

NBER WORKING PAPER SERIES

WHY STOCKS MAY DISAPPOINT

Andrew Ang
Geert Bekaert
Jun Liu

Working Paper 7783
<http://www.nber.org/papers/w7783>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2000

We have benefitted from the comments of Shlomo Benartzi, Arjan Berkelaar, John Heaton, Roy Kouwenberg, Deborah Lucas, Anthony Lynch, Michael Stutzer and seminar participants at Columbia University, New York University, and the Society of Financial Studies Conference on Market Frictions and Behavioral Finance at Northwestern University. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

© 2000 by Andrew Ang, Geert Bekaert, and Jun Liu. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Why Stocks May Disappoint
Andrew Ang, Geert Bekaert, and Jun Liu
NBER Working Paper No. 7783
July 2000

ABSTRACT

Recently much progress has been made in developing optimal portfolio choice models accomodating time-varying opportunity sets, but unless investors are unreasonably risk averse, optimal holdings include unreasonably large equity positions. One reason is that most studies assume investors behave as expected utility maximizers with power utility. In this article, we provide a formal treatment of both static and dynamic portfolio choice using the Disappointment Aversion preferences of Gul (1991). While different from the Kahneman-Tversky (1979) loss aversion utility, these preferences imply asymmetric aversion to gains versus losses and are consistent with the tendency of some people to like lottery-type gambles but dislike stock in-vestments. By calibrating a number of data generating processes to actual US data on stock and bond returns, we find very reasonable portfolios for moderately disappointment averse investors with utility functions exhibiting low curvature. Disappointment aversion preferences affect intertemporal hedging demands and the state dependence of asset allocation in such a way as to not be replicable by standard expected utility functions with higher curvature. Furthermore, it is easy to reconcile the large equity premium observed in the data with disappointment aversion utility of low curvature and reasonable disappointment aversion.

Andrew Ang
Graduate School of Business
Columbia University
3022 Broadway/805 Uris Hall
New York, NY 10027
aa610@columbia.edu

Geert Bekaert
Graduate School of Business
Columbia University
3022 Broadway/802 Uris Hall
New York, NY 10027
and NBER
gb241@columbia.edu

Jun Liu
Anderson School of Business
UCLA
405 Hilgard Avenue
Los Angeles, CA 90095
jliu@anderson.ucla.edu

1 Introduction

Several authors demonstrate that only a minority of American households actually hold stocks.¹ Overall, the portfolios of US investors including pension funds and endowments are fairly balanced across stocks and bonds/cash, most estimates yielding a 60%-40% mix of equity and bonds. Given that we observe a large equity premium in US data (Mehra and Prescott (1985)), standard portfolio choice models predict large equity positions for most investors. Recently, much progress has been made in developing optimal portfolio choice models under more realistic data generating process (DGP's) for returns (see Campbell and Viceira (1999, 1998) and Liu (1999)), but unless investors are unreasonably risk-averse, optimal holdings under these DGP's continue to include large equity positions. For example, in Campbell and Viceira (1998) an investor with risk aversion equal to 2, the value often considered to be normal, actually invests over 200% of her wealth in the stock market at unconditional mean levels of stock returns and dividend yields.

Benartzi and Thaler (1995) provide an interesting perspective on these observations. They argue that investors display “myopic loss aversion”, and that this explains both the observed portfolio holdings and the large equity premium. They model an aversion to loss using the framework of Kahneman and Tversky (1979) where the utility function is defined asymmetrically over gains and losses. They add the feature that investors evaluate gains and losses frequently, even though their investment objectives are long-term. Put together they find that most investors are largely indifferent between bonds and stocks even in the presence of the large equity premium.

In this article, we provide a formal treatment of portfolio choice in the presence of loss aversion, but rather than relying on Kahneman and Tversky (1979)'s prospect theory, we use the Disappointment Aversion (DA) framework of Gul (1991). These preferences are a one parameter extension of standard iso-elastic preferences in the usual expected utility framework and have the characteristic that good outcomes - outcomes above the certainty equivalent - are downweighted relative to bad outcomes. The larger weight given to outcomes which are bad in a relative sense gives rise to the name “disappointment-averse” preferences, but as we show they imply a sharp aversion to losses.

DA utility displays first order risk aversion, where the risk premium, the amount that makes an investor indifferent between the status quo and accepting a lottery, is proportional to volatility. In contrast, with expected utility the risk premium is proportional to variance. This feature helps DA utility to account for the phenomenon that individuals are risk averse with respect to

¹See Heaton and Lucas (1999, 1996), Vissing-Jørgensen (1997) and Mankiw and Zeldes (1991).

gambles which yield a large loss with small probability (for example a stock investment) but risk-loving with respect to gambles that involve winning a large prize with small probability (as in lottery gambles). DA utility also accommodates the violation of the independence axiom commonly observed in experiments (the Allais paradox) as shown by Gul (1991).

DA utility has a number of advantages over prospect theory while still capturing the asymmetry between losses and gains. First, in prospect theory there is no guidance about how to choose and update the reference point to which gains and losses are compared. Moreover, the portfolio weights implied by prospect theory depend very sensitively on the choice of the reference point. Benartzi and Thaler (1995) choose current wealth as the reference point, while Barberis, Huang and Santos (1999) use current wealth times the risk-free rate. In DA utility the reference point is endogenous, and can be updated over time without having to make an arbitrary exogenous choice.

The second advantage of DA utility over loss aversion is that DA utility is axiomatic (Gul (1991)) and is a normative theory. Hence formal techniques like dynamic programming can be consistently applied rather than being *ad hoc* adapted to a descriptive theory. We show how optimal asset allocations can be computed both in a static and dynamic framework, where investors with DA preferences maximize end-of-period wealth subject to an exogenous return process. We generalize DA preferences to a multi-period dynamic asset allocation set-up, which includes dynamic constant relative risk aversion (CRRA) utility as a special case.

While Kahneman and Tversky (1979)'s prospect theory differs in a myriad of ways from the standard expected utility framework, DA preferences are very closely related to the standard CRRA expected utility preferences that are prevalent in mainstream portfolio theory. In fact, standard preferences are a special case of DA preferences with the loss aversion parameter put equal to one. We consider this closeness to be the major advantage of our proposed framework: we can capture many of the asymmetric effects of loss aversion without resorting to behavioral theory. This makes our results directly comparable to the large body of empirical work that has accumulated on dynamic asset allocation. We illustrate the connection in this article, by considering a DGP where stock returns are predictable.

The sensitivity of intertemporal hedging demands to different forms of stock return predictability has been the focus of much of the recent dynamic asset allocation literature. Whereas some (Brennan, Schwartz and Lagnado (1997), Barberis (2000), Campbell and Viceira (1998) and Liu (1999)) find hedging demands to be large, others (Brandt (1999), and Ang and Bekaert (1999)) find them to be small. Our results strongly suggest that the proper specification of an investor's utility function matters as much as, if not more than, the proper specification of the stochastic environment. Given weaker evidence on the predictability of excess returns with

more recent data, especially using traditional instruments such as dividend yields (See Ang and Bekaert (2000) and Amit and Goyal (1999)), this may hopefully re-focus the direction the literature takes.

Although DA preferences are promising, it remains to be seen whether they are a viable alternative to a more exotic theory. There are two ways in which they could fail. First, DA preferences may simply be not flexible enough to generate realistic portfolio allocations. In that case, our results add to the body of work that calls for changing our standard preferences paradigm. Second, DA preferences are of little use if we can obtain the same results with a standard utility function at higher risk aversion. In this article, we will clearly demonstrate that DA preferences yield predictions that cannot be obtained by scaling up risk aversion, because they generate asset allocations which exhibit different intertemporal hedging and state dependence than what is implied by standard CRRA preferences.

Finally, we mention a growing literature analyzing the effects of loss or disappointment aversion. In equilibrium settings, Epstein and Zin (1991) consider embedding a number of alternative preferences, including DA preferences into an infinite horizon consumption model with recursive preferences. Bekaert, Hodrick and Marshall (1997) consider asset return predictability in the context of an international consumption model with DA preferences. Barberis, Huang and Santos (1999) use prospect theory in an infinite horizon consumption problem. To do this, they have to make a number of non-standard auxiliary assumptions about how prospect theory can be generalized to a dynamic setting, and how to specify and update the reference point. Other authors have used loss aversion or DA utility in partial equilibrium settings. Gomes (2000) completely characterizes stock holdings of loss aversion under two state lotteries for the stock return, and Berkelaar and Kouwenberg (2000) derive closed-form solutions for optimal loss aversion portfolio choice under more general return distributions. Benartzi and Thaler (1995) use loss aversion to try to explain the equity premium with different rebalancing horizons. All these papers are conceptually different from our dynamic asset allocation set-up with an exogenous DGP and end-of-period utility wealth optimization.²

This article is organized as follows. In Section 2 we contrast the basic DA preference framework both with standard CRRA preferences and with loss aversion in the context of a number of simple realistic gamble examples. In Section 3, we introduce the formal asset allocation framework and show how to solve for optimal asset weights for a particular DGP. We discuss static CRRA and DA problems and generalize DA to a dynamic long horizon case. The fourth

²A different treatment of an investor's asymmetric response to gains and losses is given by Roy (1952), Maenhout (1999), and Stutzer (1999). These authors model agents who first minimize the possibility of undesirable outcomes.

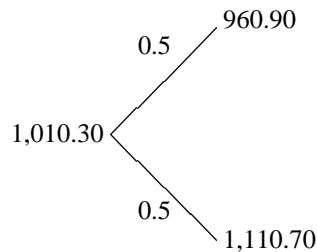
section delivers optimal asset allocation under two different DGP's, estimated from US data on stocks and interest rates, including one that embeds predictability of equity returns. We also consider what equity premiums are required to make investors hold the observed mix of equity and risk free assets. Section 5 looks at the robustness of our results by considering a DGP incorporating inflation as a state variable, and a DGP estimated on more recent data that shows stronger predictability. Section 6 concludes.

2 Why Stocks May Disappoint

In this Section, we introduce DA preferences and compare them to standard CRRA and loss-aversion preferences in the context of simple two-state atemporal gambles. First, we show that while DA investors may choose to accept lottery-type gambles but decline stock gambles, risk-averse CRRA agents will reject both gambles. Second, we show that the unrealistic risk aversion over large stakes implied by CRRA utility (Rabin (1999)) is not shared by DA utility. These examples show that DA preferences are more realistic approximations of investor's preferences than standard expected utility.

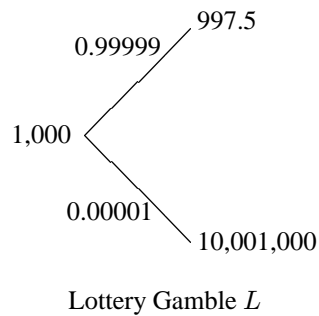
2.1 Lotteries versus Stocks

Many people do not invest in stocks but have no qualms buying lottery tickets. To investigate the consistency of such behavior with various preferences, we consider two different gambles, S and L , where S stands for "stocks" and L denotes "lottery". We use quarterly data on US stock returns and the quarterly T-bill rate to calibrate the stock gamble. Initial wealth is set at $1,000 \times (1 + 0.0103)$, where 0.0103 is the average quarterly 3 month T-bill rate over 1941-1998. There is a 50% probability of realizing $1,000 \times (1 + \mu + \sigma)$ and a 50% probability of realizing $1,000(1 + \mu - \sigma)$, where $\mu = 0.0358$ and $\sigma = 0.0749$, the mean and standard deviation of quarterly US simple stock returns over the same period. We represent the S gamble by:



Stock Gamble S

For the lottery-type payoff, we assume that the investor, with initial wealth also equal to 1,000 dollars, either loses 2.5 dollars, reflecting the small cost of lottery tickets, but has a one in a 100,000 chance of winning \$10,000,000. Hence:



For our two gambles, we assume that the relevant benchmark is either initial wealth scaled up with the risk free rate for the stock gamble or simply initial wealth for the lottery-type gamble.³

That many people will reject the first gamble but accept the second is puzzling from the perspective of standard expected utility, especially given the enormous positive expected return on the lottery-type gamble. To see this, consider standard CRRA preferences, where risk aversion is characterized by risk aversion γ . That is, the utility function over random wealth W is given by:

$$U(W) = \frac{W^{1-\gamma}}{1-\gamma} \tag{1}$$

To determine whether investors with different risk aversions will accept or reject the S or L gambles, we compute the “willingness-to-pay” (to avoid the gamble) for both gambles. The willingness-to-pay is the difference between the certain wealth the investors have available by not taking on the gamble minus the certainty equivalent of the gamble. The certainty equivalent is the certain level of wealth that generates the same utility as the gamble, hence the

³We have to assume a positive expected value to the lottery, which we interpret as the agent gaining utility above purely potential monetary gains. In reality, the physical distribution has negative expected returns in lottery or casino gambles. However, the lottery buyer may “feel lucky” or gamblers may feel that they are “better than average” so the expected returns using their subjective probability distribution may be positive. Following Benartzi and Thaler (1995) and Barberis, Huang and Santos (1999), we do not consider subjective transformations of the physical distribution, but work directly with the physical distribution which is assumed to be known by the agent. A third way to explain why people take negative expected return gambles is to assume that those people are risk seeking at very low wealth levels. In Kahneman and Tversky (1979), the utility function may be convex (risk seeking) in the loss region. Both Benartzi and Thaler (1995) and Barberis, Huang and Santos (1999) ignore the risk-seeking areas in the loss region.

certainty equivalent associated with different gambles can be used to represent the utility function. We denote the certainty equivalent for CRRA utility as $\mu_W^{CRRA} = E[U(W)]^{1/(1-\gamma)}$. If the willingness-to-pay is negative, rational agents would accept the gamble. The top plot of Figure (1) shows willingness-to-pay dollar amounts for S and L in the case of CRRA utility. In Figure (1) we divide the plots into different areas:

Area	lottery L	stock S
A	likes	likes
B	dislikes	likes
C	dislikes	dislikes
D	likes	dislikes

The puzzle is very apparent. Only investors with γ 's higher than 10 would reject the stock gamble. But such agents would never invest in lotteries; only agents close to risk neutral find the lottery payoff attractive and they also like stocks.

Although this puzzle has to our knowledge not been offered as a motivation for the use of loss aversion, we now demonstrate that Kahneman and Tversky (1979)'s prospect theory as applied by Benartzi and Thaler (1995) can be used to resolve this puzzle. This preference framework is very different from standard expected utility. First, utility is defined over gains and losses relative to a reference point, rather than wealth levels. We choose the reference level to be initial wealth as Benartzi and Thaler do. Second, the utility is asymmetric over these gains and losses. With χ representing a gain or loss relative to a reference point, the loss aversion (LA) utility of Kahneman and Tversky (1979) is given by:

$$-\lambda E[(-\chi)^{(1-\gamma_1)} \mathbf{1}_{\{\chi \leq 0\}}] + E[\chi^{(1-\gamma_2)} \mathbf{1}_{\{\chi > 0\}}], \quad (2)$$

where $\mathbf{1}$ is an indicator variable, $\chi = W - B_0$ is the gain or loss of final wealth W relative to a benchmark B_0 . Kahneman and Tversky estimated $\lambda = 2.25$, so losses are weighted 2.25 times as much as gains, and $\gamma_1 = \gamma_2 = 0.12$, implying the same amount of curvature across gains and losses. The felicity function $(-\lambda(-\chi)^{1-\gamma_1} \mathbf{1}_{\{\chi \leq 0\}} + \chi^{(1-\gamma_2)} \mathbf{1}_{\{\chi > 0\}})$, is monotone in χ if $0 \leq \gamma_1 < 1$ and $0 \leq \gamma_2 < 1$.⁴

In Kahneman and Tversky (1979), the expectation in the above equation is taken with a subjective probability distribution transformed from the objective probability distribution, as Kahneman and Tversky allow for the possibility for individuals to transform objective probabilities into “decision weights”. However, Benartzi and Thaler (1995)'s results are robust to

⁴Note that the utility function gives different preferences if χ is expressed in different units unless $\gamma_1 = \gamma_2$ or the difference between γ_1 and γ_2 is very small. Expressing χ in returns (so χ has no dimension) circumvents this problem.

this transformation. Barberis, Huang and Santos (1999) also use objective probabilities. The parameter λ governs the additional weight on losses. Benartzi and Thaler set $\lambda = 2.25$, $\gamma_1 = 0$ and $\gamma_2 = 0$, so they use a bilinear model.

