Available online at www.sciencedirect.com
SCIENCE DIRECT®

Finance Research Letters 1 (2004) 171–177

Finance Research Letters

www.elsevier.com/locate/frl

Myopic loss aversion and the equity premium puzzle reconsidered ☆

Robert B. Durand *, Paul Lloyd, Hong Wee Tee

University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia

Received 24 February 2004; accepted 4 May 2004

Available online 2 July 2004

Abstract

Benartzi and Thaler [The Quarterly Journal of Economics 110 (1995) 73–92] offer a quasi-rational explanation for the equity premium puzzle. We reconsider their methodology and, making a simple modification to it, find that their analysis is not robust.

Keywords: Equity premium; Prospect theory

The gulf between equity and risk-free returns—the equity premium—has puzzled analysts, and seemingly defied rational explanations, for almost two decades.[1] BT (Benartzi and Thaler, 1995) suggest a quasi-rational explanation for the puzzle. Using monthly returns from 1926 to 1990, they randomly draw 100,000 n-month returns and ask over what evaluation period a representative agent, whose preferences may be modeled using Prospect Theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992), might be indifferent to holding stocks or bonds.[2] They suggest that the equilibrium evaluation period they calculate, of about a year, is consistent with observed behavior.

☆ Robert Durand is a senior lecturer in Finance, Paul Lloyd a lecturer in Accounting, and Hong Wee Tee is a postgraduate student in Finance at the University of Western Australia.

* Corresponding author.
E-mail address: Robert.Durand@uwa.edu.au (R.B. Durand).

[1] Mehra and Prescott (2003) provide a recent review of literature analyzing the puzzle.

[2] There is experimental evidence that individuals behave as BT argue: investors combine loss aversion with mental accounting (Thaler, 1985). Although most studies have been conducted using students as subjects, recent evidence shows that loss aversion and mental accounting are more pronounced in a sample of professional traders from the Chicago Board of Trade (CBOT) (Haigh and List, 2004). Coval and Shumway (2004) focus on the

1544-6123/$ – see front matter
doi:10.1016/j.frl.2004.05.001

This letter reconsiders BT's methodology. We make one simple change to BT's methodology. BT take random samples of returns and ask at what evaluation-horizon is the prospective utility of stocks and bonds equal. We take samples of returns and, at each evaluation-horizon, we test the null hypothesis that the prospective utility of stocks equals the prospective utility of the risk-free instrument. This simple change has a dramatic consequence: using US data coincident with the period examined in BT, we find that BT's conclusions do not hold for the US market.

Before beginning our discussion, it is worth revisiting the analysis in BT (Benartzi and Thaler, 1995, pp. 79–80). Investors judge the outcome of investments using an asymmetric-sigmoidal function of the form:

$$v(x) = \begin{cases} x^{\alpha} & \text{if } x \geqslant 0, \\ -\lambda(-x)^{\beta} & \text{if } x < 0, \end{cases} \tag{1}$$

where $v(x)$ is the value of the outcome, λ is the coefficient of loss aversion, and α and β are estimated empirically to be 0.88. All things held equal, the "pain" of a loss $\$x$ is greater than the "joy" of a gain of $\$x$. By way of contrast to standard utility theory, outcomes are not evaluated by assessing the potential to change an investor's total wealth; outcomes are evaluated with reference to some "natural" reference point (say, a price particularly salient to an investor). The prospective utility of a risky investment (denoted G) is not quite the expected value of the expected outcome; rather the probabilities are weighted subjectively as investors have a propensity to be biased in processing probabilities. The prospective utility of G, $V(G)$, is

$$V(G) = \sum \pi_i v(x_i), \tag{2}$$

where π_i is the decision weight assigned to outcome i. When π_i is used in a cumulative probability distribution it is defined as

$$\pi_i = w(P_i) - w\left(P_i^*\right), \tag{3}$$

where P_i is the probability of obtaining an outcome at least as good as x and P_i^* is the chance of obtaining a better outcome and the probability of each outcome i, P_i, is approximately

$$w(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{1/\gamma}}. \tag{4}$$

The asymmetry, and biased weighting of probabilities, is captured by γ, which is taken to be 0.69 in both the domains of losses and gains.

