Review of Finance, 2017, 269–298
doi: 10.1093/rof/rfv062
Advance Access Publication Date: 11 February 2016

OXFORD

Reference-Dependent Preferences and the Empirical Pricing Kernel Puzzle*

Maria Grith[1], Wolfgang K. Härdle[1], and Volker Krätschmer[2]

[1]Humboldt-Universität zu Berlin and [2]Universität Duisburg-Essen

Abstract

Supported by several recent investigations, the empirical pricing kernel (PK) puzzle might be considered as a stylized fact. Based on an economic model with reference-dependent preferences for the financial investors, we emphasize a microeconomic view that explains the puzzle via state-dependent aggregate preferences. We also investigate how the shape of the PK estimated from option and stock market index returns changes in relation to the volatility risk premium.

JEL classification*:* D01, D53, C02, G13

1. Introduction

Empirical investigations of the option prices written on the underlying stock market index and realized index returns reveal puzzling features. This article addresses the empirical pricing kernel (EPK) puzzle, as it was first coined by Jackwerth (2000). To illustrate the concept, we consider a simple one-period model of a financial market.

Let $[0, T]$ be the finite investment horizon, where $t = 0$ denotes the present time and $t = T \in (0, \infty)$ the time of maturity. It is assumed that a riskless bond and a risky asset are traded in the financial market as basic securities. The price process of the riskless bond $(B_t)_{t\in[0,T]}$ is defined by $B_t = \exp(-\int_0^t r_s \ \mathrm{d}s)$ via a deterministic Riemannian-integrable interest process $(r_t)_{t\in[0,T]}$. The price process of the risky asset $(S_t)_{t\in[0,T]}$ is considered to be a non-negative semimartingale with continuously distributed marginals S_t. The return of the risky asset at maturity is defined as $R_T = S_T/S_0$.

We further assume that the financial market is arbitrage free in the sense that there exists an equivalent martingale measure, which can be identified with a risk neutral pricing rule. Under such a measure, discounted prices have the martingale property. Hence, there is

* We thank two anonymous referees, Thorsten Hens, Harald Uhlig, and the participants at seminars in Berlin, Konstanz, New York, Princeton, and Warwick for helpful comments. This research was supported by Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk."

an unknown Radon–Nikodym density π of a martingale measure such that the price at $t = 0$ of any random payoffs $\psi(R_T)$ is characterized by

$$\mathsf{E}[B_T\psi(R_T)\pi]. \tag{1}$$

By factorization with some Borel-measurable K, that we call pricing kernel (PK) (w.r.t. π), with $\mathsf{E}[\pi|R_T] = K(R_T)$ we obtain

$$\int_0^\infty B_T\psi(x)K(x)p(x)\mathrm{d}x, \tag{2}$$

where p denotes the probability density function of R_T. The PK in Equation (2) is summarized as a univariate function and can be used for the valuation of payoffs that depend only on R_T but not on other risk factors. In what follows, we will refer to the risky asset as the stock market index.

Equilibrium asset pricing models identify the discounted random variable π with the stochastic discount factor (SDF) or the intertemporal marginal rate of substitution. In general, the SDF depends on many variables. Many studies focus on aggregate consumption and customarily assume a representative agent whose preferences have an expected utility representation over consumption sets. Under standard assumptions (true knowledge of the objective probability distribution of the risk factors, positive risk aversion, non-satiation), the SDF is proportional with the von Neumann–Morgenstern marginal utility index of the representative agent and therefore nonincreasing in the aggregate consumption.

By definition, K is a projection of any admissible SDF onto the space spanned by R_T. K is not necessarily equal to π, nor does it automatically inherit its properties. However, for reasons outlined below one would expect it to be nonincreasing in R_T. Rubinstein (1976), Brown and Gibbons (1985), Breeden, Gibbons, and Litzenberger (1989), among others, provide both theoretical and practical justifications for the use of the market portfolio to proxy aggregate consumption. With classical preferences, representable by an increasing and concave utility function, the implications of such an approximation for Equation (2) are that K should be nonincreasing in the stock market index return, if the index is used to represent consumption in the economy. Even if this equivalence is removed, we can still interpret K as rationalizing the investment behavior on a segment of the market composed of the stock index and index derivatives. Hence, the aggregate investor should be risk averse over all values of the index. Departing from the spirit underlying utility-based asset pricing models, Dybvig (1988) shows that K should be nonincreasing in order to avoid statistical arbitrage relative to a direct market investment, see Constantinides and Jackwerth (2009) and Beare (2011) for discussions on stochastic dominance between the bond, stock index, and options.

In practice however, PKs estimated by empirical techniques introduced by Rubinstein (1976) and Jackwerth and Rubinstein (1996) for option prices, often appear to be locally increasing. This is what we call the EPK puzzle.

The scatter plot in Figure 1 displays noisy "realizations" of an average PK at 1-month investment horizon, over 10 years, between January 2002 and December 2011, for the DAX 30 stock market index. A detailed account of the estimation methodology is provided in Section 4. The figure summarizes two features. There is an increasing region located right before the current value of the stock index and the PK is hump-shaped around one. At the same time, there is an increasing region for states identified with high future returns that lead to a U-shaped PK in the absence of the hump. In either case, the local nonmonotonicity

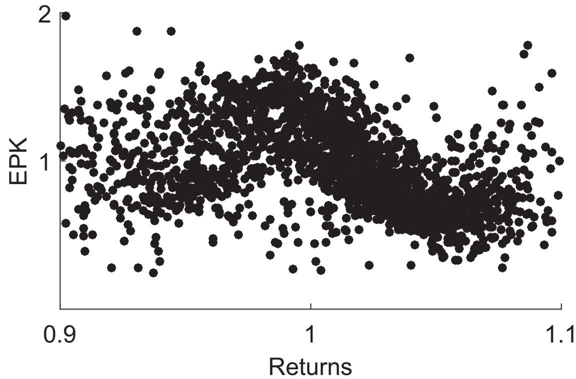

Figure 1. PK estimates for 1-month investment horizon, January 2002–December 2011.

of the empirical kernels contrasts with the standard theory of expected concave utility maximization of the representative agent under true beliefs.

We contribute to the EPK puzzle literature in two ways. First, we explain the hump-shape feature of the EPK from a preference point of view, in a generalized expected utility framework. We modify the classical preferences of the investors by allowing them to "switch" utilities depending on their individual reference or benchmark. In light of the empirical evidence, the references are assumed to lie in the space of the stock index. We focus on the technical conditions that ensure preference aggregation, provide a link between individual and market preferences, and derive the PK in equilibrium. Then, we provide the conditions that lead to our main result: the locally increasing PK in the stock index return. As a second contribution, we show how our model can help explain the two empirical features of the EPK by analyzing its shape in relation to the volatility risk premium (VRP), used as a proxy for the macroeconomic uncertainty. Our results suggest that the hump is more pronounced during periods with low uncertainty, during which investors display a more pronounced benchmarking behavior relative to the stock market index.

The remainder of the article proceeds as follows. Section 2 surveys the EPK puzzle literature and the existing theoretical explanations that attempt to rationalize it. Section 3 introduces the reference-dependent preferences model and derives the technical conditions for the nonmonotonicity of the PK. Investigations into the shape behavior of the PK are developed through a simulation study. Further, some insights into the implications of such preferences on the asset allocation are considered. Section 4 inspects the ability of the model to describe the empirically observed features of the PK. In particular, it explains the observed shape variability of the EPK in relation to the VRP and its implications for the derivative prices. Section 5 concludes. Several mathematical results and proofs of technical nature are available in Appendices A–C.

2. An Overview of the EPK Puzzle Literature

Starting with Aït-Sahalia and Lo (2000), Jackwerth (2000), and Engle and Rosenberg (2002), different econometric models have been employed to estimate PKs from the historical returns and option prices of the underlying asset. Many of the estimates have nonmonotonic shape regardless of the models and the data sets used. Typically, we find either a hump-shaped PK or a U-shaped PK.

Statistical significance of the local nonmonotonicity of the EPKs was initially investigated by Golubev, Härdle, and Timofeev (2014), who test the local concavity of the utility function implied by the DAX 30 index options and the historical index returns. In a subsequent paper, Härdle, Okhrin, and Wang (2015) build uniform confidence bands for the PK estimates and test if a decreasing PK implied by the Black–Scholes model can be fitted within the confidence corridor. Beare and Schmidt (2014) test the concavity of the ordinal dominance curve associated with the risk neutral and physical distributions associated with the S&P 500 index. In their paper, the null hypothesis of nonincreasing EPK was rejected, particularly for hump-shaped EPKs.

The (non)monotonicity of the PK has been characterized within one- or multiperiod investment models. For pure financial economies, with no other exogenous income, finite investment horizon and trading at any instant $t \in [0, T]$, the intertemporal problem translates in maximizing the utility of final consumption, which equals the final wealth. One-period models can be subsumed to the multiperiod models by restricting the initial conditions of the subsequent period by the outcomes of the prior one. The dynamic models address the time-varying patterns of the intertemporal EPKs by making use of dependence on state variables. The reason for considering finite-period investment decisions to model PKs is that European options have fixed maturities. The interpretation of the PK estimates differs in relation to the data used by the econometrician, e.g., average, conditional, or unconditional estimates.

We first review studies that investigate mechanisms through which a locally increasing region in the PK, both hump and U-shaped, can occur in equilibrium models with single investment horizon. Hens and Reichlin (2013) conduct a systematic analysis of the EPK puzzle by relaxing each of the assumptions embedded in the standard expected utility models: complete markets, risk-averse investors, and correct beliefs. They calibrate a hump-shaped PK from Jackwerth (2000) and investigate feasible explanations. Their findings suggest that incomplete markets alone can explain the puzzle. The authors rule out local risk-seeking preferences of the representative agent as they do not work in a continuous state framework. According to them, the misestimation of the objective probability distribution of stock index in isolation requires an unrealistic degree of pessimism. This finding is in line with Ziegler (2007). Hens and Reichlin (2013) also state that biased beliefs together with distortions, i.e., probability weighting functions used in rank-dependent expected utility, may reproduce the hump.

Closely related to this last point, Polkovnichenko and Zhao (2013) estimate nonparametrically the probability weighting functions implied by the PK estimates. For most of the years in their study the estimates are inverse S-shaped, consistent with the experimental findings that people tend to overweight low-probability events while they underweight the likelihood of high-probable ones. The inverse S-shaped probability weighting functions correspond to U-shaped PKs. The estimates become S-shaped during the years 2004–7 but the authors do not make further investigations about the differences in these treatments. Several other studies report results that suggest that the PK is hump-shaped for most months between 2004 and 2007. These hold for both the American S&P 500 index (Barone-Adesi, Mancini, and Shefrin, 2013; Beare and Schmidt, 2014) and the German DAX 30 index (Giacomini and Härdle, 2008; Grith, Härdle, and Park, 2013). According to Hens and Reichlin (2013), a plausible explanation for the shape of probability weighting function in the presence of hump-shaped PKs is the heterogeneity in beliefs.

Heterogeneity in beliefs about the objective distribution of future states appears in several other papers as a possible interpretation for the EPK puzzle. For instance, Bakshi and

Madan (2007) and Bakshi, Madan, and Panayotov (2010) consider an equilibrium model with short and long equity investors that is able to explain the U-shaped PK; in particular, locally increasing PKs occur when some investors are shorting equities. Misestimation of the real probability distribution, as reflected by optimism and pessimism are consistent with biases in the first moment of the objective distribution; in order to explain the empirical findings, Shefrin (2008) suggests that higher order biases should be considered and emphasizes the bias in the second moments. The time-varying shape of the EPK is explained in Barone-Adesi, Mancini, and Shefrin (2013) through a combination of optimism/pessimism and over-/underconfidence defined as differences in the first and second moments of the physical and risk neutral distributions. Accordingly, the authors find that the hump-shaped PK stems from a mix of optimistic overconfident and pessimistic underconfident agents.

Working from a preference perspective, Shefrin (2008) rationalizes the EPK puzzle in a model with mixed expected utility maximizers and agents endowed with SP/A preferences, see Lopes (1987) and Lopes and Oden (1999).

Branger, Schlag, and Zaharia (2011) describe an equilibrium model for Heston-type approaches in which a U-shaped PK can be reproduced if there exists a premium for pure variance risk. The idea of such a PK is similar to Christoffersen, Heston, and Jacobs (2013), who propose an augmented Heston and Nandi (2000) model that nests a U-shaped PK in the presence of a positive variance premium parameter.