Although the LA utility in equation (2) is defined over gains and losses, it is possible to calculate a certainty equivalent of wealth of LA for most parameter values. Since gains and losses are always evaluated relative to a benchmark, wealth is implicitly given as the gain or loss plus the reference point. Denoting the LA utility in equation (2) as U_{LA} , the certainty equivalent of LA, μ_W^{LA} , is defined by:

$$\mu_W^{LA} = \begin{cases} U_{LA}^{\frac{1}{1-\gamma_2}} + B_0 & \text{if } U_{LA} > 0, \\ - \left(-\frac{U_{LA}}{\lambda} \right)^{\frac{1}{1-\gamma_1}} + B_0 & \text{if } U_{LA} \leq 0 \end{cases} \quad (3)$$

where B_0 is the benchmark of the gamble, which is initial wealth in our case. The middle plot of Figure (1) graphs the willingness-to-pay for LA utility for $\gamma = 0.12$ and various λ values. Individuals with small λ both like stock gambles and lotteries, but when λ increases to 2.5, the stock gamble is no longer attractive but the lottery gamble remains desirable.

It is not necessary to deviate so dramatically from expected utility to obtain such results. Consider the following implicitly defined utility function, which defines DA utility:

$$U(\mu_W) = \frac{1}{K} \left(\int_{-\infty}^{\mu_W} U(W) dF(W) + A \int_{\mu_W}^{\infty} U(W) dF(W) \right) \quad (4)$$

where $U(\cdot)$ represents power utility (CRRA), that is $U(W) = W^{(1-\gamma)}/(1-\gamma)$, $A \leq 1$ is the coefficient of disappointment aversion, $F(\cdot)$ is the cumulative distribution function for wealth, and

$$K = Pr(W \leq \mu_W) + A Pr(W > \mu_W). \quad (5)$$

If $0 \leq A < 1$ the outcomes below the certainty equivalent are weighted more heavily than outcomes above the certainty equivalent. Note that these preferences are outside the standard expected utility framework because the level of utility at the optimum (or the certainty equivalent of wealth) appears on the right hand side. Although this is a non-expected utility function, CRRA preferences are a special case for $A = 1$. Moreover, this utility function can easily embed aversion to losses by letting $A \rightarrow 0$. When A is zero, individuals derive no utility at all from gains but worry only about losses.

The bottom plot of Figure (1) shows willingness-to-pay for $\gamma = 0.12$ and various A values. Clearly, this utility function resolves the stock/lottery puzzle. The fairly small losses the lottery gamble generates deter disappointment averse investors only slightly given the possibility of a

very large payoff should they win the lottery. As long as A remains above approximately 0.1 the lottery payoff generates positive utility. Stocks, however, are only attractive when people are not too disappointment averse. When A drops below 0.5, the individual does not invest in the stock market anymore fearing the loss, but does accept the lottery gamble.

2.2 Rabin Gambles

In a recent article, Rabin (1999) shows that within the expected-utility framework, anything but virtual risk neutrality over modest stakes implies manifestly unrealistic risk aversion over large stakes. His “calibration theorem” is best illustrated with an example. Suppose that for some ranges of wealth (or for all wealth levels), a person turns down gambles where she loses \$100 or gains \$110, each with equal probability. Then she will turn down 50%-50% bets of losing \$1,000 or gaining ANY sum of money. We will call such a gamble a “Rabin gamble”. Since DA preferences do not fall into the expected utility category, they do not necessarily suffer from the Rabin-gamble problem.

Figure (2) illustrates this. Imagine an investor with \$10,000 wealth. If he has CRRA preferences, a $\gamma = 10$ makes him reject the initial 100/110 gamble. The graph shows both his utility and willingness-to-pay relative to the Rabin gamble of losing 1,000 and gaining the amount on the x -axis. The last amount on the right hand side of the x -axis represents \$25,000. It is apparent from the top graph that the marginal utility of additional wealth becomes virtually zero very fast. The willingness to pay to avoid the gamble asymptotes to about \$280, even if the potential gain is over \$1,000,000. The extreme curvature in the utility function drives the continued rejection of the second gamble even as the possible amount of money to be gained increases to infinity.

With DA preferences, an investor need not display an extremely concave utility function to dislike the original 100/110 gamble, because he hates to lose \$100. In fact, an investor with $\gamma = 0$ but $A = 0.9$ will reject the original gamble, since she puts $1/0.9 = 1.11$ times more weight on the loss than on the gain. Of course, this particular investor’s willingness-to-pay to avoid the gamble will be very small, but he will nonetheless reject it. However, such an investor loves the Rabin gambles. Since there is no curvature in the utility function, the utility and willingness to pay are linear in the gain and increase (decrease) monotonically. For example, our DA investor would be willing to pay \$10,000 to enter a bet where he can gain \$25,000 but may lose \$1,000 with equal probability. From introspection, this seems a much more reasonable attitude towards risk.

3 Asset Allocation under Disappointment Aversion

In Sections 3.1 and 3.2 we first consider a simple one-period asset allocation problem for CRRA and DA preferences, then extend to consider the dynamic case in Sections 3.3 and 3.4.

3.1 Static CRRA Utility

The investment opportunity set of an investor with initial wealth W_0 consists of a risky asset and a riskless bond. The bond yields a certain return of r^f and the risky asset yields an uncertain return of y . The investor chooses the proportion of his initial wealth to invest in the risky asset α , to maximize the expected utility of end-of-period wealth W , which is uncertain. Formally, the problem is:

$$\max_{\alpha} E[U(W)] \quad (6)$$

where W is given by

$$W = \alpha W_0 (\exp(y) - \exp(r^f)) + W_0 \exp(r^f). \quad (7)$$

Since CRRA utility is homogenous in wealth, we set $W_0 = 1$.

The first-order condition (FOC) of equation (6) is solved by choosing α to equate:

$$\int_{-\infty}^{\infty} W^{-\gamma} (\exp(y) - \exp(r^f)) dF(y) = 0 \quad (8)$$

where $F(\cdot)$ is the cumulative density function of the risky asset return. This expectation can be computed by numerical quadrature as described in Tauchen and Hussey (1991). This procedure involves replacing the integral with a probability-weighted sum:

$$\sum_{s=1}^N p_s W_s^{-\gamma} (\exp(y_s) - \exp(r^f)) = 0. \quad (9)$$

The N values of the risky asset return ($\{y_s\}_{s=1}^N$) and the associated probabilities ($\{p_s\}_{s=1}^N$) are chosen by an optimal quadrature rule. W_s represents the investor's terminal wealth when the risky asset return is y_s . In the case of normally distributed asset returns, Gaussian quadrature can be used to determine the abscissae and the associated probability weights. Moreover, the approximation is very accurate with as few as five points. Quadrature approaches to asset allocation problems have been used by Ang and Bekaert (1999), Balduzzi and Lynch (1999), and Campbell and Viceira (1999), among others.

3.2 Static DA Utility

For the non-expected utility case of DA preferences, the optimization problem becomes:

$$\max_{\alpha} U(\mu_W) \quad (10)$$

where the certainty equivalent is defined in equation (4) and end of period wealth W is given by equation (7). For $U(\cdot)$ given by power utility, optimal utility remains homogenous in wealth and we set $W_0 = 1$. The implicit definition of μ_W makes the optimization problem non-trivial, and we relegate a rigorous treatment to the Appendix (See also Epstein and Zin (1989, 1991)). Here, we offer an intuitive derivation of the FOC's. To do so, we introduce an alternative preference function:

$$V(W) = \begin{cases} U(W) & \text{if } W > \mu_W, \\ U(W) - (\frac{1}{A} - 1)[U(\mu_W) - U(W)] & \text{if } W \leq \mu_W \end{cases} \quad (11)$$

Note that $U(\mu_W) = E[V(W)]$. Equation (11) clearly shows the penalty associated with disappointing outcomes (those that are worse than expectations). The lower A , the larger the penalty and when $A = 1$, the penalty is zero and expected utility results.

As equation (11) demonstrates, in DA utility, the reference point defining elating outcomes (“gains”), versus disappointing outcomes (“losses”) is endogenous. In contrast, for standard loss aversion preferences, the reference point is arbitrarily defined. For example, suppose χ in the loss aversion utility in equation (2) is end of period wealth, so the reference point B_0 is zero. If $0 \leq \alpha \leq 1$, wealth is then always positive and the loss aversion utility function reduces to $E[\chi^{(1-\gamma_2)}]$ which are standard CRRA preferences. Barberis, Huang and Santos (1999) choose to set the reference point at current wealth times the risk-free return. In this case, the gain or loss χ in equation (2) is given by $\chi = \alpha(\exp(y) - \exp(r^f))$. If x_e denotes the excess return ($\exp(y) - \exp(r^f)$) the loss aversion utility can be written as:

$$E \left[-\lambda(-\alpha x_e)^{(1-\gamma_1)} \mathbf{1}_{\{\alpha x_e \leq 0\}} + (\alpha x_e)^{(1-\gamma_2)} \mathbf{1}_{\{\alpha x_e > 0\}} \right]. \quad (12)$$

In the case of $\gamma_1 = \gamma_2 = \gamma$, as in Kahneman and Tversky (1979), the loss aversion utility simplifies to $E[k(\alpha)^{(1-\gamma)}(|x_e|)^{(1-\gamma)}]$, where k is a constant, which is a homogenous function in the excess return. The formal solution to this asset allocation problem is a corner solution where α is either zero or infinity. DA utility does not suffer from such problems as the reference point is endogenous.

The DA investor's problem can be solved by maximizing $E[V(W)]$ over α , where heuristi-

cally we ignore the constant $U(\mu_W)$:

$$\max_{\alpha} \mathbb{E}[V(W)] = \max_{\alpha} \left(\begin{array}{l} \mathbb{E}[U(W)|W > \mu_W]Pr(W > \mu_W) \\ + \frac{1}{A}\mathbb{E}[U(W)|W \leq \mu_W]Pr(W \leq \mu_W) \end{array} \right) \quad (13)$$

The FOC for this problem is:

$$\begin{aligned} \mathbb{E} \left[\frac{\partial U(W)}{\partial W} (\exp(y) - \exp(r^f)) \mid W > \mu_W \right] Pr(W > \mu_W) \\ + \frac{1}{A} \mathbb{E} \left[\frac{\partial U(W)}{\partial W} (\exp(y) - \exp(r^f)) \mid W \leq \mu_W \right] Pr(W \leq \mu_W) = 0 \end{aligned} \quad (14)$$

If μ_W were known, equation (14) could be solved for α in the same way as in the standard expected utility framework. The only difference is that for states below μ_W , the original utilities have to be scaled up by $1/A$. However, μ_W is itself a function of the outcome of optimization (that is, μ_W is a function of α). Hence, equation (14) must be solved simultaneously with equation (11) which defines μ_W .

To solve these equations numerically, we can use quadrature to approximate the definition of μ_W in equation (11) by:

$$\mu_W^{1-\gamma} = \sum_{s:W_s > \mu_W} p_s W_s^{1-\gamma} + \sum_{s:W_s \leq \mu_W} p_s \left(W_s^{1-\gamma} - \left(\frac{1}{A} - 1\right) [\mu_W^{1-\gamma} - W_s^{1-\gamma}] \right), \quad (15)$$

and the FOC in equation (14) by:

$$\sum_{s:W_s > \mu_W} p_s W_s^{-\gamma} (\exp(y_s) - \exp(r^f)) + \sum_{s:W_s \leq \mu_W} p_s \frac{W_s^{-\gamma}}{A} (\exp(y_s) - \exp(r^f)) = 0 \quad (16)$$

Equations (15) and (16) are solved simultaneously to yield the portfolio weight α that maximizes the utility of this disappointment-averse investor. The exact discretization procedure used depends on the DGP for returns and is discussed in detail in the Appendix.

The closeness between standard expected utility and DA preferences suggests another algorithm to obtain the optimal asset allocation for a discrete state space. Let $x_e = (\exp(y) - \exp(r^f))$ denote the excess stock return. With N quadrature points there are N outcomes for x_e , $\{x_{es}\}_{s=1}^N$, with probability weights $\{p_s\}_{s=1}^N$. Without loss of generality we can order x_e from low to high across states s . The utility equivalent μ_W^* corresponding to the optimal portfolio weight α^* can be in any of N intervals:

$$[\exp(r^f) + \alpha^* x_{e1}, \exp(r^f) + \alpha^* x_{e2}),$$

$$[\exp(r^f) + \alpha^* x_{e2}, \exp(r^f) + \alpha^* x_{e2}),$$

⋮

$$[\exp(r^f) + \alpha^* x_{e,N-1}, \exp(r^f) + \alpha^* x_{eN}].$$

Suppose μ_W^* lies in $[\exp(r^f) + \alpha^* x_{ei}, \exp(r^f) + \alpha^* x_{e,i+1}]$ for some state i . Then α^* solves

$$\sum_{s:W_s > \exp(r^f) + \alpha^* x_{e,i+1}} p_s (W_s^*)^{-\gamma} x_{es} + \sum_{s:W_s \leq \exp(r^f) + \alpha^* x_{e,i}} \frac{p_s}{A} (W_s^*)^{-\gamma} x_{es} = 0 \quad (17)$$

where $W_s^* = \exp(r^f) + \alpha^* x_{es}$. Equation (17) is a CRRA maximization problem with a changed probability distribution $\pi_i = \{\pi_{is}\}_{s=1}^N$, where the probabilities above the certainty equivalent are downweighted, that is:

$$\pi_i \equiv \frac{(p_1, \dots, p_i, Ap_{i+1}, \dots, Ap_N)'}{(p_1 + \dots + p_i) + A(p_{i+1} + \dots + p_N)}. \quad (18)$$

Our algorithm is as follows. We start with a state i and solve the CRRA problem with probability distribution π_i . Then we calculate the certainty equivalent $\mu_{W_i}^*$ given by:

$$\mu_{W_i}^* = \left(\sum_{s=1}^N (W_s^*)^{1-\gamma} \pi_{is} \right)^{\frac{1}{1-\gamma}} \quad (19)$$

Then we check if in this state the following is true:

$$\mu_{W_i} \in [\exp(r^f) + \alpha_i^* x_{ei}, \exp(r^f) + \alpha_i^* x_{e,i+1}]. \quad (20)$$

If this is true for $i = i^*$ then $\alpha^* = \alpha_{i^*}^*$ and $\mu_W^* = \mu_{W_{i^*}}^*$. As the states are ordered in increasing wealth across states for a given portfolio weight, it is easy to do a bisection search algorithm (with intermediate CRRA optimizations) to obtain the DA portfolios. If we start our search for i^* at the midpoint of the N states and find that $\mu_{W_i} > (<) \exp(r^f) + \alpha_i^* x_{e,i+1}$, then we begin a search in the upper (lower) half of the state space.⁵

3.3 Dynamic CRRA Utility

When the return distribution is independent and identically distributed (IID) over time, we know from Samuelson (1969) that there are no hedging demands in a standard expected utility framework with CRRA preferences. However, once returns are not IID, intertemporal hedging demands for CRRA utility appear. In the empirical section, we consider a DGP in which the

⁵A similar algorithm is given in Gul (1991)'s appendix. Both our algorithm and Gul's require the solution of an optimization problem in each discrete state. The difference is that in our algorithm we solve a simple smooth CRRA problem, whereas Gul requires a non-linear maximization involving an indicator function. For his optimization problem, gradient-based search algorithms cannot be used, and thus our algorithm is numerically more tractable.

interest rate predicts equity returns and solve the dynamic asset allocation problem under this setting.