Therefore, for any n-month period, the distribution of returns can be used to calculate $V(G)$ for both stocks and bonds. Where n is the period over which the representative agent evaluates his investments, when $V(G)_{\text{stocks},n} = V(G)_{\text{bonds},n}$ we assume the agent would have been indifferent between the two investment classes over that period of time. In this case, over evaluation period n, the investment classes would be in a type of

trading behavior of Treasury Bond market makers on the CBOT. They find evidence that traders exhibit loss aversion but that any price impact resulting from such biases is not long lasting.

equilibrium: market forces, embodied in our representative agent, would have no incentive to shift assets from one class of investments to another and prices would not change as a result of such shifting. If $V(G)_{\text{stocks},n} > V(G)_{\text{bonds},n}$ we would assume that the agent would have preferred holding stocks rather than bonds over the evaluation period n and that returns would not be in equilibrium. If $V(G)_{\text{stocks},n} < V(G)_{\text{bonds},n}$, vice versa. That $V(G)_{\text{stocks},n} = V(G)_{\text{bonds},n}$ at $n \approx 12$ (and at no other evaluation-horizon) leads BT to the conclusion that a 12-month evaluation horizon is consistent with observed equity premiums and, hence, the equity premium is consistent with the model of prospective value outlined above.

We take issue with BT over a fairly elementary concern. We contend that, when sampling, the question of when $V(G)_{\text{stocks},n} = V(G)_{\text{bonds},n}$ should not be determined by visually inspecting two lines.[3] Rather, as with all questions of statistical inference, it should be determined by the observed values of the data and their distribution. Only when there is sufficient evidence should the null that $V(G)_{\text{stocks},n} = V(G)_{\text{bonds},n}$ be rejected with confidence. If the data are parametric, a paired t-test would be appropriate to test the null hypothesis that $V(G)_{\text{stocks},n} = V(G)_{\text{bonds},n}$.[4] If the data are non-parametric, as is sometimes the case, a non-parametric matched test would be appropriate to test the null hypothesis that $V(G)_{\text{stocks},n} = V(G)_{\text{bonds},n}$: we use the Wilcoxon signed-rank test. The test statistic for the Wilcoxon signed-rank test is approximately normally distributed with a mean zero and unit variance.[5] If BT's analysis holds, we would expect to reject the null that $V(G)_{\text{stocks},n} = V(G)_{\text{bonds},n}$ at most evaluation horizons and to accept it at a, hopefully plausible, value of n.

Such tests of $V(G)_{\text{stocks},n} = V(G)_{\text{bonds},n}$, however, lead us to take issue with BT on another point. The assumption that observations are independent is made for both the paired t-test and the Wilcoxon signed-rank test (Hogg and Tanis, 1993, pp. 10, 414, and 614). Following BT's resampling methodology[6] to generate 100,000 n-month returns would result in a number of return observations that are taken over overlapping periods[7] and, therefore, they would not be independent. Consequently, any resulting estimates of the prospective utility of these returns will also not be independent. In light of this, utilizing data obtained through replicating BT's resampling methodology would be inconsistent with the assumptions behind the tests we conduct. Therefore, to ensure our data are independent, starting from our first observation January 1926, we calculate 768 values for $V(G)$ for both stocks and bonds for each discrete one-month evaluation horizon, 684 values for $V(G)$ for both stocks and bonds for each discrete two-month evaluation horizon, etc. until we have discrete observations for evaluation periods ranging from one to twenty four months. $V(G)$ is calculated in the same way as BT using the formulae presented

[3] BT's results are depicted in Figure 1 (Benartzi and Thaler, 1995, p. 84).

[4] The test computes the differences in $V(G)$ for each asset at each n and, using the t-distribution, tests whether the average differs from zero.

[5] See Siegel and Castellan (1988, pp. 87–95) for a discussion.

[6] BT drew their 100,000 n-month returns, with replacement, from the CRSP time series of returns for the sample period. Further details may be found in Benartzi and Thaler (1995, p. 82).

[7] Nelson and Kim (1993), Goetzmann and Jorion (1993), and Kirby (1997) provide detailed discussions of the utilization of overlapping data in financial econometrics.

above. We utilize data in Cornell (1999, pp. 79–100) to examine monthly returns for the S&P and two measures of the risk-free rate of return: one-month T-bills and twenty-year Treasury Bond rates (both nominal and real).

