08$ as the optimal relative risk-aversion parameter for a power utility when looking at a monthly horizon. With risk-aversion, all p-values are below 10%, except for the Berkowitz (2001) test with p-values of more than 80%. Note that we obtain similar results for pricing kernels based on exponential utility functions.[9]

Overall, for all tests but the Berkowitz (2001) test, the difference between an unrestricted and a restricted pricing kernel is significantly different from zero, and, therefore, we reject the hypothesis of a non-increasing pricing kernel. The Cramér van Mises test does not reject monotonicity in the $\gamma = 0$ setting as, being non-parametric, its power is too limited. The parametric Berkowitz (2001) test, which would be well-suited for small samples in terms of power, ignores moments higher than the second and also fails to reject.

4.2 The Shape of the Empirical Pricing Kernels

Having documented that pricing kernel monotonicity can be rejected in most reasonable settings, we turn our attention to the shape of the empirical pricing kernels. Panel A of Figure 2 repeats Figure 1 and shows the optimal unrestricted and restricted pricing kernel estimates when maximizing the average log score. The optimal pricing kernels from minimizing the Knüppel (2015), Cramér van Mises, and Berkowitz (2001) test statistics are depicted in Panels B, C, and D of Figure 2, respectively.

Except for optimizing the Berkowitz (2001) test, all empirical pricing kernels are broadly u-shaped, consistent with the parametric option pricing model of Christoffersen, Heston, and Jacobs (2013) where past returns are used for inferring the subjective density. Moreover, these three unrestricted u-shaped pricing kernels display a further non-monotonicity at-the-money, similar to the estimates of the canonical pricing kernel papers of Ait-Sahalia and Lo (2000),

---

[9]The pricing kernel derived from an exponential utility function is given by $m^{exp\cdot} = \exp(-a \cdot R_t)$. Bliss and Panigirtzoglou (2004) identified $a = 6.33$ as the optimal value when looking at a monthly horizon and so we repeated our base procedure for $a = 2, 4, 6, 8,$ and 10. Again, all p-values decrease in a. All our tests, except the Berkowitz (2001) test, reject pricing kernel monotonicity at least at the 10% level.
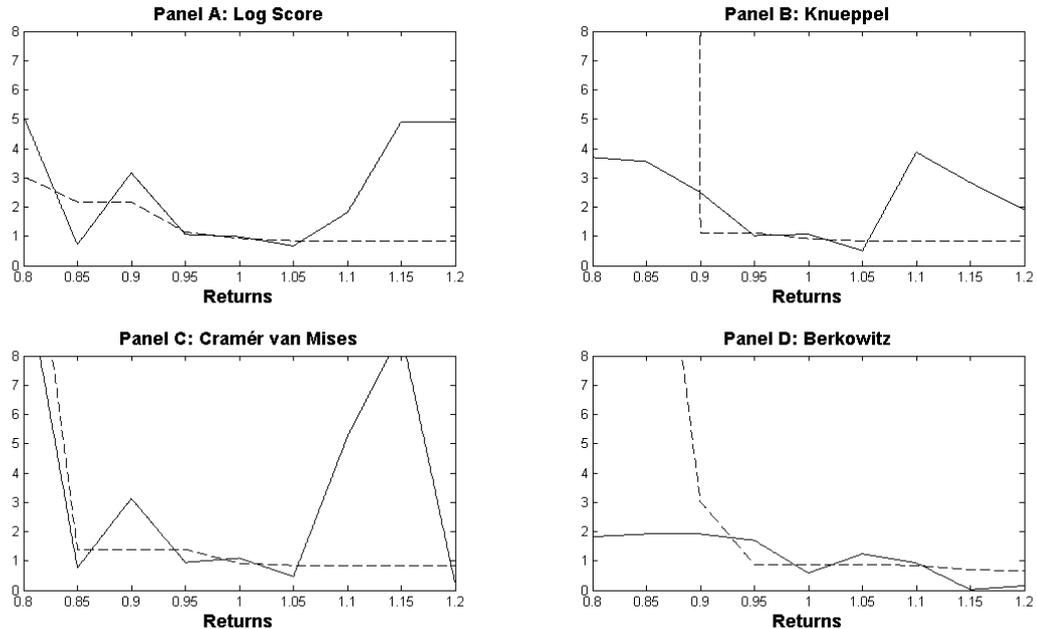
Figure 2: Optimal Pricing Kernels for Different Measures. We plot unrestricted scaled pricing kernels (solid) and a non-increasing scaled pricing kernels (dotted), constructed by our linear interpolation rule on the full sample. Panel A shows the pricing kernels when maximizing the average log score. Panel B shows the pricing kernels when minimizing the Knueppel test statistic. Similarly for Panel C and D for the Cramér van Mises and the Berkowitz test statistics. Returns are on the horizontal axis, and the values of the pricing kernel are on the vertical axis.

