

# The independence axiom and asset returns

Larry G. Epstein<sup>\*</sup>, Stanley E. Zin<sup>1</sup>

---

## Abstract

This paper integrates models of atemporal risk preference that relax the independence axiom into a recursive intertemporal asset-pricing framework. The resulting models are amenable to empirical analysis using market data and standard Euler equation methods. We are thereby able to provide the first nonlaboratory-based evidence regarding the usefulness of several new theories of risk preference for addressing standard problems in dynamic economics. Using both stock and bond returns data, we find that a model incorporating risk preferences that exhibit *first-order risk aversion* accounts for significantly more of the mean and autocorrelation properties of the data than models that exhibit only *second-order risk aversion*. Unlike the latter class of models which require parameter estimates that are outside of the admissible parameter space, e.g., negative rates of time preference, the model with first-order risk aversion generates point estimates that are economically meaningful. We also examine the relationship between first-order risk aversion and models that employ exogenous stochastic switching processes for consumption growth. © 2001 Published by Elsevier Science B.V.

*Keywords:* Independence axiom; Asset returns; Risk preference

---

## 1. Introduction

The expected utility model of decision making under risk and, particularly, its cornerstone, the independence axiom, have come under attack recently. The

---

<sup>\*</sup> Corresponding author. Department of Economics, University of Rochester, Rochester, NY 14627, USA.

<sup>1</sup> Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, PA 15213 and NBER.

empirical evidence upon which this criticism is based consists mostly of behavioral/experimental studies where subjects' choices among hypothetical and/or small scale gambles are observed (e.g., Kahneman and Tversky, 1979; Chew and Waller, 1986; Camerer, 1989a,b; Conlisk, 1989). Machina (1982) surveys much of the evidence and argues that violations of the independence axiom are both systematic and widespread. A number of new theories of choice under uncertainty have been developed in an attempt to *explain* the evidence which contradicts expected utility theory. Those upon which we focus here are due to Chew (1983, 1989) and Gul (1991).

In this paper, we use the general intertemporal asset-pricing model developed in Epstein and Zin (1989) and aggregate monthly U.S. time-series data for consumption and asset returns as the basis for an empirical examination of the generalized theories of Chew and Gul. In common with much of the empirical literature on aggregate consumption and asset returns, we assume the existence of a representative agent, the homotheticity of preferences and the rationality of expectations. We inquire whether, given these assumptions, relaxing the independence axiom in the directions defined by Chew and Gul can help account for the time-series data. To our knowledge, this is the first evidence available regarding the usefulness of these newly developed theories of choice for explaining market data. Of course, tests of a theory based on behavior in the field (as opposed to the laboratory) are prone to potentially serious problems such as errors in data measurement and unavoidable joint hypotheses. Laboratory-based methods, however, also have well-known drawbacks and the behavioral evidence is inconclusive. Some recent behavioral studies (e.g., Camerer, 1989b; Conlisk, 1989; Harrison, 1990) have cast doubt upon the extent and systematic nature of violations of expected utility theory. Thus, we feel that an analysis based upon market data would provide an important complementary piece of evidence regarding the usefulness of the generalized theories of choice. Moreover, we suspect that many economists would attach greater importance to the question of whether these new theories do (or do not) help to resolve some of the standard problems in economics, than to their consistency with laboratory behavior. The explanation of asset returns clearly qualifies as such a *standard* problem.

Representative agent models, with preferences represented by the expected value of the discounted sum of within-period utilities, have not performed well in explaining the behavior of consumption and asset returns over time (e.g., Hansen and Singleton, 1982, 1983; Mehra and Prescott, 1985). This, in part, motivates the work in Epstein and Zin (1989), which formulates a class of intertemporal utility functions which are recursive, but not necessarily additive or consistent with expected utility theory. Recursive utility permits some degree of separation in the modeling of risk aversion and intertemporal substitution. In Epstein and Zin (1991), generalized method of moments estimation procedures are applied to the Euler equations implied by a particular parametric member of this class of utility functions. The results are mixed though some support is provided for the general-

ized specification. The functional form used in that study does not conform with intertemporal expected utility. However, it satisfies the independence axiom for the set of so-called timeless wealth gambles—those for which all uncertainty is resolved before further consumption/savings decisions are made. Since these are precisely the sort of gambles that are considered in the experimental literature, the specification of intertemporal utility that is estimated in Epstein and Zin (1991) is inconsistent with the cited evidence against the independence axiom. This paper focuses on the empirical gains from further generalizations of intertemporal utility to specifications in which orderings of timeless wealth gambles conform with the theories of Chew and Gul and not necessarily with the independence axiom.

The Euler equations implied by a representative agent's consumption/portfolio selection problem form the basis for the empirical analysis. We first make graphical comparisons of the mean/variance properties of the stochastic discount rates, i.e., the marginal rates of intertemporal substitution, that are implied by the Euler equations for each model. In addition, these mean/variance properties can be used as in Hansen and Jagannathan (1991) to form an informal test of these asset-pricing models. Second, we provide a more formal statistical analysis based on the generalized method of moments. Finally, we explore the relationship between the general preference-based models and standard expected utility models in which consumption growth rates are subject to exogenous stochastic process switching.

The paper proceeds as follows. The next two sections lay the groundwork for the empirical analysis by summarizing and applying the most relevant material from the two literatures on decision theory and intertemporal asset pricing. In Section 2, the Chew (1983, 1989) and Gul (1991) utility functions are described. Section 3 outlines their integration into a recursive model of intertemporal utility based on Epstein and Zin (1989) and presents the implied model of asset returns. Section 4 contains the empirical analysis. Section 5 concludes and summarizes the paper.

## 2. Certainty equivalents for timeless wealth gambles

In this section, we summarize the relevant aspects of the utility functions proposed by Chew (1983, 1989) and Gul (1991). These functions represent preferences over atemporal or *one-shot* gambles. Integration into a multiperiod setting is described in the next section.

Consider a utility function,  $\mu$ , defined on a subset of  $D(R_{++}^1)$ , the set of cumulative distribution functions (cdf's)  $F$ , on the positive real line. If  $x > 0$ , then  $\delta_x$  represents the gamble in which the outcome,  $x$ , is certain;  $\delta_x(y) = 0$  if  $y < x$  and  $\delta_x(y) = 1$  if  $y \geq x$ . Similarly,  $F = \sum_{i=1}^n P_i \delta_{x_i}$  represents the gamble in which the outcome,  $x_i$ , is realized with probability,  $p_i$ ,  $i = 1, 2, \dots, n$ .

Only the ordinal properties of  $\mu$  are relevant. Without essential loss of generality, therefore, we can assume that:

$$\mu(\delta_x) = x \quad \text{for all } x > 0. \tag{2.1}$$

As a result,  $\mu$  assigns to any cdf  $F$  its certainty equivalent, i.e., that wealth level,  $x$ , such that receiving  $x$  with certainty is indifferent to the gamble represented by  $F$ . Thus, we refer to  $\mu$  as a certainty equivalent function.

Consistent with the relevant empirical literature on intertemporal asset pricing and consumption, we assume that  $\mu$  exhibits constant relative risk aversion. That is, for any random variable,  $\tilde{x}$ , with cdf  $F\tilde{x}$ :

$$\mu(F_{\lambda\tilde{x}}) = \lambda\mu(F_{\tilde{x}}) \quad \text{for all } \lambda > 0. \tag{2.2}$$

The functional forms for  $\mu$  considered in this paper are all special cases of the following, which represents the constant-relative-risk-averse members of the family of semiweighted utility functions studied by Chew (1989):

$$\int \phi(x/\mu_{sw}(F))dF(x) = 0. \tag{2.3}$$

where  $\phi$  has the form:

$$\phi(x) = \begin{cases} w^U(x)(v(x) - v(1)), & x \geq 1 \\ w^L(x)(v(x) - v(1)), & x \leq 1. \end{cases} \tag{2.4}$$

Under suitable assumptions for the valuation function,  $v$ , and for the lower and upper weight functions,  $w^L$  and  $w^U$ , Eq. (2.3) defines  $\mu(F)$  implicitly for each cdf  $F$ . Note that  $\mu$  defined in this way satisfies the certainty equivalent condition (2.1) if  $\phi(1) = 0$ .

Though the general form of Eq. (2.2) has an intuitive interpretation (see below), we find it instructive to restrict our attention initially to the parametric specializations we will use in our empirical investigation. Accordingly, restrict the valuation and weight functions to have the forms:

$$v(x) = \begin{cases} (x^\alpha - 1)/\alpha, & \alpha \neq 0. \\ \log(x), & \alpha = 0, \end{cases} \\ w^L(x) = x^\delta \quad \text{and} \quad w^U(x) = Ax^\delta. \tag{2.5}$$

Consider the further parametric restrictions:

$$0 < A \leq 1 \quad \text{and} \quad \alpha + 2\delta < 1. \tag{2.6}$$

In Appendix A, it is shown that the given Eqs. (2.4)–(2.6), we can find an interval  $[a, b]$ , depending on  $\alpha$  and  $\delta$ ,  $0 < a < 1 < b$ , such that Eq. (2.3) defines  $\mu(F)$  implicitly for all cdf's on the positive real line which have support in  $[a, b]$ . Furthermore, on this domain,  $\mu$  is strictly monotone in the sense of first-degree stochastic dominance, implies risk aversion in the sense of strict aversion to mean preserving spreads and exhibits constant relative risk aversion. Finally, under some

auxiliary assumptions,  $\mu$  is well defined and well behaved in the above sense for all cdf's having finite mean and support in the positive real line. Those assumptions are:

- (i)  $\delta = 0$  and  $\alpha < 1$ , or
- (ii)  $\delta < 0$  and  $0 < \alpha + \delta < 1$ , or
- (iii)  $0 < \delta < 1$  and  $\alpha + \delta < 0$ .

(2.7)

Turn to some special cases. For greater ease of interpretation, we express the functional forms for cdf's having finite support.

### 2.1. Expected utility ( $\delta = 0, A = 1$ )

With these parameter restrictions, we obtain the linearly homogeneous expected utility certainty equivalent:

$$\mu_{\text{EU}} \left( \sum_1^n p_i \delta_{x_i} \right) = \begin{cases} \left( \sum p_i x_i^\alpha \right)^{1/\alpha}, & \alpha \neq 0, \\ \exp \left( \sum p_i \log(x_i) \right), & \alpha = 0. \end{cases} \quad (2.8)$$

The Arrow-Pratt measure of relative risk aversion equals  $1 - \alpha$ .

### 2.2. Weighted utility ( $A = 1$ )

The constant relative risk averse form of Chew's (1983) weighted utility theory is given by:

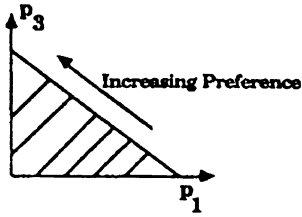
$$\mu_{\text{WU}} \left( \sum p_i \delta_{x_i} \right) = \begin{cases} \left[ \frac{\sum p_i x_i^{\delta + \alpha}}{\sum p_i x_i^\delta} \right]^{1/\alpha}, & \alpha \neq 0, \\ \exp \left[ \frac{\sum p_i x_i^\delta \log(x_i)}{\sum p_i x_i^\delta} \right], & \alpha = 0. \end{cases} \quad (2.9)$$

Note that  $\mu_{\text{WU}}$  is a single-parameter extension of Eq. (2.8). The connection between expected and weighted utility is clarified by the reference to their implied indifference curves in the probability simplex for the case of three-outcome gambles ( $n = 3$ ; see Fig. 1). It is well known that the indifference curves of  $\mu_{\text{EU}}$  are parallel straight lines. For  $\mu_{\text{WU}}$ , indifference curves remain linear; however, if  $\delta \neq 0$ , they emanate from a finite point,  $Q$ , in the plane;  $Q$  recedes to infinity as  $\delta \rightarrow 0$ .

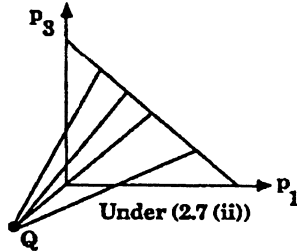
In the triangle shown, indifference curves corresponding to higher levels of utility are steeper. Such indifference curves are said to *fan out*. Machina (1982), via his Hypothesis II, points out the close connection between fanning out and the consistency with Allais-type behavior and the empirical patterns that have come to

Three Outcomes:  $x_1 < x_2 < x_3$ ; probabilities  $p_1, p_2, p_3$ ;  $p_2 = 1 - p_1 - p_3$

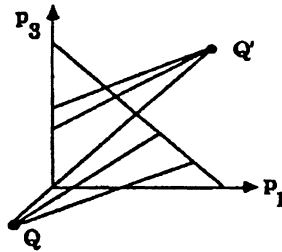
Expected Utility:



Weighted Utility: Chew (1983)



Disappointment Averse Utility: Gul (1988)



Semi-Weighted Utility: Chew (1989)

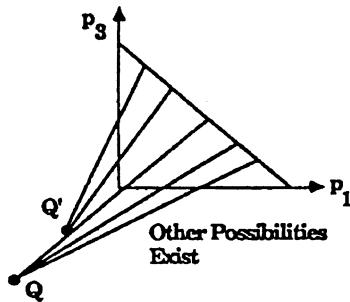


Fig. 1. Utility function for timeless wealth gambles.

be known as the common consequence effect and the common ratio effect. However, recent behavioral evidence regarding fanning out is inconclusive (see the papers by Camerer and Conlisk). Fortunately, if  $\delta > 0$ , weighted utility also admits *fanning in*, where the point  $Q$  lies to the north-east of the triangle and

higher indifference curves are flatter. In Section 4, we are able to exploit the flexibility of weighted utility to investigate whether fanning in or fanning out is indicated by financial market data.