The results of our analyses for nominal returns are reported in Table 1. We report the estimated mean $V(G)$ for each of the assets we study for each evaluation period in the first

Table 1
Tests of the equality of the prospective utility of nominal returns of US stocks and bonds

Evaluation period (months)	Mean estimated V(S&P500 nominal returns)	Mean estimated V(one-month nominal T-bill)	Mean estimated V(twenty-year nominal Treasury Bond)	S&P500 vs. the one-month T-bill rate (nominal)	S&P500 vs. the twenty-year Treasury Bond (nominal)
1	−0.0135	0.0057	−0.0046	−0.676 (W)	−1.372 (W)
2	−0.0066	0.0058	−0.0019	−0.879 (W)	−0.682 (W)
3	−0.0053	0.0058	−0.0015	−0.798 (W)	−0.804 (W)
4	−0.0049	0.0058	−0.0010	−1.094 (W)	−0.582 (W)
5	−0.0051	0.0058	−0.0009	−2.495*	−0.912
6	−0.0017	0.0059	0.0007	−1.585	−0.498
7	−0.0051	0.0058	−0.0008	2.040*	−0.830
8	−0.0049	0.0058	−0.0010	−2.442*	−0.911
9	−0.0045	0.0058	−0.0005	−2.086*	−0.813
10	−0.0052	0.0059	−0.0008	−2.433*	−0.929
11	−0.0075	0.0058	−0.0010	−2.328*	−1.149
12	−0.0085	0.0059	−0.0020	−2.958**	−1.384
13	−0.0115	0.0059	−0.0019	−2.839**	−1.579
14	−0.0115	0.0058	−0.0022	−3.272**	−1.765
15	−0.0074	0.0059	−0.0001	−2.279*	−1.263
16	−0.0186	0.0057	−0.0051	−4.938**	−2.704**
17	−0.0196	0.0056	−0.0063	−5.006**	−2.504*
18	−0.0206	0.0057	−0.0066	−4.653**	−2.507*
19	−0.0214	0.0057	−0.0070	−4.270**	−2.217*
20	−0.0220	0.0058	−0.0064	−4.966**	−2.678*
21	−0.0227	0.0057	−0.0070	−4.471**	−2.479*
22	−0.0214	0.0057	−0.0078	−4.421**	−2.190*
23	−0.0263	0.0056	−0.0078	−5.320**	−2.950**
24	−0.0266	0.0057	−0.0074	−4.870**	−3.000**

Note. This table reports estimates of the mean prospective utility—$V(G)$—for nominal returns of three US investment classes: the S&P500, twenty-year Treasury Bond returns, and 1-month Treasury Bill returns from January 1926 and finishing in December 1990. The data are from Cornell (1999, pp. 79–100). The table also reports tests of the null hypothesis that $V(G)_{\text{stocks},n} = V(G)_{\text{bonds},n}$ where n $(n = 1, \ldots, 24)$ is the evaluation period for the calculation of $V(G)$. Where a one-sample Kolmogorov–Smirnov test (not reported) could not reject the null hypothesis that the distributions of both $V(G)_{\text{stocks},n}$ and $V(G)_{\text{bonds},n}$ are compatible with a Gaussian normal distribution (i.e., where the p-value > 0.05), the test of the null hypothesis that $V(G)_{\text{stocks},n} = V(G)_{\text{bonds},n}$ is conducted using a paired t-test. Where the null of normality is rejected for one, or both of the distributions of $V(G)_{\text{stocks},n}$ and $V(G)_{\text{bonds},n}$, the test of the null hypothesis that $V(G)_{\text{stocks},n} = V(G)_{\text{bonds},n}$ is conducted using a Wilcoxon signed-rank test (Siegel and Castellan, 1988, pp. 87–95). Use of the Wilcoxon signed-rank test is indicted by "(W)" after the reported test statistic (which is approximately normally distributed with mean zero and unit variance).

