Pricing Kernel Monotonicity and Conditional Information

Matthew Linn, Sophie Shive and Tyler Shumway*

October 21, 2015

Abstract

A large literature finds evidence that pricing kernels estimated nonparametrically from option prices and historical returns are not monotonically decreasing in market index returns. We propose a new nonparametric estimator of the pricing kernel that reflects the information available to investors who set option prices. In simulations, the estimator outperforms current techniques. Our empirical estimates using S&P 500 index option data from 1996-2012 and FTSE 100 index option data from 2002-2013 suggest that the "pricing kernel puzzle" is a byproduct of econometric technique rather than a behavioral or economic phenomenon.

*University of Michigan, University of Notre Dame, and University of Michigan, respectively. We thank seminar participants at Indiana University, the University of Michigan, and the University of Washington and the WFA (2014). We also thank Robert Dittmar, Wayne Ferson, Stephen Figlewski, Benjamin Golez and Steven Heston (WFA disscussant) for helpful comments. Shumway is grateful for the financial support provided by National Institute on Aging grant 2-P01-AG026571.

1 Introduction

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 tradeoff 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 many of the methods used to estimate the pricing kernel 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.

Since the pricing kernel summarizes the attitudes of economic agents about risk, understanding its behavior is one of the primary goals of asset pricing. The research on SDF estimation from option data starts with Jackwerth (2000) and Ait-Sahalia and Lo (2000), which exploit the relation between option prices and the risk-neutral density. The risk-neutral density is proportional to the SDF multiplied by the (physical) density of the underlying asset. 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 and then they divide the resulting risk-neutral density by a nonparametric estimate of the physical density based on historical return data. The resulting ratio of densities is what we refer to as the "classic" 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. Many other researchers apply similar techniques, though sometimes with important improvements, and also find that the SDF appears to be a nonmonotonic function. More recent papers in this literature include Rosenberg and Engle (2002), Chaudhuri and Schroder (2009), Audrino and Meier (2012), Härdle, Okhrin, and Wang (2014), Beare and Schmidt (2013), Bakshi and Chabi-Yo (2013), and Song and Xiu (2014). 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, but the noise in average returns makes it difficult for them to draw strong conclusions. 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.

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. For financial economists, this is extremely counterintuitive. Even with a multidimensional state vector, it is difficult to see how a higher realized value of the market portfolio could be systematically worse than a lower value. A non-monotonic pricing kernel is so surprising that it has been coined the "implied risk aversion puzzle" or the "pricing kernel puzzle" in the literature that has developed to explain it. Ziegler (2007) attributes it to differences in beliefs among agents about the mean and variance of expected returns. Polkovnichenko and Zhao (2013) postulate a model with rank-dependent utility to explain the puzzle. Barone-Adesi, Mancini, and Shefrin (2013) explain the puzzle with overconfidence, and Grith, Härdle, and Krätschmer (2013) propose heterogeneity of investor reference points.

Another set of explanations for the puzzle relies on state dependence, generally with higher moments as additional factors. Chabi-Yo, Garcia, and Renault (2007) identify latent factors as a probable cause and propose a parametric option pricing model that can generate upward slopes. Christoffersen, Heston, and Jacobs (2013) and Song and Xiu (2014) propose models that include volatility as a factor.

One criticism of the empirical papers that find nonmonotonicity is that they do not agree

on the location of nonmonotonicities. In Ait-Sahalia and Lo (2000) and Audrino and Meier (2012) for example, positive slopes appear in the center of the distribution of returns, while Christoffersen, Heston, and Jacobs (2013) and Bakshi, Madan, and Panayotov (2010) find a U-shaped kernel. If nonmonotonicities appear in different locations when different data sets and methods are used then the pricing kernel puzzle may not be very robust. This is perhaps because naive estimators of the ratio of two separately estimated densities are known to perform poorly.[1]

Another criticism of almost all of the empirical papers that find nonmonotonicity is that they compare a conditional risk-neutral density to an essentially unconditional physical density. Since option prices, like all market-determined prices, are discounted expectations of future cash flows conditional on all information available, the risk-neutral density estimated from option prices is a conditional density. In the data, most of the moments of the estimated risk-neutral densities change substantially from one month to the next. Common practice for nonparametrically estimating physical densities conditional on all available information is to rely on the use of a rolling window of historical data to make the physical density conditional.[2] 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 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. 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.[3]

[1]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.

[2]See Foster and Nelson (1996).

[3]Note that methods that do not compute a ratio of densities, such as that suggested in Hansen and Renault (2010), for example, also assume that the objective probability distribution of returns does not change over time.

To demonstrate the problem caused by failing to account for conditional information in the denominator of the SDF, we give two examples of how nonmonotonicity can arise in an estimated pricing kernel implied by a misspecified Black-Scholes model. The first example shows that in a simple single-period setting we can get nonmonotonic ratios of risk-neutral densities to physical densities if we allow the variances of the two to differ as they may when we compare conditional and unconditional densities. In the second example we simulate data from the misspecified Black-Scholes model in order to show that using a rolling window to estimate the physical density while using strictly conditional estimates of the risk-neutral density can lead to nonmonotonic and inaccurate estimators.

We propose a new method that avoids 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 available for each density. Given these data, we can compute the integral of each of the risk-neutral densities from $-\infty$ to their corresponding actual return realizations to obtain a set of realized CDF values. If the risk-neutral density is the same as the physical density, the resulting CDF values will be uniformly distributed. To the extent that the empirical distribution of the CDF values is not uniform, we can use the distribution of the CDF values to identify the pricing kernel. This is the intuition behind our pricing kernel estimator. In simulations we find that our method substantially outperforms the classic method in recovering the SDF that generated 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 propose a stable but completely flexible unconditional SDF function, which we model with a spline estimator. We estimate the spline

using the fact that integrating the inverse of the SDF times the risk-neutral density up to each realized value should produce a set of cumulants that are uniformly distributed. In using this fact to identify our model, we follow Bliss and Panigirtzoglou (2005), who use the same fact to estimate implied risk aversion coefficients parametrically. We also allow our estimated SDF to vary with time by using the level of the VIX index as an instrument in our GMM estimation. We use a bootstrapping procedure to estimate confidence bounds for our nonparametric SDF. We refer to our method as a 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 surprisingly different characteristics. Furthermore, when we (incorrectly) follow the classic 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, which suggests they are not econometrically robust. However, when we properly account for the conditional nature of the risk-neutral densities estimated from option prices by using the CDI estimator, the resulting pricing kernel estimate is monotonically decreasing. The kernel estimate is decreasing even when we use the VIX index as an instrumental variable in our GMM estimation to allow the SDF to vary with time. Thus, we conclude that the pricing kernel puzzle is likely to be an artifact of econometric technique.

In the next section of the paper, we discuss both the classic estimation method and our new CDI method in detail. We also motivate our estimation method theoretically and show that the misspecified Black-Scholes model produces a nonmonotonic SDF. In Section 3 we report the results of simulations designed to compare the performance of the CDI method to that of the classic method. Section 4 describes the data that we use for our tests, and Section 5 reports our primary results. Section 6 concludes.

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 perhaps the biggest 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. We also discuss the classic nonparametric approach to estimating the SDF, point out its shortcomings and discuss how these can lead to economically implausible pricing kernels. In Section 5, 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 Classic Method

The classic 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.[4] 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

$$m_{t,t+s} = e^{-rs}\frac{d\mathbb{F}^Q}{d\mathbb{F}^P}, \tag{1}$$

a change of measure between two *conditional* probability measures where each probability is conditional upon the same information set, $\mathcal{F}_t$.

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

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)^{-1}. \tag{2}$$

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

Furthermore, if both $\mathbb{Q}$ and $\mathbb{P}$ are equivalent to dx, then

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{d\mathbb{Q}}{dx} \Big/ \frac{d\mathbb{P}}{dx}. \tag{3}$$

The corollary allows one to express the Radon-Nikodym derivative as a ratio of two derivatives. This corolary is implicitly invoked in the method we refer to as the classic nonparametric method of SDF estimation. The classic approach relies on the fact that the SDF is proportional to the Radon-Nikodym derivative of the risk-neutral distribution with respect the physical distribution. Furthermore, the 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, $\frac{d\mathbb{F}^Q}{dx}$ to the physical density, $\frac{d\mathbb{F}^P}{dx}$. 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.

Since the classic nonparametric method relies on estimation of the Radon-Nikodym derivative via Equation (3), it reduces to estimating the two densities separately. Theoretically, the densities in the numerator and denominator of the Radon-Nikodym derivative in Equation (3) are conditional densities; they take into account investors' beliefs at the time of investment, conditional on all information available, $\mathcal{F}_t$. As such, we ideally should take care to estimate the densities in a conditional, forward-looking manner. For estimation of the numerator, one typically relies 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 options data typically allow us to 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 which is forward-looking and hence conditional in nature.

On the other hand, there are no known methods for estimating $d\mathbb{F}_t^P$, the time t physical density, in a forward-looking manner, taking into account the information investors base their investment decisions on *at time t*. In previous studies, $d\mathbb{F}_t^P$ has been estimated by smoothing

or averaging past realized returns. In order to make the estimates reflect a conditional rather than unconditional density, a rolling window is typically used to estimate the physical density. This approach clearly leaves much to be desired. Nonparametric estimates require large amounts of data, thus forcing recent data, even if it accurately reflects beliefs about the future, to be a small part of the estimated density.

In effect, the classic method of nonparametrically estimating the SDF implicitly assumes that physical probability measures and their corresponding densities are stable over time, or that the conditional densities are the same as unconditional densities. The assumption of stable physical densities and distributions is widely believed to be implausible. The method of Breeden and Litzenberger (1978) for estimating conditional, risk-neutral densities reveals that their time series is not stable. We characterize the risk-neutral densities implied by option price data in Table 1, which is discussed in Section 5. If the risk-neutral densities vary over time, it is plausible that the conditional physical densities do as well.