Another body of literature refers to state-dependence in preferences, economic fundamentals, or beliefs in multiperiod models. In these works, the conjecture is that if properly conditioned, the PK becomes decreasing. State dependence has been traditionally used to explain the asset pricing puzzles in equilibrium models mainly based on two utility classes: habit formation, see Constantinides (1990), Campbell and Cochrane (1999), or recursive utilities, see Epstein and Zin (2001). These frameworks have been adapted to typically presume a Markov switching process for the evolution of states and derive asset-related characteristics in a consumption-based model. Garcia, Luger, and Renault (2003) investigate recursive utility functions with state-dependence in the fundamentals. Melino and Yang (2003) disentangle the roles played by state-dependent intertemporal substitution and time preference in explaining the equity premium puzzle in a model with state-dependent recursive preferences. Veronesi (2004) extends the state-dependent utility by assuming that the agents possess a probability distribution over their state and introduces the concept of "belief-dependent preferences" in a model with habit formation. Chabi-Yo, Garcia, and Renault (2008) generalize the setup of Melino and Yang (2003) by extending the set of state variables to rationalize the EPK puzzle, in its alternative formulation as a risk aversion puzzle.

More recent research indicates volatility as a state variable that helps explain the observed nonmonotonicities in the PK. Song and Xiu (2016) find that, consistent with the economic theory, the PK decreases in the market index return if it is conditioned by the market volatility. As such, under the alleged positive relationship between returns and volatility, the unconditional estimates of the PK may appear U-shaped. The evidence in Chabi-Yo (2012) also suggests the volatility as a pricing factor in the PK.

Given these prior works, it can be seen that the shape of the EPK is surrounded at times by seemingly inconsistent results. A non-exhaustive list of reasons for the various shapes found in the PK is: The estimates use only options with strikes around the current stock price as opposed to the entire range of options available (e.g., Chabi-Yo, 2012; Christoffersen, Heston, and Jacobs, 2013), estimates consider selective periods of time (e.g., Song and Xiu, 2016), the window lengths used for the estimation of the physical densities may not be adequately chosen or estimates do not account for time-varying volatility.

In the following, we introduce a one-period model in which reference-dependent individual heterogeneous preferences are able to reproduce the hump-shaped PK. This model nests the decreasing PK as a special case. The mechanism we suggest can take place together with other explanations that support a U-shaped PK, for instance, structural biases, e.g., Polkovnichenko and Zhao (2013) or different investment strategies, e.g., Bakshi, Madan, and Panayotov (2010).

3. A Microeconomic View on the EPK Puzzle

In this section we formally derive the preference-based model that links investors' risk preferences to the pricing rule of the financial market described in Section 1. We consider preferences defined with respect to a benchmark or a reference.

The inclusion of a benchmark in the utility function is achievable in various ways. Typical choices are to subtract the reference from the consumption or wealth, or to divide them by either a fixed or stochastic reference. These techniques are used for instance to model habit formation. In effect, the implied risk aversion often changes. More complicated formulations that include further components of wealth are also possible, see e.g., Cuoco and Kaniel (2011). Similar behavioral implications are obtained by multiplying individual utility functions with a constant or monotonic function in the aggregate wealth, conditional on a benchmark. In finance, such specifications can be linked to the incentive schedules that impact investors' risk taking: percentage fees, option-like contracts, bonus schemes proportional to the excess return over the benchmark, see Dybvig, Farnsworth, and Carpenter (2010). For instance, Ross (2004) shows that option-like incentive schedules change investors' risk behavior. In a similar fashion, Lewellen (2006) illustrates how options can increase the preference for volatility. This may lead to a kink—local convexity—in the utility function. Both authors use the term magnification of the utility function to describe the effect of risk-inducing payoff schedules.

Motivated by these facts, we describe the individual preferences using two regimes of utilities that "become active" if the stock index is above or below a reference point. The consideration for linking the performance of stock index to the individual preferences follows from the fact that the stock index is a natural indicator for the average performance of the investors in the market. In this regard, investors' decisions depend not only on the individual wealth but also on the wealth of their peers. Throughout the article, we work with the relative wealth. This formulation is safe in the present setup with a single period but in a multiperiod model one must proceed with care due to the endogeneity of the stock price.

3.1 Reference-Dependent Preferences

Let us assume that there are m investors with exogenous initial wealth $w_{10}, \dots, w_{m0}$ and stochastic financial wealth in form of nonnegative random variables $e_1(R_T), \dots, e_m(R_T)$. Without loss of generality we assume that all prices are discounted, i.e., the numeraire bond equals unity, $B_T = 1$. In addition, the bond is in zero net supply and the risky asset is in positive supply.

3.1.a. *The budget constraint*

The terminal wealth $w_i(R_T)$ fulfills the individual budget constraint

$$\int_0^\infty w_i(x)K(x)p(x)\ dx \leq w_{i0} + \int_0^\infty e_i(x)K(x)p(x)\ dx,\ \ i = 1, \dots, m. \tag{3}$$

Financial wealth $e_i(R_T)$ at $t = T$ depends on the initial holdings of securities and on the investment choice at $t = 0$. If we denote δ_i the fraction of assets invested in the risky asset, then $e_i(R_T) = \delta_i R_T + (1 - \delta_i)$ and δ_i expresses the risk exposure given initial wealth w_{i0}.

3.1.b. *Individual preferences*

The investors are assumed to have extended expected utility preferences within the terminology of Mas-Colell, Whinston, and Greene (1995). This framework allows us to describe a switching behavior for the investors' preference, when they are faced with a benchmark or a reference, through a utility function $u^i : \mathbb{R}_+^2 \to \mathbb{R}$ with

$$u^i\{r_T, w_i(r_T)\} = u_i^0\{w_i(r_T)\}\mathrm{I}\{r_T \in [0, x_i]\} + u_i^1\{w_i(r_T)\}\mathrm{I}\{r_T \in (x_i, \infty)\}, \tag{4}$$

for every realization r_T of R_T. The numerical representation for the preferences of investor i in Equation (4) is in terms of two basic utility indices $u_i^0, u_i^1 : [0, \infty) \to \mathbb{R}$ and a reference point $x_i > 0$ in the space of state variable R_T. u^i is assumed to satisfy

$$u^i\{r_T, w_i(r_T)\} \in \mathbb{R}\ \text{ for }\ r_T \geq 0, \tag{5}$$

$$u^i(r_T, \cdot) \quad \text{is strictly increasing and strictly concave for any}\ \ r_T \geq 0, \tag{6}$$

$$u^i\{\cdot, w_i(r_T)\} \quad \text{is Borel} - \text{measurable for every}\ \ w_i(r_T). \tag{7}$$

If $u^i(r_T, \cdot)$ is continuously differentiable the usual Inada conditions are assumed to hold

$$\lim_{w_i(r_T)\to 0} \frac{\mathrm{d}u^i(r_T, \cdot)}{\mathrm{d}y}\Big|_{y=w_i(r_T)} = \infty, \quad \lim_{w_i(r_T)\to \infty} \frac{\mathrm{d}u^i(r_T, \cdot)}{\mathrm{d}y}\Big|_{y=w_i(r_T)} = 0. \tag{8}$$

We impose a further condition on the asymptotic elasticity of the utilities that represents a minimal requirement to describe the optimal investment in terms of the marginal utilities of the individual investors and a PK

$$\limsup_{w_i(r_T)\to \infty} \frac{\mathrm{d}u^i(r_T, \cdot)}{\mathrm{d}y}\Big|_{y=w_i(r_T)} < 1\ \text{ for any}\ r_T \geq 0\ \text{and every}\ i \in \{1, \dots, m\}. \tag{9}$$

This condition follows from Kramkov and Schachermayer (1999); a similar characterization appears in Dana and Jeanblanc (2003), Duffie (1996), and Karatzas and Shreve (1998).

3.1.c. *Equilibrium conditions*

Each investor i chooses an optimal wealth plan $\overline{w}_i(R_T)$ such that the following properties are fulfilled:

i. individual optimization

$$\overline{w}_i(R_T) = \arg\max_{w_i(R_T)} \mathsf{E}\big[u^i\{R_T, w_i(R_T)\}\big] \tag{10}$$

$$\text{s.t.}\ \ w_i(R_T)\ \ \text{satisfies individual budget constraint (3).}$$

ii. market clearing

$$\sum_{i=1}^{m} \overline{w}_i(R_T) = \overline{w}(R_T). \tag{11}$$

The conditions (10) and (11) describe a weak version of a contingent Arrow Debreu equilibrium (Dana and Jeanblanc, 2003, section 7.1). As a byproduct $\overline{w}_1(R_T), \ldots, \overline{w}_m(R_T)$ are Pareto optimal, meaning that there are no $w_1(R_T), \ldots, w_m(R_T)$ with $\mathsf{E}\big[u^i\{R_T, w_i(R_T)\}\big] \geq \mathsf{E}\big[u^i\{R_T, \overline{w}_i(R_T)\}\big]$ for every i and such that $\mathsf{E}\big[u^i\{R_T, w_i(R_T)\}\big] > \mathsf{E}\big[u^i\{R_T, \overline{w}_i(R_T)\}\big]$ for at least one i.

3.1.d. *Market preferences*

With another formulation of the Pareto optimality in place, see Negishi (1960), we may express the aggregate state-dependent utility $u_\alpha : \mathbb{R}^2_+ \to \mathbb{R}$ in terms of properly defined weights $\alpha = (\alpha_1, \ldots, \alpha_m)$ with $\sum_{i=1}^m \alpha_i = m$ and every $\alpha_i > 0$ so that

$$u_\alpha\{r_T, \overline{w}(r_T)\} = \sum_{i=1}^{m} \alpha_i u^i\{r_T, \overline{w}_i(r_T)\}, \tag{12}$$

for every realization r_T of R_T and $\overline{w}_i(r_T)$ derived in equilibrium. For a further characterization of Equation (12), let us assume that the thresholds of investors are ordered $x_1 \leq \ldots \leq x_m$. Then, given the individual preference representation in Equation (4), one can build aggregate utility indices defined for $j = 1, \ldots, m+1$

$$u^j_\alpha\{\overline{w}(r_T)\} \stackrel{\text{def}}{=} \sum_{i=1}^{m} \alpha_i u^0_i\left\{\overline{w}^{0,j}_i(r_T)\right\}\mathrm{I}\{i \geq j\} + \sum_{i=1}^{m} \alpha_i u^1_i\left\{\overline{w}^{1,j}_i(r_T)\right\}\mathrm{I}\{i < j\}. \tag{13}$$

$\overline{w}^{l,j}_i(R_T)$, $l \in \{0, 1\}$ solves (i) the individual optimization (10) when investors have standard utilities representable by u^l_i under the modified constraint (3) that reflects endogenously the composition of preferences implied by u^j_α through the PK, and (ii) the market clearing condition $\sum_{i=1}^m \overline{w}^{0,j}_i(R_T)\mathrm{I}\{i \geq j\} + \sum_{i=1}^m \overline{w}^{1,j}_i(R_T)\mathrm{I}\{i < j\} = \overline{w}(R_T)$. In essence, the aggregate utility indices u^j_α differ through the varying proportions of utility indices of type 0 and 1 that enter the aggregation and inherit their properties, i.e., strictly increasing, strictly concave, and continuously differentiable. As a consequence, the aggregate state-dependent utility u_α can be interpreted as expressing the hegemony of different potential representative agents on ordered subintervals

$$\begin{aligned} u_\alpha\{r_T, \overline{w}(r_T)\} = {} & u^1_\alpha\{\overline{w}(r_T)\}\mathrm{I}\{r_T \in [0, x_1]\} \\ & + \sum_{j=1}^{m-1} u^{j+1}_\alpha\{\overline{w}(r_T)\}\mathrm{I}\big\{r_T \in (x_j, x_{j+1}]\big\} \\ & + u^{m+1}_\alpha\{\overline{w}(r_T)\}\mathrm{I}\{r_T \in (x_m, \infty)\}. \end{aligned} \tag{14}$$

3.1.e. *Pricing kernel*

The following theorem is the cornerstone for linking individual preferences to the market PK with its potential nonmonotonicities.

