Semiparametric Estimation of Probability Weighting Functions Implicit in Option Prices*

H. Peter Boswijk
Amsterdam School of Economics
University of Amsterdam
and Tinbergen Institute

Jeroen Dalderop
Department of Economics
University of Notre Dame

Roger J. A. Laeven
Amsterdam School of Economics
University of Amsterdam, EURANDOM
and CentER

Niels Marijnen
Amsterdam School of Economics
University of Amsterdam
and Tinbergen Institute

This version: March 4, 2024

Abstract

In this paper, we develop semiparametric inference theory to disentangle and identify the probability weighting and utility functions implicit in option prices. Our approach relies on kernel estimation based on suitable probability integral transforms and profile maximum likelihood. We establish the asymptotic properties of our estimation procedure. Monte Carlo simulations show the good finite sample performance of our approach. Empirical results reveal the relevance of probability weighting. The shape of the probability weighting function implicit in S&P 500 index option prices over the period 1996–2023 is found to be inverse S-shaped and appears to be robust to the specification of the utility function of wealth.

Keywords: Probability weighting; Semiparametric inference; Kernel estimation; Profile likelihood; Options.

JEL Classification: Primary: C13; C58; Secondary: G13.

*This research was funded in part by the Netherlands Organization for Scientific Research under grant NWO Vici 2020–2027 (Laeven). *E-mail addresses*: H.P.Boswijk@uva.nl, jdalderop@nd.edu, R.J.A.Laeven@uva.nl, and N.Marijnen@uva.nl.

1 Introduction

The distinction between attitudes toward wealth and attitudes toward probabilities plays an important role in understanding risky decisions. Decision weights assigned to objective probabilities are offered as an explanation for a wide variety of behavioral characteristics in economics and finance, such as those embedded in insurance contracts, savings decisions, portfolio choices and asset prices. Probability weighting is therefore at the heart of the canonical non-expected utility models for decision under risk.

Option markets, trading large numbers of contracts with non-linear payoff structure and a wide variety of strike prices, provide a rich source of information on prices of risk and the risk attitudes that generate these prices. Econometric theory to disentangle attitude toward wealth from probability weighting in option prices is, however, limited.

In this paper, we develop semiparametric inference theory to identify the probability weighting function implicit in option prices. By exploiting the comprehensive distributional information embedded in a panel of option contracts we disentangle wealth and probability attitudes. The semiparametric approach we develop allows us to estimate the probability weighting function fully nonparametrically, while taking a parametric stance on wealth attitudes.

Different from the laboratory setting in which the experimentalist has full control over the probability distributions that subjects are evaluating, the risk-neutral distributions implicit in option prices are beyond the econometrician's control. This makes disentangling attitude toward wealth from probability weighting in the wealth-probability plane a challenging econometric problem.

Our approach relies on suitably defined probability integral transforms (PITs) involving the options' probability weighted return distributions, from which we build a nonparametric kernel-based estimator of the probability weighting function. This estimator depends on a parameter vector associated with the utility function of wealth, which is estimated via profile maximum likelihood. The inference theory we develop establishes the asymptotic properties of this semiparametric approach.

We demonstrate in Monte Carlo simulations that our semiparametric approach has good finite sample properties. Furthermore, a corresponding nonparametric bootstrap procedure is shown to yield accurate coverage rates.

An empirical analysis of a panel of S&P 500 index options spanning the period 1996–2023 reveals the importance of probability weighting. The probability weighting function takes an inverse S-shape and appears to be quite robust to the specification of the utility function of wealth.

1.1 Related literature

A large literature in economics seeks to measure wealth attitudes and probability weighting. The development of decision-theoretic models allowing for probability weighting (Quiggin, 1982; Tversky and Kahneman, 1992), to accommodate Allais (1953)-type behavior, has spurred the interest in the shape of the probability weighting function. The weighting function is perhaps most conveniently studied in a laboratory setting, where the researcher has full control over the set of alternatives that subjects can consider. The studies of, *inter alia*, Wu and Gonzalez (1996), Gonzalez and Wu (1999), and Abdellaoui (2000) show the presence of non-linear probability weighting, typically in the form of an inverse S-shaped function that is concave for small probabilities and convex for large probabilities.[1] This shape represents "diminishing sensitivity" in the probability domain, a psychological phenomenon that entails that decision-makers becomes less (more) sensitive to changes in objective probabilities away from (close to) the reference points of 0 and 1. We study probability weighting in real financial option prices.

In a discrete-time dynamic setting, the physical, or real-world, distribution of financial asset returns is generally estimated in one of two ways. Perhaps most popular are parametric time series specifications, such as modern versions of the family of GARCH models after Engle (1982) and Bollerslev (1986). However, fully specifying the functional form, as is required in such models, is rather restrictive, especially considering the aim of this paper. As an alternative, one may take a nonparametric approach. Nonparametric estimators are usually kernel-based, thus effectively constituting a smoothed historical distribution (e.g., Aït-Sahalia and Lo, 2000). However, the single-dimensional nature of return observations requires extensive historical data to conduct reliable inference on the current return distribution. Nonparametric estimators typically assume that historical returns are drawn from the same distribution as future returns, which is a stringent assumption. Kernel-based methods can incorporate conditioning variables, such as volatility, to explain time variation. This effectively leads to projected, conditional densities, which are not correctly specified forward-looking return densities if relevant variables are omitted. To avoid such concerns, we do not estimate the physical return distribution. Instead, we estimate its risk-neutral counterpart, from a panel of option prices.

The information in option contracts has been used to estimate Arrow-Debreu-like state prices since the seminal work of Breeden and Litzenberger (1978), who prove a simple nonparametric relation between prices of European-style option contracts and the risk-neutral density, also called the state price density. Inference techniques exploiting this relation have been developed in, among others, Aït-Sahalia and Lo (1998, 2000), Aït-Sahalia and Duarte (2003), and Dalderop (2020). Another notable nonparametric estimator is that of Lu and Qu (2021), who also use options to estimate this density, but do not rely on the Breeden and Litzenberger (1978) result; they fit a Hermite expansion by matching expansion-implied and observed prices. In a parametric setting, risk-neutral densities can be estimated by fitting model-implied prices to those observed in the market. Typically, price dynamics are then specified in a continuous-time setting. Popular models include those of Heston (1993), Bates (1996, 2000), Pan (2002) and Broadie et al. (2007). Most specifications fall into the generalized affine class of Duffie et al. (2000), for which semi-closed form expressions exist for the characteristic function, and by extension the risk-neutral density. Estimating a risk-neutral density from the observed options data then reduces to the estimation of the parameters in the dynamic specification, and the filtering of potential latent states. Parametrically modeling the option's risk-neutral distribution in discrete time is considerably less popular, but GARCH-like models have been developed (Duan, 1995; Heston and Nandi, 2000). We consider parametric as well as non-parametric estimation of the risk-neutral density.

The risk-neutral density contains comprehensive information for the estimation of attitudes toward wealth and toward probabilities. Indeed, option prices reflect the probabilistic beliefs and attitudes toward probabilities of investors, so that, in the presence of probability weighting, their implied risk-neutral cumulative distributions depend upon the probability weighting function. Of course, these risk-neutral distributions also embed attitudes toward wealth. Without probability weighting, Bliss and Panigirtzoglou (2004) and Liu et al. (2007) estimate the risk aversion parameter in an isoelastic utility function by maximum likelihood. Furthermore,

[1] For an inverse S-shaped probability weighting function, the local index of risk attitude (Eeckhoudt and Laeven, 2022) changes sign at the inflection point.

Aït-Sahalia and Lo (2000) consider a nonparametric estimator for the Arrow-Pratt (relative) risk aversion by taking ratios of estimators for the physical and risk-neutral densities and their derivatives. In this case, the risk aversion need not be constant over different states. A different strand of literature, starting with Rosenberg and Engle (2002), has directly estimated the pricing kernel. The Arrow-Pratt result exploited by Aït-Sahalia and Lo (2000) can be used to compute the risk aversion here as well, though now expressed in terms of the pricing kernel and its derivative. This has evolved to a point where the pricing kernel is explicitly allowed to depend on conditioning variables, such as volatility (Dalderop, 2021; Schreindorfer and Sichert, 2023). Though such kernels tend to provide a better fit than their unconditional counterparts, a foundation in economic theory is typically absent.

The development of such techniques, and the subsequent empirical analysis of option markets, has led to the discovery of seemingly paradoxical local risk-seeking behavior in option markets: the "pricing kernel puzzle". This phenomenon takes the form of a locally increasing stochastic discount factor, or pricing kernel, implying that the marginal value of an additional dollar *increases* as the state of nature improves — a fact not consistent with standard economic theory. Research has addressed this apparent inconsistency since its discovery (see Cuesdeanu and Jackwerth, 2018b, for a review). Linn et al. (2018), for instance, argue that this puzzle is actually a statistical error, caused by misaligned information sets, and show that it vanishes in their data when the model is correctly specified. However, Cuesdeanu and Jackwerth (2018a) and Dalderop (2021) confirm the existence of the puzzle while staying clear of the critique of Linn et al. (2018). A different solution is offered by Polkovnichenko and Zhao (2013), exploiting the rank-dependent utility model of Quiggin (1982), a cornerstone of prospect theory (Tversky and Kahneman, 1992). In this popular non-expected utility model, economic agents transform probabilities into decision weights before choosing their investment portfolio. Polkovnichenko and Zhao (2013) argue that the inflation of tail risk observed in this model helps explain the prizing kernel puzzle. The pricing kernel puzzle is not the only financial puzzle that can be explained using probability weighting; for example, Baele et al. (2019) show that probability weighting can be used to explain the variance premium, and the unexpectedly low returns on out-of-the-money options, and Barberis et al. (2016, 2021) use it to explain stock market anomalies.

Papers considering probability weighting functions implied by financial markets, including the aforementioned Polkovnichenko and Zhao (2013) and Baele et al. (2019), but also Kliger and Levy (2009), typically consider parametric utility as well as a parametric probability weighting function. Polkovnichenko and Zhao (2013) fix the utility function and fit a nonparametric probability weighting function, which they then try to match to the parametric specification of Prelec (1998), Baele et al. (2019) jointly estimate parametric utility and probability weighting functions using GMM, whereas Kliger and Levy (2009) do the same using nonlinear least squares. Their findings are generally in line with the experimental literature referenced above.

Different from the existing literature on probability weighting in financial markets, we consider joint estimation of a parametric utility function and a nonparametric probability weighting function. The resulting estimator is therefore semiparametric. Our estimation procedure hence generalizes earlier approaches, which require parameterizing the probability weighting function (Kliger and Levy, 2009; Baele et al., 2019), or *ex ante* fix the utility function (Polkovnichenko and Zhao, 2013). Our semiparametric approach builds on existing estimation theory of Andrews (1994) and Newey (1994), who provide general conditions for consistency and asymptotic normality. In general, a semiparametric approach is suitable when nonparametric identification is not possible or not feasible, e.g., in a computational sense, see Linton et al. (2008). In our setting, full nonparametric identification is not possible, as we can only identify the product of two functions;

separating the two requires some additional specification, in our case of the utility function.

Attitudes toward probabilities are arguably as fundamental to economic and financial decision-making under risk as attitudes toward wealth. Eeckhoudt and Laeven (2022) show, for instance, that probability weighting leads to a separate term in the well-known Arrow-Pratt measure of risk aversion. Relatedly, Eeckhoudt et al. (2020) show the importance of the signs of the derivatives of the probability weighting function, which can be economically interpreted as (probability) variants of higher-order risk concepts as prudence and temperance, the nonparametric analysis of which our approach makes possible.

1.2 Outline

This paper is organized as follows. In Section 2 we introduce the model framework. In Section 3 we develop our semiparametric inference theory. Section 4 describes the results of our Monte Carlo analysis. Our empirical results are in Section 5. Section 6 contains several concluding remarks. All proofs are in the Appendix.

2 Model Framework

This section discusses, in its simplest form, the rank-dependent utility framework to which our estimation strategy is applicable. Additionally, we present a (re)derived theoretical result that is of independent interest. The section concludes with a short discussion of potential adjustments to the simple model that still fit within the scope of the econometric theory developed in Section 3.

2.1 Rank-dependent utility and the risk-neutral density

Consider a complete, arbitrage-free financial market defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Denote by $F(r) := \mathbb{P}(R_1 \leq r)$ the cumulative distribution function of next-period returns R_1 under the probability measure $\mathbb{P}$ and by $f(r) := F'(r) = \frac{\mathrm{d}}{\mathrm{d}x}F(x)\big|_{x=r}$ the associated probability density function, assumed to exist.