To investigate whether comparing a conditional density to an unconditional density can cause non-monotonicity in practice, we calculate implied pricing kernels under Black-Scholes assumptions, but with a slightly higher risk-neutral than physical variance. Our example is motivated by the fact that the risk-neutral density can change significantly from period to period while the estimate of the physical density will not change much from period to period. In some periods, the risk-neutral density may have a higher variance than the physical, while in other periods it may have a lower variance. In Panel A of Figure 1, we plot the physical and risk-neutral densities under the assumption that returns are lognormally distributed and have parameter values that correspond to our risk-neutral sample moments. Panel B plots the corresponding pricing kernel function, which is monotonically decreasing in market returns. In Panels C and D, we plot the densities and pricing kernel under the assumption that the variance of the risk-neutral density is slightly higher than that of the physical density, changing the (monthly) σ parameter from 5.26% to 5.50% percent. The pricing kernel in Panel D starts at a very high level and is first decreasing and then increasing, reflecting a pattern often found in prior work. This example only allows the second moment

to differ across these densities. In typical pricing kernel estimation, all the moments of the estimated risk-neutral density can, in principle, vary from period to period while the estimated physical density, based on historical data, is relatively stable. This shows that if the estimated physical density does not change to reflect new information as much as the risk-neutral density does, the corresponding estimated pricing kernel can be increasing over some range. This problem is inherently present in all of the nonparametric pricing kernel estimators based on option prices that we are familiar with.

2.2 Estimating Risk-Neutral Densities

In order 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 below. 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. While 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 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 large 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 which 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. By pasting the resulting generalized Pareto distribution onto the truncated risk-neutral density, we complete the estimation of the entire density.[5]

The generalized Pareto distribution is characterized by three parameters: a location parameter, a scale parameter and a shape parameter. In order 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. While Figlewski (2008) only uses two points for each tail and imposes the additional constraint

[5]Our method differs slightly from that of Figlewski (2008), which 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, since 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.

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 where the normalization has much impact, imposing a constraint on the area in the tail-matching optimization results in awkward kinks in the density which 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 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 resulting 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.

We use risk-neutral densities estimated with this method to calculate the SDF using both the classic nonparametric method and our new CDI method. Using the same set of risk-neutral densities, the classic 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.

2.3 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 classic method, which relies on Equation (3), we are left to estimate the physical densities corresponding to each of the risk-neutral densities. As described above, until now there has been 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. To obtain a conditional estimate, it is best to use as short a window as possible without compromising the integrity of the kernel estimator.[6]

As discussed earlier, 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.

2.4 CDI Approach

To simplify our estimation of the unconditional SDF with option prices observed over a period of time, we need to make an invariance assumption for the SDF. We initially assume the following:

Assumption 1. *The stochastic discount factor over our sample period is time-invariant up*

[6]Recall that our estimation of physical densites is purely illustrative. More recent work has refined this methodology, but finds similar non-monotonic pricing kernels.

to a rate of time discount factor $e^{-r_t\tau}$ and/or conditioning variables, where r_t is the risk free rate at time t and τ is the duration of the payoff period over which the SDF is discounting.

While this is a very common assumption in empirical asset pricing, it probably merits a little extra discussion in this context. It is equivalent to the assumption that the ratio of risk-neutral to physical densities is stable over time. This is a fairly plausible assumption if one believes that the representative investor's preferences are relatively stable over time, since investor preferences are responsible for the difference between the risk-neutral and physical densities. This assumption is consistent with the empirical finding that risk-neutral and physical densities change over time, but it requires that the two vary together. Mathematically, the assumption reduces to stability of $\frac{d\mathbb{F}_t^Q}{d\mathbb{F}_t^P}$, as opposed to stability of $d\mathbb{F}_t^P$. Our assumption is also the key identifying assumption made in Bliss and Panigirtzoglou (2005), where it is argued that this is a more plausible assumption than the assumption that is implicitly required for the classical estimator of the stochastic discount factor. Many of the researchers who apply the classic nonparametric SDF estimation method report an average SDF, and our unconditional estimate can easily be compared to theirs.

In addition to estimating an unconditonal SDF we allow our SDF estimates to vary with time in a specific way that is described below. Thus, the invariance assumption above is not strictly necessary.

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, allows us to circumvent the need for estimating the physical densities corresponding to each of the risk-neutral densities. The property is given in the following proposition:

Proposition 1. *For any continuous random variable, X with CDF $\mathbb{F}_x$, the random variable defined by $\mathbb{F}_x(X)$ is uniformly distributed on the interval* $[0, 1]$,

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

We let $\mathbb{F}_t^P$ be the unobserved probability measure representing investors' aggregate beliefs about returns on the S&P 500 under the physical measure at time t and let returns over the

subsequent period be given by X_t. Now it follows from Proposition 1, that

$$\int_{-\infty}^{X_t} d\mathbb{F}_t^P(x) \sim U[0,1]. \tag{5}$$

Since there are no known methods for estimating $d\mathbb{F}_t^P$ 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.[7] One would then have to find a way to use these returns to estimate forward looking beliefs about returns under the physical measure. As discussed earlier, this method would require something beyond simply smoothing a long time series of *past* returns, since that does not do a good job of estimating the *current* beliefs held by the market. In order 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 our generalized Pareto distribution tail matching procedure. Let $d\mathbb{F}_t^Q$ be the time t risk-neutral probability measure and let $\frac{d\mathbb{F}_t^P}{d\mathbb{F}_t^Q}$ denote the Radon-Nikodym derivative of time t physical distribution with respect to time t risk-neutral distribution. Then

$$\int_{-\infty}^{X_t} d\mathbb{F}_t^P = \int_{-\infty}^{X_t} \frac{d\mathbb{F}_t^P}{d\mathbb{F}_t^Q} d\mathbb{F}_t^Q = \int_{-\infty}^{X_t} \left(\frac{d\mathbb{F}_t^Q}{d\mathbb{F}_t^P} \right)^{-1} d\mathbb{F}_t^Q \sim U[0,1], \tag{6}$$

where the first equality in Equation (6) follows from Equation (1) and the second equality follows from Equation (2).

Since 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 variable $\left(\frac{d\mathbb{F}_t^Q}{d\mathbb{F}_t^P} \right)^{-1}$, which is proportional to the inverse of the stochastic discount factor. Therefore, by estimating $\left(\frac{d\mathbb{F}_t^Q}{d\mathbb{F}_t^P} \right)^{-1}$, we have essentially estimated the stochastic discount factor. It is important, however, that we first establish uniqueness of the random variable we attempt to estimate. The following proposition ensures that there is such a unique random variable.

[7] 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.

Proposition 2. *For any equivalent measures $\mathbb{Q}$ and $\mathbb{P}$ on $\mathbb{R}$ with random variable $X \sim \mathbb{P}$, there exists a unique (a.s. $\mathbb{Q}$) non-negative function $g : \mathbb{R} \to \mathbb{R}_+$ such that*

$$\int_{-\infty}^{X} g(y) d\mathbb{Q}(y) \sim U[0,1]. \tag{7}$$

A proof of this proposition appears in the Appendix.

The function denoted g 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 is not directly applicable here. The functional form of g 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 as values the random variable X can take. The outcomes of the random variable depend upon $\omega \in \Omega$ the probability space determining returns, $X = X(\omega)$. As such, the integral with respect to $d\mathbb{Q}$ can be interpreted as the integral with respect to the measure $\mathbb{Q}(\{\omega : X(\omega) \in dy\})$. In this way, $g(y) = \phi(\{\omega : X(\omega) \in dy\})$, where ϕ is a mapping from Ω to the non-negative real line, $\phi : \Omega \to \mathbb{R}_+$. So $g(y)$ represents possible realizations of the random variable $g(X(\omega)) = \phi(\omega)$. We will let g 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.

Proposition 2 establishes uniqueness of the function g that transforms the integral with respect to measure $\mathbb{Q}$, to a specific distribution. This is similar to the statement of the Radon-Nikodym Theorem. The function g, mapping realizations of returns to non-negative values is itself a random variable, much the same as the Radon-Nikodym derivative. The difference is that here we have a random domain $(-\infty, X]$. We thus estimate the functional form of g 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.

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 non-normal due to the large mass at zero payoffs, and averages of these returns could be unstable. Last, we want to follow a method that is comparable to the existing literature so that a comparison of results is possible along multiple dimensions.

2.5 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. Since any real valued function can be reproduced by a cubic spline of infinite order, this is a completely model-free estimation procedure. 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 which 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 (7) 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,

$$\int_{-\infty}^{X} g(y) d\mathbb{F}^Q(y) \approx \sum_{j=1}^{b} \theta_j \int_{-\infty}^{X} B_j(y) d\mathbb{F}^Q(y). \tag{8}$$

Since we have a linear function in θ, our estimated function $\hat{g}$ is given by

$$\hat{g} = A\hat{\theta}, \tag{9}$$

where $\theta = (\theta_1, ..., \theta_b)'$ and $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_b)'$. A is our data matrix which is expressed in terms of risk-neutral distributions estimated from options data, realized S&P 500 index returns corresponding to each risk-neutral distribution, denoted X_t and the spline basis functions.[8] We can formally represent the data matrix $A \in \mathbb{R}^{T \times b}$ by

$$A_{i,j} = \int_{-\infty}^{X_i} B_j(y) d\mathbb{F}_i^Q(y), \qquad i = 1, ..., T; \;, \quad j = 1, ..., b, \tag{10}$$

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.

