Pricing Kernel Monotonicity and Conditional Information

Matthew Linn
University of Massachusetts

Sophie Shive
University of Notre Dame

Tyler Shumway
University of Michigan

A large literature finds evidence that pricing kernels nonparametrically estimated from option prices and historical returns are not monotonically decreasing in market index returns. We argue that existing estimation methods are inconsistent and propose a new nonparametric estimator of the pricing kernel that reflects the information available to investors who set asset prices. In simulations, the estimator outperforms existing techniques. Our empirical estimates using S&P 500 index option data from 1996 to 2014 and FTSE 100 index option data from 2002 to 2014 suggest that the "pricing kernel puzzle" is due to flaws in existing estimators rather than a behavioral or economic phenomenon. (*JEL* G12, G13)

Received August 2, 2015; editorial decision April 15, 2017 by Editor Andrew Karolyi.

It is well known that the absence of arbitrage implies the existence of a positive pricing kernel, or stochastic discount factor (SDF), that prices all assets. Almost all models of the trade-off between risk and return specify a pricing kernel that decreases monotonically with the quality of the state of the world. The state of the world is often modeled as a function of the change in aggregate wealth, which is measured by the return on a broad stock market index. A number of researchers combine index option data with historical returns to estimate the pricing kernel nonparametrically, but the kernels they estimate are generally not monotonic functions of the market return. We argue that most of the methods used to estimate the pricing kernel produce estimates

We thank four anonymous referees and the editor for all their help. We also thank seminar participants at Indiana University, the University of Michigan, the University of Washington, the University of Houston, the WFA (2014) and the NTU Finance Conference (2016). We also thank Robert Dittmar, Wayne Ferson, Stephen Figlewski, Steven Heston (WFA disscussant), Stefan Nagel, and Xingyu Zhu for helpful comments. Shumway is grateful for the financial support provided by National Institute on Aging [grant 2-P01-AG026571]. Send correspondence to Tyler Shumway, Ross School of Business, University of Michigan, 701 Tappan Street, Ann Arbor, MI 48109, USA; telephone: (734) 763-4129. E-mail: shumway@umich.edu

doi:10.1093/rfs/hhx095 Advance Access publication August 21, 2017

that do not converge to the true pricing kernel as the sample size grows. Most estimates are inconsistent for the true pricing kernel because they compare a forward-looking, conditional risk-neutral density estimated with option prices to a backward-looking, essentially unconditional physical density estimated with historical returns. We propose a new, completely nonparametric pricing kernel estimator that explicitly accounts for the fact that option prices should reflect all information available. The new estimator suggests that the pricing kernel is a monotonic function of stock market return realizations.

Because the pricing kernel summarizes the attitudes of economic agents about risk, understanding its properties is one of the primary goals of asset pricing. Breeden and Litzenberger (1978) show that the second derivative of the price of a call option with respect to the strike price is proportional to the risk-neutral density. Both Jackwerth (2000) and Ait-Sahalia and Lo (2000) cleverly use this fact to estimate the risk-neutral density with market index option prices for different strike prices. Then they divide the resultant risk-neutral density by a nonparametric estimate of the physical density based on historical return data. The resultant ratio of densities is what we refer to as the "classical" nonparametric SDF estimator. Existing research has found that it is typically a decreasing function of the market return over much of its range, but it is also often increasing over part of its range. Important improvements in the classical method over time have not changed this result.[1]

If the pricing kernel is truly increasing in some range of aggregate wealth, then the marginal value of a dollar is higher when markets rise than when they fall over that range. This is the opposite of most economists' intuition. A nonmonotonic pricing kernel is so surprising that it has been coined the "implied risk aversion puzzle" or the "pricing kernel puzzle" in the long stream of papers that attempt to explain it.

One set of explanations for the pricing kernel puzzle is that additional factors that are strongly related to the market return and have large risk premiums cause the projection of the pricing kernel onto market returns to be nonmonotonic. For example, Chabi-Yo, Garcia, and Renault (2007)[2] propose a model that includes stochastic volatility as a factor and can be consistent

[1] Seminal papers in this literature include Jackwerth (2000), Ait-Sahalia and Lo (2000), and Rosenberg and Engle (2002). More recent papers include Chaudhuri and Schroder (2009), Audrino and Meier (2012), Härdle, Okhrin, and Wang (2014), and Beare and Schmidt (2016). One paper that does not appear to find an upward-sloping kernel is Barone-Adesi, Engle, and Mancini (2008). Using data from January 2002 to December 2004, and adjusting the variance of the physical distribution using a GARCH model, they find a pricing kernel that appears to be decreasing.

[2] Chabi-Yo, Garcia, and Renault (2007) incorporate a regime-switching model with two regimes corresponding to high and low volatility states. Positive returns increase wealth but they also indicate a relatively high probability of the high volatility state. The model can produce a nonmonotonic SDF only if investors are sufficiently averse to the high volatility state, which requires a large risk premium. The absolute risk aversion function for their baseline model appears to be very close to zero in the low volatility regime, but it is uniformly above 20 in the high volatility regime.

with some nonmonotonicity.[3] Another set of explanations holds that investor characteristics drive nonmonotonicity. Ziegler (2007) attributes the puzzle to differences in risk preferences among agents.[4] The nonmonotonicity to be explained is surprisingly large. The time-invariant (average) SDF estimates of Rosenberg and Engle (2002) imply that, on average, investors strongly prefer a loss of four percent over the next month to a gain of two percent. It is challenging to think of an explanation that might cause agents to prefer market losses to gains month after month. Our explanation is that existing estimators do not adequately account for the information that investors use to price assets.[5]

Our primary criticism of almost all of the existing empirical nonmonotonicity papers is that they compare mismatched information sets: a conditional risk-neutral density to an essentially unconditional physical density. Because option prices, like all market-determined prices, are discounted expectations of future cash flows conditional on all information available, the option-implied risk-neutral density is conditional on all information. Market participants have a great deal of information on which they can condition, including past and current prices of all securities, a large amount of political and business news, macroeconomic data, accounting data for all firms, the current calendar and the schedule of many future events, and trading conditions in numerous exchanges. By contrast, most nonmonotonicity papers use only past index returns to estimate the physical density. It is difficult to believe that all of the information market participants use to set prices, and hence the physical density, can be captured with one time series of past index returns. We show that unless this is the case, the classical method produces inconsistent estimates that can deviate materially from the true SDF.

It is important to note that our econometric explanation of the puzzle is quite distinct from the additional factors explanation of the puzzle. The additional factors explanation claims that the projection of the pricing kernel onto returns is truly not monotonic, and holds that a few additional factors, which must be strongly (possibly nonlinearly) related to market returns, explain the otherwise puzzling nonmonotonicity. We claim that the projection of the SDF onto returns can be decreasing and that option price setters condition on many variables, making it close to impossible to adequately model the conditional distribution

---

[3] Christoffersen, Heston, and Jacobs (2013) and Song and Xiu (2016) also use volatility as a factor to explain nonmonotonicity.

[4] Ziegler also examines differences in beliefs about the mean and variance of expected returns. Polkovnichenko and Zhao (2013) postulate a model with rank-dependent utility. Barone-Adesi, Mancini, and Shefrin (2013) explain the puzzle with overconfidence, and Grith, Härdle, and Krätschmer (2017) propose heterogeneity of investor reference points.

[5] In related research, Bakshi, Madan, and Panayotov (2010) find that average index option returns in several countries are consistent with a U-shaped pricing kernel, and Constantinides, Czerwonko, Jackwerth, and Perrakis (2011) show that index options that seem to violate stochastic dominance bounds appear to offer profitable trading opportunities. These papers are consistent with a nonmonotonic pricing kernel, but they do not directly estimate the kernel and the noise in average returns makes it difficult for them to draw strong conclusions. Furthermore, the results of our paper suggest that the risk of options is likely to vary substantially over time in potentially complex ways, so properly adjusting option returns for risk is very difficult.

of future returns. Although both explanations claim that additional conditioning information is necessary, the factors explanation is about correctly modeling risk, whereas ours is about correctly matching conditional information sets.

We propose a new method that avoids the mismatch created by comparing conditional risk-neutral densities to historical data, and creates an estimate that is fully conditional on all moments of the forward-looking distributions. Our method exploits the insight that, at any given time, the conditional density of the future market return is only the density for that particular return realization. We can think of the observations we have as a series of risk-neutral densities accompanied by a corresponding series of return realizations, with each period's risk-neutral density being different and with only one realization corresponding to each density. Given these data, we can integrate each risk-neutral density up to each of their corresponding actual return realizations to obtain a set of realized CDF values, or cumulants. If the risk-neutral density is the same as the physical density, the resultant cumulants should be uniformly distributed. To the extent that the empirical distribution of the cumulants is not uniform, we can use the distribution of the cumulants to identify the pricing kernel. This is the intuition behind our pricing kernel estimator. In simulations we find that our method substantially outperforms the classical method in recovering the SDF that generates the data.

To estimate the SDF, we use monthly S&P 500 and FTSE 100 index option data to nonparametrically estimate risk-neutral densities in the standard fashion, following Figlewski (2008) with slight improvements. We then nonparametrically estimate the SDF using a spline estimator. The estimate can be time-varying and is robust to any stochastic process that might generate returns, including processes with stochastic volatility and jumps. In using the fact that the cumulants should be uniformly distributed to identify our model, we follow Bliss and Panigirtzoglou (2005), who use the same fact to estimate implied risk aversion coefficients for parametric SDFs. Importantly, unlike Bliss and Panigirtzoglou (2005), we do not constrain our estimate to be monotonic. We also examine the importance of time variation in the SDF both by using the level of the VIX index as an instrument in our GMM estimation and by estimating the kernel in high and low VIX states. We use a bootstrapping procedure to estimate confidence bounds for our nonparametric SDF. We refer to our method as the conditional density integration (CDI) method.

We estimate risk-neutral densities from option prices and physical densities from historical returns, and find that these two sets of densities have very different characteristics. Furthermore, when we (incorrectly) follow the classical procedure by dividing option-based risk-neutral densities by historical-return-based physical densities, we also find implied pricing kernels that are nonmonotonic. These nonmonotonic pricing kernels are very sensitive to how the physical densities are estimated, a fact that suggests they are not econometrically robust. More state-of-the-art methods are also very sensitive to modeling choices. For example, changing the length of sample period

used to estimate the physical density for returns sometimes dramatically changes the shape of the estimated SDF. Furthermore, different researchers find nonmonotonicities in very different parts of the SDF. In Ait-Sahalia and Lo (2000), Audrino and Meier (2012), and Rosenberg and Engle (2002), for example, positive slopes appear in the center of the distribution of returns, whereas Christoffersen, Heston, and Jacobs (2013) and Bakshi, Madan, and Panayotov (2010) find a U-shaped kernel. Given the complex procedures required to estimate the SDF, the sensitivity of estimates to modeling choices is not surprising. In fact, researchers in statistics have found that naive estimators of the ratio of two (separately estimated) densities generally perform poorly.[6] The fact that estimates of the pricing kernel are not very robust is consistent with the pricing kernel puzzle being driven by questionable econometric methods.

When we properly account for conditional information by using the CDI estimator, the resultant S&P 500 pricing kernel estimate is monotonically decreasing. We estimate the SDF with sufficient precision to reject the SDF estimates of Rosenberg and Engle (2002). The kernel estimate is consistent with a decreasing SDF even when we allow the kernel to vary with the VIX index. We conclude that the pricing kernel puzzle is likely to be due to flaws in existing estimators.[7]

1. Existing Methods and Consistency

In this section we describe the classical nonparametric approach to estimating the SDF and point out its shortcomings.

1.1 Classical method

The classical nonparametric method of estimating the SDF of Jackwerth (2000) and Ait-Sahalia and Lo (2000) relies on a well-known result from probability theory known as the Radon-Nikodym Theorem.[8] The theorem implies that if $\mathbb{F}^Q$ and $\mathbb{F}^P$ are measures induced by the risk-neutral and physical cumulative distribution functions, the SDF can be expressed as

$$m_t(x_{t+s}) = e^{-rs} \frac{d\mathbb{F}^Q(x_{t+s}|\mathcal{F}_t)}{d\mathbb{F}^P(x_{t+s}|\mathcal{F}_t)}, \tag{1}$$

[6] Sugiyama, Suzuki, Nakajima, Kashima, von Bünau, and Kawanabe (2008), Sugiyama, Takeuchi, Suzuki, Kanamori, Hachiya, and Okanohara (2010), and Izbicki, Lee, and Schafer (2014) propose estimators that do not rely on separate density estimation. Simulations in Table 1 of Sugiyama, Takeuchi, Suzuki, Kanamori, Hachiya, and Okanohara (2010) illustrate the poor performance of the naive ratio.

[7] In reaction to our paper, Cuesdeanu and Jackwerth (2016) have acknowledged the problem with existing methods and claim to find nonmonotonicities using a modification of our method in S&P 500 options data. However, their method requires high confidence in their estimation of the tails, where no data exist.

[8] See, for example, Billingsley (2012).

a change of measure between two *conditional* probability measures where each probability is conditional on the same information set, $\mathcal{F}_t$.[9] The classical method relies on the fact that for sufficiently well behaved distributions, the Radon-Nikodym derivative in question is simply the ratio of the risk-neutral density, $\frac{d\mathbb{F}^Q(x_{t+s}|\mathcal{F}_t)}{dx}$ to the physical density, $\frac{d\mathbb{F}^P(x_{t+s}|\mathcal{F}_t)}{dx}$,

$$m_t(x_{t+s}) = e^{-rs}\frac{d\mathbb{F}^Q(x_{t+s}|\mathcal{F}_t)}{dx} \Big/ \frac{d\mathbb{F}^P(x_{t+s}|\mathcal{F}_t)}{dx}. \tag{2}$$

This fact allows econometricians to estimate the SDF by estimating the risk-neutral and physical densities separately and then taking the ratio of the densities.