Probability weighting is an important ingredient of both the rank-dependent utility (RDU) model of Quiggin (1982) and (cumulative) prospect theory of Tversky and Kahneman (1992), extensions of expected utility (EU) theory developed to address its empirical failure for decision under risk. These decision models can accommodate Allais (1953)-type behavior and related objective phenomena such as the common consequence and common ratio effects. Under RDU, a representative investor with (parametric) utility $u(\cdot;\gamma)$, henceforth assumed to be non-decreasing, continuously differentiable and concave, and next-period wealth $W_1 = R_1 W_0$, with $W_0 > 0$ the investor's initial wealth, evaluates her risk according to

$$V := \int_0^\infty u\left(rW_0;\gamma\right)\mathrm{d}\left(1 - Z(1 - F(r))\right) = \mathbb{E}^{\mathbb{P}}\left(u\left(W_1;\gamma\right)Z'(1 - F(R_1))\right), \tag{2.1}$$

with $Z : [0,1] \to [0,1]$ a non-decreasing function satisfying $Z(0) = 0$ and $Z(1) = 1$, referred to as a probability weighting function, and throughout assumed to be strictly increasing and continuously differentiable.[2] By allowing for probability weighting, outcomes are no longer weighted linearly in probabilities. The weighting of probabilities, e.g., by inflating tail probabilities, effectively transforms probabilities into decision weights, and helps explain empirical and experimental behavior in settings in which EU fails descriptively. Studies into the

[2] The non-decreasingness of the utility and probability weighting functions ensures that RDU is compatible with first-order stochastic dominance (Quiggin, 1982; Yaari, 1987).

shape of the weighting function (see, e.g., Wu and Gonzalez, 1996; Gonzalez and Wu, 1999; Abdellaoui, 2000) have indeed found evidence of non-linear weighting, with an inverse S-shaped function that is concave for small probabilities, and convex for large probabilities.[3] Prelec (1998) has axiomatized a popular functional form that is able to rationalize this behavior.

Suppose the investor has access to a risk-free asset with deterministic return R^0 and n risky assets with stochastic return R^i and allocates a fraction α_i of initial wealth to these $n+1$ assets, $i = 1, \ldots, n$. Hence, under full allocation, $R_1 = \sum_{i=0}^{n} \alpha_i R^i$ with $\alpha_0 = 1 - \sum_{i=1}^{n} \alpha_i$. We assume that the R^i have continuous distributions. Then, the first-order conditions for optimal allocation of initial wealth are given by (e.g., Ai, 2005)

$$\frac{\partial}{\partial \alpha_i} \mathbb{E}^{\mathbb{P}} \left(u \left(W_1; \gamma \right) Z'(1 - F(R_1)) \right) = \mathbb{E}^{\mathbb{P}} \left(u' \left(W_1; \gamma \right) Z'(1 - F(R_1))(R^i - R^0) \right) = 0, \qquad i = 1, \ldots, n. \tag{2.2}$$

The pricing kernel, or stochastic discount factor, m follows from the optimization problem.[4] The pricing kernel m expresses the relation between the physical probability measure $\mathbb{P}$ and an equivalent risk-neutral probability measure $\mathbb{Q}$, with associated cumulative distribution function by Q and density q. Specifically, we can write (Polkovnichenko and Zhao, 2013)

$$m = u'(W_1; \gamma) Z'(1 - F(R_1)) \tag{2.3}$$

$$q(R_1) = \frac{m}{\mathbb{E}^{\mathbb{P}}(m)} f(R_1) = \frac{u'(W_1; \gamma) Z'(1 - F(R_1)) f(R_1)}{\mathbb{E}^{\mathbb{P}} \left(u'(W_1; \gamma) Z'(1 - F(R_1)) \right)}. \tag{2.4}$$

We note that the pricing kernel is positive, induces a linear pricing rule, and hence is arbitrage-free. Thus, we have

$$q(R_1) \propto m f(R_1) = u'(W_1; \gamma) Z'(1 - F(R_1)) f(R_1),$$

and write, for later reference,

$$Z'(1 - F(R_1)) f(R_1) = c \frac{q(R_1)}{u'(W_1; \gamma)}, \tag{2.5}$$

where $c = \mathbb{E}^{\mathbb{P}} \left(u'(W_1; \gamma) Z'(1 - F(R_1)) \right)$.

Under CRRA, i.e., $u(w; \gamma) = w^{1-\gamma}/(1-\gamma)$, $\gamma \geq 0$, $\gamma \neq 1$, this simplifies to

$$q(R_1) = \frac{R_1^{-\gamma} Z'(1 - F(R_1)) f(R_1)}{\mathbb{E}^{\mathbb{P}} \left(R_1^{-\gamma} Z'(1 - F(R_1)) \right)}, \tag{2.6}$$

and, with additionally $W_0 = 1$,

$$m = R_1^{-\gamma} Z'(1 - F(R_1)). \tag{2.7}$$

[3] The inverse S-shape does not conform to second-order stochastic dominance: the RDU maximizer is (globally) strongly risk averse, in the sense of aversion to mean-preserving spreads à la Rothschild and Stiglitz (1970), and hence respects second-order stochastic dominance if and only if the utility function is concave and the probability weighting function is convex (see Yaari, 1986; Chew and Safra, 1987; Roëll, 1987; Ryan, 2006, and for higher order results, see Muliere and Scarsini, 1989; Eeckhoudt et al., 2020). For the corresponding results on local risk aversion under RDU, see Eeckhoudt and Laeven (2022).

[4] Assuming a complete, arbitrage-free market already ensures its existence.

2.2 From risk-neutral to physical distributions and densities

The physical distribution has a closed-form expression in terms of $q(r)$, which will later be presumed observable. In particular, consider the following string of identities:

$$1 - Z(1 - F(R)) = \int_0^{F(R)} Z'(1 - F)\,\mathrm{d}F = \int_0^R Z'(1 - F(r))f(r)\,\mathrm{d}r = c\int_0^R \frac{q(r)}{u'(rW_0;\gamma)}\,\mathrm{d}r, \tag{2.8}$$

using (2.5), where $c = \mathbb{E}^{\mathbb{P}}\left(u'(W_1;\gamma)Z'(1 - F(R_1))\right)$. The physical cumulative distribution $F(R)$ therefore equals

$$F(R) = 1 - Z^{-1}\left(1 - c\int_0^R \frac{q(r)}{u'(rW_0;\gamma)}\,\mathrm{d}r\right). \tag{2.9}$$

Differentiating with respect to R yields the physical density $f(R)$ as

$$f(R) = c\frac{q(R)}{u'(RW_0;\gamma)}Z^{-1\prime}\left(1 - c\int_0^R \frac{q(r)}{u'(rW_0;\gamma)}\,\mathrm{d}r\right). \tag{2.10}$$

Note that $Z^{-1\prime}(p) = \frac{1}{Z'(Z^{-1}(p))}$. Setting out to express the physical density in terms of the risk-neutral, the presence of the scaling constant c on the right hand side seems problematic, having been defined as an expectation under the physical measure. However, the identity $c \equiv \left(\int_0^\infty q(r)/u'(rW_0;\gamma)\,\mathrm{d}r\right)^{-1}$ can easily be shown (e.g., take the limit $R \to \infty$ in (2.9)), resolving this issue.

2.3 Primal and dual moments of the transformed distribution

As the analysis in Eeckhoudt and Laeven (2022) makes explicit, both the primal and the dual moments are of importance when evaluating risk premia within the RDU model. Denote by $\kappa^{(p)}$ the p-th moment of returns under the distribution $1 - Z(1 - P)$, and by $\mu^{(p)}$ its counterpart under P. Let $\mu_{r:(s)}$ denote the expectation of the r-th order statistic in a sample of size s drawn from P.

Proposition 1. *Let $\overline{P} \in (0,1)$. If $|Z^{(k)}(1-\overline{P})| \vee |\mu^{(p)}_{k:(k)}| < \infty$ for $k = 1,\dots,K+1$, and $\mathbb{E}^{\mathbb{P}}|R_1|^p < \infty$. Then, with $\overline{Z}(P) = 1 - Z(1-P)$ the dual weighting function, we have*

$$\kappa^{(p)} = \sum_{k=0}^{K} b_k \mu^{(p)}_{k+1:(k+1)} + o(2^{-K}),$$

where

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

This is a correction to Proposition 1 in Polkovnichenko and Zhao (2013), whose coefficients are incorrect.[5] The proof is identical otherwise.

[5] This can most easily be observed by example: using polynomial weighting $1 - Z(1-P) = P^n$ with $K = n-1$ should result in $\kappa = \mu_{n:(n)}$ for all n, and therefore in $b_k = \mathbb{1}_{(k=n-1)}$, which is not the case using the expression of Polkovnichenko and Zhao (2013). Some components in their expression also hint to this: $1/(j-k)!$ appears twice, and the possible products between P^k and $\overline{P}^j$ are wrong, as P^k can be multiplied by $\overline{P}^K$ for all k.

We are interested in how probability weighting affects the dual moments $\kappa_{n:(n)}$ under the risk-neutral measure, defined as:

$$\kappa_{n:(n)} := \mathbb{E}^{\mathbb{Q}}\left(\max\{R_1, \dots, R_n\}\right) = \int r \,\mathrm{d}(Q(r))^n.$$

The importance of the dual moments is revealed by Eeckhoudt and Laeven (2022). Using Proposition 1, we can write these risk-neutral dual moments as a linear combination of the physical dual moments, so long as we have linear utility. It follows directly by applying Proposition 1 to the weighting function $(1 - Z(1 - P))^n$. The required derivatives follow from Faà di Bruno's formula, the higher order chain rule. Note that, in contrast to the approximate result of Proposition 1, the following result is exact.

Corollary 1. *Suppose the assumptions of Proposition 1 hold for $K = n-1$. Then, for linear utility, we have*

$$\kappa_{n:(n)} = \sum_{k=0}^{n-1} b_k \mu_{k+1:(k+1)},$$

where, with $\mathcal{D}_k := \{\boldsymbol{d} \in \mathbb{N}_0^k : \sum_{i=1}^k i d_i = k\}$ and $d_0 := n - d_1 - \dots - d_k$,[6]

$$b_k = \frac{1}{k+1} \sum_{j=k}^{n-1} \frac{(-1)^{j-k}}{(j-k)!} \overline{P}^{j-k} \sum_{\boldsymbol{d} \in \mathcal{D}_k} \binom{n}{d_0, d_1, \dots, d_k} \overline{Z}(\overline{P})^{d_0} \prod_{j=1}^{k} \left(\frac{\overline{Z}^{(j)}(\overline{P})}{j!} \right)^{d_j}.$$

2.4 A comment on the generality of the model framework

The optimization problem (2.1) naturally extends to a dynamic setting, where the resulting investment decision is made on a period-by-period basis. Unfortunately, allowing for wealth-effects in the pricing kernel then causes nonstationarity in the time series, which complicates inference. Our framework thus requires the pricing kernel to be a function of returns. This seemingly limits the applicability of the remainder of this paper to isoelastic utility (also called power or CRRA), as other choices typically lead to a wealth effect in the pricing kernel. However, within these boundaries, there is room for some popular extensions of traditional utility specifications.

One such extension is the multiplicative habit formation model, introduced by Abel (1990). In such models, an investor does not derive utility from absolute wealth W_1 (or rather, W_{t+1} in a dynamic setting), but from wealth relative to some "habit", i.e., $W_1/\overline{W}$. The habit $\overline{W}$ can be exogenous or dependent on previous investment decisions of the representative agent. Clearly, if the habit-level is initial wealth, utility is derived from returns, and wealth would not factor into the pricing kernel. This can be extended to a specification as in Chen and Ludvigson (2009), where the habit is assumed to be a function of past returns, proportional to initial wealth. In the multiplicative habit model, this implies a pricing kernel that is a function of current and past returns, but not of wealth levels. So long as the relative wealth that enters the utility function is stationary, its functional form can be chosen freely.

The model can further be extended by allowing for external variables in the pricing kernel. A large literature

[6] With $B_{k,i}$ the Bell polynomial, this can be rewritten as

$$b_k = \frac{1}{(k+1)!} \sum_{j=k}^{K} \frac{(-1)^{j-k}}{(j-k)!} \overline{P}^{j-k} \sum_{i=1}^{k} \frac{n!}{(n-i)!} \overline{Z}(\overline{P})^{n-i} \cdot B_{k,i}\left(\overline{Z}'(\overline{P}), \overline{Z}''(\overline{P}), \dots, \overline{Z}^{(k-i+1)}(\overline{P})\right).$$

has directly specified flexible functional forms for a time-varying pricing kernel, starting from Rosenberg and Engle (2002). By modelling a pricing kernel as a function of returns, this is a projection of the pricing kernel onto the return of the considered asset, which has the same pricing implications for any asset whose payoff is determined by this return (Cochrane, 2009). In line with this literature on *ad hoc* pricing kernel specifications, we consider the following marginal utility function as a robustness check in our empirical analysis:

$$u'(R;\gamma) = \exp\left(\sum_{l=1}^{L} \gamma_l (\log R)^l\right).$$

This nests our baseline model, power utility, when $L = 1$. Though parametric, this more general utility specification should alleviate possible concerns that the results are spuriously driven by this baseline of isoelastic utility.