Theorem 1. *In addition to conditions (5)–(9) let $u^1(r_T,\cdot),\ldots,u^m(r_T,\cdot)$ be twice continuously differentiable for $r_T\geq 0$. Then $u_\alpha(r_T,\cdot)$ is continuously differentiable for every realization r_T of R_T. Furthermore for any $\alpha_i>0$ there exists some $\beta_i>0$ such that*

$$\frac{du_\alpha(r_T,\cdot)}{dy}\Big|_{y=\overline{w}(r_T)}=\alpha_i\frac{du^i(r_T,\cdot)}{dy}\Big|_{y=\overline{w}_i(r_T)}=\alpha_i\beta_iK(r_T)=\beta K(r_T),$$

where β is a constant independent of the state r_T.

A proof of Theorem 1 is provided in Appendix A.

From Equation (14) and Theorem 1 it follows immediately that

$$\begin{aligned}\beta K(r_T)=&\frac{du^1_\alpha(\cdot)}{dy}\Big|_{y=\overline{w}(r_T)}\mathrm{I}\{r_T\in[0,x_1]\}\\&+\sum_{j=1}^{m-1}\frac{du^{j+1}_\alpha(\cdot)}{dy}\Big|_{y=\overline{w}(r_T)}\mathrm{I}\{r_T\in(x_j,x_{j+1}]\}\\&+\frac{du^{m+1}_\alpha(\cdot)}{dy}\Big|_{y=\overline{w}(r_T)}\mathrm{I}\{r_T\in(x_m,\infty)\}.\end{aligned}\tag{15}$$

From Equation (15) it becomes clear that the PK is nonincreasing separately on the intervals $[0,x_1],(x_1,x_2],\ldots,(x_m,\infty)$ but it might fail to be monotone at the reference points $x_1,\ldots,x_m$. If we assume that the initial aggregate wealth sums up to zero, it is reasonable to conclude that the market final wealth becomes $\overline{w}(r_T)=mr_T$. Then we can define the scaled PK as $\tilde{K}(r_T)\overset{\text{def}}{=}\beta K(r_T)|_{\overline{w}(r_T)=mr_T}$. Its argument r_T is interpreted as average final wealth.

The following theorem establishes the nonmonotonicity of the PK for the reference-dependent preferences.

Theorem 2. *Assume that the preferences of investor $i\in\{1,\ldots,m\}$ are representable by the utility function $u^i(x,\cdot)$ given in Equation (4), which is in addition to conditions (5)–(9), is continuously differentiable. Furthermore, assume that $x_i\neq x_k$ if $i\neq k$, for all $k\in\{1,\ldots,m\}$. If there exists $z>0$ so that*

$$\frac{du^1_i(\cdot)}{dy}\Big|_{y=\overline{w}^{1,i+1}_i(r_T)}>\frac{du^0_i(\cdot)}{dy}\Big|_{y=\overline{w}^{0,i}_i(r_T)}\quad\text{for all } r_T>z,\tag{16}$$

then if $x_i>z$, K is not monotone at x_i.

Theorem 2 implies that higher risk taking occurs in the states in which the stock index (or the average final wealth) exceeds investor's reference point x_i. This switching behavior increases marginal risk aversion on the market. Therefore, u^0_i and u^1_i can be thought of as utility indices representing bearish (low risk taking) and bullish (high risk taking) attitudes, respectively. Examples of such preferences follow in the next subsection. A proof of Theorem 2 is sketched in Appendix B.

Experimental literature employs the concept of reference dependence in various contexts. In prospect theory, for instance, the focus of reference dependence is on differentiating between attitudes toward gains and losses. In contrast, our study employs the concept of reference for studying the behavior patterns of investors below and above a certain threshold that refers to the average performance. Two types of utility specifications are used for both approaches. However, in prospect theory, one uses a convex–concave value function with distinct properties on two different subsets of the potential outcomes, whereas we characterize preferences over states by two concave utility functions with different curvatures (or magnifying factors) defined on the same domain of individual final wealth. In principle, the two models address two different phenomena and their predictions relative to the reference-induced behavior differ.

The preferences used in our model fall in a more general category in which the individual marginal utility increases with the average wealth of the peer group. The economic literature incorporates this concept under the preferences known as "Keeping up with the Joneses", see Hong, Jiang, and Zhao (2014) for a comprehensive discussion. In a financial context, the stock index can be viewed as a proxy for the market performance.

Furthermore, with our choice of preferences we come close to the literature that focuses on the compensation of fund managers, in particular the remuneration schemes that target performance over a benchmark. There are several contributions that model this aspect explicitly, see for instance Brennan (1993); Carpenter (2000); Tepl (2001); Gómez and Zapatero (2003); Basak, Shapiro, and Tepl (2006); Basak, Pavlova, and Shapiro (2007); Jackwerth and Hodder (2007); and Basak and Pavlova (2013). Further, Dybvig, Farnsworth, and Carpenter (2010) show that benchmarking the fund manager against the index may emerge as an optimal contract in delegation problems. These incentives of the money managers to outperform their benchmark lead to tilting their portfolio toward riskier assets. Our findings suggest similar behavior.

One purpose of this article is to explain the hump-shaped feature of the EPK that occurs in the middle of the distribution of stock index returns. In order to reproduce this empirical feature, the reference points of the investors should be located around the zero net returns, i.e., the present value of the stock index. Our model, in its current form, does not make any restrictions on the location of the reference points. The implications of the empirical evidence for the interpretation of our model is that investors anchor their preferences to the present value of the index, as it provides information about the average level of wealth. Technically, the reference-dependent preferences model that we describe in this section can in fact generate positive jumps also for large returns and is able to generate locally increasing PKs in the right tail of the return distribution, as long as Theorem 2 is satisfied. However, the motivation for linking the reference points to the stock index seen as the average market performance is lost under such an utilization. Therefore, we attribute the increasing right tail of the EPK to a distinct mechanism that will be discussed in Section 4 and we focus in the following on reference points located around the central region of return distribution.

3.2 PK Specifications

In this section, we illustrate how analytical and graphical representations of PK derived from Equation (15) depend on the parameterization of the utility indices and on the reference points.

Assuming that investors have possibly heterogeneous reference points x_i we denote

$$F(r_T) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{I}\{r_T \in (0, x_i]\} \tag{17}$$

the cumulative distribution function (c.d.f.) of the reference points. The interpretation of F is as the share of agents that have preferences described by u^1 at the realization r_T of the state variable R_T. F will be further used to derive the PK specifications for the case

$u_i^l = u^l$, $l \in \{0, 1\}$ for all i. The formula for F in Equation (17) holds also when only a fraction m^*/m investors have reference-dependent preferences but its interpretation as c.d.f. is no longer valid because F does not integrate to 1. In this case, some investors have either preferences described by u^0, equivalent to $x_i \to 0$, or preferences described by u^1, equivalent to $x_i \to \infty$.

Our model is specialized to discrete reference points. F is a step-wise function, the PK is decreasing on ordered subintervals and has potential jumps at the reference points. Assume, for example, that all investors have an identical reference point. Then F takes values in $\{0, 1\}$ and the PK will have a single jump, see upper panels of Figure 2. When investors have heterogeneous reference points, the number of jumps in the PK is smaller or equal to the number of distinct reference points and local nonmonotonicities occur at each reference point for which Theorem 2 is satisfied. The lower displays of Figure 2 mark this point for a few reference points. The resulting PK has evident discontinuities.

Typical PKs estimates are smooth. In order to generate such a pattern, our model has to assume a large number of reference points. Figure 3 depicts simulated PKs for which the reference points have been drawn from a normal distribution. Bell-shaped distributions for reference points realizations, including asymmetric specifications, lead to aggregate utilities that are concave–convex–concave functions and generate hump-shaped PKs.

We use Equation (4) to derive formulas for the PK for two basic choices of the utility indices: (i) constant relative risk aversion (CRRA) functions with different concavity parameters and (ii) utility indices with a magnifying component for values of the stock index above the reference. Without loss of generality, in this section we assume that

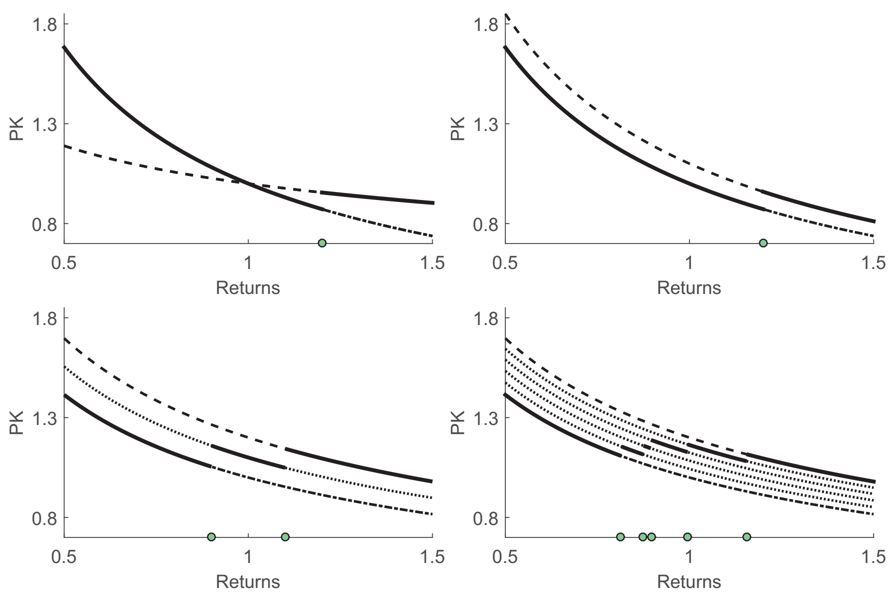

Figure 2. $\frac{du_z(\cdot)}{dy}$ (solid), $\frac{du_z^j(\cdot)}{dy}$ (dotted), $\frac{du_z^1(\cdot)}{dy}$ (dot-dashed), and $\frac{du_z^{m+1}(\cdot)}{dy}$ (dashed). Upper panels: one reference point. Lower panels: several reference points.

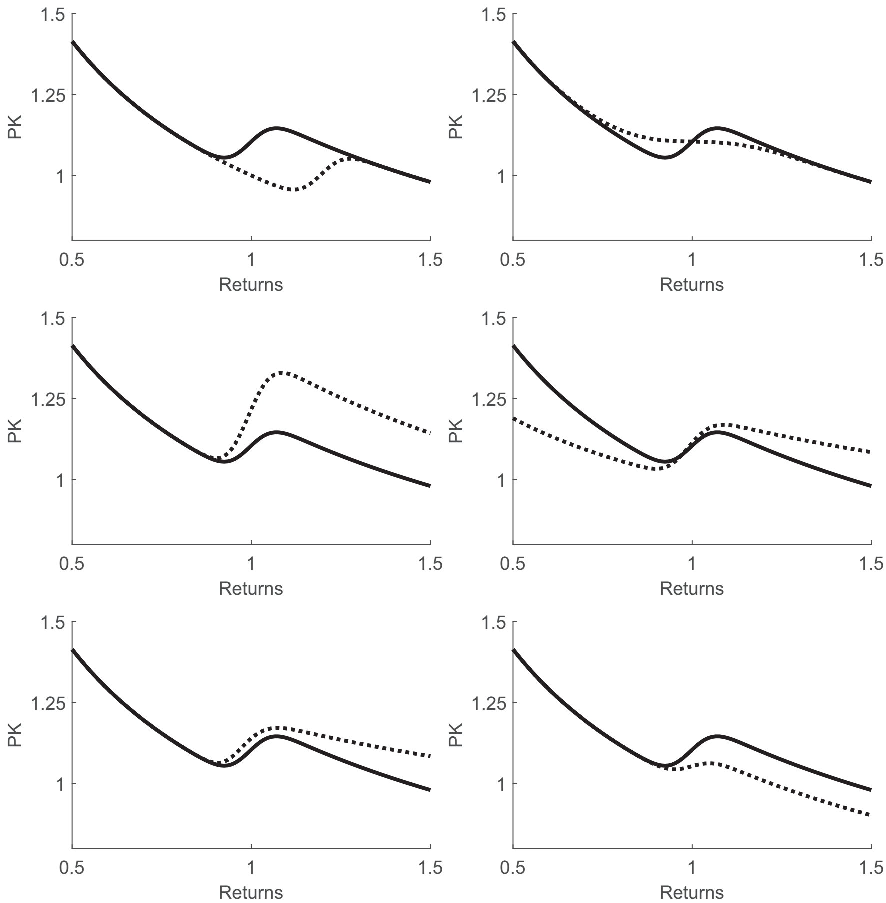

Figure 3. Impact of model parameters on the shape of PK: baseline model (solid): $\gamma = 0.5$, $b_0 = 1$, $b_1 = 1.2$, $x_i \sim N(1, 0.05)$. Comparative models (dotted): top left panel $x_i \sim N(1.2, 0.05)$; top right panel $x_i \sim N(1, 0.15)$; middle left panel $b_1 = 1.4$; middle right panel $\gamma = 0.25$; lower left panel $\gamma_0 = 0.5$, $\gamma_1 = 0.25$; lower right panel $m^*/m = 1/2$.

