Journal of Financial Economics 107 (2013) 580–609

Contents lists available at SciVerse ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Probability weighting functions implied in options prices ☆

Valery Polkovnichenko [a], Feng Zhao [b,*]

[a] Federal Reserve Board, Division of Research and Statistics, Capital Markets MS-89, 20th & C street, Washington, DC 20551, United States
[b] Naveen Jindal School of Management, University of Texas at Dallas, SM31, P.O.Box 830699, Richardson, TX 75083-0699, United States

ARTICLE INFO

Article history:
Received 17 March 2010
Received in revised form
24 January 2012
Accepted 30 January 2012
Available online 16 October 2012

Jel Classification:
G12

Keywords:
Pricing kernel
Nonparametric estimation
Probability weighting
Rank-dependent utility

ABSTRACT

The empirical pricing kernels estimated from index options are non-monotone (Rosenberg and Engle, 2002; Bakshi, Madan, and Panayotov, 2010) and the corresponding risk-aversion functions can be negative (Aït-Sahalia and Lo, 2000; Jackwerth, 2000). We show theoretically that these and several other properties of empirical pricing kernels are consistent with rank-dependent utility model with probability weighting function, which overweights tail events. We also estimate the pricing kernels nonparametrically from the Standard & Poor's 500 index options and construct empirical probability weighting functions. The estimated probability weights typically have the inverse-S shape, which overweights tail events and is widely supported by the experimental decision theory.

1. Introduction

Pricing kernels estimated from index options are non-monotone (Rosenberg and Engle, 2002; Bakshi, Madan, and Panayotov, 2010) and the corresponding risk-aversion functions can be negative (Aït-Sahalia and Lo, 2000; Jackwerth, 2000). These empirical facts are at odds with the standard pricing kernel, which is monotonically decreasing in investor wealth and has a positive risk-aversion function. In this paper we consider a pricing kernel based on the rank-dependent expected utility (RDEU) model with a probability weighting function. We show that this model is consistent with several features of the empirical pricing kernel estimated from index options and that the data imply the shape of probability weights with the emphasis on tail events. This property is widely supported by experimental research, and our results confirm it outside of laboratory settings. An important advantage of our empirical approach is that it is nonparametric, and we do not impose any a priori restrictions on the shape of the weighting function.

Economists have long recognized that decisions under risk are more sensitive to changes in probability of events that lie at the extremes of possible outcomes compared with the probabilities of typical events in the middle. The expected utility (EU) model might not be able to generate this behavior because it assumes that outcomes are weighted linearly by probability. Several alternative

☆ We are grateful to our referee Gurdip Bakshi for many detailed comments and suggestions which greatly improved the paper. We also would like to thank Fousseni Chabi-Yo, Jianqing Fan, Jens Jackwerth, George Jiang, George Korniotis, Chris Lamourex, Haitao Li, Mike Rebello, Leonidas Rompolis, Zhaogang Song, John Wald, Yexiao Xu, Harold Zhang and seminar participants at the University of Arizona, Georgia State University, Indiana University, University of Buffalo, University of Texas at Dallas, University of Texas at San Antonio, 2010 Western Finance Association meetings, 2010 Miami Behavioral Finance conference, and 2011 International Monetary Fund/George Washington University Behavioral Finance Conference for comments. We are grateful to Kerry Hennigin for editorial assistance. The views presented in this paper are solely ours and do not necessarily represent those of the Federal Reserve Board or its staff.

* Corresponding author.
E-mail addresses: valery.y.polkovnichenko@frb.gov (V. Polkovnichenko), feng.zhao@utdallas.edu (F. Zhao).

0304-405X/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jfineco.2012.09.008

models accommodate nonlinear response to changes in probabilities by introducing probability weighting function, which transforms cumulative distribution of ranked outcomes into decision weights used to compute weighted average utility. Two models that include this mechanism are the rank-dependent expected utility (Quiggin, 1982; Yaari, 1987) and the cumulative prospect theory (CPT; Tversky and Kahneman, 1992). A substantial body of experimental research supports the inverse S-shaped probability weighting, which overweights the events in the tails of the outcomes distribution.[1] Our results complement this literature and show that similar features are present in the probability weighting functions estimated from option prices.

In the theoretical part of the paper, we introduce utility with probability weights and derive the corresponding pricing kernel. We then investigate how the probability weighting function affects the moments of the risk neutral distribution and the properties of the pricing kernel. We derive necessary conditions for the probability weighting function to generate nonmonotonicity in the pricing kernel and negative expected return on out-of-the-money (OTM) calls observed in the data. For concave utility functions, these conditions imply convexity in the probability weighting in a range of higher strike prices. This suggests that the empirical facts are consistent with a weighting function that over-weights the right tail of the distribution, such as the inverse S-shaped weights. We also derive general approximation formulas for the effects of probability weights on the moments of the risk-neutral distribution. In some special cases, we can derive comparative statics for these moments with respect to the probability weighting function parameters.

In the empirical part of the paper, we estimate the pricing kernel from index options and use it to construct the implied probability weighting function. We follow a nonparametric approach to estimate the risk-neutral density from option prices (Breeden and Litzenberger, 1978; Aït-Sahalia and Lo, 1998, 2000). The existence of the stochastic discount factor (SDF) relies only on the principle of no-arbitrage and does not require specific utility assumptions. The advantage of the nonparametric method is that it does not impose any constraints on the option pricing model or the functional form of the pricing kernel.

To estimate risk-neutral distribution, we use the constrained local polynomial method from Aït-Sahalia and Duarte (2003), together with the semi-nonparametric method based on Gram-Charlier series expansion. This approach enjoys several advantages over other parametric and nonparametric methods. First, the estimator is nonparametric so the derived shapes for the probability weighting functions do not suffer from the misspecification problems of parametric models.[2] Second, the local polynomial estimators have superior performance to the traditional Nadaraya-Watson kernel regression estimators for derivatives estimation. Furthermore, the constrained estimator has better finite sample performance than the unconstrained one. Third, we use the semi-nonparametric estimates for guidance in choosing the optimal bandwidth for our nonparametric estimator in a finite sample setting. Finally, the semi-nonparametric estimates can be used in the robustness check to ensures our findings are not dependent upon a particular estimation method.

The pricing kernel for RDEU is a product of marginal utility and the derivative of the probability weighting function. Because we estimate SDF nonparametrically, both components cannot be identified nonparametrically. To estimate probability weighting function without imposing any parametric assumptions we use several standard utility functions. The majority of the estimated probability weighting functions have the inverse S-shape, which overweights the tails. The shapes of the weighting functions are robust to the assumptions about utility functions. We also use nonparametric estimates of the probability weights to fit a parametric specification of the weighting function. This approach allows us to construct time series of the parameters characterizing the shape of the probability weighting function. While the dominant shape of the weighting function is the inverse S, we also find that during some periods it changes to the S shape and underweights the tails. We believe that information in the time series of probability weighting parameters can be used to construct investor sentiment toward tail events. We leave it to future work to investigate how the probability weighting function could be linked to expected stock returns and volatility.

Our results have implications beyond explaining the shape of the empirical pricing kernel. Interest is growing in applications of utility functions with probability weighting in finance.[3] While the experimental support for the inverse S-shape weighting is significant, a typical concern is that the patterns detected in a controlled lab environment might not apply to real-life decisions. We extend empirical support for probability weighting

[1] For example, Camerer and Ho (1994) summarize evidence from several studies supporting nonlinearity of probability weights with higher sensitivity of preferences to tail events. Wu and Gonzalez (1996) identify the inverse S-shape probability weights in experimental data using a procedure that does not require assumptions about the functional form of utility or probability weighting functions. Many other authors used hypothetical and real payoff experiments to obtain insights about the probability weighting functions and concluded that the inverse S-shape fits the data well. See, for example, Quiggin (1987), Lopes (1987), Tversky and Kahneman (1992) and Prelec (1998). Berns, Capra, Moore, and Noussair (2007) present evidence from nonmonetary experiments. The surveys of the related theoretical and empirical literatures on choice under risk are provided by Shoemaker (1982), Camerer (1995), and Starmer (2000).

[2] One such problem, particularly relevant in the present context, is that the tails of the estimated distributions are affected by the prices of at-the-money options through parametric assumptions, even though these prices contain little information about the tails. Nonparametric estimation instead relies on the prices of out-of-the money options to estimate the tails of the distribution.

[3] For example, Epstein and Zin (1990) use rank-dependent utility to investigate equity risk premium. Shefrin and Statman (2000) and Polkovnichenko (2005) consider implications of probability weighting for portfolio choice and diversification. Barberis and Huang (2008) study the cross-sectional risk premia using cumulative prospect theory. Other recent examples include Levy and Levy (2004), Barberis, Huang, and Thaler (2006), and Chapman and Polkovnichenko (2009).

to nonexperimental data and provide independent validation of the assumptions about investor behavior used in this literature.

This paper is related to the theoretical research investigating the shape of the pricing kernel and the risk-aversion function. This literature explored preferences heterogeneity, time variation in beliefs, and state-dependent utility parameters. Shefrin (2001, 2005) proposes a behavioral model with heterogeneous sentiment, which generates nonmonotonicity in the pricing kernel. Ziegler (2007) considers a model with two agents and heterogeneous beliefs and concludes that heterogeneity in beliefs must be implausibly large to explain the empirical properties of the risk aversion function. Brown and Jackwerth (2004) and Chabi-Yo, Garcia, and Renault (2008) explore the models with state-dependent preferences that can generate nonmonotonic pricing kernels. Bakshi, Madan, and Panayotov (2010) show that nonmonotonic pricing kernels imply negative expected returns. We estimate probability weighting functions with monthly frequency and find the inverse-S shape weights in most cases. We would expect that utility parameters do not change considerably during such relatively short periods of time. Thus, our results are consistent with the probability weighting function being an element of the utility function instead of an implication of time aggregation of standard preferences across periods with different parameters. Empirical papers investigating probability weighting functions include Kliger and Levy (2009) and Dierkes (2009). In both of these papers, the authors assume the inverse S-shape weighting function and estimate its parameters to show that it provides a better fit than the linear weighting. Our approach does not require a priori assumptions about parametric form, or even shape, of the weighting function or the underlying option pricing model, and it is superior to parametric methods used in these papers. Parametric estimation allows to compare functions only within a chosen parametric family, and it is inconclusive about whether a given function is the best overall or only some of its features contribute to a better fit.

The rest of the paper is organized as follows. Section 2 reviews theoretical foundations for the pricing kernel with probability weighting function. In Section 3, we describe the empirical method and data used to estimate state-price density. Section 4.1 discusses the results, and Section 5 concludes.

2. Some properties of the pricing kernel and the risk neutral distribution with probability weighting function

In this section, we introduce the stochastic discount factor with a probability weighting function based on the rank-dependent expected utility (Quiggin, 1982; Yaari, 1987). The key difference between RDEU and EU is that the utility is weighted by a probability weighting function assigned to ranked outcomes. RDEU nests EU but also allows modeling of risk attitude toward probabilities and not just the levels of wealth or consumption. The analytical convenience of RDEU allows us to analyze the properties of the SDF and estimate empirical weighting functions from the state-price densities implied in index options. We also extend some of the results in Bakshi, Kapadia, and Madan (2003) to incorporate the effect of probability transformation on the moments of the risk-neutral distribution.

2.1. Basics of the utility function with probability weights

We specialize in continuous-state case RDEU here, see Quiggin (1993) for detailed exposition of discrete case and derivation of the main properties of this utility function. The RDEU is defined over outcomes that are ranked from the worst to the best. In our case, it is natural to rank possible outcomes by investor wealth w. We assume w to be a random variable with cumulative distribution function (CDF) $P(w)$ and with the density $p(w)=P'(w)$. A probability weighting function $G(P)$ is defined as a continuous, nondecreasing function $G(\cdot):[0,1]\to[0,1]$, such that $G(0)=0$ and $G(1)=1$. For convenience, we also assume that $G(\cdot)$ is differentiable. The purpose of G is to transform original probabilities into decision weights that are used to compute weighted average utility value. The continuous-state version of RDEU is given by

$$U=\int u(w)\,dG(P). \tag{1}$$

By denoting the probability weighting density as $Z(P)\equiv G'(P)\geq 0$, we can rewrite the utility function as

$$U(w)=\int u(w)G'(P)\,dP=E\{u(w)Z(P)\}. \tag{2}$$

Outcomes with $Z>(<)1$ are weighted more (less) than their objective probability. As a special case with $G(P)=P$ $(Z=1)$, RDEU nests the standard EU. Also, because the decision weights integrate to one we have $EZ\equiv\int dG(P)=1$.

While the weighting function is a transformation of the original probability measure P into $G(P)$, the decision maker is assumed to know the underlying distribution P. Intuitively, the probability weighting function is a modeling mechanism for risk attitude toward probabilities of ranked events. It transforms events' probabilities into decision weights in a way that is conceptually similar to the utility function mapping wealth or consumption into utility values. In that sense, probability weights address the criticism put forth by Allais (1988), that risk aversion should be independent of the curvature of the utility in the absence of risk.[4]

Theoretical restrictions on the weighting function are very mild and ultimately its shape has to be determined empirically. Many experimental studies (e.g., Camerer and Ho, 1994; Wu and Gonzalez, 1996; Tversky and Kahneman, 1992) find that individuals typically overweight the events in the tails of the distribution, i.e., for P near zero and one, relative to the events in the middle of the distribution. This type of behavior could be characterized by the inverse S-shaped probability weighting function G with a corresponding U-shaped density Z. However,

[4] See also a related discussion in Quiggin (1993, Section 5.6, p. 68). Also, unlike subjective beliefs, probability weights depend on the actions of the agent through cumulative distribution of ranked outcomes.

as summarized in Camerer and Ho (1994), in some experiments the estimated weighting function is S-shaped, which instead underweights tail events. In our analysis, in addition to the nonparametric estimation, we consider three alternative parametric specifications of the weighting functions that appeared in the literature. They can accommodate a wide variety of shapes including concave, convex, and inverse S-shaped and S-shaped weights. It is important to allow for this flexibility in parametric specifications as it helps to approximate the data more closely. The simplest of the three forms, originally proposed by Lopes (1987), is a weighted average power function:

$$G(P)=\lambda P^{\phi}+(1-\lambda)(1-(1-P)^{\phi}),\quad \phi\geq 1,\quad \lambda\in[0,1]. \tag{3}$$

This weighting function is useful for modeling a variety of risk attitudes. The power parameter ϕ controls overweighting of the probabilities in the tails by changing the curvature of the weighting function.[5] When ϕ is an integer, the weights have particularly simple polynomial form, which allows us to derive semi-analytically the comparative statics effects of the transformation on the moments of a distribution. The weight λ controls the mix of optimism and pessimism by changing the relative strength of the concave and convex segments. At the ends of admissible values, $\lambda=0$ corresponds to globally concave pessimistic function; $\lambda=1$ to the globally convex optimistic. Thus, λ can be intuitively understood as a sentiment parameter, and we show that it can be linked to skewness of the risk-neutral distribution.

Two other weighting functions frequently used in the literature are due to Tversky and Kahneman (1992) and Prelec (1998), respectively:

$$G_1(P)=\frac{P^{\alpha}}{(P^{\alpha}+(1-P)^{\alpha})^{1/\alpha}}$$

and

$$G_2(P)=\exp(-(-\log(P))^{\alpha}). \tag{4}$$

For both specifications, many experimental studies typically find $\alpha<1$ corresponding to the inverse S-shape overweighting probabilities in the tails, but some variation exists across studies with some estimating $\alpha>1$ (underweighting the tails) as noted in Camerer and Ho (1994). While these functions have only a single parameter, which is attractive for experimental work, this is not very convenient for our analysis. Prelec (1998) proposes the following two-parameter version, which we use in our empirical work:

$$G(P)=\exp(-(-\beta\log(P))^{\alpha})=\exp(-(-\log(P^{\beta}))^{\alpha}). \tag{5}$$

As seen from Eq. (5), this weighting function can be represented as a compound function using the one-parameter Prelec specification: $G_2(P^{\beta})$. Using the same principle, we define a two-parameter extension for the Kahneman and Tversky specification as $G_1(P^{\beta})$. The two-parameter extensions allow the weighting functions to accommodate changes in the relative strengths of concave and convex segments. This property appears to be relevant for the analysis of the empirical pricing kernels. We also consider how the parameters α and β affect the moments of the risk-neutral distribution.

2.2. SDF with probability weighting function

The atemporal version of RDEU was incorporated in an intertemporal model of preferences in Epstein and Zin (1989, 1990). Epstein and Zin integrate a general class of atemporal certainty equivalents, including RDEU, in a recursive utility model. They also show that, for a large class of homogeneous certainty equivalents, such as RDEU, and under some additional technical assumptions, one can use the representative agent approach to derive asset pricing implications of such utilities. Here we consider a simplified setting to focus on the properties of SDF with probability weighting functions.

We use a static model with utility function in Eq. (1) over terminal wealth w. Assume that markets are complete, the consumer has initial wealth w_0, and he has access to $N+1$ traded securities. Denote portfolio share of investment in security $k=0,1,\ldots,N$ as θ^k and return as R^k. The portfolio constraint implies $\theta^0=1-\sum_{k=1}^{N}\theta^k$ and the terminal wealth is given by

$$w=w_0\left(\sum_{k=0}^{N}R^k\theta^k\right). \tag{6}$$

To obtain the necessary first-order optimality conditions, we rely on the results in Ai (2005) and Carlier and Dana (2003), which show differentiability of the RDEU functional with respect to continuously distributed random variables. Ai (2005) shows that when X is continuously distributed (has a density), the derivative of the above RDEU functional with respect to a continuously distributed random variable Y is given by

$$\frac{\partial}{\partial\alpha}U(X+\alpha Y)\Big|_{\alpha=0}=E\{u'(X)Z(P_X)Y\}. \tag{7}$$

Using Eq. (7), the first order optimality condition with respect to θ^k is given by

$$E\{u'(w)Z(P)(R^k-R^0)\}=0,\quad k=1,\ldots,N. \tag{8}$$

The above Euler equation is usually written as $E\{m(R^k-R^0)\}=0$ with the stochastic discount factor m defined here as

$$m=u'(w)Z(P). \tag{9}$$

The SDF in Eq. (9) has all of the standard properties. It is positive everywhere, it is arbitrage-free, and it implies a linear pricing operator.[6] The shape of the weighting

[5] While this specification is convenient for theoretical analysis, it cannot accommodate S-shaped weights as the other two we consider.

[6] For a well-defined G, the RDEU utility functional is strictly monotone and respects first-order stochastic dominance irrespective of the shape of the weights (see Quiggin, 1982, 1993). The shape of the weighting function does matter for the second order stochastic dominance (SSD). A sufficient condition for SSD is $G(P)$ and $u(w)$ to be concave. RDEU agent could be risk-seeking with respect to mean-preserving spreads when G contains convex segments. This feature, however, does not imply arbitrage in prices because the SDF is always

function does not affect any of these properties.[7] We can write the risk-neutral probability density function (PDF) q as

$$q = \frac{m}{Em} \times p = \frac{u'(w)Z(P)p}{E\{u'(w)Z(P)\}}. \tag{10}$$

If we denote $R \equiv w/w_0$ gross return on total investor wealth and specialize to power utility $u(w) = w^{1-\gamma}/(1-\gamma)$, after normalizing initial wealth to one, we can rewrite the SDF and the price density as

$$m = u'(R)Z(P) \tag{11}$$

and

$$q = \frac{m}{Em} \times p = \frac{R^{-\gamma}Z(P)p}{E\{R^{-\gamma}Z(P)\}}, \tag{12}$$

where P and p now denote the CDF and PDF of R.[8] We can now proceed to investigate how probability weighting function affects moments of the risk-neutral distribution.