Our problem for dynamic CRRA utility is to find a series of portfolio weights $\alpha = \{\alpha_t\}_{t=0}^{T-1}$ to maximize:

$$\max_{\alpha_0, \dots, \alpha_{T-1}} E_0[U(W_T)] \quad (21)$$

where $\alpha_0, \dots, \alpha_{T-1}$ are the portfolio weights at time 0 (with T periods left), \dots , to time $T - 1$ (with 1 period left), and $U(W) = W^{1-\gamma}/(1 - \gamma)$. Wealth W_t at time t is given by $W_t = R_t(\alpha_{t-1})W_{t-1}$ with given by

$$R_t(\alpha_{t-1}) = \alpha_{t-1}(\exp(y_t) - \exp(r_{t-1}^f)) + \exp(r_{t-1}^f).$$

As in the static case, since CRRA utility is homogenous in wealth, we set $W_0 = 1$.

Using dynamic programming we can obtain the portfolio weights at each horizon t by using the (scaled) indirect utility:

$$\alpha_t^* = \arg \max_{\alpha_t} E_t[Q_{t+1,T} W_{t+1}^{1-\gamma}] \quad (22)$$

where $Q_{t+1,T} = E_{t+1} [(R_T(\alpha_{T-1}^*) \dots R_{t+2}(\alpha_{t+1}^*))^{1-\gamma}]$, and $Q_{T,T} = 1$. The variable $Q_{t+1,T}$ is the investor's (scaled) indirect utility. The FOC's of the investor's problem are:

$$E_t[Q_{t+1,T} R_{t+1}^{-\gamma}(\alpha_t) x_{e,t+1}] = 0, \quad (23)$$

where $x_{e,t+1} = (\exp(y_{t+1}) - \exp(r_t^f))$ are the excess returns at time $t + 1$. This expectation can be solved using quadrature in a similar manner to the static problem. For N quadrature points there will be N values of $Q_{t+1,T}$ to keep track at each horizon. At each horizon there will also be N portfolio weights, corresponding to each state.

In equation (22), if (y_{t+1}, r_{t+1}) is independent of (y_t, r_t) for all t , then $Q_{t+1,T}$ is independent of $W_{t+1} \equiv R_{t+1}^{1-\gamma}(\alpha_t)$, so equation (22) becomes

$$E_t[Q_{t+1,T} W_{t+1}^{1-\gamma}] = E_t[Q_{t+1,T}] E_t[R_{t+1}^{1-\gamma}(\alpha_t)] \quad (24)$$

Since $E_t[Q_{t+1,T}]$ does not depend on α_t the objective function for the optimization problem at time t is equivalently $E_t[R_{t+1}^{1-\gamma}(\alpha_t)]$. Thus the problem has been reduced to a single-period problem and there will be no intertemporal hedging component.

3.4 Dynamic DA Utility

The DA utility defined in equation (4) is atemporal. We generalize DA to a multi-period setting, which has dynamic CRRA utility as a special case. We will show that in our multi-period

version of DA utility, the portfolio weights will also be constant across investment horizon when returns are IID. The generalization of DA utility to multiple periods is less trivial than it may seem. Therefore, we first explore a two-period example, before discussing our dynamic programming algorithm.

3.4.1 Two Period Example

Suppose there are three dates $t = 0, 1, 2$ and two states u, d for the excess equity return at dates $t = 1, 2$. Without loss of generality we take the risk-free rate to be zero. The distribution is independent across time. In this special setting $R_t(\alpha_{t-1})$ is given by $1 + \alpha_{t-1}u$ in state u and $1 + \alpha_{t-1}d$ in state d . The agent chooses optimal portfolios at dates 0 and 1.

At $t = 1$ an agent will maximize μ_1 in each state u and d given by:

$$K_1\mu_1^{1-\gamma} = AE[R_2^{1-\gamma}(\alpha_1)\mathbf{1}_{\{R_2(\alpha_1) > \mu_1\}}] + E[R_2^{1-\gamma}(\alpha_1)\mathbf{1}_{\{R_2(\alpha_1) \leq \mu_1\}}], \quad (25)$$

to get the optimal portfolio weight α_1 and the corresponding optimal utility μ_1^* . The constant $K_1 = Pr(R_2(\alpha_1) \leq \mu_1) + APr(R_2(\alpha_1) > \mu_1)$. Since the distribution is IID, μ_1^* is the same across states, that is $\mu_1^*(u) = \mu_1^*(d)$.

Suppose at $t = 0$ the agent defines the DA utility as:

$$K_0\mu_0^{1-\gamma} = AE_0[(R_1(\alpha_0)R_2(\alpha_1^*))^{1-\gamma}\mathbf{1}_{\{R_1(\alpha_0)R_2(\alpha_1^*) > \mu_0\}}] + E_0[(R_1(\alpha_0)R_2(\alpha_1^*))^{1-\gamma}\mathbf{1}_{\{R_1(\alpha_0)R_2(\alpha_1^*) \leq \mu_0\}}], \quad (26)$$

where $K_0 = Pr(R_1(\alpha_0)R_2(\alpha_1^*) \leq \mu_0) + APr(R_1(\alpha_0)R_2(\alpha_1^*) > \mu_0)$. That is, he computes the certainty equivalent of end-of-period wealth, given his current information. There are four states $\{uu, ud, du, dd\}$ with portfolio returns $\{(1 + \alpha_0u)(1 + \alpha_1^*u), (1 + \alpha_0u)(1 + \alpha_1^*d), (1 + \alpha_0d)(1 + \alpha_1^*u), (1 + \alpha_0d)(1 + \alpha_1^*d)\}$. Note that the returns are not necessarily recombining (the ud return can be different from the du return) using this definition of DA utility. We must track both the return states both at $t = 1$ and $t = 0$. As a result, the number of states increases exponentially with the number of periods. Moreover, the optimization is time-dependent, so portfolio weights will depend on the horizon even when the returns are IID.

An alternative way to compute the certainty equivalent at $t = 0$ is to use the certainty equivalent at $t = 1$ rather than actual returns:

$$K_0\mu_0^{1-\gamma} = AE_0[(R_1(\alpha_0)\mu_1^*)^{1-\gamma}\mathbf{1}_{\{R_1(\alpha_0)\mu_1^* > \mu_0\}}] + E_0[(R_1(\alpha_0)\mu_1^*)^{1-\gamma}\mathbf{1}_{\{R_1(\alpha_0)\mu_1^* \leq \mu_0\}}], \quad (27)$$

where K_0 is now defined as $K_0 = Pr(R_1(\alpha_0) \leq \mu_0) + APr(R_1(\alpha_0) > \mu_0)$. In this formulation there are only two states $\{u, d\}$, and we only need to track $\{(1 + \alpha_0u)\mu_1^*, (1 + \alpha_0d)\mu_1^*\}$. This

agent uses the next period's indirect utility μ_1^* to form the DA utility this period as in a dynamic programming problem. The endogenous reference point also updates itself and depends on the future optimal return. This generalization of DA utility to a dynamic setting not only preserves computational feasibility but also preserves the property that the CRRA dynamic program (using the CRRA indirect utility) is a special case for $A = 1$. Like CRRA utility, the DA portfolio weights in this generalization of DA utility to a dynamic setting will not exhibit intertemporal hedging demands if the return DGP is IID.

3.4.2 Dynamic DA Algorithm

Building on the DA utility defined in equation (27) we present an algorithm for solving dynamic DA utility. Our problem is similar to the problem described in equation (21), but utility is now disappointment aversion. We start the dynamic program at horizon $t = T - 1$. We solve:

$$\max_{\alpha_{T-1}} \mu_{T-1}(\alpha_{T-1}) \quad (28)$$

where μ_{T-1} is defined by:

$$K_{T-1} \mu_{T-1}^{1-\gamma} \equiv \text{AE}_{T-1}[R_T^{1-\gamma}(\alpha_{T-1}) \mathbf{1}_{\{R_T(\alpha_{T-1}) > \mu_{T-1}\}}] + \text{E}_{T-1}[R_T^{1-\gamma}(\alpha_{T-1}) \mathbf{1}_{\{R_T(\alpha_{T-1}) \leq \mu_{T-1}\}}], \quad (29)$$

with $K_{T-1} = \text{Pr}(R_T(\alpha_{T-1}) \leq \mu_{T-1}) + A \text{Pr}(R_T(\alpha_{T-1}) > \mu_{T-1})$. We denote the optimal portfolio weight α_{T-1}^* , with corresponding μ_{T-1}^* , which can be solved as in the one-period problem. At this horizon, this problem is equivalent to the static problem, except it is solved for each quadrature state. The optimal utility equivalent μ_{T-1}^* will differ in each state.

At horizon $t = T - 2$ we solve:

$$\max_{\alpha_{T-2}} \mu_{T-2}(\alpha_{T-2}) \quad (30)$$

where μ_{T-2} is defined by:

$$K_{T-2} \mu_{T-2}^{1-\gamma} \equiv \text{AE}_{T-2}[R_{T-1}^{1-\gamma}(\alpha_{T-2})(\mu_{T-1}^*)^{1-\gamma} \mathbf{1}_{\{R_{T-1}(\alpha_{T-2})\mu_{T-1}^* > \mu_{T-2}\}}] + \text{E}_{T-2}[R_{T-1}^{1-\gamma}(\alpha_{T-2})(\mu_{T-1}^*)^{1-\gamma} \mathbf{1}_{\{R_{T-1}(\alpha_{T-2})\mu_{T-1}^* \leq \mu_{T-2}\}}], \quad (31)$$

with $K_{T-2} = \text{Pr}(R_{T-1}(\alpha_{T-2}) \leq \mu_{T-2}) + A \text{Pr}(R_{T-1}(\alpha_{T-2}) > \mu_{T-2})$. In this definition we need to keep track of only the states at $T - 1$ for μ_{T-1}^* to solve for optimal α_{T-2}^* and μ_{T-2}^* . We continue this process for $t = T - 3$ until $t = 0$.

If $A = 1$, then at horizon $t = T - 2$ the DA utility reduces to:

$$\mu_{T-2}^{1-\gamma} = \text{E}_{T-2}[R_{T-1}^{1-\gamma}(\alpha_{T-2})(\mu_{T-1}^*)^{1-\gamma}] = \text{E}_{T-2}[R_{T-1}^{1-\gamma}(\alpha_{T-2})Q_{T-1,T}] \quad (32)$$

which is the standard CRRA problem. Note that if returns are IID, then at each horizon, exactly the same DA problem will be solved and the portfolio weights are independent of the horizon.

To solve the DA problem at each horizon, we can solve simultaneously the maximization problem and the definition of the utility equivalent, as in the static DA problem where equations (11) and (14) are solved simultaneously. In the case where μ_t is increasing across states, it is possible to extend the bisection algorithm given in Section 3.2 to the dynamic case. This is done by defining normalized wealth $\tilde{W}_s = W_s \mu_{s,t}$ at horizon t for state s in equation (17).⁶ If the states are not able to be ordered in increasing wealth for a given portfolio weight, then a bisection algorithm is not possible, but at most $N - 1$ steps are required to find the optimal DA portfolio weight for N quadrature states at each horizon.

4 Disappointment Aversion and Stock Holdings

4.1 Data and DGP's

To examine portfolio choice under realistic DGP's, we use US data on stock returns and Treasury bills. Most of our results use nominal quarterly data from 1926 to 1998, but we also check robustness for the post 1940 period. The use of nominal data makes our study comparable to the empirical work in Benartzi and Thaler (1995) but it does make T-bills unrealistically attractive for disappointment averse agents (unless they truly exhibit money illusion). Hence, in Section 5, we estimate a DGP using real stock return data and real T-bill returns.

Table (1) summarizes some properties of the stock return and Treasury bill data, most of them well-known. The equity premium is about 6.55% (in logs) over the whole period but about a percent higher post 1940, although the average interest rate is higher then. Equity volatility is lower post 1940, but the decrease is primarily due to the exclusion of the 1929 crash. Generally, stocks look more attractive post 1940. In real terms, the average annual post-1940 real return on T-bills is 1.2%, whereas on stocks it is 8.7%. Volatility of real stock returns and real T-bill returns are higher than their nominal counterparts: 16% (5%) real versus 15% (4%) nominal for stock (bond) returns. Whereas interest rates are generally persistent processes, the persistence of the ex-post real rate (0.56) is much lower than that of nominal interest rates (0.92).

We use two main DGP's in this paper that largely conform to what has been used in the extensive literature on dynamic asset allocation.⁷ In our first model, stock returns are IID over

⁶If W_s is increasing across states for a given portfolio weight, and μ_{ts} is also increasing across states for a given portfolio weight, then \tilde{W}_s will also be increasing across states.

⁷See for example Barberis (2000), Campbell and Viceira (1999, 1998), Balduzzi and Lynch (1999), and Kandel

time and the interest rate follows a first-order autoregressive system. This is slightly more general than Benartzi and Thaler (1995) who consider only IID returns. Following most of the dynamic asset allocation literature, we consider only one possible predictor of stock returns and consider a system where an instrument linearly predicts stock returns in the conditional mean of equity returns. Whereas many authors have focussed on yield variables, we use the interest rate itself. This has the advantage of reducing the state space and introduces an interesting dynamic since the predictor itself is the return on an investable asset. We are also unlikely to lose much predictive power, since Ang and Bekaert (2000) find that the short rate is the most robust predictor of international stock return data, including the US. Goyal and Welch (1999) and Ang and Bekaert (2000) demonstrate that the dividend yield, which has been previously used by many authors to forecast returns, has no forecasting power when data of the late 1990's are added to the sample.

Our two DGP's for nominal data are special cases of a bivariate VAR on stock returns and interest rates:

$$X_t = c + \Phi X_{t-1} + \Sigma^{\frac{1}{2}} \epsilon_t, \quad (33)$$

where $X_t = (\tilde{y}_t r_t^f)'$, $\tilde{y}_t = y_t - r_{t-1}^f$ is the continuously compounded excess equity return and r_t^f is the risk-free rate, and $\epsilon_t \sim N(0, I)$.

The ‘‘No Predictability’’ model imposes all elements of Φ to equal zero except Φ_{22} , and the ‘‘Predictability’’ model constrains all elements of Φ except Φ_{12} and Φ_{22} to be zero. Estimates for these DGP's are reported in Table (2). In both systems, note the negative contemporaneous correlation between shocks to short rates and stock returns. The predictability system reveals the short rate to be a significant predictor of stock returns only in the post-1941 period. We will look at sensitivity with respect to the Φ_{12} parameter in Section 5.