$\alpha_1 = \ldots = \alpha_m = 1$. The basic examples incorporate the idea that changes in risk aversion over a benchmark can be induced either through changes in curvature or applying a magnifying factor to the utility index.

Example 1. We consider individual utility indices u^l that differ by a constant b_l, $l \in \{0, 1\}$:

$$u^l(y) = \begin{cases} b_l \dfrac{y^{1-\gamma}}{1-\gamma} & \text{if } \gamma > 0 \text{ and } \gamma \neq 1 \\ b_l \log(y) & \text{if } \gamma = 1. \end{cases}$$

The positive constants $b_0 < b_1$ retain the relationship between u^0 and u^1 in Theorem 2 for all r_T. Given our parametric specifications for the utility indices and F we can obtain the following formula for $\tilde{K}(r_T)$

$$\tilde{K}(r_T) = \left[\frac{r_T}{\{1 - F(r_T)\} b_0^{\frac{1}{\gamma}} + F(r_T) b_1^{\frac{1}{\gamma}}} \right]^{-\gamma}, \tag{18}$$

for every possible realization r_T of R_T.

Example 2. We consider the individual utility indices u^l with different coefficients of risk aversion γ_l, $l \in \{0, 1\}$:

$$u^l(y) = \begin{cases} \dfrac{y^{1-\gamma_l}}{1 - \gamma_l} & \text{if } \gamma_l > 0 \text{ and } \gamma_l \neq 1 \\ \log(y) & \text{if } \gamma_l = 1 \end{cases}.$$

For $\gamma_0 > \gamma_1 > 0$ it follows that $u_1^0, \ldots, u_m^0$ represent more risk averse attitudes than $u_1^1, \ldots, u_m^1$ as expressed by the coefficients of relative risk aversion. The market PK can be written as a power function

$$\tilde{K}(r_T) = b(r_T) r_T^{-\gamma_\alpha(r_T)} \tag{19}$$

with non-constant aggregate coefficient of relative risk aversion $\gamma_\alpha(r_T)$:

$$\gamma_\alpha(r_T) = r_T \left[\{1 - F(r_T)\} \frac{m\overline{w}^0}{\gamma_0} + F(r_T) \frac{m\overline{w}^1}{\gamma_1} \right]^{-1}$$

and non-constant multiplicative term

$$b(r_T) = \left[\{1 - F(r_T)\} \frac{m\overline{w}^0}{r_T} + F(r_T) \frac{m\overline{w}^1}{r_T} \right]^{\gamma_\alpha(r_T)},$$

where $\overline{w}^l$, $l \in \{0, 1\}$ is the optimal wealth path under preferences l, see $\overline{w}_i^l$ in Section 3.3. Notice that for this aggregation mechanism the absolute risk aversion implied by the market is always positive, contrasting with the absolute risk aversion function of the representative agent derived in previous studies, see for instance Chabi-Yo, Garcia, and Renault (2008) or Grith, Härdle, and Park (2013).

Example 3. The combination of the previous two examples results in a PK that has the same representation as in Equation (19) with the same functional form for $\gamma_\alpha(r_T)$ and

$$b(r_T) = \left[\{1 - F(r_T)\} b_0^{\frac{1}{\gamma_0}} \frac{m\overline{w}^0}{r_T} + F(r_T) b_1^{\frac{1}{\gamma_1}} \frac{m\overline{w}^1}{r_T} \right]^{\gamma_\alpha(r_T)}.$$

Further generalizations of the previous examples are possible if we consider for instance heterogeneity of agents in CRRA, γ_{li} and constants b_{li}. However, the examples above are illustrative for our case and offer enough flexibility to reproduce the main shape modifications of the PKs in terms of key characteristics of the model: concavity of preferences, the magnifying parameter, and the characteristics of the distribution of the reference points. In Figure 3, we illustrate how these characteristics influence the shape of the PK.

When using the above specifications to estimate the PKs in practice, the parameters are often only partially identified, i.e., there are several combinations of parameters that yield the same shape of the PK. Even in the presence of the identification problem, they illustrate the main mechanisms through which nonmonotonicity can occur in our model and show which features induce modifications in the shape of PK. Solving the identification problem is possible if we consider, for instance, dynamic variations of the PKs under shape restrictions. This is, however, a topic in itself and we leave it for further research. In addition, it is also possible to find simpler or more intuitive closed-form solutions of the PK in similar equilibrium models.

3.3 Asset Allocation and Investors' Preferences

Based on Theorem 1 and Appendix A we can establish the general relationship between the optimal terminal wealth of investor i and the market PK

$$\begin{aligned}\overline{w}_i(r_T) &= I_i\{r_T, \beta_i K(r_T)\}\\ &= I_i^0\{\beta_i K(r_T)\}\mathrm{I}\{r_T \in [0, x_i]\} + I_i^1\{\beta_i K(r_T)\}\mathrm{I}\{r_T \in (x_i, \infty)\}\\ &= \overline{w}_i^0(r_T)\mathrm{I}\{r_T \in [0, x_i]\} + \overline{w}_i^1(r_T)\mathrm{I}\{r_T \in (x_i, \infty)\}\\ &= \overline{w}_i^{0,1}(r_T)\mathrm{I}\{r_T \in [0, x_1]\} + \sum_{j=1}^{i-1}\overline{w}_i^{0,j+1}(r_T)\mathrm{I}\{r_T \in (x_j, x_{j+1}]\}\\ &\quad + \sum_{j=i}^{m-1}\overline{w}_i^{1,j+1}(r_T)\mathrm{I}\{r_T \in (x_j, x_{j+1}]\} + \overline{w}_i^{1,m+1}(r_T)\mathrm{I}\{r_T \in (x_m, \infty)\},\end{aligned} \tag{20}$$

where $\overline{w}_i^l(r_T) = I_i^l\{\beta_i K(r_T)\}$, $I_i^l(\cdot)$ the inverse function of $\frac{\mathrm{d}u_i^l(\cdot)}{\mathrm{d}y}$ and $\overline{w}_i^{l,j}$ defined in Section 3.1, for $l \in \{0, 1\}$ and $j = 1, \ldots, m+1$. At the same time, the optimal wealth $\overline{w}_i(r_T)$ satisfies the budget constraint (3). From Theorem 1 it follows that for any $i = 1, \ldots, m$

$$\mathsf{E}\left[u^i\{R_T, \overline{w}_i(R_T)\}' - \beta_i K(R_T)\right] = 0, \text{ where } u^i\{r_T, \overline{w}_i(r_T)\}' \stackrel{\text{def}}{=} \frac{\mathrm{d}u^i(r_T, \cdot)}{\mathrm{d}y}\Big|_{y=\overline{w}_i(r_T)}.$$

For any two investors i and $i+1$, a sufficient condition for investor i to invest more in the risky asset than investor $i+1$ is that

$$\mathsf{E}\left[u^i\{R_T, \overline{w}_{i+1}(R_T)\}' - \beta_i K(R_T)\right] > 0.$$

This follows from the fact that $u^i(r_T, \cdot)$ is strictly concave. Further, we show that if $x_i < x_{i+1}$, the equilibrium conditions imply that investor i invests more in the risky asset than investor $i+1$, all other things being kept constant, i.e., $w_{k0} = w_0$, $\alpha_k = \alpha$, $u_k^l = u^l$ $\overline{w}_k^{l,j} = \overline{w}^{l,j}$, $k \in \{i, i+1\}$. Under these assumptions, and using Equation (20), we get

$$\int_{x_i}^{x_{i+1}} u^1\{\overline{w}^{0,i+1}(x)\}' dx > \int_{x_i}^{x_{i+1}} u^1\{\overline{w}^{1,i+1}(x)\}' dx$$

if inequality (16) in Theorem 2 holds. Inequality (16) together with strict concavity of u^l implies $\overline{w}^{0,i+1}(r_T) < \overline{w}^{0,i}(r_T) < \overline{w}^{1,i+1}(r_T)$, for all $r_T \geq z$.

For u_i^0 denoting bearish and u_i^1 bullish attitudes, investor i invests a higher fraction of wealth in the risky assets when $x_i \geq z$ is relatively small. Small x_i increases investor's expected final wealth $\mathsf{E}[\overline{w}_i(R_T)]$ and hence the fraction invested in the risky assets. The explanation for this lies in the fact that $\overline{w}_i^1(r_T) > \overline{w}_i^0(r_T)$ for $r_T \geq z$ on a relatively larger interval. In essence, portfolio holding in our model is equivalent to a weighted average investment in two portfolio that are chosen under bullish and bearish attitudes, where the weights depend on the reference point. Investors with a small reference point x_i appear to be more risk taking. Conversely, the investors with a reference point further to the right seem to be more risk averse. One possible explanation for this appears in Basak, Shapiro, and Tepl (2006) who find that when investment funds are moderately behind the benchmark they increase their risk taking.

The terminal wealth for three types of agents is illustrated in Figure 4. The PK parameterization corresponds to Example 3.2 with $\gamma = 0.5$, $b_0 = 1$, $b_1 = 1.2$, $m = 100$, $\alpha_1 = \cdots = \alpha_m = 1$ and x_i drawn from a normal distribution N(1, 0.05). The 45 degree line depicts the final wealth of the average investor while the solid line corresponds to the optimal wealth of the individual investors. The shaded areas mark the difference between the individual and average investor's final wealth; if positioned above the 45 degree line the investor overperforms the market for the respective states r_T. Individual investors perform better than the market when the realizations of the index return lie between their individual and the largest reference point, i.e., $r_T \in (x_i, x_m)$. The portfolio of a less risk averse investor exceeds the market index more often and mostly for negative returns. This feature of the model can be linked to the results of Moskowitz (2000) and Kosowski (2002) who find that mutual funds on average outperform the market in recessions.

4. Real Data Analysis

PKs are not directly observable and need to be estimated from the available data. Their shapes vary with the economic conditions, reflecting changes in the variables that enter the SDF. In this section, we describe the estimation methodology for the EPKs in Figure 1. Furthermore, we study how the VRP, defined as the difference between square root of the

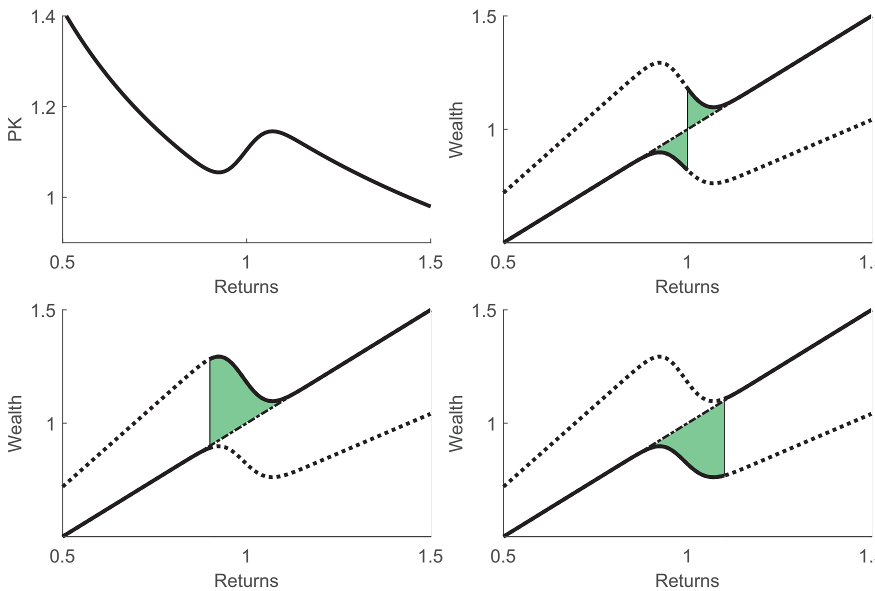

Figure 4. Market PK and (scaled) final wealth of three types of agents: mixed agent (upper right); less risk averse agent (lower left) and more risk averse agents (lower right); $m\overline{w}_i(r_T)$ (solid), $m\overline{w}_i^0(r_T)$ and $m\overline{w}_i^1(r_T)$ (dotted) for the baseline model from (Figure 3).

expected variance under the risk neutral and under the objective distribution of the index, is able to explain the variation of the hump and U-shape feature of the EPK.