Since we will be using non-overlapping 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, 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 for example Hayashi (2000)). In order to make the best use of the data available to us, we optimally choose model parameters b and m in order to balance the trade off between the number of moment restrictions and the number of parameters to be estimated. A larger number of spline basis functions, b, corresponds to a more flexible and accurate spline approximation of the function g. However, increasing the number of basis functions requires that we increase the number of moment restrictions in our estimation because identification of θ requires that the number of moment restrictions be at least as large as the dimension of θ, $b \leq m$. Arbitrarily increasing the number of moment restrictions, on the other hand, decreases our degrees of freedom in estimating θ, resulting in data limitations. Our goal is to make the best possible use of the finite data sample available to us by letting the data determine the optimal values of b and m.

[8] 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 ∞.

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

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^b}{\operatorname{argmin}} \sum_{j=1}^{m} \left(\sum_{t=1}^{T} \left(\underbrace{\sum_{j=1}^{b} \theta_j \int_{-\infty}^{X_t} B_j(y) d\mathbb{F}_t^Q(y)}_{\hat{g}(\theta)} \right)^j - \frac{1}{j+1} \right)^2, \tag{11}$$

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 the first m moments in estimating the vector θ. It is important to note that the solution to Equation (11) 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 (9) 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), $\frac{d\mathbb{F}_t^Q}{d\mathbb{F}_t^P} = \frac{1}{\hat{g}}$ for all t. So our estimated SDF is given by $e^{-r_t\tau}\frac{1}{\hat{g}(X)}$, 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 denotes returns on the S&P 500 index. This can be re-expressed as

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

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

Since we initially assume that the function g is time-invariant, it follows that M and $\hat{M} = \frac{1}{g}$ are also time-invariant. Since $\hat{M}$ is time-invariant, the SDF will be time-invariant up to the term $e^{-r_t\tau}$. The value of $e^{-r_t\tau}$ is also very stable over our sample period. So the SDF does not vary substantially over our sample under our set of assumptions. 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 5 we will discuss our empirical results based upon estimates of $M(x)$, as described above.

To allow our SDF estimate to vary with time we partially relax the time-invariance assumption by using an instrumental variable in our GMM estimation. Since several other researchers have proposed market volatility as an important conditioning variable, we use the level of the VIX index measured at the same time as our option prices as an instrument

in the estimation of our splines. This allows the spline parameters to vary with the VIX over time, and hence allows us to estimate a conditional SDF. In Section 5 we present our instrumental variable estimate of $M(x)$ as well.

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 (10). 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, in order to get a very tight confidence interval, a large amount of data is needed.

2.6 Model Selection

In order to estimate θ, we use a GMM type estimation to match the resulting estimate to the moments of the uniform distribution over the unit interval as in Equation (11). 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, which determine the SDF. In keeping with this goal, we optimally choose the m and b according to our data and we place no restrictions on θ in our estimation.

Our model selection criterion for determining b and m uses the Cramer-von Mises statistic[9] which is a common nonparametric criterion for determining the goodness of fit of an

[9]We use the Cramer-von Mises statistics as our criterion because it minimizes the mean-squared distance between CDFs as opposed to the Kolmogorov-Smirnov statistic, which minimizes the maximum distance between two CDFs,

$$KS = \sup_{x \in \mathbb{R}} |\hat{\mathbb{F}}(x) - \mathbb{F}_U(x)|.$$

This amounts to choosing the estimate which minimizes the difference over the entire range of values in a

estimated distribution. The Cramer-von Mises statistic compares an estimated distribution to a target distribution (uniform in our case) by comparing the corresponding CDFs, $\hat{\mathbb{F}}$ and $\mathbb{F}_U$ respectively. Here $\hat{\mathbb{F}}$ is the empirical distribution function. A small Cramer-von Mises statistic implies a good fit while larger statistics imply poor fit. The statistic is given by

$$CvM = \int_{x=-\infty}^{\infty} (\hat{\mathbb{F}}(x) - \mathbb{F}_U(x))^2 d\mathbb{F}_U(x).$$

In the case of the uniform distribution over the unit interval, we can express this as

$$CvM = \int_{x=0}^{1} (\hat{\mathbb{F}}(x) - x)^2 dx.$$

While we choose the model based solely on the value of the Cramer-von Mises statistic, this doesn't necessarily tell us how well our optimal model transforms the data to match the uniform distribution. We also calculate the p-values corresponding to the null hypothesis that the estimated distribution is the same as the hypothesized distribution. We calculate p-values base upon 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.

We refer to optimal selection of b and m as model selection, and we will use the optimal model to estimate θ and hence $\hat{g}$ as well as the SDF. In order to optimally select our model, we examine goodness of fit of our estimated CDF with the uniform CDF. Our estimated CDF is given by the empirical CDF corresponding to the estimated vector $\hat{\theta}$ for a given combination of b and m,

$$\hat{\mathbb{F}}_{b,m}(x) = \frac{1}{T} \sum_{t=1}^{T} \mathbb{1} \left(\sum_{j=1}^{b} \hat{\theta}_j \int_{-\infty}^{X_t} B_j(y) d\mathbb{F}_t^Q(y) \leq x \right), \tag{12}$$

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 (12) with the estimated parameter vectors and then compare the Cramer-von Mises statistics for each, keeping in mind that in order for θ to be identified

mean-squared sense, as opposed to choosing the statistic which minimizes the size of the largest error.

requires that $b \leq m$. That is, the number of moment restrictions must be at least as large as the dimension of the vector to be estimated, θ. The smallest Cramer-von Mises statistic corresponds to the model for which the CDI procedure transforms the data to a distribution closest to the uniform distribution. We refer to this as the optimal model.

3 Simulation

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. We extend the example described in Section 2.1 from a static, single period setting to a multiperiod setting with data comparable to that which we observe in the S&P 500 and FTSE 100 data. By simulating data with a known SDF, we can observe how accurately each is able to estimate the true SDF. Our simulated data assumes underlying index returns are distributed log-normally as is the case in the Black-Scholes world, but note that the CDI method is more general and does not make this assumption. We choose parameters of the distribution to fit the data generated by our risk-neutral S&P 500 densities.

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. To be consistent with our data and Assumption 1, we assume that the SDF in the economy is given by the SDF in Panel B of Figure 6. This is the SDF resulting from taking the ratio of the (risk-neutral) log-normal density with location parameter $\mu_Q = 0.00011$ and scale parameter $\sigma_Q = 0.0526$ and the (physical) log-normal density with location parameter $\mu_P = 0.0040$ and scale parameter $\sigma_P = 0.0526$. As described in Section 2.1, these parameters are chosen to match the average of the monthly distributions corresponding to those (annualized) values given in Panel A of Table 1. Notice that we have set $\sigma_Q = \sigma_P$ to be consistent with the Black-Scholes model. As in the Black-Scholes model, the location parameters μ_q and μ_p differ.

The S&P 500 risk-neutral densities described in Table 1 are time varying and it is generally accepted that both σ_P and σ_Q are time varying (but equal to each other such that the pricing kernel is stable). We fit our series of S&P 500 monthly variances described in Table 1 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 2.2, and maximizing the likelihood function to estimate the parameters of the Ornstein-Uhlenbeck process being fit to the series of variances. With the resulting 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 $d\mathbb{F}_t^P$. Since 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 is available to the econometrician.

With the series $\left(d\mathbb{F}_t^Q, X_t\right)$, we estimate true SDF using both the CDI method and the classical method. We show results of both estimation procedures for $N = 200, 500,$ and $1,000$. By comparing these estimates we can see how well each of the methods performs with small data samples. In particular, comparing the two methods allows us to see how estimates can

be affected when comparing forward-looking estimates with backward-looking estimates. We use a 60 period rolling window of realized returns X_t to compute kernel density estimates of the physical densities which are unknown to the econometrician. The results of the simulations for both estimators are shown in Figure 3. Panel A shows that for all values of N, the CDI estimator does a very good job of recovering the true SDF. While the smallest data sample recovers the true SDF fairly well over the range $[.95, 1.05]$, 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.

Panel B shows the results of the simulation 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 $N = 1,000$ simulated months exhibit extreme non-monotonicity 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. Figure 3 shows that the CDI method performs very well while the classical method performs poorly.[10] 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.

[10] Work such as Audrino and Meier (2012) and Beare and Schmidt (2013) improve on the classical approach, and their methods may produce better results.

4 Data

We start with daily S&P 500 and FTSE 100 data from OptionMetrics. For the S&P 500 index options, price midpoints are available from September, 1996 through December, 2012, for a total of 196 months. For the FTSE data, closing prices are available from January 2002 through July 2013. 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 121 total months of data. We use options with one month to maturity, giving a non-overlapping time-series of options prices. This non-overlapping 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 significant loss of information for our analysis 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 static no arbitrage conditions. We wind up using put prices for relatively low strike prices, call prices for relatively high strike prices and weighted averages for intermediate strike prices. We use 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 to determine the relative weights of puts and calls when both prices are observable. Using open interest to calculate the weighted average gives almost exactly the same result, 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 classic 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 website.

5 Results

In this section, we present the results of our estimation described in Section 2, using the data described in Section 4. We compare CDI results with the results obtained by using the classic nonparametric method over the same sample period. We argue that our estimation procedure results in economically plausible SDFs, unlike the classic 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 classic method are the same densities used for the CDI method. This allows us to compare the methods consistently.