Another source of deviation between the pricing kernel and the marginal utility function occurs when investors have subjective and/or heterogeneous beliefs, as considered by Carr and Madan (2001) and Ziegler (2007). For example, Carr and Madan (2001) specify investors' subjective beliefs $f^s(R) = f(R)m^s(R)$ as multiplicative distorted versions of the true density $f(R)$, where $m^s(R)$ is for example a polynomial or exponential-polynomial function. In this case, the distortion function $m^s(R)$ shows up in the pricing kernel; see also Hens and Reichlin (2013).

More recent pricing kernel formulations allow dependence on conditioning variables, such as conditional volatility (Dalderop, 2021; Schreindorfer and Sichert, 2023). Gordon and St-Amour (2004) take a slightly more primitive approach, and specify a time-varying risk aversion. Practically, our extension of these frameworks would be the addition of probability weighting. In the current form of this paper, we do not consider this extension yet.

3 Estimation Theory

This section discusses the identification strategy of our estimation approach, and derives the relevant asymptotic properties. All proofs can be found in Appendix A.

3.1 From static optimization to dynamic markets

The investment horizon in the RDU model of (2.1) is a single period. Translating this to dynamic markets implies that investors reset their portfolio on a period-by-period basis, using the newly available information. This expanding information set is modeled by the filtration $\{\mathcal{F}_t\}$. Let $F_t(r) := \mathbb{P}(R_{t+1} \leq r|\mathcal{F}_t)$ be the conditional counterpart of $F(R_1)$ in (2.9), and similarly define the conditional density f_t based on (2.10), and the risk-neutral CDF Q_t and density q_t. Furthermore, for ease of exposition, define the dual probability weighting function $\overline{Z}(p) := 1 - Z(1-p)$ for $0 \leq p \leq 1$. Note that $\overline{Z}$ is itself a probability weighting function and that $\overline{Z}'(p) = Z'(1-p)$. Furthermore, we assume iso-elastic utility, so that we can disregard initial wealth W_t.

The conditional physical cumulative distribution and density function of the return are therefore modeled as

$$F_t(R;\gamma,\overline{Z}) = \overline{Z}^{-1}\left(c_t(\gamma)\int_0^R \frac{q_t(r)}{u'(r;\gamma)}\,\mathrm{d}r\right) \tag{3.1}$$

$$f_t(R;\gamma,\overline{Z}) = c_t(\gamma)\frac{q_t(R)}{u'(R;\gamma)}\overline{Z}^{-1\prime}\left(c_t(\gamma)\int_0^R \frac{q_t(r)}{u'(r;\gamma)}\,\mathrm{d}r\right), \tag{3.2}$$

where $c_t(\gamma) = \left(\int_0^\infty \frac{q_t(r)}{u'(r;\gamma)}\,\mathrm{d}r\right)^{-1}$.

Each period's investment decisions result in observable traded prices, in both equity and derivative markets. As such, the observation set available to the econometrician consists of return realizations, and risk-neutral densities (i.e., estimated from options): $\{R_{t+1}, q_t\}_{t=1}^T$.

3.2 Estimating the probability weighting function via PITs

For any γ, define the random variable $U_{t+1}(\gamma) \in (0,1)$ as

$$U_{t+1}(\gamma) := c_t(\gamma)\int_0^{R_{t+1}} \frac{q_t(r)}{u'(r;\gamma)}\,\mathrm{d}r. \tag{3.3}$$

Invoking integration by parts, the latter integral can be conveniently computed using the relation

$$\int_0^{R_{t+1}} \frac{q_t(r)}{u'(r;\gamma)}\,\mathrm{d}r = \frac{Q_t(R_{t+1})}{u'(R_{t+1};\gamma)} + \int_0^{R_{t+1}} Q_t(r)\frac{u''(r;\gamma)}{(u'(r;\gamma))^2}\,\mathrm{d}r.$$

The latter expression only requires the distribution function, which can be estimated more efficiently than its associated density.

Let γ_0 be the true parameter value. By (2.8), $U_{t+1}(\gamma_0) = \overline{Z}(F_t(R_{t+1}))$, where the probability integral transform $F_t(R_{t+1})$ follows a standard uniform distribution, denoted by $\mathcal{U}$. Therefore, monotonicity of $\overline{Z}$ implies

$$\mathbb{P}(U_{t+1}(\gamma_0) \le v) = \mathbb{P}(\overline{Z}(\mathcal{U}) \le v) = \mathbb{P}(\mathcal{U} \le \overline{Z}^{-1}(v)) = \overline{Z}^{-1}(v). \tag{3.4}$$

The structure of the PITs thus implies that the $U_{t+1}(\gamma_0)$ are independent copies of a random variable whose distribution function is the (inverse, dual) probability weighting function $\overline{Z}^{-1}$.

More generally, for any γ, define the CDF of the corresponding PITs $U_{t+1}(\gamma)$ as

$$G(v;\gamma) = \mathbb{P}(U_{t+1}(\gamma) \le v).$$

At the true parameter value, $G(v;\gamma_0) = \overline{Z}^{-1}(v)$. Given any γ, the natural estimator of G is the empirical CDF:

$$\widehat{G}(v;\gamma) = T^{-1}\sum_{t=1}^T \mathbb{1}(U_{t+1}(\gamma) \le v). \tag{3.5}$$

As such, what remains is estimating γ.

3.3 Profile likelihood estimator

Define the probability density $g(v;\gamma) := \frac{\partial}{\partial v}G(v;\gamma)$, and its shorthand $g_\gamma(v)$. For any γ, $g_\gamma(v)$ can be consistently estimated by the kernel estimator

$$\widehat{g}(v;\gamma) = T^{-1}\sum_{t=1}^{T} K_h(U_{t+1}(\gamma) - v), \tag{3.6}$$

where $K_h(\cdot) = \frac{1}{h}K\left(\frac{\cdot}{h}\right)$ for some kernel function K and bandwidth h.

Using the kernel density estimator, we estimate γ by maximizing the profile, or concentrated, likelihood:

$$\hat{\gamma} : = \arg\max_{\gamma\in\Theta} \ell_T(\gamma) := \arg\max_{\gamma\in\Theta} \ell_T(\gamma, \hat{g}(\cdot;\gamma)) \tag{3.7}$$

over some parameter space Θ, where

$$\begin{aligned}\ell_T(\gamma, g) &:= \frac{1}{T}\sum_{t=1}^{T} \log f_t(R_{t+1};\gamma, q_t, g) - \log q_t(R_{t+1}) \\ &= \frac{1}{T}\sum_{t=1}^{T} \log c_t(\gamma) - \log u'(R_{t+1};\gamma) + \log g\left(U_{t+1}(\gamma)\right)\end{aligned}$$

is the likelihood criterion (minus $\log q_t(R_{t+1})$, which does not depend on any parameter), and the second equation follows from conditional density specification (3.2).

Newey (1994) shows that in general the estimation method of the unknown density g is irrelevant if the limit of the estimator $\hat{g}$ maximizes the expected log-likelihood over the space of candidate densities. The following result confirms this is the case for the density function g_γ, which is the probability limit of our kernel density estimator under general conditions.

Lemma E. *For any γ, the density function $g_\gamma = \frac{\partial}{\partial v}G(v;\gamma)$ satisfies*

$$g_\gamma = \arg\max_{g\in\mathcal{G}} \mathbb{E}\left(\log f_t(R_{t+1};\gamma, g)\right), \tag{3.8}$$

where $\mathcal{G}$ is the space of all probability density functions on $[0,1]$.

This result implies that profile ML and ML identify the same functional parameter. In particular, this implies the profile maximum likelihood estimator $\hat{\gamma}$ reaches the semiparametric efficiency bound.

3.4 Identification of γ

The identification of γ thus depends on whether the profiled log-likelihood population criterion

$$\ell(\gamma) := \ell(\gamma, g_\gamma), \tag{3.9}$$

stated in terms of the long-term mean of the unprofiled log-likelihood function

$$\ell(\gamma, g) := \mathbb{E}\left(\log c_t(\gamma) - \log u'(R_{t+1};\gamma) + \log g\left(U_{t+1}(\gamma)\right)\right),$$

has a unique maximizer γ_0.

For any γ,

$$\begin{aligned}\ell(\gamma)-\ell(\gamma_0) &= \mathbb{E}\left(\log\frac{f_t(R_{t+1};\gamma,g_\gamma)}{f_t(R_{t+1};\gamma_0,g_{\gamma_0})}\right)\\ &= \mathbb{E}\left(\mathbb{E}_t\left(\log\frac{f_t(R_{t+1};\gamma,g_\gamma)}{f_t(R_{t+1};\gamma_0,g_{\gamma_0})}\right)\right)\\ &\le \mathbb{E}\left(\log\mathbb{E}_t\left(\frac{f_t(R_{t+1};\gamma,g_\gamma)}{f_t(R_{t+1};\gamma_0,g_{\gamma_0})}\right)\right)\\ &= 0,\end{aligned}$$

where the Jensen's inequality in the third step holds with equality if and only if

$$f_t(R_{t+1};\gamma,g_\gamma) = f_t(R_{t+1};\gamma_0,g_{\gamma_0}) \quad \text{a.s.} \tag{3.10}$$

This is a trivial identification condition: no two parameters can imply the same distribution. Let $f_t(r) = f_t(r;\gamma_0,g_{\gamma_0})$ and $F_t(r) = F_t(r;\gamma_0,g_{\gamma_0})$ be shorthand for the true conditional density and cumulative distribution functions. The following assumption assures that (3.10) cannot hold for any $\gamma\neq\gamma_0$, so that the Jensen's inequality holds strictly.

Assumption I.

(i) *For any $\gamma\neq\gamma_0$, there exists two possible outcomes (r_1,r_2) such that $\mathbb{P}\left(f_t(r_j)>0\right)>0$ for $j=1,2$ and $\frac{u'(r_1;\gamma)}{u'(r_2;\gamma)}\neq\frac{u'(r_1;\gamma_0)}{u'(r_2;\gamma_0)}$.*

(ii) *$G_\gamma(u)$ is strictly monotonic in u for all $\gamma\in\Theta$.*

Intuitively, I*(i)* requires that the parameter γ changes the shape of the marginal utility function u', i.e., it cannot be a multiplicative constant. Meanwhile I*(ii)* requires that $g_\gamma(u)>0$ for $u\in(0,1)$ for all $\gamma\in\Theta$.

Lemma I. *Under Assumption I, $\ell(\gamma)$ is uniquely maximized at γ_0.*

3.5 Consistency

Besides identification, the consistency of $\hat\gamma$ requires the uniform convergence of $\ell_T(\gamma)$ to $\ell(\gamma)$ over the parameter space Θ. If g_γ were known, this would follow from the uniform convergence of $\ell_T(\gamma,g_\gamma)$ to $\ell(\gamma,g_\gamma)$. In addition, we require the uniform convergence of the density estimator to $\hat g_\gamma$ to g_γ. This depends on the properties of the PIT functional

$$U(R,q;\gamma) := c(\gamma)\int_0^R \frac{q(r)}{u'(r;\gamma)}\,\mathrm{d}r.$$

In particular, we establish consistency under the following assumptions on the log-likelihood, which are similar to those in Newey (1994, Lemma 5.2):

Assumption C. *There are a constant $\varepsilon>0$, norm $\|g\|$, and integrable functions $b(R,q),\tilde b(R,q)>0$ such that*

(i) *for all γ in the compact parameter space Θ, $\log f(R;q,\gamma,g_\gamma)$ is continuous at γ almost surely, and $|\log f(R;q,\gamma,g_\gamma)-\log q(R)|\leqslant b(R,q)$*

(ii) $|\log g(U(R,q;\gamma))-\log g_0(U(R,q;\gamma))|\leqslant \tilde b(R,q)\,\|g-g_0\|^{\varepsilon}$

(iii) $\ell_T(\gamma) \xrightarrow{p} \ell(\gamma)$ *pointwise and* $\sup_{\gamma\in\Theta} ||\hat{g}_\gamma - g_\gamma||^\varepsilon \xrightarrow{p} 0$.

Proposition C. *Suppose that Assumptions I and C hold. Then,* $\hat{\gamma} \xrightarrow{p} \gamma_0$ *when* $T \to \infty$.

Assumption C contains some high-level regularity conditions that can be broken down into more primitive ones. For example, the last part of C*(i)* follows from uniform moment conditions on the log pricing kernel. The law of large numbers in C*(iii)* allows for general forms of time series dependence, such as mixing or ergodicity.

