Title

Optimal Measure Preserving Derivatives

Author

Beare, Brendan K.

Publication Date

2010-04-27

Optimal Measure Preserving Derivatives

Brendan K. Beare*†

University of California, San Diego

April 27, 2010

Abstract: Consider the collection of all derivative contracts written on an asset that deliver the same payoff distribution as a direct investment of $1 in the asset. We refer to the cheapest such derivative as the optimal measure preserving derivative. Using the Hardy-Littlewood rearrangement inequality, we obtain an explicit solution for the optimal measure preserving derivative in terms of the payoff distribution and pricing kernel of the underlying asset. The optimal measure preserving derivative corresponds to a direct investment of $1 in the underlying asset if and only if the pricing kernel is monotone decreasing. We obtain conditions under which an estimated optimal measure preserving derivative formed from estimates of the underlying payoff distribution and pricing kernel will be consistent in a particular sense. Building on an existing empirical study, we estimate the optimal measure preserving derivative for the S&P 500 index in October 1986 and April 1992, using a 31-day time horizon. We find that the precrash optimal derivative roughly coincides with a direct investment in the index, while the postcrash optimal derivative does not. The estimated price of the postcrash optimal derivative corresponds to nearly half a percentage point increase in monthly returns compared to a direct investment in the index.

Keywords and phrases: distributional price; Hardy-Littlewood inequality; Monge optimal transportation problem; pricing kernel puzzle; risk aversion puzzle; Ryff's decomposition.

JEL classifications: primary C14, G11; secondary D53.

---

*Address: Department of Economics, UCSD, 9500 Gilman Drive, La Jolla CA 92093-0508, USA. Email: bbeare@ucsd.edu.

†I thank Chris Chambers, Philip Dybvig, Jens Jackwerth, Ivana Komunjer, Mark Machina, Bruce McCullough, Eric Renault, Andres Santos, Xiaoxia Shi, Matthew Shum, Hal White, and seminar participants at UC San Diego and at the University of Pennsylvania for helpful comments. Jens Jackwerth graciously provided the density estimates used in the empirical part of the paper.

1 Introduction

Suppose we have some underlying asset – say, a market index – upon which we are considering writing a derivative contract such as a put or call option. A direct investment of \$1 in the underlying asset yields a random nonnegative payoff after one period distributed according to the probability measure μ. Profit after one period is equal to the random payoff minus the initial investment of \$1. A derivative written on the underlying asset is simply a function that maps the underlying payoff outcomes to other payoff outcomes. For instance, a European call option written at strike price s is characterized by the map $x \mapsto \max\{0, x - s\}$, while a European put option written at strike price s is characterized by the map $x \mapsto \max\{0, s - x\}$. The prices of derivatives written on the underlying asset can be obtained as their discounted expected payoffs under a probability measure ν referred to as the risk neutral measure for the underlying payoff.

In this paper we consider the following problem: given the measures μ and ν, what is the form of the cheapest derivative contract ϑ that achieves the minimum price among all derivatives that are measure preserving with respect to μ? Under some simple technical conditions given in the formal presentation of our model in the next section, we find an explicit formula for ϑ in terms of μ and ν. We refer to ϑ as the optimal measure preserving derivative. The optimal measure preserving derivative describes the cheapest way that an investor may obtain the payoff distribution of the underlying asset by purchasing a derivative written on that asset. When the underlying asset is a market index, the optimal measure preserving derivative is of particular interest: if it does not coincide with a direct investment in the market index – i.e. if ϑ is not the identity function – then a rational agent will never invest his entire wealth in the market index. This is because he may obtain a return distribution that first-order stochastically dominates the market return by instead investing all his wealth in derivatives of the form ϑ.

Our results in this paper build on important early work by Dybvig (1988a,b). Dybvig considered the collection of all derivatives written on the underlying asset that generate a fixed payoff distribution, and showed that any derivative that is not countermonotone with respect to the underlying pricing kernel cannot achieve the minimum price over that collection. The underlying pricing kernel π is defined as the Radon-Nikodym derivative of ν with respect to μ. For each x, $\pi(x)$ can be thought of as the nondiscounted price-per-probability-unit of an Arrow security paying \$1 when the underlying payoff is equal to x, and \$0 otherwise. A derivative θ is said to be countermonotone with respect to π if we have $\theta(x) \leq \theta(y)$ whenever $\pi(x) > \pi(y)$. Dybvig's results therefore imply that any derivative that provides the cheapest way to attain a given payoff distribution must pay more in less expensive states and less in more expensive states, where the relevant notion of state prices is given by the pricing kernel. In particular, the optimal measure preserving derivative must have this property.

In Section 2 of this paper we extend and formalize Dybvig's results by providing an explicit formula for ϑ in terms of μ and ν, or equivalently μ and π, and a formal proof of its validity. Dybvig proved his results carefully in the case where μ concentrates on finitely many states, but their extension to atomless μ omitted some details that are perhaps not obvious. We give a formal and relatively straightforward demonstration of the validity of our formula for ϑ in the atomless case using the Hardy-Littlewood rearrangement inequality, which was introduced recently in the field of economics by Carlier and Dana (2005) and other authors; see Remark 2.8 below. As one might guess from our discussion of Dybvig's results in the previous paragraph, our formula for ϑ in terms of μ and π depends on π only insofar as it depends on the linear preorder induced by π. When π is a monotone decreasing function, ϑ reduces to the identity function. In other words, when the pricing kernel is monotone decreasing, the optimal measure preserving derivative corresponds to a direct investment in the underlying asset. When π is not monotone decreasing, the optimal measure preserving derivative provides a way to achieve a return distribution that first-order stochastically dominates a direct investment in the underlying asset.

In Section 3 we develop the beginnings of an asymptotic theory for the estimation of optimal measure preserving derivatives. Suppose we form an estimate of ϑ by substituting estimates of μ and π into our formula for ϑ. We prove that if the estimates of μ and π are consistent in a particular sense, then the implied estimate of ϑ will also be consistent in a particular sense. The difficulty here lies in finding suitable notions of distance with which to define consistency of the three estimates. For μ we define consistency in terms of the total variation metric. For π we employ a pseudometric that depends only on the linear preorders induced by π and its estimate. For ϑ we use a version of the deformed L^1 pseudometric introduced in Beare (2009). We discuss conditions under which estimates of μ and π will be consistent with respect to the relevant notions of distance.

In Section 4 we apply our results on optimal measure preserving derivatives to US financial data from shortly before and after the stock market crash of 1987. Our application builds directly on an important empirical study by Jackwerth (2000). Jackwerth constructed nonparametric estimates of the actual and risk neutral distributions of monthly returns on the S&P 500 index using historical return data and option prices. We focus on Jackwerth's estimates for the dates October 21, 1986 and April 15, 1992. Jackwerth observed that the pricing kernel implied by his estimated actual and risk neutral distributions appeared to be monotone decreasing in the precrash sample period, but not monotone decreasing in the postcrash sample period. In particular, the pricing kernel appeared to be increasing between -3% and +3% monthly returns. Similar nonmonotone behavior has been identified by a number of other authors using a variety of postcrash sample periods; see Section 4 for references. The apparent nonmonotonicity of the postcrash pricing kernel has been referred to variously as the pricing kernel puzzle, risk aversion puzzle, or empirical pricing kernel paradox, as it appears to be inconsistent with standard representative agent models of market behavior. In Section 4 we take Jackwerth's estimates of the pre- and postcrash distributions, with which he kindly supplied us, and use them to compute estimates of the optimal measure preserving derivatives for the two sample periods. Consistent with our theoretical results, we find that the precrash optimal derivative largely coincides with a direct investment in the S&P 500 index, while the postcrash optimal derivative does not. We estimate that the postcrash optimal derivative is priced at nearly half of one cent less than a $1 investment in the index, despite delivering the same payoff distribution after one month.

We conclude the paper in Section 5 by discussing some directions in which our results may be improved or extended. Proofs of all theorems are collected in the Appendix.

2 A formula for the optimal derivative

Let μ be a probability measure on the nonnegative real line $\mathbb{R}_+$ equipped with its Borel σ-field $\mathcal{B}(\mathbb{R}_+)$. The measure μ represents the payoff distribution after one period of an investment of one monetary unit in some underlying asset; say, a market index. We are interested in the payoff distribution of derivative contracts written on the underlying asset. Such contracts may be represented by Borel measurable functions $\theta : \mathbb{R}_+ \to \bar{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{\infty\}$. A derivative contract θ yields a payoff of $\theta(x)$ monetary units after one period when the payoff from the underlying asset is x monetary units. Implicitly, we confine our attention to contracts that expire after one period, with no possibility of early exercise. Also, by requiring the payoff functions θ to be nonnegative, we restrict attention to contracts that are "self-collateralized". This rules out a derivative contract consisting solely of a short position in the underlying asset, but does not necessarily rule out a derivative contract formed from positions in several put or call options written on the underlying asset, with some positions being short and others long.

Let ν be another probability measure on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$. The measure ν represents the risk neutral probability distribution used to price derivative contracts written on the underlying asset. The price of a contract θ is equal to $\frac{1}{1+r}\int \theta \mathrm{d}\nu$, where r denotes the one-period risk-free interest rate. This is simply the discounted expected payoff of θ under the risk neutral measure. The price of a contract θ will be infinite if θ is not ν-integrable. The contract with payoff $\theta(x) = x$ corresponds to a direct investment of one monetary unit in the underlying asset, so we must have $1 + r = \int x \mathrm{d}\nu(x) < \infty$. Further, we require that $\int x \mathrm{d}\nu(x) \geq 1$ to ensure that the risk-free interest rate is nonnegative. In fact, the results given in this paper remain formally valid in the absence of these conditions on $\int x \mathrm{d}\nu(x)$.

We place the following technical conditions on μ and ν.

Assumption 2.1. The probability measures μ and ν satisfy the following three conditions:

1. μ and ν are atomless;
2. μ and ν are mutually absolutely continuous;
3. The Radon-Nikodym derivative of ν with respect to μ, denoted $\pi : \mathbb{R}_+ \to \mathbb{R}_+$, satisfies
$$\mu\{x : \pi(x) = y\} = 0$$
for all $y \in \mathbb{R}_+$.