2.3. Probability weighting and the risk-neutral moments

Using the standard power expected utility, Bakshi, Kapadia, and Madan (2003) derive approximations for skewness of state price density up to the first order of γ. The expression for risk-neutral skewness links it to the first four moments of the physical distribution as follows:

$$SKEW_Q \approx SKEW_P - \gamma(KURT_P - 3)\sigma_P^2, \tag{13}$$

where P and Q refer to moments under physical and risk-neutral distributions, respectively. Thus, risk-neutral skew consists of inherited skewness of the physical distribution and can be further affected by the risk aversion when the underlying physical distribution is fat-tailed.

The utility function with probability weights introduces another layer between risk-neutral and physical distributions. One could equivalently think of RDEU as EU computed with respect to the transformed probability distribution of wealth (return):

$$U = E_G\{u(w)\}, \tag{14}$$

where E_G denotes the expectation operator with respect to the distribution generated by $G(P)$. Under G-distribution we have $Z \equiv 1$ and the state price density (SPD) in Eq. (12) has the standard form corresponding to the power expected utility, which is analyzed by Bakshi, Kapadia, and Madan (2003). Thus, if the underlying return distribution was generated by $G(P)$, we can use the relation from Eq. (13) replacing P by $G(P)$:

$$SKEW_Q \approx SKEW_G - \gamma(KURT_G - 3)\sigma_G^2. \tag{15}$$

Probability weighting directly affects the moments of the physical distribution, and the standard risk-aversion curvature further affects the skewness.

How does the transformation G affects the moments of the physical distribution? In Appendix A, we consider a general transformation G and a general distribution P and derive approximations to the moments. To introduce notation, let $\mu^{(k)}$ and $\kappa^{(k)}$ denote the uncentered kth moments of the random variables generated by P and $G(P)$, respectively. Also, let $\mu_{r:s}^{(k)}$ denote the kth (uncentered) moment of the order statistic r $(1 \le r \le s)$ out of sample size s drawn from independently and identically distributed (i.i.d.) random variables generated by P $(\mu^{(k)} \equiv \mu_{1:1}^{(k)})$.

Proposition 1. Let $0 < \overline{P} < 1$ be such that weighting density and its derivatives are finite; $Z^{(k)}(\overline{P}) < \infty$, and assume that $|\mu_{k:k}^{(n)}| < \infty$, for $1 \le k \le K+1$. Then, the nth uncentered moment of the transformed distribution is approximated by a linear combination of the nth moments of the maximum order statistics from the samples up to size $K+1$:

$$\kappa^{(n)} \approx \sum_{k=0}^{K} a_k \mu_{(k+1):(k+1)}^{(n)}, \tag{16}$$

$$a_k = \frac{1}{k+1} \sum_{j=k}^{K} \frac{(-1)^j}{(j-k)!} Z^{(j)}(\overline{P}) C_j^{j-k} \overline{P}^j,$$

where C_m^l is the combinatorial coefficient.

Proof. See Appendix A.

While Proposition 1 applies to a general weighting function and distribution, it does not establish unambiguous comparative statics about the moments of the risk-neutral density in a general case. We were able to obtain several results analytically by specializing in the weighting function of the type given in Eq. (3) when ϕ is a positive integer. We also consider numerically the two-parameter versions of the Kahneman and Tversky and the Prelec functions given in Eqs. (4) and (5).

For the weighting function given in Eq. (3) with integer ϕ, we can express the moments of the transformed distribution as functions of only two extreme order statistics (minimum and maximum) from the sample of size ϕ. Here we outline some main points; the details are given in Appendix A. We show that when ϕ is a positive integer we have the following result:

$$\kappa^{(k)} = \lambda \mu_{\phi:\phi}^{(k)} + (1-\lambda)\mu_{1:\phi}^{(k)}. \tag{17}$$

Eq. (17) can be used to construct central moments, and it can be used for some types of distributions to obtain approximate analytical results about the effect of the weighting function parameters on the transformed moments. The central moment of order n is a polynomial

(*footnote continued*)
positive. It simply implies that lottery-like contracts could be valued higher by an RDEU agent compared with the standard EU specification. In fact, as we show below, when G has convex segments certain call-type derivative contracts could have negative expected returns.

[7] When state-space is discrete, the RDEU functional is in general nondifferentiable. Furthermore, for weighting functions with convexities, the first-order conditions are not sufficient for extremum because the utility functional might not be globally concave. As a result, there could be multiple maxima, and sometimes optimal demand functions are discontinuous with respect to prices. Therefore, in discrete state settings, one has to be careful to examine solutions obtained by standard gradient methods and explicitly consider solutions at the allocations in which indifference curves have kinks. In the continuous state setting adopted in this paper, RDEU is smooth and differentiable (Ai, 2005).

[8] Monotonic transformations preserve rank of outcomes and, therefore, we can replace the CDF of total wealth with that of portfolio return as a state-ranking variable under the expectation.

of order n in λ with the coefficients as functions of extremum statistics. Assuming that P is symmetric ($P(t)=1-P(-t)$), using the identity $\mu^{(k)}_{\phi:\phi}=(-1)^k\mu^{(k)}_{1:\phi}$, Eq. (17) simplifies to

$$\kappa^{(k)}=\mu^{(k)}_{\phi:\phi}(1+\lambda(1+(-1)^k)). \tag{18}$$

For a symmetric distribution, the even uncentered moments do not depend on λ and depend only on ϕ, while the odd moments are all zero at $\lambda=\frac{1}{2}$.[9] For a symmetric distribution, we provide expressions for variance and skewness in Appendix A. In particular, it is interesting to consider the approximate behavior of skewness and variance for values of $\lambda=\frac{1}{2}+\epsilon$, where ϵ is small. These are probability weighting transformations that are nearly symmetric with respect to tail behavior. The exact expression for variance and the approximation of skewness up to the first order in ϵ are given by

$$\sigma^2_G=\mu^{(2)}_{\phi:\phi}-(\mu^{(1)}_{\phi:\phi})^2\epsilon^2,$$

$$SKEW_G\approx\epsilon\times\frac{\mu^{(3)}_{\phi:\phi}-3\mu^{(2)}_{\phi:\phi}\mu^{(1)}_{\phi:\phi}}{(\mu^{(2)}_{\phi:\phi})^{3/2}}. \tag{19}$$

We show in Appendix A that the sign of the coefficient on ϵ in skewness is negative. When $\epsilon<0$ the concave (pessimistic) segment of the weighting function is relatively stronger. While this lowers the mean of the transformed distribution, it also stretches the right tail and makes the skewness less negative. This property could help to account for several rare episodes in the data when the risk-neutral skewness exceeds physical skewness. The physical distribution typically exhibits fat tails so that the standard risk aversion (curvature) usually implies more negative risk-neutral skewness than physical according to Eqs. (13) and (15). However, changes in the relative importance of the inverse S-shape segments could induce additional variation of the risk neutral skewness not accounted for by standard risk aversion mechanism. The variance of the transformed symmetric distribution is at the maximum when $\lambda=0.5$. The transformation shifts and compresses the distribution either to the right or to the left, which reduces the variance. The effect of ϕ is not easily presented analytically, even for a symmetric distribution, because the behavior of the maximum statistic's moments in general depends on the shape of the CDF P.

The above discussion helps explain some basic intuition about the inverse S transformation given in Eq. (3). However, for practical purposes, we also would like to analyze the transformations in which analytical results are not possible to derive. Using numerical integration, we consider Prelec and Kahneman and Tversky transformations from Eq. (4) with their two-parameter extensions $G_1(P^\beta)$ and $G_2(P^\beta)$. We consider the effect of these transformations on the moments of the standard normal distribution and the physical distribution arising from a model with (exponential) moneyness-dependent volatility as in Aït-Sahalia and Duarte (2003). Qualitatively, the effects of transformations are similar across the two types of functions and, to conserve space, we present only Prelec transformation applied to the Gaussian distribution and to the distribution from the Aït-Sahalia and Duarte (2003) model.

Figs. 1 and 2 show the mean, variance, skewness, and kurtosis of the transformed distributions for various values of α and β. When $\alpha=\beta=1$, the moments are undistorted. β primarily affects the mean, and α primarily affects higher order moments. This is because the transformation P^β is either uniformly convex or concave and shifts the underlying distribution, and α determines the strength of the inverse S-shape and mainly affects the tails. The transformed mean is increasing in β because higher β results in more convex shape (or less concave), which shifts the probability mass to the right.

The inverse S portion of the transformation primarily affects the tails, and its effect shows up in higher order moments. The variance is decreasing in α because the transformation changes from the inverse S to S-shaped for $\alpha>1$. The S-shaped transformation underweights the tails and reduces variance. Skewness is increasing in α because the transformation is asymmetric and, as α gets larger, it reduces the weight in the left tail more rapidly than in the right tail. Similarly to the analytical case for polynomial weighting function considered previously, the transformed skewness can be either higher or lower than the underlying skewness. In particular, the more pronounced inverse S-shape (lower α) results in lower skewness and the S-shape implies higher skewness than the underlying. This pattern appears to be stable across different transformations and underlying distributions we consider. Finally, kurtosis is mostly decreasing in α for the normal case, but it can also be non-monotone for non-Gaussian underlying distribution. The effect on kurtosis in α shown in Fig. 2 is hump-shaped, and the location of the maximum depends on β. The transformed kurtosis could be higher or lower than for the original underlying distribution.

In summary, the probability weighting transformations applied to the physical density produce intuitive effects on the moments of the transformed distribution. Probability weighting function has an independent effect on risk-neutral moments, distinct from the standard risk aversion. As part of our empirical analysis, we fit some of the functional forms we introduced above and show that these transformations can be useful for understanding the connection between physical and risk neutral densities estimated from options data.

2.4. Probability weighting, the shape of the pricing kernel and expected returns

Using the expression for the pricing kernel from Eq. (11) we can write the condition for an increasing SDF as

$$\frac{dm}{dR}=u''(R)Z(P)+u'(R)Z'(P)p>0. \tag{20}$$

We can rewrite the above as

$$\frac{G''(P)}{G'(P)}p>-\frac{u''(R)}{u'(R)}, \tag{21}$$

[9] Central moments do depend on λ because they are polynomials that include both odd and even uncentered moments. As a result, skewness and kurtosis also depend on λ.

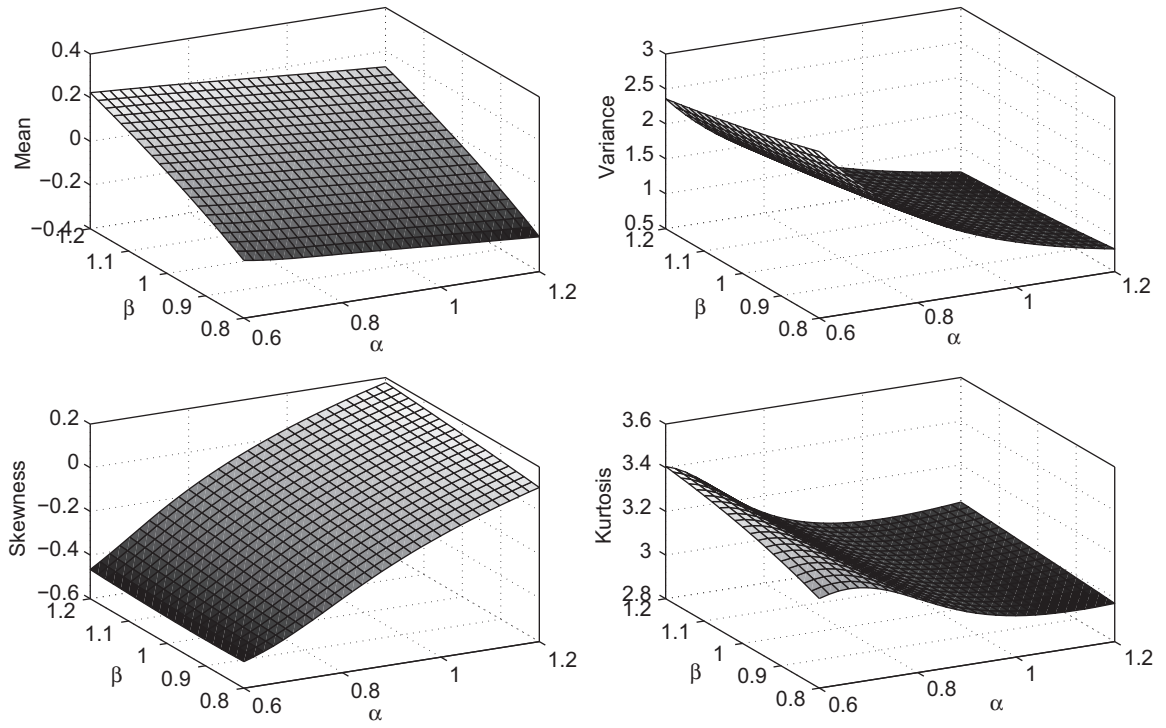

Fig. 1. Moments of the transformed distribution $G_2(P^\beta)$ [two-parameter extension of Prelec function in Eq. (4)] from the underlying standard normal P.

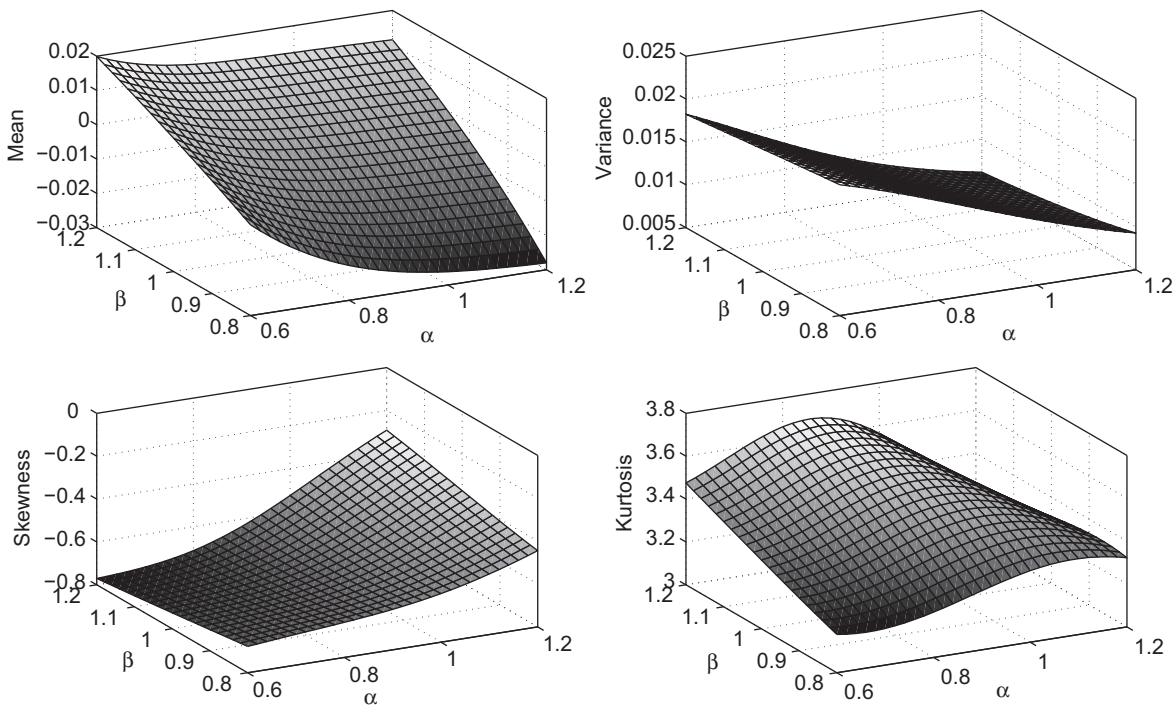

Fig. 2. Moments of the transformed distribution $G_2(P^\beta)$ [two-parameter extension of Prelec function in Eq. (4)]. The underlying P is from the model with moneyness-dependent volatility.

where R, as before, denotes the return on wealth. For the power utility in wealth, the right hand side of Eq. (21) is equal to the relative risk aversion (RRA) coefficient γ. The ratio $-G''(P)/G'(P)$ is referred to as "probabilistic risk aversion" (see Quiggin, 1993, Chapter 6).[10] Thus, the condition for nonmonotonicity is expressed in terms of relative strengths of the risk aversion over outcomes and risk seeking in the probabilities weighting function. Because all valid weighting functions are nondecreasing ($G' > 0$), the above condition requires the weighting function to have a convex segment ($G'' > 0$) to generate nonmonotonicity in the SDF.

Bakshi, Madan, and Panayotov (2010) demonstrate that general U-shaped pricing kernels imply negative expected returns for calls with moneyness in the range where SDF is increasing. Therefore, the SDF with probability weights that has sufficient convexity implies this property as well. The SDF with probability weights need not be constrained to a global U-shape. The probability weighting function allows for additional flexibility and could be used to construct an SDF that is oscillating, i.e., contains a U-shaped and a monotone declining right tail similar to that estimated by Rosenberg and Engle (2002). In fact, when the ratio $G''(P)/G'(P)$ is bounded, then

$$\exists \overline{R} : R > \overline{R} \Rightarrow \frac{G''(P)}{G'(P)} p < \gamma, \tag{22}$$

because $\lim_{R \to +\infty} p(R) = 0$. Thus, for such G the SDF eventually returns to its normal declining shape and call contracts that are sufficiently out-of-the-money (OTM) will have positive expected return.[11]

An alternative way to illustrate the connection between the expected returns on options and the weighting function is to use a decomposition of the expected payoff. Consider a contingent claim g of the call type contract such that $g(R;K)$: $\{g(R;K) > 0 \text{ for } R > K;\ g(R;K) = 0 \text{ for } R \le K\}$ (we suppress the expiration time T argument for brevity). Assume that $\Pr\{R > K\} > 0$, i.e., the probability that the contract will be in the money is strictly positive. $g(R;K)$ includes the type of claims considered by Bakshi, Madan, and Panayotov (2010) such as standard call option, digital call, and kernel call. Assume that pricing kernel has a form $m = u'(R)Z$, where $u(\cdot)$ is a strictly concave and strictly increasing utility function with $\lim_{R \to \infty} u'(R) = 0$ and $Z > 0$ is the probability weighting density. Because $g \ge 0$ and has a positive probability of positive payoff, under the no-arbitrage assumption its price is strictly positive. Then, the expected return on the contract is negative if, and only if, its expected net payoff is negative: $E_t g - E_t\{mg\} < 0$. Consider the following decomposition of the expected net payoff:

$$E_t g - E_t\{gu'Z\} = -E_t\{g(u'-1)Z\} - E_t\{g(Z-1)\}. \tag{23}$$

The first term in the decomposition is positive for sufficiently high K because u is strictly concave, such that eventually $u'(R) < 1$ for $R > \underline{K}$ for some $\underline{K} < \infty$. Therefore, a necessary condition for the negative expected return on claim g is that $Z > 1$ at least for some $R > K$, where $g > 0$. If we restrict attention to the weighting functions with convex Z, then the necessary condition for negative expected return on g is that $Z > 1$ for $R > \underline{K}$ for some $\underline{K}$. Therefore, a weighting function with convex density Z has to overweight the entire right tail of the returns distribution to imply negative expected returns on the call-type contingent claims $g(R;K)$.

From the above discussion of the properties of the pricing kernel with probability weights, we conclude that a probability weighting function that overweights the tails has a potential to reconcile with theory a number of empirical observations such as U-shaped kernels, negative implied risk aversion and negative expected returns for OTM calls. However, ultimately the shape of the weighting function can be established only empirically and preferably through a procedure that minimizes the effect of model specification and parametric assumptions about physical distribution.