* Test statistic is significant at the 5% level.
** Test statistic is significant at the 1% level.

three columns of numbers in the table. In each of the final two columns of the table, we test the null hypothesis that the hypothesis that $V(G)_{\text{stocks},n} = V(G)_{\text{bonds},n}$ for each evaluation period ($n = 1, \ldots, 24$) using either the equal-weighted or value weighted market index and the risk free rate (the 13-week Treasury note rate for the short-term risk free rate and the 10-year government bond rate for the long-term risk free rate). We examine the distribution of $V(G)$ for each investment in each evaluation period and, where the hypothesis that the data are compatible with a Gaussian-normal distribution cannot be rejected with confidence, we report the result of a paired t-test.[8] Where the hypothesis that one, or both, of the values for $V(G)$ conform to a normal distribution can be rejected with confidence (where the p-value is less than or equal to 0.05), we test the null hypothesis that $V(G)_{\text{stocks},n} = V(G)_{\text{bonds},n}$ using, as foreshadowed above, the Wilcoxon signed-rank test; we denote the use of this non-parametric test by including "(W)" after the test statistic.

BT argue that we would expect to reject the null hypothesis that $V(G)_{\text{stocks},n} = V(G)_{\text{bonds},n}$ at most evaluation horizons and to accept it at a plausible value of n. This is not what we find in the data. Examination of Table 1 reveals that, at shorter horizon evaluation periods, the hypothesis that $V(G)_{\text{stocks},n} = V(G)_{\text{bonds},n}$ cannot be rejected. At longer horizons, the hypothesis that $V(G)_{\text{stocks},n} = V(G)_{\text{bonds},n}$ can be rejected, but, at these horizons, it is bonds that seem to be preferred. We do not see the expected pattern of an insignificant difference between $V(G)_{\text{stocks}}$ and $V(G)_{\text{bonds}}$ at a particular evaluation horizon "sandwiched" between periods where the prospective utilities of each investment class are significantly different. We repeated the analysis for real returns (the results are reported in Table 2) and these findings are consistent with those for nominal returns. Thus, for both nominal and real returns, as evaluation horizons lengthen, the volatility of stock returns and the pain of the many instances of negative returns leads to lower prospective utility.[9]

In conclusion, BT proffer a highly innovative approach to solving the equity premium puzzle. Utilizing prospect theory, they examine over what evaluation period the prospective utility of stocks and bonds would be equal and, in view of their estimates of the evaluation period being in keeping with observed behavior, argue that the observed equity premium is consistent with quasi-rational behavior. Without doubt, quasi-rational, or behavioral, approaches to the equity premium puzzle should be a fruitful course of research. More than two decades of research has only shown that "the equity premium. . .[is] of an order of magnitude greater than could be rationalized in the context of the standard neoclassical paradigms of financial economics as a *premium for bearing risk*. . .unless one is willing to accept that individuals are implausibly risk averse" (Mehra and Prescott, 2003). BT provide the first attempt to consider a quasi-rational explanation to the puzzle. This letter, however, has demonstrated that BT's analysis is not robust to a simple fine-tuning of their methodology.

[8] We do not report the results of the Kolmogorov–Smirnov test we utilize to conduct these tests of the null of normality.

[9] It may be the case that BT's sampling methodology introduced bias into the analysis. We believe that exploring this possibility is an interesting area for future research.

Table 2
Tests of the equality of the prospective utility of real returns of US stocks and bonds

Evaluation period (months)	Mean estimated V(S&P500 adjusted for inflation)	Mean estimated V(one-month T-bill—adjusted for inflation)	Mean estimated V(twenty-year Treasury Bond—adjusted for inflation)	S&P500 vs. the one-month T-bill rate (inflation adjusted)	S&P500 vs. the twenty-year Treasury Bond (inflation adjusted)
1	−0.0194	−0.0028	−0.0121	−0.324 (W)	−1.640 (W)
2	−0.0126	−0.0022	−0.0092	−0.415 (W)	−0.964 (W)
3	−0.0112	−0.0022	−0.0087	−0.366 (W)	−1.044 (W)
4	−0.0106	−0.0021	−0.0083	−0.549 (W)	−0.912 (W)
5	−0.0106	−0.0022	−0.0082	−0.914 (W)	−0.535
6	−0.0075	−0.0019	−0.0064	−0.540 (W)	−0.232
7	−0.0106	−0.0023	−0.0080	−0.846 (W)	−0.497
8	−0.0106	−0.0021	−0.0080	−1.211 (W)	−0.579
9	−0.0101	−0.0022	−0.0077	−1.569	−0.476
10	−0.0106	−0.0023	−0.0080	−1.799	−0.533
11	−0.0129	−0.0024	−0.0081	−1.815	−0.838
12	−0.0140	−0.0026	−0.0096	−2.302*	−0.932
13	−0.0172	−0.0029	−0.0095	−2.286*	−1.243
14	−0.0170	−0.0030	−0.0095	−2.600*	−1.412
15	−0.0132	−0.0025	−0.0075	−1.810	−0.985
16	−0.0245	−0.0035	−0.0125	−4.195**	−2.400*
17	−0.0251	−0.0037	−0.0139	−4.227**	−2.101*
18	−0.0260	−0.0039	−0.0144	−3.855**	−2.044*
19	−0.0270	−0.0040	−0.0145	−3.541**	−1.877
20	−0.0280	−0.0040	−0.0143	−4.233**	−2.345*
21	−0.0287	−0.0041	−0.0145	−3.789**	−2.221*
22	−0.0275	−0.0042	−0.0155	−3.783**	−1.933
23	−0.0316	−0.0046	−0.0153	−4.415**	−2.619*
24	−0.0317	−0.0046	−0.0151	−4.135**	−2.606*