We now proceed to derive optimal asset allocations for various parameter configurations under the two DGP's. Since the DGP's are first-order Markov processes, they lend themselves easily to discretization (See the Appendix). In the last sub-section, we consider a different exercise. We attempt to infer what equity premium is necessary to obtain a 60% optimal asset weight for equities. Benartzi and Thaler (1995) find that loss aversion utility, where losses are weighted 2.5 times more heavily than gains and where updating happens annually, results in a 60% optimal equity position, at the historical equity premium.

and Stambaugh (1996).

4.2 No Predictability Case

In this system, the excess premium is constant and IID, while short rates are autoregressive and negatively correlated with equity returns. For a given risk aversion, portfolio allocations in this system depend on the horizon, but they do not depend on the level of the short rate (as we will show later). In Figure (3) we see that moderately risk averse CRRA agents ($\gamma = 2$) should put close to 100% of their portfolio in equities.⁸ More realistic equity allocations start to appear at γ 's between 3 and 4, which is lower than is found in some recent literature (see further below). However, even for γ 's equalling 10, a corner solution of no equity holdings is still quite far away. Figure (3) appears to suggest there are hedging demands in that the equity proportion is larger for longer horizons and hence agents gradually decrease their equity proportions as they age, but the effect is very slight.

As we will confirm later in Figure (5), in the no predictability system the intertemporal hedging demand does not depend on the short rate but depends only on the horizon. This is not surprising given that our set-up is similar to that of Liu (1999). Liu proves this result analytically in a continuous-time problem with the short rate following a Vasicek (1977) model. Under the Vasicek term structure model, excess returns of bonds have a constant risk premium, constant volatilities, and are perfectly correlated with the short rate. Similarly, in our no-predictability system the excess returns of stocks have a constant risk premium and a constant volatility. Although in our setting the correlation between equities and the short rate is not unity, Liu's results obtain. However, Liu's results will not generalize to our predictable system, where short rates predict the conditional mean of stock returns.

The horizon effect in the no predictability system arises from the persistence of the short rate, and the correlation of short rate shocks with shocks to the excess returns. It is well known, from Samuelson (1991) and others, that processes with positive persistence will exhibit negative hedging demands (they are "riskier" over longer periods), whereas negatively correlated processes will exhibit positive hedging demands.

In our no predictability system, if the correlation between short rates and stock returns is zero then the stock return is IID and independent of the short rate. In this case, we are back in the Samuelson (1969) world and portfolio weights will be constant across all horizons. However, in our empirical estimates, shocks to stock returns and short rates are slightly negatively correlated

⁸Note that if (\tilde{y}_t, r_t^f) is normally distributed, DA utility (which includes CRRA utility) may not be strictly defined for any leveraged portfolio. In this case, the formal solution is a corner solution of 100% equity. This is because wealth has a possibility of going negative in some states of the world, and the utility may not be defined for these states. In our model, leveraged positions occur when investors are not very risk averse, for example when risk aversion is less than log.

(-0.0474). As Campbell and Viceira (1999) note, as the number of periods increases, the total portfolio becomes less volatile because of the negative correlation, and this enables an investor to hold a greater proportion of stocks as the horizon increases.

Figure (4) compares optimal asset allocation under DA preferences with CRRA preferences. We focus on the 3 month horizon here since intertemporal hedging demands are very small. An explicit discussion of hedging demands in the no predictability case versus the predictability case is deferred to the next Section. The various lines in the top panel of Figure (4) correspond to different γ 's, with higher γ 's leading to lower equity proportions. Going from left to right, we decrease A from 1 (which is CRRA utility) to 0.85, which represents very modest disappointment aversion. For $\gamma = 2$, dropping A to 0.85 is sufficient to bring the equity allocation close to 0.60. The effect on asset allocation of lower A is less dramatic for higher γ . This is most clearly illustrated in the bottom panel, where, for each γ , the equity proportions for different A 's are depicted. As γ increases, risk-aversion, which diminishes the attractiveness of stocks, is more and more driven by the curvature in the utility function rather than the additional distaste of disappointing outcomes.

4.3 Predictability Case

Figure (5) contrasts portfolio allocation in the no predictability case with the predictable system under CRRA preferences. In the predictable system, asset demand is a function of the interest rate. Hence, we graph optimal asset weights for various horizons as a function of the interest rate. The top panels consider $\gamma = 2$, and the bottom panel $\gamma = 5$. The equity premium is a negative function of the interest rate in this model, so the higher the interest rate, the lower the equity allocation. The effect is quite pronounced. For example, for $\gamma = 5$, optimal equity proportions decrease from around 40% for a 4% interest rate to around 15% for a 12% interest rate. This elasticity is primarily driven by the Φ_{12} parameter in equation (33). Note that the hedging demands are larger at higher interest rates, but similar to the no predictability case they are small. Like the no predictability system, portfolio weights increase with investment horizon in the predictability case because the shocks to stock returns and short rates are negatively correlated ($\rho = -0.0475$).

Figure (6) focuses on DA preferences. In the left column we see that, like CRRA utility, hedging demands are small. Graphs in the right column illustrate the state dependence for the three month horizons at γ equal to 2 and 5 for various A . The $A = 1$ case is the CRRA case. As with CRRA preferences, the curves slope down but at a lower level. Figure (6) shows that, similar to CRRA preferences the hedging demands are positive due to the negative correlation

of the shocks. Second, as the disappointment aversion increases (A decreases), investors hold a smaller equity position as more emphasis is placed on avoiding disappointing outcomes.

The graphs in Figure (6) prompt two questions. First, is the state dependence for DA preferences any different than for CRRA preferences? The graph suggests it may not be, with the shape of the lines looking very similar for $A = 1$ and $A = 0.9$. If this is the case, we may find DA outcomes using CRRA utility with a higher risk aversion coefficient. However, Figure (6) is indeed deceptive. Figure (7) vividly illustrates. For each short rate, we start from the optimal equity weight at a horizon of one quarter for a DA investor with $\gamma = 5$ and $A = 0.95$ or $A = 0.90$. We then find a CRRA investor, characterized by γ , that chooses the same portfolio. If the above claim were true, we should find a horizontal line. In contrast, the line starts out relatively flat but then ratchets upward non-linearly for higher short rates, so the aversion of the DA investor to stocks increases non-linearly with higher interest rates. As a consequence, the state dependence of DA preferences cannot be captured by a CRRA utility function. For lower A , the effect is very dramatic, as the bottom graph in Figure (7) illustrates. The intuition behind the shape of the plots in Figure (7) is that the equity return is lower for higher short rates. The higher the short rate the more stocks can disappoint, leading to lower equity holdings for DA investors relative to CRRA preferences and consequently higher implied CRRA risk aversions.

Second, comparing Figures (5) and (6), the hedging demands also look similar across CRRA and DA preferences. Again this is deceptive. Table (3) shows that the intertemporal hedging demands delivered by DA preferences cannot be mimicked by a CRRA utility function. The table presents the CRRA risk aversion parameter that would yield the same optimal equity demand for each horizon as is true for a particular DA investor. For the portfolio weights implied by DA utility with $\gamma = 5$ and $A = 0.95$ or $A = 0.85$, we find the γ a CRRA investor must have to choose the same portfolio at a 3 month horizon. The exercise is repeated for a 1 year horizon, rebalancing quarterly, and then a 5 and 10 year horizon. The longer the horizon, the less CRRA risk aversion is required to match the DA asset demand. For example, for DA utility with $\gamma = 5$ and $A = 0.85$, a γ of 7.39 is required to produce the DA portfolio weight at a 3 month horizon, but at a 10 year horizon, the required γ drops to 6.89. The intuition behind these results is that as the horizon increases, mean reversion allows the total portfolio to become less volatile. As stocks have less room to disappoint with increasing horizon, this leads to smaller implied CRRA risk aversion.

Overall, the effect of DA preferences on intertemporal hedging demands seems small compared to its direct effect on asset allocation, which would apply in a model with a constant investment opportunity set as well. Whereas the typical $\gamma = 2$ investor puts over 90% of his wealth in stocks (See Figure (4)), a mildly disappointment averse investor with $\gamma = 2$ and

$A = 0.85$ decreases his allocation to little over 60%. Despite the large equity premium, stocks may disappoint! Whereas the primary focus of the recent literature has been on the effects of predictability on portfolio choice, our results suggest the importance of understanding the investor's attitude towards risk. Consequently, it is encouraging to see related work such as Barberis, Huang and Santos (1999) who embed prospect theory in a dynamic portfolio choice model with consumption, or Campbell and Viceira (1999) who investigate the portfolio choice implications disentangling risk aversion from intertemporal substitution in an Epstein-Zin (1989) framework. Given that some researchers find much stronger intertemporal hedging demands than we document here (See Brennan, Schwartz and Lagnado (1997), Barberis (2000), Campbell and Viceira (1998, 1999), and Balduzzi and Lynch (1999)), we revisit this issue in Section 5 with a DGP exhibiting more pronounced predictability than the system we examined so far.

4.4 Disappointment Aversion and the Equity Premium

Our paper started out wondering why many rational people do not invest in the equity market. Empirically, institutional investors hold about 50% to 60% of their portfolios in equity. Suppose we want the optimal asset allocation weight between the equities and the one period bond to be 60%-40%, what is the equity premium that delivers this outcome for different utility functions and parameters? We focus on the post-1940 period and look at both one-period quarterly and annual horizons.⁹ The latter data frequency is used in Benartzi and Thaler (1995) and Gomes (2000) in a similar exercise. The equity premiums required to hold a 60%-40% portfolio are given in Figure (8). Given the substantial sampling error in estimating the equity premium, 95% confidence bands of the empirical equity premium stretch from less than 5%, almost 4%, to somewhat less than 12%. Under CRRA preferences, a γ of 4 is enough to just make the lower boundary (annual data) or to just barely miss it (quarterly data). That is, such investors require a premium of about 4% to invest 60% of their wealth in equity. Investors with relative risk aversion coefficients between 6 and 7, require a premium of about the same as observed in the data (8%) to invest 60% of their wealth in the stock market. Note that these results are for the no predictability case. Predictability typically makes equities more attractive at low interest rates.¹⁰

In the bottom row of Figure (8), we focus on DA utility with very low curvature, setting γ either equal to 1 (log-utility) or 2. For the quarterly horizon, even moderately disappointment

⁹VAR's are fitted separately for each horizon. For quarterly horizon estimates see Table (2).

¹⁰It is important to realize that this is not to be understood as an "explanation" for the equity premium, as Benartzi and Thaler (1995) do. We simply try to understand portfolio behavior in the presence of a large equity premium, and try to make inferences about risk preferences

averse investors (with $A = 0.75$) limit their equity investments to 60%, even when the equity premium is as large as observed in the data, or larger. The results for the annual horizon are somewhat weaker. Again, only modest disappointment aversion is required for a 60% equity investment to be consistent with an equity premium in the 95% confidence interval. However, for investors to require a premium close to 8%, the observed value A must be somewhat lower than 0.5. This result is very much consistent with Benartzi and Thaler (1995) who find investors to require the observed equity premium with losses being weighted 2.25 times as much as gains. Although our utility functions are not directly comparable, the discrete solution algorithm discussed in Section 3.2 shows that DA utility implicitly over-weights disappointing outcomes by $\frac{1}{A}$, implying that an informal translation of Benartzi and Thaler's 2.25 parameter is $A = \frac{1}{2.25} = 0.44$.

5 Robustness

Agents who face practical portfolio allocation problems must confront the problem of inflation, which becomes a significant factor over long horizons. In this Section we first consider the case where agents care about returns after inflation. This introduces another state variable, inflation, into our analysis and makes the return of both stocks and bonds conditionally stochastic, where in the nominal setting only the stock return was stochastic. We also look at the case where predictability in returns is stronger than over the full 1926-1998 sample. The predictability coefficient (Φ_{12} in the VAR in equation (33)) using data from 1941-1998 is much more negative and more significant than over the full sample (-1.3167 versus -0.6049). In both these cases, the qualitative effects of DA as compared to CRRA utility are the same as in our main analysis.

5.1 Inflation

To incorporate inflation we estimate a VAR of the following form:

$$\begin{pmatrix} \tilde{y}_t \\ \tilde{r}_t \\ p_t \end{pmatrix} = c + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \Phi_{22} & 0 \\ 0 & 0 & \Phi_{33} \end{pmatrix} \begin{pmatrix} \tilde{y}_{t-1} \\ \tilde{r}_{t-1} \\ p_{t-1} \end{pmatrix} + \Sigma^{\frac{1}{2}} \epsilon_t \quad (34)$$

where $\tilde{y}_t = y_t - r_t^f$ is the nominal (or real) excess return of stocks, $\tilde{r}_t = r_{t-1}^f - p_t$ are real bond returns and p_t is inflation. All variables are continuously compounded. This system is comparable to the nominal no predictability system because it retains constant excess returns of equity. The real return on equity is given by $(\tilde{y}_t + \tilde{r}_t)$. Estimates of the system using quarterly

data from 1928-1998 are given in Table (4). Real bond returns are much less persistent than their nominal counterparts (autocorrelation of 0.56), and inflation has an autocorrelation of 0.57. Excess stock returns show almost identical slight negative correlation with real bond return shocks as they did with nominal interest rate shocks (See Table (2)). Shocks to real bond returns and inflation are strongly negatively correlated (-0.92). The negative relationship of asset returns with inflation has been documented by several authors (See Hess and Lee (1999), and Boudoukh, Richardson and Whitelaw (1994) for recent summaries).

We look at optimal portfolio holdings over a one quarter horizon for the problem:

$$\max_{\alpha} U(\mu_{\tilde{W}}),$$

where $\mu_{\tilde{W}}$, the certainty equivalent for real wealth, is defined as in equation (4). End of period real wealth \tilde{W} is given by

$$\tilde{W} = \alpha \tilde{W}_0 (\exp(\tilde{y}_{t+1} + \tilde{r}_{t+1}) - \exp(\tilde{r}_{t+1})) + \tilde{W}_0 \exp(\tilde{r}_{t+1}).$$

where α is the equity portfolio weight. Both the real stock ($\tilde{y}_{t+1} + \tilde{r}_{t+1}$) and real bond return \tilde{r}_{t+1} are now stochastic whereas in the problem with nominal returns, the bond return was known at time t . We solve for α by simultaneously solving for the FOC's and the definition of DA utility (equations (15) and (16)) using a discretization method outlined in the Appendix.¹¹

The optimal portfolio weights for CRRA and DA utility are presented in Table (5) under the headings “real weights”. Under the heading “nominal weights” we list optimal equity holdings for the nominal no-predictability system discussed in Section 4.2. The portfolio weights do not depend on the level of inflation or real interest rates as in the no-predictability nominal system.

We first focus on CRRA utility. It is no surprise that inflation risk in both stocks and bonds increases the relative attractiveness of equity compared to the nominal system where bonds are risk-free. Table (5) shows that at $\gamma = 1$, an investor levers up to obtain an equity position larger than 100% of his wealth. Nevertheless, at $\gamma = 3$, we obtain a reasonable 64% equity position, compared to a 62% equity position in the nominal system. There is only a small difference (around 2%) between the nominal and the real positions.

The DA utility results confirm our previous findings, making equity less attractive as disappointment aversion increases. However, increasing disappointment aversion (lowering A) reduces equity holdings at a slower rate than was the case in the nominal system. Since bond returns can now also disappoint, equity is relatively more attractive. For example, when $\gamma = 1$, we need to decrease A to 0.70 before we obtain an optimal equity allocation around 70% for the

¹¹The bi-section algorithm presented in Section 3.2 cannot be used as there are now two stochastic assets.

real system. If the investor maximizes nominal wealth, she will hold an equity position around 55% at the same risk aversion and disappointment aversion. For $\gamma = 2$, a DA-investor with A equal to 0.75 will invest around 50% in the equity market taking into account inflation, whereas she will invest around 40% in the nominal system. Although the inflation effects decrease the impact of DA on equity allocation, our main results appear robust. It remains true that only modest amounts of disappointment aversion are required to substantially lower equity holdings for the same level of curvature in the utility function.