Jackwerth (2000), and Rosenberg and Engle (2002).

Bakshi, Madan, and Panayotov (2010), Polkovnichenko and Zhao (2013), and Chabi-Yo (2012) develop theories for u-shaped pricing kernels. Bakshi, Madan, and Panayotov (2010) assume heterogeneity among investors, and pessimists shorting the market lead to increases in the pricing kernel for increasing returns larger than some threshold. Polkovnichenko and Zhao (2013) claim that investors overweight the subjective probability of high and low returns, which again yields a u-shaped pricing kernel. Finally, Chabi-Yo (2012) describes an economy where skewness and kurtosis preferences enter the pricing kernel. His empirical pricing kernel, estimated from the returns of 30 monthly industry portfolios displays a u-shape, too, when being projected on the market return. The Berkowitz (2001) test, ignoring moments higher than the second, could not reject pricing kernel monotonicity. Such failure might be related to the existence of higher moment preferences in the pricing kernel. In the next section we document that the u-shape is very persistent and survives various robustness checks.

17

5 Robustness

We show that our results are robust to a number of changes, namely, changing the sample period, using alternative pricing kernel specifications (spline instead of piece-wise linear), and employing different methods for the risk-neutral densities (Gatheral (2004) instead of Jackwerth (2004)). Concerning lengthening horizons, the pricing kernel puzzle weakens for longer horizons. Finally, we investigate exactly why Linn, Shive, and Shumway (2014) obtain monotonically decreasing pricing kernels.

5.1 Alternative Sub-Samples

An important issue is the choice of the sample period. As there is no evidence for the pricing kernel puzzle in the S&P 500 data before the crash of 1987, we choose January 1988 as our starting date. The largest available sample then extends until August 2015 and is used in the base case. A large sample makes our study more reliable from a statistical point of view, but the assumption of time invariant preferences might be tenuous over a period of almost 30 years. As volatility is considered as a relevant state variable for preferences, see e.g. Brown and Jackwerth (2012), we follow Bliss and Panigirtzoglou (2004) by dividing our sample into a high and a low volatility subsample. The dividing line is the median of the monthly risk-neutral volatilities from 1988 to 2015. We report the results in Table 3, columns 'High Vol' and 'Low Vol'. We simulate all p-values using the procedure described in Section 2.4. Panels A, B, and C correspond to the settings where we assume that the true pricing kernel is non-increasing and stems from a power utility function with $\gamma = 0$, 2, and 4, respectively.

For most settings in the high volatility sample, we still reject the null of a non-increasing pricing kernel, see Table 3, 'High Vol'. Compared to the base setting, only the Knüppel (2015) test now fails to reject monotonicity with a p-value of 19% in the risk-neutral simulation setting. For the low volatility sample (see Table 3, 'Low Vol'), the log score and the Cramér van Mises test turn insignificant, while the Knüppel (2015) test holds up well. Interestingly, the Berkowitz (2001) test is able to reject pricing kernel monotonicity in the low volatility sample with a p-value

Table 3:
Alternative Sub-Samples

Estimation results when fitting a piecewise linear pricing kernel, with nine points in the return space and optimized values in the pricing kernel space, to alternative samples. We show the p-values of the difference between the unrestricted and restricted test statistics mentioned in the first column. We simulate all p-values, assuming that the true pricing kernel stems from a power utility function with $\gamma = 0$ in Panel A, $\gamma = 2$ in Panel B, and $\gamma = 4$ in Panel C. Column 2 of each panel contains the basic setting from Table 2. In the other columns, we use different samples: 'High Vol' uses all days with above median volatility, and 'Low Vol' uses the remaining days. '1996-2012' and '1996-2015' use the respective years.