Researchers should take care to estimate the densities in a conditional, forward-looking manner. For estimation of the numerator, most researchers rely on the result of Breeden and Litzenberger (1978), that $\frac{d\mathbb{F}^Q}{dK} = e^{rT}\frac{\partial^2 C}{\partial K^2}$, where C represents the option price, K represents strike prices and $\frac{d\mathbb{F}^Q}{dK}$ represents the risk-neutral density over possible realizations of the underlying. Since we typically observe option prices with a number of strike prices K, we are able to estimate the derivative $\frac{d\mathbb{F}^Q}{dK}$ over a collection of points K. Various techniques for estimating or interpolating values of the density between observed strike prices have been proposed in the literature. This gives an estimate of the risk-neutral density that is forward-looking and conditional on all information investors use to set prices.

Unfortunately, there are no known methods for estimating $d\mathbb{F}^P(x_{t+s}|\mathcal{F}_t)$, the time t physical density over returns from t to $t+s$, in a forward-looking manner, or taking into account all the information investors base their investment decisions on *at time t*. In previous studies, $d\mathbb{F}^P(x_{t+s}|\mathcal{F}_t)$ has often been estimated by smoothing or averaging past realized returns. To make the estimates reflect a conditional rather than unconditional density, one must typically use a rolling window to estimate the physical density.[10] Of course, this is not really comparable to using forward-looking option prices to back out market expectations. In fact, given that from one period to the next, the

[9] A corollary to this theorem states that if probability measures $\mathbb{F}^P$ and $\mathbb{F}^Q$ are equivalent measures, then the Radon-Nikodym derivative of $\mathbb{F}^P$ with respect to $\mathbb{F}^Q$ is equal to the reciprocal of the Radon-Nikodym derivative of $\mathbb{F}^Q$ with respect to $\mathbb{F}^P$,

$$\frac{d\mathbb{F}^Q}{d\mathbb{F}^P} = \left(\frac{d\mathbb{F}^P}{d\mathbb{F}^Q}\right)^{-1}.$$

If both $\mathbb{F}^Q$ and $\mathbb{F}^P$ are equivalent to dx, then

$$\frac{d\mathbb{F}^Q}{d\mathbb{F}^P} = \frac{d\mathbb{F}^Q}{dx} \Big/ \frac{d\mathbb{F}^P}{dx}.$$

The corollary allows one to express the Radon-Nikodym derivative as a ratio of two derivatives.

[10] See Foster and Nelson (1996).

nonparametric estimate of the physical density may only change because of the inclusion of one new observation and the exclusion of one old one, the estimated physical density can often be considered almost unconditional.

In the data, most of the moments of the estimated risk-neutral densities change substantially from one month to the next. At times when the conditional risk-neutral density has a higher variance, skewness, kurtosis or other moment than the estimated physical density, the ratio of the two densities can easily display nonmonotonicity.

1.2 Consistency

To emphasize the importance of conditioning on all available information we show that the classical method produces inconsistent estimates under assumptions that are likely to hold. We start by assuming the following:

Assumption 1. The true SDF is proportional to the ratio of two densities, both of which are conditional on all available information available at t: $m_t(x_{t+s}) = e^{-rs}\frac{d\mathbb{F}^Q(x_{t+s}|\mathcal{F}_t)}{d\mathbb{F}^P(x_{t+s}|\mathcal{F}_t)}$.

Assumption 2. Econometricians only observe a subset of all information, $\mathcal{S}_t \subset \mathcal{F}_t$.

Assumption 3. Econometricians use estimators that are pointwise consistent for the risk-neutral density conditional on $\mathcal{F}_t$ and the physical density conditional on $\mathcal{S}_t$, denoted $d\mathbb{F}^{Q*}(x_{t+s}|\mathcal{F}_t)$ and $d\mathbb{F}^{P*}(x_{t+s}|\mathcal{S}_t)$ respectively.

Under these three assumptions we can prove the following proposition:

Proposition 1. Given Assumptions 1–3, the classical estimate $m_t^*(x_{t+s}) = e^{-rs}\frac{d\mathbb{F}^{Q*}(x_{t+s}|\mathcal{F}_t)}{d\mathbb{F}^{P*}(x_{t+s}|\mathcal{S}_t)}$ is pointwise consistent for the true SDF, $m_t(x_{t+s})$, if and only if $d\mathbb{F}^P(x_{t+s}|\mathcal{F}_t) = d\mathbb{F}^P(x_{t+s}|\mathcal{S}_t)$.

The proof of this proposition is very simple. Since we have assumed that we have two pointwise consistent estimators, the classical estimate converges pointwise to

$$m_t^*(x_{t+s}) \xrightarrow[\text{lim}]{} e^{-rs}\frac{d\mathbb{F}^Q(x_{t+s}|\mathcal{F}_t)}{d\mathbb{F}^P(x_{t+s}|\mathcal{S}_t)} = m_t(x_{t+s})\left[\frac{d\mathbb{F}^P(x_{t+s}|\mathcal{F}_t)}{d\mathbb{F}^P(x_{t+s}|\mathcal{S}_t)}\right]. \tag{3}$$

Thus, unless $d\mathbb{F}^P(x_{t+s}|\mathcal{F}_t) = d\mathbb{F}^P(x_{t+s}|\mathcal{S}_t)$ so that the ratio of these two densities is equal to one, the classical method is an inconsistent estimator. More generally, if the SDF is a function of $d\mathbb{F}^P(x_{t+s}|\mathcal{S}_t)$ and $d\mathbb{F}^Q(x_{t+s}|\mathcal{F}_t)$, then the inconsistency exists. Therefore, the inconsistency also applies to studies that do not explicitly estimate the ratio (e.g., Rosenberg and Engle 2002). The two

conditional densities will be equivalent when the econometrician's information set captures all relevant information for the density of future prices. Thus, if all the additional information that market participants have in addition to what the econometrician uses is irrelevant then the classical estimator will be consistent. Given that many of the nonmonotonicity papers use only past index prices to estimate the physical density, this assumption seems unlikely to hold in practice.

Some researchers try to deal with the conditioning problem by incorporating models of time-varying volatility. The inconsistency suggested above is distinct from the effect implicitly suggested by Chabi-Yo, Garcia, and Renault (2007) and Christoffersen, Heston, and Jacobs (2013) (among others) due to a few omitted factors. Unless changes in volatility capture changes in all information available, these methods will also produce inconsistent results. The estimator we propose is meant to address this conditioning information problem without adding particular conditioning variables to a model.

It is useful to think about the expected magnitude of the bias term, $\left[\frac{d\mathbb{P}^P(x_{t+s}|\mathcal{F}_t)}{d\mathbb{P}^P(x_{t+s}|\mathcal{S}_t)}\right]$. Because this term is a ratio of two probability densities, the denominator of the ratio often will be close to zero. In places in which it is very close to zero, the bias term will be arbitrarily large. It is difficult, therefore, to think of a case in which we can be confident that the econometrician's information set is sufficiently close to the full information set that we can ignore the bias term.

Several recent papers examine the shape of the estimated SDF at a relatively high frequency, often monthly. Understanding how attitudes about risk vary from month to month is extremely valuable to researchers interested in asset pricing. However, Proposition 1 implies that unless researchers have a good way to estimate conditional physical densities each month, their estimates will conflate time-varying conditional information with time-varying risk premia. Without separating the effects of these two sources of variation, it is difficult to know what we can learn from high-frequency SDF estimates.

2. Estimating the SDF

The new CDI method we use to derive an estimate of the stochastic discount factor that properly accounts for conditional information is an important contribution of our paper, so we describe it in detail in this section. Our CDI method allows an econometrician to better account for the information set available to investors at the time investment decisions are made. We carefully explain how this is achieved. In Section 4, we apply the estimation procedures described here and show that the proposed econometric method has the potential to solve the risk aversion puzzle.

2.1 Estimating risk-neutral densities

To estimate the stochastic discount factor over the horizon spanned by the OptionMetrics data, we first estimate monthly risk-neutral densities following

the method outlined in Figlewski (2008), with a few modifications that we describe in Appendix A.[11] We use risk-neutral densities estimated with this method to calculate the SDF using both the classical nonparametric method and our new CDI method. Using the same set of risk-neutral densities, the classical nonparametric method yields nonmonotonic SDF estimates but the CDI method produces monotonic estimates. Thus, our method of estimating risk-neutral densities does not seem to drive the monotonicity result that we find.

2.2 CDI approach

Our identification strategy relies on several well-known properties from statistics and probability theory. The first of these properties, which is central to our method, eliminates the need to estimate the physical densities corresponding to each of the risk-neutral densities. The property is as follows:
For any continuous random variable X with CDF $\mathbb{F}$, the random variable defined by $\mathbb{F}(X)$ is uniformly distributed on the interval $[0,1]$,

$$\mathbb{F}(X) \sim U[0,1]. \tag{4}$$

We let $\mathbb{F}^P(x_{t+s}|\mathcal{F}_t)$ be the unobserved probability measure representing investors' aggregate beliefs about returns on the S&P 500 under the physical measure between time t and $t+s$ and let returns over the subsequent period be given by x_{t+s}. Now it follows from Equation 4, that

$$\int_{-\infty}^{x_{t+s}} d\mathbb{F}^P(x_{t+s}|\mathcal{F}_t) \sim U[0,1]. \tag{5}$$

Since there are no known methods for estimating $d\mathbb{F}^P(x_{t+s}|\mathcal{F}_t)$ in a forward-looking manner, estimating Equation (5) directly from the data is not a simple task. It would presumably require obtaining a long time series of past realizations of ex-dividend returns.[12] One would then have to find a way to use these returns to estimate forward-looking beliefs at time t about returns under the physical measure. As discussed earlier, this method would require something beyond simply smoothing a long time series of *past* returns, because that does not do a good job of estimating the *current* beliefs held by the market. To circumvent this problem, we make use of the fact that we do have forward-looking estimates of market beliefs about future returns under the risk-neutral measure.

We express Equation (5) in terms of the risk-neutral densities estimated using the method described in Appendix A. Let $d\mathbb{F}^Q(x_{t+s}|\mathcal{F}_t)$ be the time t risk-neutral

[11] We thank Stephen Figlewski for suggesting these improvements.

[12] We use percentage changes in market value because option payoffs are based on the market value of the S&P 500 at expiration. This amounts to shifting the cum-dividend return density to the left, but a stable dividend yield does not affect the shape of the SDF.

probability measure and let $\frac{d\mathbb{F}^P(x_{t+s}|\mathcal{F}_t)}{d\mathbb{F}^Q(x_{t+s}|\mathcal{F}_t)}$ denote the Radon-Nikodym derivative of time t physical distribution with respect to time t risk-neutral distribution of returns between t and $t+s$. Then, using a corrolary to the Radon-Nikodym theorem (in Footnote 9, we can state:

$$\int_{-\infty}^{x_{t+s}} d\mathbb{F}^P(x_{t+s}|\mathcal{F}_t) = \int_{-\infty}^{x_{t+s}} \frac{d\mathbb{F}^P(x_{t+s}|\mathcal{F}_t)}{d\mathbb{F}^Q(x_{t+s}|\mathcal{F}_t)} d\mathbb{F}^Q(x_{t+s}|\mathcal{F}_t)$$

$$= \int_{-\infty}^{x_{t+s}} \left(\frac{d\mathbb{F}^Q(x_{t+s}|\mathcal{F}_t)}{d\mathbb{F}^P(x_{t+s}|\mathcal{F}_t)}\right)^{-1} d\mathbb{F}^Q(x_{t+s}|\mathcal{F}_t) \sim U[0,1]. \quad (6)$$

Because we can estimate the risk-neutral densities, and we observe realized returns over the periods corresponding to each density, it only remains to estimate the random function $\left(\frac{d\mathbb{F}^Q(x_{t+s}|\mathcal{F}_t)}{d\mathbb{F}^P(x_{t+s}|\mathcal{F}_t)}\right)^{-1}$, which is proportional to the inverse of the stochastic discount factor. Of course, estimating this function requires a sample of realizations of both x_{t+s} and $d\mathbb{F}^Q(x_{t+s}|\mathcal{F}_t)$. We have a sample of time-varying risk-neutral densities and for each of these densities we have a realization from the corresponding conditional physical density. Our method resolves the inconsistency of other methods because the draw, x_{t+s}, and the density, $d\mathbb{F}^Q(x_{t+s}|\mathcal{F}_t)$, are both conditional on the same information set, $\mathcal{F}_t$.

It is important that we establish uniqueness of the random variable we attempt to estimate. The following proposition ensures that there is such a unique random variable for each given t and s.

Proposition 2. For any equivalent measures $\mathbb{F}^Q$ and $\mathbb{F}^P$ on $\mathbb{R}$ with random variable $X_{t+s} \sim \mathbb{F}^P$, there exists a *unique* $(a.s.\ \mathbb{F}^Q)$ nonnegative function $g_{t,s}: \mathbb{R} \to \mathbb{R}_+$ such that

$$\int_{-\infty}^{X_{t+s}} g_{t,s}(y) d\mathbb{F}^Q(y|\mathcal{F}_t) \sim U[0,1]. \quad (7)$$

Appendix B provides a proof of this proposition.[13]

Proposition 2 establishes uniqueness of the function $g_{t,s}$ that transforms the integral with respect to measure $\mathbb{F}^Q$, to a specific distribution. This is similar to the statement of the Radon-Nikodym Theorem. The function $g_{t,s}$, mapping

[13] The function $g_{t,s}$ in Proposition 2 is similar to the Radon-Nikodym term in Equation (6), the main difference being that in Equation (7), the region of integration is itself random. So the Radon-Nikodym Theorem in its most commonly stated form is not directly applicable here. The functional form of $g_{t,s}$ defines a random variable in Proposition 2 because it is evaluated at possible values of the random outcome. We can think of inputs to the function $g_{t,s}$ as values the random variable x_{t+s} can take. The outcomes of the random variable depend on $\omega \in \Omega$ the probability space determining returns, $x_{t+s} = x_{t+s}(\omega)$. As such, the integral with respect to $d\mathbb{F}^Q$ can be interpreted as the integral with respect to the measure $\mathbb{F}^Q(\{\omega : x_{t+s}(\omega) \in dy\})$. In this way, $g_{t,s}(y) = \phi(\{\omega : x_{t+s}(\omega) \in dy\})$, where ϕ is a mapping from Ω to the nonnegative real line, $\phi : \Omega \to \mathbb{R}_+$. So $g_{t,s}(y)$ represents possible realizations of the random variable $g(X(\omega)) = \phi(\omega)$. We will let $g_{t,s}$ denote the inverse of the SDF up to a rate of time discount $e^{r_t \tau}$, where r_t is the risk-free rate at time t and τ is the time to expiration of the option. Our estimation procedure will focus on estimating $g_{t,s}$.

realizations of returns to nonnegative values is itself a random variable, much the same as the Radon-Nikodym derivative. The difference is that here we have a random domain $(-\infty, X_{t+s}]$. We thus estimate the functional form of $g_{t,s}$ that maps random outcome of percentage changes in the S&P 500 to the unique kernel that transforms the integral in Equation (7) to the uniform distribution. Note that Proposition 2 does *not* in any way require that the SDF be constant over time, and the method is valid no matter what stochastic process returns follow.