3. Estimation of probability weighting functions

Two approaches can be taken to estimate the probability weighting function G. One is to rely on Eq. (12) by estimating the ratio of risk-neutral and physical densities and, thus, the probability weighting density Z. The function G is obtained through integration over Z. The other approach is estimating the function G directly. We apply the latter approach in our implementation. Given the physical distribution function $P(\cdot)$ and its density $p(\cdot)$ and the risk-neutral distribution $Q(\cdot)$ and its density $q(\cdot)$ over the returns R, we proceed as follows. For a specific P_0 with corresponding return R_0 such that $P(R_0) = P_0$, we have

$$\begin{aligned} G(P_0) = G(P^{-1}(R_0)) &= \int_0^{P_0} Z(P)\, dP \\ &= \int_0^{R_0} Z(P(R))p(R)\, dR = c \int_0^{R_0} \frac{q(R)}{u'(R)}\, dR \\ &= c\left[\frac{Q(R_0)}{u'(R_0)} + \int_0^{R_0} Q(R)\frac{u''(R)}{u'(R)^2}\, dR\right], \end{aligned} \tag{24}$$

where $u'(\cdot)$ is the marginal utility and the normalizing constant $c = (\int_0^{\infty} (q(R)/u'(R))\, dR)^{-1}$. If the utility function is linear, we have $G(R_0) = Q(R_0)$, where the probability weighting amounts to the change of measure. We can see from Eq. (24) that this approach has three advantages over numerically integrating Z to get the weighting function G. First, the probability distribution function can be estimated more efficiently than the probability density function. This is especially important in a finite sample setting. Second, the ratio of densities is difficult to estimate at tails as the physical density tends to decay faster than the risk-neutral one. Third, because the physical distribution function affects only the upper bound of

[10] Quiggin shows that $-G''(P)/G'(P) \ge 0$ along with concavity of $u(\cdot)$ is necessary and sufficient for SSD. Therefore, increasing pricing kernel implies risk-seeking behavior toward probabilities in some range of outcomes.

[11] Some evidence of this is presented in Bondarenko (2003, Table 1). However, as a caveat, it is hard to estimate reliably the returns of the far OTM options due to low liquidity. Thus, it is relatively more difficult to establish empirically whether the expected returns of far OTM calls conform to the oscillating versus globally U-shaped form. Nevertheless, we consider valuable the ability of the probability weighting function to accommodate oscillating SDF pattern in a parsimonious way.

Table 1
Summary statistics of the out-of-the-money option quotes.

This table reports the number of quotes, the Black and Scholes implied volatility (mean, minimum and maximum), and the average trading volume and open interests. The OTM puts have moneyness below 1 and the OTM calls above 1. The sample period is from January 1996 to December 2008. For 28-day options, we pick the date with 28 days to the option maturity date each month. An adjacent date is selected if it is a non-trading day or of few observations. The same procedure is applied to 45- and 56-day options.

Moneyness	Number of quotes	Implied volatility (percent)			Average trading volume (thousands)	Average open interest (thousands)
		Average	Minimum	Maximum		
28-day options						
+50 (0,0.9]	998	37.68	19.25	139.87	2,705	16,118
+50 (0.9,0.95]	987	23.86	12.76	77.33	2,966	20,098
(0.95,0.97]	495	20.59	11.17	73.42	2,973	19,694
(0.97,0.99]	554	19.56	9.67	72.13	3,628	18,533
(0.99,1]	279	18.35	8.82	70.66	4,572	15,688
(1,1.01]	279	17.68	8.31	70.00	3,726	15,515
(1.01,1.03]	533	16.34	7.35	69.67	2,651	15,193
(1.03,1.05]	456	16.08	7.25	67.90	2,019	13,756
(1.05,1.1]	519	19.65	10.03	65.83	1,650	13,048
(1.1,+∞)	157	36.57	16.44	62.98	1,970	16,684
45-day options						
+50 (0,0.9]	997	32.82	16.66	106.02	902	13,004
+50 (0.9,0.95]	733	22.80	13.49	56.55	1,235	15,317
(0.95,0.97]	361	20.62	12.02	52.80	1,304	14,579
(0.97,0.99]	412	19.09	10.56	50.82	1,775	14,866
(0.99,1]	232	18.18	9.74	49.33	1,677	13,163
(1,1.01]	209	17.13	9.52	47.19	1,518	9,988
(1.01,1.03]	388	16.24	8.53	46.83	990	11,604
(1.03,1.05]	361	15.56	7.74	44.70	1,140	11,362
(1.05,1.1]	489	17.19	7.86	43.40	1,211	11,020
(1.1,+∞)	232	23.97	13.55	39.75	756	11,685
56-day options						
+50 (0,0.9]	957	31.67	16.03	100.38	787	12,067
+50 (0.9,0.95]	613	21.98	12.58	61.78	1,345	14,903
(0.95,0.97]	299	20.47	11.19	58.32	1,640	14,874
(0.97,0.99]	347	18.85	10.09	56.64	1,755	12,978
(0.99,1]	199	17.94	9.26	55.34	1,848	11,959
(1,1.01]	191	17.44	9.04	54.04	1,681	10,757
(1.01,1.03]	327	16.42	8.16	53.22	1,127	11,280
(1.03,1.05]	295	15.37	7.44	51.87	1,113	11,119
(1.05,1.1]	477	16.52	7.43	50.47	743	9,944
(1.1,+∞)	257	21.39	12.33	46.95	462	9,561

the integral, moderate change in the estimates of the physical distribution will not change the shape of the probability weighting function qualitatively. So our empirical findings are robust to the method for estimating the physical distribution function.

Given our desire not to impose any a priori restrictions on the elements of the SDF, we can estimate only combined SDF and cannot separately identify the marginal utility and probability weighting density. Therefore, we assume several standard utility functions commonly found in empirical and theoretical work in asset pricing and we restrict utility to be only nonconvex. This procedure allows us to obtain nonparametric estimates of the probability weights. We also approximate nonparametric probability weighting functions using parametric specifications of Prelec and Kahneman and Tversky functions.

3.1. Data

We obtain the data on the Standard & Poors (S&P) 500 index options (symbol SPX) from OptionMetrics. The market for SPX options is one of the most active index options markets in the world. These data have been the focus of many empirical investigations, including Bakshi, Cao, and Chen (1997), Aït-Sahalia and Lo (1998, 2000), Aït-Sahalia and Duarte (2003), Jackwerth (2000), and Rosenberg and Engle (2002), among many others. The options are European, have no wild card features, and can be hedged using the active market on the S&P 500 index futures. Our sample period is from January 1996 to December 2008. We select monthly option quotes with closest to 28, 45, and 56 days from expiration date. The average of bid and ask prices is taken as the option price. We also obtain the term structure of default-free interest rates from OptionMetrics.

To exclude illiquid options, we discard the in-the-money options, options with zero trading volume or open interest, and options with quotes less than 3/8. We also exclude options that allow for arbitrage across strikes.[12] Table 1 reports the summary statistics of the option

[12] Specifically, we exclude options that violate the monotonicity constraint across strikes but keep options that violate the convexity constraint, which is more frequent.

quotes over our sample period across moneyness. Specifically, we report the Black and Scholes implied volatilities, average trading volume, and open interest for calls and puts that are 10%, 5%, 3%, and 1% out-of-the-money. The average number of options is around 34 each month. There are more OTM puts than OTM calls, averaging about 16 puts that are at least 3% OTM versus about seven calls that are at least 3% OTM each month for the 28-day options. The average Black and Scholes implied volatilities exhibit the smirk shape as shown in the option pricing literature. The average trading volumes for the OTM options suggest they are liquid compared with the near-the-money options.

Next, we apply the same procedure as done in Aït-Sahalia and Lo (1998) to address the problems of non-synchronous prices between the option and underlying index (Fleming, Ostdiek, and Whaley, 1996) and the unobserved dividend process in the data. Specifically, at each day t the forward price $F_t(T)$ of maturity T and the spot price S_t are linked via the no-arbitrage argument:

$$F_t(T) = S_t e^{(r_{t,T}-\delta_{t,T})(T-t)}, \tag{25}$$

where $r_{t,T}$ is the risk-free rate and $\delta_{t,T}$ is the dividend yield from t to T. This forward price can be inferred from option prices through put-call parity, that the call price $C(t)=C(S_t,X,T-t,r_{t,T},\delta_{t,T})$ and put price $J(t)=J(S_t,X,T-t,r_{t,T},\delta_{t,T})$ of the same maturity T and strike price X satisfy

$$C(t)-J(t) = e^{-r_{t,T}(T-t)}[F_t(T)-X]. \tag{26}$$

This relation is independent of any option pricing model. Using the near-the-money call and put option prices, we can derive the implied forward price of the underlying index. This procedure removes the problem of matching the option prices and the underlying spot by their recording times. Next, we compute the in-the-money call prices from the out-of-the-money puts using the put-call parity and implied forward price. This is necessary when we later estimate the risk-neutral density by taking derivatives of the call price with respect to the strike. The index returns in our setting are the ratios of the forward prices, i.e., $F_T(T)/F_t(T)=S_T/F_t(T)$, not the spot prices S_T/S_t. For stochastic dividend processes, the returns on the forward prices are better proxies for the returns on the total wealth process by not excluding the dividends.

3.2. Empirical methods

In the following subsections we discuss the empirical procedure for estimating the moments and distribution functions under risk neutral and physical measures. Technical details of the estimation procedure are provided in Appendix B.

3.2.1. Estimation of risk-neutral density and distribution function

Our goal is to estimate the risk-neutral density of the S&P index returns with a relatively small sample of observations (SPX option quotes from a single day) without imposing parametric restrictions on the shape of the density. Two commonly applied approaches are used to this end. One is semi-nonparametric via approximating the shape of the density function by series expansion. The most popular method within this approach is the Gram-Charlier series expansion (GCSE). This includes the so-called A-type GCSE (see Jarrow and Rudd, 1982) and the C-type GCSE (see Rompolis and Tzavalis, 2008). The other approach is fully nonparametric yet customized for small samples, such as the constrained local polynomial method of Aït-Sahalia and Duarte (2003). Between the two, the GCSE based method can be more efficient than the constrained local polynomial method for smaller samples due to its semi-nonparametric nature, but it can also introduce bias to the shape of the estimated risk-neutral density if the true shape of the density function is far from what the specific type of series can approximate the constrained local polynomial method can capture various shapes of the density function, but choosing the optimal smoothing parameters for small samples can be a demanding task.

With this trade-off in mind, we apply the constrained local polynomial method with the guidance of the semi-nonparametric method. Specifically, we have three steps in our procedure. First, the risk-neutral moments are estimated based on the spanning result from Bakshi and Madan (2000) and Bakshi, Kapadia, and Madan (2003). Second, we estimate the GCSE-based semi-nonparametric risk-neutral density from the moments estimates. Finally, we estimate the density using the constrained local polynomial method in which the smoothing parameter, the bandwidth, is chosen by minimizing the simulated mean squared errors (MSEs) using the bootstrapped samples generated from the semi-nonparametric estimates. There are two advantages in this procedure. First the semi-nonparametric estimates provide a robust benchmark for choosing the bandwidth via simulation.[13] Second, the semi-nonparametric estimates themselves can be used as a robust check for the conclusions based on the nonparametric estimates.

We describe in general our procedure below, leaving technical details in Appendix B. First, given the forward price $F_t(T)$ of the underlying stock price, we define the $(T-t)$ period log-return as $R_t(T)=\ln(F_T(T)/F_t(T))$. The risk-neutral moments, $\mu_{R,n}=E_t^Q[R_t^n(T)]$, can be computed from the OTM call and put prices. Next, we use the estimates of the risk-neutral moments to construct the risk-neutral density of the index returns $R_t(T)$, through Gram-Charlier series expansions. We consider both the A-type and the C-type GCSE. The A-type GCSE has been widely used in estimating the risk-neutral densities; for example, see Jarrow and Rudd (1982). The A-type series expansion is specified around the Gaussian density to account for the skewness and excess kurtosis of the distribution. However, when the estimated distribution deviates considerably from the Gaussian distribution, the A-type GCSE estimator might not be as efficient. The C-type GCSE (see Rompolis and Tzavalis, 2008) is not specified around a particular distribution and, thus, works better when there is large deviation from the Gaussian distribution. The C-type GCSE can also guarantee the estimated density being

[13] The reason that simulation is necessary for the choice of the bandwidth is that we are dealing with small samples and finite sample bias and variance are not available especially for the constrained local polynomial method proposed in Aït-Sahalia and Duarte (2003).

positive. However, the C-type GCSE is computationally more challenging in that for computing the series up to order m the C-type requires knowing the first $(2m-2)$ risk-neutral moments, compared with the first m moments for the A-type.[14] This is especially problematic when the estimation of higher order moments is inaccurate. To this end, we compute density estimates from both the A-type and C-type GCSE and choose the best estimate based on the information criteria, such as the Akaike information criterion (AIC) or Schwarz criterion (SC).

Next we present the nonparametric method, the constrained local polynomial method used in our paper. This method follows the insight from Breeden and Litzenberger (1978), that the risk-neutral density is the scaled second derivative of the call option price with respect to the call strike price. Studies using this insight to estimate the risk-neutral density include Aït-Sahalia and Lo (1998, 2000), Jackwerth (2000), and many others. Based on the no-arbitrage argument, the option value can be computed as the expectation of payoff at maturity using the risk-neutral density. We can estimate the risk-neutral distribution function and density by taking derivatives of call price function with respect to strike. Estimating the derivative via the local polynomial method has been shown to be superior to the traditional kernel regression method in the statistics and econometrics literature (for reference, see Fan and Gijbels, 1996, Subsection 3.2). Furthermore, when working with data of small sample size such as the S&P 500 option quotes from a typical day, Aït-Sahalia and Duarte (2003) demonstrate that the constrained local polynomial method works particularly well. We, therefore, employ this method for our nonparametric estimation.

We describe the constrained local polynomial method in Appendix B. See Aït-Sahalia and Duarte (2003) for more details. One difference in our implementation is that the call price function is estimated using the GCSE method, while Aït-Sahalia and Duarte (2003) simulate data from the Black and Scholes formula with implied volatility modeled as a parametric function of moneyness. Specifically, we use the semi-nonparametric estimates of risk-neutral densities $\widehat{p}^*_{GCSE}(s)$ to compute the estimated call prices $\widehat{C}_{GCSE}(F_t,X_i)$ at time t with strike X_i. Next we compute estimated residuals $\{\hat{\varepsilon}_{t,i}= C(F_t,X_i)-\widehat{C}_{GCSE}(F_t,X_i)\}$. By resampling from the estimated residuals, we obtain a new set of residuals $\{\varepsilon^*_{t,i}\}$ and a simulated sample of call prices

$$C^*(F_t,X_i)=\widehat{C}_{GCSE}(F_t,X_i)+\varepsilon^*_{t,i}, \quad \forall t,i. \tag{27}$$

To better capture the potential heteroskedasticity of the call prices data across strikes and time, following the wild bootstrap procedure, we draw the new residuals $\varepsilon^*_{t,i}$ from a two-point distribution such that $E\varepsilon^*_{t,i}=0$, $E\varepsilon^{*2}_{t,i}=\hat{\varepsilon}^2_{t,i}$, and $E\varepsilon^{*3}_{t,i}=\hat{\varepsilon}^3_{t,i}$.[15] Repeating this procedure many times, we obtain a set of simulated samples of call prices. In our implementation we repeat the bootstrap sampling eight hundred times.

To choose the optimal bandwidth in our nonparametric estimation, we minimize the mean squared errors of the constrained locally linear estimator, which is the sum of the squared bias term and variance term. Instead of focusing on the asymptotic MSE, we choose the optimal bandwidth by minimizing the finite sample integrated MSE using the simulated call prices obtained from the bootstrap resampling. At each time t, we compute $\widehat{C}^*_h(x)$ the constrained locally linear estimator for the call price function based on a simulated sample $\{C^*(F_t,X_i),X_i\},\forall i$ with bandwidth h. Then we choose h to minimize the following integrated MSE:

$$\min_h \int E^*[\widehat{C}^*_h(x)-\widehat{C}^*_{GCSE}(x)]^2 w(x)\,dx, \tag{28}$$

where $w(x)$ is the weighting function of on the grids across strikes and the expectation E^* is taken with respect to the distribution of the simulated samples. Similar procedures can be used in obtaining the optimal bandwidth for the first and second derivatives of the call price function. The only change is that h is chosen to minimize the MSE of the first and second derivatives of $\widehat{C}^*_h(x)$.

We conduct finite sample inferences for the constrained locally linear estimator using the bootstrapped samples as well. Specifically, at each time t for each simulated sample $\{C^*(F_t,X_i),X_i\},\forall i$, we obtain an estimate of $\widehat{C}^*_{h^{opt}}(x)$ based on the optimal bandwidth h^{opt}. The estimates from all the simulated samples provide a point-wise finite sample distribution for $\widehat{C}_{h^{opt}}(x)$, the constrained locally linear estimator obtained from actual data at bandwidth h^{opt}. Similar procedures can be used to obtain finite sample distributions for the first and second derivatives of the price function $\widehat{C}_{h^{opt}}(x)$. We compute the 95% confidence interval for the nonparametric estimates of the price function and its derivatives throughout our sample period.

3.2.2. Estimation of physical distribution function

We need to estimate the distribution function under the physical measure to compute the probability weighting function. Consistent with the time-varying estimates of the risk-neutral distribution, we allow the physical distribution to vary month by month. Because we estimate the distribution from time series of the daily S&P 500 index returns, we rely on simulation to generate estimates for returns over the horizons of our interest, such as 28 and 56 days. We also want to employ the most widely used models for the data generating process of daily returns as it resembles most closely the aggregate view of the market participants. To this end, we use the exponential generalized autoregressive conditional heteroskedasticity (EGARCH) model of Nelson (1991). EGARCH models have been used widely with much success, in research and practice, to model the

[14] For example, when $m=4$, the A-type requires the first four moments while the C-type requires the first six moments. For $m=2$, both GCSE specifications coincide with Gaussian density. The C-type has explosive behavior at the positive or negative infinity for odd order of m.

[15] For more detailed discussions of wild bootstrap, see Wu (1986), Härdle and Marron (1991), Yatchewa and Härdle (2006), and many others. In our implementation, $\varepsilon^*_{t,i}=\hat{\varepsilon}_{t,i}(1-\sqrt{5})/2$ with probability $(5+\sqrt{5})/10$ and $\varepsilon^*_{t,i}=\hat{\varepsilon}_{t,i}(1+\sqrt{5})/2$ with probability $(5-\sqrt{5})/10$. Other distributions can be used as well.

time-varying features of the financial time series such as the S&P 500 index returns. Furthermore, we use the filtered innovation terms from the EGARCH model for simulation to avoid making distributional assumptions on them. Overall, our procedure follows closely Rosenberg and Engle (2002).

Specifically, there are three steps in our estimation. In the first step, we fit the daily log-return of S&P 500

Table 2
Estimation of the conditional density of daily Standard and Poor's (S&P) 500 log-returns from 1990 to 2008.

Coefficient estimates shown in the table correspond to the model in Eq. (29) [θ_1 asymmetric autoregressive conditional heteroskedasticity (ARCH), θ_2 ARCH, ϕ_1 generalized ARCH, ω_1 long run variance, and ω_2 long-run variance from 1996–2008]. We estimate an exponential ARCH(1,1) model of daily S&P 500 log-returns from 1990 to 2008 by maximizing the likelihood function. The estimated model has significant ARCH and GARCH coefficients and a significantly negative coefficient for asymmetric ARCH effect, implying negatively skewed unconditional density. We incorporate the time-varying long-run variance component to capture the time-varying physical densities in our sample. We also report the properties of the estimated standardized innovations. The normality test rejects the standardized innovations are from standard normal distribution. We use Engle (1982) ARCH test statistics to measure unexplained stochastic volatility and Ljung and Box (1978) serial correlation test statistics to measure unexplained serial correlation. The ARCH test cannot reject the nonexistence of stochastic volatility and the serial correlation test cannot reject the null of zero serial correlation. Both tests have p-values less than 0.0001 when applied to the daily log-returns.