3.6 Asymptotic normality

Define the score of the profile likelihood function

$$m(R_{t+1}, q_t, \gamma, g) := \frac{\partial}{\partial\gamma} \log f_t(R_{t+1}; \gamma, g) = \frac{c_t'(\gamma)}{c_t(\gamma)} - \frac{\partial u'(R_{t+1};\gamma)/\partial\gamma}{u'(R_{t+1};\gamma)} + \frac{\frac{\partial}{\partial\gamma} g_\gamma\left(U_{t+1}(\gamma)\right)}{g_\gamma\left(U_{t+1}(\gamma)\right)}$$

in terms of the profiled density function $g_\gamma(u)$. The F.O.C. of problem (3.7) then equals

$$\hat{m}_T(\gamma) := \frac{1}{T}\sum_{t=1}^{T} m(R_{t+1}, q_t, \gamma, \hat{g}) = 0,$$

where $\hat{g}$ is given in (3.6). This estimator belongs to the class of semiparametric M-estimators considered in Newey (1994). The latter's Theorem 2.1 provides high-level regularity conditions such that the estimation error takes the form

$$\sqrt{T}\,(\hat{\gamma} - \gamma_0) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \psi(R_{t+1}, q_t) + o_p(1),$$

for some influence function $\psi(\cdot)$ with $\mathbb{E}\left(\psi(R_{t+1}, q_t)\right) = 0$ and $\mathbb{V}\mathrm{ar}\left(\psi(R_{t+1}, q_t)\right) < \infty$. Moreover, its equation (3.10) states that

$$\psi(R_{t+1}, q_t) = -M^{-1}\left(m(R_{t+1}, q_t, \gamma_0, g_0) + \alpha(R_{t+1}, q_t)\right),$$

assuming $M := \partial\mathbb{E}\left(m(R_{t+1}, q_t, \gamma, g_0)\right)/\partial\gamma\big|_{\gamma=\gamma_0}$ is non-singular, where $\alpha(R_{t+1}, q_t)$ is the *pathwise derivative* of $m(R, q, \gamma_0, g)$ in g at g_0. Fortunately, Newey (1994, eqn. 3.14) shows that $\alpha(R_{t+1}, q_t) = 0$ when the limit of $\hat{g}_\gamma$ converges to the population likelihood maximizer, which in our case is g_γ in (3.8).

To verify this result in our setting, note that the unknown density function g only enters the score of the likelihood at γ_0 via the following score functional:

$$S(R, q, g) := \frac{\frac{\partial}{\partial\gamma} g_\gamma\left(U(R, q; \gamma)\right)\big|_{\gamma=\gamma_0}}{g_{\gamma_0}(U_0(R, q))} = \frac{\dot{g}_{\gamma_0}(U_0(R, q)) + g_{\gamma_0}'(U_0(R, q))\dot{U}_0(R, q)}{g_{\gamma_0}(U_0(R, q))},$$

where $\dot{f}_\gamma(u) := \frac{\partial}{\partial\gamma} f_\gamma(u)$ and $f_\gamma'(u) := \frac{\partial}{\partial u} f_\gamma(u)$ for any function f, and $U_0(\cdot) = U(\cdot; \gamma_0,)$. The following conditions, adapted from Newey (1994, Assumptions 5.1–5.3), ensure that the nonparametric estimation of g does not affect the limiting distribution of $\hat{\gamma}$.

Assumption A. *There is a function* $D(R, q, g)$ *that is linear in* g *such that:*

(i) for all g *with* $\|g - g_0\|$ *small enough,*

$$\|S(R,q,g)-S(R,q,g_0)-D(R,q,g-g_0)\| \leq b(R,q)\|g-g_0\|^2 \qquad \textit{(asymptotic linearity)}$$

(ii) $\mathbb{E}(b(R_{t+1},q_t))\sqrt{T}\|\hat{g}-g_0\|^2 \xrightarrow{p} 0$ *($T^{-1/4}$ rate of nonparametric estimator)*

(iii) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(D(R_{t+1},q_t,\hat{g}-g_0)-\int D(R_{t+1},q_t,\hat{g}-g_0)\,\mathrm{d}F_0(R_{t+1},q_t)\right) \xrightarrow{p} 0$ *(stochastic equicontinuity)*

Lemma A. *Under Assumption A, when $T \to \infty$,*

$$\sqrt{T}\hat{m}_T(\gamma_0) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} m(R_{t+1},q_t,\gamma_0,g_0) + o_p(1).$$

The functional S only depends on the bivariate function $g(\gamma,u)$ through the univariate functions $g_{\gamma_0}(\cdot) = g(\cdot;\gamma_0)$, $\dot{g}_{\gamma_0}(u) = \frac{\partial}{\partial\gamma}g_\gamma(u)\big|_{\gamma=\gamma_0}$, and $g'_{\gamma_0}(u) = \frac{\partial}{\partial u}g_{\gamma_0}(u)$. Similarly, the moment functional m could be written in terms of the univariate functions $(g_\gamma, \dot{g}_\gamma, g'_\gamma)$, as done in Linton et al. (2008). We could therefore follow the latter's definition of the norm for g as

$$\|g\| = \sup_{\gamma\in\Gamma}\max\{\|g_\gamma\|_2, \|\dot{g}_\gamma\|_2, \|g'_\gamma\|_2\},$$

where $\|f\|_2 = \left(\int f(u)^2 du\right)^{1/2}$ is the L^2-norm for univariate functions. Linton et al. (2008, Lemma A.8) establish the asymptotic linearity condition A *(i)* under this norm for their profile likelihood estimator. In particular, consider $D(R,q,g) = D_S(R,q,g_0)[g]$, where D_S is the pathwise derivative of S in g, defined for any direction $\bar{g}-g$ as

$$\begin{aligned}
D_S(R,q,g)[\bar{g}-g] &:= \lim_{\tau\to 0}\frac{1}{\tau}\left(S(R,q,g+\tau(\bar{g}-g))-S(R,q,g)\right)\\
&= \frac{\partial}{\partial\tau}\frac{(1-\tau)\dot{g}_{\gamma_0}(U_0)+\tau\dot{\bar{g}}_{\gamma_0}(U_0)+\left((1-\tau)g'_{\gamma_0}(U_0)+\tau\bar{g}'_{\gamma_0}(U_0)\right)\dot{U}_0}{(1-\tau)g_{\gamma_0}(U_0)+\tau\bar{g}_{\gamma_0}(U_0)}\Big|_{\tau=0}\\
&= \frac{1}{g_{\gamma_0}(U)}(\dot{\bar{g}}_{\gamma_0}(U)-\dot{g}_{\gamma_0}(U)) + \frac{\dot{U}}{g_{\gamma_0}(U)}\left(\bar{g}'_{\gamma_0}(U)-g'_{\gamma_0}(U)\right)\\
&\quad - \frac{\left(\dot{g}_{\gamma_0}(U)+g'_{\gamma_0}(U)\dot{U}\right)}{g_{\gamma_0}(U)^2}(\bar{g}_{\gamma_0}(U)-g_{\gamma_0}(U)),
\end{aligned}$$

which is linear in $\bar{g}-g$. Therefore A *(i)* can be satisfied by choosing $b(R,q)$ based on the second-order pathwise derivative $\frac{\partial^2}{\partial\tau^2}S\left(R,q,g+\tau(\bar{g}-g)\right)$.

A *(ii)* could be verified using the conditions in Linton et al. (2008) that ensure the kernel estimators $\hat{g}_\gamma$, $\hat{\dot{g}}_\gamma$ and $\hat{g}'_\gamma$ are well-behaved. One issue is that the standard second-order kernel density derivative estimator may not converge fast enough, which Linton et al. (2008) prevent by assuming higher order (at least 4) kernels.

For kernel estimators, A *(iii)* is commonly established using the V-statistic projection outlined in Newey and McFadden (1994, Section 8.3, eqn. 8.11).

The asymptotic distribution of $\hat{\gamma}$ follows from the mean-value theorem, by which there exists $\tilde{\gamma}$ in between $\hat{\gamma}$

and γ_0 such that

$$0 = \hat{m}_T(\hat{\gamma}) = \hat{m}_T(\gamma_0) + \frac{\partial}{\partial\gamma}\hat{m}_T(\tilde{\gamma})\,(\hat{\gamma} - \gamma_0) \tag{3.11}$$

$$\Rightarrow \sqrt{T}\,(\hat{\gamma} - \gamma_0) = \left(\frac{\partial}{\partial\gamma}\hat{m}_T(\tilde{\gamma})\right)^{-1} \sqrt{T}\hat{m}_T(\gamma_0). \tag{3.12}$$

The following additional assumptions therefore yield the $\sqrt{T}$-convergence rate and asymptotic normality of our profile ML estimator.

Assumption N.

(i) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m(R_{t+1}, q_t, \gamma_0, g_0) \xrightarrow{d} N(0, V)$

(ii) $\sup_{\gamma\in\Theta} \|\frac{\partial}{\partial\gamma}\hat{m}_T(\gamma) - \partial\mathbb{E}\left(m(R_{t+1}, q_t, \gamma, g_\gamma)\right)/\partial\gamma\| \xrightarrow{p} 0$ *and* $M := \partial\mathbb{E}\left(m(R_{t+1}, q_t, \gamma, g_\gamma)\right)/\partial\gamma\big|_{\gamma=\gamma_0}$ *is non-singular.*

Proposition N. *Under Assumptions C, I, A, and N, when* $T \to \infty$,

$$\sqrt{T}\,(\hat{\gamma} - \gamma_0) \xrightarrow{d} N(0, M^{-1}VM^{-1}).$$

Assumption N*(i)* is not affected by the nonparametric part of our estimation, and simply requires that the score function follows a central limit theorem. Meanwhile, N*(ii)* typically follows from applying Assumption A to $\partial m(R_{t+1}, q_t, \gamma, g_\gamma)/\partial\gamma$ rather than S.

4 Monte Carlo Simulations

We simulate $N = 1000$ replications of 25 years of monthly data ($\tau = 1/12, T = 300$), taking for the risk-neutral density that implied by a Heston model:

$$\mathrm{d}V_t = 4.5(0.02 - V_t)\,\mathrm{d}t + 0.4\sqrt{V_t}\,\mathrm{d}W_t, \tag{4.1}$$

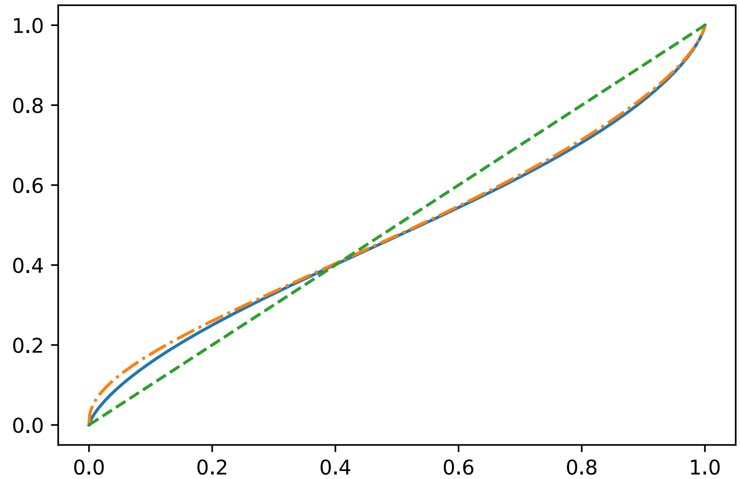
with discount rate $r = 0.03$ and leverage parameter $\rho = -0.9$. To simulate the volatility paths, we use the exact non-centered chi-squared distribution of the CIR model. We draw the initial V_0 from its stationary Gamma distribution.

As the utility function, we take CRRA with $\gamma_0 = 2$. The considered probability weighting functions are the one-parameter function proposed by Tversky and Kahneman (1992), and the two-parameter proposed by Prelec (1998):

$$Z_1(P) = \frac{P^\gamma}{(P^\gamma + (1-P)^\gamma)^{1/\gamma}}, \quad Z_2(P) = \exp(-(-\beta\log(P))^\alpha). \tag{4.2}$$

Both weighting functions, popular in the literature, allow for the typical inverse-S shape observed in experiments. The parameters are set as $\gamma = 0.75$, $(\alpha, \beta) = (0.7, 0.95)$. These weighting functions are displayed in Figure 1.

Figure 1: The probability weighting functions



Note: This figure displays the probability weighting functions Z_1 (blue, solid), Z_2 (orange, dash-dotted) and the 45-degrees line (green, dashed). In expected utility models, this 45-degrees line is the appropriate weighting function.