5.2 Stronger Predictability

Table (2) shows several differences between the coefficients of the quarterly predictability VAR over the full sample 1926-1998 compared to more recent 1941-1998 sample. First, conditional volatility for the excess returns is lower with more recent data (0.0750 versus 0.1094), and stock and bond returns are more negatively correlated (-0.17 versus -0.05). There has been a large change in the significance and magnitude of the predictability coefficient Φ_{12} from -0.6 to -1.3. This strong evidence of short-rate predictability using more recent data has been noted by many authors (See Patelis (1997), for example). The predictability of returns using the short rate decreases with horizon, and in the last column of the bottom panel of Table (2) we see that it is not significant using annual data.

Using the estimates of the quarterly VAR from 1941-1998 produces severely leveraged portfolios at low and high interest rates, similar to the highly leveraged positions found by authors who considered DGP's with strong predictability.¹² At low interest rates, excess returns are very high and agents want to short bonds and go long equity. At high interest rates, excess returns can be negative, so agents short equity and lever into bonds.

In Table (6) we present quarterly horizon portfolio weights for DA utility at the interest rate state corresponding to the unconditional mean of short rates over 1941-1998 (0.0487 annualized). At this short rate, excess returns are sufficiently attractive for investors with low γ to short bonds and go long equity. For $\gamma = 2$, a CRRA investor invests more than 200% of her portfolio in equity. Only for $\gamma = 7$ do we obtain an equity allocation of 60%. To obtain a 60% equity allocation, for a DA investor with $\gamma = 2$, now requires dropping A below 0.60. Nevertheless,

¹²Many of these authors use the dividend yield, over sample periods where dividend yield predictability was much stronger than what it is using more recent data. See Barberis (2000), Campbell and Viceira (1999, 1998), Kandel and Stambaugh (1996) and others. See the comments by Goyal and Welch (1999) on dividend predictability in more recent periods. Equity holdings using the dividend yield as a predictor also have much larger hedging demands than using the short rate as a predictor as in this study because of the large negative correlation between stock returns and the dividend yield.

the decreases in wealth allocated to equity as A is decreased, very much follow the pattern of Figure (4), but starting from a much higher initial CRRA allocation. Again our main results appear robust to stronger predictability.

6 Conclusions

In this article, we have used the disappointment aversion (DA) preference framework developed by Gul (1991) to look at the portfolio choice of US investors. Although DA preferences are very much related to loss aversion in that they treat gains and losses asymmetrically, they are fully axiomatically motivated and admit easy comparison with standard expected utility. From the perspective of the smooth concave nature of constant relative risk averse (CRRA) preferences, the behavior of many investors often appears puzzling. For example, the tendency of people to happily accept bets with small but almost certain losses, but a very small probability of very large gains (as in a lottery), but at the same time not to invest in the stock market is not consistent with standard preferences. By increasing the relative weight of bad outcomes by $1/A$, DA preferences can resolve this puzzle. DA investors may find lottery-type payoffs very attractive and stock market investments rather disappointing. Also, the curvature of CRRA utility makes investors unrealistically risk averse over large stakes, as discussed by Rabin (1999). This behavior can likewise be resolved by DA preferences.

By calibrating a number of data generating processes to actual US data on stock and bond returns, we find very reasonable portfolios for moderately disappointment averse investors with utility functions exhibiting quite low curvature. DA preferences affect intertemporal hedging demands and the state dependence of asset allocation in such a way as to not be replicable by a CRRA utility function with higher curvature. Furthermore, it is easy using these preferences as the benchmark to reconcile the large equity premium with a typical asset allocation to equities of about 60%. Our results are robust to considering stronger predictability over a more recent subsample of our data, and to incorporating inflation as another state variable.

There are a number of interesting avenues for future work. Disappointment averse agents will dislike negative skewness much more than standard CRRA agents. Hence, the regular occurrence of equity market crashes inducing such skewness may further scare investors away from equity investments or it may induce them to buy (costly) insurance against such crashes. This may account for the recent popularity of put-protected products which seem to have lured many investors into the stock market. In an international context, the occurrence of correlated bear markets (See Ang and Bekaert (1999), Longin and Solnik (1999), and Das and Uppal

(1999)) may induce home bias in asset preferences for disappointment averse investors. Disappointment aversion may help account for equity market non-participation if agents are very disappointment averse, and cross-sectional variation in portfolio holdings (See Heaton and Lucas (1999)). Although DA preferences yield portfolio allocations promisingly close to actual holdings, we must ultimately investigate whether DA preferences can be accommodated in an equilibrium model of risk.

A Appendix

A.1 First Order Conditions for DA Utility

We derive the FOC for the static DA utility problem. Given a random outcome W , the utility μ for DA preferences is defined by the following equation:

$$\mu^{1-\gamma} = \frac{1}{K} \left(AE(W^{1-\gamma} \mathbf{1}_{\{W>\mu\}}) + E(W^{1-\gamma} \mathbf{1}_{\{W\leq\mu\}}) \right), \quad (\text{A-1})$$

where $\mathbf{1}$ is an indicator function and the normalization constant K is given by:

$$K = AE(\mathbf{1}_{\{W>\mu\}}) + E(\mathbf{1}_{\{W\leq\mu\}}) = A Pr(W > \mu) + Pr(W \leq \mu).$$

For the portfolio problem studied in this paper, the random outcome is wealth W given by:

$$W = \exp(r_f) + \alpha(\exp(y) - \exp(r_f)) \equiv R_f + \alpha x_e,$$

where $x_e = (\exp(y) - \exp(r_f))$ represents the stochastic excess return on equity, r_f is the constant risk-free rate, and $R_f = \exp(r_f)$. Although we restrict attention to one risky asset, the analysis can easily be generalized to multiple assets. The formal static portfolio problem is:

$$\max_{\alpha} \mu.$$

Taking the derivative with respect to α of both sides of equation (A-1) we have:

$$\begin{aligned} (1-\gamma)\mu^{-\gamma} \frac{\partial \mu}{\partial \alpha} &= \frac{1}{K} \left(A \frac{\partial}{\partial \alpha} E(W^{1-\gamma} \mathbf{1}_{\{W>\mu\}}) + \frac{\partial}{\partial \alpha} E(W^{1-\gamma} \mathbf{1}_{\{W\leq\mu\}}) \right) \\ &\quad - \frac{AE(W^{1-\gamma} \mathbf{1}_{\{W>\mu\}}) + E(W^{1-\gamma} \mathbf{1}_{\{W\leq\mu\}})}{K^2} \left(A \frac{\partial}{\partial \alpha} E(\mathbf{1}_{\{W>\mu\}}) + \frac{\partial}{\partial \alpha} E(\mathbf{1}_{\{W\leq\mu\}}) \right) \\ &= \frac{1}{K} \left(A \frac{\partial}{\partial \alpha} E(W^{1-\gamma} \mathbf{1}_{\{W>\mu\}}) + \frac{\partial}{\partial \alpha} E(W^{1-\gamma} \mathbf{1}_{\{W\leq\mu\}}) \right) \\ &\quad - \frac{\mu^{1-\gamma}}{K} \left(A \frac{\partial}{\partial \alpha} E(\mathbf{1}_{\{W>\mu\}}) + \frac{\partial}{\partial \alpha} E(\mathbf{1}_{\{W\leq\mu\}}) \right). \end{aligned} \quad (\text{A-2})$$

Note that by definition we have:

$$\begin{aligned} E(W^{1-\gamma} \mathbf{1}_{\{W>\mu\}}) &= \int_{W>\mu} f(x_e) (R_f + \alpha x_e)^{1-\gamma} dx_e, \\ E(W^{1-\gamma} \mathbf{1}_{\{W\leq\mu\}}) &= \int_{W\leq\mu} f(x_e) (R_f + \alpha x_e)^{1-\gamma} dx_e, \\ E(\mathbf{1}_{\{W>\mu\}}) &= \int_{W>\mu} f(x_e) dx_e, \\ E(\mathbf{1}_{\{W\leq\mu\}}) &= \int_{W\leq\mu} f(x_e) dx_e, \end{aligned} \quad (\text{A-3})$$

where $f(x_e)$ is the probability density function of x_e .

This implies:

$$\begin{aligned}
\frac{\partial}{\partial \alpha} \mathbb{E}(W^{1-\gamma} \mathbf{1}_{\{W > \mu\}}) &= (1-\gamma) \int_{\frac{\mu - R_f}{\alpha}}^{\infty} f(x_e) (R_f + \alpha x_e)^{-\gamma} x_e dx_e - f\left(\frac{\mu - R_f}{\alpha}\right) \mu^{1-\gamma} \frac{\partial}{\partial \alpha} \frac{\mu - R_f}{\alpha} \\
&= (1-\gamma) \mathbb{E}(W^{-\gamma} \mathbf{1}_{\{W > \mu\}}) - f\left(\frac{\mu - R_f}{\alpha}\right) \mu^{1-\gamma} \frac{\partial}{\partial \alpha} \frac{\mu - R_f}{\alpha} \\
\frac{\partial}{\partial \alpha} \mathbb{E}(W^{1-\gamma} \mathbf{1}_{\{W \leq \mu\}}) &= (1-\gamma) \int_{\infty}^{\frac{\mu - R_f}{\alpha}} f(x_e) (R_f + \alpha x_e)^{-\gamma} x_e dx_e + f\left(\frac{\mu - R_f}{\alpha}\right) \mu^{1-\gamma} \frac{\partial}{\partial \alpha} \frac{\mu - R_f}{\alpha} \\
&= (1-\gamma) \mathbb{E}(W^{-\gamma} \mathbf{1}_{\{W \leq \mu\}}) + f\left(\frac{\mu - R_f}{\alpha}\right) \mu^{1-\gamma} \frac{\partial}{\partial \alpha} \frac{\mu - R_f}{\alpha} \\
\frac{\partial}{\partial \alpha} \mathbb{E}(\mathbf{1}_{\{W > \mu\}}) &= -f\left(\frac{\mu - R_f}{\alpha}\right) \mu^{1-\gamma} \frac{\partial}{\partial \alpha} \frac{\mu - R_f}{\alpha} \\
\frac{\partial}{\partial \alpha} \mathbb{E}(\mathbf{1}_{\{W \leq \mu\}}) &= f\left(\frac{\mu - R_f}{\alpha}\right) \mu^{1-\gamma} \frac{\partial}{\partial \alpha} \frac{\mu - R_f}{\alpha} \tag{A-4}
\end{aligned}$$

Note that the $f((\mu - R_f)/\alpha)$ terms in the above equations come from the taking the derivative of the limit of the integrals.

Using equations (A-3) and (A-4) we have:

$$\begin{aligned}
&\left(A \frac{\partial}{\partial \alpha} \mathbb{E}(W^{1-\gamma} \mathbf{1}_{\{W > \mu\}}) + \frac{\partial}{\partial \alpha} \mathbb{E}(W^{1-\gamma} \mathbf{1}_{\{W \leq \mu\}}) \right) \\
&= (1-\gamma) (A \mathbb{E}(W^{-\gamma} x_e \mathbf{1}_{\{W > \mu\}}) + \mathbb{E}(W^{-\gamma} x_e \mathbf{1}_{\{W \leq \mu\}})) \\
&\quad + (A-1) f\left(\frac{\mu - R_f}{\alpha}\right) \mu^{1-\gamma} \frac{\partial}{\partial \alpha} \frac{\mu - R_f}{\alpha}, \tag{A-5}
\end{aligned}$$

and

$$\left(A \frac{\partial}{\partial \alpha} \mathbb{E}(\mathbf{1}_{\{W > \mu\}}) + \frac{\partial}{\partial \alpha} \mathbb{E}(\mathbf{1}_{\{W \leq \mu\}}) \right) = (A-1) f\left(\frac{\mu - R_f}{\alpha}\right) \frac{\partial}{\partial \alpha} \frac{\mu - R_f}{\alpha}. \tag{A-6}$$

Using the above two equations and equation (A-2), the FOC can be written as:

$$\begin{aligned}
&\frac{1}{K} (A \mathbb{E}(W^{-\gamma} x_e \mathbf{1}_{\{W > \mu\}}) + \mathbb{E}(W^{-\gamma} x_e \mathbf{1}_{\{W \leq \mu\}})) \\
&+ \frac{1}{K} (A-1) \mu^{1-\gamma} f\left(\frac{\mu - R_f}{\alpha}\right) \frac{\partial}{\partial \alpha} \frac{\mu - R_f}{\alpha} - \frac{1}{K} (A-1) \mu^{1-\gamma} f\left(\frac{\mu - R_f}{\alpha}\right) \frac{\partial}{\partial \alpha} \frac{\mu - R_f}{\alpha} = 0
\end{aligned}$$

or,

$$\frac{1}{K} (A \mathbb{E}(W^{-\gamma} x_e \mathbf{1}_{\{W > \mu\}}) + \mathbb{E}(W^{-\gamma} x_e \mathbf{1}_{\{W \leq \mu\}})) = 0. \tag{A-7}$$

Therefore, in the FOC, we can ignore the α dependence in the indicator functions, both in the numerator and in the denominator. The intuition for this is clear. From the definition of the utility equivalent

$$\mu^{1-\gamma} = \frac{A \mathbb{E}(W^{1-\gamma} \mathbf{1}_{\{W > \mu\}}) + \mathbb{E}(W^{1-\gamma} \mathbf{1}_{\{W \leq \mu\}})}{A \mathbb{E}(\mathbf{1}_{\{W > \mu\}}) + \mathbb{E}(\mathbf{1}_{\{W \leq \mu\}})},$$

we see that if W were constant, μ would be a constant. When taking the derivative with respect to the indicator function, it is zero except at $W = \mu$. So when taking the derivative of the indicator functions in the numerator with respect to α , we can treat W as if it were a constant. Hence all the derivatives with respect to the indicator functions add up to zero.

A.2 Estimation of Data Generating Processes

We estimate the following Vector Autoregression (VAR):

$$X_t = c + \Phi X_{t-1} + u_t \quad (\text{A-8})$$

where u_t IID $N(0, \Sigma)$. For the nominal systems $X_t = (\tilde{y}_t r_t)$ where $\tilde{y}_t = y_t - r_{t-1}^f$ is the (real or nominal) equity excess return, and r_t is the short rate. For the real systems with inflation $X_t = (\tilde{y}_t \tilde{r}_t p_t)$ where \tilde{y}_t is the real equity excess return, $\tilde{r}_t = r_{t-1}^f - p_t$ is the real bond return and p_t is the inflation rate. We discuss the estimation for the nominal systems, as the real system estimation is similar.¹³

The system without predictability has $\Phi = \begin{pmatrix} 0 & 0 \\ 0 & \rho \end{pmatrix}$ and with predictability $\Phi = \begin{pmatrix} 0 & b \\ 0 & \rho \end{pmatrix}$.