Panel A: Parameter estimates

	Coefficient	Robust standard error	p-value
θ_1	−0.0985	0.0133	0.000
θ_2	0.1169	0.0139	0.000
ϕ_1	0.9800	0.0045	0.000
ω_0	−0.1900	0.0442	0.000
ω_1	0.0088	0.0046	0.029
Number of observations 4,790			
Log likelihood 15,763			

Panel B: Properties of the estimated standardized innovations

Skewness	Excess kurtosis	Normality test$p-$value	ARCH test$p-$value	Serial correlation *test$p-$value*
−0.3622	1.7224	0.001	0.344	0.152

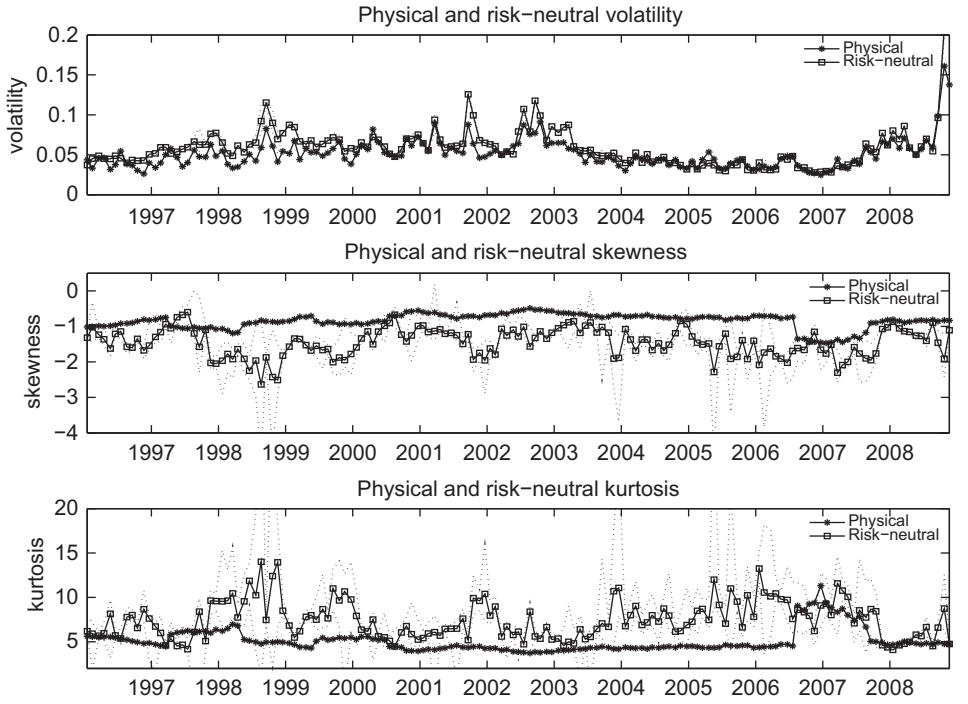

Fig. 3. Physical and risk-neutral moments from 28-day options. We plot the estimated physical and risk-neutral volatility, skewness and kurtosis of the Standard & Poor's (S&P) 500 index log returns over 28 days. The estimates are not annualized. The sample period is from January 1996 to December 2008. The physical moments are estimated using the S&P 500 daily returns. We first fit the daily returns to the EGARCH (exponential autoregressive conditional heteroskedasticity) model and simulate the returns of longer horizon. The risk-neutral moments are estimated from the out-of-the-money (OTM) options using the model-free approach of Bakshi, Kapadia, and Madan (2003). The confidence intervals of the risk-neutral moments, in dotted line, are obtained by bootstrap resampling of the OTM options.

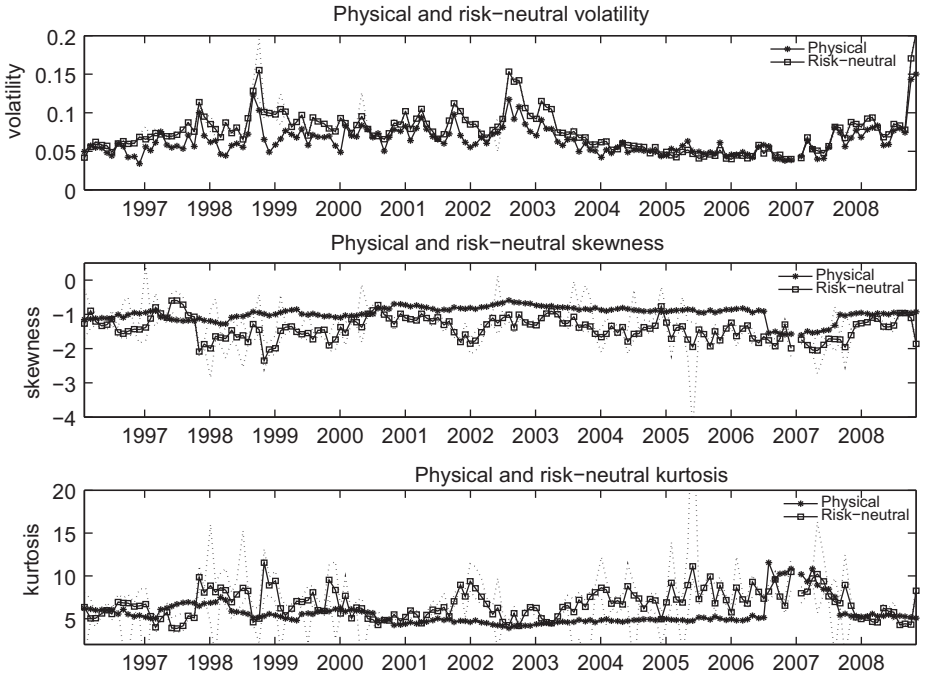

Fig. 4. Physical and risk-neutral moments from 45-day options. We plot the estimated physical and risk-neutral volatility, skewness and kurtosis of the Standard & Poor's (S&P) 500 index log returns over 45 days. The estimates are not annualized. The sample period is from January 1996 to December 2008. The physical moments are estimated using the S&P 500 daily returns. We first fit the daily returns to the EGARCH (exponential autoregressive conditional heteroskedasticity) model and simulate the returns of longer horizon. The risk-neutral moments are estimated from the out-of-the-money (OTM) options using the model-free approach of Bakshi, Kapadia, and Madan (2003). The confidence intervals of the risk-neutral moments, in dotted line, are obtained by bootstrap resampling of the OTM options.

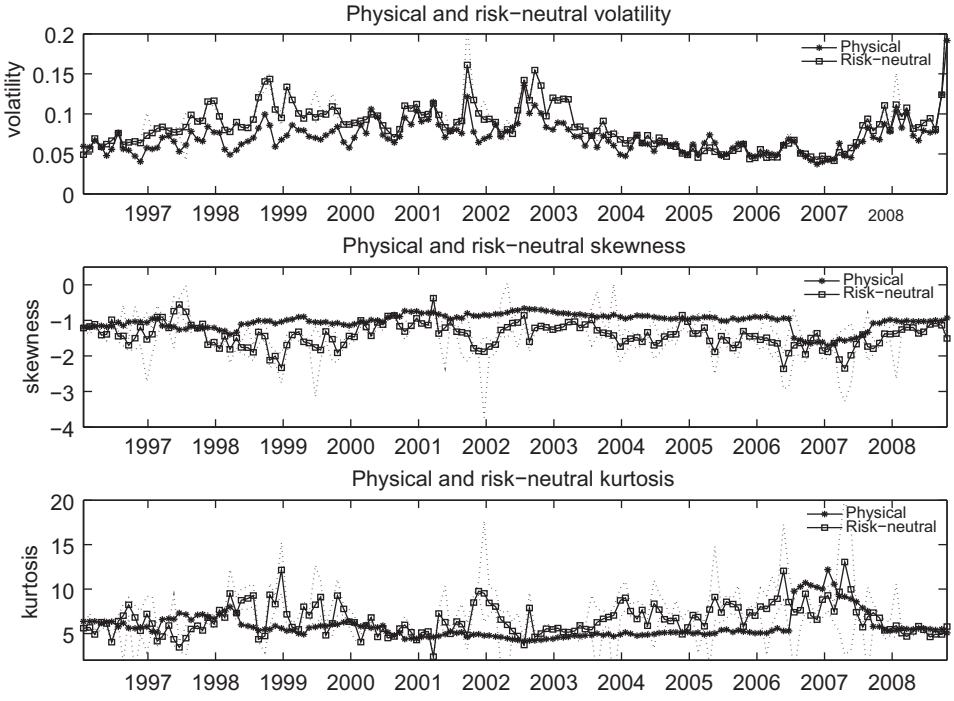

Fig. 5. Physical and risk-neutral moments from 56-day options. We plot the estimated physical and risk-neutral volatility, skewness and kurtosis of the Standard & Poor's (S&P) 500 index log returns over 56 days. The estimates are not annualized. The sample period is from January 1996 to December 2008. The physical moments are estimated using the S&P 500 daily returns. We first fit the daily returns to the EGARCH (exponential autoregressive conditional heteroskedasticity) model and simulate the returns of longer horizon. The risk-neutral moments are estimated from the out-of-the-money (OTM) options using the model-free approach of Bakshi, Kapadia, and Madan (2003). The confidence intervals of the risk-neutral moments, in dotted line, are obtained by bootstrap resampling of the OTM options.

Table 3
Summary statistics of fitted coefficient α of the two-parameter Prelec's weighting function.

We report the minimum, maximum, quartiles, and fraction below one for the fitted coefficient α of Prelec's weighting function. We estimate the Prelec parameters α and β by minimizing the distance between the Prelec function implied values and the nonparametric estimates of the probability weighting functions. The sample period is from January 1996 to December 2008. The maturities are 28, 45, and 56 days. The relative risk aversion coefficient is set to $\gamma=\{0,2\}$. The nonparametric estimates are obtained from the constrained local linear estimator with bandwidth chosen based on the simulated samples of the out-of-the-money options prices. We use the A-type and C-type Gram-Charlier series expansion semi-nonparametric estimates for simulation.

Maturity	Minimum	25th percentile	Median	75th percentile	Maximum	Percent of $\alpha<1$
$\gamma=0$						
28	0.596	0.836	0.947	1.145	1.640	59.4
45	0.534	0.854	0.953	1.071	1.518	59.7
56	0.552	0.837	0.943	1.092	1.497	60.8
$\gamma=0$						
28	0.563	0.804	0.924	1.127	1.617	63.1
45	0.493	0.810	0.928	1.041	1.503	68.0
56	0.494	0.796	0.899	1.060	1.461	66.8

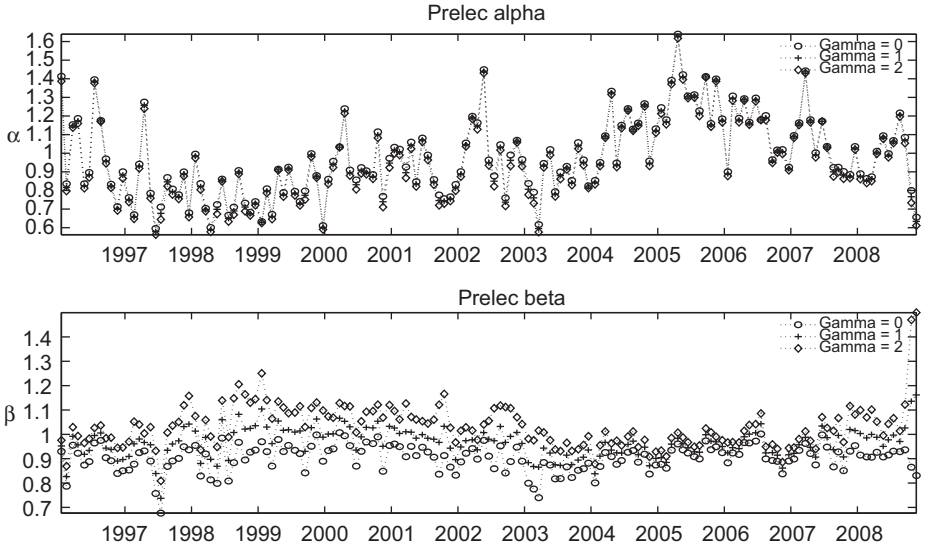

Fig. 6. Coefficient values for the two-parameter Prelec function approximating the nonparametric estimates of the probability weighting functions obtained from 28-day options. We estimate the Prelec parameters α and β by minimizing the distance between the Prelec function implied values and the nonparametric estimates of the probability weighting functions. The sample period is from January 1996 to December 2008. The relative risk aversion parameter is set to $\gamma=\{0,1,2\}$. The nonparametric estimates are obtained from the constrained local linear estimator with bandwidth chosen based on the simulated samples of the out-of-the-money options prices. We use the A-type and C-type Gram-Charlier series expansion semi-nonparametric estimates for simulation.

index to the following EGARCH(1,1) model:

$$\begin{cases} \log(S_t/S_{t-1})=\mu+\sqrt{h_t}e_t, \\ \log(h_t)=\omega_0+\omega_1 1_{96-08}+\theta_1 e_{t-1}, \\ \qquad +\theta_2[|e_{t-1}|-E(|e_{t-1}|)]+\phi_1\log(h_{t-1}), \end{cases} \tag{29}$$

where μ is the expected daily return, h_t is the conditional variance of the residuals, and e_t is the innovations that are i.i.d. with zero mean and unit variance and the unconditional expectation $E(|e_{t-1}|)=\sqrt{2/\pi}$. The conditional variance dynamics has an asymmetric term that allows for the leverage effect. ω_0 and ω_1 are, respectively, the long-run variances for the whole sample and the sample period 1996–2008 during which we estimate the physical distribution. The whole sample period is from January 2, 1990 to December 31, 2008. We use the sample before 1996 to reduce the impact of the starting values on GARCH filtering.

The EGARCH model is parsimonious and yet offers considerable flexibility in modeling the conditional distribution of the innovations. The conditional distribution of returns depends on the time-varying variance h_t and the distribution of the standardized innovations e_t. The GARCH framework permits flexible distributional assumptions for the innovations e_t. We use standard normal distribution for the innovations in the estimation step and draw the filtered innovations $\hat{e}_t$ in the simulation step.[16] This is essentially using the empirical distribution

[16] Bollerslev and Wooldridge (1992) show that the consistent estimates of the model parameters can be obtained under certain conditions even if the true innovation distribution is not normal.

of the innovations to generate returns and it helps capture the skewness and excess kurtosis of the data.

The estimates are reported in Table 2. The parameter ϕ_1 is 0.9800, suggesting persistent conditional volatility, still different from unit at 0.1% statistical significance level. The parameter θ_1 is significantly negative and, therefore, the unconditional distribution is more negatively skewed than the conditional one. The long-run variance is slightly higher after 1996. The filtered innovations pass Engle's ARCH (autoregressive conditional heteroskedasticity) test and the serial correlation test of Ljung and Box (1978) and have non-normal distribution, negatively skewed and heavy-tailed.

In the second step of our estimation, with the estimated model and the estimated conditional variance and

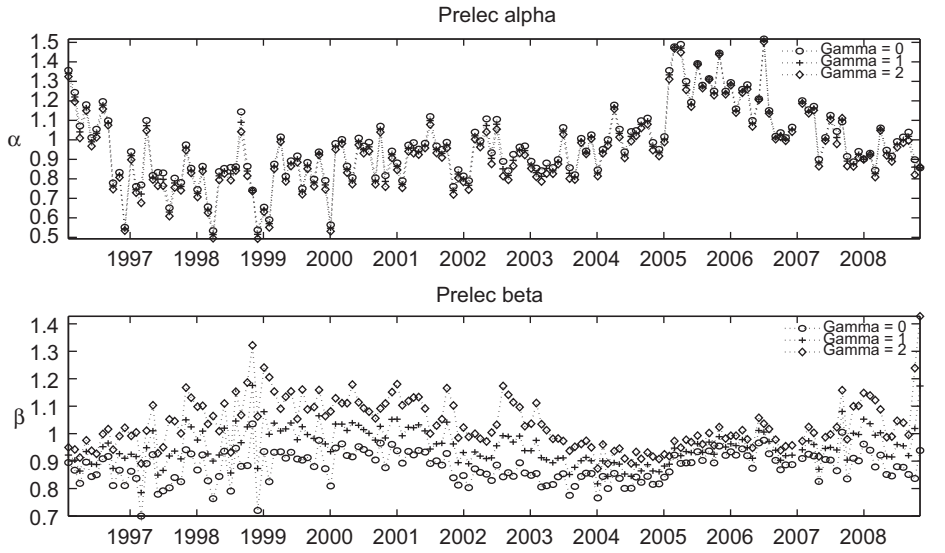

Fig. 7. Coefficient values for the two-parameter Prelec function approximating the nonparametric estimates of the probability weighting functions obtained from 45-day options. We estimate the Prelec parameters α and β by minimizing the distance between the Prelec function implied values and the nonparametric estimates of the probability weighting functions. The sample period is from January 1996 to December 2008. The relative risk aversion parameter is set to $\gamma = \{0,1,2\}$. The nonparametric estimates are obtained from the constrained local linear estimator with bandwidth chosen based on the simulated samples of the out-of-the-money options prices. We use the A-type and C-type Gram-Charlier series expansion semi-nonparametric estimates for simulation.

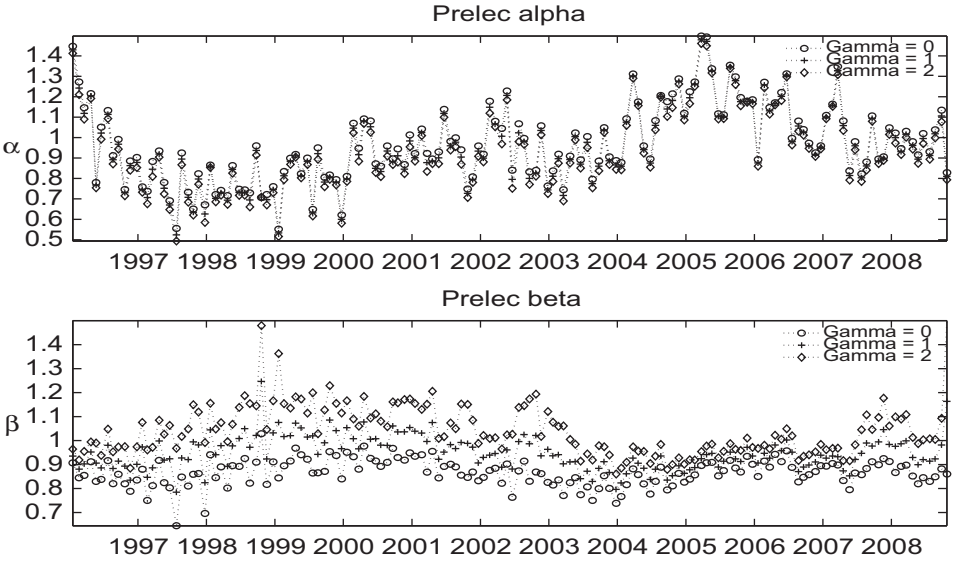

Fig. 8. Coefficient values for the two-parameter Prelec function approximating the nonparametric estimates of the probability weighting functions obtained from 56-day options. We estimate the Prelec parameters α and β by minimizing the distance between the Prelec function implied values and the nonparametric estimates of the probability weighting functions. The sample period is from January 1996 to December 2008. The relative risk aversion parameter is set to $\gamma = \{0,1,2\}$. The nonparametric estimates are obtained from the constrained local linear estimator with bandwidth chosen based on the simulated samples of the out-of-the-money options prices. We use the A-type and C-type Gram-Charlier series expansion semi-nonparametric estimates for simulation.

innovations $\{\widehat{h}_t,\widehat{e}_t\}$, we are ready to simulate the path of daily index log returns at each time t. For each path, the innovations are drawn from its empirical distribution. To better capture the time-varying features in the distribution of the innovations terms $\widehat{e}_t$, we draw with replacement from the innovations within a rolling window of time t. We experiment with different sizes of the rolling window. A large window size introduces bias by tilting toward the unconditional distribution of the innovations, and a small window size induces too much variability. We use one-year rolling window to achieve balance between the two. Our qualitative conclusions are nonetheless little affected by the choice of window size. We generate 250,000 paths for each time t. The reason for having a large simulation sample is to reduce the variance for the estimates of the moments and distribution function under the physical measure to the extent that the variability of the probability weighting function is mainly from the risk-neutral estimates.