It is interesting to note that our method is similar to examining the average returns of butterfly spreads. We believe the CDI method is superior to this for several reasons. First, our method of estimating the left tails of the risk-neutral distribution is more palatable than a method that would assume left and right end points for the series of butterfly spread returns. Second, butterfly spread returns are highly nonnormal because of the large mass at zero payoffs, and the averages of these returns could be unstable. Third, we want to follow a method that is comparable to the existing literature so that a comparison of results is possible along multiple dimensions.

2.3 CDI approach estimation and inference

Our goal is to estimate the SDF in a way that reflects investors' beliefs as accurately as possible. For this reason, we do not impose any parametric restriction on the form of the stochastic discount factor. Instead, we use a cubic spline to obtain nonparametric estimates of the inverse SDF. Because any real valued function can be reproduced by a cubic spline of infinite order, this is a completely model-free estimation procedure. Because our risk-neutral densities are model-free, they reflect any stochastic process that market participants believe in. Thus, it is robust to stochastic volatility and jumps in the underlying asset price process. Appendix C provides the details of our estimation procedure.

2.4 Standard approach to estimating physical densities

Once we have the forward-looking risk-neutral densities, we can proceed with estimating the stochastic discount factor. For the classical method, which relies on Equation (2), we are left to estimate the physical densities corresponding to each of the risk-neutral densities. As described above, there is no known way to estimate the physical density in a forward-looking manner, and the solution proposed in the literature is to use a rolling window of data to nonparametrically estimate the physical densities. We use a Gaussian kernel density estimator with a rolling window. In the spirit of a conditional estimate, it is best to use as short a window as possible without compromising the integrity of the kernel estimator.[14]

[14] Recall that our estimation of physical densities is purely illustrative. More recent work uses variations on this methodology, but finds similar nonmonotonic pricing kernels.

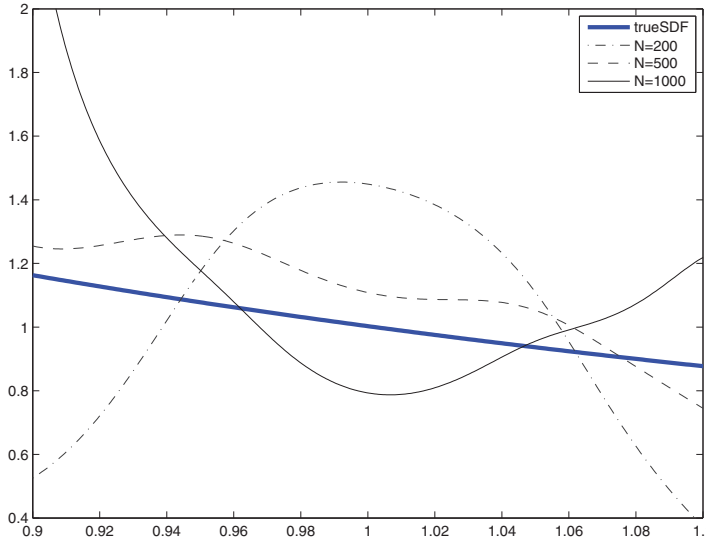
In theory, the physical and risk-neutral densities should have the same support. Empirically, using a rolling window of data to estimate the kernel density often results in estimates of the physical density with different (machine measurable) support from the risk-neutral density for the same period. This is itself a sign that there is a problem with the estimation procedure. This is a result of improperly matching conditional information in the numerator and denominator of the Radon-Nikodym derivative. If, for instance, previous returns within the rolling window tend to be low but recently the market received news suggesting high returns in the future, then the upper tail of the forward-looking risk-neutral density may have support beyond the range of positive support for the physical density estimate. Similarly, we observe instances where the physical density has wider support than the risk-neutral density. In practice, when this happens, we need to truncate the densities such that they have the same region of positive support, to avoid dividing a positive density by zero for some returns. To avoid this problem, for each date, we estimate the pricing kernel over the range between the maximum of the lower bounds of support for the densities and the minimum of the upper bound.

Using a nonparametric density estimator with data from a rolling window to estimate the physical density makes it possible to estimate an SDF every month. In fact, it has become popular in the empirical SDF literature to estimate a time series of SDFs and discuss their properties. Unfortunately, Proposition 1 implies that such high-frequency estimates are likely to be econometrically inconsistent and thus uninformative.

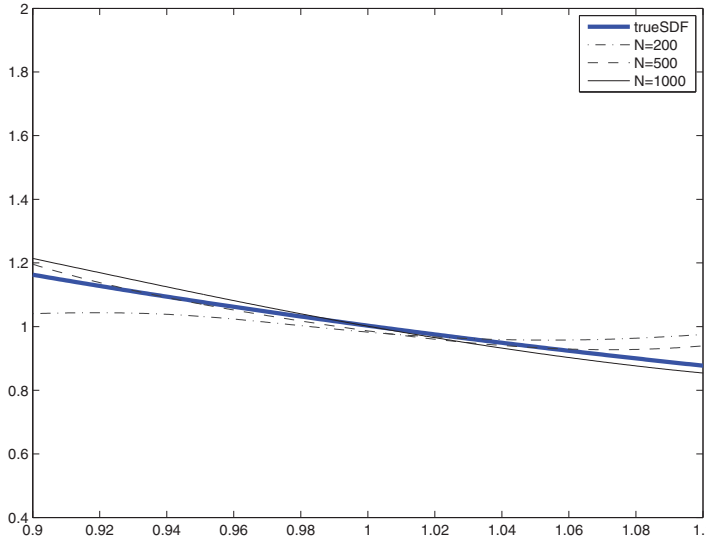
3. CDI and Classical Methods in Simulations

This section examines the efficacy of the CDI method in sample sizes typical of those in the empirical literature on pricing kernel estimation, and contrasts this with the efficacy of the classical estimator in the same sample. By simulating data with a known SDF, we can observe how accurately each is able to estimate the "true" SDF. Our simulated data draw underlying index returns from a log-normal distribution like in the Black-Scholes world. Of course, the CDI method is more general and does not assume any particular distribution. We choose parameters of the distribution to fit the data generated by our risk-neutral S&P 500 densities, being careful to allow the volatility of the densities to vary with time. With the simulated series $\left(d\mathbb{F}_t^Q, x_{t+s}\right)$, we estimate true SDF using both methods for $N = 200, 500$, and $1,000$. For the classical estimator we use a 60 period rolling window of realized returns x_{t+s} to compute kernel density estimates of the physical densities that are unknown to the econometrician. Figure 1 illustrates the results of the simulations for both estimators, and Appendix D provides the details of our simulation procedure.

Panel A shows the results of the simulations performed for the classical method. It is clear from the figure that none of the estimates are able to recover the true SDF with any accuracy. The estimates resulting from $N = 200$ and



(A) Classical method estimates



(B) CDI estimates

Figure 1
Estimated and true SDFs from simulations
The plots compare the performance of the CDI method and classical method of nonparametric estimates of the SDF. Our simulated data are generated using the true SDF depicted by the bold line in each panel. The estimates of each method are depicted with the true SDF. The level of the SDF is on the *y*-axis, and the moneyness is on the *x*-axis.

$N = 1,000$ simulated months exhibit extreme nonmonotonicity and do not come close to recovering the true SDF. The estimate when $N = 500$ does far better than the other two estimates using the classical method. However, if we compare the classical method with $N = 500$ to the poorest performing CDI estimator, that with $N = 200$, it is clear that the the poorest performing CDI estimate significantly outperforms the best estimate using the classical method.

Panel B shows that for all values of N, the CDI estimator does a very good job of recovering the true SDF. The smallest data sample recovers the true SDF fairly well over the range $[0.95, 1.05]$, but outside of the range $[0.95, 1.05]$, the CDI estimator veers away from the true SDF when $N = 200$. This is hardly surprising given that there are relatively few realized observations outside this range. For $N = 500$ and $N = 1,000$, the CDI estimator does a very good job of recovering the true SDF over the entire range depicted, $[0.9, 1.1]$. This is made possible by the fact that larger samples have a larger number of observations near both 0.9 and 1.1, allowing the spline to accurately estimate the SDF near those values of returns. Figure 1 shows that the CDI method performs very well, but the classical method performs poorly.[15] The reason is that the CDI method properly accounts for conditional information whereas the classical method uses the ratio of a forward-looking estimate to a backward-looking estimate, thus failing to take account of conditional information.

4. CDI and Classical Methods in Market Data

4.1 Data

We use daily S&P 500 and FTSE 100 index option data from OptionMetrics. For the S&P 500 index options, price midpoints are available from January, 1996 through December, 2014, for a total of 228 months. For the FTSE data, closing prices are available from January 2002 through December 2014. Prior to 2006, FTSE data was collected from the exchange directly. After 2006, Optionmetrics began receiving tick data with more limited availability until 2007. As a consequence, several months are unavailable in 2006 and 2007 and we are left with 138 total months of data. We use options with one month to maturity, giving a nonoverlapping time-series of options prices. These nonoverlapping data allow us to obtain independent observations for beliefs about the coming month and an independent realization of returns. Using monthly rather than higher frequency data does not cause a loss of information because we only have one option expiration per month.

We also use OptionMetrics implied volatilities for each strike price at each date in our set. We remove data for which there is no available implied volatility as these violate no arbitrage conditions. We use put prices for relatively low strike prices, call prices for relatively high strike prices and weighted

[15] Audrino and Meier (2012) and Beare and Schmidt (2016) improve on the classical approach, and their methods may produce better results.

averages for intermediate strike prices. The relative weights of puts and calls are determined using a logistic function that is centered at the closing index value with a volatility parameter that is half of the range of observable option prices. Using open interest to calculate the weighted average gives almost identical results, but the logistic function is slightly smoother.

We obtain S&P 500 closing prices for monthly trading dates and for option expiration dates from CRSP, and closing FTSE 100 values from OptionMetrics Europe. To estimate the SDF with the classical procedure, we also use prices from up to ten years prior to the start of our OptionMetrics sample for our rolling window estimations of the physical density. Finally, we calculate the risk-free rate from continuously compounded yields on secondary market 3-month Treasury Bills. This data is from the Federal Reserve report H.15.

We use the level of the VIX index as an instrumental variable in some of our GMM estimations. The level of the VIX index is available for download on the Chicago Board Options Exchange Web site.

4.2 Results

In this section, we present the results of our estimation described in Section 2, using the data described in Section 4.1. We compare CDI results with the results obtained by using the classical nonparametric method over the same sample period. We argue that our estimation procedure results in economically plausible SDFs, unlike the classical method, which does not properly account for conditional information and suggests the existence of a pricing kernel puzzle. Throughout this section, it is important to recall that the risk-neutral densities used for estimation of the SDF with the classical method are the same densities used for the CDI method. This allows us to compare the methods directly.

Table 1 presents sample averages of the mean, variance, skewness and kurtosis associated with both the risk-neutral and physical densities estimated for each of the 228 months from January, 1996 through December 2014 for the S&P 500 and the 138 available months from January 2002 to December 2014 for the FTSE 100. The physical densities described in Table 1 are estimated with a kernel density method using the prior 48 months of index returns. Looking first at the means of both the risk-neutral and physical densities, we see that the average means are about the same, but the physical density means are much more variable than the risk-neutral density means.

Theory dictates that the expected value of the risk-neutral density should equal the risk-free rate, r_t, for all t. The average of the annualized expected return associated with the estimated risk-neutral S&P 500 densities is 3.33% with a sample standard deviation of 0.96%. This is close to the value we obtain when we plug in the mean value for r_t over our sample period, $\bar{r} = 2.42\%$. Of course, this is not exactly the correct comparison to make, as one would want to compare $e^{r_t \tau}$ with the expected value of each risk-neutral density in our sample. We calculate the absolute value of this difference for each month in our sample.

Table 1
Summary statistics for risk-neutral densities

A. S&P 500

	Risk-neutral densities from options prices			
	Annualized mean ret	Annualized SD	Monthly skewness	Monthly kurtosis
Sample average	3.33%	21.58%	−1.27	9.12
Sample SD	0.96%	9.22%	1.01	5.98
	Physical densities from 48 months of historical data			
	Annualized mean ret	Annualized SD	Monthly skewness	Monthly kurtosis
Sample average	7.70%	17.33%	−0.455	4.039
Sample SD	7.01%	4.94%	0.448	1.298

B: FTSE 100

	Risk-neutral densities from options prices			
	Annualized mean ret	Annualized SD	Monthly skewness	Monthly kurtosis
Sample average	3.55%	19.93%	−1.016	8.607
Sample SD	3.74%	8.71%	0.707	5.798
	Physical densities from 48 months of historical data			
	Annualized mean ret	Annualized SD	Monthly skewness	Monthly kurtosis
Sample average	3.87%	17.76%	−0.5863	3.7341
Sample SD	5.31%	3.62%	0.2567	0.6133

For each of the months in our sample (228 months: from January 1996 through December 2014 for S&P 500 data. 138 months: from January 2002 to December 2014 for FTSE 100 data), we estimate both a risk-neutral density based on option prices and a physical density based on historical data. The physical densities are estimated with a Gaussian kernel density estimator using 48 months of past returns, and the risk-neutral densities are estimated as described in Section 2.1. This table reports summary statistics on the moments of these densities. The table reports both sample averages and sample standard deviations of the first four centralized moments in terms of returns: mean, standard deviation, skewness, and kurtosis. The average means and standard deviations are annualized to ease interpretation.