Remark 2.1. As measures on $\mathcal{B}(\mathbb{R}_+)$, μ and ν are atomless if and only if the probability distribution functions they induce on $\mathbb{R}_+$ are continuous. This condition does not imply the existence of probability density functions with respect to Lebesgue measure. Such existence requires the absolute continuity of distribution functions.

Remark 2.2. The requirement that μ and ν are mutually absolutely continuous can be interpreted as a no arbitrage condition. If μ was not absolutely continuous with respect to ν, there would exist a set $B \in \mathcal{B}(\mathbb{R}_+)$ such that $\mu B > 0$ and $\nu B = 0$. One could then write a derivative contract with payoff function equal to the indicator function for B, and such a contract would have zero price but positive payoff with nonzero probability. A similar arbitrage opportunity arises if ν is not absolutely continuous with respect to μ.

Remark 2.3. The Radon-Nikodym derivative π is defined uniquely only up to μ-a.e. equivalence. This detail is of no importance to our results. We will refer to π as the pricing kernel for the underlying asset. Dybvig (1988a) refers to $(1 + r)\pi$ as the state-price density, and does not use the term pricing kernel. In more recent literature (see e.g. Aït-Sahalia and Lo, 1998) the term state-price density is used to refer to the Radon-Nikodym derivative of the risk neutral measure ν with respect to Lebesgue measure, while the term pricing kernel is used to refer to π, the Radon-Nikodym derivative of ν

with respect to μ. We follow the recent literature in our use of the term pricing kernel, and avoid using the term state-price density.

Remark 2.4. The requirement that $\mu\{x : \pi(x) = y\} = 0$ for all $y \in \mathbb{R}_+$ ensures that π is not flat on any set of positive μ-measure. This is important because, following Dybvig (1988a), we will use π to generate a ranking of the different points in $\mathbb{R}_+$. Intuitively, π describes the Arrow-Debreu price of securities paying one unit in exactly one "state of the world" $x \in \mathbb{R}_+$, normalized according to the probability of that state occurring. The ranking induced by π orders states according to how cheaply a unit of payoff may be obtained in that state, per unit of probability. Our nonflatness condition on π ensures that the induced ranking is essentially unique.

Our interest in this paper concerns a particular family of derivative contracts. Specifically, we are interested in derivatives that are measure preserving. A derivative θ is said to be measure preserving if the payoff distribution of θ is the same as the payoff distribution of an investment of one monetary unit in the underlying asset; that is, μ. Formally, for a Borel measurable map $\theta : \mathbb{R}_+ \to \bar{\mathbb{R}}_+$ to be measure preserving (with respect to μ) we require that $\mu = \mu\theta^{-1}$, where $\mu\theta^{-1}$ is the measure on $\mathcal{B}(\mathbb{R}_+)$ that assigns measure $\mu\{x \in \mathbb{R}_+ : \theta(x) \in B\}$ to each set $B \in \mathcal{B}(\mathbb{R}_+)$. Let Θ_μ denote the set of all such measure preserving maps. Note that the identity function on $\mathbb{R}_+$, which corresponds to a direct investment of one monetary unit in the underlying asset, is necessarily an element of Θ_μ.

Though all measure preserving derivatives yield the same payoff distribution, it is in general not the case that all measure preserving derivatives are priced equally by the market. If a rational agent were to invest all his wealth in one of the measure preserving derivatives written on some underlying asset, we would expect him to choose whichever measure preserving derivative is the cheapest. We refer to the cheapest measure preserving derivative as the optimal measure preserving derivative, or occasionally as simply the optimal derivative. The price of the optimal measure preserving derivative is what Dybvig (1988a) calls the distributional price of μ. If the distributional price of μ is less than one, then the optimal measure preserving derivative costs less than one monetary unit to purchase, and yields the same payoff distribution as an investment of one monetary unit directly in the underlying asset. Thus, an agent who invests all his wealth in the optimal measure preserving derivative will obtain a payoff distribution that first-order stochastically dominates the payoff distribution he would obtain by investing all his wealth directly in the underlying asset. We shall see shortly that the distributional price of μ is less than one whenever the pricing kernel π is not monotone decreasing.

Before stating our main result, we require some additional notation. Given a finite measure m on $\mathcal{B}(\mathbb{R}_+)$, define the distribution function $F_m : \bar{\mathbb{R}}_+ \to [0,1]$ of m by

$$F_m(x) = m\{y \in \mathbb{R}_+ : y \leq x\},$$

and define the quantile function $Q_m : [0,1] \to \bar{\mathbb{R}}_+$ of m by

$$Q_m(u) = \inf\{y \in \mathbb{R}_+ : F_m(y) \geq u\}.$$

Our main result is as follows.

Theorem 2.1. *Let $\vartheta : \mathbb{R}_+ \to \bar{\mathbb{R}}_+$ be given by*

$$\vartheta(x) = Q_\mu(1 - F_{\mu\pi^{-1}}(\pi(x))). \tag{2.1}$$

Then, under Assumption 2.1,

1. $\vartheta \in \Theta_\mu$*;*
2. $\int \vartheta \, d\nu = \inf_{\theta \in \Theta_\mu} \int \theta \, d\nu$*; and*
3. $(\vartheta(x) - \vartheta(y))(\pi(x) - \pi(y)) \leq 0$ *for all* $(x, y) \in \mathbb{R}_+^2$*.*

Remark 2.5. Theorem 2.1 gives an explicit formula for the optimal measure preserving derivative ϑ in terms of the underlying measure μ and pricing kernel π. This is equation (2.1). Note that empirical methods exist for estimating both μ and π; we will discuss this issue further in subsequent sections. Part 1 of Theorem 2.1 says that ϑ is indeed a measure preserving derivative. Part 2 of Theorem 2.1 says that ϑ achieves the minimum price among all measure preserving derivatives. Part 3 of Theorem 2.1 says that ϑ and π are countermonotone; that is, ϑ assigns payoffs to states in such a way that the payoff is greater when the pricing kernel is lower. The fact that the cheapest measure preserving derivative must assign payoffs in this fashion is the content of Theorem 1 of Dybvig (1988a). Theorem 2.1 extends Dybvig's result by giving an explicit formula for ϑ in terms of μ and π, and a complete proof of its validity.

Remark 2.6. Theorem 3 of Dybvig (1988a) gives a formula for the cheapest price at which a given payoff distribution may be obtained using a derivative written on the underlying asset. This quantity is referred to as the distributional price of the payoff distribution in question. Dybvig's formula implies that the distributional price of the underlying payoff distribution is $\frac{1}{1+r} \int_0^1 Q_\mu(u) Q_{\mu\pi^{-1}}(1-u) \mathrm{d}u$. It is straightforward to verify that our optimal measure preserving derivative achieves this price: using the change of variables $\pi(x) \mapsto y$, we can see that $1+r$ times the price of ϑ is given by

$$\int_0^\infty \vartheta(x)\mathrm{d}\nu(x) = \int_0^\infty Q_\mu(1 - F_{\mu\pi^{-1}}(\pi(x)))\pi(x)\mathrm{d}\mu(x) = \int_0^\infty Q_\mu(1 - F_{\mu\pi^{-1}}(y))y\mathrm{d}\mu\pi^{-1}(y).$$

With a second change of variables $F_{\mu\pi^{-1}}(y) \mapsto u$, we obtain

$$\int_0^\infty Q_\mu(1 - F_{\mu\pi^{-1}}(y))y\mathrm{d}\mu\pi^{-1}(y) = \int_0^1 Q_\mu(u)Q_{\mu\pi^{-1}}(1-u)\mathrm{d}u,$$

as claimed. This was pointed out to me by Dybvig in private communication.

Remark 2.7. Though our formula for the optimal measure preserving derivative ϑ depends on both μ and π, the dependence of ϑ on π extends only to the linear preorder on $\mathbb{R}_+$ induced by π. This linear preorder can be represented by the set

$$L_\pi = \{(x, y) \in \mathbb{R}_+^2 : \pi(x) \leq \pi(y)\}.$$

The quantity $F_{\mu\pi^{-1}}(\pi(x))$ appearing in equation (2.1) satisfies

$$F_{\mu\pi^{-1}}(\pi(x)) = \mu\{y \in \mathbb{R}_+ : \pi(y) \leq \pi(x)\} = \mu\{y \in \mathbb{R}_+ : (y, x) \in L_\pi\}$$

for each $x \in \mathbb{R}_+$, and therefore depends on π only through L_π.

Remark 2.8. The main tool used to prove Theorem 2.1 is the Hardy-Littlewood inequality. The Hardy-Littlewood inequality was originally stated in chapter 10 of the classic book on inequalities by Hardy, Littlewood and Pòlya (1934), and appears to have first been explicitly applied in economics in the early 21st century: see e.g. Renou and Carlier (2003), Carlier and Dana (2005), and Carlier, Dana and Galichon (2009). An earlier application, made apparently without knowledge of the existing mathematical literature on the subject, appears in Becker (1973): the results on optimal sorting in the first section of the appendix to the paper constitute a version of the Hardy-Littlewood inequality for discrete sums of smooth supermodular bivariate functions. Becker attributes the proofs in that section to William Brock.

We now state a version of the Hardy-Littlewood inequality, which we will apply in the proof of Theorem 2.1 given in the Appendix. Given nonnegative extended real valued measurable functions f, g defined on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), m)$, with m a finite atomless measure, define the nonincreasing rearrangement of f by

$$f_*(x) = Q_{mf^{-1}}(1 - F_m(x)),$$

and the nondecreasing rearrangement of g by

$$g^*(x) = Q_{mg^{-1}}(F_m(x)),$$

where we let x range over $\mathbb{R}_+$. The Hardy-Littlewood inequality states that

$$\int f_* g^* \mathrm{d}m \leq \int fg\mathrm{d}m.$$

We refer the reader to Chong and Rice (1971) for a more detailed account of monotone rearrangements and the Hardy-Littlewood inequality; many other references also exist.