The degree of risk aversion implicit in  $\mu_{WU}$  is of interest. It follows from Chew (1983, p. 1083) that the risk premium for a random variable with mean  $x$  and a small variance  $\sigma^2 x^2$  is approximately  $x\sigma^2 (1 - \alpha - 2\delta)/2$ . Thus,  $(1 - \alpha - 2\delta)$  is a measure of relative risk aversion for small gambles about certainty. Risk aversion increases as  $\alpha$  or  $\delta$  falls.

2.3. Disappointment aversion ( $\delta = 0$ )

If we set  $\delta = 0$  in Eq. (2.4), then Eq. (2.3) implies the following functional form, which is the contrast-relative-risk-averse specialization of the utility functions axiomatized by Gul (1991):  $\mu_{DA} (\sum P_i \delta_{x_i})$  is the unique solution to:

$$\begin{aligned} \mu_{DA}^\alpha / \alpha &= \sum p_i x_i^\alpha / \alpha + (A^{-1} - 1) \sum_{x_i < \mu_{DA}} p_i (x_i^\alpha - \mu_{DA}^\alpha) / \alpha, & \alpha \neq 0, \\ \log \mu_{DA} &= \sum p_i \log x_i + (A^{-1} - 1) \sum_{x_i < \mu_{DA}} p_i (\log x_i - \log \mu_{DA}), & \alpha = 0. \end{aligned} \tag{2.10}$$

This functional form provides an alternative single parameter extension of expected utility for which  $A = 1$ . The generalization to  $A < 1$  admits the following psychological interpretation. Refer to an outcome  $x_1$  as *disappointing* if it is worse than expected in the sense of being smaller than the certainty equivalent of the gamble according to  $\mu_{DA}$ . In Eq. (2.10), disappointing outcomes generate negative values for the second summations on the right sides,  $A^{-1} - 1 > 0$ . Thus, the certainty equivalent is smaller than it would be if  $A = 1$ , reflecting an aversion to disappointment.<sup>2</sup>

The indifference curves for  $\mu_{DA}$  in the two-dimensional simplex are shown in Fig. 1. They are linear and emanate from two distinct points  $Q$  and  $Q'$ , both of which recede to infinity as  $A \rightarrow 1$ . Thus, indifference curves *fan out* in the lower part of the triangle and otherwise *fan in*. It is interesting to note in this regard that the behavioral evidence supporting fanning out is weaker in the upper triangle than in the lower region (see Conlisk, for example.)

<sup>2</sup> There is a similarity in spirit between the structure of  $\mu_{DA}$  and the hypothesis of Kahneman and Tversky (1979) that individuals evaluate risky prospects in terms of gains and losses relative to a reference position. Here the reference position is the certainty equivalent of the gamble and the gains and losses are treated differently in computing the utility of the prospect. Since the reference point is endogenous and depends on the gamble in question, one obtains an ordering based on final wealth positions. In the cited sources, an exogenously specified reference position is taken to apply to all gambles under consideration and the preference ordering is defined on deviations from that position.

Say that individual 1 is *more risk averse* than individual 2 if any gamble that is rejected by 2 in favor of a certain prize is also rejected by individual 1. With *greater risk aversion* defined accordingly, Gul shows that risk aversion increases as  $\alpha$  or  $A$  falls. An important feature of the risk aversion embedded in  $\mu_{DA}$  is portrayed in Fig. 2, which shows the indifference curves in outcome space for binary gambles with fixed probabilities  $p_1$  and  $p_2$ . For the expected utility functions  $\mu_{EU}$ , the corresponding indifference curves are tangent at the certainty line to the actuarially fair market line of slope  $-p_1/p_2$ . Thus,  $\mu_{EU}$  is risk neutral to the first order Arrow (1965) and the risk premium for a small gamble is proportional to its variance. In contrast, the tangency fails for  $\mu_{DA}$  if  $A < 1$ , in which case the risk premium for a small binary gamble is proportional to the standard deviation and we say that  $\mu_{DA}$  exhibits first-order risk aversion. (See Segal and Spivak, 1990 for a general treatment of first-order risk aversion and Epstein and Zin, 1990 for an application to asset pricing.)

The most general specification considered in our empirical analysis is Eqs. (2.3)–(2.6). These functional forms represent a two-parameter,  $\delta \neq 0$  and  $A \neq 1$ , extension of expected utility which contains both weighted utility and disappointment-averse utility. The semiweighted certainty equivalent,  $\mu_{SW} (\sum P_1 \delta_{x_1})$ , satisfies:

$$\sum p_i x_i^\delta (x_i^\alpha - \mu_{SW}^\alpha) / \alpha + (A^{-1} - 1) \sum_{x_i < \mu_{SW}} p_i x_i^\delta (x_i^\alpha - \mu_{SW}^\alpha) / \alpha = 0,$$

$$\alpha \neq 0, \tag{2.11}$$

and similarly for  $\alpha = 0$ .

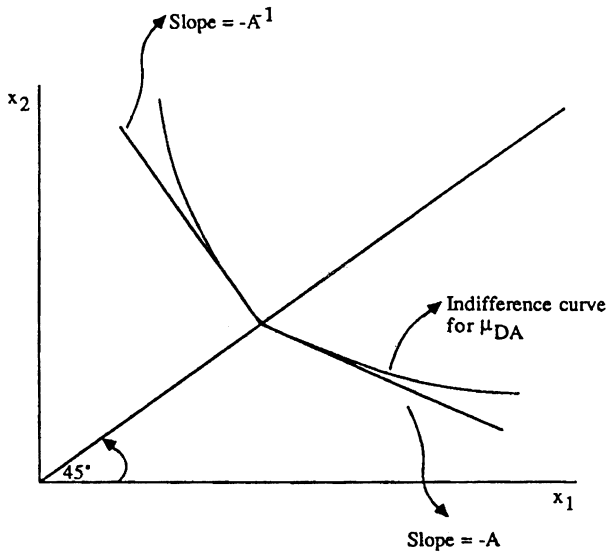


Fig. 2. Indifference Curves in State Space.



It is straightforward to provide a disappointment-aversion interpretation for  $A < 1$  and also to show that the latter implies first-order risk aversion. Risk aversion increases as  $\alpha$ ,  $\delta$  or  $A$  falls. Finally, a probability simplex indifference map for  $\mu_{\text{SW}}$  is shown in Fig. 1. (Other configurations for  $Q$  and  $Q'$  are possible though, in all cases, they are collinear with the vertex  $p_2 = 1$ ). All of the above certainty equivalents share the property that indifference curves in the three-outcome probability simplex are straight lines and more generally are hyperplanes in higher dimensional simplices. Thus, they satisfy the axiom of betweenness, which is a weakening of the independence axiom proposed and studied by Fishburn (1983), Dekel (1986) and Chew (1983, 1989).

There exist in the literature alternative generalizations of expected utility which can explain some of the accumulated behavioral evidence. Two such alternatives are prospect theory (Kahneman and Tversky, 1979) and rank-dependent or anticipated utility theory (e.g., Quiggin, 1982; Yaari, 1987; Segal, 1989). However, in each case, there are serious difficulties associated with adopting the corresponding functional forms in the analysis to follow. For example, prospect theory violates first-degree stochastic dominance unless potential violations are eliminated in a preliminary editing phase; however, a satisfactory specification of the latter is not apparent to us. On the other hand, the central role played by the rank ordering of outcomes in the structure of rank-dependent theory makes it computationally intractable in our multiple asset portfolio choice context. In contrast, there are no such difficulties associated with the betweenness-conforming utility functions adopted here. Whether alternative generalizations of the independence axiom and expected utility can help to explain the data we study remains a subject for future research.

### 3. Intertemporal asset pricing with recursive utility

The certainty equivalent functions of the last section are now integrated into an infinite-horizon, intertemporal setting. Then, we describe the restrictions implied for consumption and asset returns by the optimizing behavior of a representative agent. The reader is referred to Epstein and Zin (1989) for the details which support the discussion in this section.

There is a single consumption good in each period. In period  $t$ , current consumption,  $c_t$ , is known with certainty; however, future consumption levels are generally uncertain. Thus, intertemporal utility is defined over random consumption sequences. It is assumed that the intertemporal utility function is recursive in the sense that the utility,  $U_t$ , derived from consumption in period  $t$  and beyond, satisfies the recursive relation:

$$U_t = W(c_t, \mu_t), \quad t \geq 0, \quad (3.1)$$

where  $\mu_t = \mu(\tilde{U}_{t+1})$  is the certainty equivalent of random future utility,  $\tilde{U}_{t+1}$ , conditional upon period  $t$  information.<sup>3</sup> The function  $W$  is called an aggregator since it aggregates current consumption,  $c_t$ , with a certainty-equivalent index of the future in order to determine current utility.

We restrict  $W$  to have the CES form:

$$W(c, y) = \begin{cases} [(1 - \beta)c^\rho + \beta y^\rho]^{1/\rho}, & 0 \neq \rho < 1, \\ \exp[(1 - \beta)\log(c) + \beta\log(y)], & \rho = 0, \end{cases} \tag{3.2}$$

where  $0 < \beta < 1$ . The utility of deterministic consumption paths is given by the CES intertemporal utility function:

$$U_0 = \left[ (1 - \beta) \sum_{t=0}^{\infty} \beta^t c_t^\rho \right]^{1/\rho},$$

with the elasticity of intertemporal substitution given by  $\sigma = (1 - \rho)^{-1}$  and the constant rate of time preference given by  $\gamma = (1/\beta) - 1$ . We, therefore, interpret  $\rho$  as an intertemporal substitution parameter.

For the certainty equivalent function,  $\mu$ , we take the semiweighted form  $\mu_{SW}$ . In our earlier work, we show that  $\mu$  represents the implied preference ordering over timeless wealth gambles, i.e., gambles in which all uncertainty is resolved before further consumption takes place. Moreover, we have already noted that all of the behavioral evidence referred to above regarding individual choice under uncertainty is based on choices among timeless gambles. Thus, the preceding discussion of  $\mu_{SW}$  is pertinent. In particular, risk aversion with respect to timeless wealth gambles is inversely related to each of the parameters  $\alpha$ ,  $A$  and  $\delta$ .

The specializations of  $\mu_{SW}$  discussed above imply corresponding subclasses of recursive intertemporal utility functions shown in Fig. 3. Note that the expected utility certainty equivalent specification,  $\mu_{EU}$  (see Eq. (2.8)), does not imply the standard intertemporal utility specification:

$$U_0 = \begin{cases} \left[ (1 - \beta) E_0 \sum_{t=0}^{\infty} \beta^t c_t^\alpha \right]^{1/\alpha}, & \alpha \neq 0, \\ \exp[(1 - \beta) E_0 \sum \beta^t \log(c_t)], & \alpha = 0. \end{cases} \tag{3.3}$$

Rather,  $\mu_{EU}$  leads to the infinite-horizon generalization of the intertemporal utility function of Kreps and Porteus (1978), explored empirically in Epstein and Zin (1991). The specification Eq. (3.3) corresponds to the added restriction that  $\alpha = \rho$ , leaving only a single parameter to model both intertemporal substitution and risk aversion. Our earlier work examined the empirical gains from relaxing this

---

<sup>3</sup> Here, we write  $\mu(\tilde{U}_{t+1})$  as shorthand for  $\mu(F\tilde{U}_{t+1})$ , where  $F\tilde{U}_{t+1}$  is the cdf for  $\tilde{U}_{t+1}$  conditional upon period  $t$  information.