Equation (A-8) can be written in compact form as:

$$X = B * Z + U \quad (\text{A-9})$$

where $X = (X_1 \dots X_T)$ ($2 \times T$), $B = [c \Phi]$ (2×3), $U = (u_1 \dots u_T)$ ($2 \times T$), $Z = (z_0 \dots z_{T-1})$ ($3 \times T$) with $z_t = [1 X_t']$ (3×1). The restrictions are written as $R\beta = r$ with $\beta = \text{vec}(B)$. The unrestricted maximum likelihood estimator, where Φ is unconstrained is given by:

$$\hat{\beta} = ((ZZ')^{-1}Z \otimes I)Y,$$

where $Y = \text{vec}(X)$. The restricted maximum likelihood estimator is given by:

$$\hat{\beta}^c = \hat{\beta} + ((ZZ')^{-1} \otimes I) R' (R((ZZ')^{-1} \otimes I)R')^{-1} (r - R\hat{\beta}). \quad (\text{A-10})$$

and $\hat{B} = \text{devec}(\hat{\beta}^c)$, as in Lütkepohl (1993).

The estimate of Σ is given by $\hat{\Sigma} = 1/T (\hat{U}'\hat{U})$, where $\hat{U} = X - \hat{B}Z$. The estimated covariance of $\hat{\beta}_c$ is given by:

$$\widehat{\text{cov}}(\hat{\beta}_c) = \Gamma \otimes \hat{\Sigma} - (\Gamma \otimes \hat{\Sigma})R' \left(R(\Gamma \otimes \hat{\Sigma})R' \right)^{-1} R(\Gamma \otimes \hat{\Sigma}) \quad (\text{A-11})$$

where $\Gamma = (ZZ')^{-1}$. The estimated covariance of $\text{vech}(\hat{\Sigma})$ is given by:

$$\widehat{\text{cov}}(\text{vech}(\hat{\Sigma})) = \frac{2}{T} D^{-1} \left(\hat{\Sigma} \otimes \hat{\Sigma} \right) (D^{-1})' \quad (\text{A-12})$$

where D^{-1} is the Moore-Penrose inverse of D , the duplication matrix which makes $\text{vec}(C) = D \text{vech}(C)$ for a symmetric matrix C .

A.3 Discretization of VAR's

We construct an approximate discrete Markov chain to the VAR in equation (A-8) using the quadrature-based methods of Tauchen (1990) and Tauchen and Hussey (1991). We first detail the Markov discretization for the nominal systems (with excess equity return and short rates), and then discuss the extension to the real systems (which also include inflation).

For the system for $X_t = (\tilde{y}_t r_t)$, with $\tilde{y}_t = y_t - r_{t-1}^f$ the excess equity return and r_t^f the short rate, \tilde{y}_t may be dependent on lagged r_t^f but not vice-versa, so r_t^f is the driving variable in the system. We choose $N = 50$ points for the short rate over a uniform grid and denote these as $\{r_i^f\}$. The short rate is very persistent, so many points

¹³The optimal lag choice in all systems by the Bayesian Information Criteria (BIC) is 1 lag.

are necessary for an accurate approximation (See Tauchen and Hussey (1991)). We use a uniform grid because points chosen by Gaussian-Hermite quadrature perform poorly in optimization as they are too widely spaced. We construct the transition probabilities Π_r ($N \times N$) for going from state r_i^f to r_j^f , $1 \leq i, j \leq N$ by evaluating the conditional density of r_j^f (which is conditionally Normal) and then normalizing the densities so that they sum to unity. This is the driving process of the discretized system.

We choose $M = 30$ discrete states for \tilde{y}_t . These states are chosen using Gaussian-Hermite points approximating the unconditional distribution of \tilde{y}_t implied by equation (A-8). To include \tilde{y}_t in the discretization we note that for each state r_i^f , an $N \times M$ vector π_i can be constructed giving the transition probabilities going from state r_i^f ($1 \leq j \leq N$) to (r_j^f, \tilde{y}_j) ($1 \leq j \leq N \times M$). The distribution of \tilde{y}_t conditional on r_i^f is Normal, and constructed by evaluating the distribution of \tilde{y}_t conditional on r_i^f for going from state r_i^f to state (r_j^f, \tilde{y}_j) . A Choleski decomposition is used to take account of the contemporaneously correlated error terms u_t in equation (A-8). The vectors π_i can be stacked to give a $N \times NM$ probability transition matrix Π_{ry} giving the probabilities from $\{r_i^f\}$, $1 \leq i \leq N$ to $\{r_j^f, \tilde{y}_j\}$, $1 \leq j \leq NM$. The Markov chain constructed this way matches first and second moments of the VAR in equation (A-8) to 3-4 significant figures. It is possible to also construct a square Π matrix, but this matrix will have repeated rows.

For the inflation systems, our state variables are the real bond return \tilde{r}_t and the inflation rate p_t . We choose $N = 20$ points for each of \tilde{r}_t and p_t to construct a $N^2 \times N^2$ transition matrix Π_{rp} for the states $\{\tilde{r}_i, p_i\}$, $1 \leq i \leq N^2$, in a similar manner to Π_r above. These points are each chosen on a uniform grid. We choose $M = 5$ discrete points for the real equity return \tilde{y}_t using Gaussian quadrature. Then a $N^2 \times N^2 M$ transition matrix Π_{rpy} can be constructed similar to Π_{ry} .

References

- [1] Ang, A., and G. Bekaert, 2000, "Stock Return Predictability - Does it Exist?", working paper.
- [2] Ang, A., and G. Bekaert, 1999, "International Asset Allocation with Time-Varying Correlations", NBER working paper 7056.
- [3] Balduzzi, P., and A. W. Lynch, 1999, "Transaction Costs and Predictability: Some Utility Cost Calculations", *Journal of Financial Economics*, 52, 47-78.
- [4] Barberis, N., 2000, "Investing for the Long Run when Returns are Predictable", *Journal of Finance*, 55, 1, 225-264.
- [5] Barberis, N., M. Huang, and J. Santos, 1999, "Prospect Theory and Asset Prices", NBER working paper 7220.
- [6] Bekaert, G., R. J. Hodrick, and D. A. Marshall, 1997, "The Implications of First-Order Risk Aversion for Asset Market Risk Premiums", *Journal of Monetary Economics*, 40, 3-39.
- [7] Benartzi, S., and R. Thaler, 1995, "Myopic Loss Aversion and the Equity Premium Puzzle", *Quarterly Journal of Economics*, 110, 73-92.
- [8] Berkelaar, A., and R. Kouwenberg, 2000, "Optimal Portfolio Choice Under Loss Aversion", working paper.
- [9] Boudoukh, J., M. Richardson, and R. F. Whitelaw, 1994, "Industry Returns and the Fisher Effect", *Journal of Finance*, 49, 1595-1615.
- [10] Brandt, M., 1999, "Estimating Portfolio and Consumption Choice: A Conditional Euler Equations Approach", *Journal of Finance*, 54, 5, 1609-1645.
- [11] Brennan, M. J., E. S. Schwartz, and R. Lagnado, 1997, "Strategic Asset Allocation", *Journal of Economic Dynamics and Control*, 21, 1377-1403.
- [12] Campbell, J. Y., and L. M. Viceira, 1999, "Consumption and Portfolio Decisions When Expected Returns are Time Varying", forthcoming *Quarterly Journal of Economics*.
- [13] Campbell, J. Y., and L. M. Viceira, 1998, "Who Should Buy Long-Term Bonds?", working paper.
- [14] Das, S. J., and R. Uppal, 1999, "The Effect of Systemic Risk on International Portfolio Choice", working paper.
- [15] Epstein, L. G., and S. E. Zin, 1991, "The Independence Axiom and Asset Returns", NBER Technical Working Paper 109.
- [16] Epstein, L. G., and S. E. Zin, 1990, " 'First-Order' Risk Aversion and the Equity Premium Puzzle", *Journal of Monetary Economics*, 387-407.
- [17] Epstein, L. G., and S. E. Zin, 1989, "Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns", *Econometrica*, 57, 937-969.
- [18] Gomes, F. J., 2000, "Loss Aversion and the Demand for Risky Assets", working paper.
- [19] Goyal, A., and I. Welch, 1999, "The Myth of Predictability: Does the Dividend Yield Forecast the Equity Premium?", working paper.
- [20] Gul, F., 1991, "A Theory of Disappointment Aversion", *Econometrica*, 59, 3, 667-686.
- [21] Heaton, J., and D. J. Lucas, 1999, "Portfolio Choice in the Presence of Background Risk," working paper.

- [22] Heaton, J., and D. J. Lucas, 1996, "Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing", *Journal of Political Economy*, 104, 31, 443-487.
- [23] Hess, P. J., and B. S. Lee, 1999, "Stock Returns and Inflation with Supply and Demand Disturbances", *Review of Financial Studies*, 12, 5, 1203-1218.
- [24] Kahneman, D., and A. Tversky, 1979, "Prospect Theory: An Analysis of Decision Under Risk", *Econometrica*, 47, 263-291.
- [25] Kandel, S., and R. F. Stambaugh, 1996, "On the Predictability of Stock Returns: An Asset-Allocation Perspective", *Journal of Finance*, 51, 2, 385-424.
- [26] Liu, J., 1999, "Portfolio Selection in Stochastic Environments", working paper.
- [27] Longin, F., and B. Solnik, 1999, "Correlation Structure of International Equity Markets During Extremely Volatile Periods", working paper.
- [28] Lütkepohl, H., 1993, *Introduction to Time Series Analysis*, 2nd edition.
- [29] Maenhout, P. J., 1999, "Robust Portfolio Rules and Asset Pricing", working paper.
- [30] Mankiw, N. G., and S. P. Zeldes, 1991, "The Consumption of Stockholders and Nonstockholders", *Journal of Financial Economics*, 29, 1, 97-112.
- [31] Mehra, R., and E. C. Prescott, 1985, "The Equity Premium: A Puzzle", *Journal of Monetary Economics*, 145-61.
- [32] Patelis, A. D., 1997, "Stock Return Predictability and the Role of Monetary Policy", *Journal of Finance*, 52, 5, 1951-72.
- [33] Rabin, M., 1999, "Risk Aversion and Expected Utility Theory: A Calibration Theorem", *working paper*.
- [34] Roy, A. D., 1952, "Safety First and the Holding of Assets", *Econometrica*, 20, 431-449.
- [35] Samuelson, P. A., 1969, "Lifetime Portfolio Selection by Dynamic Stochastic Programming", *Review of Economics and Statistics*, 51, 3, 239-246.
- [36] Samuelson, P. A., 1991, "Long-Run Risk Tolerance When Equity Returns are Mean Regressing: Pseudoparadoxes and Vindication of Businessmen's Risk", in Brainard, W. C., W. D. Nordhaus, and H. W. Watts, eds., *Money, Macroeconomics and Economic Policy*, MIT Press, 181-200.
- [37] Stutzer, M. A., 1999, "A Portfolio Performance Index and its Implications", working paper.
- [38] Tauchen, G., and R. Hussey, 1991, "Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models", *Econometrica*, 59, 2, 371-396.
- [39] Vasicek, O., 1977, "An Equilibrium Characterization of the Term Structure", *Journal of Financial Economics*, 5, 177-188.
- [40] Vissing-Jørgensen, A., 1997, "Limited Stock Market Participation," working paper.

Table 1: Summary Statistics of Data

Nominal Data

		stock	bond	excess
1926-1998	mean	0.1063	0.0408	0.0655
	std	0.2193	0.0173	0.2197
	ρ	-0.0575	0.9273	-0.0532
1941-1998	mean	0.1299	0.0487	0.0812
	std	0.1494	0.0170	0.1520
	ρ	0.0861	0.9191	0.1142

Real Data

		real stock	real bond	inflation
1926-1998	mean	0.0758	0.0103	0.0305
	std	0.2205	0.0273	0.0273
	ρ	-0.0060	0.5613	0.5742

All data is quarterly. Stock data is S&P 500 returns, with dividends. The bond data are 90 day bond returns and 1 year bond returns, both from CRSP. Nominal excess returns refer to stock returns in excess of bond returns. Inflation is changes in CPI. All returns (and inflation) are compounded logarithmic price changes. The data is annualized by multiplying the mean by 4, the standard deviation by 2. The first autocorrelation is denoted by ρ .

Table 2: VAR Estimates

No Predictability System		
	1926-1998	1941-1998
c_1	0.0161 (0.0064)	0.0202 (0.0087)
c_2	0.0008 (0.0003)	0.0011 (0.0004)
Φ_{22}	0.9273 (0.0219)	0.9156 (0.0257)
σ_1	0.1095 (0.0032)	0.0758 (0.0033)
σ_2	0.0015 (0.0000)	0.0010 (0.0000)
ρ	-0.0474 (0.0585)	-0.1674 (0.0645)
Predictability System		
	1926-1998	1941-1998
c_1	0.0223 (0.0099)	0.0362 (0.0087)
c_2	0.0008 (0.0003)	0.0011 (0.0004)
Φ_{12}	-0.6049 (0.7416)	-1.3167 (0.5867)
Φ_{22}	0.9273 (0.0219)	0.9156 (0.0261)
σ_1	0.1094 (0.0032)	0.0750 (0.0033)
σ_2	0.0015 (0.0000)	0.0010 (0.0000)
ρ	-0.0475 (0.0585)	-0.1693 (0.0645)

The model is:

$$X_t = \mu + \Phi X_{t-1} + \Sigma^{\frac{1}{2}} \epsilon_t$$

with $X_t = (y_t \ r_t)'$, y_t the excess one-period return, and r_t the short rate. All elements of Φ are constrained to be zero except for Φ_{22} in the No Predictability System. In the Predictability System Φ_{11} and Φ_{21} are constrained to be zero. The correlation between the errors of y_t and r_t is denoted ρ .

Table 3: Implied CRRA Risk Aversion for DA Hedging Demands

Horizon	DA $\gamma = 5$	
	$A = 0.95$	$A = 0.80$
3 months	5.57	7.39
1 year	5.51	7.21
5 years	5.36	6.92
10 years	5.33	6.89

Implied CRRA risk aversion for DA with $\gamma = 5$ and various A for the system with no predictability. The optimal T horizon portfolio weights for dynamically rebalanced DA utility are calculated, and then we solve for the implied CRRA risk aversion to give the same portfolio weight as the DA utility.

Table 4: Inflation System VAR Estimates

	Estimate	SE
c_1	0.0161	0.0064
c_2	0.0011	0.0007
c_3	0.0033	0.0007
Φ_{22}	0.5608	0.0279
Φ_{33}	0.5735	0.0276
σ_1	0.1095	0.0015
σ_2	0.0113	0.0000
σ_3	0.0112	0.0000
ρ_{12}	-0.0375	0.0585
ρ_{13}	0.0110	0.0586
ρ_{23}	-0.9203	0.0090

The model is:

$$X_t = c + \Phi X_{t-1} + \Sigma^{\frac{1}{2}} \epsilon_t$$

with $X_t = (y_t \tilde{r}_t p_t)'$, y_t the nominal excess equity return, and \tilde{r}_t the real bond return and p_t the inflation rate. All elements of Φ are constrained to be zero except for Φ_{22} and Φ_{33} (so there is no predictability of equity). The correlation between the errors of variable i and j is denoted ρ_{ij} .