Remark 2.9. Theorem 2.1 is closely related to Ryff's decomposition (Ryff, 1970; see also Theorem 6.2 in Chong and Rice, 1971, and Proposition 2 in Carlier and Dana, 2005). Ryff's decomposition states that, for every extended real valued measurable function f on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), m)$, with m finite and atomless, there exists a measure preserving map $\phi_f : \mathbb{R}_+ \to \mathbb{R}_+$ such that $f = f_* \circ \phi_f$ m-a.e. The optimal measure preserving derivative ϑ in Theorem 2.1 is the measure preserving map ϕ_π appearing in Ryff's decomposition of π, and satisfies $\pi = \pi_* \circ \vartheta$ μ-a.e. This fact is established in the proof of Theorem 2.1 given in the Appendix.

Remark 2.10. The optimization problem solved in Theorem 2.1 is a special case of the Monge optimal transportation problem. Given measures m_1 and m_2 on the real line with equal total mass, and a cost function $c : \mathbb{R} \times \mathbb{R} \to \bar{\mathbb{R}}_+$, the Monge optimal transportation problem requires us to minimize $\int c(x, T(x))\mathrm{d}m_1(x)$ over all measurable maps T such that $m_1 T^{-1} = m_2$. This problem has a physical interpretation: it involves finding the cheapest way to move a mass of particles distributed according to m_1 so that it is instead distributed according to m_2, when the per-unit cost of moving mass from point x to point y is given by $c(x, y)$. Refer to Villani (2003) for further discussion of optimal transportation problems. If we set $m_1 = m_2 = \mu$ and $c(x, y) = \pi(x)y$, the Monge optimal transportation problem reduces to our own. No use of transportation theory is made in this paper; we solve our optimization problem directly using the Hardy-Littlewood inequality.

As stated earlier, the identity function on $\mathbb{R}_+$ is necessarily an element of Θ_μ. Since this derivative corresponds to a direct investment of one monetary unit in the underlying asset, one might wonder under what conditions our optimal measure preserving derivative

ϑ corresponds to the identity function. In fact, there is a very simple answer to this question: neglecting qualifications about sets of measure zero, ϑ is the identity function if and only if the pricing kernel is monotone decreasing. We will state this result as a separate theorem.

Theorem 2.2. *Under Assumption 2.1, the function ϑ defined in (2.1) satisfies*

$$\mu\{x \in \mathbb{R}_+ : \vartheta(x) = x\} = 1$$

if and only if

$$\mu \otimes \mu\{(x, y) \in \mathbb{R}_+^2 : (\pi(x) - \pi(y))(x - y) \leq 0\} = 1.$$

Remark 2.11. If the underlying asset with payoff distribution μ is taken to be a market index, then a monotone decreasing pricing kernel is precisely what basic economic intuition would suggest. Output is more scarce when the market index is low, so Arrow securities yielding payoffs in such states should command a high price per unit of probability. Conversely, Arrow securities yielding payoffs when output is plentiful should command a lower price per unit of probability. In this scenario, the pricing kernel is monotone decreasing, and so Theorem 2.2 implies that the optimal measure preserving derivative corresponds to a direct investment in the market index. Other measure preserving derivatives are priced more highly because they allow an investor to hedge against poor market performance.

Remark 2.12. As discussed in the first section of this paper, there is substantial recent empirical evidence that the pricing kernel corresponding to monthly returns for a major US market index – the S&P 500 – is not monotone decreasing. Jackwerth (2000) found that the pricing kernel for monthly S&P 500 returns during 1988-1995 was nonmonotone, exhibiting an increasing region around the middle of the return distribution. Similar findings have been documented by other researchers, cited in Section 4, for various market indices over a range of timeframes post-1987. If it is indeed true that the pricing kernel for a market index is nonmonotone, Theorems 2.1 and 2.2 imply that the one-period return obtained by investing one monetary unit in the index can be first-order stochastically dominated by investing one monetary unit in derivatives written on the index in accordance with equation (2.1). This will be discussed in more detail in Section 4.

3 Asymptotic theory for optimal derivative estimation

In the previous section we gave an explicit formula, equation (2.1), for the optimal measure preserving derivative written on some underlying asset. The formula is determined entirely by two objects: the underlying measure μ and the pricing kernel π. Though these objects are typically unknown, they are estimable. One may thus consider forming an estimate for the optimal measure preserving derivative by substituting estimates of μ and π into equation (2.1). In this section we will identify conditions on the estimates of μ and π that are sufficient to ensure that the estimated optimal derivative is well behaved. Readers bored by asymptotic theory may skip to Section 4 without loss of continuity.

Let $(\Omega, \mathcal{A}, P)$ be a probability space. Let $\mathcal{M}$ be the set of all probability measures on $\mathcal{B}(\mathbb{R}_+)$, and suppose that for each $n \in \mathbb{N}$ we have a map $\hat{\mu}_n : \Omega \to \mathcal{M}$. The sequence of maps $\hat{\mu}_n$ is intended to represent a sequence of estimates of μ. We also need to define a sequence of estimates of π. Let $\mathcal{F}$ denote the set of all Borel measurable functions from $\mathbb{R}_+$ to $\bar{\mathbb{R}}_+$. Suppose that for each $n \in \mathbb{N}$ we have a map $\hat{\pi}_n : \Omega \to \mathcal{F}$. The sequence of maps $\hat{\pi}_n$ is intended to represent a sequence of estimates of π. Note that we have not required the maps $\hat{\mu}_n : \Omega \to \mathcal{M}$ and $\hat{\pi}_n : \Omega \to \mathcal{F}$ to satisfy a measurability condition. Our main result in this section, Theorem 3.3 below, is stated in terms of convergence in outer probability.

One may interpret the subscript n as the sample size used to form the estimators $\hat{\mu}_n$ and $\hat{\pi}_n$ if this is helpful, but formally we abstract from the notion of data and define our estimators directly on the underlying probability space. Note that in applications the estimates $\hat{\mu}_n$ and $\hat{\pi}_n$ are typically formed using both historical data on the payoff of the underlying asset, and current data on the prices of options written on the underlying asset at a range of strike prices, so it is not entirely clear that the size of the sample is best described by a single number.

In order to characterize the approximation of μ and π by $\hat{\mu}_n$ and $\hat{\pi}_n$, we require a notion of distance on $\mathcal{M}$ and $\mathcal{F}$. We equip $\mathcal{M}$ with the total variation metric d_1, which assigns distance

$$d_1(m_1, m_2) = \sup_{B \in \mathcal{B}(\mathbb{R}_+)} |m_2 B - m_1 B|$$

to measures $m_1, m_2 \in \mathcal{M}$. The notion of distance we attach to $\mathcal{F}$ is less standard. Recall from Remark 2.7 that the optimal measure preserving derivative ϑ depends on π only through the induced linear preorder L_π. For this reason, we will not require our estimated pricing kernel $\hat{\pi}_n$ to directly approximate π. Rather, we will require that the linear preorder induced by $\hat{\pi}_n$ provide a good approximation to L_π, in the sense that the symmetric difference of these two sets has $\mu \otimes \mu$-measure tending to zero in outer probability as $n \to \infty$. The linear preorder induced by a function $f \in \mathcal{F}$ is the set $L_f := \{(x, y) \in \mathbb{R}_+^2 : f(x) \leq f(y)\}$. We equip $\mathcal{F}$ with the pseudometric d_2 given by

$$d_2(f_1, f_2) = \mu \otimes \mu(L_{f_1} \triangle L_{f_2})$$

for $f_1, f_2 \in \mathcal{F}$, where $\triangle$ is the symmetric difference operation. $(\mathcal{F}, d_2)$ may be viewed as a metric space rather than a pseudometric space if we identify functions in $\mathcal{F}$ that generate linear preorders whose symmetric difference is of zero $\mu \otimes \mu$-measure.

The following assumption describes the way in which we require $\hat{\mu}_n$ and $\hat{\pi}_n$ to approximate μ and π.

Assumption 3.1. As $n \to \infty$, the maps $\hat{\mu}_n$ and $\hat{\pi}_n$ satisfy

1. $\hat{\mu}_n \rightsquigarrow \mu$ under d_1, and

2. $\hat{\pi}_n \rightsquigarrow \pi$ under d_2,

where "$\rightsquigarrow$" denotes convergence in outer probability.

Remark 3.1. We follow Definition 1.9.1(i) in van der Vaart and Wellner (1996) in defining convergence in outer probability. Let P^* be the outer measure corresponding to P. The statement that $\hat{\mu}_n \rightsquigarrow \mu$ under d_1 is equivalent to the claim that

$$P^* \{\omega \in \Omega : d_1(\hat{\mu}_n(\omega), \mu) > \varepsilon\} \to 0$$

for any $\varepsilon > 0$, while the statement that $\hat{\pi}_n \rightsquigarrow \pi$ under d_2 is equivalent to the claim that

$$P^* \{\omega \in \Omega : d_2(\hat{\pi}_n(\omega), \pi) > \varepsilon\} \to 0$$

for any $\varepsilon > 0$. Convergence in outer probability is equivalent to weak convergence to the probability measure assigning unit mass to the limit point, with weak convergence defined in the sense of Definition 1.3.3 in van der Vaart and Wellner (1996).

Remark 3.2. Requiring that $\hat{\mu}_n \rightsquigarrow \mu$ in the total variation metric rules out some simple estimators of μ, but is not a particularly strong condition. If $\hat{\mu}_n$ is the empirical measure formed from a sample of n draws from μ, then we will not have $\hat{\mu}_n \rightsquigarrow \mu$ because $\hat{\mu}_n$ is concentrated on a finite set whereas μ is atomless. Consequently the total variation distance between $\hat{\mu}_n$ and μ will be one for all n. Well specified parametric likelihood-based estimates of μ will typically satisfy $\hat{\mu}_n \rightsquigarrow \mu$ under standard conditions, as the total variation distance between probability measures is bounded by the square root of twice their Kullback-Leibler divergence, by Pinkser's inequality. The total variation distance between probability measures that are absolutely continuous with respect to Lebesgue measure is equal to half the L^1 distance between their probability density functions, and so if μ admits a density function then an estimate of μ formed from an L^1-consistent nonparametric estimate of the density of μ will satisfy $\hat{\mu}_n \rightsquigarrow \mu$. Scheffé's theorem implies that any estimate of the density of μ that is pointwise strongly consistent on a set of μ-measure one will also satisfy L^1-consistency.