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 196 months from September, 1996 through December 2012 for the S&P 500 and the 121 available months from January 2002 to July, 2013 for the FTSE 100. The physical densities described in Table 1 are estimated with a kernel density method using the past 60 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 2.76% with a sample standard deviation of 0.97%. This is remarkably close to the value we obtain when we plug in the mean value for r_t over our sample period, $\bar{r} = 2.64\%$. 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. The mean absolute monthly difference is 0.18% with a standard deviation of 0.17%. This suggests that our estimation procedure does very well in terms of matching the risk-free

rate. This is rather remarkable 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 classic procedure smoothes out any extreme returns and has no 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 12.96% 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 which have annualized standard deviations of 18% for both days. While the physical densities certainly respond to the extreme returns during the financial crisis, 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 which forecast variances (e.g. Rosenberg and Engle (2002)) will likely fail to miss variation in skewness or kurtosis. This highlights a major advantage of the CDI method over existing methods.

5.1 Classic Method Results

We first present the results of estimating the average of a series of estimated SDFs using the classic nonparametric method similar to those of Jackwerth (2000) and Ait-Sahalia and Lo (2000). We should point out that while our classic method estimates are similar to those of other papers, they are not exactly the same as any particular paper. We use monthly data over a longer time span than most other papers, and other papers often have slightly different ways to model the SDF. Nevertheless, our classic method results should be 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 2.2, 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 upon 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 do not seem to vary much 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.

Figures 4 and 5 present estimation results using the classic nonparametric method. The panels of Figures 4 and 5 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. In both figures the four panels look similar over lower returns, while there is some 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. The FTSE 100 estimates appear almost flat for some window lengths, and again bumps appear and disappear as the window length changes. It is surprising how much these classic estimates vary as we change the window length. An estimator that changes our inference about nonmonotonicity as we alter the window length for estimating physical densities does not seem very robust.

The figures include pointwise 95% bootstrapped confidence intervals. Since 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 both Figures 4 and 5, we are able to obtain statistically significant non-monotonicity. We define a non-monotonicity 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 each panel of Figure 4, the lower confidence bound at 1.02 on the returns axis exceeds the upper confidence bound at 0.98. Therefore the estimates exhibit a statistically significant non-monotonicity. As one would expect, using a longer window of returns allows us to identify non-monotonicity at higher confidence levels. In Panel A of Figure 4, the non-monotonicity 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 non-monotonicities are more pronounced and thus are significant at even higher levels of confidence.

5.2 CDI Results

The upward sloping portions of the SDF in Figures 4 and 5 cannot be easily reconciled with standard economic theory of risk averse investors, and similar estimates in the literature have perpetuated the pricing kernel puzzle. The remainder of the paper investigates whether properly accounting for investors' information sets can eliminate the non-monotonicities of estimated SDFs as functions of the index returns.

In order to simultaneously select the optimal model and estimate θ, we evaluate Equation (12) for different numbers of moment restrictions and spline bases and then compare the Cramer-von Mises statistics for each of the 1081 combinations of b and m satisfying $5 \leq b \leq m \leq 50$. The smallest Cramer-von Mises statistic occurs when $b = m = 9$, for both the FTSE and S&P data, with values of 0.00016 (p = 0.976) and 0.00047 (p = 0.836), respectively. This means that the optimal model we choose will solve Equation (11) when using the first nine moment restrictions of the $U[0, 1]$ distribution to estimate the coefficients for a spline with nine bases. The null hypothesis of each associated goodness of fit test is that the estimated distribution comes from the hypothesized distribution of $U[0, 1]$.

We also calculate the Cramer-von Mises statistic and corresponding p-value for our data in the case of no transformation. These statistics indicate the form the results would take if we did not transform the data by estimating a pricing kernel. More specifically, the case

of no transformation means that we take $g(y) \equiv 1$ in Equation (8). So the non-transformed data we use to calculate the Cramer-von Mises statistic is given by the vector V with

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

For the S&P 500 the untransformed data produce a statistic of 0.00057 (p = 0.534), while for the FTSE 100 the statistic is 0.0012 (p = 0.481). These numbers imply that our estimation procedure succeeds in transforming the S&P 500 data to a $U[0, 1]$ sample quite well. We are not able to fit the FTSE data to the uniform distribution quite as well as we can the S&P data. We can also see from the results that even prior to our estimation, the data are not statistically different from $U[0, 1]$ at accepted significance levels. These results should not be considered a formal test comparing the transformed model to the non-transformed data. That being said, our transformation does appear to improve the fit and according to the Cramer-von Mises criterion the fit is very good for the S&P data. For the FTSE data, the fit is not quite as strong but is still good. Figure 6 displays histograms of our data before and after the transformation. Panels A and B clearly show the Cramer-von Mises results for the S&P 500 are confirmed. The transformed S&P 500 data appears very close to a uniform distribution over the unit interval and it does appear more uniform than the non-transformed data. Panels C and D, on the other hand, show that we are not able to fit the uniform distribution of with the FTSE data nearly as well as we can with the S&P data. Furthermore, the histograms in Panels C and D do not visually display the improvement in fit suggested by the Cramer-von Mises statistics. This is simply due to the fact that the histogram with fairly thick bars is not always a good indication of fit. Both the Cramer-von Mises results and Table 2, which we discuss below, show a significant improvement in fit from the non-transformed to the transformed FTSE data. The vertical axis in Figure 6 counts the number of data points falling within each bin as opposed to the density, which is simply a normalization of the count.

We focus on the functional form of the inverse of the function $\hat{g}$ whose estimation is described in Section 2. Below, we plot the estimated functional form of $\hat{M}(x) = \frac{1}{\hat{g}(x)}$ which

we will refer to as the SDF since $e^{-r_t\tau}$ 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 SDF estimated with the CDI approach is a downward sloping function of S&P 500 index realizations. Figure 8 shows the estimated SDF for the FTSE data is downward sloping over the returns ranging from 0.88 to 1.03, but is upward sloping at returns larger than 1.03. However, there are relatively few observed returns larger than 1.05 in the FTSE data set. As a result, our nonparametric estimator is bound to be imprecise at larger values of index returns. The SDF estimates based on the FTSE data look similar to the $N = 200$ estimates in Panel A of Figure 3. This could suggest that the true SDF is actually downward sloping everywhere while our estimate shows non-monotonicity in the right tail only as a result of insufficient data. In order to investigate whether there are indeed non-monotonicities in the SDF, we need to determine whether the non-monotonicity 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 non-parametric estimation, an extremely large set of data is required for one to achieve tight confidence intervals. Since options data does not go back very far, we don't have many extreme observed returns within the time series of realized returns corresponding to the options data. 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 are not very tight in regions that correspond to far out-of-the-money options. This is to be expected as we have only 196 months worth of S&P data and 121 months for the FTSE data.

Our estimate of the time-varying SDF, using the VIX index as in instrumental variable in our GMM estimation, appears in Figure 9. As can be seen from the figure, conditioning on the level of the VIX does not alter our estimated pricing kernel very much.

We note that the estimated SDF based on the S&P 500 data, which has 33% more observations than the FTSE data, is clearly downward sloping and the pointwise confidence

intervals, while wide at certain points, 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 have found the SDF to be increasing. Our estimated downward sloping M is in agreement with mainstream financial and economic theory that risk averse investors' marginal rates of substitution should be downward sloping as a function of states of the world. While the FTSE 100 SDF appears upward sloping in the region of large positive returns, the 95% confidence intervals show that this non-monotonicity is not statistically significant. Thus, our evidence suggests that avoiding the mixture of forward-looking and historical data is a solution to the pricing kernel puzzle.

Since the CDI method is related to the estimation method of Bliss and Panigirtzoglou (2005), we report results of the Berkowitz test, which is the main test used in Bliss and Panigirtzoglou (2005) to assess parametric estimates of the risk aversion function. The test involves two separate likelihood ratio tests. The first, with a test statistic denoted LR_3 is a joint test of the hypothesis that our observed cumulants, $\int_{-\infty}^{X_t} \hat{g}(y) d\mathbb{F}_t^Q(y)$, $t = 1, 2, ..., T$, are i.i.d. and are uniformly distributed over the interval $[0, 1]$. The LR_3 test statistic is distributed χ_3^2 asymptotically. The second likelihood ratio test, with test statistic LR_1, tests the null hypothesis that our observations are iid. The LR_1 statistic follows a χ_1^2 asymptotic distribution. The two likelihood ratio tests are complementary in that if we reject the joint test based upon LR_3, but we do not reject the test of independence based upon LR_1, then it must be the case that we reject the null hypothesis of a uniform distribution. Rejecting the hypothesis of a uniform distribution after the transformation would mean that we do not have the correct SDF, whose inverse transforms our data to a uniform distribution. The results of the Berkowitz test are given in Table 2. We report the results of the test for both the transformed data as well as the non-transformed data, $\int_{-\infty}^{X_t} d\mathbb{F}_t^Q(y)$, $t = 1, 2, ..., T$.

We can see in Panel A of Table 2, that for the untransformed S&P 500 data, we can reject the joint hypothesis at the 90% confidence level, with a p-value of 0.0732. This result, along with the fact that we cannot reject the test of independence, implies that the non-transformed data cannot be rejected as independent but we can reject the hypothesis of a

uniform distribution. On the other hand, the transformed data has a p-value of 0.8777 for the joint test, confirming the results of the Cramer-von Mises statistics and suggesting that the transformation gives a valid SDF. Panel B of Table 2 shows that the transformation of the FTSE data is not able to match the uniform distribution as well as that of the S&P data. Again the LR_1 statistics for both the non-transformed data and the transformed data are small enough that the the corresponding p-values are 0.9348 and 0.8544 respectively. This means that the data appear to be convincingly independent. However, the LR_3 statistics of 6.4839 and 2.4112 with corresponding p-values of 0.0903 and 0.4916 suggest that we can reject the uniform distribution of the non-transformed FTSE 100 data but we cannot reject the uniform distribution for the transformed data. However, the p-value of 0.4916 corresponding to the LR_3 statistic in Panel B does not suggest that we have a very great fit of the data to the uniform distribution over the unit interval.