To simulate returns, we draw a standard uniform random variable $\mathcal{U}$ at each t, and calculate for which R_{t+1} the following equation holds:

$$1 - Z(\mathcal{U}) = c_t(\gamma_0) \int_{-\infty}^{R_{t+1}} \frac{q_t(r)}{u'(r;\gamma_0)} \, \mathrm{d}r, \tag{4.3}$$

calculating $c_t(\gamma_0)$ and the integral by a Riemann-discretization, and $q_t(r)$ by Fourier-inverting the known Heston CF. We truncate numerically negative density values at zero, and rescale such that the density sums to 1. Throughout, we treat (q, R) as observed.

4.1 Profile ML

Using the techniques developed in Section 3, we estimate $\hat{\gamma}$ for each of the $N = 1000$ replications. Recall that, though the data is generated using parametric probability weighting functions, the profile likelihood uses a nonparametric, smoothed estimate of the weighting function. In fact, we use a Gaussian kernel to estimate the profiled g_γ as in (3.6), and use a variety of bandwidths. Taking a different bandwidth results in a different $\hat{\gamma}$, see Table 1. The histograms in Figure 2 show the distribution of these estimates in deviation of the true γ_0, overlaid with a normal density, for a given bandwidth. The resulting $\hat{\gamma}$ can then be used to nonparametrically estimate a non-smoothed weighting function by exploiting (3.5). For the same bandwidth as used in Figure 2, results of this estimation of the weighting function are displayed in Figure 3.

Table 1: Profile ML performance with different bandwidths

Bandwidth	Bias	St. Dev.	MSE
0.05	-0.108	2.718	7.397
0.10	-0.056	1.863	3.475
0.15	0.218	1.567	2.505
0.20	0.597	1.417	2.365
0.25	0.991	1.331	2.754

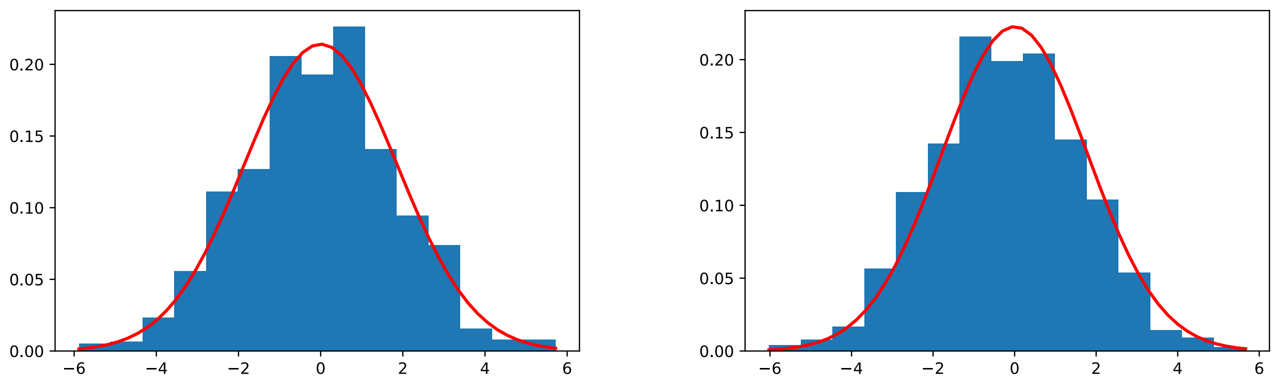
(a) True PWF is Z_1

Bandwidth	Bias	St. Dev.	MSE
0.05	-0.452	2.572	6.821
0.10	-0.204	1.792	3.252
0.15	0.190	1.531	2.381
0.20	0.589	1.414	2.345
0.25	0.955	1.333	2.688

(b) True PWF is Z_2

Note: These subtables display the estimation results of the profile likelihood over the $N = 1000$ replications. The profile likelihood uses a normal kernel with varying bandwidth, displayed in the first column. The two subpanels are based on simulations with differing DGP, in the form of a different true probability weighting function.

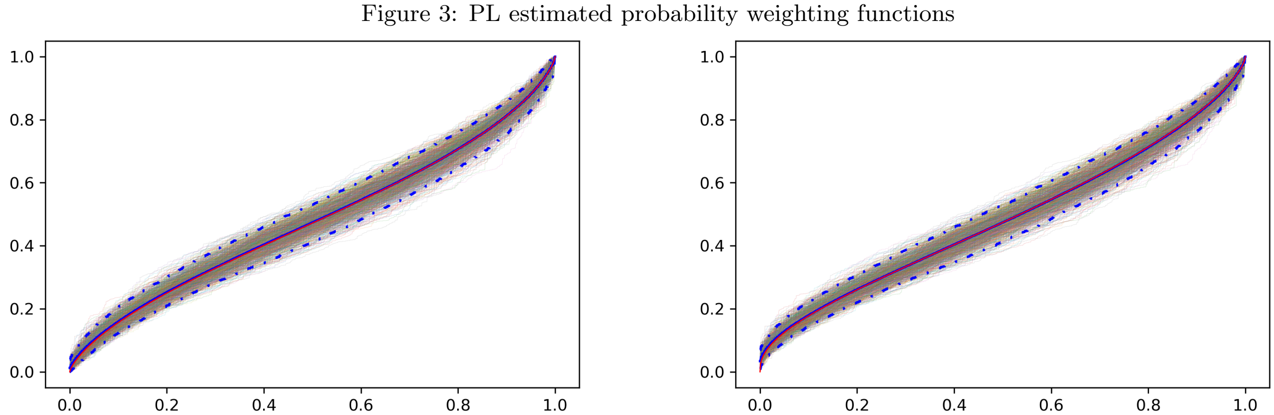
Figure 2: Histogram of simulation estimates $\hat{\gamma} - \gamma_0$



(a) True PWF is Z_1 (b) True PWF is Z_2

Note: These figures display histograms of the estimates $\hat{\gamma}$ resulting from the simulations, in deviation of the true γ_0. The profile likelihood uses a normal kernel with bandwidth $h = 0.1$. The red curve is a normal density with mean zero and standard deviation equal to that of $\hat{\gamma} - \gamma_0$ over the $N = 1000$ replications. The two subpanels are based on simulations with differing DGP, in the form of a different true probability weighting function.

Figure 3: PL estimated probability weighting functions



(a) True PWF is Z_1 (b) True PWF is Z_2

Note: These figures display the nonparametric estimates of the probability weighting function, given the respective value of $\hat{\gamma}$. The profile likelihood uses a normal kernel with bandwidth $h = 0.1$. Each of the $N = 1000$ lines represents a single simulation. The mean (blue, solid), lower and upper 2.5% percentiles (blue, dash-dotted), and the true PW function (red, solid), are also displayed. The mean and the true function almost completely overlap. The two subpanels are based on simulations with differing DGP, in the form of a different true probability weighting function.

4.1.1 Bootstrap estimation of PL confidence interval

To compute the lower and upper α-quantiles of $\hat{\gamma} - \gamma_0$, i.e., of the distribution of the PL estimator in deviation from its true value, we consider a bootstrap algorithm, see Algorithm 1. The bandwidth used for the kernel in the PL step is 0.1. We use $K = 500$ bootstrap repetitions for the $N = 1000$ simulations. In both DGPs, the bootstrap leads to a correctly centered, seemingly normal distribution for $\tilde{\gamma} - \hat{\gamma}$: the means are 0.04 and 0.02 under Z_1 and Z_2 respectively. The coverage rates of the bootstrap-implied confidence interval are 95.5% and 95.0%, and the average width of this 95% interval are 7.792 and 7.451, respectively.

Algorithm 1: Nonparametric bootstrap

Data: $\hat{\gamma}$; $(q_t, R_{t+1})_{t=1}^T$; $K \in \mathbb{N}$

Result: $(\hat{\gamma} - \gamma_0)_\alpha$; $(\hat{\gamma} - \gamma_0)_{1-\alpha}$

for $k = 1, \ldots, K$ **do**

  Construct $(\tilde{q}_t, \tilde{R}_{t+1})_{t=1}^T$ by drawing T times from $(q_t, R_{t+1})_{t=1}^T$ with replacement;

  $\tilde{\gamma}_k \leftarrow \text{ProfileML}\Big((\tilde{q}_t, \tilde{R}_{t+1})_{t=1}^T\Big)$

end

$\{(\hat{\gamma} - \gamma_0)_\alpha, (\hat{\gamma} - \gamma_0)_{1-\alpha}\} \leftarrow \text{quantile}(\{\alpha, 1-\alpha\}, \tilde{\gamma} - \hat{\gamma})$;

5 Empirical Results

We illustrate our method using option prices on the S&P 500 index obtained from OptionMetrics over the period January 1996 to February 2023. In particular, we consider options expiring on the third Friday of each month, and obtain their prices on the last trading day with at least $\tau = 29$ days to maturity.

The risk-neutral densities q_t can be estimated from cross-sections of option prices with varying strike prices, for which we consider two methods. The first estimation method fits the parametric Bates (1996) single-factor stochastic volatility jump-diffusion model to each monthly cross-section. The second estimation methods uses the local cubic kernel estimator of Dalderop (2020), with a varying plug-in bandwidth based on the fitted Bates (1996) model. Since nonparametric methods would be unstable in the tails of the density due to sparse trading of deep OTM options, we 'paste' the tails of the Bates (1996) model to match at the lower and upper moneyness thresholds that leave 15 observed option prices in either tail.[7] Figure 4 shows that our estimation method produces smooth densities throughout the sample.

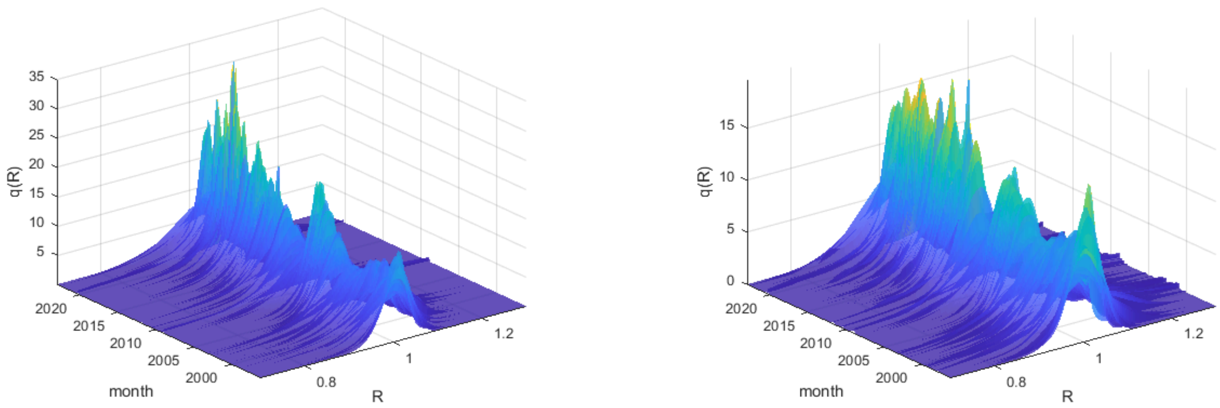
First, we consider the estimation of the parameter γ of the CRRA utility function *without* probability weighting. Figure 5 shows the log likelihood functions for varying γ for both the parametric and nonparametric risk-neutral density estimation methods. Their maxima are obtained at $\hat{\gamma} = 2.16$ and 1.44, respectively. These values are on the lower end of the range of estimates in the literature.

Figure 6 shows the profile log likelihood functions as a function of the parameter γ of the CRRA utility function, now allowing for probability weighting. The optimal $\hat{\gamma}$ are now even lower, and range within $(-1, 1)$ for both methods and bandwidth choices considered.

Finally, Figure 7 shows the resulting estimates of the inverse dual probability weighting densities. For both methods and bandwidth choices, the density show a clear hill-shape, rather than a uniform shape. The leveling off to values well below one near the boundaries suggests that both left and right tail cumulative

[7]There were 13 periods with fewer than 15 OTM put or call options, all occurring in the early years of the sample, for which we only use the parametric fit.

Figure 4: Estimated risk-neutral densities for monthly S&P 500 index returns

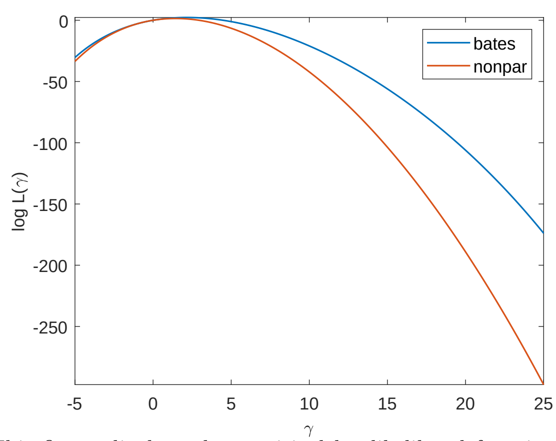


(a) Bates (1996) parametric fit.

(b) Local cubic kernel-smoother.

Note: These figures display nonparametric estimates of the risk-neutral density estimated from options. The two subpanels present the results of different estimators.