The use of variance to study PKs can be linked to models that aim to describe the data generating process of the risky stock, in stylized cases. A geometric Brownian motion for the representation of the risky asset dynamics, for instance, is consistent with a representative agent model with a power utility index. The corresponding coefficient of relative risk aversion is constant and equal to the expected excess return over variance. This implies an inverse relationship between the variance of the stock and the risk aversion coefficient. This simple framework has been adapted to incorporate dynamic volatility models, with the same implications for the relationship between the preference parameters and stock variance, see Gordon and St-Amour (2004). High variance is associated with low risk aversion ceteris paribus.

More recent empirical investigations of the equity and derivative prices suggest that the volatility estimates under the risk neutral and objective measure, often referred to as implied and realized volatility, are not equal, see for instance Bollerslev, Gibson, and Zhou (2011), Christoffersen, Heston, and Jacobs (2013), Polkovnichenko and Zhao (2013), and Barone-Adesi, Mancini, and Shefrin (2013) for results on S&P 500 index. Figure 5 illustrates similar behavior for the volatilities estimated from DAX 30 Index and option data. There is a strong co-movement between the two estimated volatilities but their difference is time varying. Bliss and Panigirtzoglu (2004) and Alonso, Blanco, and Rubio (2005) report negative correlations between the coefficient of relative risk aversion implied from stock index and option returns, and the implied volatility. Other studies find a negative correlation between the VRP and the risk aversion coefficient, see Bollerslev, Gibson, and Zhou (2011). The reported results suggest that implied volatility is more sensitive to the changes in the economic environment than the realized volatility, and its changes account for most of the variability in the VRP. We look at this aspect in the following sections. Moreover, the studies mentioned above consider overall decreasing PKs. In contrast, we evaluate the influence of the VRP on the shape of the PK in the presence of local nonmonotonicities.

4.1 Estimation Methodology

The risk neutral pricing rule implies the existence of an equivalent martingale measure with a density that we denote q, such that $K(x) = q(x)/p(x)$ in Equation (2). We propose an estimator of the PK that is constructed as a ratio of two estimated densities. We also incorporate the idea that the intertemporal PKs are time varying and consider conditional densities at different moments in time.

For the estimation of conditional risk neutral density q we use the results of Breeden and Litzenberger (1987) and Banz and Miller (1978), who show that for a continuum of strikes the risk neutral density is proportional to the quotient of the European call options with respect to the strike price. We implement an equivalent method based on Rookley (1997). Its operationalization involves the estimation of smooth Black–Scholes implied volatility functions and their first two derivatives from discrete stock index option prices. We perform local polynomial regressions of degree three and use a unique bandwidth across all curves in the sample. This bandwidth is defined as the maximum of the individual optimal bandwidths computed through the leave-one-out cross-validation criterion, see Stone (1974), for each curve. Grith, Härdle, and Schienle (2012) show that oversmoothing the implied volatility improves the tails of the density with negligible effects on the bias in the middle of the

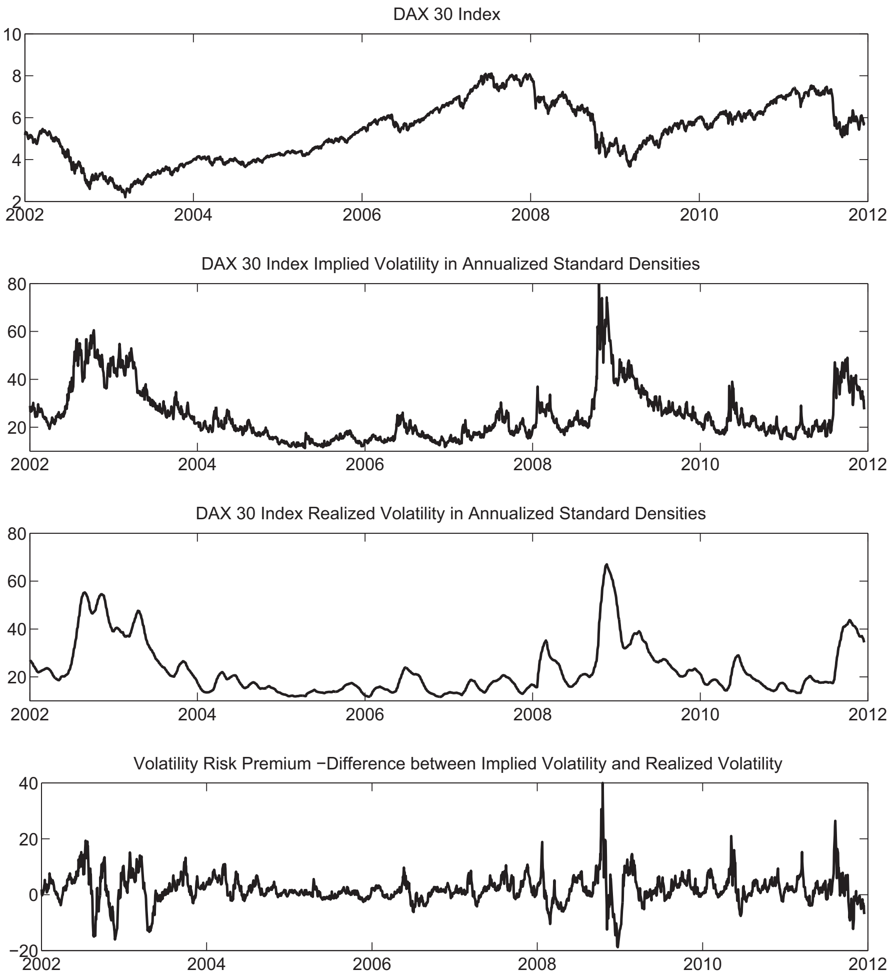

Figure 5. Time series of DAX 30 Index (in thousands), implied and realized volatility, and VRP.

distribution. A detailed account on the estimation procedure is given in Grith, Härdle, and Park (2013).

To estimate the conditional physical density p we take into account the time-varying nature of the volatility. For each time point, we approximate the 1-month ahead expected variance under the physical measure by the quadratic variation of the stock index over the last twenty-eight calendar days. A simple estimator involves the sum of daily squared index returns. The realized volatility is then calculated as the square root of its annualized value. In addition, at each moment in time, we are able to compute in advance the corresponding realized monthly index return. With the availability of two time series of ex-post realized volatility and ex-ante 1-month returns, we can estimate the conditional density of returns using the factorization of the joint probability density function in conditional and marginal. The joint density of returns and volatility is obtained as a two-dimensional kernel density estimator. We use a bivariate Gaussian kernel function and the rule-of-thumb bandwidth suggested by

Bowman and Azzalini (1997). For the marginal density of realized volatility we use an univariate kernel density estimator and Silverman's rule-of-thumb for the optimal bandwidth.

The main difference between our estimation procedure for p and previous studies stems in the fact that we use the entire range of returns realized during the time interval considered, while most of the other procedures look at the historical realizations up to the estimation point. The underlying assumption for our approach is that p is stationary conditional on the volatility during the period considered. The investors are rational in the sense that they use the returns of the last month to build estimates of the volatility and have true beliefs about the conditional distribution of returns.

If we assume that both distributions are stationary, conditional on the respective volatilities, we can use the VRP as a conditioning variable to capture the changes in risk preferences implied by the PK estimates.

4.2 Data Description

For the estimation of the risk neutral density we use daily out-of-the-money European call and put settlement prices written on the DAX 30 index between the January 2, 2002, and the December 30, 2011. We address the bid-ask gap for each maturity by shifting the daily strings of option prices to intersect at the strike value that equals the current price. After this operation, we translate the puts in corresponding calls using the Black–Scholes put-call parity formula. The value of the index on a particular day is the settlement price. Our proxies for the risk-free interest rates equal the daily quoted EURIBOR for 1-month maturity. During 1 day, option contracts with only a few maturities, on average five, are traded over the counter due to the fact that expiration dates are fixed on every third Friday of the month. To obtain densities at desired twenty-eight calendar day maturities we interpolate two estimates with maturities closest to the one of interest. For the volatility under q we use the implied volatility index VDAX–NEW daily time series. This represents the squared root annualized expected risk neutral volatility over the next 30 days and is derived directly from the option prices.

We estimate conditional physical densities of the index returns at each date for which we compute the risk neutral density using DAX 30 index returns. Stock index return series include additional twenty-eight calendar days before and after the reference period. Not for each day at which we can estimate q there exists a corresponding realized index price value and we delete these days from our sample. We obtain 2,348 daily density estimates.

Next, for each day we extract the values of the risk neutral and physical density corresponding to the 1-month ahead stock index return and we pair them with the corresponding implied and realized volatility at the estimation day. We can now calculate the EPK by taking the ratio of the two estimated densities at the corresponding values of the realized return, at each date in our sample.

4.3 Empirical Results

The results are displayed in Figure 1. The observed patterns are consistent when we vary the parameters used in the estimation, e.g., different bandwidths for the conditional probability density function, or if we use univariate kernel density estimators to compute the conditional objective probability density functions, by selecting the realized returns associated with VRP values within a small interval.

Figure 1 provides a measure of the average PK over the entire sample period. Figure 6 displays the yearly estimates and offers some insights into the time variability of the PK. We find that for years 2004–07, as well as 2010 the PKs look hump-shaped, with an increasing region for high returns. For the rest of the years, the estimates are very volatile. These periods correspond in fact to empirically reported U-shaped PKs for the S&P 500 index, e.g., Polkovnichenko and Zhao (2013) and Song and Xiu (2016). The yearly results are not conclusive and we proceed further with the analysis.

We investigate the behavior of the PK estimates conditional on the VRP. This approach has some practical advantages derived from the possibility of combining two variables, implied and realized volatility, into a single indicator with relevant economic content. In Figure 7 we display the estimated PK for four equally distant intervals spanning the interquartile range of the VRP, for increasing levels from one to four. The solid lines are the local constant estimates for the bandwidth equal with 0.1. The estimates display a hump that is prominent mostly for low values of the volatility premium. The increasing slopes, consistent with the positive probability mass function of the reference points, correspond mostly to returns smaller than one and suggest that the investors increase their risk appetite when their benchmark is moderately behind the current stock index value.

In Figure 8 we plot the risk neutral and physical density estimates used to compute the corresponding empirical values for the PKs in Figure 7. The graphs show that the physical probability density is located slightly at the right due to the risk premium. Both densities are left-skewed but the risk neutral density is more peaked for low values of VRP. The shape of the physical density is different from Jackwerth and Rubinstein (1996), who report right skewness of the probability density function. The PK hump arises from the relatively more peaked risk neutral densities. In general, for smooth densities, locally increasing PK occurs for $p'/p - q'/q < 0$, where the numerators refer to the derivatives of the functions.

To get further insights into the EPK variation, we investigate the relationship between returns, implied and realized volatility, and the VRP, for the data that fall within the

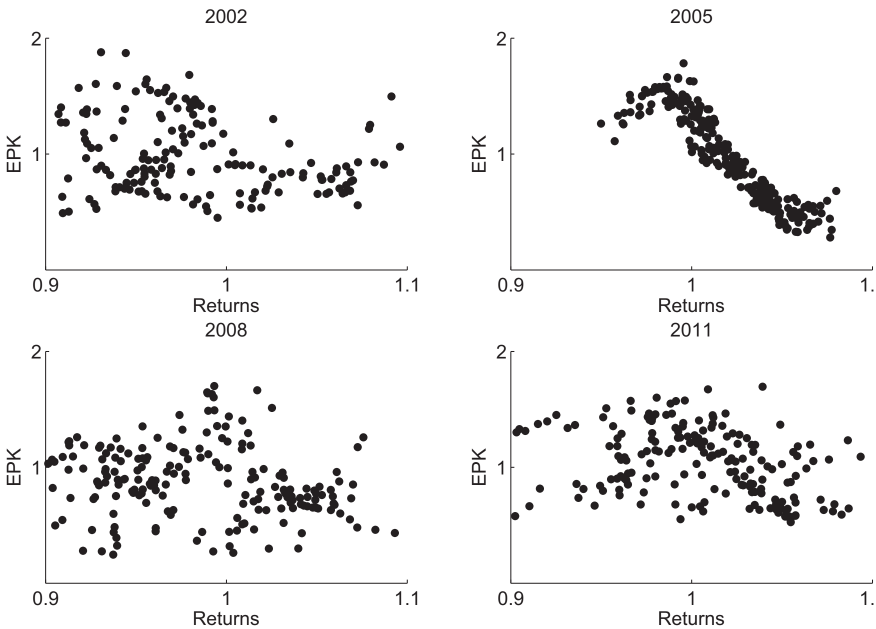

Figure 6. Yearly PK estimates.