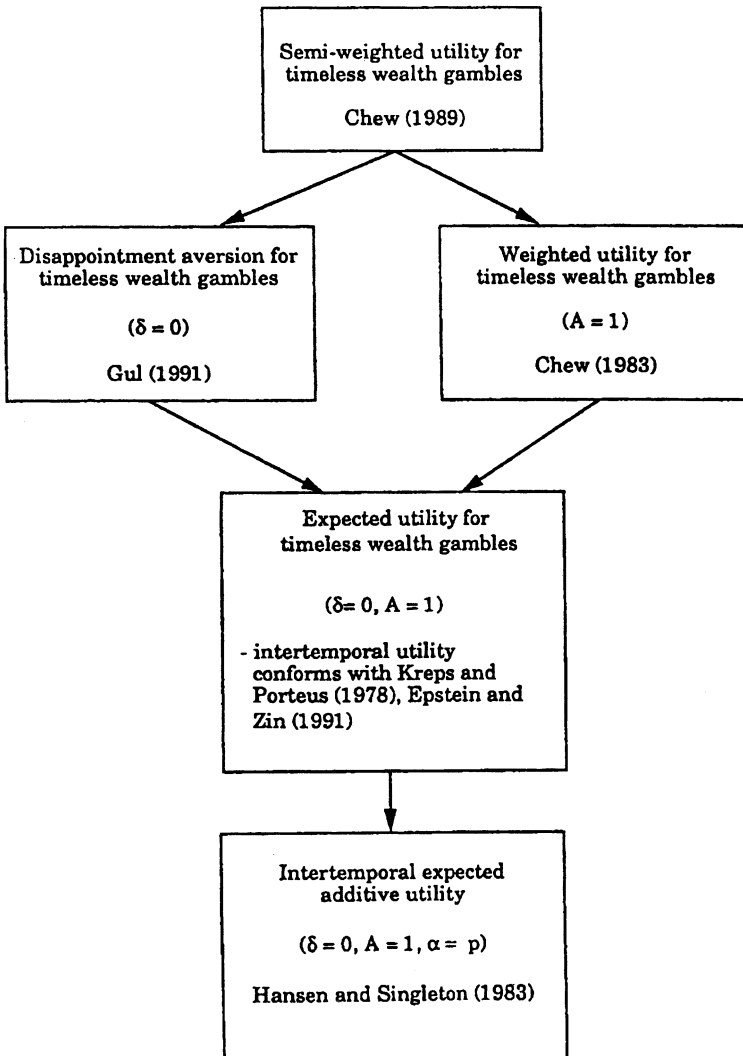


Fig. 3. Classes of recursive intertemporal utility functions.

constraint. This paper is concerned with the further gains from relaxing the parameter restrictions  $\delta = 0$  and  $A = 1$ .

From the perspective of the latter objective, it is interesting to note the following result due to Duffie and Epstein (1990). When recursive utility is suitably formulated in a continuous time setting with a Brownian information environment, weighted utility and expected utility certainty equivalents are obser-

tionally equivalent to one another, whereas (and this is strongly suggested, though not proven by their analysis) they are distinguishable from  $\mu_{DA}$ . Moreover, the essential reason for this difference between  $\mu_{DA}$  and  $\mu_{WU}$  seems to be that the former alone satisfies first-order risk aversion. Since in the present paper we are not assuming a Brownian-information, continuous time setting, we cannot rule out the potential empirical importance of the generalization from  $\mu_{EU}$  ( $\delta = 0, A = 1$ ), to  $\mu_{WU}$  ( $A = 1$ ). On the other hand, this result suggests that an empirical analysis, such as ours, should include consideration of certainty equivalents like  $\mu_{DA}$  that exhibit first-order risk aversion.

We now describe the implications for asset returns and consumption of a representative agent having recursive intertemporal utility. The agent is assumed to operate in a standard competitive environment. There are  $N$  assets and the  $i$ th asset has positive gross real return,  $\tilde{r}_{i, t+1}$  when held over the interval  $\{t, t + 1\}$ . Denote by  $\tilde{M}_{t+1}$  the return to the market portfolio over the same interval.

In Appendix A, we derive the following Euler equations which represent first-order conditions for the representative agent’s consumption/portfolio choice problem:

$$E_t \left[ h(\tilde{z}_{t+1}) I_A(\tilde{z}_{t+1}) \left( \frac{\tilde{r}_{i, t+1} - \tilde{r}_{j, t+1}}{\tilde{M}_{t+1}} \right) \right] = 0, \quad i \neq j = 1, \dots, N, \quad (3.4)$$

and

$$E_t [\phi(\tilde{z}_{t+1})] = 0, \quad (3.5)$$

where  $\phi$  is defined by Eqs. (2.4) and (2.5):

$$z_{t+1} = \beta^{1/\rho} (c_{t+1}/c_t)^{(\rho-1)/M_{t+1}^{1/\rho}}, \quad \rho \neq 0, \quad (3.6)$$

and

$$h(x) = \begin{cases} x^\delta \{ x^\alpha (1 + \delta/\alpha) - \delta/\alpha \}, & \alpha \neq 0 \\ x^\delta [1 + \delta \log(x)], & \alpha = 0. \end{cases} \quad (3.7)$$

Above,  $E_t$  is the expected value operator conditional on period  $t$  information and  $I_A$  is the indicator function:

$$I_A(x) = \begin{cases} 1, & x < 1, \\ A, & x \geq 1. \end{cases}$$

For these Euler equations to be valid in the general case where  $A \neq 1$ , we must restrict the probability distribution of consumption growth and asset returns as described in Appendix A. Sufficient conditions are that for each information set at  $t$ :

- (a) the conditional distribution of  $[r_{1, t+1}, \dots, r_{N, t+1}]$  has compact support in the positive orthant; and
- (b) the conditional distribution of the random variable  $(c_{t+1}/c_t)^{(\rho-1)/M_{t+1}^{1/\rho}}$  has a bounded density function.

Typo: M should be in the denominator, not the exponent.

According to the model of asset returns represented by Eq. (3.4), mean excess returns are determined by covariances of returns with a function of consumption growth and the return to the market. Of course, if  $A = 1$  and  $\delta = 0$ , then one obtains the model of asset returns studied in Epstein and Zin (1991), which has as special cases both the *static* CAPM when  $\alpha = 0$ , and the *consumption* CAPM when  $\alpha = \rho$ . These latter restrictions correspond to the standard expected utility model studied by Hansen and Singleton (1982, 1983).

To obtain a set of restrictions that apply directly to the levels of individual asset returns, multiply Eq. (3.4) by the portfolio weights and sum over all  $N$  securities to get:

$$E_t \left[ I_A(\tilde{z}_{t+1}) h(\tilde{z}_{t+1}) \left( \frac{\tilde{r}_{i,t+1}}{\tilde{M}_{t+1}} \right) \right] = E_t [ I_A(\tilde{z}_{t+1}) h(\tilde{z}_{t+1}) ], \quad i = 1, \dots, N. \quad (3.8)$$

Now use Eq. (3.5) to rewrite Eq. (3.8) in the form:

$$E_t \left[ I_A(\tilde{z}_{t+1}) h(\tilde{z}_{t+1}) \left( \frac{\tilde{r}_{i,t+1}}{\tilde{M}_{t+1}} \right) \right] = E_t [ I_A(\tilde{z}_{t+1}) \tilde{z}_{t+1}^\delta ], \quad (3.9)$$

which is an equation restricting the level of each return,  $i = 1, \dots, N$ .

## 4. Empirical analysis

### 4.1. The marginal rate of intertemporal substitution

The asset-pricing model derived in the previous sections has the geometric structure studied by Hansen and Richard (1987). That is, the model predicts that equilibrium asset prices are determined by a *marginal rate of intertemporal substitution*, which discounts future asset payoffs before they are averaged across states. For example, we can rewrite our model's asset-pricing equation for excess returns given in Eq. (3.4) as:

$$E_t [ \text{MRS}_{t+1,t} (\tilde{r}_{i,t+1} - \tilde{r}_{j,t+1}) ] = 0, \quad (4.1)$$

where the marginal rate of intertemporal substitution of  $c_{t+1}$  for  $c_t$  is:

$$\text{MRS}_{t+1,t} = \tilde{M}_{t+1}^{-1} h(\tilde{z}_{t+1}) I_A(\tilde{z}_{t+1}), \quad (4.2)$$

for  $z_{t+1}$  defined in Eq. (3.6). The marginal rate of intertemporal substitution for an individual asset return, rather than an excess return, is defined in Eq. (3.9) as:

$$\text{MRS}_{t+1,t}^* = \frac{h(\tilde{z}_{t+1}) I_A(\tilde{z}_{t+1}) \tilde{M}_{t+1}^{-1}}{E_t [ I_A(\tilde{z}_{t+1}) \tilde{z}_{t+1}^\delta ]}.$$

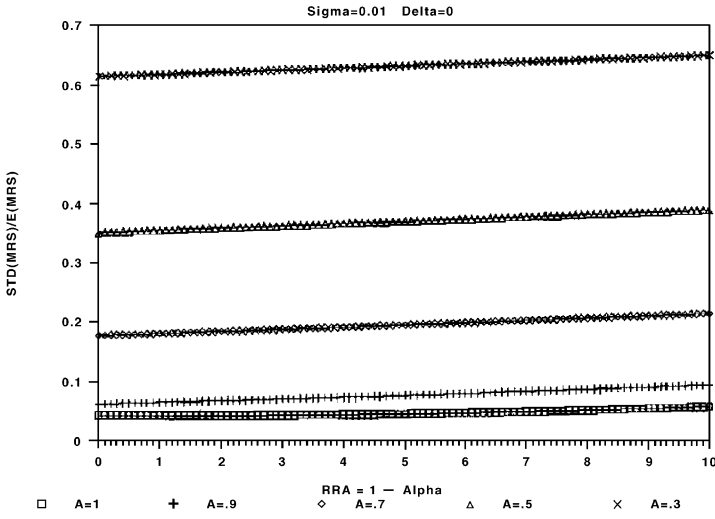


Fig. 4. Hansen-Jagannathan Statistic as RRA varies:  $\sigma = 0.01$  and  $\delta = 0$ .

We focus on the properties of MRS rather than on  $MRS^*$  since the latter involves a conditional expectation which is difficult to compute.<sup>4</sup>

Figs. 4–11 plot the ratio of the estimated standard deviation of MRS to its mean for various parameter values. Consumption is measured as monthly *per capita* expenditure in nondurables and services and the monthly NYSE value-weighted return is used to measure the return on the aggregate wealth portfolio<sup>5</sup> for the time period 1959:3 to 1986:12. Each figure has five graphs, one for each of the five choices for the first-order risk aversion parameter,  $A$ . These values range from first-order risk neutrality,  $A = 1$ , to first-order risk aversion corresponding to  $A = 0.3$ . The elasticity to intertemporal substitution,  $\sigma = (1 - \rho)^{-1}$ , varies across the figures and takes on the values  $\{0.01, 0.1, 0.5, 10\}$  to reflect varying degrees of substitutability. For Figs. 4–7, the additional risk-aversion parameter for the weighted utility specification,  $\delta$ , is held fixed at zero. Therefore, these figures pertain to the special case of disappointment aversion. The parameter  $\alpha$  is varied between  $-29$  and  $1$  so that (second-order) relative risk aversion varies from  $0$  to  $30$ . In Figs. 8–11,  $\alpha$  is fixed at  $-1$  and  $\delta$  is allowed to vary so that the measure of local (second-order) risk aversion,  $1 - \alpha - 2\delta$ , still varies from  $0$  to  $30$ . In all of the figures, the discount factor,  $\beta$ , is fixed at  $0.9975$  corresponding to a 3% annual constant rate of time preference.

<sup>4</sup> The methods of Gallant et al. (1991) could be used to evaluate  $MRS^*$  by first fitting a semi-nonparametric estimator for the joint distribution of consumption growth and asset returns.

<sup>5</sup> The consumption data (and corresponding implicit price deflator) are from Citibase and the value-weighted return is from CRSP.

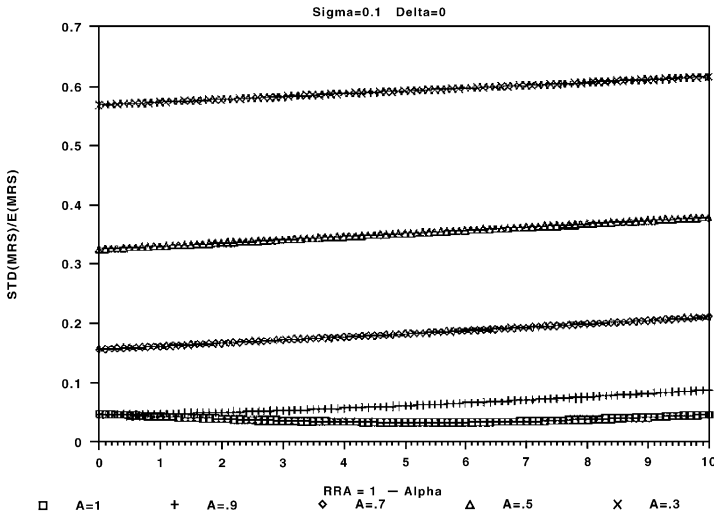


Fig. 5. Hansen-Jagannathan Statistic as RRA varies:  $\sigma = 0.1$  and  $\delta = 0$ .