6 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 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 time by using the level of the VIX index as an instrumental variable does not affect our estimate very much.

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

Proof of Proposition 2: 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 $(a.s\ \mathbb{Q})$ unique random variable $\frac{d\mathbb{P}}{d\mathbb{Q}}$ such that

$$\mathbb{P}((-\infty, x]) = \mathbb{P}(x) = \int_{-\infty}^{X_t} \frac{d\mathbb{P}}{d\mathbb{Q}}(y) d\mathbb{Q}(y) \quad \forall x \in \mathbb{R}. \tag{13}$$

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

$$\mathcal{G}_t(X_t) := \int_{-\infty}^{X_t} g(y) d\mathbb{Q}(y), \tag{14}$$

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

Next we establish uniqueness. Since we can only show almost sure $(\mathbb{Q})$ uniqueness, we reduce the space in question by removing all $\mathbb{Q}-$null sets. Call this reduced space over the real line $\mathbb{R}'$. Since g is non-negative, the function $\mathcal{G}$ uniquely determines where g must be zero over $\mathcal{B}(\mathbb{R}')$. So any functions satisfying the criteria of the proposition must take the value zero over the exact same subsets of $\mathcal{B}(\mathbb{R}')$. Now it only remains to show that over the sets where $g \neq 0$, the functional form is unique. Let $\mathcal{N}$ denote the set in $\mathcal{B}(\mathbb{R}')$ where $g > 0$. Over this set, the function $\mathcal{G}$ is invertible because $g > 0$.

Suppose there is another function g' satisfying Equation (14) over $\mathcal{N}$. Define $\mathcal{G}'_t$ as

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

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

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

and

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

Since P and Q are equivalent by assumption, and $\mathcal{N}$ does not contain any $\mathbb{Q}$-null sets, it follows that $\mathcal{N}$ does not contain any $\mathbb{P}$-null sets. This implies that $\mathbb{P}(X \leq \cdot)$ is a strictly increasing function and hence

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

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

$$\int_E g'(y) d\mathbb{Q}(y) = \int_E g(y) d\mathbb{Q}(y) \quad \forall E \subset \mathcal{B}(\mathcal{N}). \tag{15}$$

Now we can apply the Radon-Nikodym Theorem on $(\mathcal{N}, \mathcal{B}(\mathcal{N}), \mathbb{Q})$. From Equation (15), the Radon-Nikodym Theorem implies $g' = g$ $a.s.$ $\mathbb{Q}$ on $\mathcal{N}$. Since the values of g and g' must be zero on non-null subsets of $\mathcal{N}^c$, we have that $g' = g$ $a.s.$ $\mathbb{Q}$ and hence g is unique ($a.s.$ $\mathbb{Q}$). ■

References

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

Audrino, Francesco, and Pirmin Meier, 2012, Empirical pricing kernel estimation using a functional gradient descent algorithm based on splines, *Working paper*.

Bakshi, Gurdip, and Fousseni Chabi-Yo, 2013, Variance bounds on the permanent and transitory components of stochastic discount factors, *Forthcoming, Journal of Financial Economics*.

Bakshi, Gurdip, Dilip Madan, and George Panayotov, 2010, Returns of claims on the upside and the viability of u-shaped pricing kernels, *Journal of Financial Economics* 97, 130–154.

Balkema, August A, and Laurens De Haan, 1974, Residual life time at great age, *The Annals of Probability* pp. 792–804.

Barone-Adesi, Giovanni, Robert F. Engle, and Loriano Mancini, 2008, A garch option pricing model with filtered historical simulation, *Review of Financial Studies* 21, 1223–1258.

Barone-Adesi, Giovanni, Loriano Mancini, and Hersh Shefrin, 2013, A tale of two investors: Estimating optimism and overconfidence, *Working paper*.

Beare, Brendan K., and Lawrence Schmidt, 2013, An empirical test of pricing kernel monotonicity, *Working paper*.

Billingsley, Patrick, 2012, *Probability and Measure* (John Wiley & Sons).

Bliss, Robert R, and Nikolaos Panigirtzoglou, 2005, Option-implied risk aversion estimates, *The Journal of Finance* 59, 407–446.

Breeden, Douglas T, and Robert H Litzenberger, 1978, Prices of state-contingent claims implicit in option prices, *Journal of Business* pp. 621–651.

Chabi-Yo, Fousseni, Rene Garcia, and Eric Renault, 2007, State dependence can explain the risk aversion puzzle, *Review of Financial Studies* 21, 973–1011.

Chaudhuri, Ranadeb, and Mark Schroder, 2009, Monotonicity of the stochastic discount factor and expected option returns, *Working paper (SSRN 1344061)*.

Christoffersen, Peter, Steven Heston, and Kris Jacobs, 2013, Capturing option anomalies with a variance-dependent pricing kernel, *Review of Financial Studies* 26, 1962–2006.

Efron, Bradley, and Robert J. Tibshirani, 1993, *An introduction to the bootstrap* . vol. 57 of *Monographs on Statistics and Applied Probability* (Chapman and Hall: New York) 1 edn.

Embrechts, Paul, Claudia Klüppelberg, and Thomas Mikosch, 1997, *Modelling extremal events: for insurance and finance* . vol. 33 of *Stochastic Modelling and Applied Probability* (Springer Verlag).

Figlewski, Stephen, 2008, *Volatility and Time Series Econometrics: Essays in Honor of Robert Engle* . chap. 8. Estimating the Implied Risk Neutral Density for the U.S. Market Portfolio (Oxford University Press: Oxford, UK).

Foster, Dean P, and Daniel B Nelson, 1996, Continuous Record Asymptotics for Rolling Sample Variance Estimators, *Econometrica* 64, 139–74.

Grith, Maria, Wolfgang K. Härdle, and Volker Krätschmer, 2013, Reference dependent preferences and the epk puzzle, *Working paper*.

Hansen, Lars Peter, and Eric Renault, 2010, *Pricing Kernels* (John Wiley & Sons, Ltd).

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

Hayashi, Fumio, 2000, *Econometrics* . vol. 1 (Princeton University Press).

Izbicki, Rafael, Ann B. Lee, and Chad M. Schafer, 2014, High-dimensional density ratio estimation with extensions to approximate likelihood computation, *Proceedings of the 17th International Con- ference on Arti cial Intelligence and Statistics* 33.

Jackwerth, Jens Carsten, 2000, Recovering risk aversion from option prices and realized returns, *Review of Financial Studies* 13, 433–451.

Pickands III, James, 1975, Statistical inference using extreme order statistics, *the Annals of Statistics* 3, 119–131.

Polkovnichenko, Valery, and Feng Zhao, 2013, Probability weighting functions implied in options prices, *Journal of Financial Economics* 107, 580–609.

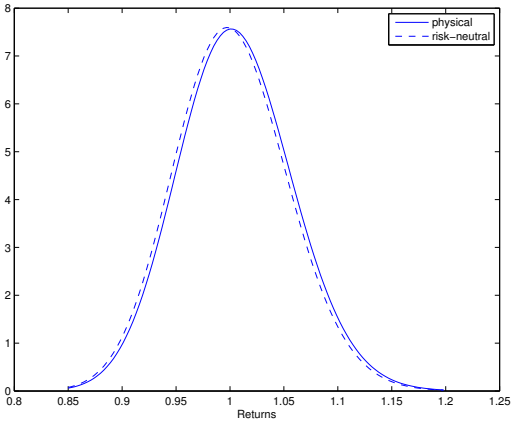
Rosenberg, Joshua V, and Robert F Engle, 2002, Empirical pricing kernels, *Journal of Financial Economics* 64, 341–372.

Song, Zhaogang, and Dacheng Xiu, 2014, A tale of two option markets: Pricing kernels and volatility risk, *Working Paper*.

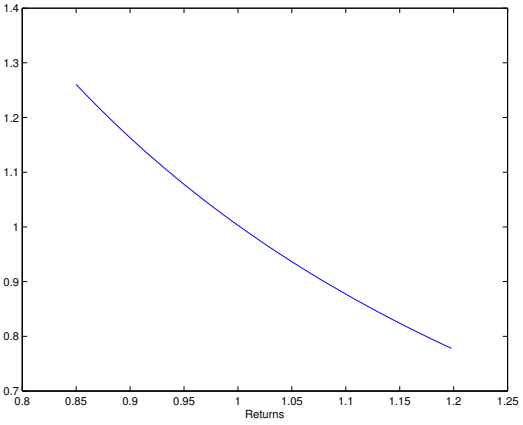
Sugiyama, Masashi, Taiji Suzuki, Shinichi Nakajima, Hisashi Kashima, Paul von Bünau, and Motoaki Kawanabe, 2008, Direct importance estimation for covariate shift adaptation, *Ann Inst Stat Math* 60, 699–746.

Sugiyama, Masashi, Ichiro Takeuchi, Taiji Suzuki, Takafumi Kanamori, Hirotaka Hachiya, and Daisuke 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* 9, 781–788.

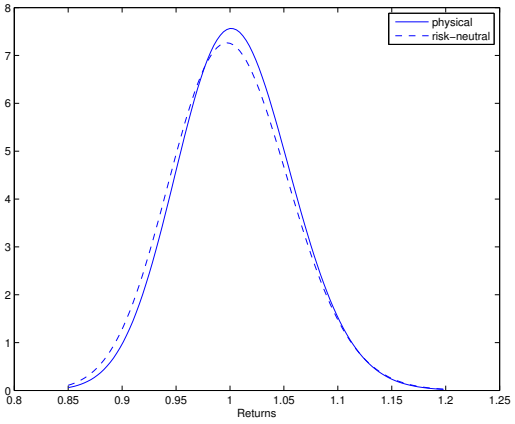
Ziegler, Alexandre, 2007, Why does implied risk aversion smile?, *Review of Financial Studies* 20, 859–904.



