



# 1 Introduction

If a risky asset or portfolio does not "dominate" a "risk free" alternative, a premium will be demanded for holding it. The "right" premium would depend on the agent's risk assessment which, in turn, depends on both the agent's utility function and the returns distribution. An on-going challenge in finance is to devise theoretical asset pricing models that are consistent with the "stylized fact" concerning the observed premium between real returns on investments in equity and the real yields from bonds. Mehra and Prescott (1985) are the first to estimate the equity premium at about 6% p.a., using annual data for the U.S. over the period 1889 to 1978. They argue that the "size" of the premium implies unacceptably high levels of risk aversion when based on standard financial models. Subsequently, they label this phenomenon the equity premium puzzle (EPP).<sup>1</sup> What makes the puzzle enduring is that it appears to arise in different sample periods, occurs for a broad selection of assets and is characteristic of some international financial markets (Mehra (2003)).

As the observed premium is a self evident fact in need of replication/calibration with any model, the equity premium puzzle can be and has been viewed as a conflict between a priori views about, and the actual estimates of the risk aversion parameter arising from incorrectly specifying either the form of the utility function, or the probability distribution of returns, or both. The explosion of the literature since the Mehra and Prescott (1985) paper can be interpreted as a specification search over a range of models with the aim of deriving empirically "sensible" estimates of the risk aversion parameter. This specification search can be categorized into three broad groups. The first class of models focuses on preferences. This class of models looks at extending existing parametric utility functions by allowing for generalized expected utility (Epstein and Zin (1991)); habit formation (Constantinides (1990)); relative consumption (Abel (1990)); and subsistence consumption (Campbell and Cochrane (1999)).<sup>2</sup> The second class of models focuses on the specification of the probability distributions underlying the processes. The majority of the proposed models assume lognormality. Some exceptions are Rietz (1988) who specifies an augmented probability

---

<sup>1</sup>An associated puzzle is the risk free rate puzzle (Weil (1989)) whereby the implied risk free rate predicted by theoretical models is too high relative to the observed rate. Whilst the focus of the current paper is on the equity premium puzzle, the alternative models proposed in the literature in general, attempt to explain both puzzles.

<sup>2</sup>A related class of explanations would be those based on behavioural finance. For example Benartzi and Thaler (1995) suggest that the equity premium can be explained by recognising that investors are more sensitive to losses than gains and that they evaluate their portfolios frequently.

distribution that allows for extreme events, and Hansen and Singleton (1983) who do not specify any probability distribution. In general, there is strong empirical evidence to reject the lognormality assumption as it is well documented that empirical returns distribution are highly non-normal being characterized by higher order moments including both skewness and kurtosis. The third class of models relaxes the assumptions concerning complete and frictionless asset markets. Some of the main suggestions consist of allowing for incomplete markets (Weil (1992)); the inclusion of trading costs through borrowing constraints (Heaton and Lucas (1995)); transaction costs (Aiyagari and Gertler (1991)); liquidity premium (Bansal and Coleman (1996)); and taxes (McGrattan and Prescott (2001)). Put another way, the puzzle is "why can a given model not be calibrated to replicate the observed stylized fact"?

An important characteristic of the proposed theoretical models is that they adopt parametric specifications of either the preference functions or the probability distribution, or both. The fact that the search still continues suggests that no parametric specification has been uncovered that yields a priori "satisfactory" estimates of risk aversion. The complimentary strategy adopted in this paper is to circumvent these problems and adopt a nonparametric framework which imposes a minimal set of conditions on preferences and the underlying probability distribution. These conditions consist of non-satiation, risk aversion, a preference for skewness and an aversion to kurtosis.<sup>3</sup> The approach consists of couching the equity premium puzzle in terms of testing for various levels of Stochastic Dominance (SD) between the returns on equities and bonds. This is of intrinsic interest, of course, but can also shed light on the EPP literature. If equities dominate bonds, especially at lower orders, there is indeed a puzzle whatever utility or other functionals within the associated class utility functionals. The non-existence of first or second order stochastic dominance, say, means that for agents with Von Neumann-Morgenstern concave utility functions, investment in equity, for example, is not sufficiently attractive without a premium. The expected utility paradigm suggests that. To quantify what is a reasonable "size" for the premium requires specific utility functions and special values for their coefficients, as well as a knowledge of the probability laws governing these returns. This suggests that evidence of a "premium puzzle" may be an artifact of the specific functionals chosen if there is no dominance. Non-dominance, or "maximality", implies that there is no uniform (weak) ranking over the risk free asset,