Panel A: $\gamma = 0$

Statistic	Basic Setting	High Vol	Low Vol	1996-2012	1996-2015
Log Score	0.02	0.03	0.26	0.01	0.01
Knüppel	0.02	0.19	0.04	0.01	0.04
Cramér van Mises	0.34	0.21	0.25	0.15	0.31
Berkowitz	0.98	0.95	0.09	0.73	0.96

Panel B: $\gamma = 2$

Statistic	Basic Setting	High Vol	Low Vol	1996-2012	1996-2015
Log Score	0.01	0.01	0.16	0.00	0.00
Knüppel	0.00	0.06	0.02	0.00	0.01
Cramér van Mises	0.06	0.09	0.17	0.05	0.11
Berkowitz	0.92	0.85	0.05	0.43	0.82

Panel C: $\gamma = 4$

Statistic	Basic Setting	High Vol	Low Vol	1996-2012	1996-2015
Log Score	0.01	0.01	0.12	0.00	0.00
Knüppel	0.00	0.03	0.01	0.00	0.00
Cramér van Mises	0.01	0.03	0.09	0.01	0.03
Berkowitz	0.82	0.77	0.03	0.33	0.68

of 9% for $\gamma = 0$ and even stronger results for higher γ.[10]

Another way of relaxing the assumption on time invariant preferences is to calibrate our model to the sub-sample used by Linn, Shive, and Shumway (2014): September 1996 to December 2012, so that we can relate our results to their claim of monotonically decreasing pricing kernels. The results are reported in Table 3, '1999-2012'. Still, we reject non-increasing pricing kernels at the 1% level for the log score and the Knüppel (2015) test. The Cramér van Mises test even improves in this sample and generates a marginally insignificant p-value of 15% in the

---

[10]The Berkowitz (2001) test might gain in power as the higher moments of the returns realized in this subsample are less extreme. Compared to the full sample, sample kurtosis drops from 7.18 to 3.74 and sample skewness is less negative: -1.24 in the full sample versus -0.66 in the low volatility subsample.
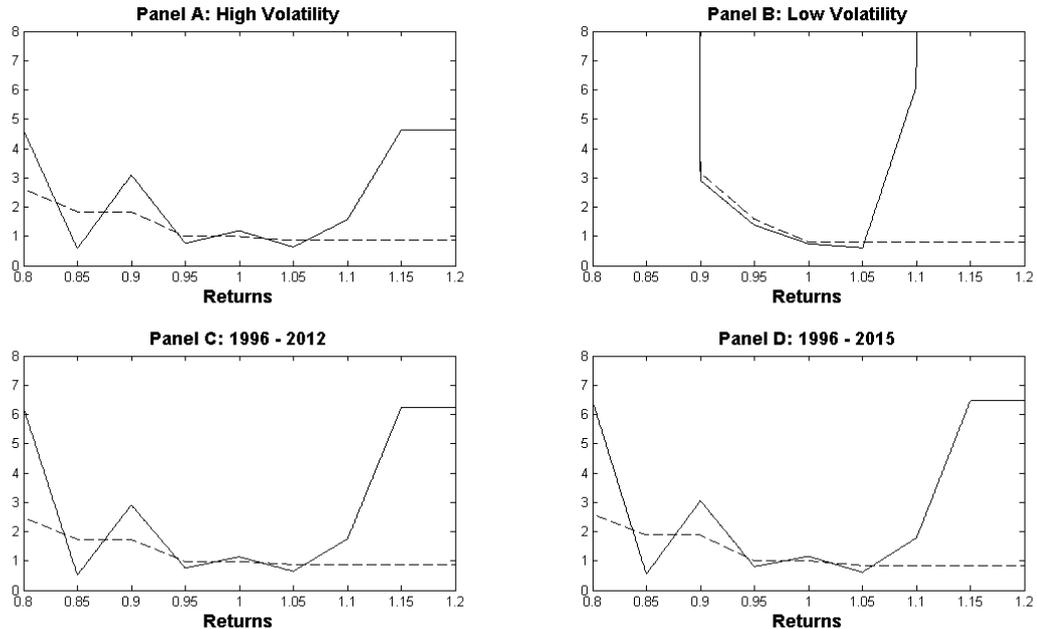
Figure 3: Optimal Pricing Kernels on Different Sub-Samples. We plot unrestricted scaled pricing kernels (solid) and a non-increasing scaled pricing kernels (dotted), constructed by our linear interpolation rule and maximizing the average log score as given in equation (4). The Panels A and B show the pricing kernels estimated on the high and low volatility samples. Panel C shows the pricing kernels estimated on the January 1996 - December 2012 sample; Panel D shows the longer sample until August 2015. Returns are on the horizontal axis, and the values of the pricing kernel are on the vertical axis.