Table 5: Portfolio Weights: Nominal versus Inflation

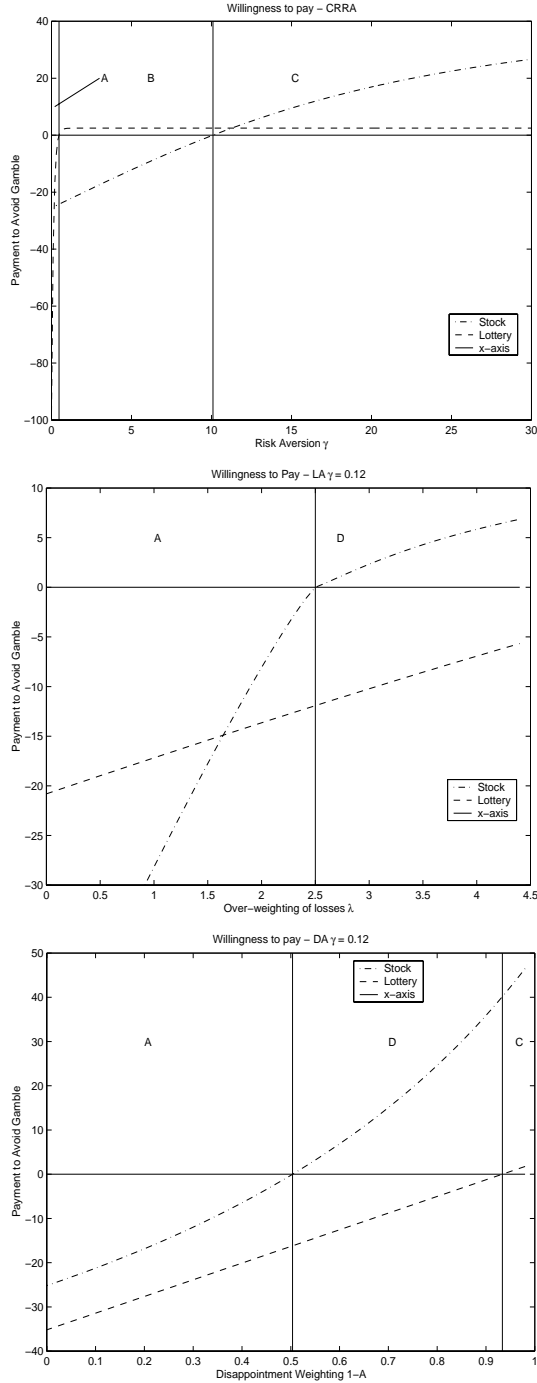
γ	CRRA Utility		A	DA Utility					
	real weight	nominal weight		$\gamma = 1$		$\gamma = 2$		$\gamma = 3$	
				Real	Nominal	Real	Nominal	Real	Nominal
1	1.8835	1.8274	0.95	1.7244	1.6510	0.8757	0.8331	0.5868	0.5550
2	0.9580	0.9270	0.90	1.5539	1.4608	0.7888	0.7337	0.5268	0.4884
3	0.6402	0.6181	0.85	1.3709	1.2559	0.6970	0.6283	0.4655	0.4181
4	0.4807	0.4632	0.80	1.1747	1.0354	0.5998	0.5165	0.4007	0.3436
5	0.3851	0.3703	0.75	0.9641	0.7984	0.4966	0.3976	0.3320	0.2645
6	0.3214	0.3084	0.70	0.7382	0.5443	0.3867	0.2710	0.2594	0.1803
7	0.2759	0.2642							

We present the proportion of the investor's portfolio invested in equity for the system with real equity and real bond returns (Table (4)). These are presented as "real weights". We compare these weights with the weights from a nominal system (with non-stochastic bond returns) for the system in Table (2). These are referred to as "nominal weights". The investment horizon is one quarter in both cases.

Table 6: Portfolio Weights with Stronger Predictability

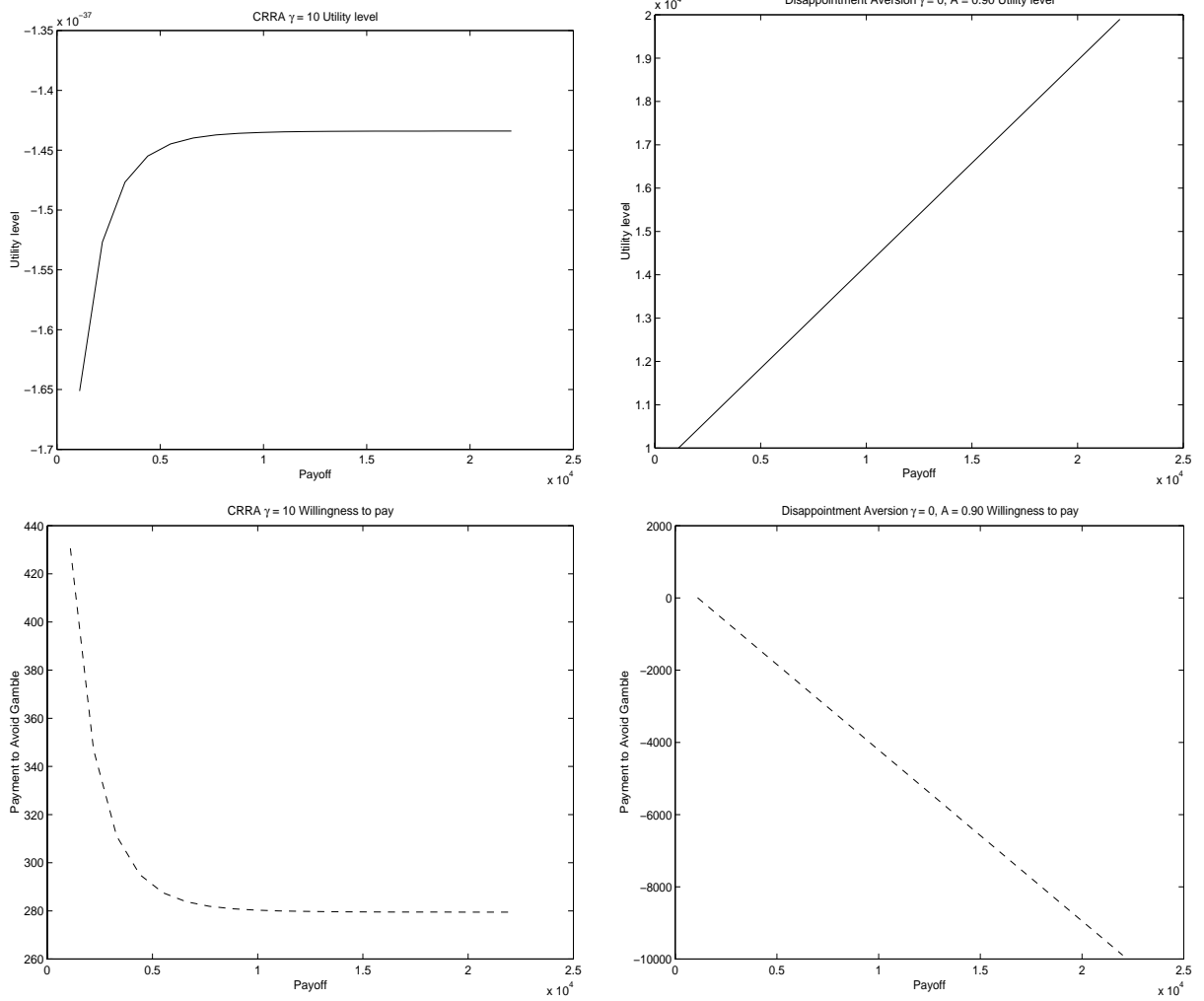
A	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$	$\gamma = 5$	$\gamma = 6$	$\gamma = 7$	$\gamma = 8$
1	3.8325	2.0319	1.3649	1.0254	0.8207	0.6839	0.5861	0.5128
0.9	3.3975	1.7674	1.1830					
0.8	2.8712	1.4656	0.9781					
0.7	2.2223	1.1182	0.7446					
0.6	1.4317	0.7140	0.4750					
0.5	0.4771	0.2378	0.1583					

We present DA portfolio weights for the period 1941-1998 using quarterly data at a short rate state equal to the unconditional mean of T-bill returns over this period (0.0487). Our horizon is one quarter. The case of CRRA is given with $A = 1$.



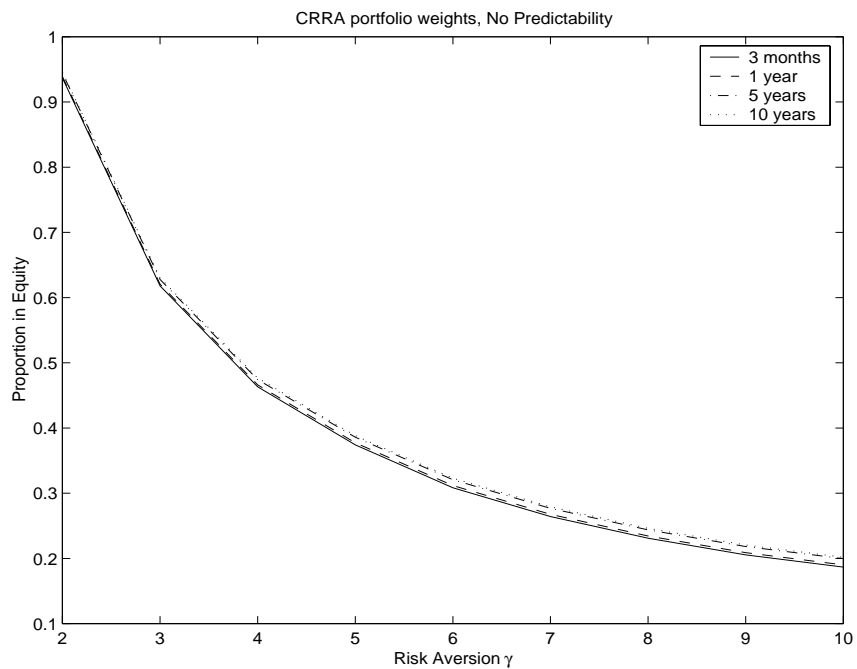
Willingness-to-pay amounts for CRRA (top plot) as a function of risk aversion γ , for loss aversion (LA) (middle plot) with $\gamma = 0.12$ as a function of loss weight λ , and for disappointment aversion (DA) (bottom plot) with $\gamma = 0.12$ as a function of disappointment level A . The willingness to pay (to avoid the stock or lottery gamble) is the difference between the certain wealth the investors have available by not taking on the gamble minus the certainty equivalent of the gamble. We divide the graphs into 4 possible areas: Area A is where agents like lotteries and stocks, Area B is where agents dislike lotteries but like stocks, Area C is where agents dislike both lotteries and stocks, and Area D is where agents like lotteries and dislike stocks.

Figure 1: Willingness to Pay



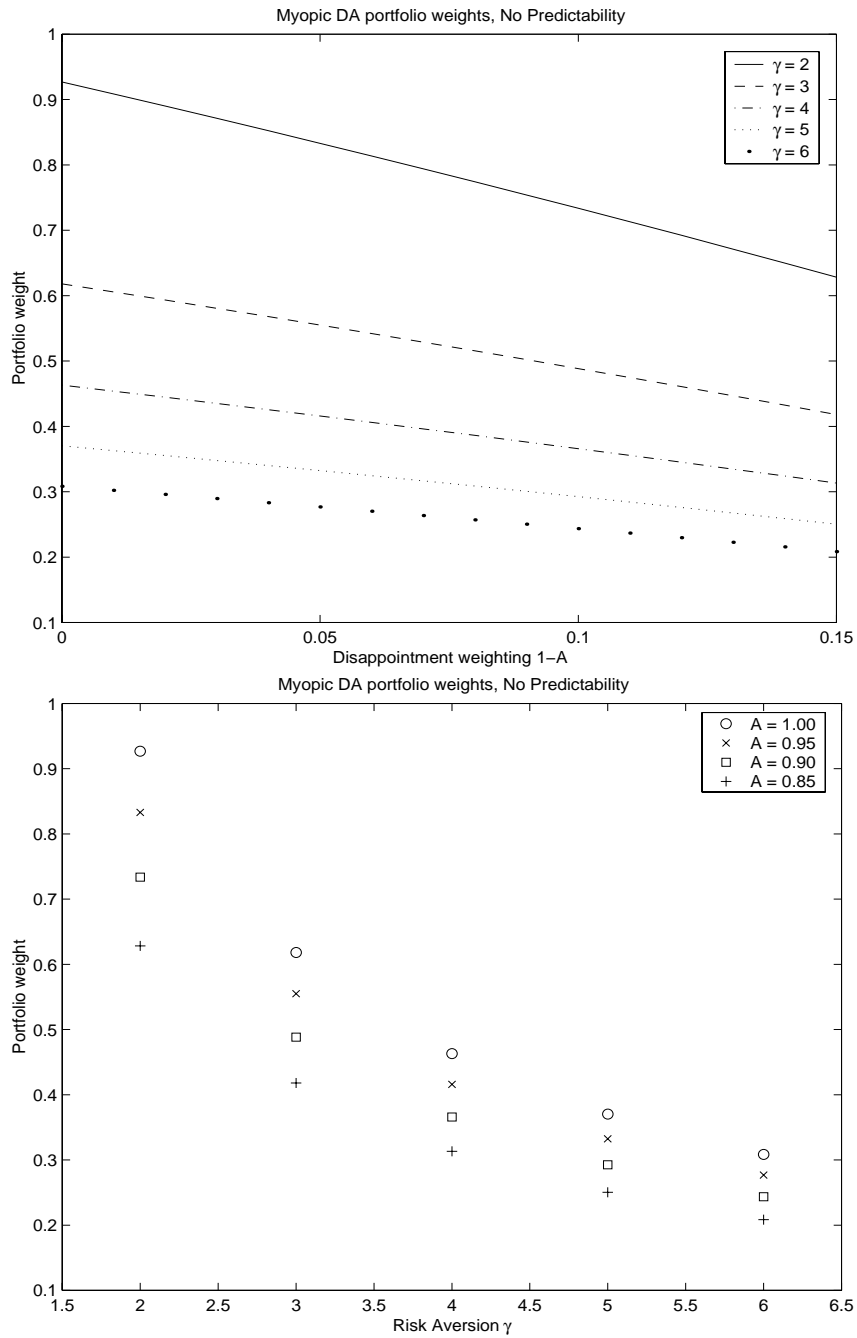
Utility levels (top row) and willingness to pay (bottom row) for a gamble with the following characteristics: an initial wealth of \$10,000, a 50% probability of a loss of \$1000 and a 50% probability of a gain represented on the x -axis. The left side considers CRRA utility with $\gamma = 10$, and the right side DA utility with $\gamma = 0$ and $A = 0.90$.