Remark 3.3. The assumption that $\hat{\pi}_n \rightsquigarrow \pi$ under d_2 is also fairly mild, and likely to be satisfied by reasonable estimators of π under standard conditions. In fact, under an additional measurability condition, it suffices that $\hat{\pi}_n$ is pointwise strongly consistent on a set of μ-measure one. We state this result as a separate theorem.

Theorem 3.1. *Suppose that, for each $n \in \mathbb{N}$, $(\omega, x) \mapsto \hat{\pi}_n(\omega)(x)$ is a measurable map from $(\Omega \times \mathbb{R}_+, \mathcal{A} \otimes \mathcal{B}(\mathbb{R}_+))$ to $(\bar{\mathbb{R}}_+, \mathcal{B}(\bar{\mathbb{R}}_+))$. Suppose further that, for μ-a.e. $x \in \mathbb{R}_+$, we have $\hat{\pi}_n(\omega)(x) \to \pi(x)$ as $n \to \infty$ for P-a.e. $\omega \in \Omega$. Then, under Assumption 2.1, we have $\hat{\pi}_n \rightsquigarrow \pi$ under d_2 as $n \to \infty$.*

The ability to move from μ-a.e. pointwise strong consistency to weak convergence under d_2 is very useful. For instance, under the assumption that ν and μ admit densities with respect to Lebesgue measure, Jackwerth (2000) proposed to estimate π by taking the ratio of individual estimates of those densities. The global asymptotic properties of such a ratio are difficult to describe because of aberrant behavior in regions where the density of μ is zero or close to zero. If, however, the two density estimates individually satisfy pointwise strong consistency, then their ratio is pointwise strongly consistent for π on the set where μ has positive density, which is automatically of μ-measure one.

Our main result in this section establishes the consistency of an estimate of ϑ formed by substituting $\hat{\mu}_n$ and $\hat{\pi}_n$ for μ and π in equation (2.1). The pseudometric d_2 is not suitable for characterizing the approximation of ϑ by $\hat{\vartheta}_n$, as we seek to estimate more than merely the linear preorder induced by ϑ, so we require a new notion of distance on $\mathcal{F}$. Define the pseudometric d_3 on $\mathcal{F}$ by

$$d_3(f_1, f_2) = \int_{\mathbb{R}_+} |F_\mu(f_2(x)) - F_\mu(f_1(x))| \mathrm{d}\mu(x)$$

for $f_1, f_2 \in \mathcal{F}$. Again, we may view $(\mathcal{F}, d_3)$ as a metric space rather than a pseudometric space if we identify functions $f_1, f_2 \in \mathcal{F}$ for which $d_3(f_1, f_2) = 0$. A more general version of d_3 has been employed by the author in a related context (Beare, 2009), where it is used to characterize the approximation of a function that transforms one measure to another. The suitability of d_3 for the present problem lies in the following result.

Theorem 3.2. *Suppose we have a sequence of functions $\vartheta_1, \vartheta_2, \ldots$ in $\mathcal{F}$ such that $d_3(\vartheta_n, \vartheta) \to 0$ as $n \to \infty$. Under Assumption 2.1, as $n \to \infty$ the following statements are true:*

1. *The measure $\mu\vartheta_n^{-1} \in \mathcal{M}$ converges weakly to μ; and*
2. *$\mu\{x \in \mathbb{R}_+ : |\vartheta_n(x) - \vartheta(x)| > \varepsilon\} \to 0$ for any $\varepsilon > 0$.*

Remark 3.4. The first part of Theorem 3.2 says that if $d_3(\vartheta_n, \vartheta) \to 0$, then we can expect ϑ_n to be approximately measure preserving for large n. The second part of Theorem 3.2 says that if $d_3(\vartheta_n, \vartheta) \to 0$, then ϑ_n converges in μ-measure to ϑ. Note that we have not required ϑ_n or ϑ to be bounded or even μ-integrable. The pseudometric d_3 allows us to ignore this issue, as $F_\mu \circ f$ is necessarily bounded and measurable for any $f \in \mathcal{F}$.

Define the map $h : \mathcal{M} \times \mathcal{F} \to \mathcal{F}$ by

$$h(m, f)(\cdot) = Q_m(1 - F_{mf^{-1}}(f(\cdot))).$$

Our sequence of estimated optimal measure preserving derivatives $\hat{\vartheta}_n : \Omega \to \mathcal{F}$ is defined by

$$\hat{\vartheta}_n(\omega) = h(\hat{\mu}_n(\omega), \hat{\pi}_n(\omega)).$$

The main result of this section establishes that $\hat{\vartheta}_n$ is a consistent estimator of ϑ, in a particular sense.

Theorem 3.3. *Under Assumptions 2.1 and 3.1, we have $\hat{\vartheta}_n \rightsquigarrow \vartheta$ under d_3 as $n \to \infty$.*

Remark 3.5. Let d_{12} be the pseudometric on $\mathcal{M} \times \mathcal{F}$ given by

$$d_{12}((m_1, f_1), (m_2, f_2)) = \max\{d_1(m_1, m_2), d_2(f_1, f_2)\}$$

for $(m_1, f_1), (m_2, f_2) \in \mathcal{M} \times \mathcal{F}$. In view of Assumption 3.1, the continuous mapping theorem implies that $\hat{\vartheta}_n = h(\hat{\mu}_n, \hat{\pi}_n) \rightsquigarrow h(\mu, \pi) = \vartheta$ under d_3 provided that h is continuous at (μ, π) with respect to d_{12} on its domain and d_3 on its range. The demonstration of such continuity is the content of the proof of Theorem 3.3 given in the Appendix.

4 Empirical illustration

In this section we apply our formula for the optimal measure preserving derivative to monthly returns for the S&P 500 index. In an important paper, Jackwerth (2000) applied nonparametric methods using option price data to study the actual and risk neutral distributions of monthly S&P 500 returns. He found that the shape of the pricing kernel changed dramatically around the time of the 1987 stock market crash: precrash, the pricing kernel appeared to be monotone decreasing, while postcrash it appeared to be increasing in a region around the center of the return distribution, and decreasing elsewhere. This pattern appeared consistently across the postcrash sample period, spanning 1988-1995. Aït-Sahalia and Lo (2000) found a similar pattern in the pricing kernel for six month S&P 500 returns in 1993 (see Figure 3 in their paper), as did Rosenberg and Engle (2002) for monthly returns in 1991-1995 (see Figures 5 and 6 in their paper; note that the power specification in Figure 6 is decreasing by construction). Chabi-Yo, Garcia and Renault (2008) re-examined Jackwerth's original data set, confirming his findings (Figure 4 in their paper). Jackwerth (2004) identified nonmonotone pricing kernel behavior in 2003 for the S&P 500, German DAX 30, United Kingdom FTSE 100 and Japanese Nikkei 225 indices (Figure 11 in his paper; note that the estimated pricing kernel for the Nikkei 225 appears very different to the others, and is monotone decreasing at non-extreme return levels.). Violations of pricing kernel monotonicity for the DAX 30 in the early 2000s have also been documented by Golubev, Härdle and Timonfeev (2008) and Härdle, Okhrin and Wang (2010). For further discussion, see Brown and Jackwerth (2004) and Constantinides, Jackwerth and Perrakis (2009).

The apparent nonmonotonicity of the pricing kernel for market indices has been referred to variously as the pricing kernel puzzle, risk aversion puzzle, and empirical pricing kernel paradox. Nonmonotonicity is puzzling from a theoretical perspective because the results of Dybvig (1988a) imply that a rational agent will not invest his entire wealth in the market portfolio. The idea of a representative agent therefore seems paradoxical. Jackwerth (2000) and other authors cited in the previous paragraph have observed that the absolute and relative risk aversion coefficients implied by first-order conditions for optimization by a representative agent are negative in regions where the pricing kernel increases. Further, the implied risk aversion coefficients return to positive levels beyond the region in which the pricing kernel is increasing, thereby contradicting the assumption of nonincreasing absolute risk aversion that is basic to much of the theoretical literature on choice under uncertainty; see e.g. Machina (1982).

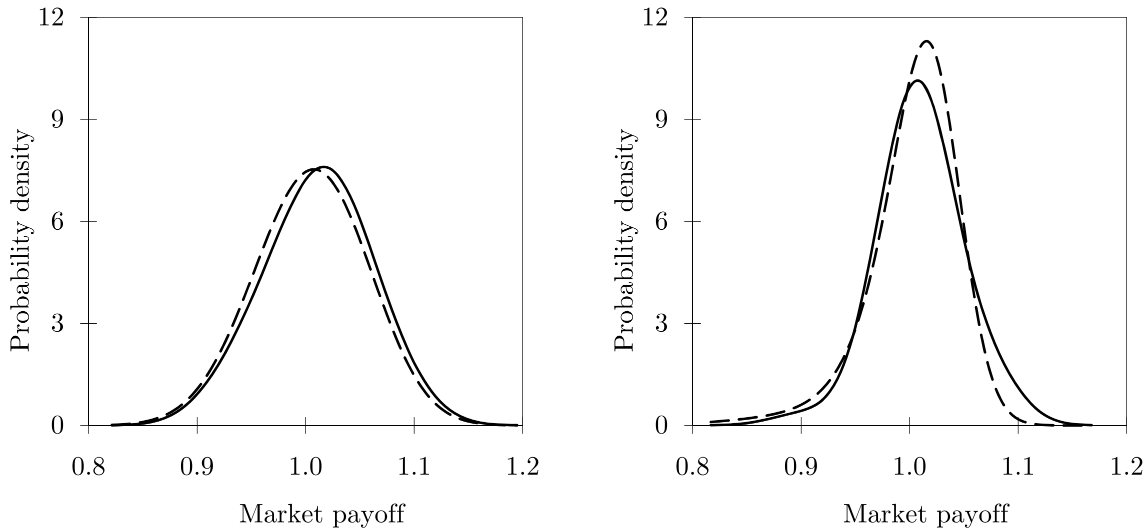
Various explanations have been proposed for the pricing kernel puzzle. Brown and

Jackwerth (2004) and Chabi-Yo, Garcia and Renault (2008) argued that nonmonotone pricing kernel estimates may reflect a failure to condition on relevant state variables used by investors to form their beliefs about the distribution of market returns. Ziegler (2007) showed that a nonmonotone pricing kernel can arise in a model where agents have heterogeneous beliefs about the distribution of market returns. Detlefsen, Härdle and Moro (2007) and Härdle, Krätschmer and Moro (2009) proposed a model in which agents switch between bullish and bearish attitudes about the market return depending on the current market outcome, with the switching point differing between agents, and showed that this model can generate a nonmonotone pricing kernel. On a more basic level, one may question whether the S&P 500 index or other market indices can truly be thought of as market aggregates in the stylized sense used in theoretical models.