In the final step of our estimation, we sum the simulated daily log returns over the horizon of our interests. We still need to replace the current spot price with the current forward price within the log returns. The difference between the two is the cost of carry for the forward contract under the no-arbitrage condition. So, we need to adjust the mean for the log-returns with spot prices by the amount of the cost of carry, which we compute from the risk-free rate and dividend yield for time t. Given the simulated sample of 28-, 45- and 56-day returns, we estimate the physical distribution function $P(\cdot)$ using the empirical cumulative distribution function. Due to our large simulation sample size, the empirical distribution function is rather smooth so we do not apply further smoothing to it. We also estimate the physical moments from the simulated sample.

We plot the estimates of physical and risk-neutral (nonannualized) volatility, skewness, and kurtosis in Figs. 3, 4 and 5 for 28-, 45-, and 56-day maturities.

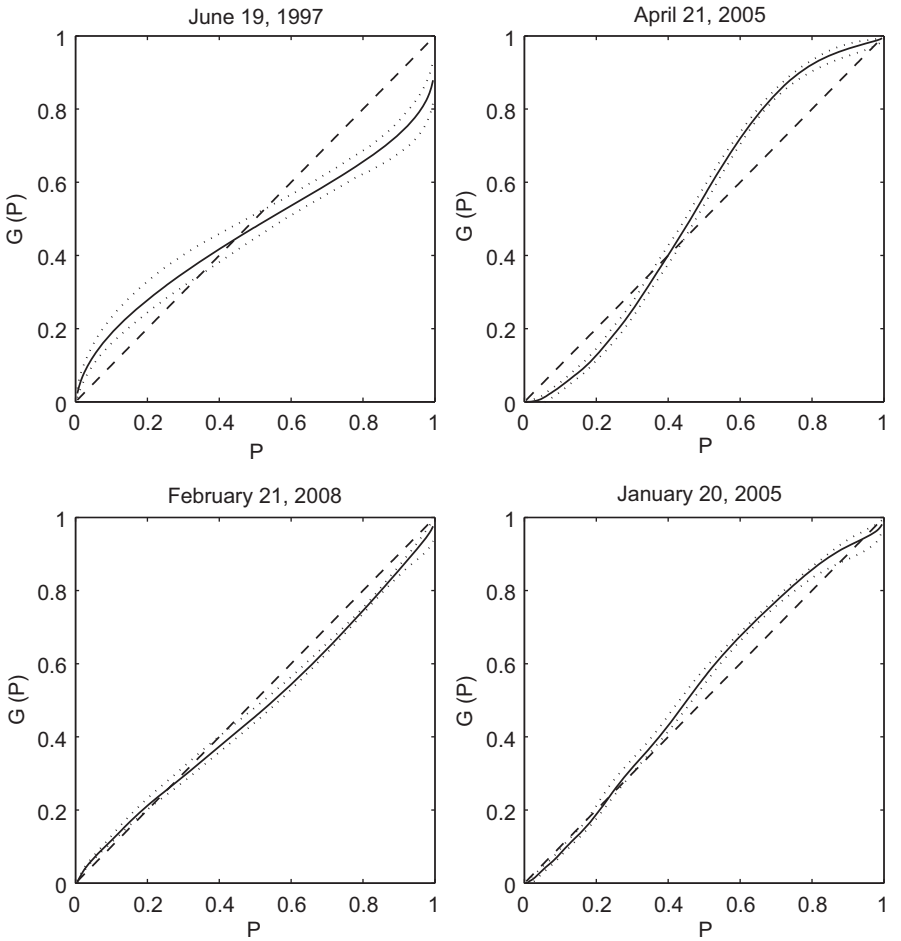

Fig. 9. Estimation of probability weighting functions. The graphs present some commonly observed shapes of the estimates of the probability weighting functions from four selected dates in our sample. The sample period is from January 1996 to December 2008. The options maturity is 28 days. The relative risk aversion coefficient is set to $\gamma = 2$. We present the nonparametric estimates obtained from the constrained local linear estimator with bandwidth chosen based on the simulated samples of the out-of-the-money (OTM) options prices. We use the A-type and C-type Gram-Charlier series expansion semi-nonparametric estimates for simulation. The four dates shown are June 19, 1997 (Prelec $\alpha = 0.56$, $\beta = 0.93$), April 21, 2005 (Prelec $\alpha = 1.6$, $\beta = 1.0$), February 21, 2008 (Prelec $\alpha = 0.84$, $\beta = 1.1$), and January 20, 2005 (Prelec $\alpha = 1.2$, $\beta = 0.93$). The confidence intervals of the nonparametric estimates, in dotted line, are obtained by bootstrap resampling of the OTM options.

The estimates of the volatilities are similar to those reported in previous studies, such as Bakshi and Madan (2006). The risk-neutral volatility is generally higher than the physical volatility, especially during the first half of the sample before 2004 and the recent financial crisis of 2008–2009. The physical and risk-neutral skewness are both negative, with the risk-neutral being more negative except for a few episodes. The estimated kurtosis are above 3 under both measures, suggesting heavier tails than Gaussian distribution. The risk-neutral kurtosis is generally higher than the physical one. The above patterns are similar for 28-, 45- and 56-day returns and consistent with previous studies such as Aït-Sahalia and Lo (1998, 2000).

We also perform several robustness checks of the estimated physical distribution. We repeat the estimation using rolling samples and samples of different lengths. For the rolling samples, we choose the rolling window of 10 and 20 years and estimate the EGARCH model for each estimation month and simulate the empirical distribution as in the base case. We find that the estimated volatilities using rolling and original samples are close, while the estimated skewness and kurtosis exhibit only moderate differences. In computing the probability weighting function, as in Eq. (24), the robustness of our results relies on the estimated inverse physical cumulative distribution function. For this purpose, we compare the estimated inverse CDF from the rolling samples with that from the original sample. Because we have an estimate of the function for each month, we consider the average over the original sample period and the time series of the 5th, 10th, 25th, 75th, 90th, and 95th percentiles from the estimates. The average inverse CDFs from the two approaches were indistinguishable, and the time series of the percentiles show very little difference as well. The robustness of the estimated inverse CDF is not surprising. The CDF function is less sensitive than the probability density function to small deviations in the estimates (in our case almost identical variance) and moderate difference in skewness and kurtosis. We also investigate the effects of different sample sizes. To this end, we fix the end date of our original sample, December 2008, and extend the start date from January 1990 to January 1980, January 1970, and the earliest available date, July 1962.

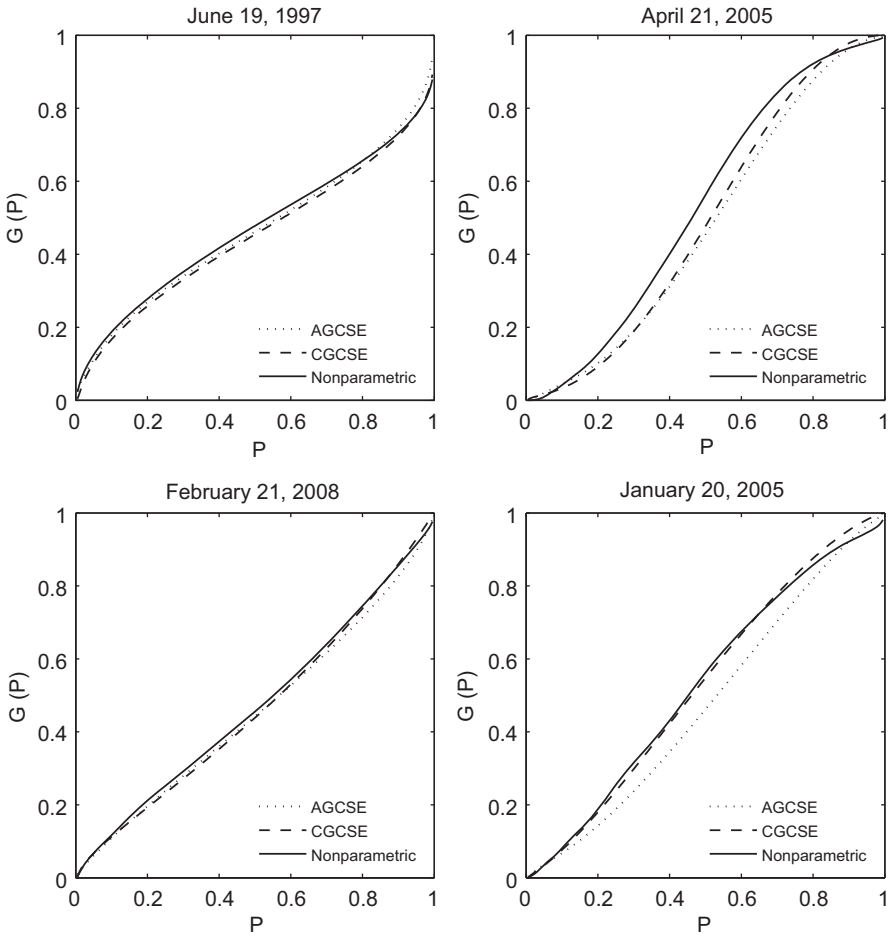

Fig. 10. Estimation of probability weighting functions. We plot the nonparametric estimates along with the semi-nonparametric estimates obtained from A-type Gram-Charlier series expansion (AGCSE) and C-type (CGCSE). They produce qualitatively similar shape. The options maturity is 28 days. The relative risk aversion coefficient is set to $\gamma = 2$.

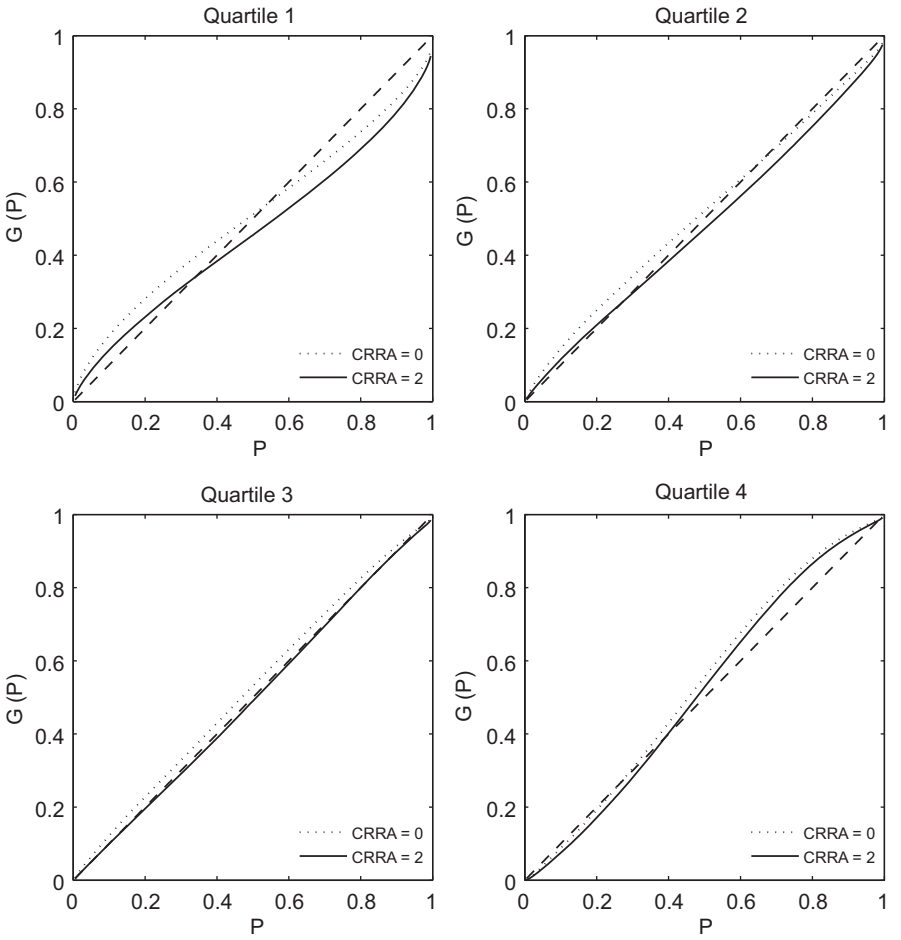

Fig. 11. Average weighting functions obtained from 28 days to maturity options. We present the estimates of the probability weighting functions average based on the quartiles of the Prelec α coefficient. The sample period is from January 1996 to December 2008. The maturities are 28 days. The relative risk aversion parameter is set to $\gamma = \{0,2\}$. We present the nonparametric estimates obtained from the constrained local linear estimator with bandwidth chosen based on the simulated samples of the out-of-the-money options prices.

We again compare with the base case the moments of the physical distribution and the average of the inverse CDF and time series of its percentiles. We found that our results are unaffected by the choice of sample length.

4. Empirical results

In this section, we first present our estimates of the probability weighting function along with the estimates of the Prelec model in fitting the nonparametric estimates. The Prelec parameters provide a concise categorization of the various shapes of the nonparametric estimates. We then define a measure of the slope of the pricing kernel and probability weighting function and present estimates of the slopes.

4.1. Probability weighting functions

Using the estimated risk-neutral and physical densities (28, 45, and 56 days), we construct probability weighting functions for every month in our sample from January 1996 to December 2008. As can be seen from Eq. (24), we can estimate the probability weighting function for any given utility function. We use the standard constant relative risk aversion (CRRA) utility functions here. Our nonparametric estimates of the weighting function vary month by month. To characterize succinctly its various shapes, we proceed through several steps.

We begin with approximations to the nonparametric estimators by fitting the two-parameter Prelec function. For each month, we use pricing kernels estimated from options with 28, 45, and 56 days to maturity. Assuming $\gamma \in \{0,1,2\}$, we construct nonparametric estimators of the weighting function and then approximate them with the best-fit Prelec function $G(P) = G(P^{\beta};\alpha)$ from Eq. (5).[17] This specification is very flexible and accommodates well all the shapes we observe in the data. When $\alpha = 1$ it corresponds to uniformly concave ($\beta < 1$) or convex ($\beta > 1$)

[17] We also use the Kahneman and Tversky two-parameter extension for such approximations and find qualitatively similar results.

weights. When $\alpha<1$, it has the inverse S-shape and changes to the S-shape for $\alpha>1$. Also, when $\alpha\neq1$, β controls the relative importance of the concave and convex segments of the function with $\beta<1$, implying more pronounced concave portion relative to the $\beta=1$ case and vice versa. The most frequently observed shape in our data is the inverse S, i.e., $\alpha<1$. In Table 3 we show the distributions of coefficient α assuming $\gamma=\{0,2\}$ obtained by fitting probability weights estimated from 28-, 45-, and 56-day options. We find that $\alpha<1$ in 60–68% of the months, depending on maturity and assumptions about risk aversion. In other months, the probability weights are nearly linear or S-shaped.

To see the variation in the shapes over time, we report time series of the Prelec function parameters in Figs. 6, 7 and 8 for 28-, 45-, and 56-day maturities, respectively. Each month has three values for α and β corresponding to different value of utility curvature γ. The value of risk aversion does not affect significantly the extent to which the probability weights are inverse S- or S-shaped. We can see that α's in every month are close for all three values of γ. In contrast, there is a substitution between γ and β, assuming higher γ implies higher values for β. This is because both of them control global concavity or convexity of the pricing kernel but act in the opposite direction: Introducing more concavity through higher γ requires a convex correction through $\beta>1$ to maintain the approximation close to the nonparametric estimate. As seen in Figs. 6–8, a utility function with $\gamma=2$ often requires $\beta>1$ to counteract the risk aversion effect from γ. Therefore, γ is already sufficiently high given the introduction of the probability weighting function. For this reason, we do not consider even higher γ in our empirical implementation. However, we observe that α is mainly responsible for matching the kernel's shape in the tails and the assumptions on γ have very small effect on it.

To see directly the nonparametric estimates of the probability weighting functions we select four typical shapes we observe in our sample and show them in Fig. 9. To illustrate these shapes, we pick the results from kernels estimated using 28 days to maturity and use power utility with risk aversion $\gamma=2$. Once again, the most frequently observed patterns is that of inverse S, shown on the top left panel. It implies overweighting of probabilities in the tails. As evident from the time series of fitted Prelec's α, this pattern

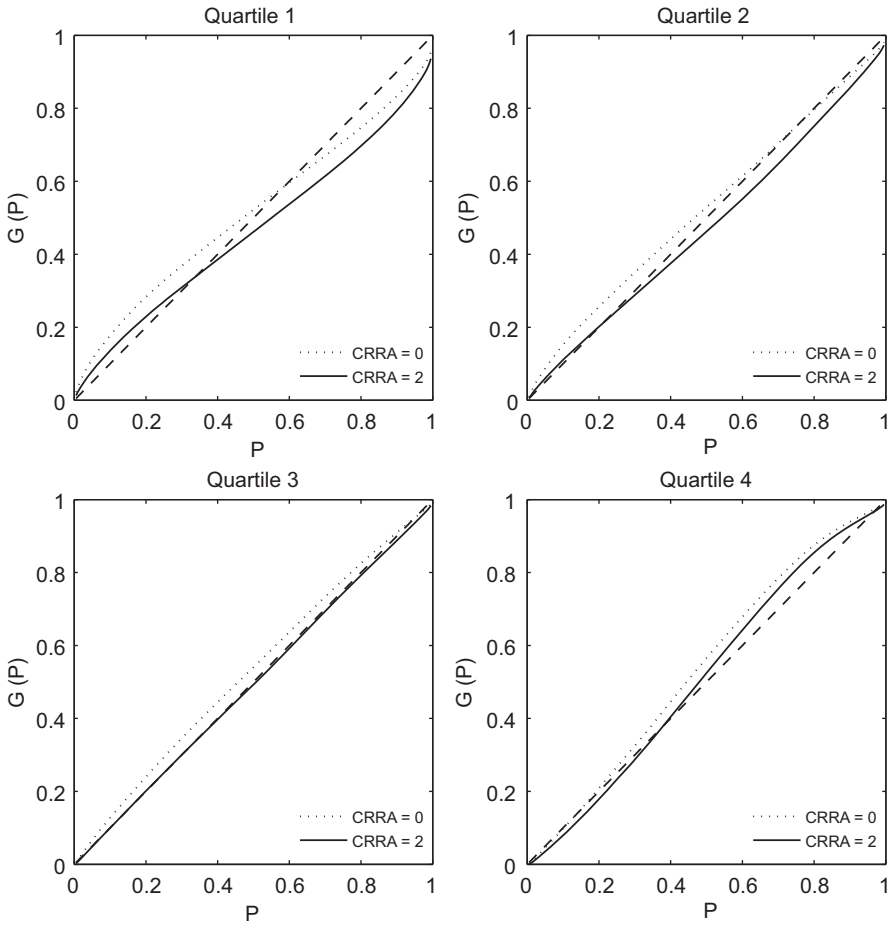

Fig. 12. Average weighting functions obtained from 45 days to maturity options. We present the estimates of the probability weighting functions average based on the quartiles of the Prelec α coefficient. The sample period is from January 1996 to December 2008. The maturities are 45 days. The relative risk aversion parameter is set to $\gamma=\{0,2\}$. We present the nonparametric estimates obtained from the constrained local linear estimator with bandwidth chosen based on the simulated samples of the out-of-the-money options prices.