The mean absolute monthly difference is 0.25% with a standard deviation of 0.58%. This suggests that our estimation procedure does very well in matching the risk-free rate. This confirms that our estimates are reasonable given that our estimation does not constrain the mean of the distributions in any way. It is interesting to note that even during the crisis, the risk-neutral densities have means that are close to the risk-free rate. The risk-neutral annualized mean returns for the S&P 500 index on September 18th and October 23rd of 2008 are estimated to be −2.81% and 7.14%, respectively. The estimated risk-neutral annualized mean returns on September 17th and October 22nd of 2008 for the FTSE 100 are −5.7% and 13%. It may be that the risk-neutral means are generally close to the risk-free rate because most option traders use some variant of the Black-Scholes model, which sets the risk-neutral mean equal to the risk-free rate.

Considering next the annualized standard deviations of risk-neutral and physical densities, the risk-neutral densities have higher average standard

deviations than the physical densities for both indices. Their standard deviations are also much more variable than those of the physical densities. This difference is presumably driven by the conditional nature of the risk-neutral densities. When investors believe the market will be volatile in the future, this belief is immediately reflected by the risk-neutral density. However, the kernel density estimator used in the classical procedure smoothes out any extreme returns and has no other way to incorporate investors' beliefs. For the S&P 500, the estimated risk-neutral annualized standard deviations for September 18th and October 23rd of 2008 are 61% and 77%, respectively. The corresponding values for the physical density are 13.0% and 16.4%. The FTSE 100 risk-neutral densities on September 17th and October 22nd of 2008 have annualized standard deviations of 38% and 56%, also much higher than the estimates under our rolling window physical density estimates that have annualized standard deviations of 18% for both days. The physical densities certainly respond to the extreme returns during the financial crisis, but their response is much smaller than the response of the risk-neutral densities.

The monthly skewness and kurtosis values are quite different for risk-neutral densities than they are for physical densities. The results on these higher moments combined with those for the means and standard deviations suggest that using a smoothing method to estimate the conditional physical densities is misguided. As discussed earlier, the implicit assumption made in order to use rolling window estimates for the physical densities is that the physical densities are stable over time. In our data, neither the physical nor the risk-neutral densities appear stable over time. Furthermore, if the pricing kernel is time invariant, then the physical and the risk-neutral densities should be related to each other. In fact, in a Black-Scholes world, the variance, skewness and kurtosis of the risk-neutral density are equal to those of the physical density. However, in our data the moments of the risk-neutral densities are not very close to those of the physical densities. Even using models that forecast variances (e.g., Rosenberg and Engle (2002)) may fail to miss variation in skewness or kurtosis. This highlights a major advantage of the CDI method over existing methods.

4.2.1 Classical method results. We first present the results of estimating the average of a series of estimated SDFs using the classical nonparametric method similar to those of Jackwerth (2000) and Ait-Sahalia and Lo (2000). We should point out that although our classical method estimates are similar to those of other papers, they are not exactly the same as any particular paper. We simply use them to illustrate how sensitive ratios are to various models for estimating the physical densities. We use monthly data over a longer time span than earlier papers, and other papers often have slightly different methods. Nevertheless, our classical method results are very similar to those of other papers. For both the FTSE 100 and the S&P 500 data, we use the same risk-neutral densities that

are used in the CDI method. These risk-neutral densities are estimated using the procedure described in Section 2.1, and an example of a risk-neutral density estimate appears in Figure 2. We then estimate the corresponding physical densities using a Gaussian kernel density estimator based on a rolling window of past returns. We use a bandwidth of $h = n^{-\frac{1}{5}} \times \sigma_{data}$, where σ_{data} denotes the standard deviation of all the data used in the kernel estimation for all time periods. The results vary little with different choices of h. When using the kernel density estimator, there is a trade-off between the number of data points available and the temporal proximity of the data points. A larger number of data points improves the mechanical estimation of the kernel density estimator, but does not solve the real problem, which is the use of backward-looking data to estimate conditional beliefs. By taking realized returns further back, we are using older, possibly irrelevant data as far as investors' time t decision making is concerned.

Figure 3 presents estimation results using the classical nonparametric method for the S&P 500. The panels of Figures 3 use different window lengths when calculating the physical densities of returns. In all panels, the same general pattern appears but significant variations arise across different window lengths. The SDF is sharply decreasing over states with low returns before displaying nonmonotonicity and sometimes gradual increasing as returns increase. The four panels look similar over lower returns but there is significant variation across the panels as returns increase. We are not able to estimate the mean SDF with any precision for gross index returns outside of the range of 0.9 to 1.1. Even though index realizations of 0.9 (-10% change) are rare, they do exist and we would like to be able to identify the form of the pricing kernel at such low return values. As we look toward larger returns, in the S&P 500 panels we see a portion of the estimated SDF that is increasing in returns between 0.95 and 1.0. We also see at least one bump that appears for short rolling windows but not for long windows. Estimates using FTSE 100 data display similar properties. An estimator that changes our inference about monotonicity as we alter the window length for estimating physical densities does not seem very robust.

The figure includes pointwise 95% bootstrapped confidence intervals. Because we use a rolling window of historical data to estimate the physical densities, we are able to obtain tighter confidence intervals than we will using the CDI method, which does not use a window of previous returns. Accordingly, the intervals become tighter as we increase the length of the rolling window for both the FTSE 100 and the S&P 500 estimates. The confidence intervals are in fact tight enough so that in every panel in Figure 3 we report statistically significant nonmonotonicity. We define a nonmonotonicity to be statistically significant in the estimated SDF if at any point on the returns (horizontal) axis, the lower confidence bound exceeds the upper bound of any confidence interval at a lower level of returns. For example, in three of the panels of Figure 3, the lower confidence bound at 1.02 on the returns axis exceeds the

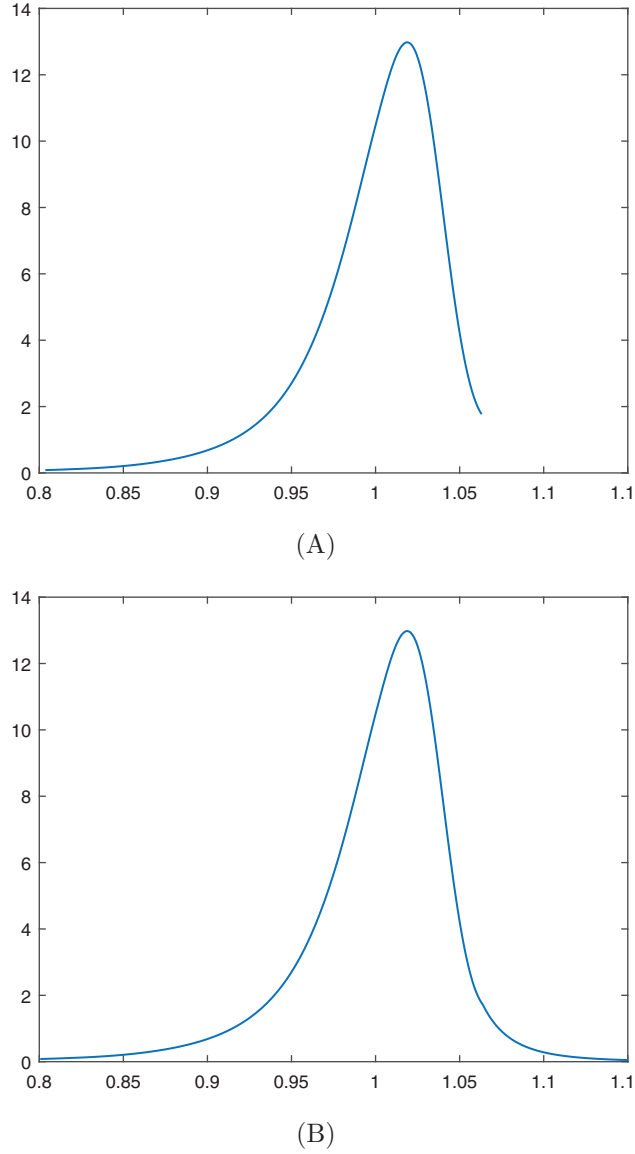


Figure 2
Example of a risk-neutral density
This example of a risk-neutral density was estimated using option prices from August 22, 2013, with best bids exceeding \$3/8. For August 22, 2013, 88 valid option prices correspond to 82 unique strike prices. Panel A depicts the density we directly estimate from option prices, and panel B depicts the complete density with tails. In panel A, The density appears more truncated on the right than on the left because there are fewer options with a price above \$3/8 in this range. Each month we use option prices to estimate a risk-neutral density like this one. In panel B, we estimate the tails of the distribution by matching a generalized Pareto distribution to the slope of the density close to where we can no longer estimate it. Section 2.1 describes the method for estimating the risk-neutral densities.

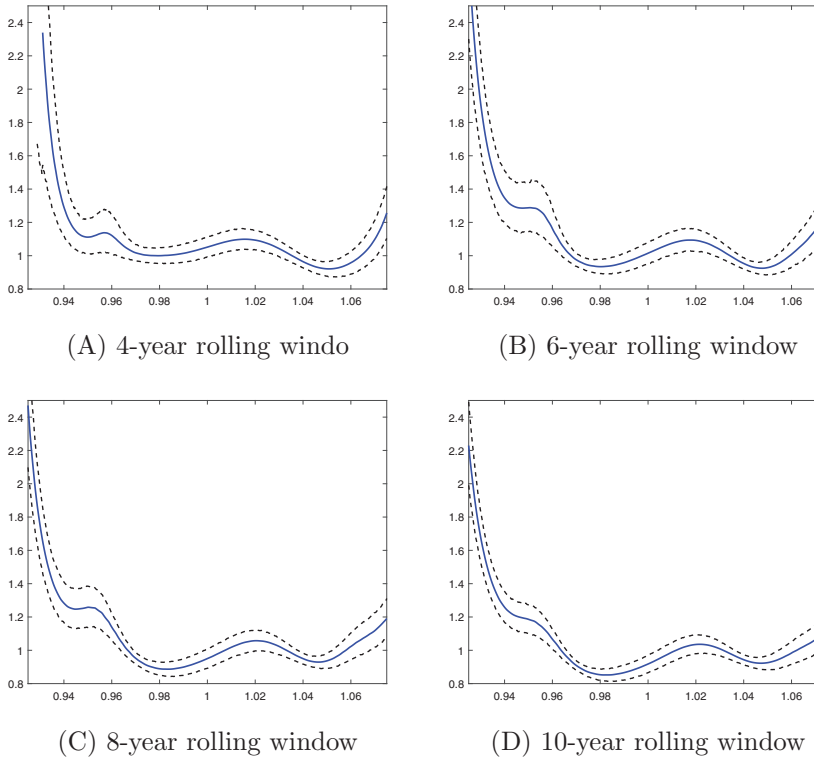
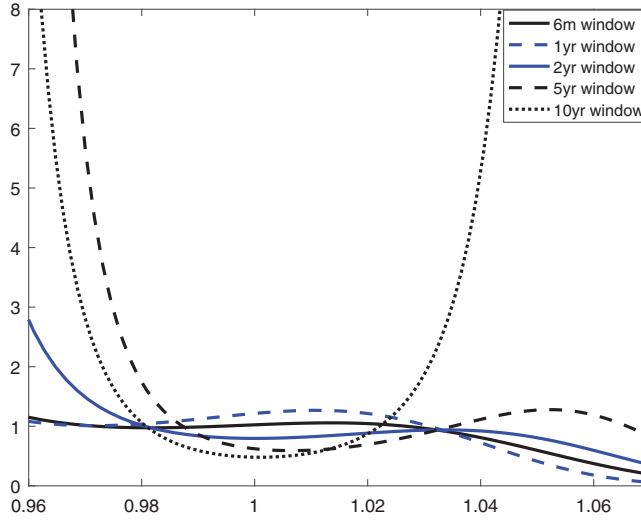


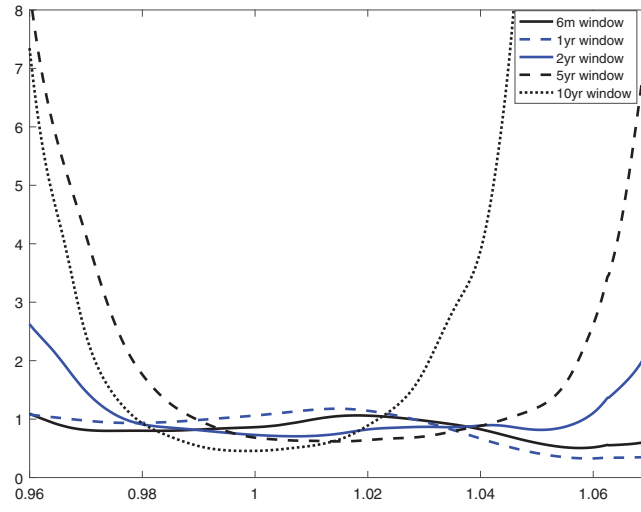
Figure 3
Estimated SDFs using classical procedure: S&P 500
The plots display classical nonparametric estimates of the stochastic discount factor as the average of monthly SDF estimates with pointwise bootstrap 95% confidence intervals. Each monthly SDF is the ratio of a risk-neutral density to a physical density estimate of returns on the S&P 500 index. Each panel represents the resultant estimate when a different widow is used to estimate the physical density using a Gaussian kernel estimator.

upper confidence bound at 0.98. Therefore the estimates exhibit a statistically significant nonmonotonicity. As one would expect, using a longer window of returns allows us to identify nonmonotonicity at higher confidence levels. In panel B, the nonmonotonicity is just significant at the 95% level. However, as we increase the length of the rolling windows used in our estimates, the confidence intervals become tighter and the nonmonotonicities are more pronounced and thus are significant at even higher levels of confidence. In fact, it looks clear that you could reject much of the SDF estimator using a four year rolling window with the estimate using a ten-year rolling window. This illustrates that increasing power for an inconsistent estimator is not very beneficial.

In Figure 4 we report estimates using the methods of Rosenberg and Engle (2002) and Barone-Adesi, Engle, and Mancini (2008), who use model estimates



(A) Rosenberg and Engle



(B) Barone Adesi et al.