In this section we present the optimal measure preserving derivatives corresponding to the estimates given by Jackwerth (2000) for the actual and risk neutral densities of 31-day S&P 500 returns before and after the 1987 stock market crash. Jackwerth's estimates for October 21, 1986, and April 15, 1992, displayed in Figures 1 and 2 of his paper, are reproduced in Figure 4.1 of this paper. In Figure 4.2 we display the pricing kernels implied by Jackwerth's estimates, and in Figure 4.3 we display the implied optimal measure preserving derivatives. In all our graphs, the horizontal axes measure the 31-day payoff from a $1 investment in the index. In Figures 1 and 2 of Jackwerth (2000), the units of the horizontal axes were normalized so that the pre- and postcrash actual distributions implied an expected payoff of exactly 1. Here, we rescale the horizontal axes so that the pre- and postcrash risk neutral distributions imply a risk neutralized expected payoff of exactly $1+r$, where the 31-day risk-free interest rate r is set equal to the one month LIBOR rate for the appropriate date. This makes the distributions in Figure 4.1 consistent with our definitions of μ and ν in this paper.

Panel (a) in Figure 4.1 shows the estimated actual and risk neutral distributions for 31-day S&P 500 returns on October 21, 1986. Jackwerth estimated the actual distribution by applying a kernel smoother to nonoverlapping 31-day returns over the four years prior to that date, and the risk neutral distribution by applying a variation of the method of Jackwerth and Rubinstein (1996) to current 31-day option prices. Jackwerth's estimates of the same quantities for April 15, 1992, are shown in panel (b) of Figure 4.1. As observed by Jackwerth, the risk neutral distribution appears to change shape substantially pre- and postcrash, becoming left-skewed in the latter period. The actual return distribution appears to concentrate more tightly around its mean in the postcrash period.

In Figure 4.2 we display the pricing kernels implied by the actual and risk neutral distributions in Figure 4.1. Each pricing kernel in Figure 4.2 is simply the ratio of the

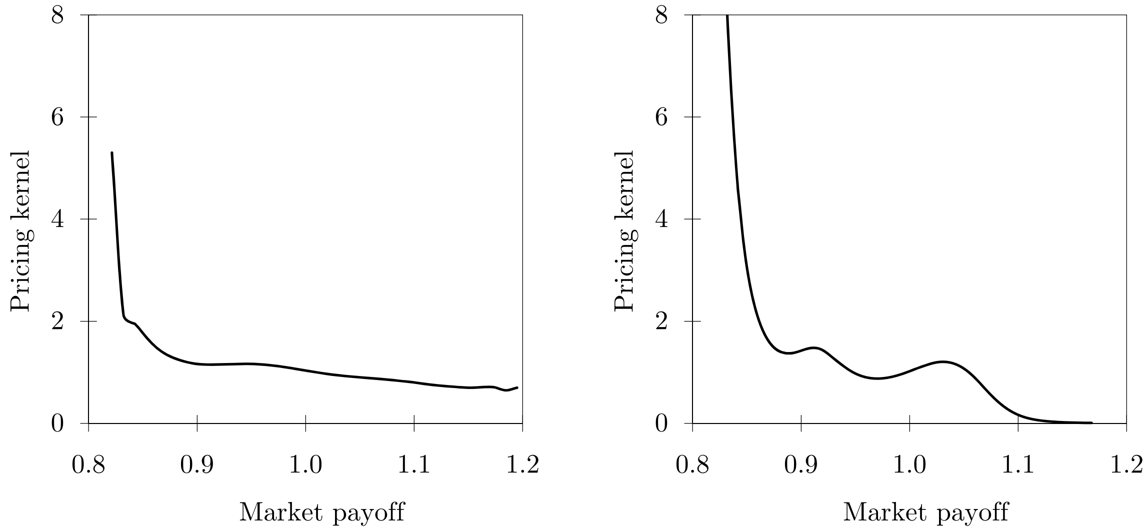


(a) October 21, 1986 (b) April 15, 1992

FIGURE 4.1: Estimated actual and risk neutral distributions for the 31-day payoff from a \$1 investment in the S&P 500 index. Actual distributions are given by solid lines and risk neutral distributions by dashed lines. Taken from Figures 1 and 2 in Jackwerth (2000).

risk neutral density to actual density from the corresponding panel in Figure 4.1. In panel (a), we see that the estimated precrash pricing kernel appears to be nonincreasing, consistent with standard economic models. In fact, the estimated pricing kernel increases from approximately 1.15 to 1.16 between -8% and -5% 31-day returns, but this increase is too small to be visually discernible and may plausibly be attributed to estimation error. There is also a minor violation of monotonicity at very high return levels. In panel (b), the estimated postcrash pricing kernel exhibits more pronounced nonmonotonicity. The estimated pricing kernel rises from approximately 1.37 to 1.48 between -11% and -9% 31-day returns, and from approximately 0.88 to 1.20 between -3% and +3% 31-day returns. It is the increasing region between -3% and +3% that Jackwerth focuses on in his discussion of the risk aversion puzzle. This region has also been identified by other empirical studies, cited at the beginning of this section, using a variety of postcrash sample periods and estimation methods. The smaller increasing region between -11% and -9% 31-day returns does not appear to have been identified by other studies, and is not discussed in the paper by Jackwerth.

In Figure 4.3 we display the optimal measure preserving derivatives implied by the actual and risk neutral distributions in Figure 4.1, calculated using equation (2.1) in Theorem 2.1 above. Numerical computation was implemented using Ox version 5.10; see Doornik (2007). Panel (a) displays the precrash optimal derivative. If the pricing kernel



(a) October 21, 1986 (b) April 15, 1992

FIGURE 4.2: Pricing kernels implied by estimated distributions in Figure 4.1.

in panel (a) of Figure 4.2 was monotone decreasing, the optimal derivative in panel (a) of Figure 4.3 would be the 45° line, by Theorem 2.2. The minor departure of the estimated precrash pricing kernel from monotonicity causes the estimated precrash optimal derivative to deviate slightly from the 45° line, with relatively small fluctuations appearing around the regions where the pricing kernel increases. In panel (b), the estimated postcrash optimal derivative exhibits much more substantial departures from the 45° line. In particular, it decreases sharply in the region between -3% and +3% 31-day returns where the estimated pricing kernel is increasing, and increases more rapidly than the 45° line to either side of this region. There is also a smaller fluctuation corresponding to the more mildly increasing segment of the pricing kernel between -11% and -9% 31-day returns.

By construction, the estimated optimal measure preserving derivatives in Figure 4.3 deliver the same payoff distribution as a direct investment in the S&P 500 index under the actual return distributions given in Figure 4.1. The nonmonotonicity of the pricing kernels in Figure 4.2 implies that the optimal derivatives in Figure 4.3 should be priced at less than \$1 under the risk neutral distributions given in Figure 4.1, by Theorem 2.1. We can calculate the prices of the optimal derivatives in Figure 4.3 by taking their discounted expected payoffs under the risk neutral distributions given in Figure 4.1. Doing so, and rounding to four decimal places, we find that the precrash optimal derivative is priced at \$1.0000, while the postcrash optimal derivative is priced at \$0.9954. The fact that the precrash optimal derivative is effectively priced at \$1 reflects the fact that the precrash

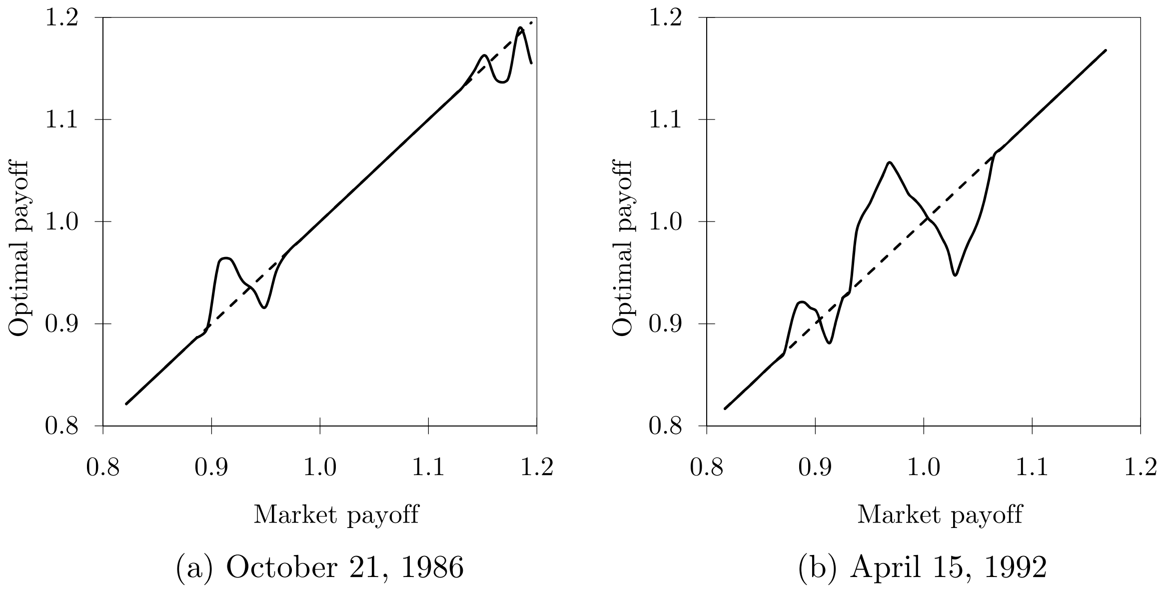


FIGURE 4.3: Optimal measure preserving derivatives implied by estimated distributions in Figure 4.1. Dashed lines are 45° lines.

pricing kernel is effectively nonincreasing. By comparison, the price of the postcrash optimal derivative is substantially less than \$1, reflecting the pronounced nonmonotonicity of the postcrash pricing kernel. To put the figure of \$0.9954 in context, observe that the price differential of \$0.0046 represents nearly half a percentage point increase in monthly returns, compared to a direct investment in the index. This is a very substantial gain.