Hansen and Jagannathan (1991) show that the restrictions on the covariances between MRS and asset returns, implied by equations such as Eq. (4.1), generate an inequality restriction for the ratio of the standard deviation to the mean of MRS. They estimate this bond using various stock and bond return data including two cases that correspond to that data we use in Figs. 4–11. Using the value-weighted return on the New York Stock Exchange and the return on one-month

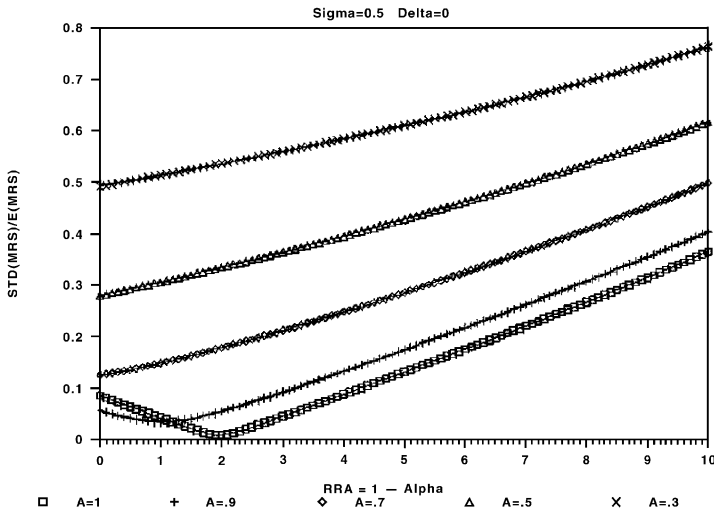


Fig. 6. Hansen-Jagannathan Statistic as RRA varies:  $\sigma = 0.5$  and  $\delta = 0$ .

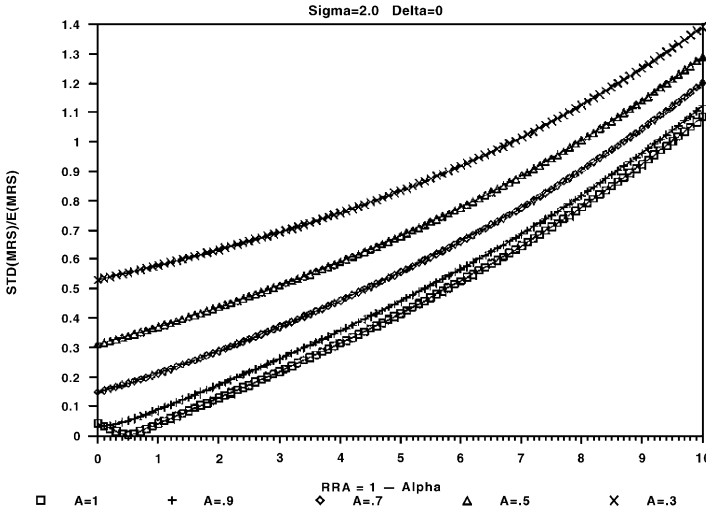


Fig. 7. Hansen-Jagannathan Statistic as RRA varies:  $\sigma = 10$  and  $\delta = 0$ .

Treasury bills,<sup>6</sup> they estimate a bound of 0.14. Using one-month holding period returns for Treasury bills with one-to-six-month maturities yield a substantially larger lower estimated bound for the standard deviation–mean ratio of 0.79. It is clear from the sizable difference in these numbers that term-structure evidence typically provides more of a challenge for asset-pricing models, e.g., Backus et al. (1989).

The first four figures allow us to identify some special cases as well as see some patterns in the behavior of the first two moments of MRS as the preference parameters change. The expected utility hypothesis is depicted by the point on each  $A = 1$  graph where  $1 - 1/\sigma = \alpha$ , i.e., the point where  $RRA = 10$  in Fig. 5,  $RRA = 2$  in Fig. 6, and  $RRA = 0.1$  in Fig. 7.<sup>7</sup> None of these points satisfies the Hansen–Jagannathan bound which is consistent with existing empirical rejections of this theory. The Kreps–Porteus model examined in Epstein and Zin (1991) is the remainder of the  $A = 1$  graph. Increasing (second-order) relative risk aversion has a small effect on the standard deviation relative to the mean when deterministic consumption is highly complementary (Figs. 4 and 5) and the Hansen–Jagannathan bounds are violated for even the most extreme cases of risk aversion. When complementarity is smaller or when deterministic consumption is substitutable (Figs. 6 and 7), increasing relative risk aversion has a much larger effect and the

<sup>6</sup> Real returns are computed using the implicit deflator for monthly consumption of nondurables and services.

<sup>7</sup> In Fig. 4, the expected utility model corresponds to a relative risk aversion parameter of 100 on the  $A = 1$  graph and is not shown.



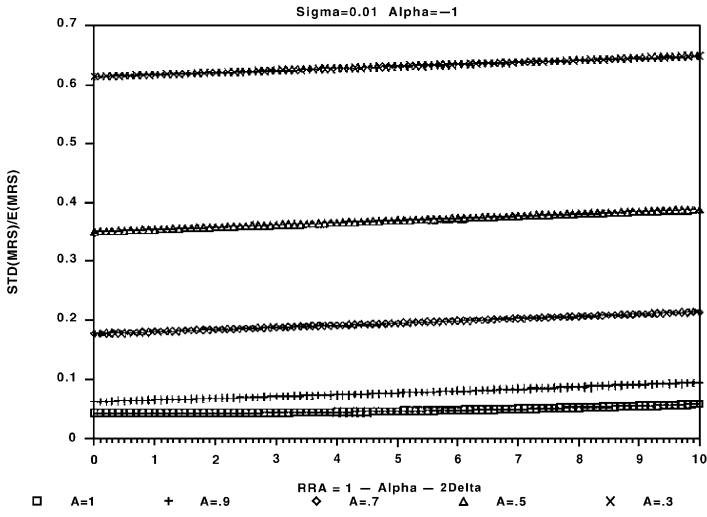


Fig. 8. Hansen-Jagannathan Statistic as RRA varies:  $\sigma = 0.01$  and  $\alpha = -1.0$ .

Hansen–Jagannathan bounds are satisfied for a range of relatively large values for RRA. Another important pattern to emerge from these figures is the impact of smaller values of  $A$  on the behavior of MRS. The standard deviation always gets larger relative to the mean as this parameter decreases. That is, the MRS gets more volatile relative to its mean and, hence, closer to the predictions of assets returns data, as  $A$  gets smaller. The Hansen–Jagannathan bounds are satisfied for a large

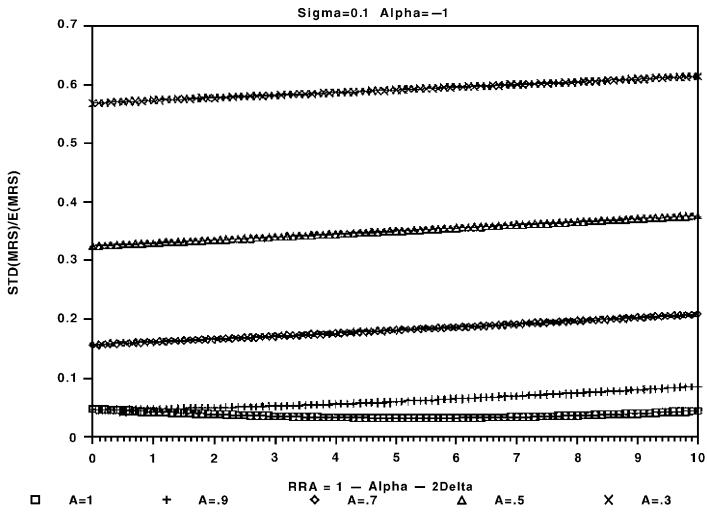


Fig. 9. Hansen-Jagannathan Statistic as RRA varies:  $\sigma = 0.1$  and  $\alpha = -1.0$ .

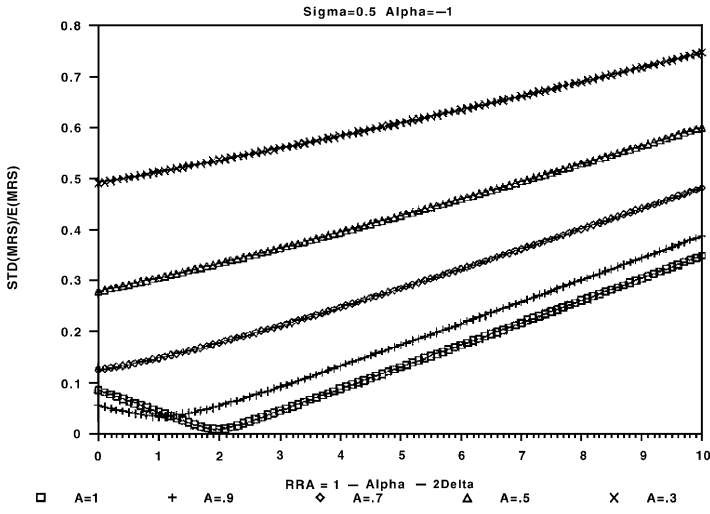


Fig. 10. Hansen-Jagannathan Statistic as RRA varies:  $\sigma = 0.5$  and  $\alpha = -1.0$ .

set of values for the parameters when  $A < 1$ . First-order risk aversion, therefore, helps move the theory closer to observed behavior. In addition, unlike second-order risk aversion, increases in first-order risk aversion continue to have a large effect on the ratio of the standard deviation to the mean as intertemporal complementarity increases.

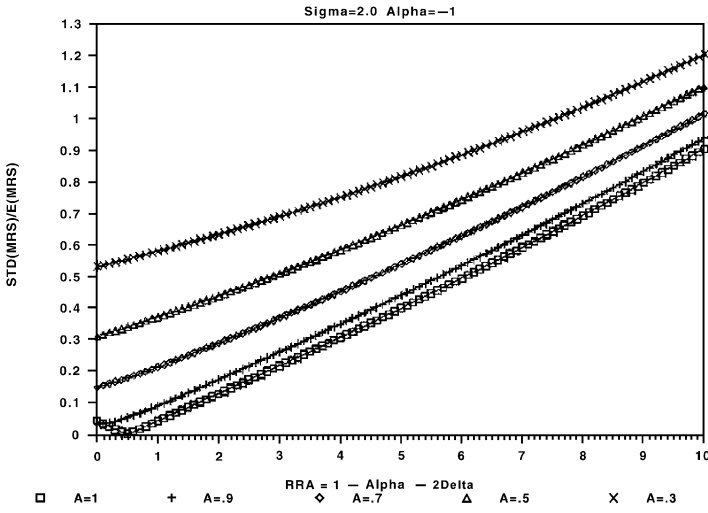


Fig. 11. Hansen-Jagannathan Statistic as RRA varies:  $\sigma = 10$  and  $\alpha = -1.0$ .

The next four figures correspond to weighted and semiweighted utility generalizations in the  $\delta$  dimension. Recall that in these figures, the parameter  $\alpha$  is constrained to equal  $-1$  and the parameter  $\delta$  varies so that  $1 - \alpha - 2\delta$  (the local measure or second-order relative risk aversion when  $A = 1$ ) varies from 0 to 30. Values along the horizontal axis greater than 2, therefore, imply fanning out, while values less than 2 imply fanning in. The most striking feature of these figures is the way they closely mimic the previous four figures in which  $\delta = 0$ . This provides some confirmation of the theoretical prediction of Duffie and Epstein (1990) regarding the observational equivalence of weighted utility and Kreps–Porteus utility.

4.2. GMM estimation

Eqs. (3.5) and (3.9) form the basis of our statistical analysis of the models discussed above. We concentrate on two assets, the value-weighted return on stocks traded on the NYSE, and the return on a one-month Treasury bill.<sup>8</sup> The existing body of empirical works suggests that the joint behavior of these two securities provides a challenge for any asset-pricing model. We treat the value-weighted return as a measure of the return on the aggregate wealth portfolio and use Eq. (3.5) to model its behavior. We restrict the T-bill return with Eq. (3.9). Rewrite ex post versions of these equations as:

$$I_A(\tilde{z}_{t+1})(\tilde{z}_{t+1}^\delta(\tilde{z}_{t+1}^\alpha - 1)/\alpha) = \tilde{\varepsilon}_{t+1}^M, \tag{4.4}$$

and

$$I_A(\tilde{z}_{t+1}) \left[ \tilde{z}_{t+1}^\delta(\tilde{z}_{t+1}^\alpha(1 + \delta/\alpha) - \delta/\alpha) \left( \frac{\tilde{r}_{t+1}^B}{\tilde{M}_{t+1}} \right) - \tilde{z}_{t+1}^\delta \right] = \tilde{\varepsilon}_{t+1}^B, \tag{4.5}$$

where  $\tilde{z}_{t+1}$  is defined in Eq. (3.6). These equations define the vector of forecast errors,  $\tilde{\varepsilon}_{t+1} = [\tilde{\varepsilon}_{t+1}^M, \tilde{\varepsilon}_{t+1}^B]^T$ , which as discussed in Section 4.1, has the property that:

$$E_t[\tilde{\varepsilon}_{t+1}] = 0. \tag{4.6}$$

Therefore, any variable that is in the agent’s information set in period  $t$  is orthogonal to the forecast error and can be used to construct an unconditional moment restriction. These unconditional moment restrictions can then be used to form a GMM estimator for the preference parameters.