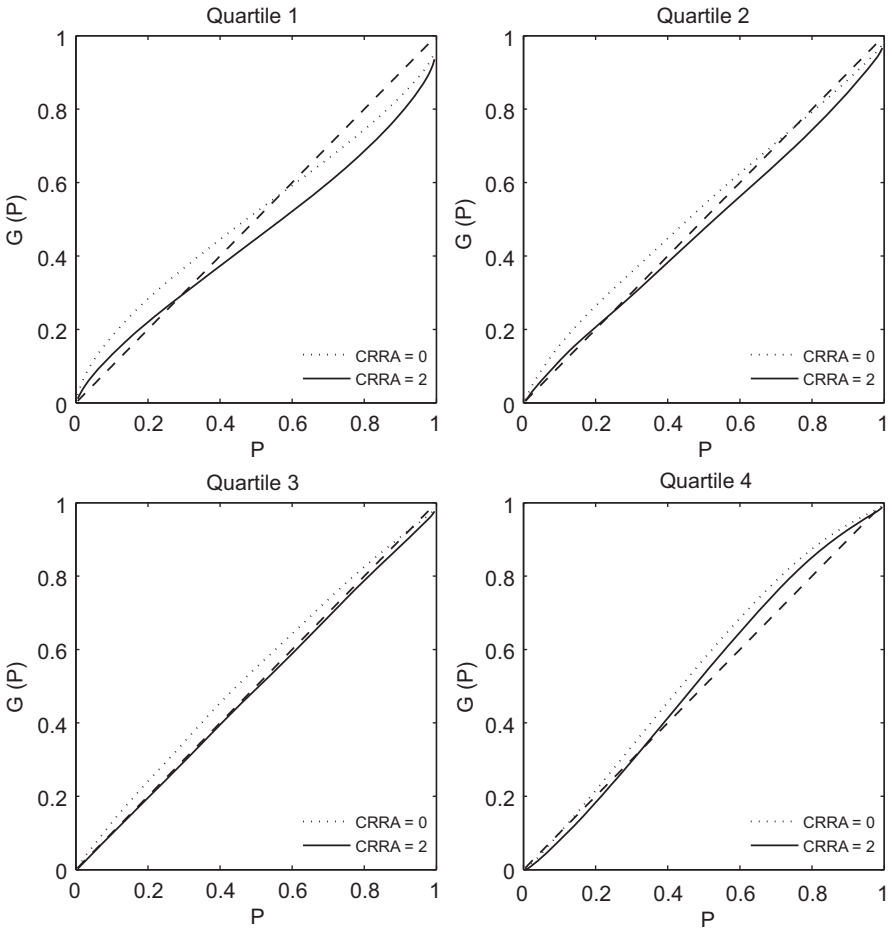

Fig. 13. Average weighting functions obtained from 56 days to maturity options. We present the estimates of the probability weighting functions average based on the quartiles of the Prelec α coefficient. The sample period is from January 1996 to December 2008. The maturities are 56 days. The relative risk aversion parameter is set to $\gamma = \{0,2\}$. We present the nonparametric estimates obtained from the constrained local linear estimator with bandwidth chosen based on the simulated samples of the out-of-the-money options prices.

is typically observed in mid- to late 1990s and in early 2000s and from second half of 2007–2008. During 2004–2006, probability weighting function estimates often resemble the S-shaped curve shown in the top right panel. These weights overweight the probabilities in the middle of the distribution and underweight the tails. For both inverse S- and S-shape forms, there is time variation in how pronounced they are. The lower panels of the figure show flatter versions of both shapes. Variation also exists in the relative size of the concave and convex segments. In a few instances, the weighting function shifts to predominantly convex or concave such that overweighting of probabilities is concentrated in only one of the tails.

We also compare probability weighting functions obtained by nonparametric estimates shown in Fig. 9 with the semi-nonparametric estimates obtained from A-type GCSE (AGCSE) and C-type GCSE (CGCSE).[18] We show the results in Fig. 10 and observe that qualitatively the shapes of the weighting functions are robust to the estimation method. For the reported results in the paper we use nonparametric estimates.

Finally, we also present the averages of nonparametric estimators of the weighting functions conditional on fitted Prelec's α. We use quartile cutoffs for α from Table 3 to group the weighting functions. The averages for each quartile for two risk aversion assumptions $\gamma = 0,2$ are presented in Figs. 11, 12, and 13 for 28-, 45-, and 56-day maturities, respectively. Comparing the average weighting functions under different risk-aversion assumptions (solid and dotted lines), we observe that higher risk aversion implies less pronounced overweighting on the left and more pronounced overweighting on the right side of the distribution. This is consistent with the substitution between risk aversion γ and Prelec's β we observe in Figs. 6–8. A related effect is also observed when we vary the risk aversion and compute the slopes of the probability weighting functions in Section 4.2.

Our results are generally consistent across different maturities we used to construct pricing kernels, although parameter approximations based on shorter maturity options

[18] We use the same 28-day options and $\gamma = 2$ for semi-nonparametric methods.

show somewhat higher volatility. Figs. 16–18 show that probability weights have the inverse S-shape in the majority of months. There are, however, instances, which are more frequent during 2004–2007, in which the probability weighting changes to the S-shape resembling that shown on top right panel in Fig. 9. During these months, $\alpha > 1$ implies underweighting in the tails of the distribution. Once again, this shape is not very sensitive to the assumption about value of γ. A casual observation suggests that the period with frequently observed S-shape probability weights coincides with the times of low volatility and abundant liquidity in the markets. Thus, probability weights during these times suggest relative complacency on the part of the investors with respect to the tail events. Notably, this phenomenon is not one-sided and applies both to the right and to the left tails of the index distribution. Also, as the financial crisis of 2008–2009 approaches, the shape of the probability weights begins to shift back to the inverse S. This time variation in the shape of the implied probability weights could prove useful in quantifying investor sentiment. Because the main purpose of our paper is to identify probability weights implied in option-based pricing kernels, we leave further investigation of time variation in weights for future work.

4.2. The slopes of the pricing kernel, expected returns on options, and the probability weighting functions

To further investigate the shape of the weighting functions, we construct alternative measures of the slope of the pricing kernel. We evaluate the shape of the kernel at different levels of moneyness and characterize its slope with respect to the cumulative physical distribution function.

We construct the slopes as follows. Given the returns distribution function under physical measure, $P(R)$, we define the slope via the area under the pricing kernel with respect to probability P. That is, for a given return R_0 and cumulative probability $P_0 = P(R_0)$, the area is $\int_0^{P_0} m(P)\,dP$ and the left slope is defined as $\int_0^{P_0} m(P)\,dP/P_0$. The right

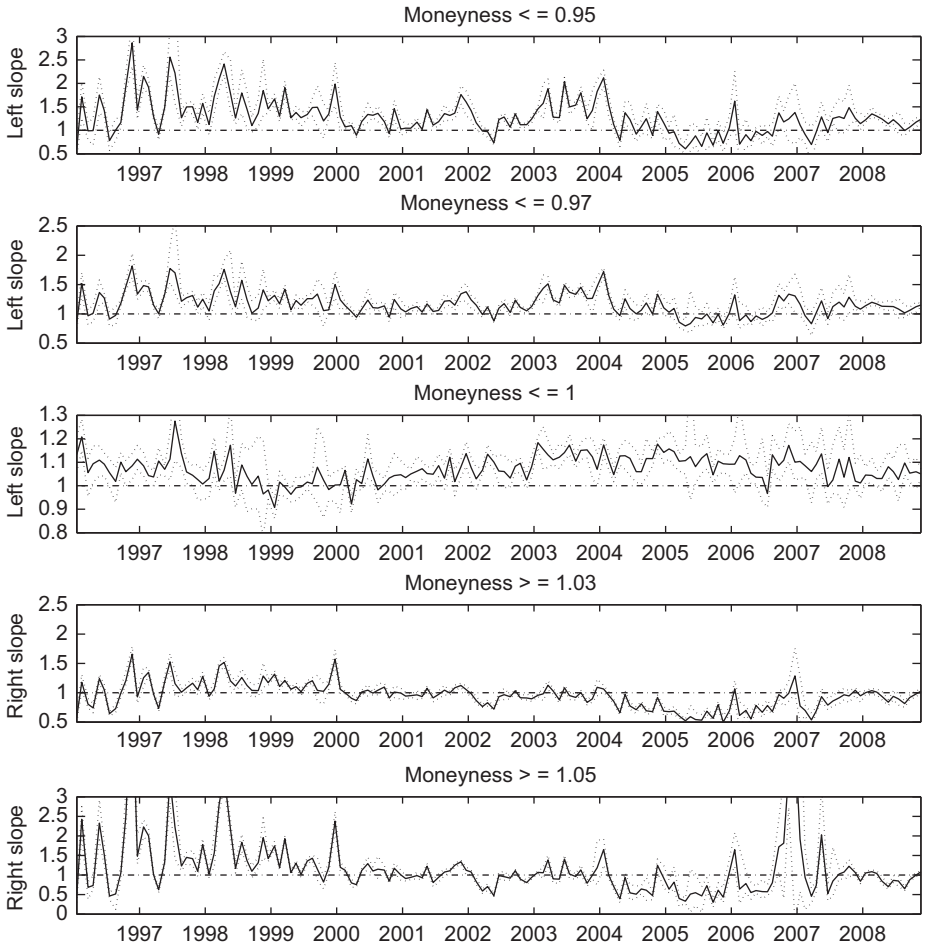

Fig. 14. The slope of the pricing kernel, 28-day options. We plot the slopes of the pricing kernel m obtained from 28-day options at different levels of moneyness. The sample period is from January 1996 to December 2008. Given the returns distribution function under physical measure, $P(R)$, we define the slope via the area under the pricing kernel with respect to probability P. That is, for a given return R_0 and cumulative probability $P_0 = P(R_0)$, the area is $\int_0^{P_0} m(P)\,dP$ and the (left) slope is defined as $\int_0^{P_0} m(P)\,dP/P_0$. The right slope is defined similarly as $\int_{P_0}^1 m(P)\,dP/(1-P_0)$. The pricing kernel is scaled so that $\int_0^1 m(P)\,dP = 1$. For the left slope, we choose the returns corresponding to the moneyness of 0.95, 0.97 and 1. For the right slope, we choose the returns corresponding to the moneyness of 1.03 and 1.05. The confidence intervals of the slope estimates, in dotted line, are obtained by bootstrap resampling of the out-of-the-money options.

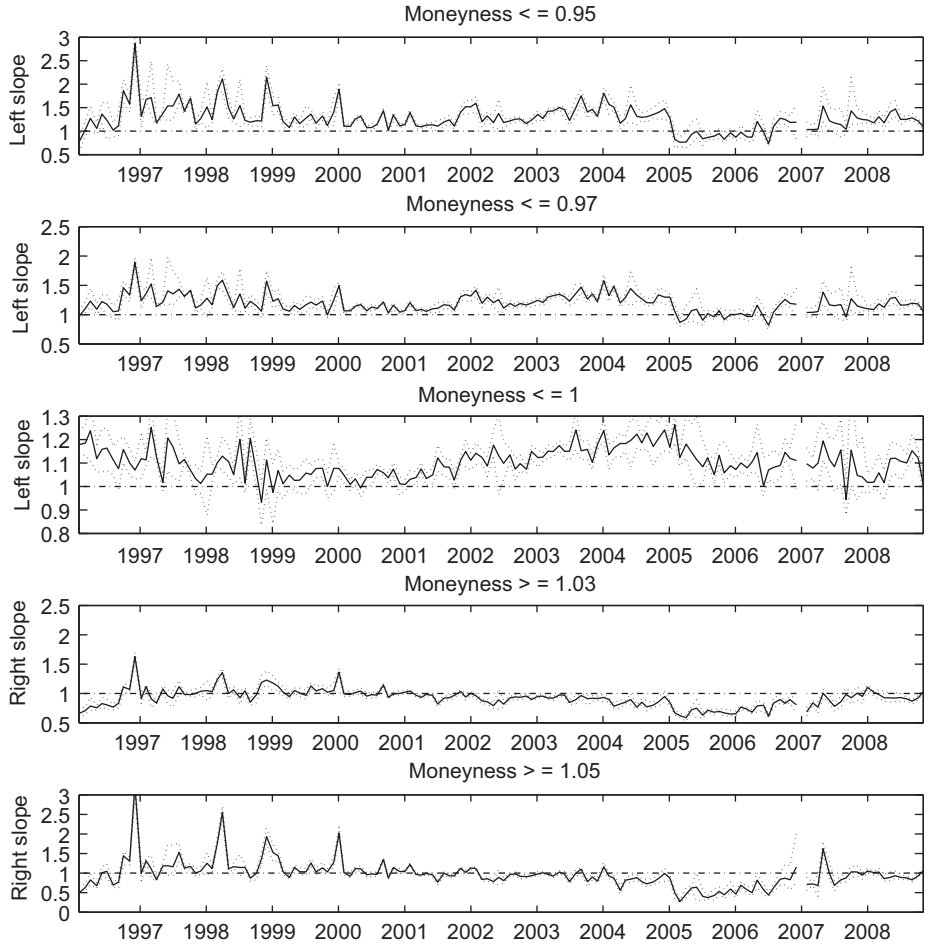

Fig. 15. The slope of the pricing kernel, 45-day options. We plot the slopes of the pricing kernel m obtained from 45-day options at different levels of moneyness. The sample period is from January 1996 to December 2008. Given the returns distribution function under physical measure, $P(R)$, we define the slope via the area under the pricing kernel with respect to probability P. That is, for a given return R_0 and cumulative probability $P_0 = P(R_0)$, the area is $\int_0^{P_0} m(P)\,dP$ and the (left) slope is defined as $\int_0^{P_0} m(P)\,dP/P_0$. The right slope is defined similarly as $\int_{P_0}^1 m(P)\,dP/(1-P_0)$. The pricing kernel is scaled so that $\int_0^1 m(P)\,dP = 1$. For the left slope, we choose the returns corresponding to the moneyness of 0.95, 0.97 and 1. For the right slope, we choose the returns corresponding to the moneyness of 1.03 and 1.05. The confidence intervals of the slope estimates, in dotted line, are obtained by bootstrap resampling of the out-of-the-money options.

slope is defined similarly as $\int_{P_0}^1 m(P)\,dP/(1-P_0)$. The pricing kernel is scaled so that $\int_0^1 m(P)\,dP = 1$. These definitions have an intuitive interpretation. We can write

$$\int_0^{P_0} m(P)\,dP = \int_0^{R_0} m(P)p\,dR = Q(R_0). \tag{30}$$

Thus, our definitions of the slopes correspond to the ratio of risk neutral to physical CDF in the left tail and the ratio of the risk-neutral and physical de-cumulative probabilities in the right tail.[19] The slopes measure how much risk neutral probability mass is concentrated in the tails relative to the underlying physical probability. A value above one corresponds to overweighting and below one corresponds to underweighting, and the value above one for the right tail measure indicates the U-shaped pricing kernel over some range of moneyness.

We construct the slopes corresponding to moneyness of 0.95, 0.97, and 1 for the left tail and 1.03 and 1.05 for the right tail. We use points on moneyness scale instead of cumulative probability because physical and risk neutral distributions are time varying while constant moneyness allows for comparison of slopes across different months. In Figs. 14, 15, and 16 we show the results from 28, 45 and 56 days to maturity options, respectively. Comparing the slopes across different moneyness, slopes further out-of-the money are typically higher and more volatile. Tails with moneyness below 0.95 or above 1.05 are generally overweighted with the exception of some periods mainly concentrated in 2004–2007. This is consistent with our results on the probability weighting

[19] These slopes can also be interpreted using the probability weighting function G corresponding to the constant marginal utility. In general, the pricing kernel is given by $m = u'(R)Z(P_R)$. If we set $u' = 1$, then $m = Z$ and we obtain

$$\int_0^{P_0} m(P)\,dP = \int_0^{P_0} Z(P)\,dP \equiv G(P_0).$$

Therefore, our pricing kernel slopes coincide with slopes of G defined as $G(P_0)/P_0$ on the left side and $(1-G(P_0))/(1-P_0)$ on the right side of the distribution.

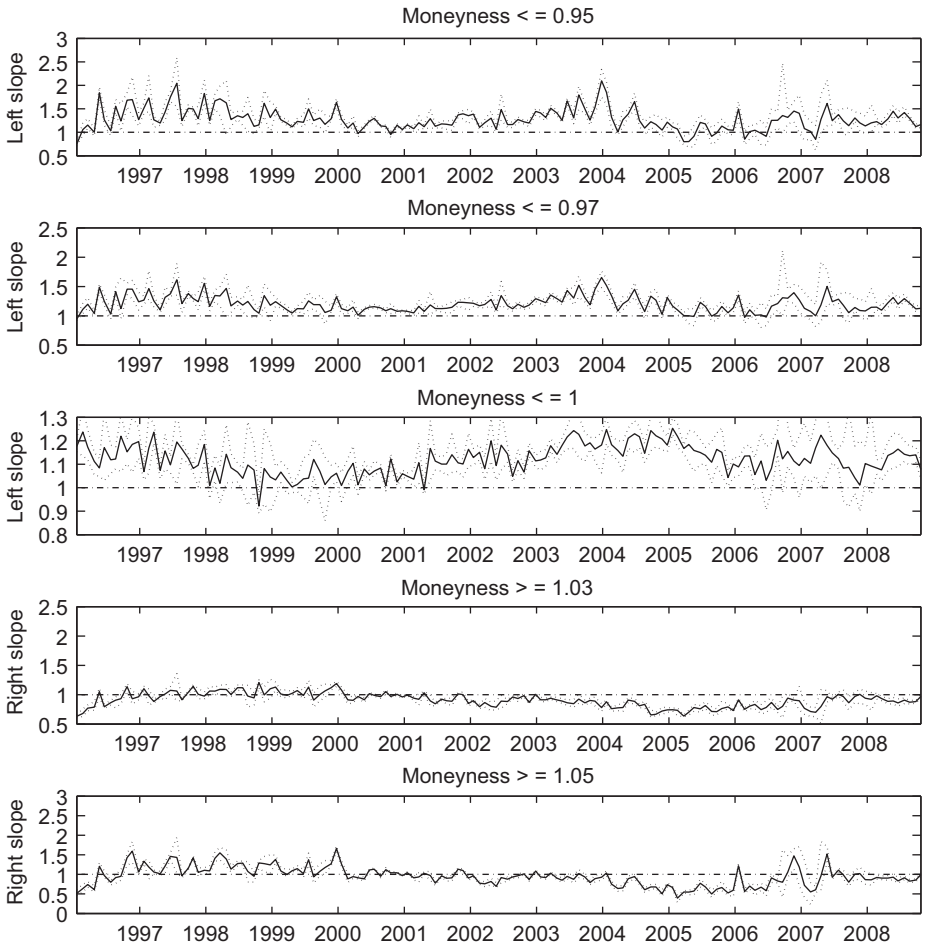

Fig. 16. The slope of the pricing kernel, 56-day options. We plot the slopes of the pricing kernel m obtained from 56-day options at different levels of moneyness. The sample period is from January 1996 to December 2008. Given the returns distribution function under physical measure, $P(R)$, we define the slope via the area under the pricing kernel with respect to probability P. That is, for a given return R_0 and cumulative probability $P_0 = P(R_0)$, the area is $\int_0^{P_0} m(P)\,dP$ and the (left) slope is defined as $\int_0^{P_0} m(P)\,dP/P_0$. The right slope is defined similarly as $\int_{P_0}^1 m(P)\,dP/(1-P_0)$. The pricing kernel is scaled so that $\int_0^1 m(P)\,dP = 1$. For the left slope, we choose the returns corresponding to the moneyness of 0.95, 0.97 and 1. For the right slope, we choose the returns corresponding to the moneyness of 1.03 and 1.05. The confidence intervals of the slope estimates, in dotted line, are obtained by bootstrap resampling of the out-of-the-money options.

functions, which mostly exhibit inverse S-shape but changes to S-shape in the years 2004–2007. Comparing the figures across maturities, we can also see that the pricing kernels based on longer maturity options exhibit less tail overweight and less volatility of the slopes.