$\gamma = 0$ setting. We return to the reasons for our disagreement with Linn, Shive, and Shumway (2014) in Section 5.5. The results for the whole OptionMetrics sample in Table 3, '1999-2015', are rather similar.

In conclusion, monotonicity violations seem to be present in all subsamples. The moderate increases in p-values could well be due to the smaller subsamples. In Figure 3, we depict the optimal pricing kernels (based on the log score) for our subsamples. Their shapes hardly change compared to the full sample, depicted in Figure 1, except for the low volatility case, where the unrestricted pricing kernel is even more u-shaped.

5.2 Alternative Pricing Kernel Modeling

As most theoretical pricing kernels are assumed to be smooth functions and to mimic the methodology of Linn, Shive, and Shumway (2014) even more closely, we replace the linear interpolation of the pricing kernel with cubic splines. To guarantee the non-negativity of all

constructed pricing kernels, we first interpolate the logarithm of the eligible pricing kernel values and afterwards take the exponential of the interpolated values. The results of this calibration are given in Table 4, 'Cubic Spline'. Our findings do not change much even though the test is more conservative when using cubic splines: In order to have simple constraints for estimating the restricted version of our pricing kernels, we label a tentative pricing kernel as non-increasing whenever the values of the pricing kernel at the spline knots are non-increasing. Hence, all non-increasing pricing kernels are consistent with our definition of a restricted kernel, but a restricted kernel constructed by cubic splines can have slightly increasing segments due to the curvature of the cubic splines.[11] A restricted pricing kernel can somewhat mimic the shape of an unrestricted pricing kernel and improve its fit to the respective criteria. This is for example the case when optimizing the Knüppel (2015) test: The optimal restricted cubic spline pricing kernel is slightly increasing in the right end, raising the p-value to 9% in the $\gamma = 0$ setting. Rejection of non-increasing pricing kernels based on the cubic spline interpolation is therefore even more meaningful as we can clearly distinguish between small fluctuations and pronounced increasing segments.

Our results are also not driven by over-fitting, as we can see from the results from estimating the pricing kernel with only 5 points, see Table 4, '5 Points'. Changing the location of our 9 points in the base case does not affect results, either, see Table 4, 'Different Grid'. In that case, we use the midpoints of the basic grid (0.825 0.875 0.925 0.975 1.025 1.075 1.125 1.175) and only the average log score turns out to be marginally insignificant with a p-value of 12% in the $\gamma = 0$ setting.

5.3 Alternative Risk Neutral Density Modeling

In the base case, we extrapolate the implied volatility curve by the fast and stable method of Jackwerth (2004). As an alternative, we use the Stochastic Volatility Inspired (SVI) model of Gatheral (2004) instead. As the name already indicates, its structural form is related to models such as Heston (1993). The model is fully parameterized and does not have a smoothness

---

[11]In case of linear interpolation, a pricing kernel is non-decreasing if and only if its values at the spline knots are non-increasing.

Table 4:
Alternative Modeling

Estimation results when fitting alternative pricing kernels or risk-neutral densities obtained by the SVI model on the full sample. We show the p-values of the difference between the unrestricted and restricted test statistics mentioned in the first column. We simulate all p-values, assuming that the true pricing kernel stems from a power utility function with $\gamma = 0$ in Panel A, $\gamma = 2$ in Panel B, and $\gamma = 4$ in Panel C. Column 2 of each panel contains the basic setting from Table 2. We present results for cubic spline interpolation instead of linearly interpolated pricing kernels ('Cubic Spline'), for pricing kernels spanned by only 5 instead of 9 points ('5 Points'), for a shifted set of grid values on the horizontal axis ('Different Grid'), and for using the SVI method of Gatheral (2004) instead of the fast and stable method of Jackwerth (2004) ('Gatheral SVI').

Panel A: $\gamma = 0$

Statistic	Basic Setting	Cubic Spline	5 Points	Different Grid	Gatheral SVI
Log Score	0.02	0.02	0.01	0.12	0.02
Knüppel	0.02	0.09	0.01	0.04	0.03
Cramér van Mises	0.34	0.38	0.27	0.46	0.39
Berkowitz	0.98	0.99	0.95	0.99	0.93

Panel B: $\gamma = 2$

Statistic	Basic Setting	Cubic Spline	5 Points	Different Grid	Gatheral SVI
Log Score	0.01	0.00	0.00	0.04	0.00
Knüppel	0.00	0.02	0.00	0.01	0.01
Cramér van Mises	0.06	0.08	0.06	0.14	0.08
Berkowitz	0.92	0.93	0.83	0.96	0.74