Figure 4
Estimated SDFs using more sophisticated procedures
The plots display estimated SDFs using the method of Rosenberg and Engle (2002) (panel A) and Barone-Adesi, Engle, and Mancini (2008) (panel B). The physical density is estimated using GJR estimates as described in Rosenberg and Engle (2002), with parameter estimates based on daily returns data with rolling windows of varying lengths: 6 months, 1 year, 2 years, 5 years, and 10 years. For the estimates in panel B, we follow Barone-Adesi, Engle, and Mancini (2008), who estimate the risk-neutral density using S&P 500 options on March 20, 2014, with 30 days to maturity. We follow our risk-neutral density estimation procedure as described in the appendix.

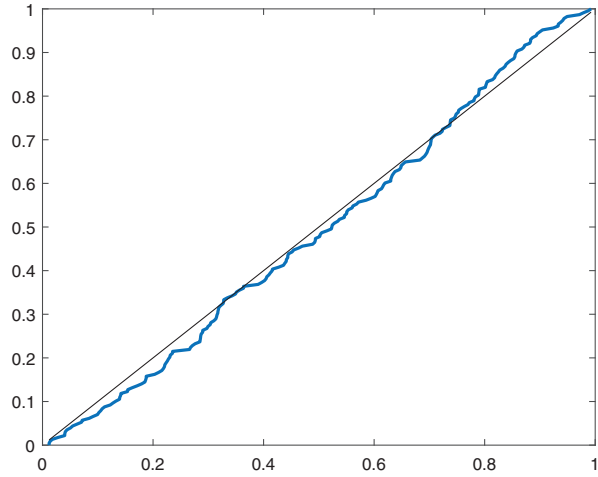
that are much more sophisticated than our classical method estimates. However, like the classical method, they do not account for conditional information. The physical density is estimated using Glosten, Jagannathan, and Runkle (1993) (GJR) as described in Rosenberg and Engle (2002), with parameter estimates based on daily returns data with rolling windows of varying lengths. Like in the classical method case, the estimates appear to depend critically on the assumed data window length.

4.2.2 CDI results. The CDI method is fairly general, allowing for both conditional and time unconditional SDF estimation. We first estimate an unconditional SDF, which is the focus of much of the nonmonotonic SDF research. In some ways, a nonmonotonic unconditional SDF is more surprising because it is not subject to fluctuations in estimation errors or investor preferences. Then we check how important time variation is likely to be in two different ways, reporting the results in the next subsection.

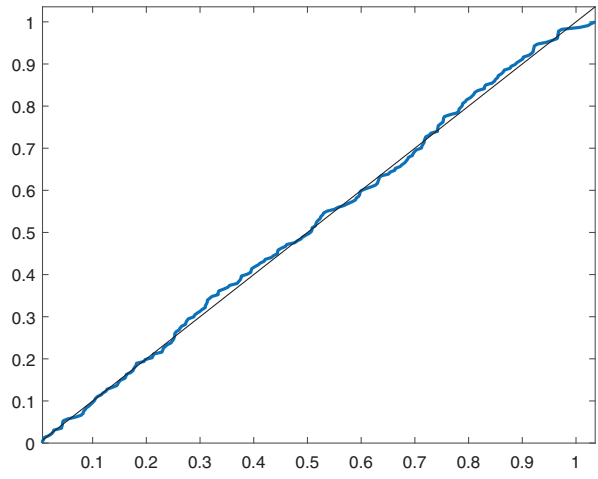
Implementing the CDI method requires that we choose how many basis functions, b and moment conditions, m, we will use. For our models to be identified, we will need to have at least as many moments as basis functions, so that $m \geq b$. In unreported preliminary estimates with a subset of the sample, we find that models with $b = m$, which are exactly identified, fit the data better than over identified alternatives. To be sure that our choices of m and b do not affect our estimates in an important way, we report SDF estimates for exactly identified models with values of $m = b$ that vary from 4 to 9. We set $m = b$ so that we maximize the flexibility of the estimator for each given number of moment conditions, m.

We graphically display how the unconditional CDI models fit the data by plotting empirical CDF functions. Our results for the S&P 500 appear in Figure 5 and those for the FTSE 100 appear in Figure 6. The line with a 45-degree slope is the uniform CDF, which will fit the data exactly if our models are perfect. Panel A in both figures shows what the data look like with a constant kernel specification. Specifically, these plots contain the empirical CDF of risk-neutral densities integrated up to realized market return values. They show clear departures from the uniform density, which is to be expected because they allow for no risk premiums. Both plots show some tendency for the untransformed cumulants to occur more frequently than the uniform for values above the median.

In the case of the S&P 500 in Figure 5, notice that below the median the untransformed cumulants occur less frequently than we would expect in the case of a uniform distribution. This combined with the fact that the untransformed cumulants occur *more* frequently above the median suggests that our transforming function g should be increasing. Because g estimates the inverse of the pricing kernel, panel A of Figure 5 suggests a decreasing pricing kernel. Panel B in both figures depicts the empirical CDF of cumulants

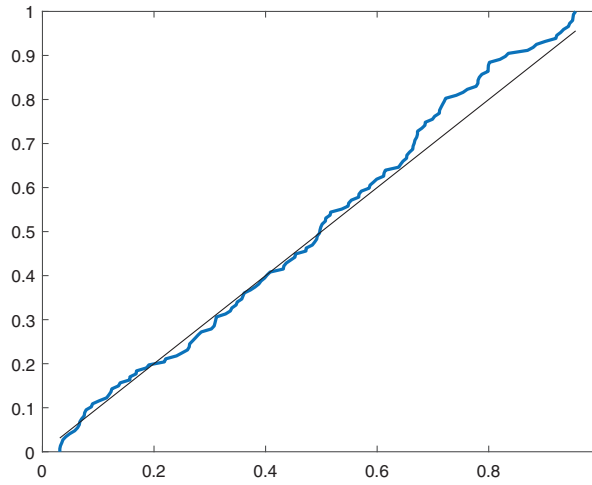


(A) No pricing kernel S&P 500

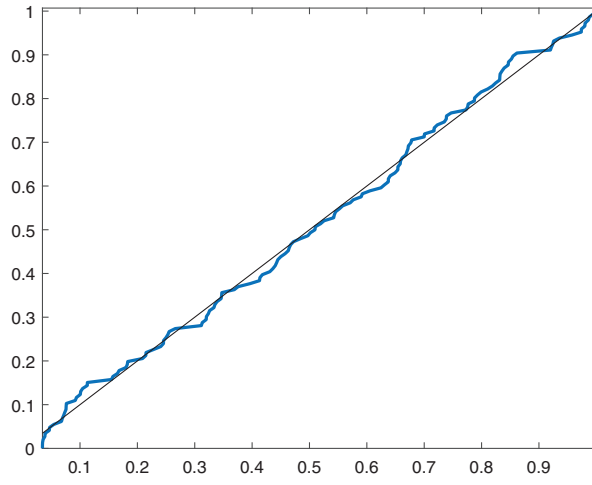


(B) Pricing kernel S&P 500: Seven bases

Figure 5
Empirical distribution functions of S&P 500 cumulants
This figure plots the empirical distribution functions (EDFs) of the cumulants that result from integrating the estimated inverse of the SDF with the risk-neutral densities up to their corresponding realized values of the S&P 500. Panel A shows the EDF of the cumulants without a transform from risk-neutral to physical densities. In this case $\widehat{g_{t,s}}(y) \equiv 1$, so that the resultant cumulants are given by $\int_{-\infty}^{x_{t+s}} d\mathbb{F}_t^Q(y), \;\; t = 1, 2, ..., T$. If the true pricing kernel is constant (or there is no compensation for risk) then we would expect the density to be close to a uniform [0,1] density, or for the EDF to follow the 45-degree line. Panel B shows the EDF of cumulants resulting from the CDI estimation method with 7 spline basis functions used in estimating the inverse of the SDF. Specifically, they show the EDF of $\int_{-\infty}^{x_{t+s}} \widehat{g_{t,s}}(y) d\mathbb{F}_t^Q(y), \;\; t = 1, 2, ..., T$, where $\widehat{g_{t,s}}(y)$ is the CDI estimate of the inverse of the pricing kernel. Other choices of the number of bases produce similar results. The EDF in panel B closely follows the 45-degree line, meaning that the cumulants are approximately uniformly [0,1] distributed.



(A) No pricing kernel FTSE 100



(B) Pricing kernel FTSE 100: Seven bases

Figure 6
Empirical distribution functions of FTSE 100 cumulants
This figure plots the empirical distribution functions (EDFs) of the cumulants that result from integrating the estimated inverse of the SDF with the risk-neutral densities up to their corresponding realized values of the FTSE 100. Panel A shows the EDF of the cumulants without a transform from risk-neutral to physical densities. In this case $\widehat{g_{t,s}}(y) \equiv 1$, so that the resultant cumulants are given by $\int_{-\infty}^{x_{t+s}} d\mathbb{F}_t^Q(y), \;\; t = 1, 2, ..., T$. If the true pricing kernel is constant (or there is no compensation for risk), then we would expect the density to be close to a uniform [0,1] density, or for the EDF to follow the 45-degree line. Panel B shows the EDF of cumulants resulting from the CDI estimation method with seven spline basis functions used in estimating the inverse of the SDF. Specifically, they show the EDF of $\int_{-\infty}^{x_{t+s}} \widehat{g_{t,s}}(y) d\mathbb{F}_t^Q(y), \;\; t = 1, 2, ..., T$, where $\widehat{g_{t,s}}(y)$ is the CDI estimate of the inverse of the pricing kernel. Other choices of the number of bases produce similar results. The EDF in panel B closely follows the 45-degree line, meaning that the cumulants are approximately uniformly [0,1] distributed.

transformed by a pricing kernel estimated with $b=m=7$. These empirical CDFs are much closer to uniform[16].

We focus on the functional form of the inverse of the function $\widehat{g_{t,s}}$ whose estimation is described in Section 2. Below, we plot the estimated functional form of $\widehat{M}(x)=\frac{1}{\widehat{g_{t,s}}(x)}$, which we will refer to as the SDF because $e^{-r_t\tau}$ is approximately equal to one for our entire sample. Furthermore, multiplying $M(x)$ by a constant will not change the qualitative aspects of the SDF we are attempting to capture.

It is easily seen from Figure 7 that the conditional SDF estimated with the CDI approach is a downward-sloping function of S&P 500 index realizations, regardless of the number of moment conditions used to estimate the SDF. Figure 8 shows that the estimated SDF for the FTSE data is generally downward sloping for different choices of $m=b$, but in some cases it exhibits nonmonotonicity over higher returns. There are relatively few observed returns larger than 1.05 in the FTSE data set, and the data set underlying the FTSE estimates is much smaller than that for the S&P. As a result, our nonparametric estimator is bound to be imprecise at larger values of index returns.

To investigate whether there are nonmonotonicities in the FTSE SDF, we need to determine whether the upward-sloping portion of the estimated SDF is statistically significant. We include bootstrap confidence intervals based on 20,000 resamples in Figures 7 and 8. In virtually all forms of nonparametric estimation, an extremely large set of data is required for one to achieve tight confidence intervals. Because options data do not go back very far, we do not have many extreme observed returns in the time series. As a result, confidence intervals for our estimates are not very tight at the extreme ends of the estimated SDFs. It can be seen in Figures 7 and 8 that the pointwise 95% confidence intervals for the SDF widen for index values that are far away from the current value. This is to be expected as we have only 228 months' worth of S&P data and 138 months for the FTSE data. Other researchers are subject to the same data constraints.

Table 2 provides the results of three goodness of fit tests, the Berkowitz, Cramér-von Mises, and Kolmogorov-Smirnov tests. The null hypothesis of each goodness of fit test is that the estimated distribution is the hypothesized distribution of $U[0,1]$. Table 2 also contains test statistics and corresponding p-values for our data in the cases of no transformation. Again, the case of no transformation means that we take $g(y)\equiv 1$ in Equation (7). So the nontransformed data we use to calculate the test statistics is given by the vector V with

$$V_i=\int_{-\infty}^{X_i} d\mathbb{F}_i^Q(y), i=1,\ldots T.$$

[16] Note that we do not constrain our method of moments estimator to produce estimates that fall within the interval $[0,1]$. As a result, some of the transformed cumulants fall slightly outside the unit interval. The estimated SDF functions are not sensitive to constraining the estimation.

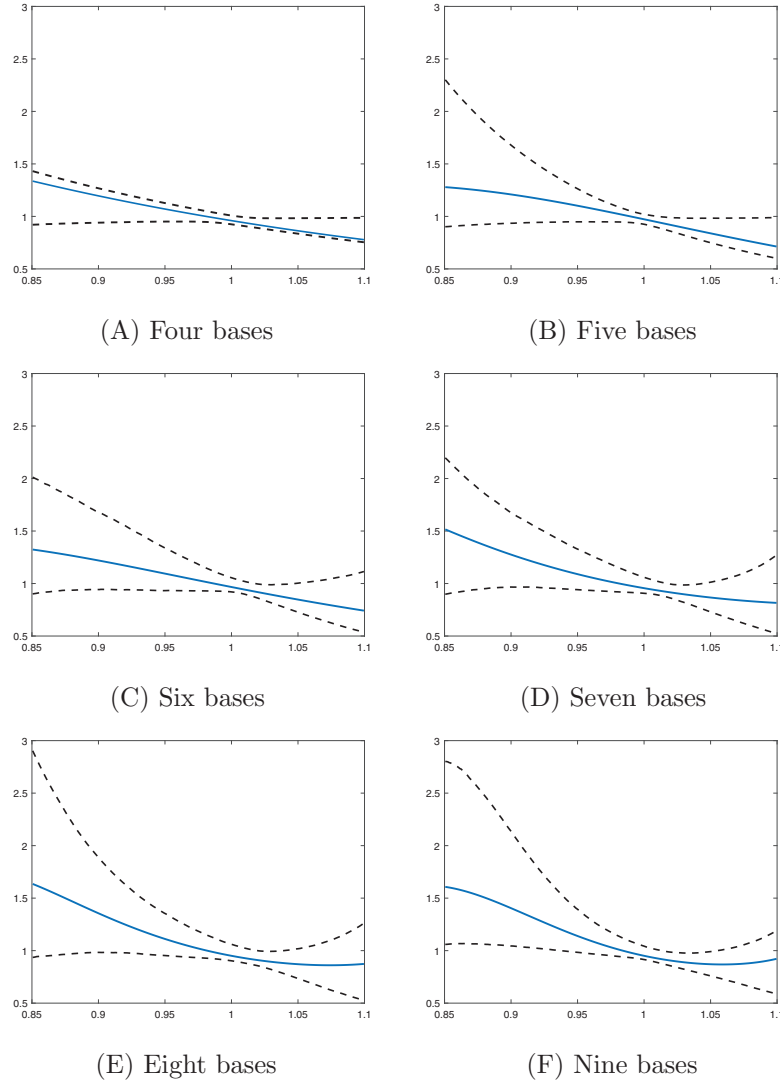


Figure 7
Unconditional stochastic discount factor using CDI method: S&P 500
The figure shows our CDI estimation of the pricing kernel for the S&P 500 with bootstrapped confidence intervals. Each panel corresponds to an estimate with a different number of spline basis functions used in estimating the inverse of the SDF. Estimates are clearly monontonically decreasing on the interval over which we can estimate the SDF with some precision. The CDI method estimates the pricing kernel by matching the moments of the distribution of the cumulants,

$$\int_{-\infty}^{x_{t,s}} \widehat{g_{t,s}}(y) d\mathbb{F}_t^Q(y), \quad t = 1, 2, ..., T,$$

to the moments of the uniform distribution by nonparametrically estimating the function $g(\cdot)$. The SDF in this formulation is actually the inverse of $g(\cdot)$, so that is what we plot above. Ninety-five percent confidence intervals, which are plotted with dashed lines, are based on 20,000 bootstrap iterations of the CDI method, sampling our set of dates with replacement.