We use a subset of the conditional moment restrictions implied by Eq. (4.6) summarized by:

$$E[\tilde{\varepsilon}_{t+1} | 1, \varepsilon_t, \varepsilon_{t-1}] = 0, \tag{4.7}$$

i.e., we impose restrictions on the dynamic behavior of the forecast error process. The linear restrictions implied by Eq. (4.7) can be summarized by ten uncondi-

---

<sup>8</sup> The data span the 1959:3–1986:12 time period. A complete description of the data set can be found in Epstein and Zin (1991).

tional moments, the means, the first- and second-order autocovariances, and the first and second-order cross-covariances, which must all be zero. The use of lagged forecast errors as instruments is, in part, motivated by the diagnostic testing of Euler equation residuals advocated by Singleton (1990). In addition, the fact that these restrictions retain their precise meaning for all model specifications permits a more straightforward comparison across models than does the more common practice of using arbitrary functions of ex ante information as instruments.<sup>9</sup> Tables 1–4 contain GMM parameter estimates and diagnostic tests for a variety of special cases of the general specifications in Eqs. (4.4) and (4.5). In each case, we use two different measures of consumption. Consistent estimates are obtained in a first round by minimizing the unweighted sum of squared errors from the ten estimating equations. These estimates are used to construct a consistent estimate of a weighting matrix which is used to obtain the consistent and (relatively) efficient estimates (and their standard errors) presented in the tables.<sup>10</sup> We turn now to a discussion of these results.

#### 4.2.1. Expected utility

Our starting point for GMM estimation and tests of the over-identifying restrictions implied by Eq. (4.7) is the standard time-additive expected utility model. Table 1 contains estimates and asymptotic standard errors for the rate of time preference parameter,  $\gamma = \beta^{-1} - 1$ , the elasticity of substitution/relative risk aversion parameter,  $\sigma$ , and Hansen's  $\chi^2$  test of the eight over-identifying restrictions (ten equations less two estimated parameters). This model corresponds to Eqs. (4.4) and (4.5) with the additional restrictions of  $A = 1$  (first-order risk neutrality),  $\delta = 0$  (no fanning in or out) and  $\alpha = \rho = 1 - 1/\sigma$  (substitution and risk aversion tied together).

The results are not favorable to the model. The over-identifying restrictions test indicates that this model can be rejected at almost any significance level. In

<sup>9</sup> Instrument choice also determines the relative efficiency of GMM estimators. Optimal instrument choice requires knowledge of the conditional expectation of the derivatives of the forecast errors with respect to the model's parameters. Without explicit distributional assumptions, this choice is typically infeasible. A casual argument for our choice of instruments along these lines can be made given the exponential functional forms that we are using. However, as with almost any instrument choice, this type of argument is necessarily very informal.

<sup>10</sup> The regularity conditions sufficient to yield consistent and asymptotically normal GMM estimators based on Eq. (4.7) when  $A = 1$  are covered by Hansen (1982). The fact that the instruments depend on unknown parameters is of no consequence for these properties. When  $A \neq 1$ , the functions defining the forecast errors are not differentiable; hence, the standard results do not necessarily apply. We can, however, appeal to arguments from the *empirical process* literature that hold for nondifferentiable functions to establish asymptotic normality of the GMM estimators based on the model with  $A \neq 1$ . These arguments are outlined in Appendix A3.

Table 1  
Expected utility

Parameter	Nondurables		Nondurables and services	
$\gamma$	−0.0043 (0.0022)		−0.0132 (0.0022)	
$\sigma$	0.2073 (0.0094)		0.1217 (0.0093)	
$\alpha$	{1 − 1/σ}		{1 − 1/σ}	
$\delta$	{0}		{0}	
$A$	{1}		{1}	
$\chi^2(8)$	35.91 [0.00002]		28.08 [0.0005]	

Moment	Fitted value	Weight ( $\times 10^7$ )	Fitted value	Weight ( $\times 10^7$ )
$E(\varepsilon^M)$	−0.1355 (0.0554)	0.0012	−0.0697 (0.0297)	0.0038
$E(\varepsilon^B)$	0.1285 (0.0702)	0.0001	0.1224 (0.0710)	0.0001
$E(\varepsilon^M \varepsilon_{-1}^M)$	−0.0047 (0.0011)	12.530	−0.0011 (0.0003)	125.71
$E(\varepsilon^B \varepsilon_{-1}^B)$	−0.0483 (0.0089)	0.1595	−0.0292 (0.0062)	0.1909
$E(\varepsilon^M \varepsilon_{-1}^B)$	0.0180 (0.0035)	1.4398	0.0077 (0.0017)	4.8541
$E(\varepsilon^B \varepsilon_{-1}^M)$	0.0145 (0.0031)	1.3587	0.0055 (0.0014)	5.0839
$E(\varepsilon^M \varepsilon_{-2}^M)$	−0.0003 (0.0010)	10.464	−0.0001 (0.0003)	101.83
$E(\varepsilon^B \varepsilon_{-2}^B)$	0.0022 (0.0096)	0.1413	0.0043 (0.0079)	0.1501
$E(\varepsilon^M \varepsilon_{-2}^B)$	−0.0027 (0.0029)	1.3898	−0.0004 (0.0014)	4.4613
$E(\varepsilon^B \varepsilon_{-2}^M)$	0.0026 (0.0030)	1.3898	−0.0001 (0.0015)	4.2351

Asymptotic standard errors, asymptotic p-values, and constrained parameter values are denoted by  $(\cdot)$ ,  $[\cdot]$ , and  $\{\cdot\}$ , respectively.  $\varepsilon^M$  and  $\varepsilon^B$  denote the Euler equation errors for the stock and bond returns. Estimated moments and their standard errors are multiplied by 100. “Weight” is the diagonal element of the weighting matrix used to compute the estimator.

addition, the estimated rate of time preference is negative and significantly so for one of the consumption measures.

The lower part of the table contains the estimated moments (and their asymptotic standard errors) and the diagonal elements of the weighting matrix. The former give us an additional set of tests of the restrictions of the model and the latter provide some information about how each equation is weighted in the computation of the GMM estimator. These help to identify the directions in which the model is rejected. The forecast errors exhibit significant first-order autocorrelation for both the stock and the bond return equations. Note that the moments involving the bond equation error have a corresponding diagonal element of the weighting matrix that is much smaller than do those for the market equation error. The GMM estimator, therefore, attaches relatively greater weight to the moments for the market equation which accounts for the fact that their fitted values are typically closer to zero than the others.

#### 4.2.2. Kreps–Porteus utility

By dropping the constraint that  $\alpha = \rho$ , but retaining the restrictions  $A = 1$  and  $\delta = 0$ , we obtain the Kreps–Porteus model from Epstein and Zin (1991). The

results in Table 2 differ numerically from our previous study given the different moment restrictions imposed on the estimation problem; however, they exhibit the same patterns. The substitution elasticity is small as is the coefficient of relative risk aversion. The rate of time preference is again large and negative. For nondurable consumption, the  $\chi^2$  test of the seven over-identifying restrictions has a  $p$ -value of approximately 13%, so it would not be rejected at, say, the 10% level. The nondurables and services measure provides a stronger rejection with a  $p$ -value of 1.2%.

The estimated moments for this model have substantially larger standard errors than for the expected utility model. As a result, marginal  $t$ -tests of the hypothesis that the moments are individually equal to zero do not reject. The relatively large values of the joint test statistic, therefore, come from the correlations across these moment equations. Even though these moments are not estimated very precisely, there are a number of changes in the point estimates from those obtained for the preceding model. The large negative autocorrelation in the bond equation error in the expected utility model is now a small positive correlation. In fact, almost all of the moments involving the bond error have changed sign. The positive average error for the bond equation in the expected utility model is now a large negative average error. However, the pattern of attaching relatively more weight to the

Table 2  
Kreps–porteus utility

Parameter	Nondurables	Nondurables and services
$\gamma$	-0.0077 (0.0012)	-0.0110 (0.0022)
$\sigma$	0.1550 (0.0736)	0.1416 (0.0632)
$\alpha$	0.4117 (0.2407)	1.7113 (0.6518)
$\delta$	{0}	{0}
$A$	{1}	{1}
$\chi^2(7)$	11.25 [0.128]	18.08 [0.012]

Moment	Fitted value	Weight ( $\times 10^8$ )	Fitted value	Weight ( $\times 10^8$ )
$E(\varepsilon^M)$	-0.0949 (0.1643)	0.0003	-0.0474 (0.0704)	0.0013
$E(\varepsilon^B)$	-0.2316 (0.4584)	0.00004	-0.2386 (0.4319)	0.00004
$E(\varepsilon^M \varepsilon_{-1}^M)$	-0.0038 (0.0074)	5.7739	-0.0012 (0.0016)	17.674
$E(\varepsilon^B \varepsilon_{-1}^B)$	0.0083 (0.0544)	0.1310	0.0041 (0.0538)	0.1650
$E(\varepsilon^M \varepsilon_{-1}^B)$	-0.0018 (0.0203)	0.8460	-0.0021 (0.0094)	5.0732
$E(\varepsilon^B \varepsilon_{-1}^M)$	-0.0049 (0.0198)	0.9062	-0.0039 (0.0090)	5.7790
$E(\varepsilon^M \varepsilon_{-2}^M)$	-0.0002 (0.0072)	6.7000	-0.0001 (0.0014)	17.525
$E(\varepsilon^B \varepsilon_{-2}^B)$	-0.0039 (0.0526)	0.1334	-0.0063 (0.0504)	0.1477
$E(\varepsilon^M \varepsilon_{-2}^B)$	-0.0039 (0.0195)	0.9603	-0.0014 (0.0086)	5.2289
$E(\varepsilon^B \varepsilon_{-2}^M)$	0.0013 (0.0195)	0.9324	-0.0010 (0.0085)	4.9152

See Table 1.

moments involving the stock return equation and, hence, providing a better fit for these moments, is still present.

#### 4.2.3. Weighted utility

Table 3 presents results for a model with risk preferences that exhibit first-order risk neutrality ( $A = 1$ ), but are allowed to fan out ( $\delta < 0$ ) or fan in ( $\delta > 0$ ). As one might anticipate, given the similarities of the weighted utility model and the Kreps–Porteus model found in Section 4.1, the additional parameter introduced by this specification is not well identified. The value of the objective function typically changes very little with substantial changes in the values of  $\alpha$  and  $\delta$ . The large standard errors for the estimates of these parameters id further evidence of this weak identification. The point estimates indicate fanning out for both data sets, however, the large standard errors do not allow fanning in to be rejected. Further, the Kreps–Porteus model cannot be rejected using either a  $t$ -test of  $\delta = 0$  or a likelihood-ratio-type comparison of  $\chi^2$  values. The estimate of the rate of time preference is positive; however, the estimate of the elasticity of substitution is significantly negative indicating convex utility of deterministic paths. Given these results, it is difficult to consider this parameterization to be much of an improve-

Table 3  
Weighted utility

Parameter	Nondurables		Nondurables and services	
$\gamma$	0.0072 (0.0055)		0.0067 (0.0038)	
$\sigma$	-0.3169 (0.0450)		-0.3656 (0.0680)	
$\alpha$	2.4105 (9.7503)		8.1409 (8.0639)	
$\delta$	-0.6480 (9.4916)		-4.0372 (7.1304)	
$A$	{1}		{1}	
$1 - \alpha - 2\delta$	-0.1145 (9.2563)		0.9335 (6.3145)	
$\chi^2(6)$	11.78 [0.067]		7.28 [0.296]	
$\chi^2(7) - \chi^2(6)$	0.0272 [0.869]		0.7953 [0.373]	

Moment	Fitted value	Weight ( $\times 10^7$ )	Fitted value	Weight ( $\times 10^7$ )
$E(\varepsilon^M)$	0.0145 (0.0085)	0.0010	0.0453 (0.0234)	0.0011
$E(\varepsilon^B)$	-0.2283 (0.0711)	0.0001	-0.0222 (0.0535)	0.0002
$E(\varepsilon^M \varepsilon_{-1}^M)$	0.0005 (0.0004)	7.9989	0.0013 (0.0006)	8.5998
$E(\varepsilon^B \varepsilon_{-1}^B)$	-0.0047 (0.0054)	0.0720	-0.0026 (0.0011)	0.5406
$E(\varepsilon^M \varepsilon_{-1}^B)$	-0.0019 (0.0020)	0.6876	0.0022 (0.0015)	1.6651
$E(\varepsilon^B \varepsilon_{-1}^M)$	-0.0055 (0.0015)	0.8933	-0.0000 (0.0009)	2.4646
$E(\varepsilon^M \varepsilon_{-2}^M)$	0.0006 (0.0006)	9.2146	-0.0002 (0.0006)	9.8977
$E(\varepsilon^B \varepsilon_{-2}^B)$	-0.0023 (0.0069)	0.0694	0.0019 (0.0021)	0.4633
$E(\varepsilon^M \varepsilon_{-2}^B)$	-0.0010 (0.0021)	0.8382	0.0010 (0.0013)	2.5605
$E(\varepsilon^B \varepsilon_{-2}^M)$	0.0024 (0.0021)	0.7633	0.0010 (0.0013)	1.9340

See Table 1.  $\chi^2(7) - \chi^2(6)$  is a likelihood ratio-type test of the  $\delta = 0$  restriction.

ment over the Kreps–Porteus model. In addition, our market data seem to impute little importance to the fanning pattern of indifference curves, which as noted earlier, has received considerable attention in the experimental literature.