Panel C: $\gamma = 4$

Statistic	Basic Setting	Cubic Spline	5 Points	Different Grid	Gatheral SVI
Log Score	0.01	0.00	0.00	0.02	0.00
Knüppel	0.00	0.01	0.00	0.00	0.00
Cramér van Mises	0.01	0.01	0.01	0.03	0.02
Berkowitz	0.82	0.84	0.76	0.93	0.62

parameter, such as λ from equation (2) in the case of the fast and stable method, which can be chosen by the econometrician. The parameterization is as follows:

$$\sigma_{SVI}^2(K|a, b, \rho, m, \theta) = a + b\{\rho(\log(K/S_t) - m) + \sqrt{(\log(K/S_t) - m)^2 + \theta^2}\} \qquad (16)$$

Here, a is the average squared implied volatility and has to be positive. The non-negative parameter b measures the angle between the slope of in-the-money to out-of-the-money volatilities. ρ and m are location parameters, and θ controls the curvature around at-the-money. In order to fit the implied volatility curve (16) to the data, we numerically minimize the squared

difference between model implied volatilities and implied volatilities observed in the market:

$$\{a^*, b^*, \rho^*, m^*, \theta^*\} = \underset{\{a,b,\rho,m,\theta\}}{\arg\min} \sum_{i=1}^{I} \left(\sigma_{SVI}(K_i|a,b,\rho,m,\theta) - \bar{\sigma}_i\right)^2. \tag{17}$$

The smooth implied volatility curve is given by

$$\{\sigma_{SVI}(K_j|a^*, b^*, \rho^*, m^*, \theta^*)\}_{j=0}^{J}, \tag{18}$$

and the risk-neutral densities are again obtained by means of Breeden and Litzenberger (1978).

In Table 4, 'Gatheral SVI', we report the results for a piece-wise linear pricing kernel and risk-neutral densities stemming from the SVI model. Again, assuming risk-aversion, non-increasing pricing kernels can be rejected at the 10% level with the exception of the Berkowitz (2001) test. For the conservative version with $\gamma = 0$, the Cramér van Mises statistic fails to reject with a p-value of 39%, similar to the base case when using the fast and stable method.

5.4 Different Horizons

The pricing kernel puzzle arises from the different shapes of subjective and risk-neutral distributions as the former tend to be more (log-)normally shaped while the latter are heavily leptokurtic and left-skewed. At longer horizons, the risk-neutral distributions look more (log-)normally shaped and we thus suspect that the pricing kernel puzzle might be weaker.

To check, we repeat our base procedure for horizons of two months, $T = 51$, and three months, $T = 79$. As we want to stick to non-overlapping returns, we can choose either January or February as the starting months when looking at a two-month horizon. For the three months setting, we can choose between January, February, and March as the starting months. This results in two sets of $N = 164$ and three sets of $N = 95$ non-overlapping risk-neutral densities and their corresponding non-overlapping two-month and three-month returns. Table 5 displays the corresponding results for the case of $\gamma = 4$. As expected, monotonicity is the harder to reject, the longer the considered horizon is. We do not report the results for lower γ values, as all p-values tend to increase in that case; further weakening the results. For $\gamma = 0$, all p-values

23

are larger than 10%.

We conclude that the pronounced skew of the short term implied volatilities of S&P 500 index options, inducing left-skewness and leptokurtosis, paired with the time series properties of the monthly returns, causes the non-decreasing pricing kernel estimates at the short, monthly horizon. The pricing kernel puzzle is much attenuated at the longer two-month horizon and hardly detectable at the three-month horizon.

Table 5:
Different Horizons

Estimation results when fitting a piecewise linear pricing kernel, with nine points in the return space and optimized values in the pricing kernel space, on different horizons. We show p-values of the difference between the unrestricted and restricted test statistics mentioned in the first column. Column 2 contains the basic setting from Table 2. We report the results for non-overlapping two months returns in the columns '2 Months, Jan' and '2 Month, Feb'. We report the results for non-overlapping three months returns in the columns '3 Months, Jan', '3 Months, Feb', and '3 Months, Mar'. We simulate all p-values, assuming that the true pricing kernel stems from a power utility function with $\gamma = 4$.