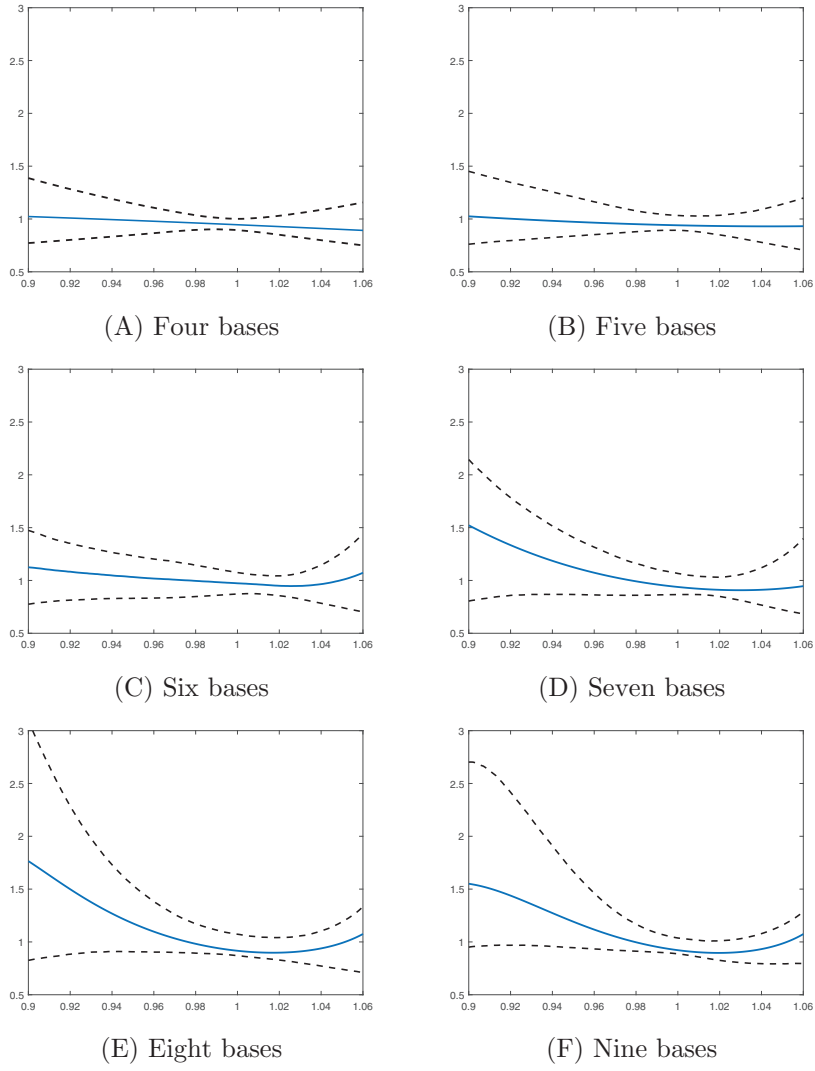


Figure 8
Unconditional stochastic discount factor using CDI method: FTSE 100
The result of our CDI estimation of the pricing kernel for the FTSE 100 is plotted above. The estimate exhibits some nonmonotonicity at the end of the interval over which we can estimate it with some precision. The nonmonotonicity is not statistically significant according to the 95% bootstrapped confidence intervals. The CDI method estimates the pricing kernel by matching the moments of the distribution of the cumulants,

$$\int_{-\infty}^{x_{t+s}} \widehat{g_{t,s}}(y) d\mathbb{F}_t^Q(y), \quad t = 1, 2, ..., T,$$

to the moments of the uniform distribution by nonparametrically estimating the function $g(\cdot)$. The SDF in this formulation is actually the inverse of $g(\cdot)$, so that is what we plot above. Ninety-five percent confidence intervals, which are plotted with dashed lines, are based on 20,000 bootstrap iterations of the CDI method, sampling our set of dates with replacement.

Table 2
Test statistics and *p*-values

A. S&P 500

Model	Berkowitz	*p*-value	Cramér-von Mises	*p*-value	Kolmogorov-Smirnov	*p*-value value
No pricing kernel	9.399	0.0244	0.156	0.377	0.052	0.543
$b=m=4$	2.690	0.442	0.029	0.979	0.031	0.975
$b=m=5$	4.424	0.219	0.033	0.964	0.030	0.981
$b=m=6$	3.614	0.306	0.031	0.973	0.033	0.964
$b=m=7$	1.751	0.626	0.026	0.988	0.026	0.996
$b=m=8$	4.294	0.231	0.029	0.978	0.028	0.992
$b=m=9$	4.737	0.192	0.030	0.969	0.029	0.990
B. FTSE 100						
No pricing kernel	6.177	0.103	0.083	0.249	0.147	0.404
$b=m=4$	2.036	0.565	0.055	0.744	0.065	0.781
$b=m=5$	2.356	0.502	0.052	0.796	0.065	0.786
$b=m=6$	0.994	0.803	0.051	0.830	0.061	0.811
$b=m=7$	1.249	0.741	0.042	0.945	0.042	0.923
$b=m=8$	0.353	0.950	0.049	0.853	0.051	0.872
$b=m=9$	0.296	0.961	0.047	0.885	0.055	0.845

This table reports three different tests of uniformity of the cumulants with no transformation and of the cumulants transformed by models with different numbers of moment conditions. It reports Berkowitz, Cramér-von Mises and Kolmogorov-Smirnov tests and associated *p*-values. For the Berkowitz tests, all of the cumulants are scaled by the maximum cumulant so no observation takes a value greater than one. The table shows that we have enough power to reject the $U[0,1]$ distribution in the untransformed case (at least according to the Berkowitz test) and that our models substantially improve the fit of the cumulants.

For the S&P 500 the untransformed data produce a Berkowitz statistic of 9.4 (p = 0.02), rejecting the hypothesis that the untransformed cumulants are uniformly distributed. For the FTSE 100 the Berkowitz statistic is 6.2 (p = 0.10), which, again, suggests that a nontrivial pricing kernel is needed to fit the data. The Cramér-von Mises and Kolmogorov-Smirnov tests, which make weaker assumptions than does the Berkowitz test, do not reject uniformity for the untransformed cumulants. Allowing for a nontrivial kernel always improves each of the test statistics we report. For the S&P 500 data, the model with $b=m=7$ appears to fit the data best.

The test statistics in Table 2 imply that our estimation procedure succeeds in transforming the S&P 500 data to a $U[0,1]$ sample quite well regardless of the number of moment conditions chosen. These results should not be considered a formal test comparing the transformed data to the nontransformed data. Still, our transformation appears to improve the fit and according to all of the test statistics the fit is very good for both indexes.

We note that the estimated SDF based on the S&P 500 data, which offer more observations than do the FTSE data, is clearly downward sloping and the pointwise confidence intervals do not allow us to reject monotonicity. Furthermore, the confidence intervals are rather tight between 0.95 and 1.05, a region where many previous studies and our own classical method estimates find the SDF to be increasing. Our estimated downward-sloping SDF is in agreement with mainstream financial and economic theory that risk-averse investors' marginal rates of substitution should be downward

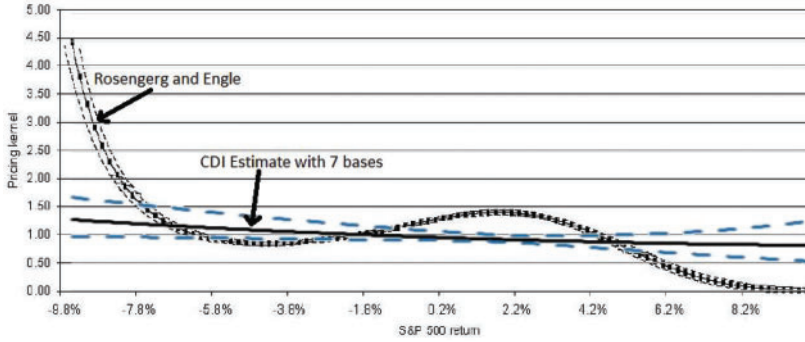


Figure 9
Power of the CDI estimate to reject nonmonotonicities
This figure illustrates the relative power of the CDI estimate by plotting the CDI estimate with seven bases and its corresponding confidence intervals over the time-invariant pricing kernel estimate of Rosenberg and Engle (2002). The figure is generated by matching the scale and range of Figure 6 from Rosenberg and Engle and then superimposing the CDI estimate over the Rosenberg and Engle Figure. It shows that the CDI estimate can easily reject large portions of the Rosenberg and Engle estimate and that the CDI confidence bounds are approximately as tight as those of Rosenberg and Engle.

sloping as a function of states of the world. Although the estimated FTSE 100 SDF is upward sloping in the region of large positive returns, the 95% confidence intervals show that this nonmonotonicity is not statistically significant. Thus, our evidence suggests that avoiding the mixture of forward-looking and historical data results in a solution to the pricing kernel puzzle.

Our SDF estimates do not display any significant nonomonotonicity, but one potential concern is that our method may lack power. To gauge the relative power of the CDI method given the type of data that are available, we directly compare our S&P 500 estimate and its confidence intervals to the average SDF estimate from Rosenberg and Engle (2002). Specifically, we take Figure 6 from Rosenberg and Engle (2002), and we scale a plot of our SDF estimate with $m = b = 7$ to exactly match the scale of the image and then we position our plot over the figure from Rosenberg and Engle (2002). Figure 9 provides the result. Although our "cut-and-paste" method is clearly not sufficiently precise for formal testing, it provides a very direct way to compare the power of our CDI estimates to the power of an influential alternative. Figure 9 shows that the CDI method confidence intervals from our bootstrap procedure are not very different from the confidence intervals in Rosenberg and Engle (2002). Thus, although the CDI method does sacrifice some power for the sake of consistency, its power is comparable to that of other methods. Figure 9 further shows that the CDI method has sufficient power to reject the type of nonmonotonicity that has been found in other papers. The confidence interval of Rosenberg and Engle (2002) falls outside of our CDI confidence intervals for most of the length of the plot.

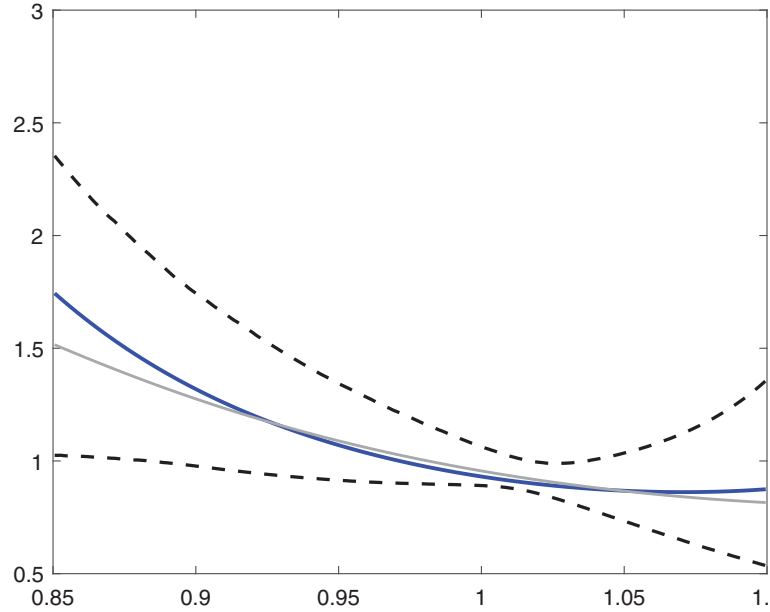
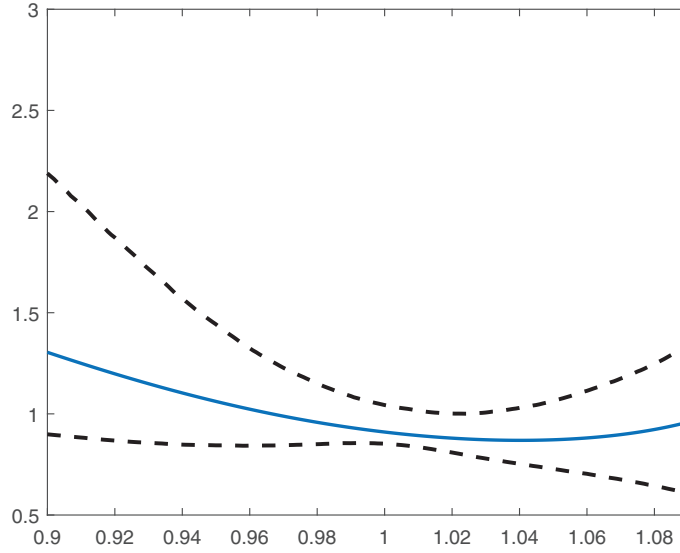


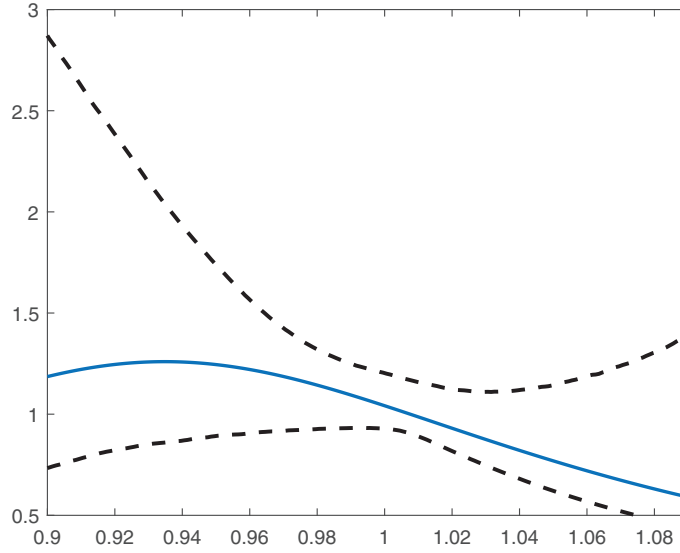
Figure 10
IV estimate of the stochastic discount factor using the CDI method: S&P 500
The result of our IV CDI estimation of the pricing kernel for the S&P 500 is plotted above. The function is estimated using the level of the VIX index as an instrument in the GMM estimation procedure. The level of the VIX index is observed at the same time as the option prices that are used to estimate risk-neutral densities. The dark blue line represents the conditional estimate, whereas the lighter gray line displays the unconditional estimate. Ninety-five percent confidence intervals, which are plotted with dashed lines, are based on 20,000 bootstrap iterations of the CDI method, sampling our set of dates with replacement.