---

<sup>3</sup>Harvey and Siddique (2000) provide a recent discussion of the importance of skewness in asset pricing, while Lim, Martin and Martin (2004) highlight the importance of skewness and kurtosis in the pricing of options.

and there are indeed *some* functionals, utility functions and probability distributions such as those adopted by Mehra and Prescott, that might present a puzzle. But, according to some functionals, even the 6 % differential initially observed by Mehra and Prescott (1985) may be too small, and almost surely so for some risk averse individuals. Stochastic Dominance testing helps to make clear that none of the functionals that are consistent with 6% or more premia is either irrational or puzzling. It provides a birds eye view of how the twin and very demanding obstacles of cardinal utility identification/estimation and heterogeneity (among individuals and in asset returns) has been handled in the EPP literature.

Our approach is applied to two data sets. The first is the original Mehra-Prescott annual data for the U.S. The second is daily observations on one risk-free and two risky asset indices for the U.S. The empirical results show little or no evidence of stochastic dominance in both data sets. There is some generally insignificant evidence of third or higher order dominance of equities over bonds in the Mehra and Prescott data, but at 1% nominal size of the test and not at 5%. The daily data reveal no first or second order dominance between Treasury bills and S&P500. There is weak evidence of third order stochastic dominance of Treasury bills over S&P500, suggesting that some agents rank the risk free asset over the risky asset when pricing skewness! This result may suggest that the observed equity premium has been too small to compensate agents adequately for bearing higher risk associated with S&P500. Finally, there was no evidence of either first or second order stochastic dominance between the two "risky" indices, S&P500 and NASDAQ. However, there was some evidence that S&P500 third and fourth order stochastically dominated NASDAQ. Given that S&P500 exhibits negative skewness and NASDAQ positive skewness, this suggests that the observed premium between the two assets would be even higher if they exhibited the same skewness characteristics. In view of these findings, we recommend the most flexible forms of utility functions, returns distributions that easily allow a role for higher order moments, and models that allow for heterogeneity, combined with very reliable inference techniques. Attribution of cardinal utility functions to individuals is not for the faint at heart!

The rest of the paper proceeds as follows. Empirical evidence of the equity premium and estimates of the risk aversion parameter using existing parametric models are reported in Section 2. The nonparametric testing framework based on stochastic dominance is presented in Section 3. This framework is applied in Section 4 to re-examine the Mehra-Prescott original data set, as well as to a more recent data set that uses daily equity returns and

bond yields. The main empirical results point to a lack of stochastic dominance amongst the financial returns series investigated. Section 5 provides some concluding comments and suggestions for future research.

## 2 Empirical Evidence of the Equity Premium

The equity premium puzzle is commonly demonstrated in one of two ways. The first is based on descriptive statistics that compare the average returns of different financial assets. The second involves estimating the risk aversion parameter for a chosen theoretical model. To highlight both of these approaches, the Mehra and Prescott (1985) original data set is adopted. This data consists of annual US data on real asset prices and aggregate real consumption expenditure beginning in 1889 and ending in 1979, a total of 91 observations.

### 2.1 Descriptive Statistics of the Premium

Some descriptive statistics on real equity returns ( $R_{s,t}$ ), real bond yields ( $R_{b,t}$ ) and real consumption growth rate ( $R_{c,t}$ ), are given in Table 1. The size of the equity premium between equities and bonds is approximately 6% p.a. The higher mean return on equity is associated with higher "risk", traditionally indicated by the higher value of the standard deviation for equity compared to bonds, 16.541 compared to 5.730. Further evidence of the higher risk from investing in equities is highlighted by observing that the extreme returns in equities are more than twice the extreme returns experienced by real bonds. The relatively higher volatility of real equity returns over real bond yields is also demonstrated in Figure 1 which plots the two series over the sample period, 1889 to 1978.

The strength of the contemporaneous linear relationships amongst the three series is highlighted Table 2, which gives the covariances in the lower triangle and the correlations in the upper triangle. Consumption and equities have a positive association (correlation of 0.375), as does equities and bonds (correlation of 0.113), whilst consumption and bonds have a negative association (correlation of  $-0.107$ ).