Figure 5: Log-likelihood function, without probability weighting



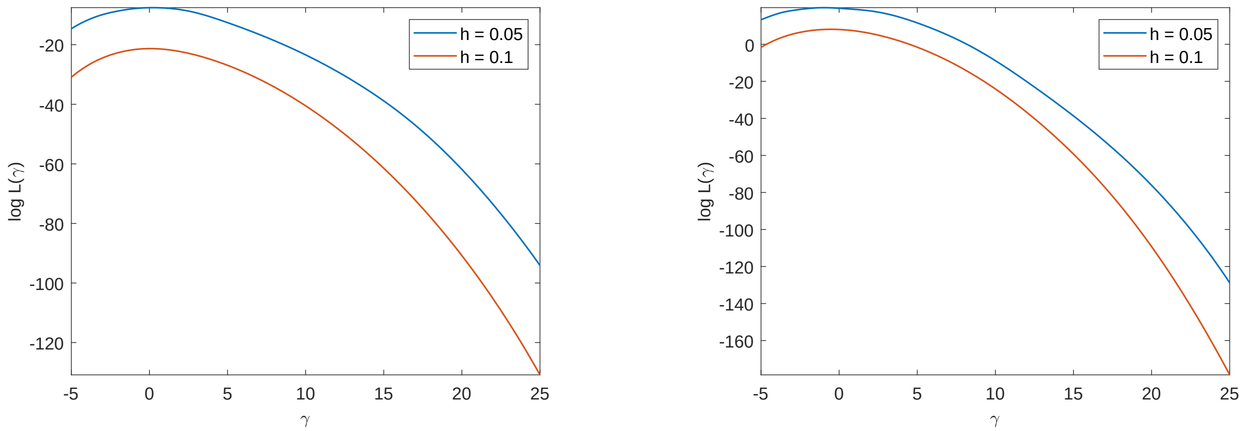
Note: This figure displays the empirical log-likelihood function for varying CRRA parameter γ, without allowing for probability weighting. The two differently colored curves are based on different estimators for the risk-neutral density.

probabilities are overestimated by the risk-neutral distribution, with the effect the strongest for the left tail. While the shape of the probability weighting function is qualitatively similar for both choices of initial risk-neutral density estimation methods, its effect is more pronounced for the nonparametric method.

The deviations from uniformity in Figure 7 suggest pronounced probability weighting. However, these estimates are subject to the assumption of CRRA utility over return outcomes. An alternative explanation for the findings is that the marginal utility function is misspecified and should be upward-sloping for positive returns, in line with the U-shaped pricing kernels reported elsewhere in the literature. Therefore we consider a wider class of pricing kernels, based on utility specifications of the form

$$u'(R;\gamma) = \exp\left(\sum_{l=1}^{L} \gamma_l (\log R)^l\right),$$

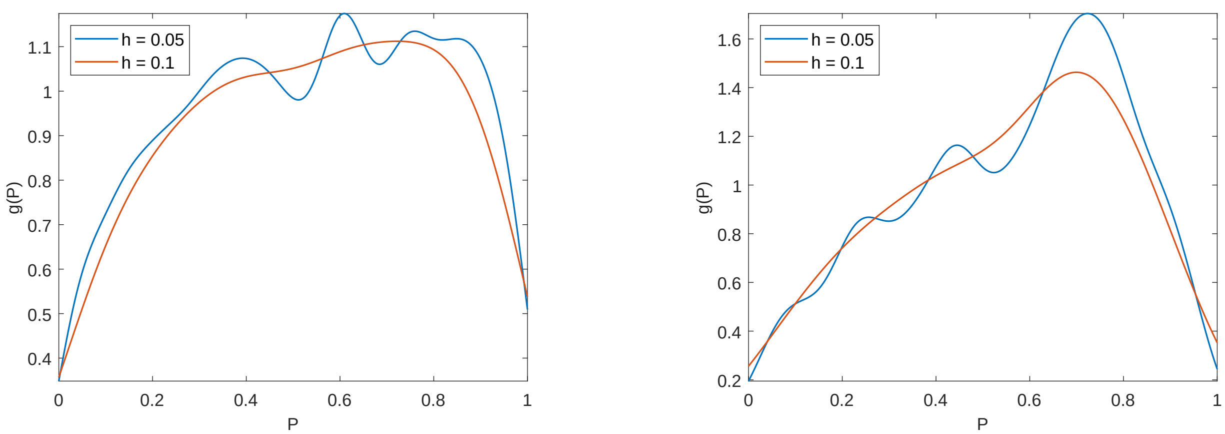
Figure 6: Profile log-likelihood function



(a) Bates (1996) parametric fit. (b) Local cubic kernel-smoother.

Note: These figures display the empirical profile likelihood for varying CRRA parameter γ. This uses the kernel-estimator $\hat{\hat{g}}(\cdot;\gamma)$ of the density of the inverse dual probability probability weighting function as input. The two differently colored curves represent different choices for the kernel-bandwidth h. The two subpanels present the results of different estimators for the risk-neutral density.

Figure 7: Kernel-based PL estimates of the inverse dual probability weighting density



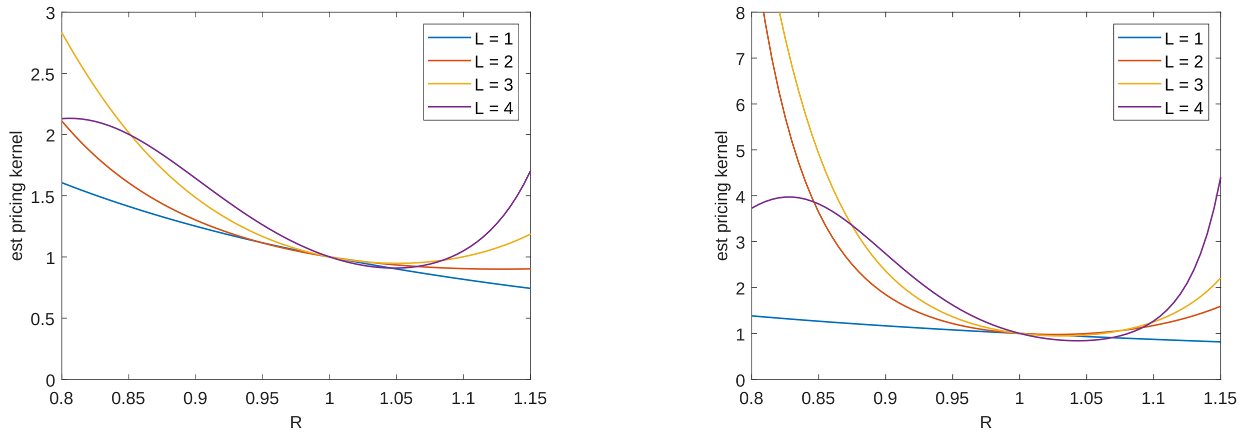
(a) Bates (1996) parametric fit. (b) Local cubic kernel-smoother.

Note: These figures display the smoothed estimates $\hat{\hat{g}}(\cdot;\hat{\gamma})$ of the density of the inverse dual probability weighting function, given the optimal value for the CRRA parameter. The two differently colored curves represent different choices for the kernel-bandwidth h. The two subpanels present the results of different estimators for the risk-neutral density.

which nests power utility when $L = 1$. Rosenberg and Engle (2002) consider pricing kernel specifications of this form. First, assuming no probability weighting, Figure 8 shows the estimated exponential-polynomial pricing kernels without probability weighting, based on standard maximum likelihood for varying order L. When $L \geq 2$, the pricing kernels display clear non-monotonicity, in particular U-shapes.

Figure 9 shows the estimated pricing kernels when probability weighting is allowed for. The curves now become mostly flat, or even increasing in the return outcome. Figure 10 shows the resulting kernel-based density estimates of the inverse dual probability weighting function. The curves show similar estimates as in Figure 7 for power utility, and are relatively insensitive to the order L. Thus, using profile ML estimation, probability weighting appears robust to the utility function specification.

Figure 8: ML-estimated exponential-polynomial pricing kernels, without probability weighting

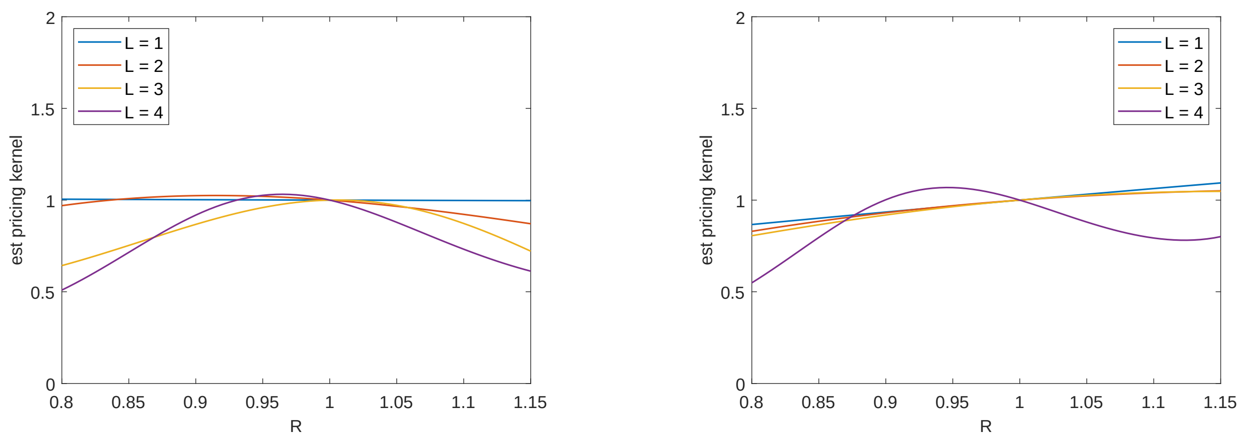


(a) Bates (1996) parametric fit.

(b) Local cubic kernel-smoother.

Note: These figures display maximum likelihood-estimated pricing kernel for the exponential-polynomial utility of varying order L, while not allowing for probability weighting. The four differently colored curves represent different choices for the polynomial length L. The two subpanels present the results of different estimators for the risk-neutral density.

Figure 9: PL-estimated exponential-polynomial pricing kernels



(a) Bates (1996) parametric fit.

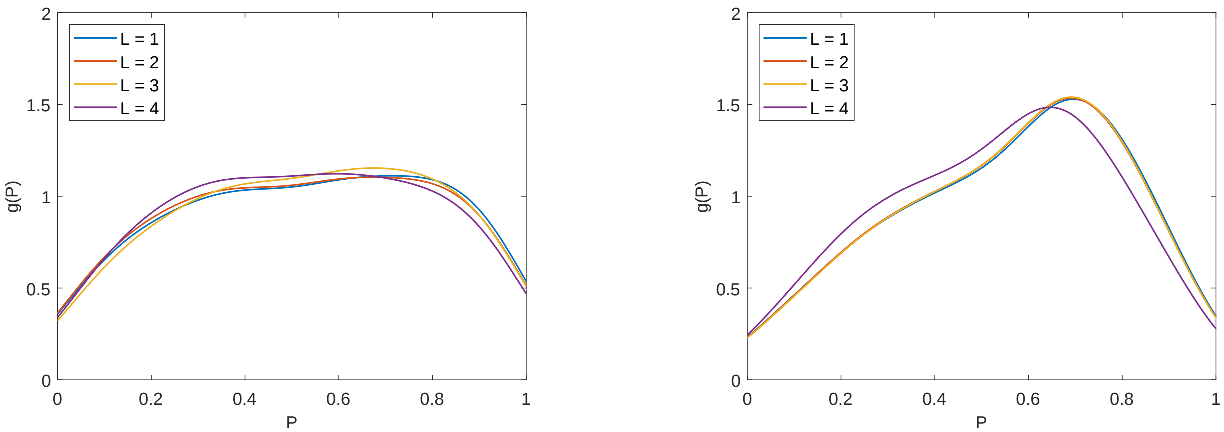
(b) Local cubic kernel-smoother.

Note: These figures display profile maximum likelihood-estimated pricing kernel for the exponential-polynomial utility of varying order L, while allowing for probability weighting. The four differently colored curves represent different choices for the polynomial length L. The two subpanels present the results of different estimators for the risk-neutral density.

Figure 11 shows the resulting probability weighting functions, under CRRA. The top row displays the kernel-smoothed cumulative distributions of the PIT functional $U_{t+1}(\hat{\gamma})$, which show a clear S-shape. Meanwhile, the bottom row displays the corresponding estimates of the original probability weighting function Z over $1-F_t(R)$, which show pronounced inverse S-shapes. The findings appear robust to the method of risk-neutral density estimation and the choice of bandwidth.