4.2.3 Conditional SDF. We have performed two additional analyses to determine whether the SDF varies with time in an important way. In one analysis, we used the level of the VIX index as an instrumental variable in our GMM estimation of the S&P SDF. Figure 10 illustrates the result of this analysis. As can be seen from the figure, the level of the VIX, when used as an instrument, does not alter our estimated pricing kernel very much. This estimate does not literally allow the SDF to vary with time, but, by weighting the moment conditions more heavily when volatility is high, it gives an indication of how much the SDF might move with volatility. Because the estimate with the VIX instrument is not very different from the estimate without the instrument, we infer that the SDF is not likely to be very different in high and low volatility states.

In a second analysis, we separately estimated the S&P SDF for high and low values of the VIX index. Specifically, we used data for all of the months in which the VIX is above the median VIX value to estimate our "High VIX SDF," and we use data for below median VIX months for our "Low VIX SDF." This gives us a simple conditional or time-varying SDF estimate. We have



(A) High VIX SDF



(B) Low VIX SDF

Figure 11
High and low VIX conditional pricing kernels
The plots above report the SDF for the S&P 500 conditional on above-median VIX observations and below-median VIX observations. Each of the SDF estimates uses 114 months of data.

produced two SDF plots, with each plot based on 114 monthly observations. Figure 11 illustrates the results of this analysis. Although both of the plots in Figure 11 show a little bit of nonmonotonicity, the nonmonotonicity is not close to statistically significant. Both estimates are clearly downward sloping around the current value of the index. We have concluded that allowing the SDF to vary with the level of the VIX index is unlikely to generate significant nonmonotonicity.

Several authors have recently tried to make sense of relatively high-frequency classical SDF estimates. Our results suggest that although risk-neutral densities fairly dramatically change from month to month, the pricing kernel does not change very much with volatility. They further cast doubt on the validity of high-frequency SDF estimates because such estimates are likely to be inconsistent for the underlying SDF.

5. Conclusion

The pricing kernel puzzle is the finding that the stochastic discount factor implied by option prices and historical returns data is not monotonically decreasing in market returns. We argue that this finding is an artifact of econometric technique, driven particularly by comparing two estimates of densities that condition on different information sets. We prove that under realistic assumptions the classical technique for estimating the SDF is asymptotically inconsistent. We show with simulations that the classical technique fails to produce an estimate that is close to the SDF used to generate the simulation data. We argue that high-frequency estimates of the SDF are generally not informative because of this problem.

We propose a new nonparametric pricing kernel estimator that properly reflects all the information that option investors use when they set option prices. Our estimator outperforms the classical method in simulations. In S&P 500 and FTSE index option data, our estimator suggests that the pricing kernel is monotonically decreasing in market returns. Allowing our pricing kernel estimates to vary with the level of the VIX index does not change the monotonicity of our estimate.

It is important to confirm that the stochastic discount factor is monontonically decreasing in market returns because a discount factor that increases in returns over some range implies that the representative agent prefers lower returns (or higher risk) over that range. It is unnatural to think of the representative agent exhibiting risk-loving behavior over any range of market returns. Explaining the pricing kernel puzzle therefore lends credence to standard risk and return theory.

Appendix A. Risk-Neutral Densities

Each month, for the options data with best bids (or last prices when bids are not available) exceeding \$3/8, we fit a fourth-degree spline to implied volatilities associated with each

observed strike price. This is done by placing a single knot at the close price on the day the option is traded, with the remainder of the required knots placed at the minimum and maximum strike prices within our sample. This creates a continuous curve in the implied volatility space. We then convert the implied volatility curve back to the price space by inverting the transformation used to obtain implied volatilities. With the given prices we apply the result of Breeden and Litzenberger (1978) that $\frac{d\mathbb{F}^Q}{dK} = e^{rT}\frac{\partial^2 C}{\partial K^2}$, where $\mathbb{F}^Q$ represents the risk-neutral CDF and $\frac{d\mathbb{F}^Q}{dK}$ represents the density over prices, K. Because we smooth implied volatilities, our estimation procedure always results in reasonable density functions with positive values.

The practice of removing options data with very small prices is standard in the options literature as options with extremely low prices tend to provide misleading data because they are so far out of the money. Although extremely small prices can often give rise to misleading data, leaving them out of our data poses a problem as well because our estimated densities are generally truncated in the tails, especially in the upper tail because far out-of-the-money call options are relatively thinly traded. The densities obtained by taking second derivatives over strike prices will often look like that in panel A of Figure 2. We refer to this part of the density as the *truncated* density. It is clear from the figure that truncating the data in our sample can potentially cause us to miss out on a portion of the density. We circumvent this problem by applying the method of Figlewski (2008) to estimate the tails of the risk-neutral distributions in our sample.

The tail estimation method relies on results from Pickands III (1975) and Balkema and De Haan (1974), both of whom show that for an independent, identically distributed sequence of random variables, the conditional distribution given that the variable exceeds some threshold approaches a generalized Pareto distribution as the specified threshold becomes large. Following the logic of this result, we find the parameters from a generalized Pareto distribution that give the closest match to the truncated risk-neutral density close to the truncation. By pasting the resultant generalized Pareto distribution onto the truncated risk-neutral density, we complete the estimation of the entire density.[17]

The generalized Pareto distribution is characterized by three parameters: a location parameter, a scale parameter, and a shape parameter. To fit the tail distribution, we choose three points on each side of the truncated distribution. With these three points, we then find the three parameter values of the generalized Pareto distribution that lies closest to the truncated distribution at the three points. By choosing three points, we are able to identify the three parameter values. We do this for each tail of the distribution. Whereas Figlewski (2008) only uses two points for each tail and imposes the additional constraint that the area under the curve must equal one, we find that the optimization gives smoother transitions between the truncated density and the tails if we do not include the constraint on the area. Instead, we match three points in each tail and then normalize our estimate to ensure that the area of the density is equal to one. In most cases, this normalization does not change the curve estimation much at all as the tail matching itself gives densities whose area is nearly equal to one. In the few cases in which the normalization has much impact, imposing a constraint on the area in the tail-matching optimization results in awkward kinks in the density that are clearly just an artifact of the optimization and its constraints.

In a small number of cases, the truncated part of the distribution does not go far enough into the tail of the distribution to allow the tail-matching procedure to fit well. This happens when the upper

[17] Our method slightly differs from that of Figlewski (2008), who uses a generalized extreme value distribution rather than a generalized Pareto distribution to estimate the tails of the risk-neutral density. The use of generalized extreme value distribution comes from similar theory of statistics of extremes. The Fisher-Tippett theorem (see for example Embrechts, Klüppelberg, and Mikosch (1997)) states that the sample maximum of an independent, identically distributed sequence of random variables approaches a generalized extreme value distribution as the sample size approaches infinity. However, because we are looking at matching the tail of the distribution beyond some extreme point determined by our data, we feel that an application of the results in Pickands III (1975) and Balkema and De Haan (1974) is most appropriate. So we use a generalized Pareto distribution as opposed to a generalized extreme value distribution when estimating the tails of the risk-neutral densities. In more recent work Figlewski also adopts the generalized Pareto distribution.

end of the central distribution, which is determined by our data, does not extend far enough past the peak of the distribution. In these cases, we interpolate the implied volatility curve to larger return values using cubic spline interpolation. The resultant implied volatility curve is then transformed back to the option price space so that we can take the second derivative to obtain the truncated part of the risk-neutral distribution. This extends the truncated part of the distribution just far enough that the tail-matching procedure gives a meaningful upper tail.

Appendix B. Proof of Proposition 2

Here, we supress the time dimension for simplicity. It suffices to prove that for any equivalent measures $\mathbb{F}^Q$ and $\mathbb{F}^P$ on $\mathbb{R}$ with random variable $X \sim \mathbb{F}^P$, there exists a *unique* $(a.s.\ \mathbb{F}^Q)$ nonnegative function $g:\mathbb{R}\to\mathbb{R}_+$ such that

$$\int_{-\infty}^{X} g(y)d\mathbb{F}^Q(y) \sim U[0,1]. \tag{B1}$$

We first prove existence. We can apply the Radon-Nikodym Theorem on the probability space $(\mathbb{R},\mathcal{B}(\mathbb{R}))$, where $\mathcal{B}(\mathbb{R})$ is the Borel σ-field generated on $\mathbb{R}$. Then by the Radon-Nikodym Theorem, there exists $(a.s\ \mathbb{F}^Q)$ unique random variable $\frac{d\mathbb{F}^P}{d\mathbb{F}^Q}$ such that

$$\mathbb{F}^P((-\infty,x]) = \mathbb{F}^P(x) = \int_{-\infty}^{x} \frac{d\mathbb{F}^P}{d\mathbb{F}^Q}(y)d\mathbb{F}^Q(y) \quad \forall x\in\mathbb{R}. \tag{B2}$$

Now if we define $\mathcal{G}(X)$ by

$$\mathcal{G}(X) \equiv \int_{-\infty}^{X} g(y)d\mathbb{F}^Q(y), \tag{B3}$$

we know from Equation 4, that if we take $g(y) = \frac{d\mathbb{F}^P}{d\mathbb{F}^Q}(y)$, then we have $\mathcal{G}(X)\sim U[0,1]$. This establishes existence.

Next we establish (almost sure $\mathbb{F}^Q$) uniqueness. Because g is nonnegative, the function $\mathcal{G}$ uniquely determines nonnull sets in $\mathcal{B}(\mathbb{R})$ where g must be zero. Any functions satisfying the criteria of the proposition must take the value zero over the exact same subsets of $\mathcal{B}(\mathbb{R})$. It only remains to show that over the sets, where $g\neq 0$, the functional form is unique. Let $\mathcal{N}$ denote $\mathbb{F}^Q$ nonnull sets, where $g>0$. Over such sets, the function $\mathcal{G}$ is invertible because $g>0$.

Suppose another function, g', satisfies Equation (B3) over the (possibly reduced) space $\mathcal{N}$. Define $\mathcal{G}'_t$ as

$$\mathcal{G}'(X) \equiv \int_{-\infty}^{X} g'(y)d\mathbb{F}^Q(y),$$

where, by our assumption on g', we know $\mathcal{G}'(X)\sim U[0,1]$. Because $\mathcal{G}$ and $\mathcal{G}'$ are invertible over $\mathcal{N}$, we know that on the restricted domain, for a fixed x,

$$\mathbb{F}^P(\mathcal{G}'(X)\le x) = \mathbb{F}^P(X\le \mathcal{G}'^{-1}(x))$$

and

$$\mathbb{F}^P(\mathcal{G}(X)\le x) = \mathbb{F}^P(X\le \mathcal{G}^{-1}(x)).$$

Because $\mathbb{F}^P$ and $\mathbb{F}^Q$ are equivalent by assumption, and $\mathcal{N}$ sets are $\mathbb{F}^Q$ nonnull, it follows that $\mathcal{N}$ sets are $\mathbb{F}^P$ nonnull. This implies that $\mathbb{F}^P(X\le\cdot)$ is a strictly increasing function over $\mathcal{N}$ sets and hence

$$\mathcal{G}'^{-1}(x) = \mathcal{G}^{-1}(x)$$

for a fixed x. It follows that for deterministic sets E (e.g., $E=(-\infty,x]$)

$$\int_E g'(y)d\mathbb{F}^Q(y) = \int_E g(y)d\mathbb{F}^Q(y) \quad \forall E\subset\mathcal{B}(\mathcal{N}). \tag{B4}$$

Now we can apply the Radon-Nikodym Theorem on $\mathcal{N}$. From Equation (B4), the Radon-Nikodym Theorem implies $g'=g\ (a.s.\ \mathbb{F}^Q)$ over $\mathcal{N}$. Because the values of g and g' must be zero on nonnull subsets of the compliment $\mathcal{N}^c$, we have that $g'=g\ (a.s.\ \mathbb{F}^Q)$ and hence g is unique $(a.s.\ \mathbb{F}^Q)$. ■

Appendix C. Estimation Details

We use finite order cubic B-splines to approximate the function g. We use cubic B-splines as opposed to polynomials because they offer more flexibility in estimating functional forms. The use of splines of order b requires that we first choose the placement of knots that will determine the bases to be used for estimation purposes. We simply use equally spaced knots over our range of returns. The minimum of the range is set to the minimum value for which our estimated risk-neutral densities, over all months in the sample, have a positive (machine measurable) support. The maximum of the range is the maximum realized return within our sample. This range corresponds to the values over which the integral in Equation (B1) is taken, once we replace $-\infty$ with the minimum value for which $d\mathbb{F}^Q$ has positive support. The cubic B-spline of order b is a linear combination of b basis functions,

$$g(y) \approx \sum_{j=1}^{b} \theta_j B_j(y),$$

where $B_j(\cdot)$ denotes the jth basis function of the spline. Using this approximation to the function g, we can also approximate the integral in Equation (6) as a linear combination of integrals,

$$\int_{-\infty}^{X} g(y) d\mathbb{F}^Q(y) \approx \sum_{j=1}^{b} \theta_j \int_{-\infty}^{X} B_j(y) d\mathbb{F}^Q(y). \tag{C1}$$

Because we have a linear function in θ, our estimated function $\widehat{g_{t,s}}$ is given by

$$\widehat{g_{t,s}} = \hat{\theta}' B, \tag{C2}$$

where $\theta = (\theta_1, ..., \theta_b)'$ and $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_b)'$. B is a data matrix where each row is a spline basis function evaluated over the support of our return data.[18] We can formally represent the data matrix $A \in \mathbb{R}^{T \times b}$ by

$$A_{i,j} = \int_{-\infty}^{X_i} B_j(y) d\mathbb{F}_i^Q(y), \qquad i = 1, ..., T; \;\; j = 1, ..., b, \tag{C3}$$

where T represents the number of monthly estimates of $\mathbb{F}^Q$ available and b is the number of basis functions included in our estimated spline approximation of g.