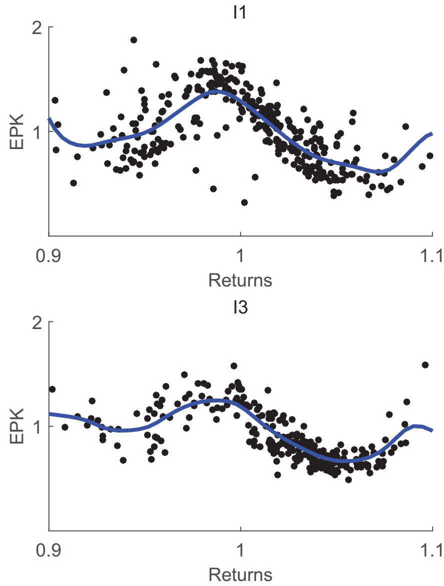

Figure 7. PK estimates on VRP intervals.

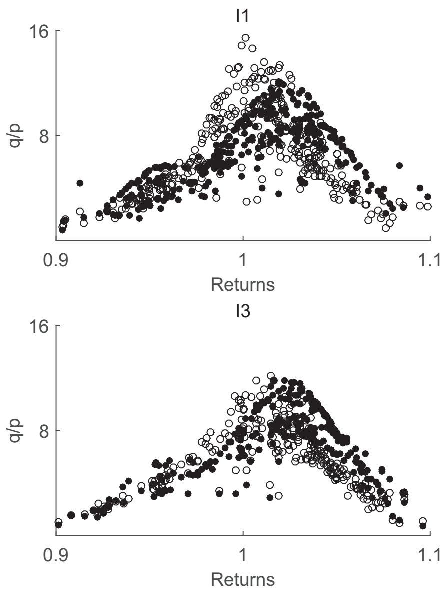

Figure 8. *q* and *p* estimates on VRP intervals.

interquartile range of VRP in Figure 9. The conditional mean return does not depend on the level of risk premium and we find that the slope parameter for regressing returns on the premium is not significant. On average, the realized volatility does not depend on the VRP either (the slope coefficient is close to zero and not significant in Table I); on the other hand, the results show a positive relationship between the implied volatility and VRP. These results suggest that the risk neutral density is more sensitive to the changing economic conditions leading to variations in the shape of the PK. Furthermore, we find an asymmetric U-shaped relationship between volatility, both implied and realized, and returns when we run a nonparametric regression; the convexity is more pronounced at the right and hence we find a positive yet not significant slope parameter when we run a linear regression. Based on our data, we do not find significant statistical evidence to support a positive relationship between returns and volatility.

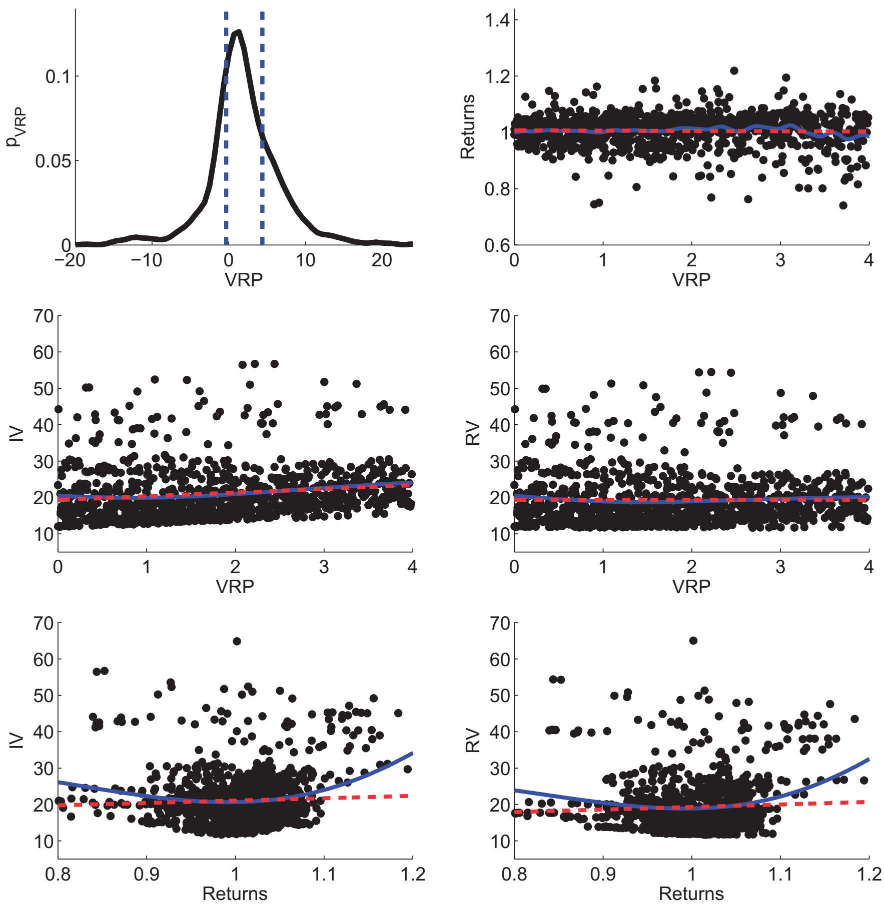

Figure 9. Top left panel: kernel density estimate for the VRP and the lower and upper quartile lines. Other panels: scatter plots and fitted curves with local linear and linear regressions.

Table I. Linear regression coefficients

d.v.	Ret.	IV	RV	IV	RV
i.v.	VRP	VRP	VRP	Ret.	Ret.
Const.	1.00*	19.30*	19.36*	14.55*	12.30*
Slope	−0.00	1.02*	0.02	6.51	6.99

*5% significance level.

4.4 PK and Market Uncertainty

In this section, we use the VRP as an indicator for the macroeconomic uncertainty risk. Buraschi, Trojani, and Vedolin (2014) find that index VRP is related to the market uncertainty and to the degree of the disagreement of the investors. In Figure 7 we see that the hump of the PK is more pronounced when market uncertainty is low. These results are consistent with Grith, Härdle, and Park (2013) who relate the time variability of the PK hump to the business cycle indicators and find that it is more prominent during the economic expansion.

In our model, the local increasing PK in the central region of the return distribution is due to the heterogeneity of reference points. In Figure 4 we illustrate that the more leveraged investors are identified through their reference points at the left side of the hump, i.e., for strikes lower than the current price. A plausible interpretation for this is that investors increase their positions in the risky assets when the returns of their portfolios fall below the market benchmark. This phenomenon is more pronounced when the market conditions are favorable and the average investor is performing well (e.g., low volatility and higher returns on the market stock).

In order to understand this feature from option price perspective, we assume that the buying pressure is what drives the prices of options up. This is consistent with portfolio insurance, see Bollen and Whaley (2004) and Constantinides *et al.* (2011) for out-of-the-money index puts. This suggests that the prices of options for a moderate range of strikes smaller than the current stock price correspond to an investment strategy in the index through benchmarking. As a consequence, we observe more expensive out-of-the-money puts around the current value of the stock index. In contrast, during high uncertainty, the risky investments are less attractive and the proportion of investors pursuing an investment strategy relative to the benchmark is lower (or there are less incentives to outperform the market). Therefore, the hump is less pronounced.

We can interpret these results also by looking at the relative shape of the risk neutral and physical density. For a fixed p, the shape of the PK will depend on the relative shape of q. Focusing on the bond option market, Beber and Brandt (2006) find that announcements with positive content for the economic prospects lead to less negatively skewed and less fat-tailed risk neutral densities. In addition, they report that all announcements, that can be interpreted as reducing uncertainty, decrease the second moment of the risk neutral distribution. Negative skewness leads to higher prices of the puts relative to calls of the same strike prices; moreover, fatter tails of the risk neutral density compared with the physical density tails are consistent with overpriced far out-of-the-money options. Focusing on the relative behavior of the tails of the two distributions, these observations suggest that the need for insurance against tail events is higher whenever the market conditions deteriorate, driving up the observed prices of far out-of-the-money options. Our results are further confirmed by

Pena, Rubio, and Serna (1999) who analyze the options written on the Spanish IBEX-35 index and find that the stable periods with increasing levels of the stock index tend to be associated with a higher degree of curvature of the volatility smile, consistent with more leptokurtic risk neutral densities. Fatter tails of the risk neutral density induce a more pronounced U-shaped PK. Relating this to models that use heterogeneity of investors to explain U-shaped feature of the PK implies that during uncertainty investors disagree to a higher extend in their behavior.

The analysis in this section suggests that the behavior in the tails and in the central part of the PK is connected: when uncertainty is low the hump is more pronounced, when it is high the U-shaped PK is prevalent. From a risk aversion perspective the results suggest that the overall investment in the risky stock is higher during the economic expansion. Under contraction or high uncertainty the investors are more cautious and seek insurance against high moves in the price of the underlying. A formalization of these findings is left for future research.

5. Conclusions

Our theoretical model encompasses a fixed investment horizon in which we attempt to explain the hump-shaped EPK. In this setting, investors' preferences depend on heterogeneous, exogenously given reference points in the space of the market return. In equilibrium the von Neumann–Morgenstern utility index of the aggregate agent may switch between different regimes depending on the reference points, possibly leading to jumps in the PK. The natural extension to our model is to build a dynamic equilibrium model in which the formation of reference points is endogenous. The history of previous gains and losses, e.g., Barberis, Huang, and Santos (2001) and learning, e.g., Benzoni, Collin-Dufresne, and Goldstein (2011); Bernales and Guidolin (2015) offer possible extensions for the implementation of the model in a dynamic setting.

By looking at the empirical data we find that the distribution of the reference points is located at the left of the current value of the stock index value. Moreover, the hump of the PK is more pronounced when the macroeconomic uncertainty is low. Then it is more likely that the investors are targeting the performance relative to the market index. Both of these findings justify the use of the reference-dependent preferences to explain the PK puzzle. In addition, we reconcile the findings about the U-shaped and hump-shaped PK by suggesting the VRP as a state variable related to the macroeconomic uncertainty that can explain the changes in the PK. The U-shape feature may be present in varying degrees during all periods, while the hump is most likely to occur during the economic conditions characterized by low uncertainty.

Our model can be adapted to other markets than the stock index to ascertain whether similar behavior occurs, see for example Dittmar (2002) for cross-section of equity returns; Li and Zhao (2009) for interest rate cap prices; Tang and Xiong (2012) for commodity futures; and Hamerle, Igl, and Plank (2012) for credit derivatives.

Appendix A

In this section we provide a proof for Theorem 1. Its conditions are assumed to hold throughout this section of the appendix. First, we characterize the optimal terminal wealth $\overline{w}_1(R_T), \ldots, \overline{w}_m(R_T)$ of the individual investor.

The Inada conditions together with Equation (6) imply that for any $i \in \{1, \ldots, m\}$ and every $r_T \geq 0$ the mapping $\frac{du^i(r_T, \cdot)}{dy}|_{(0,\infty)}$ is injective onto $(0, \infty)$ with continuously differentiable, strictly decreasing inverse say $I_i(r_T, \cdot)$. This enables us to apply the dominated convergence theorem to show

(A.1) continuity of mappings

$$g^i_{r_T} : (0, \infty) \to \mathbb{R}, y \to I_i\{r_T, \beta_i K(r_T)\}K(r_T);$$

(A.2) $\lim_{\beta_i \to 0} g^i_{r_T}(\beta_i) = \infty$ and $\lim_{\beta_i \to \infty} g^i_{r_T}(\beta_i) = 0.$

We are now ready to extend the classical characterization of the optimal terminal wealth to the case of extended expected utility preferences.