4.2.4. Disappointment aversion

We now turn to the results for a model with first-order risk aversion. We were unable to identify the model with all five parameters ( $\gamma$ ,  $\sigma$ ,  $\alpha$ ,  $\delta$  and  $A$ ), unrestricted. This is not surprising given the nature of the results for the weighted utility model. Thus, we fix  $\delta = 0$  and estimate the remaining four parameters for the disappointment aversion model.

The evidence in Table 4 indicates that allowing for first-order risk aversion greatly improves the performance of the model on many different dimensions. The point estimates of  $A$  are roughly three standard deviations from  $A = 1$ , hence, are significantly different from first-order risk neutrality. These values for  $A$  indicate that the representative consumer attaches roughly three times more weight to disappointing outcomes than to favorable outcomes. That is, a substantial amount of first-order risk aversion appears to be necessary to rationalize observed consumption and asset returns data. The estimates of the substitution elasticity are very small—substantially smaller than in either the expected utility or Kreps–Porteus models. The rate of time preference is positive even though the substitu-

Table 4  
Disappointment aversion

Parameter	Nondurables	Nondurables and services
$\gamma$	0.0036 (0.0030)	0.0094 (0.0028)
$\sigma$	0.0048 (0.1028)	0.0032 (0.1291)
$\alpha$	-0.9763 (0.5983)	-6.4672 (1.0657)
$\delta$	{0}	{0}
$A$	0.3800 (0.1830)	0.2872 (0.1687)
$\chi^2(7)$	2.19 [0.902]	0.15 [0.999]

Moment	Fitted value	Weight ( $\times 10^7$ )	Fitted value	Weight ( $\times 10^7$ )
$E(\varepsilon^M)$	-0.1218 (0.0300)	0.0004	-0.0144 (0.0449)	0.0004
$E(\varepsilon^B)$	-0.1148 (0.0760)	0.0002	-0.0011 (0.0381)	0.0004
$E(\varepsilon^M \varepsilon_{-1}^M)$	-0.0008 (0.0019)	1.0073	-0.0001 (0.0027)	1.0137
$E(\varepsilon^B \varepsilon_{-1}^B)$	0.0013 (0.0051)	0.2282	0.0002 (0.0011)	1.9470
$E(\varepsilon^M \varepsilon_{-1}^B)$	0.0003 (0.0035)	0.4152	0.0003 (0.0021)	1.0731
$E(\varepsilon^B \varepsilon_{-1}^M)$	-0.0009 (0.0035)	0.5218	0.0001 (0.0019)	1.6679
$E(\varepsilon^M \varepsilon_{-2}^M)$	0.0002 (0.0022)	1.3457	0.0000 (0.0024)	1.3624
$E(\varepsilon^B \varepsilon_{-2}^B)$	0.0003 (0.0057)	0.2187	-0.0003 (0.0017)	1.8554
$E(\varepsilon^M \varepsilon_{-2}^B)$	-0.0005 (0.0037)	0.5823	-0.0001 (0.0022)	1.8445
$E(\varepsilon^B \varepsilon_{-2}^M)$	-0.0002 (0.0036)	0.4978	-0.0005 (0.0022)	2.4838

See Table 1.



tion elasticity is small. A negative rate of time preference is not needed in this model to match the average levels of returns. In other words, the model appears to be generating sensible results with economically meaningful values of the parameters.

The over-identifying restrictions test almost never rejects this model. The potential for singularity in a system of moment restrictions is always a concern. However, the matrix in the quadratic form used to construct the  $\chi^2$  test had a substantially larger condition number (ratio of the smallest to largest eigenvalue) in this case than for any of the other models. It is, therefore, unlikely that the risk of a singularity is greater than for any of the other models (which typically reject). The estimated moments and their standard errors indicate that this model generates moments that are much closer to their theoretical value than any of the other specifications. The pattern of fitting stock restrictions more closely than bond restrictions is no longer present. The mean of each error appears to be receiving approximately the same weight and the same is true for the autocovariances. Overall, the empirical performance of this model appears to be very good.

The degree of risk aversion implied by the point estimates for the disappointment aversion model cannot be summarized by a single number as in the expected utility model. We can, however, make some risk aversion comparisons by computing certainty for simple lotteries for different parameter values. Table 5 contains some “willingness-to-pay” calculations, i.e., the difference between the mean and the certainty equivalent, for three expected utility certainty equivalents ( $\alpha = -1, -9$  and  $-29$ ) and two disappointment aversion certainty equivalents (corresponding to the estimates from Table 4). The gamble considered is timeless and has two equally likely outcomes with an expected value of 75,000; the standard deviation of the gamble increase as one moves down any column of the

Table 5  
Some “willingness-to-pay” calculations

$\varepsilon$	Risk preferences				
	$\mu_{EU}^{\alpha}$ ( $\alpha = -1$ )	$\mu_{EU}^{\alpha}$ ( $\alpha = -9$ )	$\mu_{EU}^{\alpha}$ ( $\alpha = -29$ )	$\mu_{DA}^{A,\alpha}$ ( $A = 0.38, \alpha = -1$ )	$\mu_{DA}^{A,\alpha}$ ( $A = 0.28, \alpha = -6.5$ )
250	1	4	12	113	140
2500	83	410	1091	1189	1575
25,000	8333	21,009	23,791	17,017	23,028
40,000	21,333	37,198	39,153	31,707	38,602
50,000	33,333	47,999	49,395	42,937	49,001
60,000	48,000	58,799	59,637	55,139	59,401
74,000	73,013	73,920	73,976	73,624	73,953

Entries give the willingness to pay to avoid a gamble with equally likely outcomes  $\pm \varepsilon$ , given initial wealth equal to 75,000. Thus, for each  $\mu$  and  $\varepsilon$ , the appropriate entry is  $75,000 - \mu(\bar{x})$ , where  $\bar{x}$  equals  $75,000 \pm \varepsilon$  with probability  $1/2$ .

table. From these calculations, we can see that even for a small gamble, the expected utility model does not differ much from risk neutrality, even for a ‘large’ degree of relative risk aversion, while the disappointment aversion model displays substantial risk aversion. However, for a range of moderate to large gambles, the disappointment aversion certainty equivalent corresponding to  $A = 0.38$  and  $\alpha = -1$  exhibits less risk aversion than an expected utility certainty equivalent with a ‘moderate’ value of relative risk aversion,  $\alpha = -9$ . (See Epstein and Zin (1990) for additional comparisons of risk aversion along these lines.)

We note that what constitutes a ‘moderate’ or ‘large’ magnitude for the degree of relative risk aversion (RRA) is not at all clear. While the conventional wisdom has until recently held that only values of RRA less than 10 are reasonable (e.g., Mehra and Prescott, 1985), Kandel and Stambaugh (1990) have raised serious doubts about the validity of arguments typically invoked to support this view. The disappointment aversion model, of course, is not immune to this debate. Whether or not one should view the degree of risk aversion embodied in the results in Table 4 to be implausibly large is equally unclear. However, it is clear that since ‘large’ second-order risk aversion alone still leads to rejections of the theory, it is the ‘large’ first-order risk aversion that improves the empirical performance of the model. Although this property is structural in the context of the representative agent formulation of this paper, there is some reason to believe that it may be reflecting a feature of the market structure rather than individual preferences. For example, Epstein (1991) presents a simple example in which an exogenously given, suboptimal risk-sharing rule leads to community preferences that exhibit first-order risk aversion, even though all individuals have standard expected preferences. There is also an interesting link between the disappointment aversion model and stochastic process switching models, to which we now turn.

#### 4.3. *Switching models for consumption and MRS*

When consumption growth exhibits stochastic process switching, expected utility models generates more plausible asset-pricing predictions (e.g., Cecchetti et al., 1989; Kandel and Stambaugh, 1990). This is similar to the inclusion of so-called “disaster states” in the consumption growth process (e.g., Reitz, 1988; Backus et al., 1989). We now examine an interesting empirical similarity between these models and the disappointment aversion model.

Fig. 12 plots the ex post realizations for the log of the MRS defined by Eq. (4.2) for the disappointment aversion model using nondurable consumption and the point estimates of the parameters that correspond to this measure of consumption from Table 4.<sup>11</sup> Fig. 13 plots a histogram for this series. From these, we can

<sup>11</sup> A comparable plot nondurables and services provides very similar evidence and is not shown.

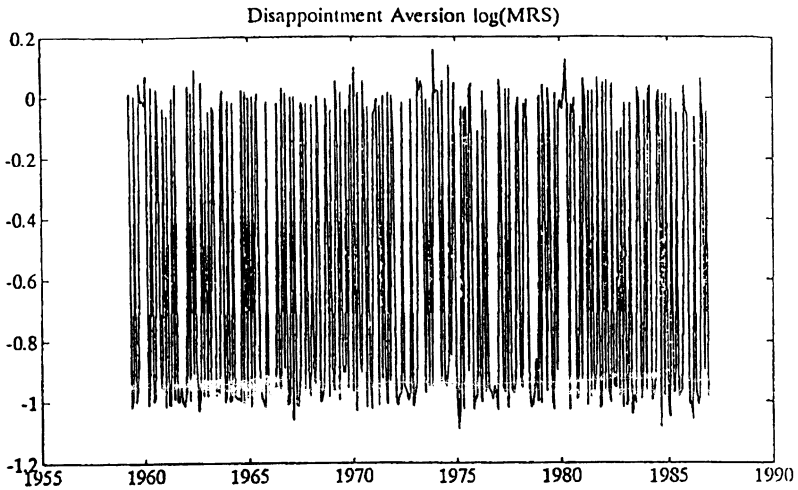


Fig. 12. Disappointment aversion log(MRS).

see two obvious properties of this MRS: it exhibits substantial negative autocorrelation and its unconditional distribution is strongly bimodal. A simple reduced form time-series representation with these two properties is a mixture of normal distributions with a two-state Markov chain generating the mixture. That is a

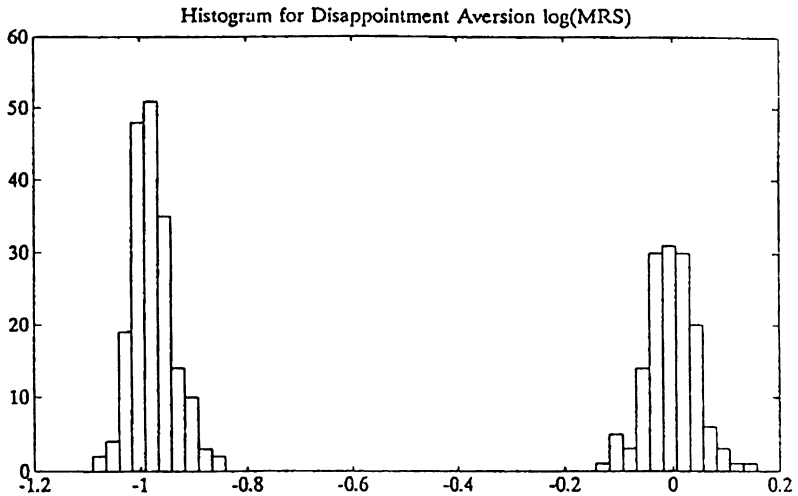


Fig. 13. Histogram for disappointment aversion log(MRS).

Table 6  
Switching models for log(MRS)

Parameter	Nondurables		Nondurables and services	
	log(MRS)	log( $c/c_{-1}$ )	log(MRS)	log( $c/c_{-1}$ )
$P$	0.4342 (0.0220)	1.0050 (0.1134)	0.3135 (0.0230)	0.9963 (0.0619)
$\theta$	-0.2068 (0.0532)	0.1444 (0.2252)	-0.0959 (0.0526)	0.1794 (0.1198)
$\mu_1$	-0.0010 (0.0039)	0.0011 (0.0004)	0.0162 (0.0052)	0.0017 (0.0003)
$\sigma_1$	0.0467 (0.0027)	0.0079 (0.0003)	0.0532 (0.0036)	0.0045 (0.0002)
$\mu_2$	-0.9763 (0.0029)	0.0014 (0.0025)	-1.2791 (0.0029)	0.0020 (0.0006)
$\sigma_2$	0.0400 (0.0021)	0.0016 (0.0050)	0.0434 (0.0020)	0.0010 (0.0003)
loglik	355.4	1144.4	344.5	1332.8
(switching)				
loglik	-231.0	1142.9	-302.9	1330.7
(i.i.d.)				