Because we will be using nonoverlapping data on monthly options from OptionMetrics, which only goes back to 1996 for the S&P 500 and 2002 for the FTSE, as described in Section 4.1, our sample is not extremely large. For this reason, we use a GMM type optimization with only the first stage optimization. This has been shown to perform best when one does not have extremely large data sets with which to perform GMM estimation (see, e.g., Hayashi (2000)).

To estimate θ, we solve the first stage GMM optimization,

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^b}{\operatorname{argmin}} \sum_{j=1}^{m} \left(\sum_{t=1}^{T} \left(\underbrace{\sum_{i=1}^{b} \theta_i \int_{-\infty}^{X_t} B_i(y) d\mathbb{F}_t^Q(y)}_{\hat{g}(\theta)} \right)^j - \frac{1}{j+1} \right)^2, \tag{C4}$$

where we use the fact that the jth moment of the uniform distribution over the unit interval is equal to $\frac{1}{j+1}$ and we use the first m moments in estimating the vector θ. It is important to note that the

[18] We do not impose restrictions on g such that it is always strictly positive or that the right-hand side of (8) converges to 1 as X approaches ∞.

solution to Equation (C4) is found by minimizing over $\mathbb{R}^b$, in other words we place no restrictions on our estimate of θ.

Once we have the estimated $\hat{\theta}$, it is straight forward to estimate g. We simply need to substitute $\hat{\theta}$ into Equation (C2) to obtain our estimate for g, the inverse of the Radon-Nikodym derivative, $\frac{d\mathbb{F}_t^Q}{d\mathbb{F}_t^P}$, for all t. By Equation (2), $\frac{d\mathbb{F}_t^Q}{d\mathbb{F}_t^P} = \widehat{\frac{1}{g_{t,s}}}$ for all t. So our estimated SDF is given by $e^{-r_t\tau}\widehat{\frac{1}{g(x_{t+s})}}$, where r_t denotes the risk-free rate at time t, τ represents time to maturity of time t index options on the S&P 500 index and x_{t+s} denotes returns on the S&P 500 index between time t and time s. This can be re-expressed as

$$m_{t,t+\tau}(X) = e^{-r_t\tau}M(X),$$

where $M(x) \equiv \frac{1}{g(x)}$.

The value of $e^{-r_t\tau}$ is very stable over our sample period, and the SDF does not substantially vary over our sample period. We focus only on the estimation of $\hat{M}$ because the time discount factor $e^{-r_t\tau}$ does not tell us anything about investors' preferences over states of the world and returns on market indices. In Section 4, we will discuss our empirical results based on estimates of $M(x)$, as described above.

For the purposes of inference, we calculate pointwise confidence intervals for the estimated SDF. We resample with replacement from the set of rows of the data matrix in Equation (C3). This is equivalent to sampling with replacement from the set of dates associated with each risk-neutral density we estimate. For each sample, we can re-calculate the SDF estimate using the CDI method. We then calculate the accelerated bias-corrected (BCa) percentile bootstrap confidence intervals as described in Efron and Tibshirani (1993). This gives us a virtual continuum of pointwise confidence intervals if we take a fine partition of the return space. However, as is the case with most nonparametric methods, many observations are needed to obtain a very tight confidence interval.

C.1 Model Evaluation

To estimate θ, we use a GMM type estimation to match the resultant estimate to the moments of the uniform distribution over the unit interval like in Equation (C4). This requires that we choose the number of moment restrictions m as well as b, the dimension of θ. As we do throughout the paper, we wish to impose as little structure as possible on the estimation. This allows us to estimate the SDF in a manner we feel best approximates the market's beliefs and risk preferences that determine the SDF. In keeping with this goal, we estimate the SDF with several different values of m and b and report all of our estimates. To keep our model from being unidentified, we need to keep b less than the number of moment conditions, m. In some preliminary estimation using a smaller data set, models with $b=m$ seem to fit the data better than alternatives with fewer basis functions, so we explore a series of models with $b=m$.

We evaluate models with different values of b and m with three common distributional test statistics. Table 2 reports results from the Berkowitz, Cramér-von Mises, and Kolmogorov-Smirnov tests. We also calculate the p-values corresponding to the null hypothesis that the estimated distribution is uniform over $[0,1]$. We calculate p-values based on simulated outcomes as opposed to asymptotic distributions. This gives us a sense of exactly how well our model selection and subsequent optimization perform given our finite sample size. Our estimated CDF is given by the empirical CDF corresponding to the estimated vector $\hat{\theta}$ for a given combination of b and m,

$$\hat{\mathbb{F}}_{b,m}(x) = \frac{1}{T}\sum_{t=1}^{T}\mathbb{1}\left(\sum_{j=1}^{b}\hat{\theta}_j\int_{-\infty}^{X_t}B_j(y)d\mathbb{F}_t^Q(y) \leq x\right), \tag{C5}$$

where $\mathbb{1}(E)$ represents the indicator function taking value 1 in the where event E is true and the value zero otherwise.

We evaluate Equation (C5) with the estimated parameter vectors and then compare the test statistics for each, keeping θ exactly identified so that $b=m$. That is, we require the number of

moment restrictions to be equal to the dimension of the vector to be estimated, θ. The smallest test statistics correspond to the model for which the CDI procedure transforms the data to a distribution closest to the uniform distribution. A model with the smallest statistics might be considered the optimal model.

Appendix D. Simulation Details

We begin by defining an SDF that will be used to generate our data. As we have done throughout the paper, we refer to the SDF as the Radon-Nikodym derivative of the risk-neutral with respect to the physical measure and we ignore the rate of time discount factor. We specify that the SDF in the economy is given by a risk-neutral log-normal density with location parameter $\mu_Q = 0.00011$ and scale parameter $\sigma_Q = 0.0526$ and a physical log-normal density with location parameter $\mu_P = 0.0040$ and scale parameter $\sigma_P = 0.0526$. These parameters are chosen to match the average of the monthly distributions corresponding to a subset of the S&P 500 sample (ending in 2012). Notice that we have set $\sigma_Q = \sigma_P$ to be consistent with the Black-Scholes model. Like in the Black-Scholes model, the location parameters μ_q and μ_p differ.

We fit a series of S&P 500 monthly variances to an Ornstein-Uhlenbeck process. This is done by simply taking the variance of each risk-neutral density estimated using the method described in Section 2.1 for the years up to 2012 and maximizing the likelihood function to estimate the parameters of the Ornstein-Uhlenbeck process being fit to the series of variances. With the resultant estimated parameters of the process, we simulate a series of N risk-neutral variances. Along with the fixed location parameter μ_Q and the assumption of log-normality, this variance process gives us a series of N risk-neutral densities. Both the CDI method and the classical method use these densities to recover the SDF estimates. Once we have the risk-neutral densities we can use the true stochastic discount factor to get the physical densities corresponding to each risk-neutral density. Recall that $d\mathbb{F}_t^P = \left(\frac{d\mathbb{F}_t^Q}{d\mathbb{F}_t^P}\right)^{-1} d\mathbb{F}_t^Q$. We use this fact to get the physical densities corresponding to each risk-neutral density. We then take a single random draw from each of the physical densities in the series. This is done by first recovering the CDF, $\mathbb{F}_t^P$ from each physical density $d\mathbb{F}_t^P$. Next we generate a series of draws from a uniform distribution over the unit interval, $u_t \sim U[0,1]$, for $t \in \{1,2,...,N\}$. Draws from the physical density $d\mathbb{F}_t^P$ are given by $(\mathbb{F}_t^P)^{-1}(u_t)$, which has exactly the distribution of our physical density $d\mathbb{F}_t^P$. Each of these draws from the physical distribution correspond to the realized monthly returns we observe in the data. Now we have a series $\left(d\mathbb{F}_t^Q, X_t\right)$ for $t \in \{1,2,...,N\}$, where X_t represents the time t realization of a draw from the time t physical density $d\mathbb{F}_t^P$. Because the physical density and the true SDF are unobservable to the econometrician, this series of risk-neutral densities and single realizations from physical densities replicates the data that are available to the econometrician.

References

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

Audrino, F., and P. Meier. 2012. Empirical pricing kernel estimation using a functional gradient descent algorithm based on splines. Working Paper.

Bakshi, G., D. Madan, and G. Panayotov. 2010. Returns of claims on the upside and the viability of U-shaped pricing kernels. *Journal of Financial Economics* 97:130–54.

Balkema, A. A., and L. De Haan. 1974. Residual life time at great age. *Annals of Probability* 2:792–804.

Barone-Adesi, G., R. F. Engle, and L. Mancini. 2008. A garch option pricing model with filtered historical simulation. *Review of Financial Studies* 21:1223–58.

Barone-Adesi, G., L. Mancini, and H. Shefrin. 2013. A tale of two investors: Estimating optimism and overconfidence. Working paper.

Beare, B. K., and L. Schmidt. 2016. An empirical test of pricing kernel monotonicity. *Journal of Applied Econometrics* 31:338–56.

Billingsley, P. 2012. *Probability and measure*. New York: John Wiley & Sons.

Bliss, R. R., and N. Panigirtzoglou. 2005. Option-implied risk aversion estimates. *Journal of Finance* 59:407–46.

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

Chabi-Yo, F., R. Garcia, and E. Renault. 2007. State dependence can explain the risk aversion puzzle. *Review of Financial Studies* 21:973–1011.

Chaudhuri, R., and M. Schroder. 2009. Monotonicity of the stochastic discount factor and expected option returns. Working paper.

Christoffersen, P., S. Heston, and K. Jacobs. 2013. Capturing option anomalies with a variance-dependent pricing kernel. *Review of Financial Studies* 26:1962–2006.

Constantinides, G. M., M. Czerwonko, J. C. Jackwerth, and S. Perrakis. 2011. Are options on index futures profitable for risk-averse investors? empirical evidence. *Journal of Finance* 66:1407–37.

Cuesdeanu, H., and J. Jackwerth. 2016. The pricing kernel puzzle in forward looking data. Working paper.

Efron, B., and R. J. Tibshirani. 1993. An introduction to the bootstrap. In *Monographs on statistics and applied probability*, eds. B. Efron and R. J. Tibshirani. Chapman and Hall: New York.

Embrechts, P., C. Klüppelberg, and T. Mikosch. 1997. Modelling extremal events: for insurance and finance In *Stochastic modelling and applied probability*, eds. P. W. Glynn and Y. Le Jan. New York: Springer.

Figlewski, S. 2008. Estimating the implied risk neutral density for the U.S. market portfolio. *Volatility and time series econometrics: Essays in honor of Robert Engle*, eds. T. Bollerslev, J. Russell, and M. Watson. Oxford, UK: Oxford University Press.

Foster, D. P., and D. B. Nelson. 1996. Continuous record asymptotics for rolling sample variance estimators. *Econometrica* 64:139–74.

Glosten, L. R., R. Jagannathan, and D. E. Runkle. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48:1779–801.

Grith, M., W. K. Härdle, and V. Krätschmer. 2017. Reference dependent preferences and the empirical pricing kernel puzzle. *Review of Finance* 21:269–98.

Härdle, W., Y. O., and W. Wang. 2014. Uniform confidence bands for pricing kernels. *Journal of Financial Econometrics* 12:1–38.

Hayashi, F. 2000. *Econometrics*, vol. 1. Princeton: Princeton University Press.

Izbicki, R., A. B. Lee, and C. M. Schafer. 2014. High-dimensional density ratio estimation with extensions to approximate likelihood computation. Proceedings of the 17th International Conference on Artificial Intelligence and Statistics.

Jackwerth, J. C. 2000. Recovering risk aversion from option prices and realized returns. *Review of Financial Studies* 13:433–51.

Pickands, J., III. 1975. Statistical inference using extreme order statistics. *Annals of Statistics* 3:119–31.

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* 64:341–72.

Song, Z., and D. Xiu. 2016. A tale of two option markets: Pricing kernels and volatility risk. *Journal of Econometrics* 190:176–96.

Sugiyama, M., T. Suzuki, S. Nakajima, H. Kashima, P. von Bünau, and M. Kawanabe. 2008. Direct importance estimation for covariate shift adaptation. *Annals of the Institute of Statistical Mathematics* 60:699–746.

Sugiyama, M., I. Takeuchi, T. Suzuki, T. Kanamori, H. Hachiya, and D. Okanohara. 2010. Conditional density estimation via least-squares density ratio estimation. Proceedings of Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS2010), JMLR Workshop and Conference Proceedings.

Ziegler, A., 2007. Why does implied risk aversion smile? *Review of Financial Studies* 20:859–904.