Theorem A. *Assuming that relations (5)–(9) hold, there exists $\beta_i > 0$ such that*

$$\overline{w}_i(R_T) = I_i\{R_T, \beta_i K(R_T)\} \quad \text{for every} \quad i = 1, \ldots, m.$$

***Proof*:** Let us fix $i \in \{1, \ldots, m\}$ and denote $z_i \stackrel{\text{def}}{=} w^i_0 + \mathsf{E}[e_i(R_T)K(R_T)]$. Since $z_i > 0$ we may find in view of Equations (A.1), (A.2) some $\beta_i > 0$ with $\int_0^\infty g^i_{r_T}(\beta_i) dr_T = z_i$.

Let $w(R_T)$ be a nonnegative random variable with $\mathsf{E}[w(R_T)K(R_T)] \leq z_i$. Then

$$\begin{aligned}
\mathsf{E}[u\{R_T, w(R_T)\}] + \beta_i\{z_i - \mathsf{E}[w(R_T)K(R_T)]\} \\
&= \beta_i z_i + \mathsf{E}[u\{R_T, w(R_T)\} - \beta_i w(R_T)K(R_T)] \\
&\leq \beta_i z_i + \sup\nolimits_{x \geq 0} \mathsf{E}[u(R_T, x) - \beta_i x K(R_T)] \\
&= \beta_i z_i + \mathsf{E}[u[R_T, I_i\{R_T, \beta_i K(R_T)\}] - \beta_i I_i\{R_T, \beta_i K(R_T)\}K(R_T)] \\
&= \mathsf{E}[u\{R_T, I_i(R_T, \beta_i K(R_T))\}].
\end{aligned}$$

Therefore $I_i\{R_T, \beta_i K(R_T)\}$ solves the optimization problem of investor i. Moreover, the numerical representation u^i of investor i's preferences is strictly concave in view of strict concavity of $u^i(r_T, \cdot)$ for every $r_T \geq 0$. In particular, $I_i\{R_T, \beta_i K(R_T)\}$ is the unique solution, hence being identical with $\overline{w}_i(R_T)$.

***Proof of Theorem 1*:** Without loss of generality, let us set $g_i \stackrel{\text{def}}{=} \alpha_i u^i$ and $g \stackrel{\text{def}}{=} u_\alpha$ (cf. Appendix C). Then, in view of Lemma 1 and Proposition 1 we obtain

$$u_\alpha\{r_T, \overline{w}(r_T)\} = \sum_{i=1}^m \alpha_i u^i\{r_T, \overline{w}^i(r_T)\}$$

for every realization r_T of R_T.

On the one hand, by Theorem A, there exists $\beta_i > 0$ such that

$$\overline{w}_i(R_T) = I_i\{R_T, \beta_i K(R_T)\} > 0.$$

On the other hand, due to Proposition 1, $u_\alpha(r_T, \cdot)|_{(0,\infty)}$ is differentiable, satisfying

$$\alpha_i \frac{du^i(r_T, \cdot)}{dy}\Big|_{y=\overline{w}_i(r_T)} = \frac{du_\alpha(r_T, \cdot)}{dy}\Big|_{y=\overline{w}(r_T)}$$

for any realization r_T and $i \in \{1, \ldots, m\}$. Now, the statement of Theorem 1 is clear.

Appendix B

From Equation (13) one can show that u_α^j, $j = 1, \ldots, m+1$, is continuous, strictly increasing and strictly concave, twice continuously differentiable. For $j \in \{i, i+1\}$ we obtain

$$\frac{du_\alpha^{i+1}(\cdot)}{dy}\bigg|_{y=\overline{w}(r_T)} - \frac{du_\alpha^{i}(\cdot)}{dy}\bigg|_{y=\overline{w}(r_T)} \geq \alpha_i \frac{du_i^1(\cdot)}{dy}\bigg|_{y=\overline{w}_i^{1,i+1}(r_T)} - \alpha_i \frac{du_i^0(\cdot)}{dy}\bigg|_{y=\overline{w}_i^{0,i}(r_T)} \text{ for every } r_T > z,$$

where z is the value of r_T for which the left hand side of the previous inequality is zero, or it is zero if the equality does not have a solution. From Equation (16) it follows immediately that

$$\frac{du_\alpha^{i+1}(\cdot)}{dy}\bigg|_{y=\overline{w}(r_T)} > \frac{du_\alpha^{i}(\cdot)}{dy}\bigg|_{y=\overline{w}(r_T)} \quad \text{for every } r_T > z.$$

The derivative of u_α^j has an inverse $I_\alpha^j : \mathbb{R}_+ \to \mathbb{R}_+$ that is continuously differentiable and strictly decreasing. The application of Lemma 1 and Proposition 1 in Appendix C yields

$$\overline{w}(r_T) = I_\alpha^1\left(\frac{du_\alpha^1(\cdot)}{dy}\bigg|_{y=\overline{w}(r_T)}\right) = \ldots = I_\alpha^{m+1}\left(\frac{du_\alpha^{m+1}(\cdot)}{dy}\bigg|_{y=\overline{w}(r_T)}\right).$$

For $x_i > z$

$$I_\alpha^i\left(\frac{du_\alpha^1(\cdot)}{dy}\bigg|_{y=\overline{w}(x_i)}\right) = \overline{w}(x_i) = I_\alpha^{i+1}\left(\frac{du_\alpha^{i+1}(\cdot)}{dy}\bigg|_{y=\overline{w}(x_i)}\right) > I_\alpha^i\left(\frac{du_\alpha^{i+1}(\cdot)}{dy}\bigg|_{y=\overline{w}(x_i)}\right).$$

This means that if $x_i \neq x_k$ for $i \neq k$ then K is not monotone at x_i. If the reference point x_i is not unique, e.g., $x_{i-1} \neq x_i = x_{i+1} = \cdots = x_{i+p} \neq x_{i+p+1}$ then inequality (16) is replaced by

$$\sum_{k=i}^{i+p} \frac{du_k^1(y)}{dy}\bigg|_{y=\overline{w}_k^{1,k+1}(r_T)} > \sum_{k=i}^{i+p} \frac{du_k^0(y)}{dy}\bigg|_{y=\overline{w}_k^{0,k}(r_T)},$$

and the proof follows the same arguments as above for $j \in \{i, i+p+1\}$.

Appendix C

Throughout this section let the mappings $g_1, \ldots g_m : \mathbb{R}_+^2 \to \mathbb{R} \cup \{-\infty\}$ satisfy the following conditions:

(B.0) $g_1(x,y), \ldots, g_m(x,y) \in \mathbb{R}$ for $x \geq 0, y > 0$;
(B.1) $g_1(x,\cdot), \ldots, g_m(x,\cdot)$ are continuous, strictly increasing, and strictly concave for $x \geq 0$;
(B.2) $g_1(\cdot,y), \ldots, g_m(\cdot,y)$ are Borel-measurable for $y \geq 0$.

Furthermore, let $g : \mathbb{R}_+^2 \to \mathbb{R}$ be defined by

$$g(x,y) = \sup_{y_i}\left\{\sum_{i=1}^m g_i(x,y_i) | y_1, \ldots, y_m \geq 0, \sum_{i=1}^m y_i \leq y\right\}.$$

Indeed $g(x,0) = \sum_{i=1}^m g_i(x,0) \in \mathbb{R}$ for $x \geq 0$, and

$$\sum_{i=1}^{m} g_i(x, \frac{y}{m}) \leq g(x,y) \leq \sum_{i=1}^{m} g_i(x,y)$$

for $x \geq 0, y > 0$ due to (B.0), (B.1).

Lemma 1. *For any $x, y \geq 0$ there is some unique $\phi(x,y) = (\phi_1(x,y), \ldots, \phi_m(x,y)) \in \mathbb{R}_+^m$ such that $\sum_{i=1}^{m} \phi_i(x,y) \leq y$ and*

$$\sum_{i=1}^{m} g_i\{x, \phi_i(x,y)\} = g(x,y).$$

Furthermore, $\sum_{i=1}^{m} \phi_i(x,y) = y$.

Proof: Let $x, y \geq 0$. For $y = 0$ the statement of Lemma 1 is obvious. So let $y > 0$, which means $g(x,y) \in \mathbb{R}$. Due to (B.1), the mapping

$$f: \left\{ (y_1, \ldots, y_m) \in \mathbb{R}_+^m \mid \sum_{i=1}^{m} y_i \leq y, \ \sum_{i=1}^{m} g_i(x, y_i) \geq g(x,y) \right\} \rightarrow \mathbb{R}, f(y_1, \ldots, y_m) = \sum_{i=1}^{m} g_i(x, y_i)$$

is continuous, strictly concave, for every $x \geq 0$, and defined on a nonvoid convex compact set. Therefore, f attains its maximum at a unique $\phi(x,y)$. Obviously, $\sum_{i=1}^{m} \phi_i(x,y) = y$ because f is strictly increasing too by Equation (B.1). The proof is complete.

Lemma 1 defines a mapping $\phi = (\phi_1, \ldots, \phi_m) : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+^m$. In order to characterize ϕ in terms of derivatives of the functions $g_1(x, \cdot), \ldots, g_m(x, \cdot)$, it is customary to impose the Inada conditions, i.e.

(B.3) for any $x \geq 0$ the mappings $g_1(x, \cdot)|_{(0,\infty)}, \ldots, g_m(x, \cdot)|_{(0,\infty)}$ are assumed to be continuously differentiable satisfying

$$\lim_{\epsilon \rightarrow 0} \frac{dg_i(x, \cdot)}{dy}|_{y=\epsilon} = \infty, \ \lim_{\epsilon \rightarrow \infty} \frac{dg_i(x, \cdot)}{dy}|_{y=e} = 0, \ i = 1, \ldots, m.$$

The Inada conditions together with condition (B.1) imply that for any $i \in \{1, \ldots, m\}$ and every $x \geq 0$ the mapping $\frac{dg_i(x,\cdot)}{dy}|_{(0,\infty)}$ is injective onto $(0, \infty)$ with continuously differentiable, strictly decreasing inverse say $I_i(x, \cdot)$.

Proposition 1. *Let the assumptions (B.0)–(B.3) be fulfilled, and let $g_1(x, \cdot)|_{(0,\infty)}, \ldots, g_m(x, \cdot)$ $|_{(0,\infty)}$ be twice continuously differentiable.*

Then for any $x \geq 0$ the mapping $g(x, \cdot)|_{(0,\infty)}$ is differentiable satisfying

$$\phi(x,y) = \left[I_1\left\{x, \frac{dg(x, \cdot)}{dy}|_y\right\}, \ldots, I_m\left\{x, \frac{dg(x, \cdot)}{dy}|_y\right\} \right] \text{ for } y > 0.$$

Proof: Let for $x \geq 0$ the mapping $F_x : (0, \infty) \times (0, \infty) \rightarrow \mathbb{R}$ be defined by $F_x(y,z) = \sum_{i=1}^{m} I_i(x,z) - y$.