## 3 Stochastic Dominance Testing

This section outlines the framework for conducting stochastic dominance tests. The approach is based on the work of Linton, Maasoumi and Whang (2004) who propose nonparametric

Table 1:

Descriptive statistics on real equity returns ( $R_{s,t}$ ), real bond yields ( $R_{b,t}$ ), and real consumption growth rate ( $R_{c,t}$ ): expressed as percentage per annum for the period 1889 to 1978 (Mehra-Prescott data).

| Statistic | Equity<br>( $100 \times R_{s,t}$ ) | Bonds<br>( $100 \times R_{b,t}$ ) | Consump.<br>( $100 \times R_{c,t}$ ) |
|-----------|------------------------------------|-----------------------------------|--------------------------------------|
| Mean      | 6.980                              | 1.036                             | 1.826                                |
| Median    | 5.664                              | 0.412                             | 2.156                                |
| Maximum   | 50.983                             | 20.062                            | 11.111                               |
| Minimum   | -37.038                            | -18.510                           | -9.091                               |
| Std. Dev. | 16.541                             | 5.730                             | 3.587                                |
| Skewness  | 0.101                              | 0.001                             | -0.338                               |
| Kurtosis  | 2.980                              | 4.707                             | 3.721                                |
| BJ (p.v.) | 0.925                              | 0.004                             | 0.160                                |

Table 2:

Covariances (lower triangle) and correlations (upper triangle) of real equity returns ( $R_{s,t}$ ), real bond yields ( $R_{b,t}$ ), and real consumption growth rate ( $R_{c,t}$ ): percentage per annum: 1889 to 1978, Mehra-Prescott data.

|         | Equity<br>$R_t^s$ | Bonds<br>$R_t^b$ | Consump.<br>$R_t^c$ |
|---------|-------------------|------------------|---------------------|
| $R_t^s$ | 270.576           | 0.113            | 0.375               |
| $R_t^b$ | 10.577            | 32.468           | -0.107              |
| $R_t^c$ | 22.011            | -2.166           | 12.722              |

Table 3:

Alternative estimates of the relative risk aversion parameter,  $\gamma$ : 1889 to 1978, Mehra-Prescott data.<sup>(a)</sup>

| Model | Method and source                                     | $\gamma$ |
|-------|---|----------|
| 1     | Mehra (2003, equation 15) <sup>(a)</sup>              | 26.085   |
| 2     | Mehra (2003, equation 16) <sup>(b)</sup>              | 46.926   |
| 3     | Campbell et al (1997, equation 8.2.9) <sup>(c)</sup>  | 1.799    |
| 4     | Campbell et al (1997, equation 8.2.10) <sup>(d)</sup> | 11.062   |
| 5     | Campbell et al (1997, equation 8.2.9) <sup>(e)</sup>  | 1.823    |
| 6     | Campbell et al (1997, equation 8.2.10) <sup>(f)</sup> | 3.351    |
| 7     | Hansen and Singleton (1983): GMM <sup>(g)</sup>       | 15.397   |
| 8     | Grossman, Melino and Shiller (1987) <sup>(h)</sup>    | 24.755   |

The following definitions are used. For log returns:  $\hat{\mu}_s$  and  $\hat{\mu}_b$  are the respective sample means of  $\ln(1 + R_{s,t})$  and  $\ln(1 + R_{b,t})$ ,  $\hat{\sigma}_s^2$  is the sample variance of  $\ln(1 + R_{s,t})$  and  $\hat{\sigma}_{s,c}$  is the sample covariance of  $\ln(1 + R_{s,t})$  and  $\ln(1 + R_{c,t})$ . For returns:  $\tilde{\mu}_s$ ,  $\tilde{\mu}_b$  and  $\tilde{\mu}_c$  are respectively the sample means of  $R_{s,t}$ ,  $R_{b,t}$  and  $R_{c,t}$ ;  $\tilde{\sigma}_{s,c}$  is the sample covariance of  $R_{s,t}$  and  $R_{c,t}$ , and  $\tilde{\sigma}_{b,c}$  is the sample covariance of  $R_{b,t}$  and  $R_{c,t}$ .

(a) Computed as  $\hat{\gamma}_1 = (\hat{\mu}_s - \hat{\mu}_b + 0.5\hat{\sigma}_s^2) \hat{\sigma}_{s,c}^{-1}$ .

(b) Computed as  $\hat{\gamma}_1 = (\hat{\mu}_s - \hat{\mu}_b + 0.5\hat{\sigma}_s^2) \hat{\sigma}_c^{-2}$ .