Statistic	Basic Setting	2 Months		3 Months		
		Jan	Feb	Jan	Feb	Mar
Log Score	0.01	0.13	0.14	0.06	0.13	0.43
Knüppel	0.00	0.07	0.17	0.16	0.30	0.17
Cramér van Mises	0.01	0.09	0.17	0.18	0.14	0.20
Berkowitz	0.82	0.39	0.02	0.04	0.03	0.11

5.5 Degenerate Densities

Last, we are interested in why Linn, Shive, and Shumway (2014) obtain a monotonically decreasing pricing kernel. We apply two methodological changes which we believe cause their results. First, Linn, Shive, and Shumway (2014) do not force the resulting subjective densities to integrate to one; we follow them by no longer scaling the subjective density. Second, they utilize a two-step model selection procedure: The first step consists of fitting the pricing kernel to the first L moments of the uniform distribution by minimizing the following objective function[12]

$$GMM(L) = \sum_{l=1}^{L} \left(\frac{1}{N} \sum_{t=1}^{N} \widetilde{u}_t^l - \frac{1}{l+1} \right)^2 . \tag{19}$$

---

[12]This GMM approach is equivalent to setting $\widetilde{\Omega}_L$ to the identity matrix in Knüppel (2015).
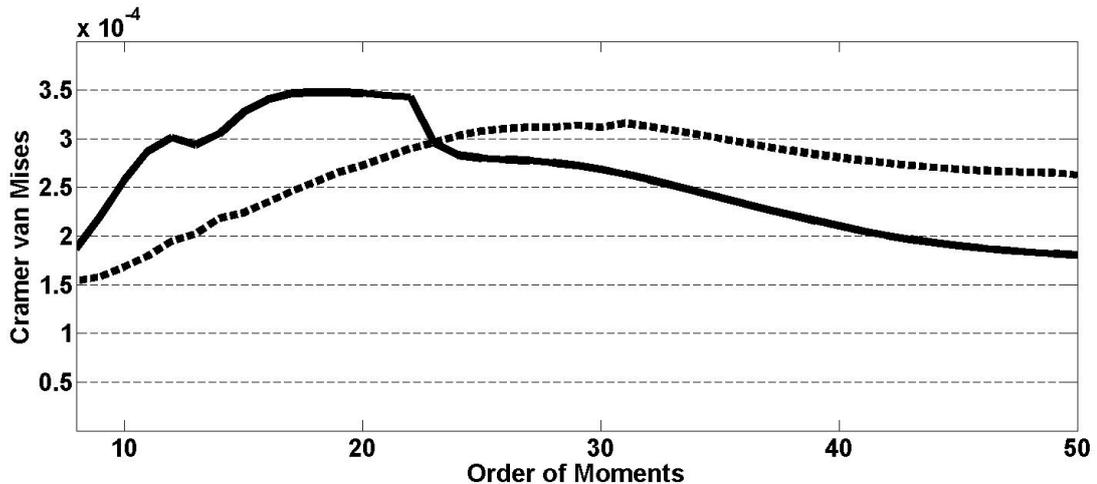
Figure 4: Moment Fit and Cramér van Mises Statistic. We estimate pricing kernels by fitting the first L moments of the uniform distribution without scaling the resulting subjective densities. For each resulting pricing kernel, we calculate the corresponding Cramér van Mises statistic. The Cramér van Mises values stemming from the unrestricted (restricted) pricing kernels are plotted with a solid (dashed) line. The number of moment conditions L is on the horizontal axis.

This optimization is implemented for up to $L = 50$ moment conditions, providing a variety of potential pricing kernels. For each of these pricing kernels, the Cramér van Mises statistic can be calculated, and the second model selection step consists of choosing the pricing kernel delivering the lowest Cramér van Mises statistic.

We again mimic their procedure with our linear pricing kernels without scaling the subjective densities, and we depict the relationship between the number of fitted moments L and the associated Cramér van Mises statistic in Figure 4. We also follow Linn, Shive, and Shumway (2014) by ignoring the fit to moments of order smaller than the number of pricing kernel parameters as, according to the authors, this leads to an under-identified GMM equation system. As our linear pricing kernel is a function of eight parameters, we start at $L = 8$ and fit unrestricted (solid line) as well as restricted (dashed line) pricing kernels to our full sample, leaving us with $86 \ (= 43 \cdot 2)$ potential estimates.