Figure 2: Rabin Gambles



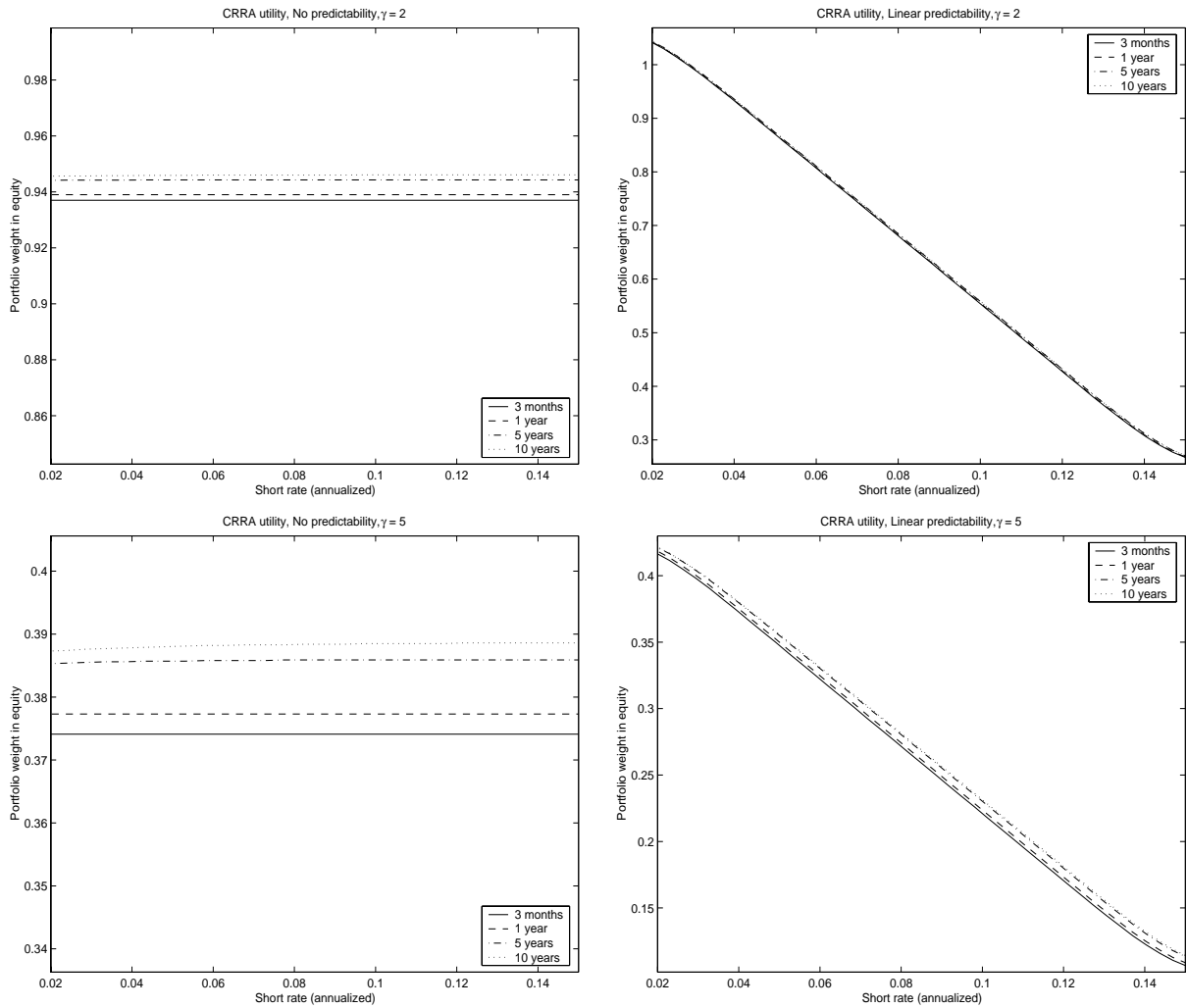
CRRA utility weights for the system with no predictability as a function of risk aversion γ . The portfolio weight is on the y -axis. We show the portfolio weight for a horizon of 3 months, 1 year, 5 year and 10 years (dynamically rebalancing each quarter) is shown on the vertical axis. Portfolio weights are the same for all short rate states.

Figure 3: CRRA Utility Weights for No Predictability



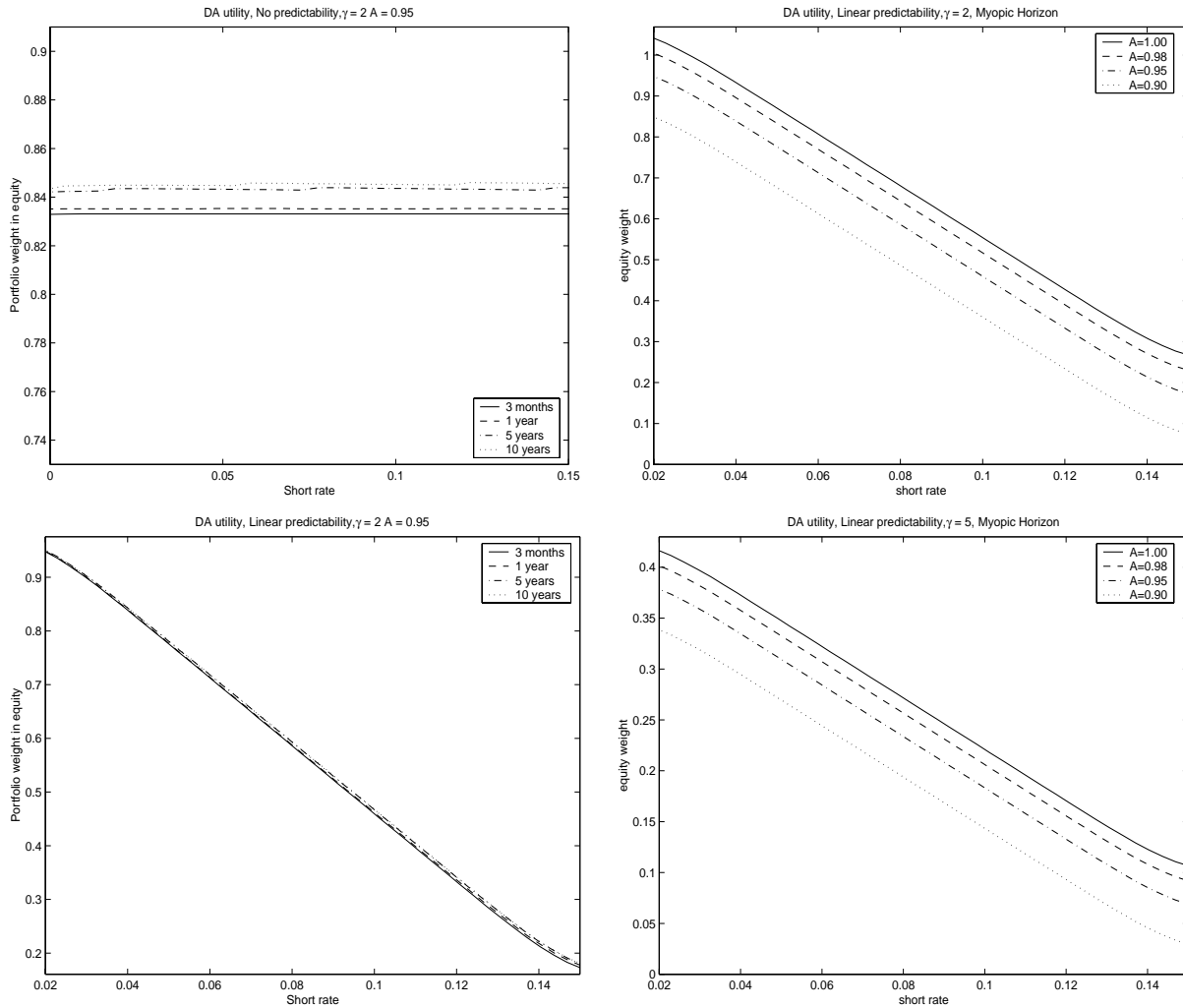
Quarterly myopic DA utility weights for the system with no predictability. In the top plot portfolio weights appear as a function of disappointment level A with various γ . The bottom plot shows the portfolio weights as a function risk aversion γ with various A . Portfolio weights are the same for all short rate states. The case of $A = 1$ corresponds to CRRA utility.

Figure 4: Myopic DA Utility Portfolio Weights



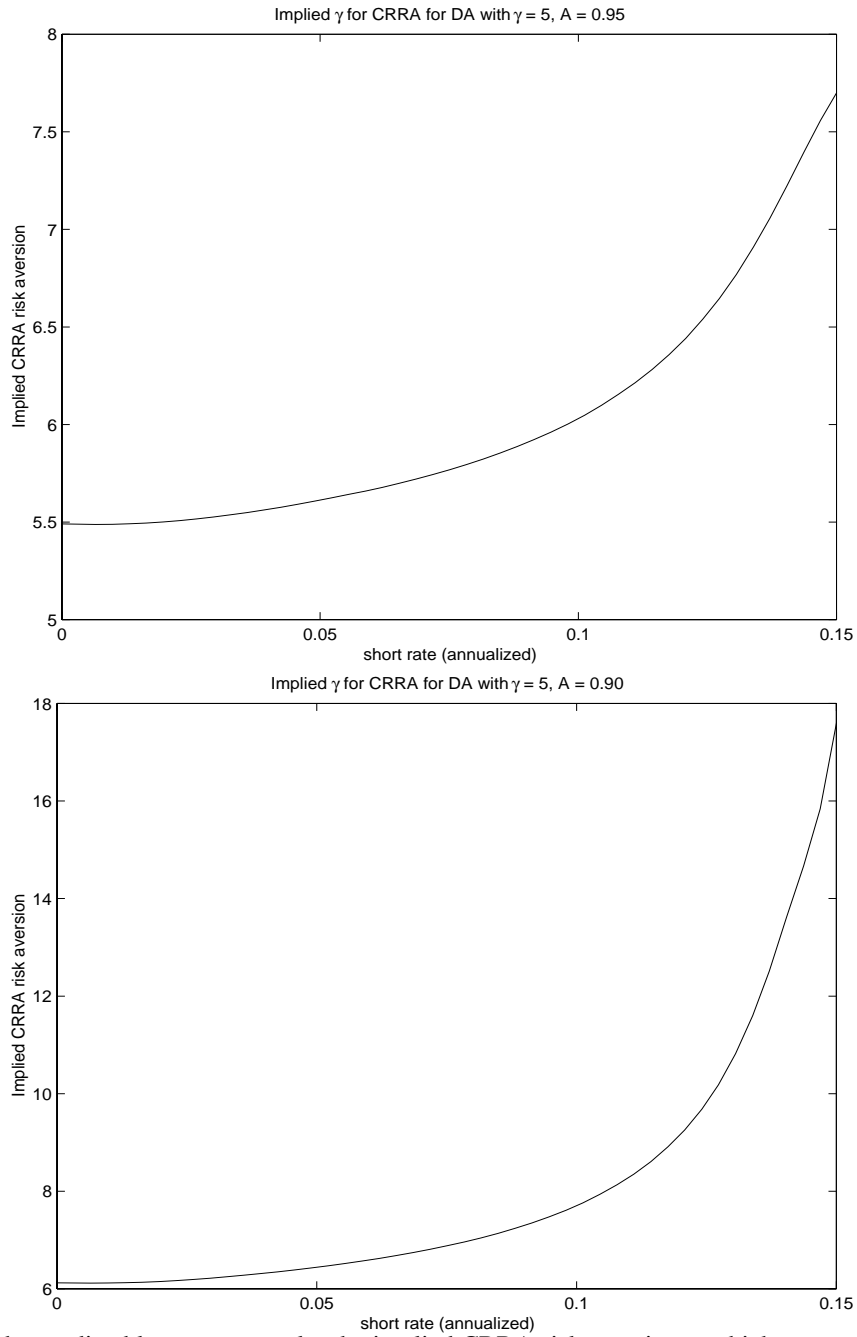
Portfolio weights for CRRA Utility for the system without predictability (left column) and with short rate predictability (right column), as a function of the short rate. We dynamically rebalance every quarter.

Figure 5: CRRA Portfolio Weights



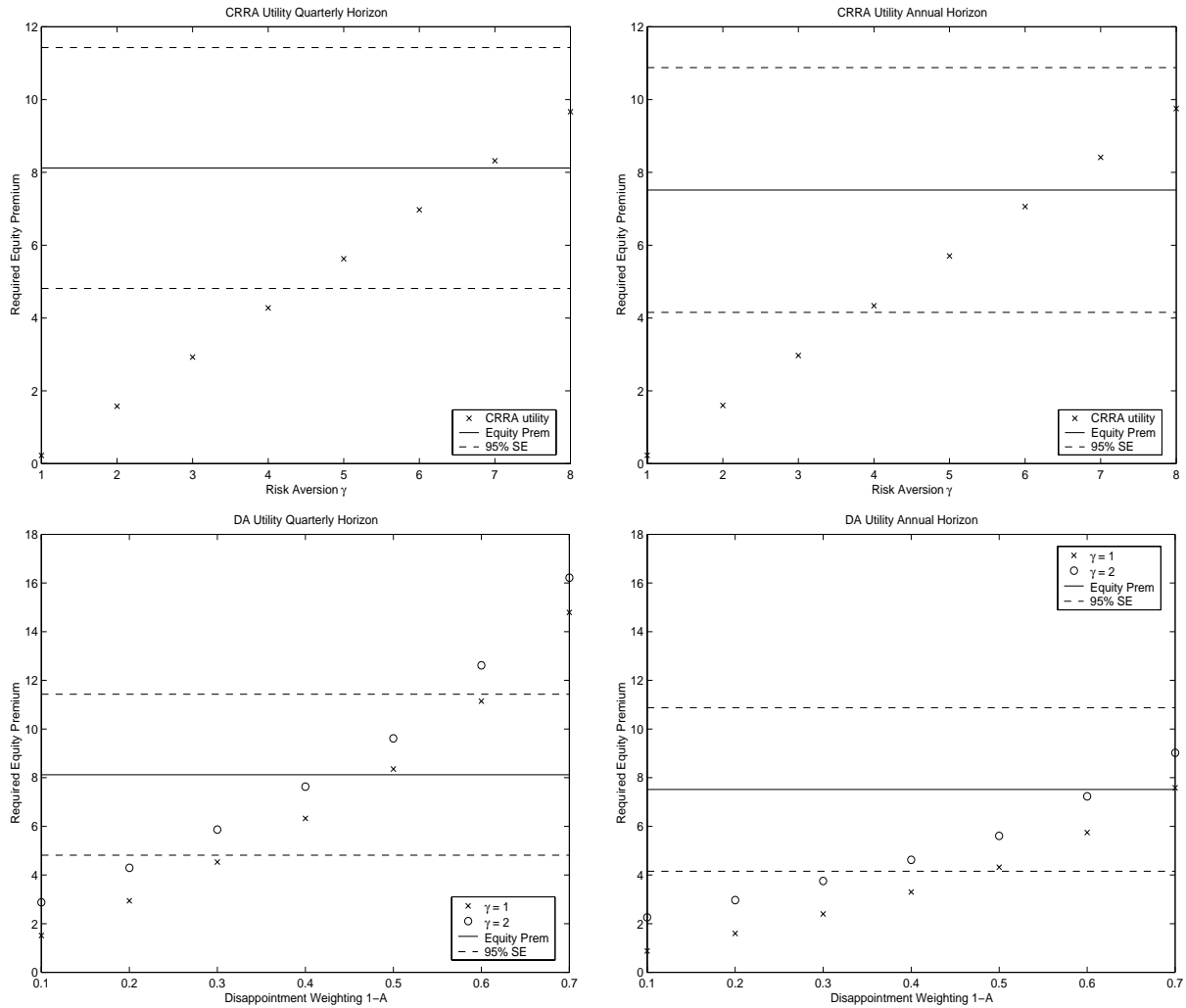
Consider DA utility for the predictable system with $\gamma = 5$ and $A = 0.95$ in the top plot and $A = 0.90$ in the bottom plot. The graph shows the γ that yields the same optimal equity allocation when utility is CRRA. The x -axis shows (annualized) short rates, and the y -axis shows the implied CRRA. We work with a quarterly horizon. We dynamically rebalance every quarter.

Figure 6: DA Utility Portfolio Weights



For the predictable system we plot the implied CRRA risk aversion γ which corresponds to the same portfolio weight for DA utility. The horizon is one quarter. On the x -axis we show (annualized) short rates, and on the y -axis implied CRRA γ .

Figure 7: Implied CRRA Risk Aversion of DA Utility



Equity premium required to hold a 60% weight in equities with a VAR without predictability (1941-1998). The historical equity premium is given by the straight line, with 95% SE bounds in dotted lines. The left hand column displays the required equity premium using a quarterly horizon, the right hand column using an annual horizon. The top row shows CRRA utility, with risk aversion γ on the x -axis. The bottom row shows DA utility, with $\gamma = 1$ and $\gamma = 2$, with disappointment weight $1 - A$ on the x -axis. The required equity premium is presented as an annualized figure.

Figure 8: Required Equity Premium for 60% Equity