(c) Computed as  $\hat{\gamma}_3 = \hat{\sigma}_{s,c} \hat{\sigma}_c^{-2}$ , by regressing  $\ln(1 + R_{s,t})$  on a constant and  $\ln(1 + R_{c,t})$ .

(d) Computed as  $\hat{\gamma}_4 = \hat{\sigma}_s^2 \hat{\sigma}_{s,c}^{-1}$ , by regressing  $\ln(1 + R_{c,t})$  on a constant and  $\ln(1 + R_{s,t})$ .

(e) Same as (c) but use an IV estimator with instruments  $\{const, R_{s,t-1}, R_{b,t-1}, R_{c,t-1}\}$ .

(f) Same as (d) but use an IV estimator with instruments  $\{const, R_{s,t-1}, R_{b,t-1}, R_{c,t-1}\}$ .

(g) The moment conditions are

$$E[\delta(1 + R_{c,t})^{-\gamma}(1 + R_{b,t}) - 1], \quad E[\delta(1 + R_{c,t})^{-\gamma}(1 + R_{s,t}) - 1],$$

and instruments as in (e) and (f).

(h) Computed as  $\hat{\gamma}_8 = (\tilde{\mu}_s - \tilde{\mu}_b)(1 + \tilde{\mu}_c)(\tilde{\sigma}_{s,c} - \tilde{\sigma}_{b,c})^{-1}$ .

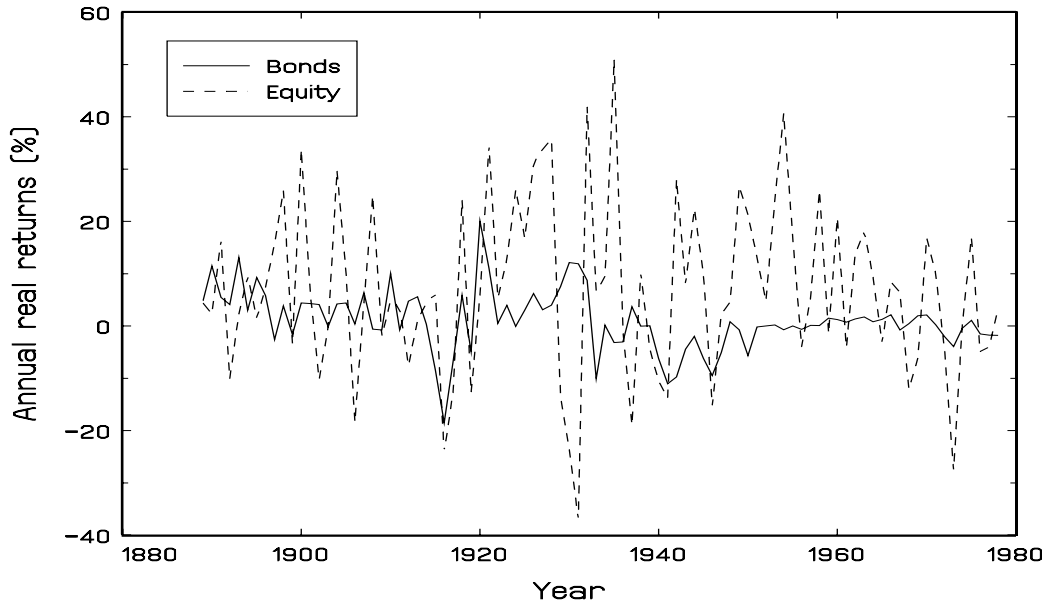


Figure 1: Bond yields and equity returns: real, percentage per annum, 1889 to 1978.

tests of stochastic dominance based on extended Kolmogorov-Smirnov statistics as formulated in McFadden (1989). Inference is performed by using "subsampling" to construct p-values as well as recentered bootstrap methods. A related approach is by Barrett and Donald (2003) who propose a set of tests with the sampling distribution of the test statistic constructed via simulation methods. An important difference between the two approaches is that unlike the resampling schemes of Linton, Maasoumi and Whang, the Barrett and Donald method for constructing critical values assumes that (i) returns are independently and identically distributed (*iid*), and (ii) different assets are independent. As these assumptions are unlikely to be satisfied in the case of financial returns which exhibit conditional volatility (Bollerslev, Chou and Kroner (1992)) and possibly higher order moment dependence structures (Harvey and Siddique (2000)), attention is restricted to the Linton, Maasoumi and Whang tests.<sup>4</sup>