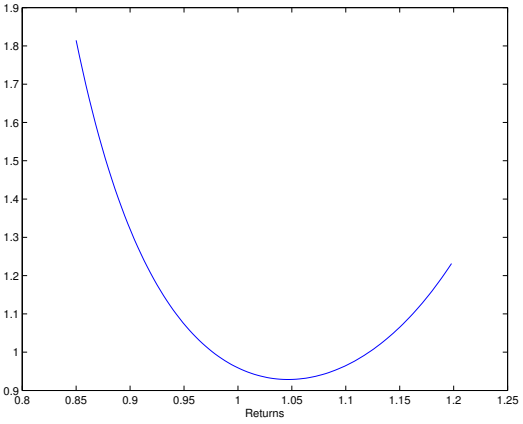
A) log-normal densities $\sigma_P = \sigma_Q$



B) SDF corresponding to panel A



C) log-normal densities $\sigma_P < \sigma_Q$



D) SDF corresponding to panel C

Figure 1: Black-Scholes-implied densities

Panel A plots the log-normal risk-neutral (dashed) and physical (solid) densities that arise under the Black-Scholes model. We choose location parameters to match those of our samples for monthly returns. The physical location parameter is thus set to $\mu_P = 0.0040$ and the risk-neutral location parameter is set equal to $\mu_Q = 0.00011$. Under the Black-Scholes model, both distributions have the same scale parameter, σ, so we set these both equal to the scale parameter for our sample of (monthly physical) returns, $\sigma_P = \sigma_Q = 0.0526$. Panel B plots the SDF corresponding to the densities in Panel A. In Panel C we slightly increase σ_Q to $\sigma_Q = 0.055$, and in Panel D we plot the corresponding SDF.

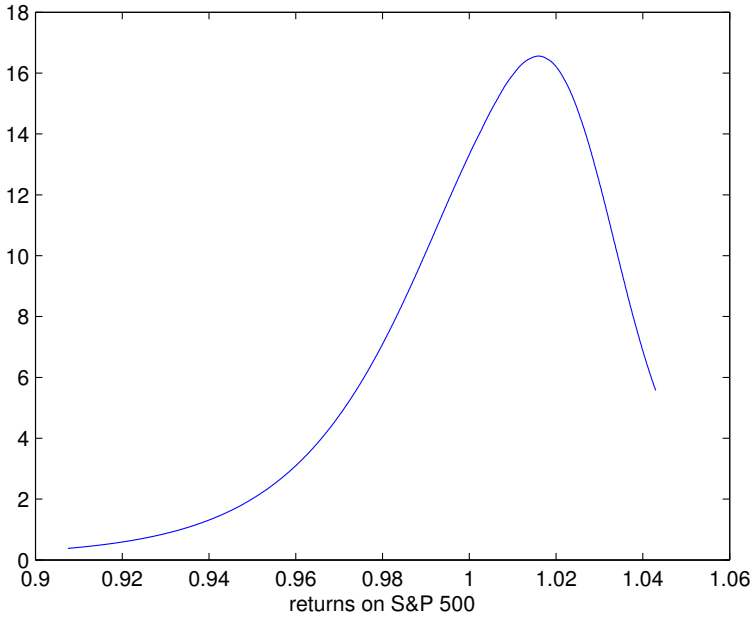
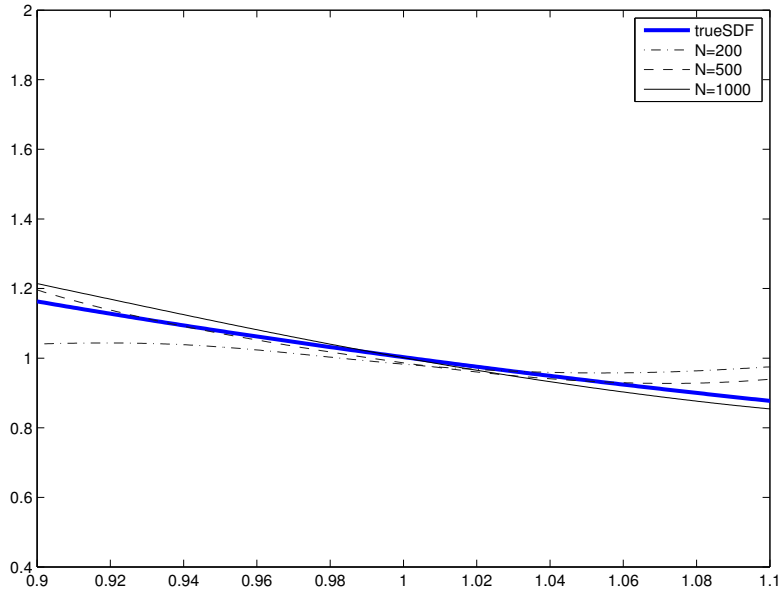
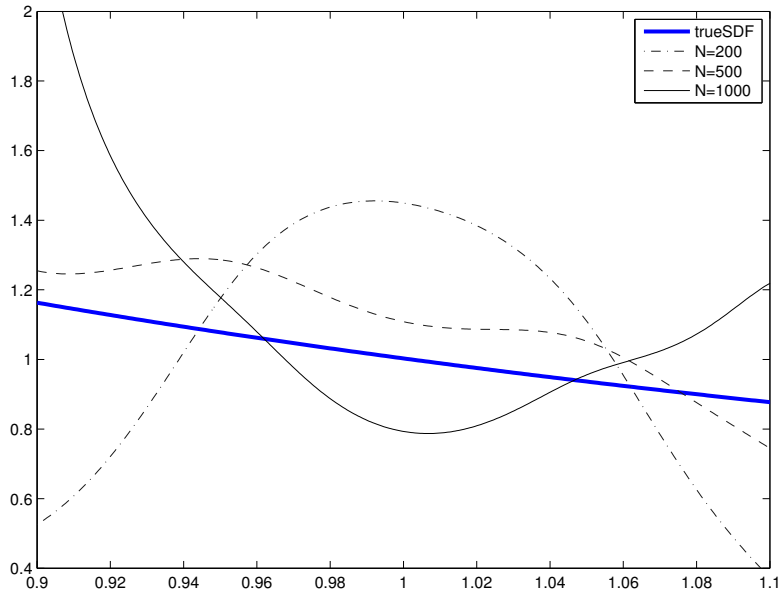


Figure 2: Example risk-neutral density

This risk-neutral density was estimated using option prices from April 20, 2006 with best bids exceeding \$3/8. For April 20, 2006, there are 43 valid option prices which we use, corresponding to 37 unique strike prices. Each month we use option prices to estimate a risk-neutral density like this one. We estimate the tails of the distribution by matching a generalized Pareto Distribution to the slope of the density very close to where we can no longer estimate it. The method for estimating the risk-neutral densities is described in detail in Section 2.2.

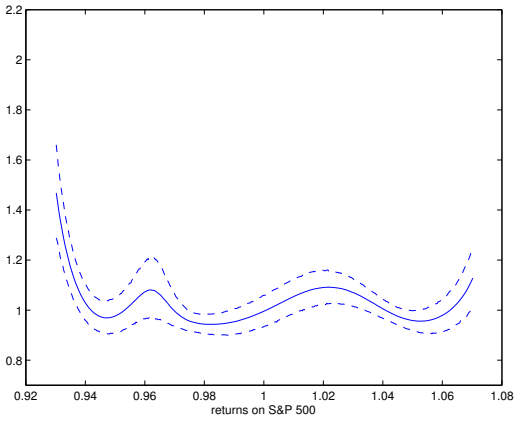


A)CDI estimates

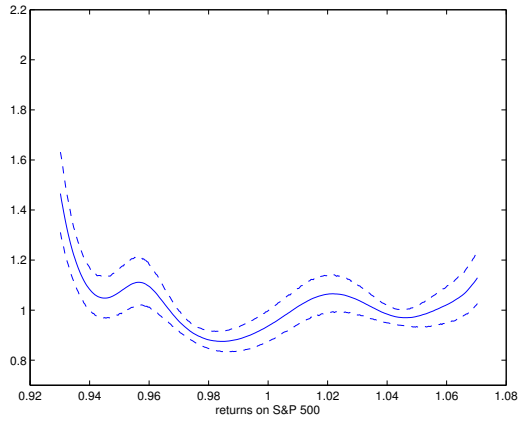


B) Classical method estimates

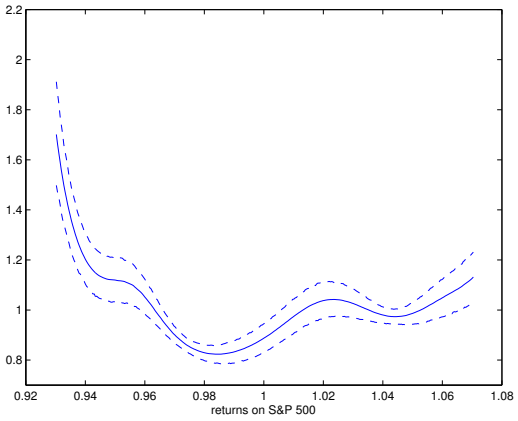
Figure 3: Estimated and true SDFs from simulations
The side by side plots compare the performance of the CDI method and classical method of non-parametric estimates of the SDF. Our simulated data is generated using the true SDF depicted by the bold line in each panel. The estimates of each method are depicted with the true SDF.