To conclude, recall that, following Eeckhoudt et al. (2020), the signs of the probability weighting function's derivatives are of interest, as these constitute probabilistic variants of risk concepts as prudence and temperance. In particular, its third derivative can be interpreted as dual prudence. As $Z(P) = 1 - \overline{Z}(1-P)$, the odd derivatives of the probability weighting function and its dual have the same sign. Additionally, standard

Figure 10: Kernel-based PL estimates of the inverse dual probability weighting density under exponential-polynomial utility



(a) Bates (1996) parametric fit.

(b) Local cubic kernel-smoother.

Note: These figures display the smoothed estimates $\hat{\hat{g}}(\cdot;\hat{\gamma})$ of the density of the inverse dual probability weighting function, given the optimal value for the exponential-polynomial utility parameters. The bandwidth is set at $h = 0.1$. The four differently colored curves represent different choices for the polynomial length L. The two subpanels present the results of different estimators for the risk-neutral density.

calculus tells us that

$$\operatorname{sign}\Big(\overline{Z}^{(3)}(\overline{Z}^{-1}(P))\Big) = \operatorname{sign}\big(3(g'(P))^2 - g''(P)g(P)\big)\,,$$

using that $g > 0$. Hence, we have dual prudence at $1 - \overline{Z}^{-1}(P)$ if $3(g'(P))^2/g(P) > g''(P)$. As the left-hand side of this inequality is positive, we can surely conclude dual prudence if g is concave. Based on Figure 7, we observe that g is indeed concave, except around $P \approx 0.5$. There seems to be a hint of convexity around the extremes $P \approx 0$ and $P \approx 1$, but the kernel estimators are notably less precise around the borders, so this should be taken *cum grano salis*.[8] Informally, we thus find evidence of dual prudence for almost all values of P, except perhaps in the center of the distribution. Note that even in the center, we cannot rule out dual prudence: concavity of g is sufficient, but stronger than necessary. This argument remains valid under the more general exponential-polynomial utility; see Figure 10.

6 Conclusion

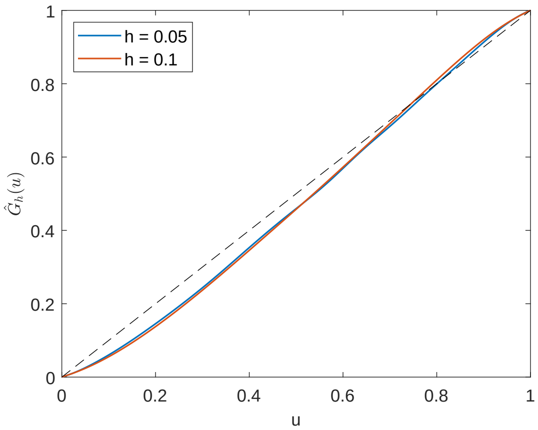
The intertwined nature of attitude toward wealth and attitude toward probabilities embedded in a panel of option prices poses a challenging econometric problem. In this paper, we have devised semiparametric inference theory to bear on this problem, and disentangle and identify the probability weighting and utility functions implicit in the risk-neutral distribution of option contracts. Using Monte Carlo simulations we have demonstrated the favorable performance of our approach in finite samples. Our empirical analysis of a large sample of S&P 500 index option prices unveils the importance of probability weighting. The probability weighting function implicit in option prices is found to be inverse S-shaped, a finding that is robust to the parametric specification of the utility function.

Our results, and the nonlinearities in probabilities that they entail, may have important implications for

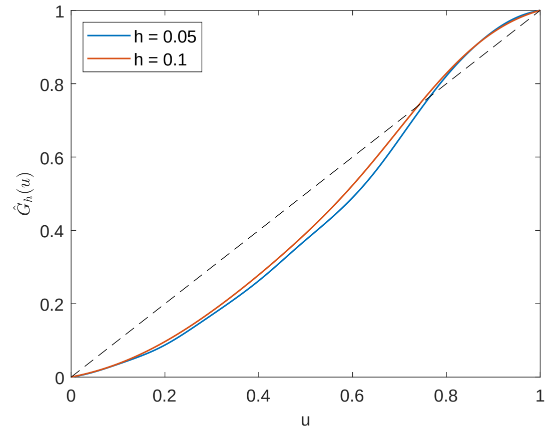
[8] In fact, the related Figures 11c and 11d do not show the thence expected decrease in (absolute) second order derivative around the borders; the curvature seems to increase as we move to the extremes.

hedging strategies using derivatives, which constitutes a promising avenue for future research.

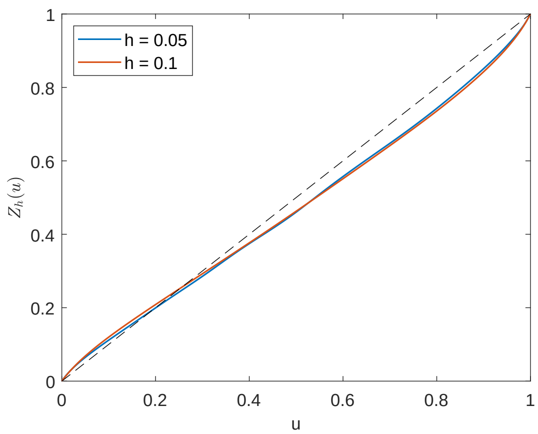
Figure 11: Smoothed PL estimates of the probability weighting function under CRRA



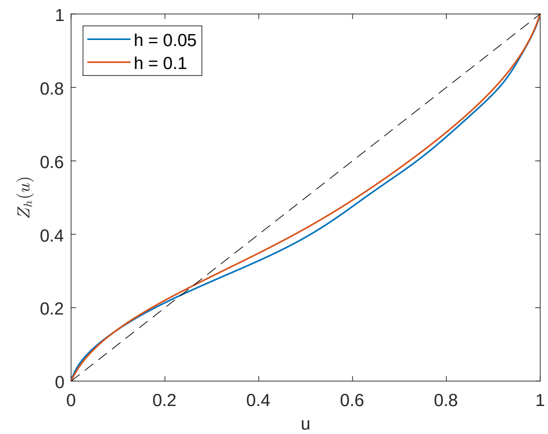
(a) Bates (1996) parametric fit.



(b) Local cubic kernel-smoother.



(c) Bates (1996) parametric fit.



(d) Local cubic kernel-smoother.

Note: These figures display the smoothed estimates of quantities related to the probability weighting function, under CRRA utility. The dual inverse G is displayed in the top panels, and the weighting function Z itself is displayed in the bottom panels. The two differently colored curves represent different choices for the bandwidth h. The dashed, black curve is the 45-degrees line. The left and right subpanels present the results of different estimators for the risk-neutral density.

Appendix

A Proofs

Proof of Lemma E. Note that

$$\mathbb{E}\left(\log f_t(R_{t+1};\gamma,g)\right)=\mathbb{E}\left(\log c_t(\gamma)+\log q_t(R_{t+1})-\log u'(R_{t+1};\gamma)+\log g\left(U_{t+1}(\gamma)\right)\right).$$

To prove (3.8), let $\tilde{g}_\gamma$ be an arbitrary density for each γ. Then,

$$\mathbb{E}\left(\log\frac{f_t(R_{t+1};\gamma,\tilde{g}_\gamma)}{f_t(R_{t+1};\gamma,g_\gamma)}\right)=\mathbb{E}\left(\log\frac{\tilde{g}_\gamma(U_{t+1}(\gamma))}{g_\gamma(U_{t+1}(\gamma))}\right)\leq\log\mathbb{E}\left(\frac{\tilde{g}_\gamma(U_{t+1}(\gamma))}{g_\gamma(U_{t+1}(\gamma))}\right)=\log\int_0^1\frac{\tilde{g}_\gamma(u)}{g_\gamma(u)}g_\gamma(u)\,\mathrm{d}u=0,$$

with equality if and only if $g_\gamma:=\tilde{g}_\gamma$ *a.s.*[9] Hence, g_γ must be the maximizing function. □

Proof of Lemma I. Equation (3.10) is equivalent to any of the following statements:

$$\begin{aligned}
&f_t(r;\gamma,g_\gamma)=f_t(r;\gamma_0,g_{\gamma_0})\text{ a.s. for any } r \text{ such that } \mathbb{P}\left(f_t(r)>0\right)>0\\
\Leftrightarrow\ &F_t(r;\gamma,g_\gamma)=F_t(r;\gamma_0,g_{\gamma_0})\text{ a.s. for any } r \text{ such that } \mathbb{P}\left(f_t(r)>0\right)>0\\
\Leftrightarrow\ &F_t(F_t^{-1}(u);\gamma,g_\gamma)=u\text{ a.s. for any } u\in(0,1)\\
\Leftrightarrow\ &G_\gamma^{-1}\left(c_t(\gamma)\int^{F_t^{-1}(u)}q_t(r)/u'(r;\gamma)\,\mathrm{d}r\right)=u\text{ a.s. for any } u\in(0,1)\\
\Leftrightarrow\ &c_t(\gamma)\int^{F_t^{-1}(u)}q_t(r)/u'(r;\gamma)\,\mathrm{d}r=G_\gamma(u)\text{ a.s. for any } u\in(0,1).
\end{aligned}$$

Since the RHS of the final statement is constant, it can only hold if the LHS

$$\begin{aligned}
U_t(u;\gamma)&:=c_t(\gamma)\int^{F_t^{-1}(u)}q_t(r)/u'(r;\gamma)\,\mathrm{d}r\\
&=c_t(\gamma)\int^{u}q_t(F_t^{-1}(v))/u'(F_t^{-1}(v);\gamma)/f_t(F_t^{-1}(v))\,\mathrm{d}v\\
&=c_t(\gamma)c_t^{-1}\int^{u}u'(F_t^{-1}(v);\gamma_0)/u'(F_t^{-1}(v);\gamma)z(v)\,\mathrm{d}v
\end{aligned}$$

is a constant function of u over time. However,

$$\frac{\partial\log U_t(u;\gamma)}{\partial u}=\frac{u'(F_t^{-1}(u);\gamma_0)/u'(F_t^{-1}(u);\gamma)z(y)}{\int^u u'(F_t^{-1}(v);\gamma_0)/u'(F_t^{-1}(v);\gamma)z(v)\,\mathrm{d}v}.$$

Under Assumption I*(i)*, $u'(F_t^{-1}(u);\gamma_0)/u'(F_t^{-1}(u);\gamma)$ cannot be a constant function of u, as the ratio varies with r with non-zero probability. Therefore $\frac{\partial\log U_t(u;\gamma)}{\partial u}$ and thus $U_t(u;\gamma)$ are not deterministic. As a result, for any $\gamma\neq\gamma_0$, (3.10) does not hold, so that $\ell(\gamma)<\ell(\gamma_0)$. □

Proof of Proposition C. If (i) for every neighborhood Θ_0 of γ_0, $\max_{\gamma\in\Theta/\Theta_0}\ell(\gamma)<\ell(\gamma_0)$ and (ii) $\sup_{\gamma\in\Theta}\big|\ell_T(\gamma)-$

[9] Jensen's *inequality* $\mathbb{E}f(X)\geq f(\mathbb{E}X)$ is an *equality* if and only if f is affine on the set A with $\mathbb{P}(X\in A)=1$. Clearly, as $f(x)=\log x$ is strictly concave, $\tilde{g}_\gamma(U_{t+1}(\gamma))/g_\gamma(U_{t+1}(\gamma))$ must be constant *a.s.* As both numerator and denominator must integrate to one, they thus have to be equivalent on any set in the support of U_γ that has positive probability.