$p$  is the unconditional probability of state 1,  $\theta$  is the simple persistence parameter,  $\mu_1, \sigma_1$  are the mean and standard deviation in state  $i, i = 1, 2$ . Standard errors are in parentheses.

stochastic switching process of the form considered in the papers cited above. Maximum likelihood estimates of such a model for the log of MRS are presented in Table 6 for both consumption measures.<sup>12</sup> The parameters in this switching model are the unconditional probability of being in the high state,  $p$ , the mean and standard deviation in the low state,  $(\mu_1, \sigma_1)$ , the mean and standard deviation of the high state,  $(\mu_2, \sigma_2)$ , and the autocorrelation parameter for the Markov chain,  $\theta$ .<sup>13</sup> The negative autocorrelation and bimodality noted in Figs. 12 and 13 are also present in the maximum likelihood estimates in Table 6. There is substantial separation in the estimates of the two means. The low mean state is slightly more probable and has a larger variance than the high state. This pattern is shared by the log MRS for both consumption measures, though the autocorrelation is not as strong for the nondurables and services MRS. For comparison, the log of the likelihood function is computed for an *i.i.d.* normal model as well, which has four fewer parameters.

One could construct an expected utility model that would deliver exactly this behavior in the marginal rate of intertemporal substitution and, hence, the same predictions for asset pricing as the disappointment aversion model. Since the log of the expected utility MRS is a linear function of the log of consumption growth,

<sup>12</sup> Since the discussion in this section is at a somewhat casual level, we ignore the complications for inference that may be introduced by the fact that this series is *generated* using parameter estimates computed from the entire sample.

<sup>13</sup> Without loss of generality, the transition probabilities are parameterized as  $\text{Prob}(\text{state}_t = i | \text{state}_{t-1} = j) = p_i(1 - \theta) + \delta_{ij}\theta$ , where  $p_i$  is the unconditional probability of state 1 and  $\delta_{ij} = 1$  when  $i = j$  and 0 otherwise.

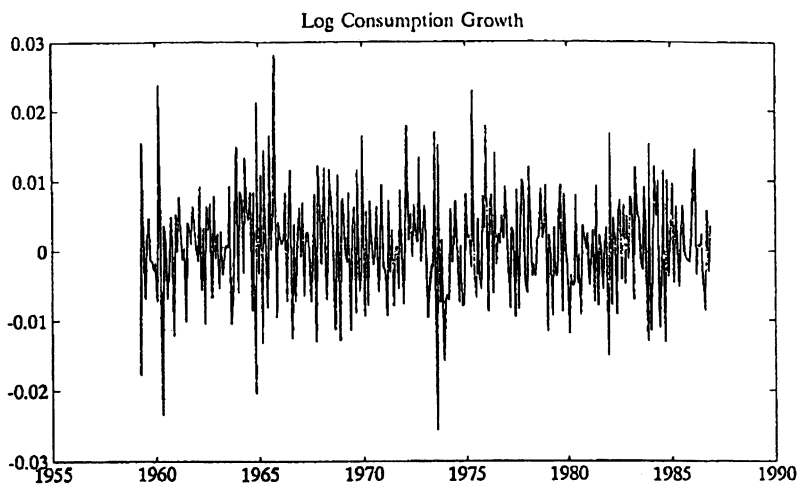


Fig. 14. Log consumption growth.

applying the inverse of this linear function to the log of the disappointment aversion MRS produces a consumption growth process that generates an observationally equivalent expected utility MRS. For example, postulating a switching process for the log of nondurable consumption growth with  $p = 0.4342$ ,  $\theta = -0.2068$ ,  $\mu_1 = 0.0026$ ,  $\mu_2 = 0.1110$ ,<sup>14</sup>  $\sigma_1 = 0.0052$  and  $\sigma_2 = 0.0044$  generates an expected utility marginal rate of substitution process with  $\beta = 0.9975$  and  $\alpha = -9$ , which exhibits exactly the same behavior as the disappointment aversion marginal rate of substitution process in the first column of Table 6.

For comparison, Figs. 14 and 15 plot the log of observed monthly nondurable consumption growth and its histogram. The log of any expected utility MRS is a linear function of this process. The second and fourth columns of Table 6 estimate the parameters of the switching model that fits observed consumption growth. From the figures and the parameter estimates, it is clear that observed consumption growth does not exhibit the persistence and bimodality properties that could generate an expected utility MRS that is similar to the disappointment aversion MRS. The evidence from these data for the type of stochastic switching process for consumption growth used in previous expected utility studies, e.g., Cecchetti et al. (1989), is quite weak. However, we have shown that the marginal rate of intertemporal substitution can exhibit switching-type behavior even if consumption

<sup>14</sup> Note that the low state for MRS translates into the high state for consumption growth.

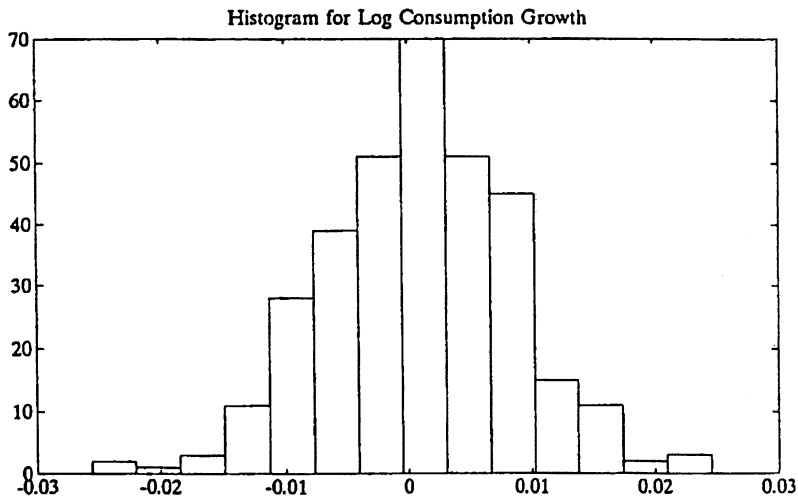


Fig. 15. Histogram for log consumption growth.

growth rates do not. The requisite persistence and bimodality of the MRS can be generated instead by the specification of disappointment averse utility.

## 5. Conclusion

We have shown that some recent developments in the theory of choice under uncertainty due to Chew (1983, 1989) and Gul (1991), which relax the independence axiom of expected utility, can be incorporated into a recursive asset-pricing model without sacrificing either theoretical or empirical tractability. We are able to provide the first tests of these new theories using actual market data rather than laboratory experiments. Our evidence indicates that the generalization to weighted utility does not enhance the explanatory power of our asset-pricing model for the data we consider. On the other hand, we find that using preferences that exhibit first-order risk aversion provides a substantial improvement in the empirical performance of a representative agent, intertemporal asset-pricing model relative to models that maintain only second-order risk aversion. A model in which the representative agent's preferences exhibit positive time preference, low substitutability in deterministic consumption and nonnegligible aversion to small risks, is very difficult to reject. Future research will examine the robustness of these results by considering other securities such as individual stock returns and multiperiod bond returns.

## Acknowledgements

The first author gratefully acknowledges the financial support of the Social Sciences and Humanities Research Council of Canada. We are also indebted to Chew Soo Hong, Robert Miller and Fallaw Sowell for valuable discussions. This paper is part of NBER's research program in Financial Markets and Monetary Economics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

## Appendix A

### A.1. Properties of $\mu_{sw}$

Recall the definition Eq. (2.3) of  $\mu_{sw}$  where  $\phi$  is given by Eqs. (2.4) and (2.5). Note that  $\phi'(1) > 0$  and  $\phi''(1) < 0$ . Thus,  $\phi$  is strictly increasing and concave on an interval  $[a, b]$  containing 1. Note also that  $\phi(1) = 0$ . (Under Eq. (2.7), the preceding is true for *any* positive numbers  $a$  and  $b$ .)

Let  $F$  have support in  $[a, \infty]$ . Then:

$$\int \phi(x/a) dF(x) \geq \int \phi(1) dF(x) = 0.$$

On the other hand, if  $F$  has finite mean  $m$ , then by Jensen's inequality:

$$\int \phi(x/m) dF(x) \leq \phi(1) = 0.$$

Thus, there exists a (unique)  $\lambda$  such that  $\int \phi(x/\lambda) dF(x) = 0$ . We define  $\mu_{sw}(F) = \lambda$ .

Properties (2.1) and (2.2) are immediate. Consistency with first-degree stochastic dominance and risk aversion follows from the monotonicity and concavity of  $\phi$ . For example, if  $G$  is a mean preserving spread of  $F$ , then:

$$\begin{aligned} & \int \phi(x/\mu_{sw}(F)) dF(x) \\ &= 0 = \int \phi(x/\mu_{sw}(G)) dG(x) \leq \int \phi(x/\mu_{sw}(G)) dF(x) \\ &\Rightarrow \mu_{sw}(F) \geq \mu_{sw}(G). \end{aligned}$$

### A.2. Derivation of the Euler equations

Turn to the derivation of the Euler Eq. (3.4). It follows from Eq. (2.3) and Epstein and Zin (1989, pp. 957–958) that the optimal portfolio share vector  $\omega_t^*$  solves:

$$\max_{\omega} E_t \phi \left[ \beta^{1/\rho} \tilde{M}_{t+1}^{(1-\rho)/\rho} (\tilde{c}_{t+1}/c_t)^{(\rho-1)/\rho} \omega^T \tilde{\mathbf{r}}_{t+1} \right], \quad (\text{A.1})$$

where  $\tilde{r}_{t+1}$  is the vector of gross real returns,  $\tilde{M}_{t+1} = \omega_t^* T \tilde{r}_{t+1}$ , the return to the market portfolio, and where  $\omega = (\omega_1, \dots, \omega_N)$  varies over the simplex  $\{\omega: \omega_i \geq 0 \text{ for all } i \text{ and } \sum \omega_i = 1\}$ . We will show that, under additional assumptions, the objective function in Eq. (A.1) is differentiable at  $\omega_t^*$  and the associated first-order conditions are given by Eq. (3.4). The existence of an interior optimum is assumed.

Recall the definition Eq. (3.6) of  $z_{t+1}$  and define:

$$g(x) = \begin{cases} x^\delta (x^\alpha - 1) / \alpha, & \alpha \neq 0, \\ x^\delta \log(x), & \alpha = 0. \end{cases}$$

The optimization problem (Eq. (A.1)) can be written:

$$\begin{aligned} \max_{\omega} \int g(z_{t+1} M_{t+1}^{-1} \omega^T r_{t+1}) dF_t \\ + (A - 1) \int_{\{z_{t+1} M_{t+1}^{-1} \omega^T r_{t+1} \geq 1\}} g(z_{t+1} M_{t+1}^{-1} \omega^T r_{t+1}) dF_t, \end{aligned}$$

where  $F_t$  is the appropriate conditional cdf. Since  $g$  is differentiable, the only potential difficulty arises because the domain of the second integral depends upon  $\omega$  and because  $A \neq 1$ , in general.

Define:

$$\psi(\omega) = \int_{\{z_{t+1} M_{t+1}^{-1} \omega^T r_{t+1} \geq 1\}} g(z_{t+1}) dF_t. \tag{A.2}$$

When there is positive probability associated with  $z_{t+1} = 1$ ,  $\psi$  can fail to be differentiable at  $\omega = \omega_t^*$ . However, under Assumptions 1 and 2,  $\psi$  is differentiable there and:

$$\psi_i(\omega_t^*) - \psi_j(\omega_t^*) = 0, \quad i, j = 1, \dots, N. \tag{A.3}$$

It then follows that Eq. (3.4) represents the first-order conditions for Eq. (A.1).

**Assumption 1.** For each information set at time  $t$ , the conditional distribution of  $r_{t+1}$  has compact support in the positive orthant.

**Assumption 2.** For each information set at  $t$ , there exists an  $\bar{\varepsilon} > 0$  such that:

$$\sup_{0 < \varepsilon < \bar{\varepsilon}} \varepsilon^{-1} \text{Prob}_t(\{\tilde{z}_{t+1} : 1 - \varepsilon < \tilde{z}_{t+1} < 1 + \varepsilon\}) < \infty.$$

A necessary condition for Assumption 2 is that the conditional probability that  $\tilde{z}_{t+1} = 1$  be zero. A sufficient condition is that  $\tilde{z}_{t+1}$  have bounded conditional density function.



One can apply elementary arguments and the fact that  $g(1) = 0$  to the difference quotient corresponding to Eq. (A.3) to prove the latter. Details are omitted.

### A.3. Asymptotic properties of the GMM estimator

Turn now to the asymptotic properties of the GMM estimator when  $A \neq 1$ . In this case, the objective function is not differentiable. We, therefore, employ the results in Andrews (1989a,b) to show how the arguments in Hansen (1982) can be modified to deal with this nonstandard situation. The application of these results to our problem is straightforward so we simply outline the steps involved in deriving asymptotic properties for the GMM estimator. Note that unlike Andrews (1989a,b), we do not have the added complication of infinite dimensional nuisance parameter estimation; all of the parameters in our model are finite dimensional.