Regarding Figure 4, it is not only surprising that the overall minimum is obtained by a restricted pricing kernel for $L = 8$, but the restricted versions tend to provide better Cramér van Mises statistics for L less than 23. This is surprising as one would normally expect a restricted optimum to perform at most as well as an unrestricted optimum. But, by ignoring the scaling of the subjective densities and using the two-step model selection, which mixes two

25

different measures of fit, a monotonically decreasing pricing kernel can be identified as the best forward looking pricing kernel estimate by Linn, Shive, and Shumway (2014).

Looking back at the integral of equation (8), we can see why Linn, Shive, and Shumway (2014) might not want to scale the subjective densities: In order to obtain the transformed realized returns series $\widetilde{u}_t$, one has to evaluate the integrand only up to the maximal realized return in the sample.[13] The advantage of their approach is that they do not have to model the pricing kernel for returns where no realized return is observed. Moreover, their pricing kernel is literally constant through time, whereas our pricing kernel is scaled by a time varying factor.

However, the two modifications have to be seen critically: not scaling the subjective densities clashes with our intuition that probabilities should sum up to one and can make the $\widetilde{u}_t$ series lie outside the unit interval. Moreover, for consistency, the same loss function should be used for estimating and evaluating a model, see Christoffersen and Jacobs (2004), which speaks against their two-step procedure of GMM and Cramér van Mises in combination.

We want to emphasize that only the combination of not scaling the subjective densities and the two-step procedure leads to a monotonically decreasing pricing kernel. First, when replicating the results shown in Figure 4 with scaled subjective densities, the surprising finding disappears: For any number L of moment conditions, the unrestricted pricing kernel is now delivering a lower Cramér van Mises statistic than the restricted pricing kernel, i.e., the solid line is always below the dotted line. Second, when leaving the subjective densities unscaled but minimizing the GMM objective function, an unrestricted pricing kernel again delivers a lower GMM value than a restricted kernel. The same holds if we use only the Cramér van Mises statistic instead.

In conclusion, both, the two-step procedure with scaled subjective densities and a single objective function minimization with unscaled subjective densities would identify a non-monotonically decreasing pricing kernel as the best estimate. This is in line with the results of our own procedure, when optimizing a single criteria and scaling the subjective densities.

---

[13]The maximal realized return in our sample is at about 1.11.

6 Conclusion

By using forward looking information only (i.e., option prices and future realized returns), we confirm the existence of the pricing kernel puzzle in the S&P 500 data. This contradicts the results in Linn, Shive, and Shumway (2014), which we attribute to a lack of scaling and a particular choice of two-step optimization in their paper. Potential pricing kernels are modeled as piece-wise linear functions and cubic splines. Forward looking subjective densities are constructed by dividing the forward looking risk-neutral density by a time invariant pricing kernel. We make use of the average log score, as well as three alternative distributional tests, to asses if the forward looking subjective densities are compatible with the future realized returns of the S&P 500. By varying the potential pricing kernel, we can optimize the unrestricted test statistics. We also optimize the test statistics in a restricted version, where the pricing kernel needs to be non-increasing.

We formally test for monotonicity based on the difference between the unrestricted and the restricted test statistics. The average log score and the Knüppel (2015) test reject pricing kernel monotonicity with a p-value of 2%. Non-rejection with the latter two criteria occurs only very rarely for subsamples. The Berkowitz (2001) test is the only test that does not reject non-increasing pricing kernels. This tests takes into account only the first two moments and ignores all higher moments. Yet, the importance of non-normal skewness and kurtosis is well documented in the financial time series literature. Due to its low power in small samples, the Cramér van Mises statistic does not reject pricing kernel monotonicity in the most conservative setting, where we simulate a risk-neutral economy.

Overall, our results are robust to variations in the methodology and the sample. We show that pricing kernel monotonicity cannot be rejected when looking at longer horizons, where subjective and risk-neutral densities are less negatively skewed and less kurtotic. Our findings suggest that solutions to the pricing kernel puzzle should therefore consider higher moment preferences.

References

AIT-SAHALIA, Y., AND A. W. LO (2000): "Nonparametric Risk Management and Implied Risk Aversion," *Journal of Econometrics*, 94, 9–51.

ANDREWS, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

BAKSHI, G., D. B. MADAN, AND G. PANAYOTOV (2010): "Returns of Claims on the Upside and the Viability of U-Shaped Pricing Kernels," *Journal of Financial Economics*, 97, 130–157.

BARONE-ADESI, G., AND H. DALL'O (2010): "Is the Pricing Kernel Monotone?," Working paper, Swiss Finance Institute.