We also construct slopes of the probability weighting function in the tails under the assumption $\gamma = 2$. Here we define the slopes following the same convention as for the pricing kernel: $G(P_0)/P_0$ for the left tail, and $(1-G(P_0))/(1-P_0)$ for the right tail. These results are shown in Figs. 17–19. Once again we observe that slopes tend to increase and often exceed one moving further out-of-the-money in both directions. This occurs even as we allow for standard risk aversion ($\gamma = 2$) and confirms our previous results that implied probability weighting functions tend to overweight the tails. When we assume concave u, the probability weighting slopes are lower on the left and higher on the right compared to their counterparts shown in Figs. 14–16. This is because standard marginal utility is decreasing and the probability weighting compensates for the remaining differences between u' and the empirical pricing kernel.

As discussed in Section 2.4, a U-shaped pricing kernel implies negative expected returns for contracts sufficiently out-of-the-money. We now consider quantitative implications of the pricing kernels with probability weighting function for expected returns of OTM contracts. We use the average of the physical density over our sample period to compute the implied (log of) expected returns for 28-day OTM puts and calls. To compute pricing kernel, we use Prelec two-parameter probability weighting function and assume linear utility. Table 4 presents the results for several values of the weighting function parameters. The inverse S-shaped probability weighting function implies substantially negative expected returns for OTM puts and calls. Furthermore, the expected returns values show wide variation with respect to the weighting function parameters α and β.

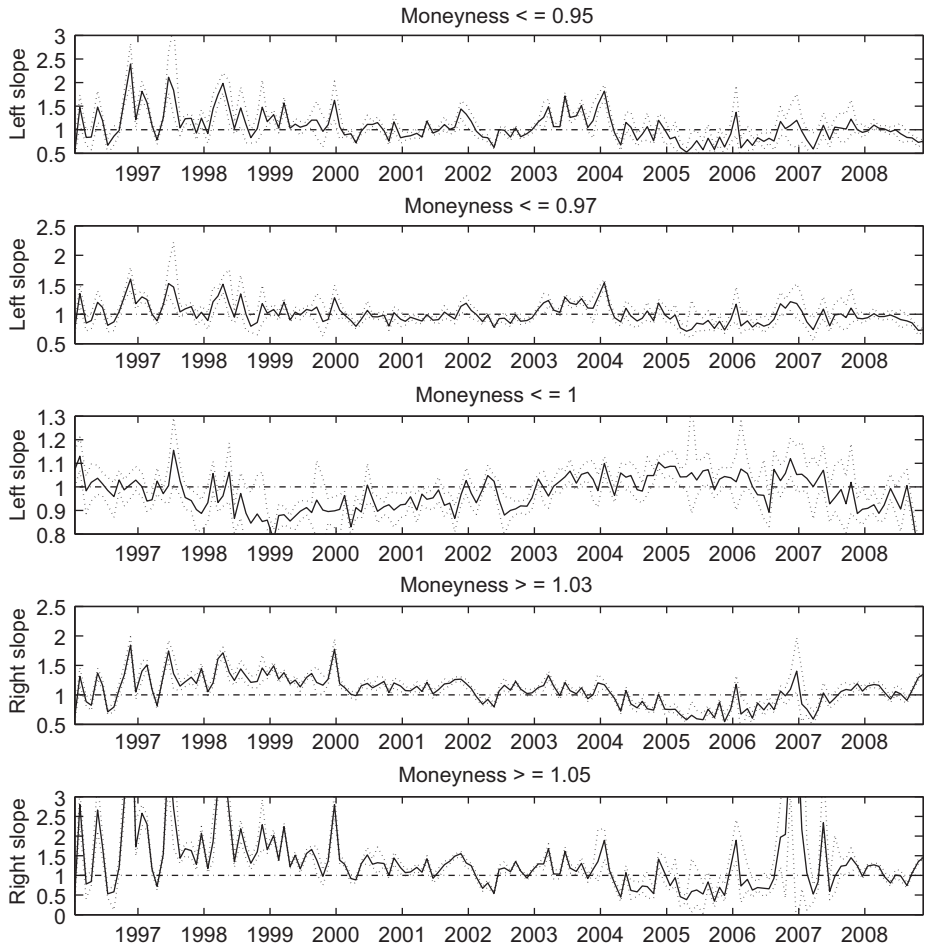

Fig. 17. The slope of the probability weighting function, 28-day options. We plot the slopes of the probability weighting function G estimated from 28-day options at various levels of moneyness. The sample period is from January 1996 to December 2008. The relative risk aversion coefficient is set to $\gamma = 2$. Given the returns distribution function under physical measure, $P(R)$, we define the slope via the area using the weighting function with respect to probability P. That is, for a given return R_0 and cumulative probability $P_0 = P(R_0)$, the area is $G(P_0)$ and the left slope is defined as $G(P_0)/P_0$. The right slope is defined similarly as $(1-G(P_0))/(1-P_0)$. For the left slope, we choose the returns corresponding to the moneyness of 0.95, 0.97, and 1. For the right slope, we choose the returns corresponding to the moneyness of 1.03 and 1.05. The confidence intervals of the slope estimates, in dotted line, are obtained by bootstrap resampling of the out-of-the-money options.

This suggests the model is flexible and can accommodate empirically estimated expected returns, such as those in Bakshi, Madan, and Panayotov (2010). The corresponding slopes of the pricing kernel, shown in the bottom half of the table, indicate that the implied risk-neutral probabilities in the right and left tails could be considerably overweighted relative to physical probabilities and the values of the slopes are consistent with our estimates shown in Fig. 14 for 28-day maturity options.

To summarize, our empirical results indicate that probability weighting functions implied in index options tend to overweight tail outcomes so that the weighting functions are frequently inverse S-shaped. The overweighting of the right tail is consistent with non-monotonic pricing kernel and implied negative risk aversion in the standard pricing kernel. There are also episodes when the probability weighting underweights the tails. Our results are consistent across a variety of measures of the shape of the weighting functions, both parametric and non-parametric, and across options of different maturity.

4.3. *Aggregate pricing kernel and individual preferences and beliefs*

Several possible interpretations can be made of the shape of the empirical pricing kernel. The kernel could be a compound result of aggregation of preferences or beliefs, or both. A standard representative agent argument can be applied within a recursive utility model to a general class of homogeneous certainty equivalents, such as RDEU, to obtain aggregate implications (see Epstein and Zin, 1989). Thus, a U-shaped pricing kernel could arise from individual risk preferences with probability weighting. Meanwhile, Bakshi and Madan (2007) construct an economy with heterogeneous beliefs in which

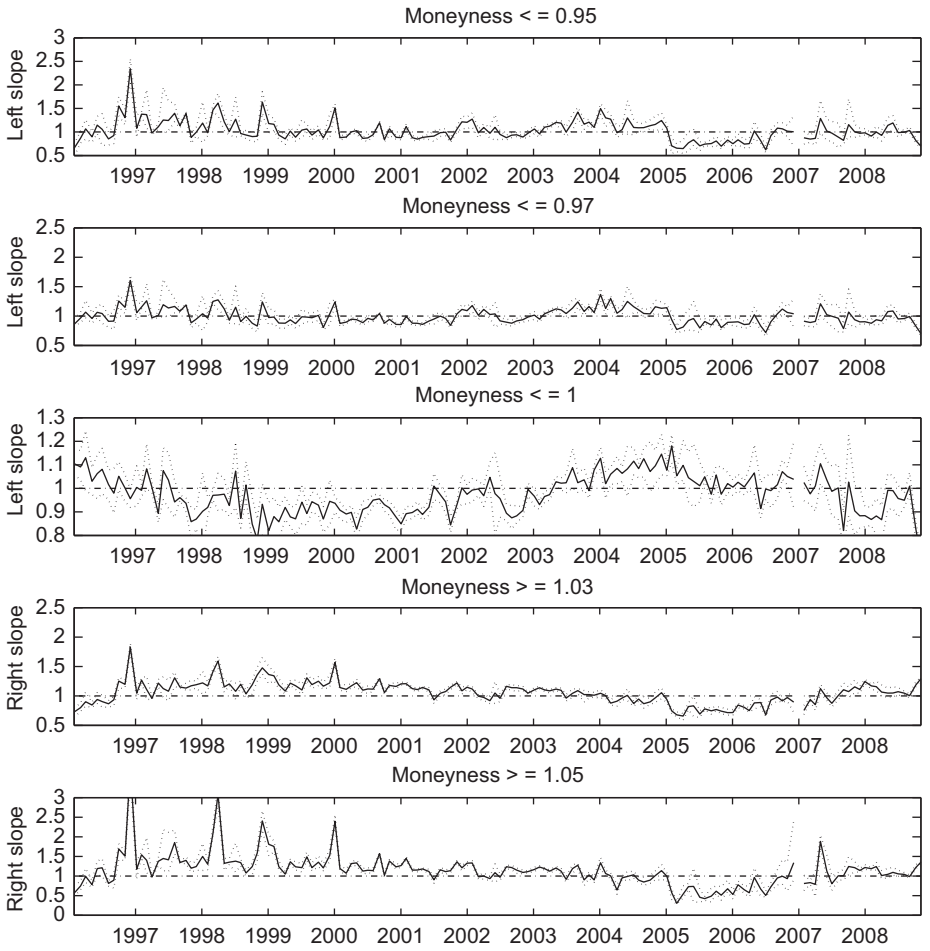

Fig. 18. The slope of the probability weighting function, 45-day options. We plot the slopes of the probability weighting function G estimated from 45-day options at various levels of moneyness. The sample period is from January 1996 to December 2008. The relative risk aversion coefficient is set to $\gamma = 2$. Given the returns distribution function under physical measure, $P(R)$, we define the slope via the area using the weighting function with respect to probability P. That is, for a given return R_0 and cumulative probability $P_0 = P(R_0)$, the area is $G(P_0)$ and the left slope is defined as $G(P_0)/P_0$. The right slope is defined similarly as $(1-G(P_0))/(1-P_0)$. For the left slope, we choose the returns corresponding to the moneyness of 0.95, 0.97, and 1. For the right slope, we choose the returns corresponding to the moneyness of 1.03 and 1.05. The confidence intervals of the slope estimates, in dotted line, are obtained by bootstrap resampling of the out-of-the-money options.

agents with standard EU preferences take different short and long positions that translate to the U-shaped aggregate kernel.[20] Aggregate preference parameters can in principle change over time either due to changes in investor composition or due to changes in the underlying individual preferences. Likewise, changes in aggregate beliefs can be driven from the individual level or from investor composition. Thus, we would find it hard to observe independent shocks to preferences or beliefs that would allow us to identify the factors affecting aggregate pricing kernel.

Empirically the two channels are not mutually exclusive, and the most likely scenario is that both preferences and beliefs affect risk-neutral probabilities. However, disentangling one from the other using only options prices and underlying index is not feasible. From the pricing perspective, the two approaches are observationally equivalent if we rely only on aggregate information. There are, however, differences in the implications of these two approaches at the individual portfolio level. Investors with the inverse S-shape probability weighting function would prefer portfolios with payoffs exposure to OTM calls and puts because they allow them to concentrate on outcomes in which they are most concerned or hopeful. This type of investors would not find linear short or long positions attractive. However, if the aggregate U-shaped pricing kernel is generated by a mix of investors with individually monotone discount factors but different beliefs, their portfolios would exhibit more linear payoffs as a function of the overall market. To evaluate the relative strength of these two alternative motivations, one would have to examine the time variation in the aggregate pricing kernel and link it to the changes in

[20] In a related result, Jouini and Napp (2009, Propositions 4 and 5) show that representative agent pricing kernel implies the inverse S-shaped probability weights when agents, in addition to beliefs, are heterogeneous along their discount factors.

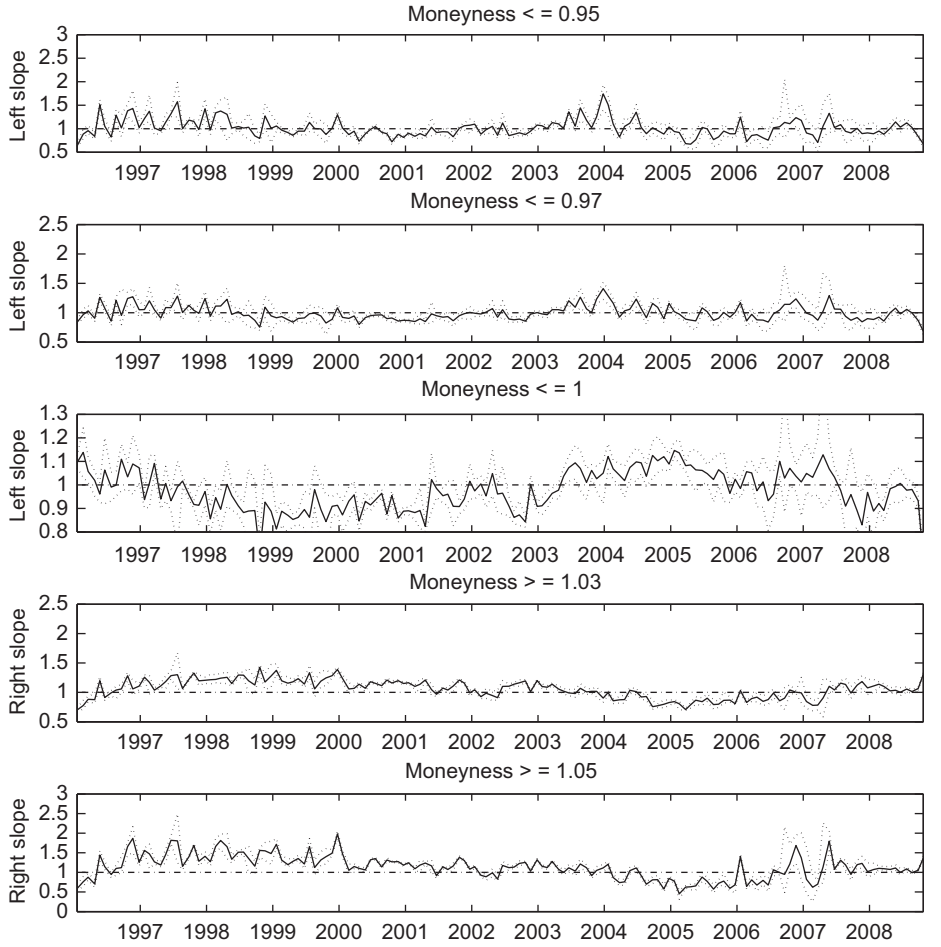

Fig. 19. The slope of the probability weighting function, 56-day options. We plot the slopes of the probability weighting function G estimated from 56-day options at various levels of moneyness. The sample period is from January 1996 to December 2008. The relative risk aversion coefficient is set to $\gamma = 2$. Given the returns distribution function under physical measure, $P(R)$, we define the slope via the area using the weighting function with respect to probability P. That is, for a given return R_0 and cumulative probability $P_0 = P(R_0)$, the area is $G(P_0)$ and the left slope is defined as $G(P_0)/P_0$. The right slope is defined similarly as $(1-G(P_0))/(1-P_0)$. For the left slope, we choose the returns corresponding to the moneyness of 0.95, 0.97, and 1. For the right slope, we choose the returns corresponding to the moneyness of 1.03 and 1.05. The confidence intervals of the slope estimates, in dotted line, are obtained by bootstrap resampling of the out-of-the-money options.

cross-sectional measures of nonlinearity of payoffs in portfolios of active institutional investors such as hedge funds and trading desks of investment banks.[21] While completely disaggregated data for such an exercise does not exist, collecting some information to investigate this issue could be possible. Such an exercise would also allow a better understanding of the sources of time variation in the aggregate pricing kernel. Other differences also could exist in the implications of the preferences and beliefs for the relation between volatility and time variation in risk premiums. However, a thorough investigation of these issues requires a formal dynamic model and goes beyond the framework of our paper. We leave this theme for future research along with other questions about the dynamics of the pricing kernel and probability weighting functions.

5. Conclusion

The properties of the empirical pricing kernels estimated from options prices are explained by the rank-dependent expected utility with concave utility and the inverse S-shaped probability weighting function. The weighting function overweights the tails of the distribution and is consistent with nonmonotonic pricing kernel. This type of weighting function is also consistent with a large experimental literature in economics. An important feature of our analysis is that we do not impose any a priori restrictions on the shape of the weighting functions, yet our estimates are often remarkably similar to those observed in various independent experiments. We also find that empirical probability weighting is time varying and at times could imply underweighted tails. It would be

[21] The majority of individual investors (by numbers) are holding plain vanilla portfolios which do not involve shorting, options, or other derivatives. Thus, for the purposes of identifying the impact of marginal investor on pricing kernel, one would have to look at the active institutional traders.

Table 4
The expected 28-day option returns and stochastic discount factor (SDF) slopes implied by Prelec probability weighting function.

We compute the expected returns and the SDF slopes based on the average physical density of the Standard & Poor's 500 index during our sample and assuming Prelec weighting function and linear utility function. OTM=out-of-the-money; ATM= at-the-money.

α	β	OTM put (0.95)	OTM put (0.97)	ATM put	OTM call (1.03)	OTM call (1.05)
Log expected return						
0.95	1.00	−19.1%	−13.9%	−7.4%	−10.2%	−14.1%
	0.95	−34.0%	−26.4%	−16.5%	−5.7%	−9.5%
	0.90	−49.3%	−39.3%	−26.0%	−1.0%	−4.5%
0.85	1.00	−59.3%	−44.2%	−24.9%	−31.5%	−43.4%
	0.95	−72.4%	−55.5%	−33.5%	−27.5%	−39.2%
	0.90	−85.7%	−67.1%	−42.5%	−23.3%	−34.9%
SDF slope						
0.95	1.00	1.07	1.03	0.99	1.05	1.09
	0.95	1.18	1.11	1.03	1.00	1.04
	0.90	1.30	1.19	1.08	0.96	0.99
0.85	1.00	1.23	1.10	0.98	1.14	1.28
	0.95	1.33	1.17	1.01	1.10	1.23
	0.90	1.44	1.24	1.05	1.06	1.18

interesting to investigate how this time variation is related to expected stock returns and volatility.

Our results confirm that probability weighting is an important and empirically relevant element for understanding asset prices. The utility with probability weighting has been applied to a wide range of problems in finance related to risk premium, cross section of expected returns, and portfolio diversification. Our findings provide encouraging nonexperimental evidence for the assumptions of investor behavior used in this literature.

Appendix A. Moments of the transformed distribution

In this appendix, we first construct approximations to moments for a general weighting function and distribution and then we proceed to specialized cases with clear comparative statics. We maintain the same notation as in the text: $\mu^{(n)}$ and $\kappa^{(n)}$ denote the uncentered nth moments of the random variables generated by P and $G(P)$ respectively. Also, $\mu^{(n)}_{r:s}$ denotes the nth (uncentered) moment of the order statistic r $(1\le r\le s)$ out of sample size s drawn from i.i.d. random variables generated by P. We begin by presenting a proof of Proposition 1.