For simplicity, consider first the just-identified case. Let  $fT(\mathbf{b}) \equiv T^{-1} \sum_{t=1}^T f(\mathbf{x}_t, \mathbf{b})$  be the sample analogue (for a sample of size  $T$ ) of the population orthogonality restrictions based on Eq. (4.7), where  $\mathbf{x}_t$  is the vector of variables and  $\mathbf{b}$  is the vector of parameters. The GMM estimator,  $b_T$ , is defined such that  $f_T(b_T) = 0$  and the true value of the parameter vector,  $\mathbf{b}_0$ , is such that  $Ef_T(\mathbf{b}_0) = 0$ . If a uniform law of large numbers holds for  $f_T(\mathbf{b})$  and if  $\mathbf{b}_0$  is identified, then a standard proof (e.g., Hansen, 1982; or Andrews, 1989a, Theorem 1.1) of the consistency of  $b_T$  can be constructed when  $Ef_T(\mathbf{b})$  is continuous. The continuity of this expectation follows from the discussion in Appendix A2 and will be assumed from here on.

Since the functions defined in Eqs. (4.4) and (4.5) are not differentiable in the parameters of the model, the mean value expansion of  $f_T(b_T)$  at  $\mathbf{b}_0$ , which is typically used to establish asymptotic normality for the GMM estimator, does not necessarily exist. We can, however, expand  $Ef_T(\mathbf{b}_0)$  at  $b_T$  since this function is differentiable. This expansion has the form:

$$Ef_T(\mathbf{b}_0) = Ef_T(b_T) - \left[ \frac{\partial Ef_T(\bar{b}_T)}{\partial \mathbf{b}} \right] (b_T - \mathbf{b}_0).$$

where  $\bar{b}_T$  is between  $b_T$  and  $\mathbf{b}_0$ . This implies that  $T^{1/2} (b_T - \mathbf{b}_0)$  is asymptotically normal distributed if  $T^{1/2} Ef_T(b_T)$  has this property, since  $Ef_T(\mathbf{b}_0)$  equals 0 by definition. Note the difference in this result from the standard argument which says that  $T^{1/2} (b_T - \mathbf{b}_0)$  is asymptotically normally distributed if  $T^{1/2} f_T(\mathbf{b}_0)$  has this property.

Define the empirical process,  $v_T(\mathbf{b}) \equiv T^{1/2} [f_T(\mathbf{b}) - Ef_T(\mathbf{b})]$ . If this process has the property of *stochastic equicontinuity* (as defined in Andrews, 1989b) at the  $\mathbf{b}_0$ , and if  $b_T$  is consistent for  $\mathbf{b}_0$ , then  $v_T(b_T)$  is asymptotically equivalent to  $v_T(\mathbf{b}_0)$ . This in turn implies that  $T^{1/2} Ef_T(b_T)$  is asymptotically equivalent to  $-T^{1/2} f_T(\mathbf{b}_0)$ , since both  $Ef_T(\mathbf{b}_0)$  and  $f_T(b_T)$  are zero by definition. When a

central limit theorem holds for  $T^{1/2} f_T(\mathbf{b}_0)$  as in the standard case, it follows that  $T^{1/2} f_T(b_T - \mathbf{b}_0)$  is asymptotically normally distributed with covariance matrix:

$$\left[ \frac{\partial E f_T(\mathbf{b}_0)}{\partial \mathbf{b}} \right]^{-1} \text{Var} [T^{1/2} f_T(\mathbf{b}_0)] \left[ \frac{\partial E f_T(\mathbf{b}_0)}{\partial \mathbf{b}} \right]^{-1}.$$

An estimator of the matrix of derivatives in this expression can be constructed using a finite difference of the sample average (as in Pakes and Pollard, 1989) evaluated at a consistent estimator for  $\mathbf{b}_0$ , provided the finite difference converges to zero at an appropriate rate as the sample size gets large. The variance of the orthogonality conditions in the middle of this expression can be consistently estimated in the standard way by using the sample average of  $f(\mathbf{x}_t, \mathbf{b})f(\mathbf{x}_t, \mathbf{b})^T$ , evaluated at a consistent estimator for  $\mathbf{b}_0$ .

When the model is over-identified,  $b_T$  is defined as the minimizer of the quadratic form  $f_T(\mathbf{b})^T \mathbf{W}_T f_T(\mathbf{b})$ , where  $\mathbf{W}_T$  is a positive definite matrix that can depend on the sample size and has a probability limit of  $W$ . As in the just-identified case, asymptotic normality follows from an expansion of  $E f_T(\mathbf{b}_0)$  around  $b_T$ . Premultiply both sides of this expansion by  $[(\partial E f_T(\mathbf{b}_0))/\partial \mathbf{b}]^T W$ . It follows that when  $[(\partial E f_T(\mathbf{b}_0))/\partial \mathbf{b}]^T W T^{1/2} f_T(b_T)$  is  $o_p(1)$  and the stochastic equicontinuity condition holds,  $T^{1/2}(b_T - \mathbf{b}_0)$  is asymptotically normally distributed with covariance matrix:

$$(D_0^T W D_0)^{-1} D_0^T W (\text{Var} [T^{1/2} f_T(\mathbf{b}_0)]) W D_0 (D_0^T W D_0)^{-1},$$

where  $D_0 = [(\partial E f_T(\mathbf{b}_0))/\partial \mathbf{b}]$ . When  $W = V^{-1}$  where  $V$  is equal to  $\text{Var}[T^{1/2} f_T \mathbf{b}_0]^{-1}$ , the asymptotic covariance matrix of the GMM estimator is  $(D_0^T V^{-1} D_0)^{-1}$ . This defines an efficient estimator in the sense that this covariance matrix differs from the covariance matrix of an estimator defined for an arbitrary choice of weighting matrix,  $\mathbf{W}_T$ , by a positive semidefinite matrix. Further, for this efficient estimator,  $T f_T(b_T)^T V^{-1} f_T(b_T)$  has a limiting  $\chi^2$  distribution with degrees of freedom equal to the number of restrictions less the number of estimated parameters. Note that this result differs from the just-identified case by requiring that  $[(\partial E f_T(\mathbf{b}_0))/\partial \mathbf{b}]^T W T^{1/2} f_T(b_T)$  converge to zero in probability as the sample size gets arbitrarily large. When the objective function is differentiable, the analogous condition to this follows from the necessary conditions for the optimization problem that defines the estimator. Alternatively, one can think of the over-identified model as corresponding to a particular just-identified model (as in Hansen, 1982), by defining it as the solution to  $a_T f_T(b_T) = 0$ , where  $a_T$  has column dimension equal to the number of orthogonality restrictions and row dimension equal to the number of parameters, i.e., the estimator sets a particular linear combination of the orthogonality restrictions equal to zero. In the differentiable case,  $a_T$  can be thought of as defining the first-order conditions. In the nondifferentiable case, when  $a_T$  converges to  $[(\partial E f_T(\mathbf{b}_0))/\partial \mathbf{b}]^T W$ , the asymptotic results discussed above will hold without any additional assumptions. Therefore,

constructing a just-identified GMM estimator for the linear combinations of the orthogonality restrictions given by a consistent estimator of  $[(\partial Ef_T(\mathbf{b}_0))/\partial \mathbf{b}]^T W$  results in an estimator that is asymptotically equivalent to the estimator based on the minimization of the quadratic form and the requirement that  $[(\partial Ef_T(\mathbf{b}_0))/\partial \mathbf{b}]^T W T^{1/2} f_T(\mathbf{b}_T) = o_p(1)$ .

The results discussed above rely on the stochastic equicontinuity of the empirical process. Sufficient conditions for this in a time-series context that can be applied to our model are given in Andrews (1989b). Note that the functions that we deal with involve his type I and type III functions (indicator functions and power functions).

## References

- Andrews, D.W.K., 1989. Asymptotics for semiparametric econometric models: Estimation, Cowles discussion paper no. 908, Yale University.
- Andrews, D.W.K., 1989. Asymptotics for semiparametric econometric models: II. Stochastic equicontinuity, Cowles discussion paper no. 909, Yale University.
- Arrow, K.J., 1965. The theory of risk aversion. *Aspects of the Theory of Risk Bearing*. Yrjo Jahnsen Saatio, Helsinki, Chap. 2.
- Backus, D.K., 1989. Risk premiums in the term structure: evidence from artificial economies. *Journal of Monetary Economics* 24, 371–399.
- Camerer, C.F., 1989a. An experimental test of several generalized utility theories. *Journal of Risk and Uncertainty* 2, 61–104.
- Camerer, C.F., 1989b. Recent of generalizations of expected utility theory, mimeo. University of Pennsylvania, Philadelphia.
- Cecchetti, S.G., Lam, P., Mark, N.C., 1989. The equity premium and the risk free rate: matching the moments, mimeo. Ohio State University, Columbus.
- Chew, S.H., 1983. A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the Allais paradox. *Econometrica* 51, 1065–1092.
- Chew, S.H., 1989. Axiomatic utility theories with the betweenness property. *Annals of Operations Research* 19, 273–298.
- Chew, S.H., Waller, W.S., 1986. Empirical tests of weighted utility theory. *Journal of Mathematical Psychology* 30, 55–72.
- Conlisk, J., 1989. Three variants on the Allais example. *American Economic Review* 79, 392–407.
- Duffie, D., Epstein, L.G., 1990. Stochastic differential utility and asset pricing, mimeo. Stanford University, Stanford.
- Epstein, L.G., 1991. Behaviour under risk: recent development in theory and applications. In: Laffont, J.J. (Ed.), *Forthcoming in Advances in Economic Theory*. Cambridge Univ. Press, Cambridge.
- Epstein, L.G., Zin, S.E., 1989. Substitution, risk aversion and the temporal behavior of consumption and asset returns: a theoretical framework. *Econometrica* 57, 937–969.
- Epstein, L.G., Zin, S.E., 1990. ‘First-order’ risk aversion and the equity premium puzzle. *Journal of Monetary Economics* 26, 387–407.
- Epstein, L.G., Zin, S.E., 1991. Substitution, risk aversion and the temporal behavior of consumption and asset returns: an empirical analysis. *Journal of Political Economy* 99, 263–286.
- Fishburn, P.C., 1983. Transitive measurable utility. *Journal of Economic Theory* 31, 293–317.
- Gallant, R., Hansen, L.P., Tauchen, G., 1991. Using conditional moments of asset payoffs to infer the volatility of intertemporal marginal rates of substitution. *Journal of Econometrics* 45, 141–180.

- Gul, F., 1991. A theory of disappointment aversion. *Econometrica* 59 (3), 667–689.
- Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.
- Hansen, L.P., Jagannathan, R., 1991. Implications of security market data for models of dynamic economies. *Journal of Political Economy* 99, 225–262.
- Hansen, L.P., Richard, S., 1987. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica* 55, 587–613.
- Hansen, L.P., Singleton, K.J., 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica* 50, 1269–1286.
- Hansen, L.P., Singleton, K.J., 1983. Stochastic consumption, risk aversion and the temporal behavior of asset returns. *Journal of Political Economy* 91, 249–265.
- Harrison, G., 1990. Expected utility theory and the experimentalists, working paper B-90-04. University of South Carolina.
- Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of choice under risk. *Econometrica* 47, 263–291.
- Kandel, S., Stambaugh, R.F., 1990. Modelling expected stock returns for long and short horizons, mimeo. University of Pennsylvania, Philadelphia.
- Kandel, S., Stambaugh, R.F., 1990. Asset returns and intertemporal preferences, NBER working paper no. 3633.
- Kreps, D.M., Porteus, E.L., 1978. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica* 46, 185–200.
- Machina, M.J., 1982. ‘Expected utility’ analysis without the independence axiom. *Econometrica* 50, 277–323.
- Mehra, R., Prescott, E., 1985. The equity premium: a puzzle. *Journal of Monetary Economics* 15, 145–161.
- Pakes, A., Pollard, D., 1989. Simulation and asymptotics of optimization. *Econometrica* 57, 1027–1057.
- Quiggin, J.C., 1982. A theory of anticipated utility. *Journal of Economic Behavior and Organization* 3, 323–343.
- Reitz, T., 1988. The equity premium: a solution. *Journal of Monetary Economics* 22, 117–131.
- Segal, U., 1989. Axiomatic representation of expected utility with rank-dependent probabilities. *Annals of Operations Research* 19, 359–373.
- Segal, U., Spivak, A., 1990. ‘First-order’ versus ‘second-order’ risk aversion. *Journal of Economic Theory* 51, 111–125.
- Singleton, K.J., 1987. Specification and estimation of intertemporal asset pricing models. In: Friedman, B., Hahn, F. (Eds.), *Handbook in Monetary Economics*.
- Yaari, M.E., 1987. The dual theory of choice under risk. *Econometrica* 55, 95–115.