BARONE-ADESI, G., R. F. ENGLE, AND L. MANCINI (2008): "A GARCH Option Pricing Model with Filtered Historical Simulation," *Review of Financial Studies*, 21, 1223–1258.

BEARE, B., AND J.-M. MOON (2015): "Nonparametric Tests of Density Ratio Ordering," *Econometric Theory*, 31, 471–492.

BEARE, B., AND L. SCHMIDT (2014): "An Empirical Test of Pricing Kernel Monotonicity," *Journal of Applied Econometrics*, 31, 338–356.

BERKOWITZ, J. (2001): "Testing Density Forecasts with Applications to Risk Management," *Journal of Business and Economic Statistics*, 19, 465–474.

BLISS, R. R., AND N. PANIGIRTZOGLOU (2004): "Option-Implied Risk Aversion Estimates," *Journal of Finance*, 59, 407–446.

BREEDEN, D. T., AND R. H. LITZENBERGER (1978): "Prices of State-Contingent Claims Implicit in Option Prices," *Journal of Business*, 51, 621–651.

BROWN, D. P., AND J. C. JACKWERTH (2012): "The Pricing Kernel Puzzle: Reconciling Index Option Data and Economic Theory," *Contemporary Studies in Economic and Financial Analysis*, 94, 155–183.

CHABI-YO, F. (2012): "Pricing Kernels with Stochastic Skewness and Volatility Risk," *Management Science*, 58, 624–640.

CHRISTOFFERSEN, P., S. HESTON, AND K. JACOBS (2013): "Capturing Option Anomalies with a Variance-Dependent Pricing Kernel," *Review of Financial Studies*, 26, 1962–2006.

CHRISTOFFERSEN, P., AND K. JACOBS (2004): "The Importance of the Loss Function in Option Valuation," *Journal of Financial Economics*, 72, 291–318.

COCHRANE, J. H. (2000): *Asset Pricing.* Princeton University Press, Princeton, USA, 1 edn.

CRAMER, H. (1928): "On the Composition of Elementary Errors," *Scandinavian Actuarial Journal*, 59, 141 ff.

CUESDEANU, H., AND J. C. JACKWERTH (2016): "The Pricing Kernel Puzzle: Survey and Outlook," Working paper, University of Konstanz.

GATHERAL, J. (2004): "A Parsimonious Arbitrage-Free Implied Volatility Parameterization with Application to the Valuation of Volatility Derivatives," Working paper, Presentation at Global Derivatives.

GNEITING, T., AND A. RAFTERY (2007): "Strictly Proper Scoring Rules, Prediction, and Estimation," *Journal of the American Statistical Association*, 102, 359–378.

GOLUBEV, Y., W. K. HAERDLE, AND R. TIMOFEEV (2014): "Testing Monotonicity of Pricing Kernels," *Advances in Statistical Analysis*, 98, 305–326.

HESTON, S. L. (1993): "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies*, 6, 327–343.

JACKWERTH, J. C. (2000): "Recovering Risk Aversion from Option Prices and Realized Returns," *Review of Financial Studies*, 13, 433–451.

——— (2004): "Option-Implied Risk-Neutral Distributions and Risk-Aversion," *Research Foundation of AIMR*, Charlotteville, USA.

KNÜPPEL, M. (2015): "Evaluating the Calibration of Multi-Step-Ahead Density Forecasts Using Raw Moments," *Journal of Business & Economic Statistics*, 33, 270–281.

LINN, M., S. SHIVE, AND T. SHUMWAY (2014): "Pricing Kernel Monotonicity and Conditional Information," Working paper, University of Michigan.

MITCHELL, J., AND G. S. HALL (2005): "Evaluating, Comparing and Combining Density Forecasts Using the KLIC with an Application to the Bank of England and NIESR Fan Charts of Inflation," *Oxford Bulletin of Economics and Statistics*, 67, 995–1033.

POLKOVNICHENKO, V., AND F. ZHAO (2013): "Probability Weighting Functions Implied in Options Prices," *Journal of Financial Economics*, 107, 580–609.

ROSENBERG, J. V., AND R. F. ENGLE (2002): "Empirical Pricing Kernels," *Journal of Financial Economics*, 64, 341–372.

SHACKLETON, M. B., S. J. TAYLOR, AND P. YU (2010): "A Multi-Horizon Comparison of Density Forecasts for the S&P 500 Using Index Returns and Option Prices," *Journal of Banking and Finance*, 34, 2678–2693.