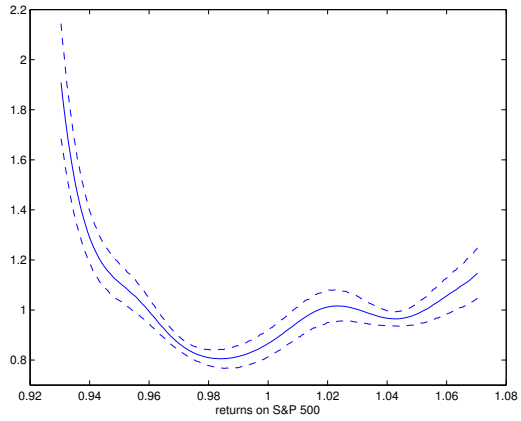
A) 4 year rolling window



B) 6 year rolling window

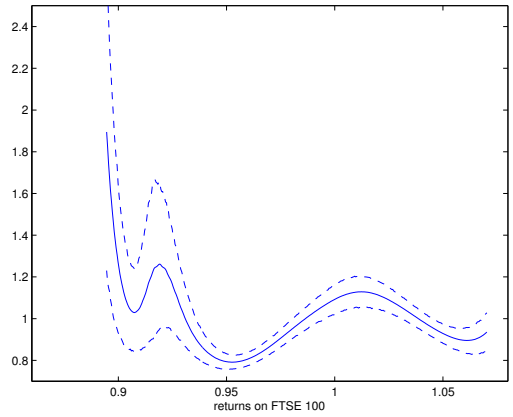


C) 8 year rolling window

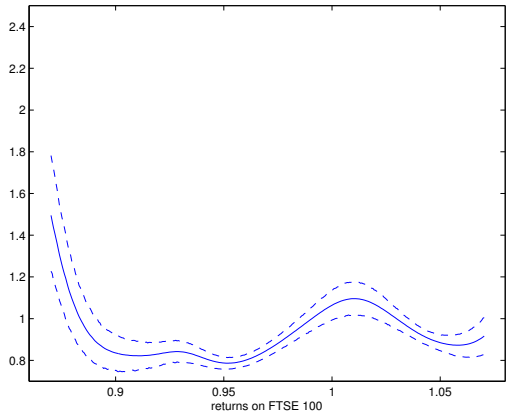


D) 10 year rolling window

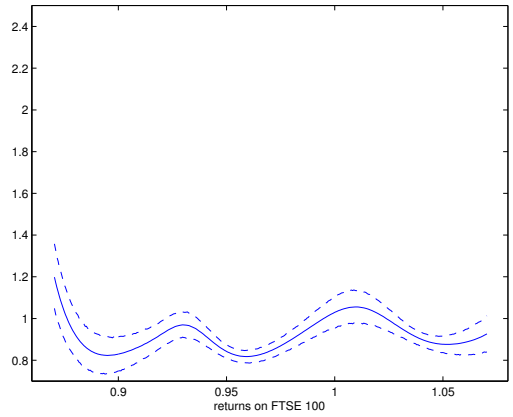
Figure 4: Estimated SDFs using classic procedure: S&P 500
Version of the classic nonparametric estimates of the stochastic discount factor as the average of monthly SDF estimates with pointwise bootstrap 95% confidence intervals. Each monthly SDF is the ratio of a risk-neutral density to a physical density estimate of returns on the S&P 500 index. Each panel represents the resulting estimate when a different widow is used to estimate the physical density using a Gaussian kernel estimator.



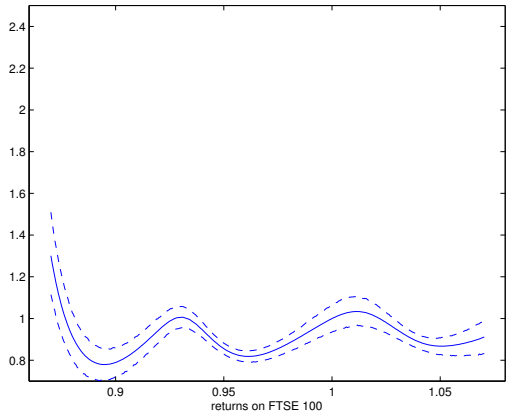
A) 4 year rolling window



B) 6 year rolling window

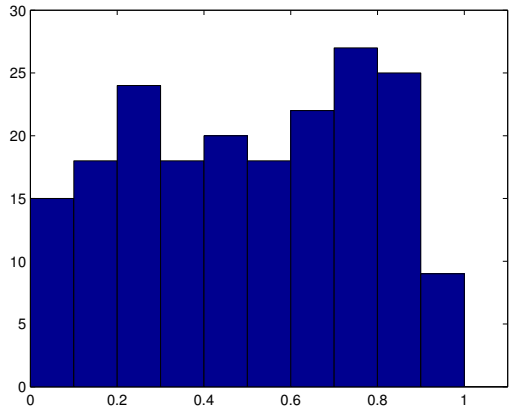


C) 8 year rolling window

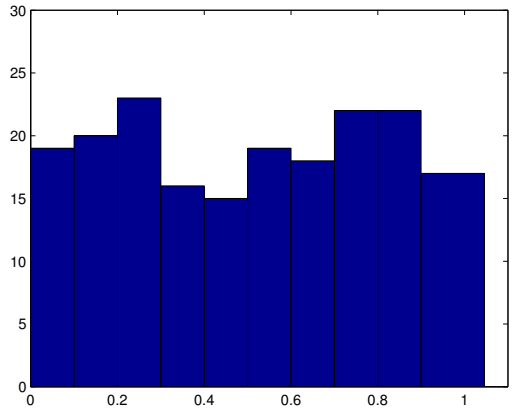


D) 10 year rolling window

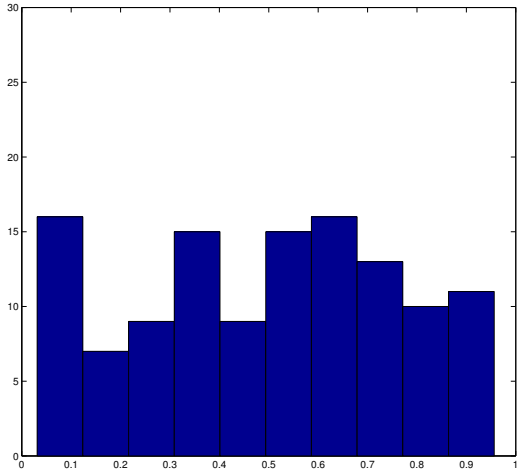
Figure 5: Estimated SDFs using classic procedure: FTSE 100
Version of the classic nonparametric estimates of the stochastic discount factor as the average of monthly SDF estimates with pointwise bootstrap 95% confidence intervals. Each monthly SDF is the ratio of a risk-neutral density to a physical density estimate of returns on the FTSE 100 index. Each panel represents the resulting estimate when a different widow is used to estimate the physical density using a Gaussian kernel estimator.



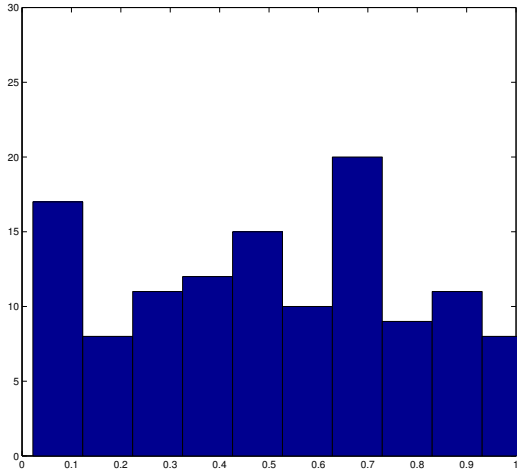
A) No pricing kernel S&P 500



B) Optimal pricing kernel S&P 500 ($b = m = 9$)



C) No pricing kernel FTSE 100



D) Optimal pricing kernel FTSE 100 ($b = m = 9$)

Figure 6: Histograms of cumulants with and without a pricing kernel
The histograms plotted in Panels A and C are estimates of the density of the cumulants that result from integrating risk-neutral densities up to their corresponding realized values of the S&P 500 and FTSE 100 data respectively, or $\int_{-\infty}^{X_t} d\mathbb{F}_t^Q(y), \;\; t = 1, 2, ..., T$. If the pricing kernel is constant (or there is no compensation for risk) then we would expect this histogram to be close to a uniform [0,1] density. The histograms in Panels B and D are estimates of the density of corresponding cumulants resulting from our CDI estimation method. Specifically, it is a histogram of $\int_{-\infty}^{X_t} \hat{g}(y) d\mathbb{F}_t^Q(y), \quad t = 1, 2, ..., T$, where $\hat{g}(y)$ is the CDI estimate of the inverse of the pricing kernel. The fact that the histogram in Panel B appears to be approximately uniformly [0,1] distributed shows that the CDI pricing kernel fits the S&P 500 data very well. The histogram in Panel D shows that the CDI pricing kernel fits the FTSE 100 data only moderately well.