### 3.1 Definitions

Consider two stationary time series of returns,  $R_{i,t}$  and  $R_{j,t}$ ,  $t = 1, 2, \dots, T$ , with respective cumulative distribution functions (CDFs),  $F_i(r)$  and  $F_j(r)$ , over the support  $r$ . The returns

---

<sup>4</sup>Abhyankar and Ho (2003) provide a recent application to financial data comparing the Linton, Massoumi and Whang (2003) and Barrett and Donald (2003) approaches.

are not expected to be *iid*, but can exhibit some dependency structures in the moments of the distribution. The null hypotheses that  $R_{i,t}$  stochastically dominates  $R_{j,t}$ , for various orders are defined below:

$$\begin{aligned}
H_0 : \text{ (First order)} \quad & F_i(r) \leq F_j(r) \\
H_0 : \text{ (Second order)} \quad & \int_0^r F_i(t) dt \leq \int_0^r F_j(t) dt \\
H_0 : \text{ (Third order)} \quad & \int_0^r \int_0^t F_i(s) ds dt \leq \int_0^r \int_0^t F_j(s) ds dt \\
H_0 : \text{ (Fourth order)} \quad & \int_0^r \int_0^t \int_0^s F_i(u) du ds dt \leq \int_0^r \int_0^t \int_0^s F_j(u) du ds dt.
\end{aligned} \tag{1}$$

The null hypotheses in this paper are unambiguous since we will test for "dominance" which combines the above with their reverses ( $j$  over  $i$ ). The alternative hypothesis is that there is no stochastic dominance. Mathematically, lower order dominance implies all higher order dominance rankings. In the case of first order dominance, the distribution function of  $R_{i,t}$  lies everywhere to the right of the distribution function of  $R_{j,t}$ , except for a finite number of points where there is strict equality. This implies that for first order stochastic dominance the probability that returns of the  $i^{th}$  asset are in excess of  $r$  say, is higher than the corresponding probability associated with the  $j^{th}$  asset

$$\Pr(R_{i,t} > r) \geq \Pr(R_{j,t} > r). \tag{2}$$

An important feature of the definitions of stochastic dominance is that they impose minimalist conditions on the preferences of agents within the class of von Neumann-Morgenstern utility functions that form the basis of expected utility theory. The different orders of dominance correspond to increasing restrictions on the shape of the utility function and the attitude towards risk of agents to higher order moments. These restrictions are nonparametric and do not require specific parametric functional forms.

Let  $u(\cdot)$  represent a utility function. For First Order Stochastic Dominance (FSD) of  $R_{i,t}$  over  $R_{j,t}$ , expected utility from holding asset  $i$  is generally greater than the expected utility from holding asset  $j$ , within the class of utility functions with positive first derivatives

$$E[u(R_{i,t})] \geq E[u(R_{j,t})], \text{ where } u' \geq 0. \tag{3}$$

That is, agents prefer higher returns on average than lower returns when preferences exhibit non-satiation. In the case of CCAPM with power utility and lognormality, the relationship

between the returns on equity ( $R_{s,t}$ ) and bond yields ( $R_{b,t}$ ) is given by (Campbell, Lo and MacKinlay (1997))

$$\ln E_t \left[ \frac{(1 + R_{s,t+1})}{(1 + R_{b,t+1})} \right] = \gamma \sigma_{s,c}, \quad (4)$$

where  $\gamma$  is the relative risk aversion parameter and  $\sigma_{s,c}$  is the covariance between  $\ln(C_t/C_{t-1})$  and  $\ln(1 + R_{s,t+1})$ . The size of the risk premium is  $\gamma \sigma_{s,c}$ , which constitutes a rightward shift in the empirical distribution of  $R_{s,t+1}$  for  $\gamma \sigma_{s,c} > 0$ .

For Second Order Stochastic Dominance (SSD), expected utility from holding asset  $i$  is generally greater than the expected utility from holding asset  $j$ , within the class of utility functions with positive first derivatives and negative second derivatives  $u' \geq 0, u'' \leq 0$ . This class of agents is characterized by risk aversion whereby a risk premium is needed to compensate investors from holding assets where the returns exhibit relatively higher "volatility".