Since the mappings $g_1(x, \cdot)|_{(0,\infty)}, \ldots, g_m(x, \cdot)|_{(0,\infty)}$ are assumed to be strictly concave and twice continuously differentiable, their second derivatives are strictly negative. Then by local inverse theorem the mappings $I_1(x, \cdot), \ldots, I_m(x, \cdot)$ are continuously differentiable, having strictly negative derivatives. In particular F_x is continuously differentiable, satisfying

$$\frac{dF_x}{dz}|_{(y,z)} = 0 \ \text{ for } \ y, z > 0.$$

Furthermore, since $I_1(x,\cdot),\ldots,I_m(x,\cdot)$ are continuous and strictly decreasing mappings onto $(0,\infty)$, we may find for any $y>0$ a unique $\varphi(y)>0$ with $F_x\{y,\varphi(y)\}=0$. Drawing on the implicit function theorem, $y\to\varphi(y)$ defines a differentiable mapping $\varphi:(0,\infty)\to(0,\infty)$. Moreover, for $y>0$ and $y_1,\ldots,y_m\geq 0$ with $\sum_{i=1}^m y_i\leq y$, we may conclude

$$\begin{aligned}
\sum_{i=1}^m g_i(x,y_i)+\varphi(y)(y-\sum_{i=1}^m y_i) &= \varphi(y)y+\sum_{i=1}^m\{g_i(x,y_i)-\varphi(y)y_i\}\\
&\leq \varphi(y)y+\sum_{i=1}^m \sup_{z\geq 0}\{g_i(x,z)-\varphi(y)z\}\\
&= \varphi(y)y+\sum_{i=1}^m[g_i\{x,I_i(x,\varphi(y))\}-\varphi(y)I_i\{x,\varphi(y)\}]\\
&= \sum_{i=1}^m g_i[x,I_i\{x,\varphi(y)\}]-\varphi(y)F_x\{y,\varphi(y)\}\\
&= \sum_{i=1}^m g_i[x,I_i\{x,\varphi(y)\}].
\end{aligned}$$

This means

$$g(x,y)=\sum_{i=1}^m g_i[x,I_i\{x,\varphi(y)\}],$$

and hence by Lemma 1

$$(^*)\ \phi(x,y)=(I_1\{x,\varphi(y)\},\ldots,I_m\{x,\varphi(y)\}).$$

As a further consequence $g(x,\cdot)|_{(0,\infty)}$ is differentiable satisfying

$$\frac{\mathrm{d}g(x,\cdot)}{\mathrm{d}y}\bigg|_y=\sum_{i=1}^m\varphi(y)\ \frac{\mathrm{d}I_i(x,\cdot)\circ\varphi(y)}{\mathrm{d}y}=\varphi(y)\ \frac{\mathrm{d}\left(\sum_{i=1}^m I_i(x,\cdot)\circ\varphi(y)\right)}{\mathrm{d}y}=\varphi(y).$$

In view of $(^*)$ the proof is complete.

References

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

Alonso, F., Blanco, R., and Rubio, G. (2005) Testing the forecasting performance of IBEX 35 option-implied risk-neutral densities. Working paper No. 0505, Banco de España.

Bakshi, G. and Madan, D. (2007) Investor heterogeneity and the non-monotonicity of the aggregate marginal rate of substitution in the market index. Working paper, University of Maryland.

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

Banz, R. W. and Miller, M. H. (1978) Prices for state-contingent claims: Some estimates and applications, *The Journal of Business* **51**, 653–672.

Barberis, N., Huang, M., and Santos, T. (2001) Prospect theory and assets prices, *The Quarterly Journal of Economics* **116**, 1–53.

Barone-Adesi, G., Mancini, L., and Shefrin, H. M. (2013) A tale of two investors: Estimating risk aversion, optimism, and overconfidence. 26th Australasian Finance and Banking Conference. Available at SSRN: http://ssrn.com/abstract=2319260.

Basak, S. and Pavlova, A. (2013) Asset prices and institutional investors, *American Economic Review* **103**, 1728–1758.

Basak, S., Shapiro, A., and Tepl, L. (2006) Risk management with benchmarking, *Management Science* **52**, 542–557.

Basak, S., Pavlova, A., and Shapiro, A. (2007) Optimal asset allocation and risk shifting in money management, *Review of Financial Studies* **20**, 1583–1621.

Beare, B. K. (2011) Measure preserving derivatives and the pricing kernel puzzle, *Journal of Mathematical Economics* **47**, 689–697.

Beare, B. K. and Schmidt, L. D. W. (2014) An empirical test of pricing kernel monotonicity, *Journal of Applied Econometrics*. http://dx.doi.org/10.1002/jae.2422.

Beber, A. and Brandt, M. W. (2006) The effect of macroeconomic news on beliefs and preferences: Evidence from the options market, *Journal of Monetary Economics* **53**, 1997–2039.

Benzoni, L., Collin-Dufresne, P., and Goldstein, R. S. (2011) Can standard preferences explain the prices of out-of-the-money S&P 500 put options? Working paper No. 2011-11, Federal Reserve Bank of Chicago.

Bernales, A. and Guidolin, M. (2015) Learning how to smile: Can rational learning explain the predictable dynamics in the implied volatility surface? *Journal of Financial Markets* **26**, 1–37.

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

Bollen, N. P. B. and Whaley, R. E. (2004) Does net buying pressure affect the shape of implied volatility functions? *The Journal of Finance* **59**, 711–753.

Bollerslev, T., Gibson, M., and Zhou, H. (2011) Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *Journal of Econometrics* **160**, 235–245.

Bowman, A. W. and Azzalini, A. (1997) *Applied Smoothing Techniques for Data Analysis*. Oxford Statistical Science Series, Oxford.

Branger, N., Schlag, C., and Zaharia, S. (2011) An equilibrium foundation for the Heston stochastic volatility model and U-shaped pricing kernels. Working paper, Universitt Münster.

Breeden, D. T. and Litzenberger, R. H. (1987) Prices of state-contingent claims implicit in option prices. *Journal of Business* **51**, 621–651.

Breeden, D. T., Gibbons, M. R., and Litzenberger, R. H. (1989) Empirical test of the consumption-oriented CAPM, *The Journal of Finance* **44**, 231–262.

Brennan, M. J. (1993) Agency and asset pricing. Working paper, UCLA.

Brown, D. P. and Gibbons, M. R. (1985) A simple econometric approach for utility-based asset pricing models, *The Journal of Finance* **40**, 359–381.

Buraschi, A., Trojani, F., and Vedolin, A. (2014) When uncertainty blows in the orchard: Comovement and equilibrium volatility risk premia, *The Journal of Finance* **69**, 101–137.

Campbell, J. Y. and Cochrane, J. H. (1999) By force of habit: A consumption based explanation of aggregate stock market behavior, *The Journal of Political Economy* **107**, 205–251.

Carpenter, J. N. (2000) Does option compensation increase managerial risk appetite? *Journal of Finance* **55**, 2311–2331.

Chabi-Yo, F. (2012) Pricing kernels with stochastic skewness and volatility risk, *Management Science* **58**, 624–640.

Chabi-Yo, Y., Garcia, R., and Renault, E. (2008) State dependence can explain the risk aversion puzzle, *Review of Financial Studies* **21**, 973–1011.

Christoffersen, P., Heston, S., and Jacobs, K. (2013) Capturing option anomalies with a variance-dependent pricing kernel, *The Review of Financial Studies* **26**, 1963–2006.

Constantinides, G. M. (1990) Habit formation: A resolution of the equity puzzle, *The Journal of Political Economy* **98**, 519–543.

Constantinides, G. M. and Jackwerth, J. C. (2009) Mispricing of S&P 500 index options, *The Review of Financial Studies* **22**, 1247–1277.

Constantinides, G. M., Czerwonko, M., Jackwerth, J. C., and Perrakis, S. (2011) Are options on index futures profitable for risk-averse investors? Empirical evidence, *The Journal of Finance* **66**, 1407–1437.

Cuoco, D. and Kaniel, R. (2011) Equilibrium process in the presence of delegated portfolio management, *Journal of Financial Economics* **101**, 264–296.

Dana, R. A. and Jeanblanc, M. (2003) *Financial Markets in Continuous Time*. Berlin: Springer.

Dittmar, R. F. (2002) Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns, *The Journal of Finance* **57**, 369–403.

Duffie, D. (1996) *Dynamic Asset Pricing Theory*, 2nd ed. Princeton University Press.

Dybvig, P. H. (1988) Distributional analysis of portfolio choice, *Journal of Business* **61**, 369–393.

Dybvig, P. H., Farnsworth, H. K., and Carpenter, J. N. (2010) Portfolio performance and agency, *Review of Financial Studies* **23**, 1–23.

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

Epstein, L. G. and Zin, S. E. (2001) The independence axiom and asset returns, *Journal of Empirical Finance* **8**, 537–572.

Garcia, R., Luger, R., and Renault, E. (2003) Empirical assessment of an intertemporal option pricing model with latent variables, *Journal of Econometrics* **116**, 49–83.

Giacomini, E. and Härdle, W. (2008) Dynamic semiparametric factor models in pricing kernel estimation, in S. Dabo-Niang and F. Ferraty (eds.), *Functional and Operational Statistics, Contributions to Statistics*, Springer Verlag, pp. 181–187.

Golubev, Y., Härdle, W. K., and Timofeev, R. (2014) Testing monotonicity of pricing kernels, *Advances in Statistical Analysis* **98**, 305–326.

Gómez, J. P. and Zapatero, F. (2003) Asset pricing implications of benchmarking: A two-factor CAPM, *European Journal of Finance* **9**, 343–357.

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

Grith, M., Härdle, W. K., and Schienle, M. (2012) Nonparametric estimation of risk-neutral densities, in J. C., Duan, J. E. Gentle and W. K. Härdle (eds.), *Handbook of Computational Finance*, Springer Verlag, pp. 277–305.

Grith, M., Härdle, W. K., and Park, J. (2013) Shape invariant modeling of pricing kernels and risk aversion, *Journal of Financial Econometrics* **11**, 370–399.

Hamerle, A., Igl, A., and Plank, K. (2012) Correlation smile, volatility skew, and systematic risk sensitivity of tranches, *The Journal of Derivatives* **19**, 8–27.

Härdle, W. K., Okhrin, Y., and Wang, W. (2015) Uniform confidence bands for pricing kernels, *Journal of Financial Econometrics* **13**, 376–413.

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

Heston, S. and Nandi, S. (2000) A closed-form GARCH option pricing model, *Review of Financial Studies* **13**, 585–626.

Hong, H. G., Jiang, W., and Zhao, B. (2014) Trading for status, Working paper.

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

Jackwerth, J. C. and Hodder, J. E. (2007) Incentive contracts and hedge fund management, *Journal of Financial and Quantitative Analysis* **42**, 811–826.

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

Karatzas, I. and Shreve, S. E. (1998) *Methods of Mathematical Finance*. New York: Springer.

Kosowski, R. (2002) Do mutual funds perform when it matters most to investors? U.S. mutual fund performance and risk in recessions and booms 1962–2000, Working paper, INSEAD.

Kramkov, D. and Schachermayer, W. (1999) The asymptotic elasticity of utility functions and optimal investment in incomplete markets, *The Annals of Applied Probability* **9**, 904–950.

Lewellen, K. (2006) Financing decisions when managers are risk averse, *Journal of Financial Economics* **82**, 551–589.

Li, H. and Zhao, F. (2009) Nonparametric estimation of state-price densities implicit in interest rate cap prices, *Review of Financial Studies* **22**, 4335–4376.

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

Lopes, L. L. and Oden, G. C. (1999) The role of aspiration level in risk choice: A comparison of cumulative prospect theory and SP/A theory, *Journal of Mathematical Psychology* **43**, 286–313.

Mas-Colell, A., Whinston, M. D., and Greene, J. R. (1995) *Microeconomic Theory*. Oxford University Press.

Melino, A. and Yang, A. X. (2003) State dependent preferences can explain the equity premium puzzle, *Review of Economic Dynamics* **6**, 806–830.

Moskowitz, T. J. (2000) Discussion: Mutual fund performance: An empirical decomposition into stock-picking talent, style, transaction costs, and expenses, *The Journal of Finance* **55**, 1695–1703.

Negishi, T. (1960) Welfare economics and existence of an equilibrium for a competitive economy, *Metroeconomica* **12**, 92–97.

Pena, I., Rubio, G., and Serna, G. (1999) Why do we smile? On the determinants of the implied volatility function, *Journal of Banking and Finance* **23**, 1151–1179.

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

Rookley, C. (1997) Fully exploiting the information content of intra day option quotes: Applications in option pricing and risk management, Working paper, University of Arizona.

Ross, S. A. (2004) Compensation, incentives, and the duality of risk aversion and riskiness, *Journal of Finance* **59**, 207–225.

Rubinstein, M. (1976) The valuation of uncertain income streams and the pricing of options, *Bell Journal of Economics* **7**, 407–425.

Shefrin, H. (2008) *A Behavioral Approach to Asset Pricing*, 2nd ed. Amsterdam: Academic Press/Elsevier.

Song, Z. and Xiu, D. (2016) A tale of two option markets: Pricing kernels and volatility risk, *Journal of Econometrics* **190**, 176–196.

Stone, M. (1974) Cross-validatory choice and assessment of statistical predictions, *Journal of the Royal Statistical Society* **36**, 111–147.

Tang, K. and Xiong, W. (2012) Index investment and financialization of commodities, *Financial Analysts Journal* **68**, 54–74.

Tepl, L. (2001) Optimal investment with minimum performance constraints, *Journal of Economic Dynamics and Control* **25**, 1629–1645.

Veronesi, P. (2004) Belief-dependent utilities, aversion to state-uncertainty, and asset prices, Working paper, University of Chicago.

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