Note. This table reports estimates of the mean prospective utility—$V(G)$—for real returns of three US investment classes: the S&P500, twenty-year Treasury Bond returns, and 1-month Treasury Bill returns from January 1926 and finishing in December 1990. The data are from Cornell (1999, pp. 79–100). The table also reports tests of the null hypothesis that $V(G)_{\text{stocks},n} = V(G)_{\text{bonds},n}$ where n ($n = 1, \ldots, 24$) is the evaluation period for the calculation of $V(G)$. Where a one-sample Kolmogorov–Smirnov test (not reported) could not reject the null hypothesis that the distributions of both $V(G)_{\text{stocks},n}$ and $V(G)_{\text{bonds},n}$ are compatible with a Gaussian normal distribution (i.e., where the p-value > 0.05), the test of the null hypothesis that $V(G)_{\text{stocks},n} = V(G)_{\text{bonds},n}$ is conducted using a paired t-test. Where the null of normality is rejected for one, or both of the distributions of $V(G)_{\text{stocks},n}$ and $V(G)_{\text{bonds},n}$, the test of the null hypothesis that $V(G)_{\text{stocks},n} = V(G)_{\text{bonds},n}$ is conducted using a Wilcoxon signed-rank test (Siegel and Castellan, 1988, pp. 87–95). Use of the Wilcoxon signed-rank test is indicted by "(W)" after the reported test statistic (which is approximately normally distributed with mean zero and unit variance).

* Test statistic is significant at the 5% level.
** Test statistic is significant at the 1% level.

Acknowledgments

Participants at the Australasian Finance and Banking Conference 2003 have provided helpful comments and feedback. We thank Ross Maller and Alex Szimayer for their advice and assistance. Thanks also to our research assistants: Alex Juricev and Jade Glenister.

References

Benartzi, S., Thaler, R.H., 1995. Myopic loss aversion and the equity premium puzzle. The Quarterly Journal of Economics 110, 73–92.
Cornell, B., 1999. The Equity Risk Premium. Wiley, New York.
Coval, J.D., Shumway, 2004. Do behavioral biases affect prices? Journal of Finance 59. In press.
Goetzmann, W.N., Jorion, P., 1993. Testing the predictive power of dividend yields. Journal of Finance 48, 663–679.
Haigh, M.S., List, J.A., 2004. Do professional traders exhibit myopic loss aversion? An experimental analysis. Journal of Finance 59. In press.
Hogg, R.V., Tanis, E.A., 1993. Probability and Statistical Inference. Macmillan, New York.
Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk. Econometrica 47, 263–291.
Kirby, C., 1997. Measuring the predictable variation in stock and bond returns. Review of Financial Studies 10, 579–630.
Mehra, R., Prescott, E.C., 2003. The equity premium in retrospect. In: Constantides, M.H., Stulz, R. (Eds.), Handbook of the Economics of Finance. Elsevier/North-Holland, Boston, pp. 888–936.
Nelson, C.R., Kim, M.J., 1993. Predictable stock returns: The role of small sample bias. Journal of Finance 48, 641–661.
Siegel, S., Castellan, N.J., 1988. Nonparametric Statistics For The Behavioral Sciences. McGraw–Hill, New York.
Thaler, R.H., 1985. Mental accounting and consumer choice. Marketing Science 4, 199–214.
Tversky, A., Kahneman, D., 1992. Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty 5, 297–323.