Proof. Consider a Taylor series approximation up to the order K to $Z(P)=G'(P)$ at some point $\overline{P}\in(0,1)$ such that $Z^{(k)}(\overline{P})<\infty$ for $k\le K$:

$$Z(P)=\sum_{k=0}^{K}\frac{1}{k!}Z^{(k)}(\overline{P})\left(\sum_{j=0}^{k}(-1)^j C_k^j P^{k-j}\overline{P}^j\right)+\mathrm{o}[|P-\overline{P}|^K], \tag{A.1}$$

where $Z^{(k)}$ is kth order derivative and C_k^j is the combinatorial coefficient. We assume that K is sufficiently large and approximation is correspondingly sufficiently close so that the approximation to the density Z is always positive. Rearranging the indexing yields

$$Z(P)=\sum_{k=0}^{K}P^k a_k+\mathrm{o}[(P-\overline{P})^K],$$
$$a_k=\frac{1}{k+1}\sum_{j=k}^{K}\frac{(-1)^j}{(j-k)!}Z^{(j)}(\overline{P})C_j^{j-k}\overline{P}^j. \tag{A.2}$$

Using the definition of order statistic, we obtain

$$\int X^n(a_k(k+1)P^k)p\,dx = a_k\int(k+1)X^n pP^k\,dx = a_k\mu^{(n)}_{(k+1):(k+1)}. \tag{A.3}$$

Also, we can assume without loss of generality $\overline{P}=\frac{1}{2}$ and, therefore, the bound on the precision of approximation is

$$\left|\int X^n(P-\overline{P})^K\,dP\right|\le\int|X|^n|P-\overline{P}|^K\,dP \le 2^{-K}\mu^n(|X|), \tag{A.4}$$

where we use $\mu^n(|X|)$ to denote nth absolute moment.[22] Thus, when the nth moment exists, the approximation error converges to zero in K for a given n. Up to the approximation error, the moments of the transformed distribution are linear combinations of the moments of the maximum order statistic drawn from samples of size $1,\ldots,K+1$:

$$\kappa^{(n)}\approx\sum_{k=0}^{K}\int X^{(n)}(a_k(k+1)P^k)p\,dx = \sum_{k=0}^{K}a_k\mu^{(n)}_{(k+1):(k+1)}. \qquad \square \tag{A.5}$$

Now consider the density of the weighting function given in Eq. (3) with restriction that ϕ is an integer and $\phi\ge1$:

$$Z(P)=\lambda\phi P^{\phi-1}+(1-\lambda)\phi(1-P)^{\phi-1}. \tag{A.6}$$

In this case, we have the exact expression for the moments:

$$\kappa^{(n)}=\lambda\int X^n\phi P^{\phi-1}p\,dx+(1-\lambda)\int X^n\phi(1-P)^{\phi-1}p\,dx = \lambda\mu^{(n)}_{\phi:\phi}+(1-\lambda)\mu^{(n)}_{1:\phi}. \tag{A.7}$$

Thus, the moments are a weighted average of the minimum and maximum out of sample size ϕ. It is straightforward to substitute these expressions to obtain central moments, skewness, and kurtosis. We specialize in the case of symmetric distribution, i.e., $P(x)=1-P(-x)$, to obtain comparative statics behavior of skewness. Because for symmetric distributions we have an identity $\mu^{(n)}_{\phi:\phi}=(-1)^n\mu^{(n)}_{1:\phi}$, therefore

$$\kappa^{(n)}=\mu^{(n)}_{\phi:\phi}(\lambda+(1-\lambda)(-1)^n). \tag{A.8}$$

[22] For an arbitrary approximation point, the speed of convergence is governed by $\max(\overline{P},1-\overline{P})^K<1$. Thus, the speed of convergence is maximized by choosing $\overline{P}=0.5$.

The variance of the transformed distribution is given by

$$\sigma_G^2 = \kappa^{(2)} - (\kappa^{(1)})^2 = \mu_{\phi:\phi}^{(2)} - (\mu_{\phi:\phi}^{(1)})^2 (2\lambda - 1)^2. \quad (A.9)$$

The transformed variance can be higher or lower than the original, depending on the specifics of the underlying distribution P. One can show, for a general distribution, that the variance as a function of λ is maximized at the point of inflection in skewness or for a symmetric distribution at $\lambda = \frac{1}{2}$. For the skewness of the transformed symmetric distribution, we have the expression

$$SKEW_G = \frac{\kappa^{(3)} - 3\kappa^{(2)}\kappa^{(1)} + 2(\kappa^{(1)})^3}{(\kappa^{(2)} - (\kappa^{(1)})^2)^{(3/2)}}. \quad (A.10)$$

When $\lambda = \frac{1}{2} + \epsilon$, we obtain up to the first order in ϵ

$$SKEW_G \approx \frac{\mu_{\phi:\phi}^{(3)} - 3\mu_{\phi:\phi}^{(2)}\mu_{\phi:\phi}^{(1)}}{(\mu_{\phi:\phi}^{(2)})^{(3/2)}} \times \epsilon. \quad (A.11)$$

To see that the numerator of this expression is negative, we use approximations to the moments of order statistics from David (1970, p. 65); the original source is David and Johnson, 1954). The approximations are in terms of derivatives of the inverse CDF Q evaluated at the point $p_r = r/(n+1)$ [denoted as $Q_r \equiv Q(p_r)$]. In our case $r = n$, where n is the sample size (ϕ in the above cases). Using additional notation $q_r = 1 - p_r$ and $\sigma_r^2 = p_r q_r/(n+2)$, we have for the first three moments of order statistic r approximated up to σ_r^2:

$$\begin{aligned} \mu_{r:n}^{(1)} &\approx Q_r + \tfrac{1}{2}\sigma_r^2 Q_r'', \\ \mu_{r:n}^{(2)} &\approx Q_r^2 + \sigma_r^2 (Q_r')^2 + \sigma_r^2 Q_r Q_r'', \\ \mu_{r:n}^{(3)} &\approx Q_r^3 + 3\sigma_r^2 Q_r (Q_r')^2 + \tfrac{3}{2}\sigma_r^2 Q_r^2 Q_r''. \end{aligned} \quad (A.12)$$

Using these expressions and setting $n = \phi$ and $r = \phi$, we obtain

$$\mu_{\phi:\phi}^{(3)} - 3\mu_{\phi:\phi}^{(2)}\mu_{\phi:\phi}^{(1)} \approx -2Q_\phi^3 - \tfrac{3}{2}\sigma_\phi^2 Q_\phi^2 Q_\phi'' - \tfrac{1}{2}\sigma_\phi^2 Q_\phi'' EX_\phi^2 < 0. \quad (A.13)$$

The inequality follows because all derivatives involved are non-negative and the value of the inverse CDF is positive for $p_\phi = \phi/(\phi+1) > \frac{1}{2}$ (the expressions for the derivatives of the inverse CDF are provided in Arnold, Balakrishnan, and Nagaraja (1992, p. 131). Under some constraints on the extreme statistics moments, the derivative of skewness with respect to λ is negative at the skewness inflection point (with respect to λ) for a general nonsymmetric distribution. These constraints are not very intuitive, and we do not proceed to analyze them. We instead examine numerically the transformed skewness and other moments for two types of distributions. In the main text, we discuss the effects of transformations on standard normal and the physical distribution from the model with exponential moneyness-dependent volatility as in Aït-Sahalia and Duarte (2003).

Appendix B. Details of empirical method

In this appendix, we provide the technical details for estimating the risk-neutral densities.

B.1. Semi-nonparametric density estimation by Gram-Charlier series expansions

First, we denote the forward price $F_t(T)$ for the index. Define the $(T-t)$ period log-return of the underlying as $R_t(T) = \ln(F_T(T)/F_t(T))$; the risk-neutral moments, $\mu_{R,n} \triangleq E_t^Q[R_t^n(T)]$. Define the forward OTM call and put prices $\tilde{C}(F_t,X) \triangleq C(F_t(T),X)/D(t,T)$ and $\tilde{P}(F_t,X) \triangleq P(F_t(T),X)/D(t,T)$, where $D(t,T)$ is the risk-free discount factor between t and T. Following Rompolis and Tzavalis (2008), we have for $n \geq 2$

$$\begin{aligned} \mu_{R,n} = &\int_{F_t}^{+\infty} \frac{n}{X^2}\left[\ln\left(\frac{X}{F_t}\right)\right]^{n-2} \times \left[n-1-\ln\left(\frac{X}{F_t}\right)\right]\tilde{C}(F_t,X)\,dX \\ &+ \int_0^{F_t} \frac{n}{X^2}\left[\ln\left(\frac{X}{F_t}\right)\right]^{n-2} \times \left[n-1-\ln\left(\frac{X}{F_t}\right)\right]\tilde{P}(F_t,X)\,dX \end{aligned} \quad (B.1)$$

and

$$\mu_{R,1} = D(t,T) - \int_{F_t}^{+\infty} \frac{1}{X^2}\tilde{C}(F_t,X)\,dX - \int_0^{F_t} \tilde{P}(F_t,X)\,dX - 1. \quad (B.2)$$

In computing the above integrals, we follow prior research, for example, Jiang and Tian (2005), to apply curve-fitting method to the Black and Scholes implied volatility curve across strikes. For each day, we have the available strike prices between $X_{\min}$ and $X_{\max}$, and we choose the integral bounds between $0.9X_{\min}$ and $1.1X_{\max}$. The implied volatility curve outside the available strikes is set flat using the end point values.

Next, we use the estimates of the risk-neutral moments to construct the risk-neutral density based on the A-type and C-type Gram-Charlier series expansions. Define the standardized deviation of the return $\sigma_R = \sqrt{\mu_{R,2} - \mu_{R,1}^2}$. We then can approximate the density of R, $f_R(x)$ through series expansion. For the A-type Gram-Charlier series, we have

$$f_R(x) = N(x)\left[1 + \sum_{m=3}^{\infty} \frac{1}{m!} E_{R,m} H_m\left(\frac{x-\mu_{R,1}}{\sigma_R}\right)\right], \quad (B.3)$$

where $N(x)$ is the density for normal distribution with mean $\mu_{R,1}$ and standard deviation σ_R; $H_m(\cdot)$, the mth order Hermite polynomial; and the Hermite moments, $E_{R,m} = E(H_m((R-\mu_{R,1})/\sigma_R))$. For the C-type Gram-Charlier expansion, the risk-neutral density is defined as

$$f_R(x) = C \exp\left[\sum_{m=1}^{\infty} \frac{1}{m} \delta_m H_m\left(\frac{x-\mu_{R,1}}{\sigma_R}\right)\right], \quad (B.4)$$

where C is the normalizing constant and δ_m depends on the Hermite moments through a system of linear equations, as detailed in Rompolis and Tzavalis (2008). The polynomial order m in empirical implementation is chosen by using the information criteria, such as AIC or SC.

B.2. Nonparametric density estimation by constrained local polynomial method

Suppose we have a set of n observations of call prices and strike prices $\{y_i,x_i\}_{i=1}^n$. For the rest of this subsection, we assume that the observations are ordered by the first explanatory variable, i.e., $x_i \geq x_j$ if $i>j$, $1 \leq i,j \leq n$. As pointed out by Aït-Sahalia and Duarte (2003), the original data could violate the following slope and convexity constraints:

$$-1 \leq \frac{y_{i+1}-y_i}{x_{i+1}-x_i} \leq 0, \quad i=1,\ldots,n-1,$$
$$\frac{y_{i+1}-y_i}{x_{i+1}-x_i} \geq \frac{y_i-y_{i-1}}{x_i-x_{i-1}}, \quad i=2,\ldots,n-1. \tag{B.5}$$

As a result, nonparametric estimates of the call price function using the original data might not be arbitrage-free in finite samples. Aït-Sahalia and Duarte (2003) suggest the raw data should be first filtered by solving the following constrained optimization problem:

$$\min_m \sum_i (m_i-y_i)^2 \tag{B.6}$$

subject to the slope and convexity constraints. That is, if we replace y_i by m_i, the observations should satisfy the above shape restrictions. In the second stage of the procedure, we use the filtered data $\{m_i,x_i\}_{i=1}^n$ in our nonparametric estimation of the call price function.

In the second stage, we need to first determine the order of local polynomials in the nonparametric regression. Although asymptotically it is optimal to use a $(k+1)$th polynomial to estimate the kth derivative, Aït-Sahalia and Duarte (2003) show that the locally linear estimator has better finite sample performance for estimating risk-neutral densities. Therefore , for each given x, we choose $\beta_0(x)$ and $\beta_1(x)$ to minimize the following weighted sum of squared errors:

$$\sum_{i=1}^n \{m_i-\beta_0(x)-\beta_1(x)\times(x_i-x)\}^2 K_h(x_i-x), \tag{B.7}$$

where $K_h(\cdot)$ is a kernel function and h is the bandwidth. Consequently, the first and second partial derivatives of the estimated call price function $\widehat{m}(x)$ with respect to x are given as, respectively,

$$\frac{d\widehat{m}(x)}{dx} = \beta_1(x) \quad \text{and} \quad \frac{d\widehat{m}^2(x)}{dx^2} = \frac{d\beta_1(x)}{dx}. \tag{B.8}$$

In our implementation, we use the Gaussian kernel for the strikes to obtain smooth derivatives.

References

Ai, H., 2005. Smooth nonexpected utility without state independence. Unpublished working paper. University of Minnesota, Minneapolis, St. Paul, MN.

Aït-Sahalia, Y., Duarte, J., 2003. Nonparametric option pricing under shape restrictions. Journal of Econometrics 116, 9–47.

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

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

Allais, M., 1988. The general theory of random choices in relation to the invariant cardinal utility function and the specific probability function. In: Munier, B. (Ed.), Risk, Decision and Rationality. Dordrecht, the Netherlands, pp. 233–289.

Arnold, B.C., Balakrishnan, N., Nagaraja, H.N., 1992. A First Course in Order Statistics. John Wiley and Sons, New York, NY.

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

Bakshi, G., Kapadia, N., Madan, D., 2003. Stock return characteristics, skew laws, and the differential pricing of individual equity options. The Review of Financial Studies 16, 101–143.

Bakshi, G., Madan, D., 2000. Spanning and derivative-security valuation. Journal of Financial Economics 55, 205–238.

Bakshi, G., Madan, D., 2006. A theory of volatility spreads. Management Science 52, 1945–1956.

Bakshi, G., Madan, D., 2007. Investor heterogeneity, aggregation, and the non-monotonicity of the aggregate marginal rate of substitution in the price of market-equity. Unpublished working paper. University of Maryland, College Park, MD.

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

Barberis, N., Huang, M., 2008. Stocks as lotteries: the implications of probability weighting for security prices. American Economic Review 98, 2066–2100.

Barberis, N., Huang, M., Thaler, R.H., 2006. Individual preferences, monetary gambles, and stock market participation: a case of narrow framing. American Economic Review 96, 1069–1090.

Berns, G.S., Capra, C.M., Moore, S., Noussair, C., 2007. A shocking experiment: new evidence on probability weighting and common ratio violations. Judgment and Decision Making 2, 234–242.

Bollerslev, T., Wooldridge, J.M., 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. Econometric Reviews 11, 143–172.

Bondarenko, O., 2003. Why are put options so expensive?. Unpublished working paper. University of Illinois, Chicago, IL.

Breeden, D., Litzenberger, R., 1978. Prices of state contingent claims implicit in option prices. Journal of Business 51, 621–657.

Brown, D.P., Jackwerth, J.C., 2004. The pricing kernel puzzle: reconciling index option data and economic theory. Unpublished working paper. Universität Konstanz, Konstanz, Germany.

Camerer, C.F., 1995. Individual decision making. In: Kagel, J.H., Roth, A.E. (Eds.), Handbook of Experimental Economics, Princeton University Press, Princeton, NJ, pp. 587–703.

Camerer, C.F., Ho, T., 1994. Violations of the betweenness axiom and nonlinearity in probability. Journal of Risk and Uncertainty 8, 167–196.

Carlier, G., Dana, R.A., 2003. Core of convex distortions of a probability. Journal of Economic Theory 113, 199–222.

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

Chapman, D., Polkovnichenko, V., 2009. Preferences heterogeneity and asset markets outcomes. Journal of Finance 64, 1863–1887.

David, F.N., Johnson, N.L., 1954. Statistical treatment of censored data. Part i. Fundamental formulae. Biometrika 41, 228–240.

David, H.A., 1970. Order Statistics. Wiley, New York, NY.

Dierkes, M., 2009. Option-implied risk attitude under rank-dependent utility. Unpublished working paper. University of Münster, Münster, Germany.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987–1007.

Epstein, L.G., Zin, S.E., 1989. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework. Econometrica 57, 937–969.

Epstein, L.G., Zin, S.E., 1990. 'First-order' risk aversion and the equity premium puzzle. Journal of Monetary Economics 26, 387–407.

Fan, J., Gijbels, I., 1996. Local Polynomial Modeling and its Applications. Chapman and Hall, London, UK.

Fleming, J., Ostdiek, B., Whaley, R.E., 1996. Trading costs and the relative rates of price discovery in stock, futures, and option markets. Journal of Futures Markets 16, 353–387.

Härdle, W.K., Marron, S.J., 1991. Bootstrap simultaneous error bars for nonparametric regression. Annals of Statistics 19, 778–796.

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

Jarrow, R., Rudd, A., 1982. Approximate option valuation for arbitrary stochastic processes. Journal of Financial Economics 10, 349–369.

Jiang, G.J., Tian, Y.S., 2005. The model-free implied volatility and its information content. The Review of Financial Studies 18, 1305–1342.

Jouini, E., Napp, C., 2009. Behavioral properties of the representative agent. Unpublished working paper. Université Paris Dauphine, Paris, France.

Kliger, D., Levy, O., 2009. Theories of choice under risk: insights from financial markets. Journal of Economic Behavior and Organization 71, 330–346.

Levy, H., Levy, M., 2004. Prospect theory and mean–variance analysis. The Review of Financial Studies 17, 1015–1041.

Ljung, G.M., Box, G.E.P., 1978. On a measure of a lack of fit in time series models. Biometrika 65, 297–303.

Lopes, L.L., 1987. Between hope and fear: the psychology of risk. Advances in Experimental Social Psychology 20, 255–295.

Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59, 347–370.

Polkovnichenko, V., 2005. Household portfolio diversification: a case for rank-dependent preferences. The Review of Financial Studies 18, 1467–1502.

Prelec, D., 1998. The probability weighting function. Econometrica 66, 497–527.

Quiggin, J., 1982. A theory of anticipated utility. Journal of Economic and Behavioral Organization 3, 323–343.

Quiggin, J., 1987. Decision weights in anticipated utility theory. Journal of Economic Behavior and Organization 8, 641–645.

Quiggin, J., 1993. Generalized Expected Utility Theory: The Rank-Dependent Model. Kluwer Academics, Norwell, MA, Dordrecht, the Netherlands.

Rompolis, L.S., Tzavalis, E., 2008. Recovering risk neutral densities from option prices: a new approach. Journal of Financial and Quantitative Analysis 43, 1037–1053.

Rosenberg, J.V., Engle, R.F., 2002. Empirical pricing kernels. Journal of Financial Economics 133, 341–372.

Shefrin, H., 2001. On kernels and sentiment. Unpublished working paper. Santa Clara University, Santa Clara, CA.

Shefrin, H., 2005. A Behavioral Approach to Asset Pricing. Elsevier Academic Press, Burlington, MA.

Shefrin, H., Statman, M., 2000. Behavioral portfolio theory. Journal of Financial and Quantitative Analysis 35, 127–151.

Shoemaker, P., 1982. The expected utility model: its variants, purposes, evidence and limitations. Journal of Economic Literature 20, 529–563.

Starmer, C., 2000. Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk. Journal of Economic Literature 38, 332–382.

Tversky, A., Kahneman, D., 1992. Advances in prospect theory: cumulative representation of uncertainty. Journal of Risk and Uncertainty 5, 297–323.

Wu, C., 1986. Jackknife, bootstrap and other resampling methods in regression analysis. Annals of Statistics 14, 1261–1295.

Wu, G., Gonzalez, R., 1996. Curvature of the probability weighting function. Management Science 42, 1676–1690.

Yaari, M.E., 1987. The dual theory of choice under risk. Econometrica 55, 95–115.

Yatchewa, A., Härdle, W.K., 2006. Nonparametric state price density estimation using constrained least squares and the bootstrap. Journal of Econometrics 133, 579–599.

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