It is natural to consider a piecewise linear approximation to the postcrash optimal derivative using a portfolio formed from investments in the underlying index and in 31-day European options. The most striking feature of the postcrash optimal derivative is the N-shaped fluctuation around the middle of the return distribution. This feature could be reproduced in an approximating portfolio by combining a direct investment in the index with heavy short and long positions in 31-day call options with strikes at -3% and +3% respectively, and more moderate long and short positions in 31-day call options with strikes at -6% and +6% respectively. Of course, in-the-money call options are rarely traded, but can be replicated by positions in out-of-the-money put options, the underlying index, and risk-free bonds. A simple portfolio of this kind may be expected to perform rather well in the presence of a pricing kernel of the kind shown in Figure 4.1(b).

5 Conclusion

In this paper we have given an explicit formula for the optimal measure preserving derivative written on some underlying asset. The formula depends on the pricing kernel and payoff distribution for the underlying asset, and corresponds to a direct investment in the asset if and only if the pricing kernel is nonincreasing. We have also given simple sufficient conditions for consistent estimation of the optimal measure preserving derivative using estimates of the pricing kernel and underlying payoff distribution. Building on an empirical study by Jackwerth (2000), we estimated the optimal measure preserving derivative for the S&P 500 index in October 1986 and April 1992, using a 31-day time horizon. Consistent with our expectations, we found that the precrash optimal derivative roughly coincides with a direct investment in the index, while the postcrash optimal derivative does not. The estimated price of the postcrash optimal derivative corresponds to nearly half a percentage point increase in monthly returns compared to a direct investment in the index.

We will conclude by discussing several directions in which our results may be extended. In Section 2, we built our results in a framework where there is a single underlying asset, so that the actual and risk neutral distributions are defined on the nonnegative real line. It may be useful to consider a more general framework in which there are multiple underlying assets, so that the actual and risk neutral distributions are defined on the positive orthant of multidimensional Euclidean space. The objective would then be to find the cheapest real valued function of several variables achieving some target return distribution. Dybvig's (1988a) results continue to apply in a multidimensional setting, implying that the cheapest such function must be countermonotone with respect to the multidimensional pricing kernel. Derivation of the precise form of the cheapest function is, however, complicated by difficulties that arise in dealing with probability integral transforms and quantile functions when working with a multivariate underlying distribution.

In Section 3, we obtained conditions under which an estimate of the optimal measure preserving derivative satisfies a version of consistency. This result is reassuring because it suggests that good estimates of the pricing kernel and underlying payoff distribution can be expected to yield a good estimate of the optimal derivative. It does not provide any means by which confidence intervals for the optimal derivative may be obtained, or hypotheses about the optimal derivative tested. The development of an asymptotic distributional theory for optimal derivative estimation seems an important but difficult task. Even the pointwise behavior of the estimated derivative must depend on the global behavior of the estimated pricing kernel and underlying payoff distribution. Much work remains to be done in this direction.

In Section 4, a number of criticisms may be leveled at our empirical results. Our optimal derivative estimates were computed using Jackwerth's (2000) estimates of the pre- and postcrash actual and risk neutral return distributions for the S&P 500 index. Jackwerth obtained his estimates for the actual return distribution by applying a simple kernel smoother to the previous four years of returns. His estimate is thus perhaps best regarded as an estimate of the unconditional actual distribution, whereas the more relevant object may be the actual distribution conditional on currently available information. In particular, an estimate that takes advantage of the well-known predictability of volatility may be preferable to Jackwerth's simple kernel estimate. We note, however, that Rosenberg and Engle (2002) obtained postcrash pricing kernel estimates similar to those of Jackwerth using an asymmetric GARCH specification for the actual return distribution; in particular, they found that the pricing kernel was increasing around the middle of the distribution. The issue of conditional versus unconditional estimation may therefore not be of critical importance. Nevertheless, it would be desirable to use more sophisticated methods to estimate the pricing kernel and return distribution. The development of such methods is an active research area: see e.g. Gagliardini, Gouriéroux and Renault (2005).

Our empirical study may also be criticized because of the stylized nature of the model it is based on. Specifically, we implicitly assume the presence of complete markets and the absence of transaction costs. If a continuum of Arrow securities written on the S&P 500 index were somehow traded in real financial markets, an investor could purchase a portfolio of those securities delivering the payoff function corresponding to our optimal measure preserving derivative. Clearly this is not actually the case. A more realistic approach would be to consider approximating the optimal derivative using a portfolio of European put and call options written at a finite collection of strike prices, as discussed briefly at the end of Section 4. Our formula for the optimal derivative could perhaps be used as a starting point for constructing a portfolio of options delivering a return distribution that first-order stochastically dominates that obtained by investing directly in the index. Transaction costs could easily be incorporated in such a setting. This seems a promising direction for future research. It would also be of interest to extend our analysis to more recent sample periods. The results in Section 4 should be taken as a first step towards more serious empirical research in this area, and largely illustrative in nature.

A Mathematical appendix

Proof of Theorem 2.1. We shall prove the three parts of Theorem 2.1 in turn.

1. To verify that $\vartheta \in \Theta_\mu$ we observe that, for any $w \in \mathbb{R}_+$,

$$\begin{aligned}\mu\{x : \vartheta(x) \leq w\} &= \mu\{x : Q_\mu(1 - F_{\mu\pi^{-1}}(\pi(x))) \leq w\} \\ &= \mu\{x : 1 - F_{\mu\pi^{-1}}(\pi(x)) \leq F_\mu(w)\} \\ &= \mu\pi^{-1}\{x : F_{\mu\pi^{-1}}(x) \geq 1 - F_\mu(w)\}.\end{aligned}$$

   Our assumption that $\mu\{x : \pi(x) = y\} = 0$ for each $y \in \mathbb{R}_+$ is equivalent to the statement that $\mu\pi^{-1}$ is atomless. Thus, the well-known property of probability integral transforms ensures that the function $x \mapsto F_{\mu\pi^{-1}}(x)$ maps the measure $\mu\pi^{-1}$ to the uniform measure on $[0, 1]$, and so

$$\mu\pi^{-1}\{x : F_{\mu\pi^{-1}}(x) \geq 1 - F_\mu(w)\} = F_\mu(w) = \mu\{x : x \leq w\}.$$

   Consequently we have $\mu\{x : \vartheta(x) \leq w\} = \mu\{x : x \leq w\}$ for all $w \in \mathbb{R}_+$, and hence $\mu = \mu\vartheta^{-1}$. ϑ is trivially Borel measurable and nonnegative, so it follows that $\vartheta \in \Theta_\mu$.

2. The measure preserving property of the functions in Θ_μ ensures that each of them has nondecreasing rearrangement equal to the identity function. The Hardy-Littlewood inequality (see Remark 2.8) thus implies that $\int \theta\pi \mathrm{d}\mu \geq \int x\pi_*(x)\mathrm{d}\mu(x)$ for all $\theta \in \Theta_\mu$, where $\pi_*(\cdot) = Q_{\mu\pi^{-1}}(1 - F_\mu(\cdot))$. Observe that, for μ-a.e. x,

$$\begin{aligned}\pi_*(\vartheta(x)) &= Q_{\mu\pi^{-1}}(1 - F_\mu(\vartheta(x))) \\ &= Q_{\mu\pi^{-1}}(1 - F_\mu(Q_\mu(1 - F_{\mu\pi^{-1}}(\pi(x))))) \\ &= Q_{\mu\pi^{-1}}(F_{\mu\pi^{-1}}(\pi(x))) \\ &= \pi(x),\end{aligned}$$

   establishing the claim made in Remark 2.9 regarding Ryff's decomposition. It follows that

$$\int \vartheta\pi \mathrm{d}\mu = \int \vartheta(x)\pi_*(\vartheta(x))\mathrm{d}\mu(x) = \int x\pi_*(x)\mathrm{d}\mu\vartheta^{-1}(x),$$

   and so the measure preserving property of ϑ yields

$$\int \vartheta\pi \mathrm{d}\mu = \int x\pi_*(x)\mathrm{d}\mu(x).$$

This shows that ϑ achieves the Hardy-Littlewood lower bound for $\int \theta\pi \mathrm{d}\mu$ over Θ_μ.

3. If $(x, y) \in \mathbb{R}^2$ is such that $\pi(x) \leq \pi(y)$ then the monotonicity of $F_{\mu\pi^{-1}}$ and Q_μ imply that $\vartheta(x) \geq \vartheta(y)$. Similarly, if $(x, y) \in \mathbb{R}^2$ is such that $\pi(x) \geq \pi(y)$ then we must have $\vartheta(x) \leq \vartheta(y)$.