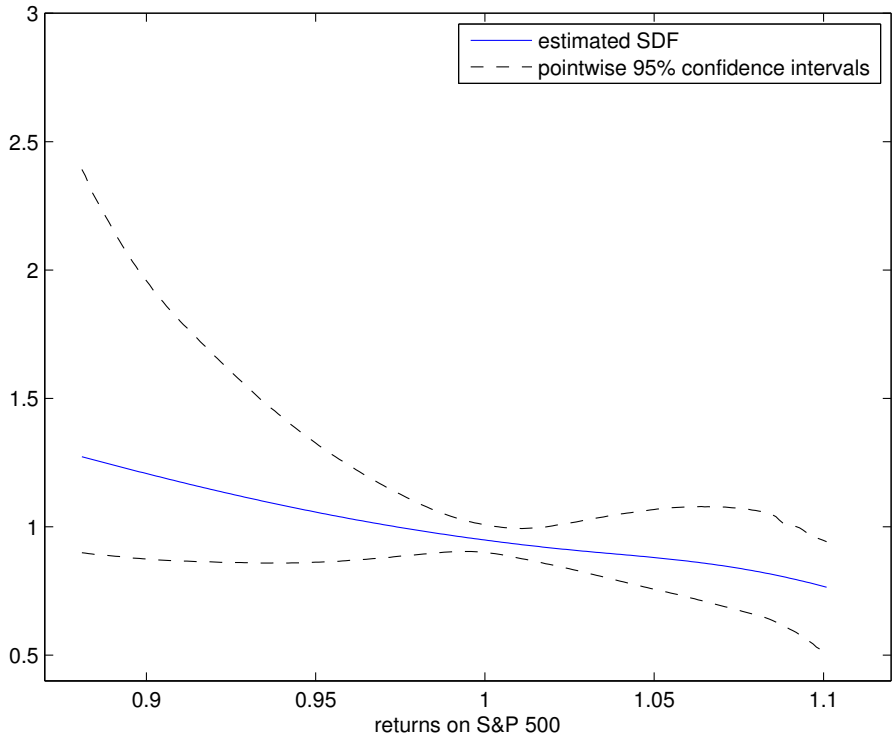


Figure 7: Estimated stochastic discount factor using CDI method: S&P 500
The result of our CDI estimation of the pricing kernel for the S&P 500 is plotted above. It is clearly monontonically decreasing on the interval over which we can estimate it with some precision. The CDI method estimates the pricing kernel by matching the moments of the distribution of the cumulants,

$$\int_{-\infty}^{X_t} \hat{g}(y) d\mathbb{F}_t^Q(y), \quad t = 1, 2, ..., T,$$

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. 95% 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.

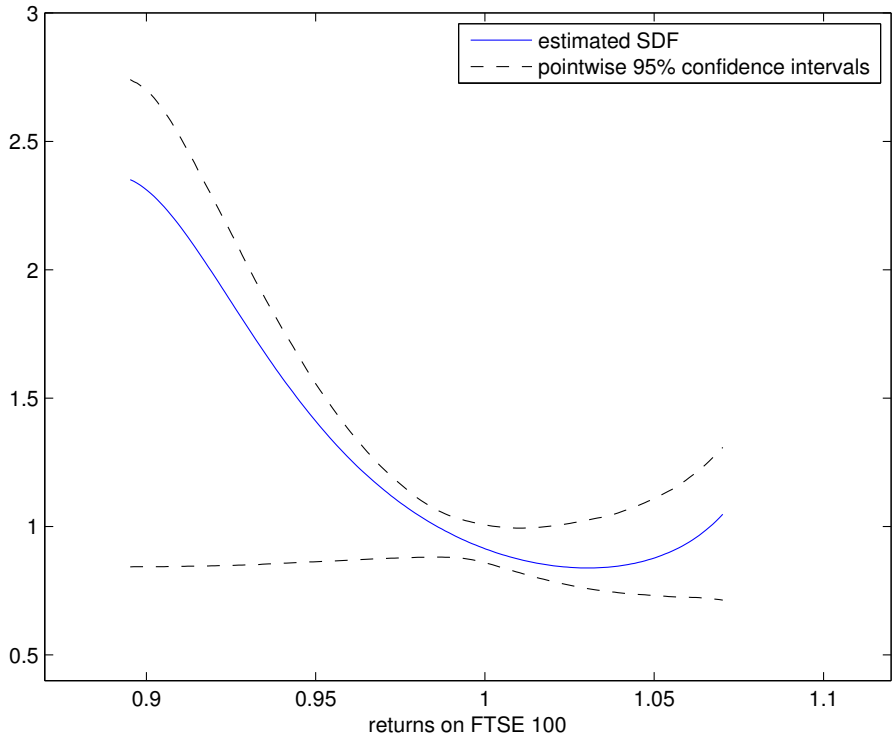


Figure 8: Estimated stochastic discount factor using CDI method: FTSE 100
The result of our CDI estimation of the pricing kernel for the FTSE 100 is plotted above. The estimate exhibits some non-monotonicity at the end of the interval over which we can estimate it with some precision. The non-monotonicity 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,

$$\int_{-\infty}^{X_t} \hat{g}(y) d\mathbb{F}_t^Q(y), \quad t = 1, 2, ..., T,$$

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. 95% 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.

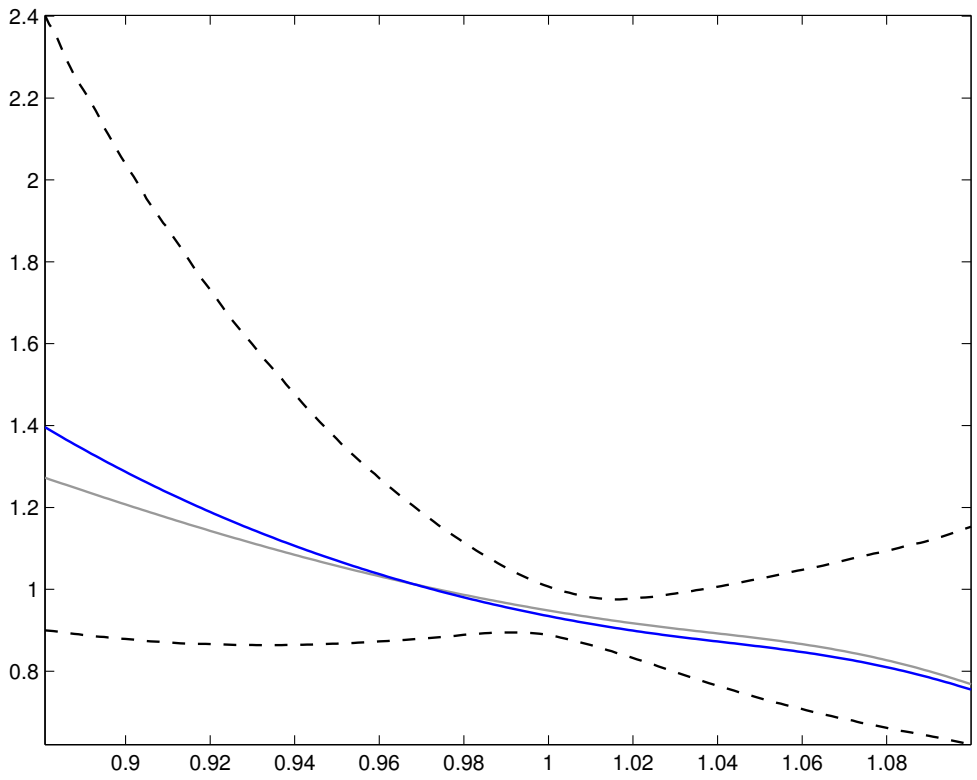


Figure 9: Time-varying stochastic discount factor using CDI method: S&P 500
The result of our CDI estimation of the time-varying pricing kernel for the S&P 500 is plotted above. The function is estimated using the level of the VIX index as an instrument in the GMM estimation procedure. The level of the VIX index is observed at the same time as the option prices which are used to estimate risk-neutral densities. The dark blue line represents the conditional estimate while the lighter gray line displays the unconditional estimate. 95% 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 1: Summary statistics for risk-neutral densities

For each of the months in our sample (196 months: from September 1996 through December 2012 for S&P 500 data. 121 months: from January 2002 to December 2012 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 60 months of past returns, and the risk-neutral densities are estimated as described in Section 2.2. 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.

Panel A: S&P 500

	Risk-Neutral Densities from Options Prices			
	Annualized Mean Ret	Annualized Standard Dev	Monthly Skewness	Monthly Kurtosis
Sample average	2.76%	22.98%	-1.1814	6.2283
Sample standard deviation	0.97%	24.00%	0.4844	2.1343
	Physical Densities from 60 months of Historical Data			
	Annualized Mean Ret	Annualized Standard Dev	Monthly Skewness	Monthly Kurtosis
Sample average	6.48%	18.33%	-0.4661	4.1178
Sample standard deviation	2.29%	11.49%	0.3537	1.4569

Panel B: FTSE 100

	Risk-Neutral Densities from Options Prices			
	Annualized Mean Ret	Annualized Standard Dev	Monthly Skewness	Monthly Kurtosis
Sample average	3.57%	21.43 %	-1.0365	7.9168
Sample standard deviation	4.01%	9.00%	0.6857	4.1581
	Physical Densities from 60 months of Historical Data			
	Annualized Mean Ret	Annualized Standard Dev	Monthly Skewness	Monthly Kurtosis
Sample average	3.45%	16.71 %	-0.5273	3.4905
Sample standard deviation	3.08%	1.10%	0.1047	0.2389

Table 2: Berkowitz statistics and p-values

The first line reports likelihood ratio test statistics and corresponding p-values for Berkowitz tests of the transformed data using the optimal model, $b = m = 9$. The second line reports likelihood ratio test statistics and corresponding p-values for Berkowitz tests of the non-transformed data, or the data without a pricing kernel. The LR_3 statistic tests the joint hypothesis that data is iid and $U[0, 1]$. The LR_1 statistic tests the hypothesis that the data are independent. Rejection based upon the LR_3 statistic can come from the data not being independent or the data not being uniformly distributed. If we reject base upon the LR_3 statistic but fail to reject based upon the LR_1 statistic, this implies that the data does a poor job fitting the $U[0, 1]$ distribution.

Panel A: S&P 500

Model	LR_3	p-value	LR_1	p-value
Optimal model ($b = m = 9$)	0.6808	0.8777	0.0054	0.9412
No pricing kernel	6.9593	0.0732	0.0447	0.8326

Panel B: FTSE 100

Model	LR_3	p-value	LR_1	p-value
Optimal model ($b = m = 9$)	2.4112	0.4916	0.0337	0.8544
No pricing kernel	6.4839	0.0903	0.0067	0.9348