$\ell(\gamma)\big| \xrightarrow{p} 0$, then $\hat{\gamma} \xrightarrow{p} \gamma_0$ (e.g., Andrews, 1994, Lemma A-1). Condition (i) follows from the identification argument discussed above: given Assumption I, $\ell(\gamma)$ is uniquely maximized at γ_0 over Θ, see Lemma I. To verify the uniform convergence condition (ii), write

$$\sup_{\gamma\in\Theta}\big|\ell_T(\gamma)-\ell(\gamma)\big| \le \sup_{\gamma\in\Theta}\big|\ell_T(\gamma,\hat{g}_\gamma)-\ell_T(\gamma,g_\gamma)\big| + \sup_{\gamma\in\Theta}\big|\ell_T(\gamma,g_\gamma)-\ell(\gamma,g_\gamma)\big|.$$

For the first term, C*(ii)* implies

$$\begin{aligned}
|\ell_T(\gamma,\hat{g}_\gamma)-\ell_T(\gamma,g_\gamma)| &= \left|\frac{1}{T}\sum_{t=1}^{T}\log\hat{g}_\gamma(U_{t+1}(\gamma)) - \log g_\gamma(U_{t+1}(\gamma))\right| \\
&\le \frac{1}{T}\sum_{t=1}^{T}\left|\log\hat{g}_\gamma(U_{t+1}(\gamma)) - \log g_\gamma(U_{t+1}(\gamma))\right| \\
&\le \frac{1}{T}\sum_{t=1}^{T}\tilde{b}(R_{t+1},q_t)\left\|\hat{g}_\gamma - g_\gamma\right\|^{\varepsilon},
\end{aligned}$$

which vanishes uniformly over γ. The second term vanishes as under C*(i)*, $\ell_T(\gamma)$ follows a uniform law of large numbers, e.g., Andrews (1987). □

Proof of Lemma A. The remainder term equals

$$\begin{aligned}
R_{T1} &:= \sqrt{T}\hat{m}_T(\gamma_0) - \frac{1}{\sqrt{T}}\sum_{t=1}^{T} m(R_{t+1},q_t,\gamma_0,g_0) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(m(R_{t+1},q_t,\gamma_0,\hat{g}) - m(R_{t+1},q_t,\gamma_0,g_0)\right) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\frac{\partial}{\partial\gamma}\hat{g}\left(U_{t+1}(\gamma);\gamma\right)\big|_{\gamma=\gamma_0}}{\hat{g}\left(U_{t+1}(\gamma_0);\gamma_0\right)} - \frac{\frac{\partial}{\partial\gamma}g_0\left(U_{t+1}(\gamma);\gamma\right)\big|_{\gamma=\gamma_0}}{g_0\left(U_{t+1}(\gamma_0);\gamma_0\right)}\right) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(S(R_{t+1},q_t,\hat{g}) - S(R_{t+1},q_t,g_0)\right).
\end{aligned}$$

Under Assumption A, it follows that

$$\begin{aligned}
\|R_{T1}\| &\le \left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}D(R_{t+1},q_t,\hat{g}-g_0)\right\| + \left\|\frac{1}{T}\sum_{t=1}^{T}b(R_{t+1},q_t)\right\|\sqrt{T}\|\hat{g}-g_0\|^2 \\
&= \left\|\sqrt{T}\int D(R_{t+1},q_t,\hat{g}-g_0)\,\mathrm{d}F_0(R_{t+1},q_t)\right\| + o_p(1),
\end{aligned}$$

We conjecture and verify that $\int D(R_{t+1},q_t,\hat{g}-g_0)\,\mathrm{d}F_0(R_{t+1},q_t) = 0$. In particular, we write

$$\int D(R,q,\hat{g}-g_0)\,\mathrm{d}F_0(R,q) = (*) + (**),$$

where

$$
\begin{aligned}
(*) &= \int \frac{1}{g_0(u)} \left(\frac{1}{T} \sum_{t=1}^{T} K_h'(U_{t+1}(\gamma_0) - u) U_{t+1}'(\gamma_0) - \dot{g}_0(u) \right) g_0(u) \, \mathrm{d}u \\
&= \frac{1}{T} \sum_{t=1}^{T} U_{t+1}'(\gamma_0) \int K_h'(U_{t+1}(\gamma_0) - u) \, \mathrm{d}u - \int \dot{g}_0(u) \, \mathrm{d}u \\
&= \frac{1}{Th} \sum_{t=1}^{T} U_{t+1}'(\gamma_0) \int K'(z) \, \mathrm{d}z - \frac{\partial}{\partial \gamma} \int g_0(\gamma, u) \, \mathrm{d}u \Big|_{\gamma=\gamma_0} = 0 - 0 = 0,
\end{aligned}
$$

and

$$
\begin{aligned}
(**) &= \int \frac{\dot{U}}{g_0(U)} \left(\frac{1}{T} \sum_{t=1}^{T} K_h'(U_{t+1}(\gamma_0) - U) - g_0'(U) \right) \mathrm{d}F_0(R, q) \\
&\quad - \int \frac{\left(\dot{g}_0(U) + g_0'(U) \dot{U} \right)}{g_0(U)^2} \left(\frac{1}{T} \sum_{t=1}^{T} K_h(U_{t+1}(\gamma_0) - U) - g_0(U) \right) \mathrm{d}F_0(R, q) \\
&= \frac{1}{T} \sum_{t=1}^{T} \int \frac{1}{g_0(U_0)^2} \left(\frac{\partial}{\partial \gamma} K_h(U_{t+1}(\gamma_0) - U(\gamma)) \Big|_{\gamma=\gamma_0} g_0(U_0) \right. \\
&\qquad \left. - \frac{\partial}{\partial \gamma} g_0(\gamma, U(\gamma)) \Big|_{\gamma=\gamma_0} K_h(U_{t+1}(\gamma_0) - U_0) \right) \mathrm{d}F_0(R, q) \\
&= \frac{1}{T} \sum_{t=1}^{T} \int \frac{\partial}{\partial \gamma} \frac{K_h(U_{t+1}(\gamma_0) - U(\gamma))}{g_0(U(\gamma))} \Big|_{\gamma=\gamma_0} \mathrm{d}F_0(R, q) \\
&= \frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \gamma} \int K_h(U_{t+1}(\gamma_0) - u) \, \mathrm{d}u \Big|_{\gamma=\gamma_0} \\
&= 0,
\end{aligned}
$$

since K_h integrates to one. Therefore, $\|R_{T1}\| = o_p(1)$. □

Proof of Proposition N. This is immediate from Proposition C, Lemma A, Assumption N, and the mean-value expansion in (3.12). □

References

Abdellaoui, M. (2000): "Parameter-free elicitation of utility and probability weighting functions," *Management Science*, 46, 1497–1512.

Abel, A. B. (1990): "Asset prices under habit formation and catching up with the Joneses," *American Economic Review*, 80, 38–42.

Ai, H. (2005): "Smooth nonexpected utility without state independence," *Working paper, Federal Reserve Bank of Minneapolis.*

Aït-Sahalia, Y. and J. Duarte (2003): "Nonparametric option pricing under shape restrictions," *Journal of Econometrics*, 116, 9–47.

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

——— (2000): "Nonparametric risk management and implied risk aversion," *Journal of Econometrics*, 94, 9–51.

Allais, M. (1953): "Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école Américaine," *Econometrica*, 21, 503–546.

Andrews, D. W. (1987): "Consistency in nonlinear econometric models: A generic uniform law of large numbers," *Econometrica*, 55, 1465–1471.

——— (1994): "Asymptotics for semiparametric econometric models via stochastic equicontinuity," *Econometrica*, 62, 43–72.

Baele, L., J. Driessen, S. Ebert, J. M. Londono, and O. G. Spalt (2019): "Cumulative prospect theory, option returns, and the variance premium," *The Review of Financial Studies*, 32, 3667–3723.

Barberis, N., L. J. Jin, and B. Wang (2021): "Prospect theory and stock market anomalies," *The Journal of Finance*, 76, 2639–2687.

Barberis, N., A. Mukherjee, and B. Wang (2016): "Prospect theory and stock returns: An empirical test," *The Review of Financial Studies*, 29, 3068–3107.

Bates, D. S. (1996): "Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options," *The Review of Financial Studies*, 9, 69–107.

——— (2000): "Post-'87 crash fears in the S&P 500 futures option market," *Journal of Econometrics*, 94, 181–238.

Bliss, R. R. and N. Panigirtzoglou (2004): "Option-implied risk aversion estimates," *The Journal of Finance*, 59, 407–446.

Bollerslev, T. (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

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

Broadie, M., M. Chernov, and M. Johannes (2007): "Model specification and risk premia: Evidence from futures options." *The Journal of Finance*, 62, 1453–1490.

Carr, P. and D. Madan (2001): "Optimal positioning in derivative securities," *Quantitative Finance*, 1, 19–37.

Chen, X. and S. C. Ludvigson (2009): "Land of addicts? An empirical investigation of habit-based asset pricing models," *Journal of Applied Econometrics*, 24, 1057–1093.

Chew, S.H., E. K. and Z. Safra (1987): "Risk aversion in the theory of expected utility with rank dependent probabilities," *Journal of Economic Theory*, 42, 370–381.

Cochrane, J. (2009): *Asset pricing: Revised edition*, Princeton University Press.

Cuesdeanu, H. and J. C. Jackwerth (2018a): "The pricing kernel puzzle in forward looking data," *Review of Derivatives Research*, 21, 253–276.

——— (2018b): "The pricing kernel puzzle: survey and outlook," *Annals of Finance*, 14, 289–329.

Dalderop, J. (2020): "Nonparametric filtering of conditional state-price densities," *Journal of Econometrics*, 214, 295–325.

——— (2021): "Efficient Estimation of Pricing Kernels and Market-Implied Densities," *Working Paper*.

Duan, J.-C. (1995): "The GARCH option pricing model," *Mathematical finance*, 5, 13–32.

Duffie, D., J. Pan, and K. Singleton (2000): "Transform Analysis and Asset Pricing for Affine Jump-diffusions," *Econometrica*, 68, 1343–1376.

Eeckhoudt, L. R. and R. J. A. Laeven (2022): "Dual moments and risk attitudes," *Operations Research*, 70, 1330–1341.

Eeckhoudt, L. R., R. J. A. Laeven, and H. Schlesinger (2020): "Risk apportionment: The dual story," *Journal of Economic Theory*, 185, 104971.

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

Gonzalez, R. and G. Wu (1999): "On the shape of the probability weighting function," *Cognitive Psychology*, 38, 129–166.

Gordon, S. and P. St-Amour (2004): "Asset returns and state-dependent risk preferences," *Journal of Business & Economic Statistics*, 22, 241–252.

Hens, T. and C. Reichlin (2013): "Three solutions to the pricing kernel puzzle," *Review of Finance*, 17, 1065–1098.

Heston, S. L. (1993): "A closed-form solution for options with stochastic volatility with applications to bond and currency options," *The Review of Financial Studies*, 6, 327–343.

Heston, S. L. and S. Nandi (2000): "A closed-form GARCH option valuation model," *The Review of Financial Studies*, 13, 585–625.

Kliger, D. and O. Levy (2009): "Theories of choice under risk: Insights from financial markets," *Journal of Economic Behavior & Organization*, 71, 330–346.

Linn, M., S. Shive, and T. Shumway (2018): "Pricing kernel monotonicity and conditional information," *The Review of Financial Studies*, 31, 491–531.

Linton, O., S. Sperlich, and I. Van Keilegom (2008): "Estimation of a semiparametric transformation model," *The Annals of Statistics*, 36, 686–718.

Liu, X., M. B. Shackleton, S. J. Taylor, and X. Xu (2007): "Closed-form transformations from risk-neutral to real-world distributions," *Journal of Banking & Finance*, 31, 1501–1520.

Lu, J. and Z. Qu (2021): "Sieve estimation of option-implied state price density," *Journal of Econometrics*, 224, 88–112.

Muliere, P. and M. Scarsini (1989): "A note on stochastic dominance and inequality measures," *Journal of Economic Theory*, 49, 314–323.

Newey, W. K. (1994): "The asymptotic variance of semiparametric estimators," *Econometrica*, 62, 1349–1382.

Newey, W. K. and D. McFadden (1994): "Large sample estimation and hypothesis testing," *Handbook of Econometrics*, 4, 2111–2245.

Pan, J. (2002): "The jump-risk premia implicit in options: evidence from an integrated time-series study," *Journal of Financial Economics*, 63, 3–50.

Polkovnichenko, V. and F. Zhao (2013): "Probability weighting functions implied in options prices," *Journal of Financial Economics*, 107, 580–609.

PRELEC, D. (1998): "The probability weighting function," *Econometrica*, 66, 497–527.

QUIGGIN, J. (1982): "A theory of anticipated utility," *Journal of Economic Behavior & Organization*, 3, 323–343.

ROËLL, A. (1987): "Risk aversion in Quiggin and Yaari's rank-order model of choice under uncertainty," *The Economic Journal*, 97, 143–159.

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

ROTHSCHILD, M. AND J. STIGLITZ (1970): "Increasing risk: I. A definition," *Journal of Economic Theory*, 2, 225–243.

RYAN, M. (2006): "Risk aversion in RDEU," *Journal of Mathematical Economics*, 42, 675–697.

SCHREINDORFER, D. AND T. SICHERT (2023): "Volatility and the pricing kernel," *Swedish House of Finance Research Paper*.

TVERSKY, A. AND D. KAHNEMAN (1992): "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and Uncertainty*, 5, 297–323.

WU, G. AND R. GONZALEZ (1996): "Curvature of the probability weighting function," *Management Science*, 42, 1676–1690.

YAARI, M. (1986): "Univariate and multivariate comparisons of risk aversion: A new approach," in *Uncertainty, Information, and Communication. Essays in honor of Kenneth J. Arrow. Volume III. 1st Ed.*, ed. by W. Heller, R. Starr, and D. Starrett, Cambridge: Cambridge University Press, 173–188.

——— (1987): "The dual theory of choice under risk," *Econometrica*, 55, 95–115.

ZIEGLER, A. (2007): "Why does implied risk aversion smile?" *The Review of Financial Studies*, 20, 859–904.