□

Proof of Theorem 2.2. Suppose first that $\vartheta(x) = x$ for μ-a.e. x. We then have

$$\mu \otimes \mu\{(x, y) : (x - y)(\vartheta(x) - \vartheta(y)) > 0\} = 1,$$

and so Part 3 of Theorem 2.1 implies that

$$\mu \otimes \mu\{(x, y) : (x - y)(\pi(x) - \pi(y)) \leq 0\} = 1.$$

Suppose next that $\mu \otimes \mu\{(x, y) : (x - y)(\pi(x) - \pi(y)) \leq 0\} = 1$. In view of the nonflatness condition on π, it must then be the case that $\mu\{y : \pi(y) \leq \pi(x)\} = \mu\{y : y > x\}$ for μ-a.e. x by Tonelli's theorem. Consequently, the atomless property of μ ensures that

$$F_{\mu\pi^{-1}}(\pi(x)) = \mu\{y : \pi(y) \leq \pi(x)\} = \mu\{y : y > x\} = 1 - F_\mu(x)$$

for μ-a.e. x, and so $\vartheta(x) = Q_\mu(F_\mu(x)) = x$ for μ-a.e. x. □

Proof of Theorem 3.1. Under the hypotheses of the theorem we may choose a set $B \in \mathcal{B}(\mathbb{R}^2_+)$ with $\mu \otimes \mu B = 1$ such that

$$P\{\omega : \hat{\pi}_n(\omega)(x) \to \pi(x) \text{ and } \hat{\pi}_n(\omega)(y) \to \pi(y)\} = 1$$

and $\pi(x) \neq \pi(y)$ for all $(x, y) \in B$. The $\mathcal{A} \otimes \mathcal{B}(\mathbb{R}_+)$-measurability of the maps $(\omega, x) \mapsto \hat{\pi}_n(\omega)(x)$ implies that the maps $A_n : \Omega \times \mathbb{R}^2_+ \to \{0, 1\}$ given by

$$A_n(\omega, (x, y)) = |1(\hat{\pi}_n(\omega)(x) \leq \hat{\pi}_n(\omega)(y)) - 1(\pi(x) \leq \pi(y))|$$

are $\mathcal{A} \otimes \mathcal{B}(\mathbb{R}^2_+)$-measurable. It is easy to see that $P\{\omega : A_n(\omega, (x, y)) \to 0\} = 1$ for all $(x, y) \in B$. An application of the dominated convergence theorem therefore yields $\int A_n(\omega, (x, y))\mathrm{d}P(\omega) \to 0$ for each $(x, y) \in B$. Since $\mu \otimes \mu B = 1$, a second application of the dominance convergence theorem yields $\iint A_n(\omega, (x, y))\mathrm{d}P(\omega)\mathrm{d}\mu \otimes \mu(x, y) \to 0$, and so from Tonelli's theorem we deduce that $\iint A_n(\omega, (x, y))\mathrm{d}\mu \otimes \mu(x, y)\mathrm{d}P(\omega) \to 0$. Noting that $A_n(\omega, \cdot)$ is the indicator function of the set $L_{\hat{\pi}_n(\omega)} \Delta L_\pi$, we may write $d_2(\hat{\pi}_n(\omega), \pi) =$

$\int A_n(\omega,(x,y))\mathrm{d}\mu\otimes\mu(x,y)$. Thus we have shown that $\int d_2(\hat{\pi}_n(\omega),\pi)\mathrm{d}P(\omega)\to 0$, from which it follows via Markov's inequality that $P\{\omega : d_2(\hat{\pi}_n(\omega),\pi)>\varepsilon\}\to 0$ for all $\varepsilon>0$. This proves that $\hat{\pi}_n \rightsquigarrow \pi$ under d_2, as claimed. □

Proof of Theorem 3.2. We shall prove the two parts of Theorem 3.2 in turn. The arguments used to prove the first part are similar to those used to prove Theorems 2.2 and 2.4 of Beare (2009).

1. It suffices for us to show that $F_{\mu\vartheta_n^{-1}}(x)\to F_\mu(x)$ as $n\to\infty$, for each $x\in\mathbb{R}_+$. Begin by observing that

$$\begin{aligned}F_{\mu\vartheta_n^{-1}}(x)-F_\mu(x) &= F_{\mu\vartheta_n^{-1}}(x)-F_{\mu\vartheta^{-1}}(x)\\ &= \mu\{y:\vartheta_n(y)\le x<\vartheta(y)\}-\mu\{y:\vartheta(y)\le x<\vartheta_n(y)\}\end{aligned}$$

for each $x\in\mathbb{R}_+$. Since μ is atomless, Tonelli's theorem implies that

$$\begin{aligned}\int\mu\{y:\vartheta_n(y)\le x<\vartheta(y)\}\mathrm{d}\mu(x) &= \int\mu\{x:\vartheta_n(y)\le x<\vartheta(y)\}\mathrm{d}\mu(y)\\ &= \int\max\{F_\mu(\vartheta(y))-F_\mu(\vartheta_n(y)),0\}\mathrm{d}\mu(y),\end{aligned}$$

and similarly

$$\int\mu\{y:\vartheta(y)\le x<\vartheta_n(y)\}\mathrm{d}\mu(x) = \int\max\{F_\mu(\vartheta_n(y))-F_\mu(\vartheta(y)),0\}\mathrm{d}\mu(y).$$

Consequently, we have

$$\int|F_{\mu\vartheta_n^{-1}}-F_\mu|\mathrm{d}\mu\le\int|F_\mu(\vartheta_n(y))-F_\mu(\vartheta(y))|\mathrm{d}\mu(y)=d_3(\vartheta_n,\vartheta),$$

which tends to zero as $n\to\infty$. We now prove by contradiction that $\int|F_{\mu\vartheta_n^{-1}}-F_\mu|\mathrm{d}\mu\to 0$ implies $F_{\mu\vartheta_n^{-1}}\to F_\mu$ pointwise. Suppose $F_{\mu\vartheta_n^{-1}}(c_0)\nrightarrow F_\mu(c_0)$ for some $c_0\in\mathbb{R}_+$. Then there must be a sequence of natural numbers $n_1,n_2,\ldots$ increasing to infinity, and a real number $\varepsilon>0$ (or $\varepsilon<0$), such that $F_{\mu\vartheta_{n_k}^{-1}}(c_0)\ge F_\mu(c_0)+\varepsilon$ (resp. $F_{\mu\vartheta_{n_k}^{-1}}(c_0)\le F_\mu(c_0)+\varepsilon$) for all k. Suppose $\varepsilon>0$; we may use an analogous argument when $\varepsilon<0$. Since F_μ is continuous, we may choose $c_1>c_0$ such that $F_\mu(c_1)=F_\mu(c_0)+\varepsilon/2$. Monotonicity of $F_{\mu\vartheta_{n_k}^{-1}}$ and F_μ then ensures that $F_{\mu\vartheta_{n_k}^{-1}}(x)\ge F_\mu(x)+\varepsilon/2$ for all $x\in[c_0,c_1]$ and all k. Consequently, we have

$$\int|F_{\mu\vartheta_{n_k}^{-1}}-F_\mu|\mathrm{d}\mu\ge\frac{\varepsilon}{2}\int_{c_0}^{c_1}\mathrm{d}\mu=\frac{\varepsilon}{2}(F_\mu(c_1)-F_\mu(c_0))=\frac{\varepsilon^2}{4}>0$$

for all k, implying that $\int |F_{\mu\vartheta_n^{-1}} - F_\mu| \mathrm{d}\mu \nrightarrow 0$.

2. Markov's inequality implies that, for any $\varepsilon > 0$,

$$\mu\{x : |F_\mu(\vartheta_n(x)) - F_\mu(\vartheta(x))| > \varepsilon\} \leq \varepsilon^{-1} d_3(\vartheta_n, \vartheta) \to 0$$

as $n \to \infty$. In this sense, $F_\mu \circ \vartheta_n$ converges in μ-measure to $F_\mu \circ \vartheta$. The measure preserving property of ϑ implies that

$$\mu\{x : Q_\mu \text{ is continuous at } F_\mu(\vartheta(x))\} = \mu\{x : Q_\mu \text{ is continuous at } F_\mu(x)\}.$$

Since μ is atomless, this last quantity is simply the Lebesgue measure of the set of points in $[0, 1]$ at which Q_μ is continuous, which must be one since Q_μ is nondecreasing and therefore has at most countably many discontinuities. Thus, for x in a set of μ-measure one, we have Q_μ continuous at $F_\mu(\vartheta(x))$. The continuous mapping theorem therefore implies that $Q_\mu \circ F_\mu \circ \vartheta_n$ converges in μ-measure to $Q_\mu \circ F_\mu \circ \vartheta$. Since $Q_\mu(F_\mu(x)) = x$ for μ-a.e. x, the measure preserving property of ϑ implies that $Q_\mu(F_\mu(\vartheta(x))) = \vartheta(x)$ for μ-a.e. x. Consequently, $Q_\mu \circ F_\mu \circ \vartheta_n$ converges in μ-measure to ϑ. That is, for any $\varepsilon > 0$,

$$\mu\{x : |Q_\mu(F_\mu(\vartheta_n(x))) - \vartheta(x)| > \varepsilon\} \to 0$$

as $n \to \infty$. We wish to show that ϑ_n converges in μ-measure to ϑ. Observe that

$$\begin{aligned} \mu\{x : |\vartheta_n(x) - \vartheta(x)| > \varepsilon\} \leq\ & \mu\{x : |Q_\mu(F_\mu(\vartheta_n(x))) - \vartheta(x)| > \varepsilon\} \\ & + \mu\{x : F_\mu(\vartheta_n(x)) = F_\mu(y) \text{ for some } y \neq \vartheta_n(x)\}. \end{aligned}$$

We have shown that the first term on the right-hand side of the above inequality converges to zero as $n \to \infty$. We wish to show that the second term also converges to zero. This term satisfies

$$\begin{aligned} & \mu\{x : F_\mu(\vartheta_n(x)) = F_\mu(y) \text{ for some } y \neq \vartheta_n(x)\} \\ = \ & \mu\vartheta_n^{-1}\{x : F_\mu(x) = F_\mu(y) \text{ for some } y \neq x\}. \end{aligned}$$

Since μ is atomless, the set $\{x : F_\mu(x) = F_\mu(y) \text{ for some } y \neq x\}$ is the union of at most countably many closed intervals, and therefore has boundary μ-measure zero.

Part 1 of this Theorem states that $\mu\vartheta_n^{-1}$ converges weakly to μ, so we must have

$$\mu\vartheta_n^{-1}\{x : F_\mu(x) = F_\mu(y) \text{ for some } y \neq x\} \to \mu\{x : F_\mu(x) = F_\mu(y) \text{ for some } y \neq x\}$$

as $n \to \infty$. Since μ is atomless, $\mu\{x : F_\mu(x) = F_\mu(y) \text{ for some } y \neq x\} = 0$. This completes the proof.

□

Proof of Theorem 3.3. Recalling Remark 3.5, our desired result follows from Theorem 1.3.6 of van der Vaart and Wellner (1996), a version of the continuous mapping theorem, if we can show that h is continuous at (μ, π) with respect to d_{12} on its domain and d_3 on its range. Consider a sequence of elements $(\mu_n, \pi_n) \in \mathcal{M} \times \mathcal{F}$, $n \in \mathbb{N}$, such that $d_{12}((\mu_n, \pi_n), (\mu, \pi)) \to 0$ as $n \to \infty$. Clearly $d_1(\mu_n, \mu) \to 0$ and $d_2(\pi_n, \pi) \to 0$ as $n \to \infty$. Observe that

$$\begin{aligned} & d_3(h(\mu_n, \pi_n), h(\mu, \pi)) \\ = & \int_{\mathbb{R}_+} |F_\mu(Q_{\mu_n}(1 - F_{\mu_n\pi_n^{-1}}(\pi_n(x)))) - F_\mu(Q_\mu(1 - F_{\mu\pi^{-1}}(\pi(x))))| \mathrm{d}\mu(x) \\ \leq & \int_{\mathbb{R}_+} |F_{\mu_n\pi_n^{-1}}(\pi_n(x)) - F_{\mu\pi^{-1}}(\pi(x))| \mathrm{d}\mu(x) + \sup_{u\in(0,1)} |F_\mu(Q_{\mu_n}(u)) - u|. \end{aligned} \tag{A.1}$$

We need to show that the two terms on the right-hand side of (A.1) converge to zero as $n \to \infty$. The second term satisfies

$$\sup_{u\in(0,1)} |F_\mu(Q_{\mu_n}(u)) - u| \leq \sup_{x\in\mathbb{R}_+} |F_{\mu_n}(x) - F_\mu(x)| + \sup_{u\in(0,1)} |F_{\mu_n}(Q_{\mu_n}(u)) - u|. \tag{A.2}$$

Since $d_1(\mu_n, \mu) \to 0$, the first term on the right-hand side of (A.2) converges to zero. The second term on the right-hand side of (A.2) is equal to the size of the largest discontinuity in F_{μ_n}. This quantity must be converging to zero since $d_1(\mu_n, \mu) \to 0$ implies that F_{μ_n} converges uniformly to F_μ, which is continuous since μ is atomless.

It remains to show that the first term on the right-hand side of (A.1) converges to zero. We have

$$\begin{aligned} & \int_{\mathbb{R}_+} |F_{\mu_n\pi_n^{-1}}(\pi_n(x)) - F_{\mu\pi^{-1}}(\pi(x))| \mathrm{d}\mu(x) \\ \leq & \int_{\mathbb{R}_+} |F_{\mu_n\pi_n^{-1}}(\pi_n(x)) - F_{\mu\pi_n^{-1}}(\pi_n(x))| \mathrm{d}\mu(x) \\ & + \int_{\mathbb{R}_+} |F_{\mu\pi_n^{-1}}(\pi_n(x)) - F_{\mu\pi^{-1}}(\pi(x))| \mathrm{d}\mu(x). \end{aligned} \tag{A.3}$$

For each $x \in \mathbb{R}_+$ the integrand in the first term on the right-hand side of (A.3) satisfies

$$\begin{aligned}|F_{\mu_n\pi_n^{-1}}(\pi_n(x)) - F_{\mu\pi_n^{-1}}(\pi_n(x))| &= |\mu_n\{y : \pi_n(y) \leq \pi_n(x)\} - \mu\{y : \pi_n(y) \leq \pi_n(x)\}| \\ &\leq \sup_{B\in\mathcal{B}(\bar{\mathbb{R}}_+)} |\mu_n(B) - \mu(B)|,\end{aligned}$$

which converges to zero since $d_1(\mu_n, \mu) \to 0$. Thus the first term on the right-hand side of (A.3) converges to zero by dominated convergence. We now need only show that the second term on the right-hand side of (A.3) converges to zero. Using Tonelli's theorem, we have

$$\begin{aligned}&\int_{\mathbb{R}_+} |F_{\mu\pi_n^{-1}}(\pi_n(x)) - F_{\mu\pi^{-1}}(\pi(x))|\mathrm{d}\mu(x) \\ =& \int_{\mathbb{R}_+} |\mu\{x : \pi_n(x) \leq \pi_n(y)\} - \mu\{x : \pi(x) \leq \pi(y)\}|\mathrm{d}\mu(y) \\ \leq& \int_{\mathbb{R}_+} \mu(\{x : \pi_n(x) \leq \pi_n(y)\} \triangle \{x : \pi(x) \leq \pi(y)\})\mathrm{d}\mu(y) \\ =& \ \mu \otimes \mu(\{(x, y) : \pi_n(x) \leq \pi_n(y)\} \triangle \{(x, y) : \pi(x) \leq \pi(y)\}),\end{aligned}$$

which is equal to $d_2(\pi_n, \pi)$ and therefore converges to zero. We have now shown that $d_3(h(\mu_n, \pi_n), h(\mu, \pi)) \to 0$, which proves that h is continuous at (μ, π), as claimed. □

References

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

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

Becker, G. S. (1973). A theory of marriage: part I. *Journal of Political Economy* **81** 813–846.

Beare, B. K. (2009). Distributional replication. UC San Diego Economics Working Paper No. 2009-05.

Brown, D. P. and Jackwerth, J. C. (2004). The pricing kernel puzzle: reconciling index option data and economic theory. Working paper, Wisconsin School of Business.

Carlier, G. and Dana, R. -A. (2005). Rearrangement inequalities in non-convex insurance models. *Journal of Mathematical Economics* **41** 483–503.

CARLIER, G., DANA, R. -A. AND GALICHON, A. (2009). Pareto efficiency for the concave order and multivariate comonotonicity. arXiv:0912.0509v1.

CHABI-YO, F., GARCIA, R. AND RENAULT, E. (2008). State dependence can explain the risk aversion puzzle. *Review of Financial Studies* **21** 973–1011.

CHONG, K. M. AND RICE, N. M. (1971). Equimeasurable rearrangements of functions. *Queen's Papers in Pure and Applied Mathematics* **28**.

CONSTANTINIDES, G. M., JACKWERTH, J. C. AND PERRAKIS, S. (2009). Mispricing of S&P 500 index options. *Review of Financial Studies* **22** 1247–1277.

DETLEFSEN, K., HÄRDLE, W. AND MORO, R. (2007). Empirical pricing kernels and investor preferences. SFB 649 Discussion Paper 2007-017, Humboldt-Universität zu Berlin.

DOORNIK, J. A. (2007). *Object-Oriented Matrix Programming Using Ox,* 3rd ed. Timberlake Consultants Press, London.

DYBVIG, P. H. (1988a). Distributional analysis of portfolio choice. *Journal of Business* **63** 369–393.

DYBVIG, P. H. (1988b). Inefficient dynamic portfolio strategies or how to throw away a million dollars in the stock market. *Review of Financial Studies* **1** 67–88.

GAGLIARDINI, P., GOURIÉROUX, C. S. AND RENAULT, E. (2005). Efficient derivative pricing by extended method of moments. Working paper, University of St. Gallen Department of Economics.

GOLUBEV, Y., HÄRDLE, W. AND TIMONFEEV, R. (2008). Testing monotonicity of pricing kernels. SFB 649 Discussion Paper 2008-001, Humboldt-Universität zu Berlin.

HÄRDLE, W., KRÄTSCHMER, V. AND MORO, R. (2009). A microeconomic explanation of the EPK paradox. SFB 649 Discussion Paper 2009-010, Humboldt-Universität zu Berlin.

HÄRDLE, W., OKHRIN, Y. AND WANG, W. (2010). Uniform confidence bands for pricing kernels. SFB 649 Discussion Paper 2010-003, Humboldt-Universität zu Berlin.

HARDY, G. H., LITTLEWOOD, J. E. AND PÒLYA, G. (1934). *Inequalities.* Cambridge University Press, Cambridge.

JACKWERTH, J. C. (2000). Recovering risk aversion from option prices and realized returns. *Review of Financial Studies* **13** 433–451.

Jackwerth, J. C. (2004). *Option-Implied Risk-Neutral Distributions and Risk Aversion.* Research Foundation of AIMR, Charlotteville.

Jackwerth, J. C. and Rubinstein, M. (1996). Recovering probability distributions from option prices. *Journal of Finance* **51** 1611–1631.

Machina, M. J. (1982). A stronger characterization of declining risk aversion. *Econometrica* **50** 1069–1079.

Renou, L. and Carlier, G. (2003). Existence and monotonicity of optimal debt contracts in costly state verification models. *Economics Bulletin* **7** 1–9.

Rosenberg, J. V. and Engle, R. F. (2002). Empirical pricing kernels. *Journal of Financial Economics* **64** 341–372.

Ryff, J. V. (1970). Measure preserving transformations and rearrangements. *Journal of Mathematical Analysis and Applications* **31** 449–458.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New York.

Villani, C. (2003). *Topics in Optimal Transportation.* Graduate Studies in Mathematics **58**. American Mathematical Society, Providence.

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