The condition for Third Order Stochastic Dominance (TSD) implies that the expected utility from holding asset  $i$  is generally greater than the expected utility from holding asset  $j$ , within the class of utility functions with positive first and third derivatives and negative second derivatives,  $u' \geq 0, u'' \leq 0, u''' \geq 0$ . This class of agents increasingly prefers positively skewed returns as they are prepared to trade-off lower average returns for the chance of an extreme positive return. See McFadden (1989) for definitions and more detail on the equivalence of various conditions for SD rankings.

Fourth order Stochastic Dominance (FOSD) incorporates the fourth moment of the returns distribution. For fourth order stochastic dominance of asset  $i$  over asset  $j$ , the expected utility from holding asset  $i$  is generally greater than the expected utility from holding asset  $j$ , for all utility functions with  $u' \geq 0, u'' \leq 0, u''' \geq 0, u'''' \leq 0$ . This class of agents is adverse to assets that exhibit extreme negative as well as positive returns. As agents prefer thinner-tailed distributions to fat-tailed distributions, to hold assets that exhibit the latter property they need to be compensated with higher average returns. Even where two assets exhibit the same volatility, the asset returns distributions may nevertheless exhibit differing kurtosis resulting in a risk premium between the two assets.

Figure 2 highlights the stochastic dominance features of a number of hypothetical asset return distributions. All distributions are assumed to be normal,  $N(\mu, \sigma^2)$  with mean  $\mu$  and volatility  $\sigma^2$ . The distributions are  $F_1 = N(1, 6^2)$ ,  $F_2 = N(7, 6^2)$ ,  $F_3 = N(1, 12^2)$ ,  $F_4 = N(6, 12^2)$ . The first column of Figure 2 gives the stochastic dominance properties between  $F_1$  and  $F_2$ . The two returns distributions exhibit the same volatility,  $\sigma_1 = \sigma_2 = 6$ ,

but have different means  $\mu_1 = 1$  and  $\mu_2 = 6$ .  $F_2$  first (and higher) order dominates  $F_1$  as asset 2 yields a higher mean return than asset 1 ( $\mu_2 > \mu_1$ ) for the same level of risk ( $\sigma_2 = \sigma_1$ ). The equity premium of  $\mu_2 - \mu_1 = 5$ , in this case would represent a puzzle as the relatively higher return earned from investing in asset 2 comes without any additional risk.

The second column of Figure 2 gives the stochastic dominance properties of  $F_1$  and  $F_3$ . Both distributions have the same mean, but have differing volatilities. In this example, there is no first order stochastic dominance. However,  $F_1$  second order dominates  $F_3$ , as asset 1 has lower risk than asset 2 ( $\sigma_1 < \sigma_3$ ), whilst the mean returns are the same ( $\mu_1 = \mu_3$ ). Within the class of concave utility functions, asset 1 stochastically dominates asset 3. The expected return on asset 3 is too low relative to the higher risk associated with this asset. This is further demonstrated in the third column of Figure 2 where now  $F_4$  exhibits a higher average return to compensate for the higher risk (compare the distribution of asset 3 in the second column of Figure 2 with the distribution of asset 4 in the third column). There is no SD of any order between the two assets. The higher expected return in this case is indeed appropriate compensation for bearing the higher risk. The equity premium of  $\mu_2 - \mu_1 = 5$ , in this case does not represent a puzzle.

## 3.2 Testing

### 3.2.1 First Order

We combine the empirical versions of two tests. The first statistic is for the null hypotheses that  $R_{i,t}$  first order dominates  $R_{j,t}$ :

$$SD_{1,i,j} = \sqrt{T} \sup_r \left( \widehat{F}_i(r) - \widehat{F}_j(r) \right), \quad (5)$$

while the second statistic is for the reverse test where the null hypothesis is that  $R_{j,t}$  first order stochastically dominates  $R_{i,t}$

$$SD_{1,j,i} = \sqrt{T} \sup_r \left( \widehat{F}_j(r) - \widehat{F}_i(r) \right). \quad (6)$$

Here  $T$  is the sample size, and  $\widehat{F}_k(r)$  is the empirical cumulative distribution functions (CDF) of  $R_{k,t}$ ,  $k = i, j$ ,

$$\widehat{F}_k(r) = \frac{1}{T} \sum_{t=1}^T I(R_{k,t} \leq r) \quad (7)$$

where

$$I(R_{k,t} \leq r) = \begin{cases} 1 & : R_{k,t} \leq r \\ 0 & : R_{k,t} > r \end{cases}, \quad (8)$$

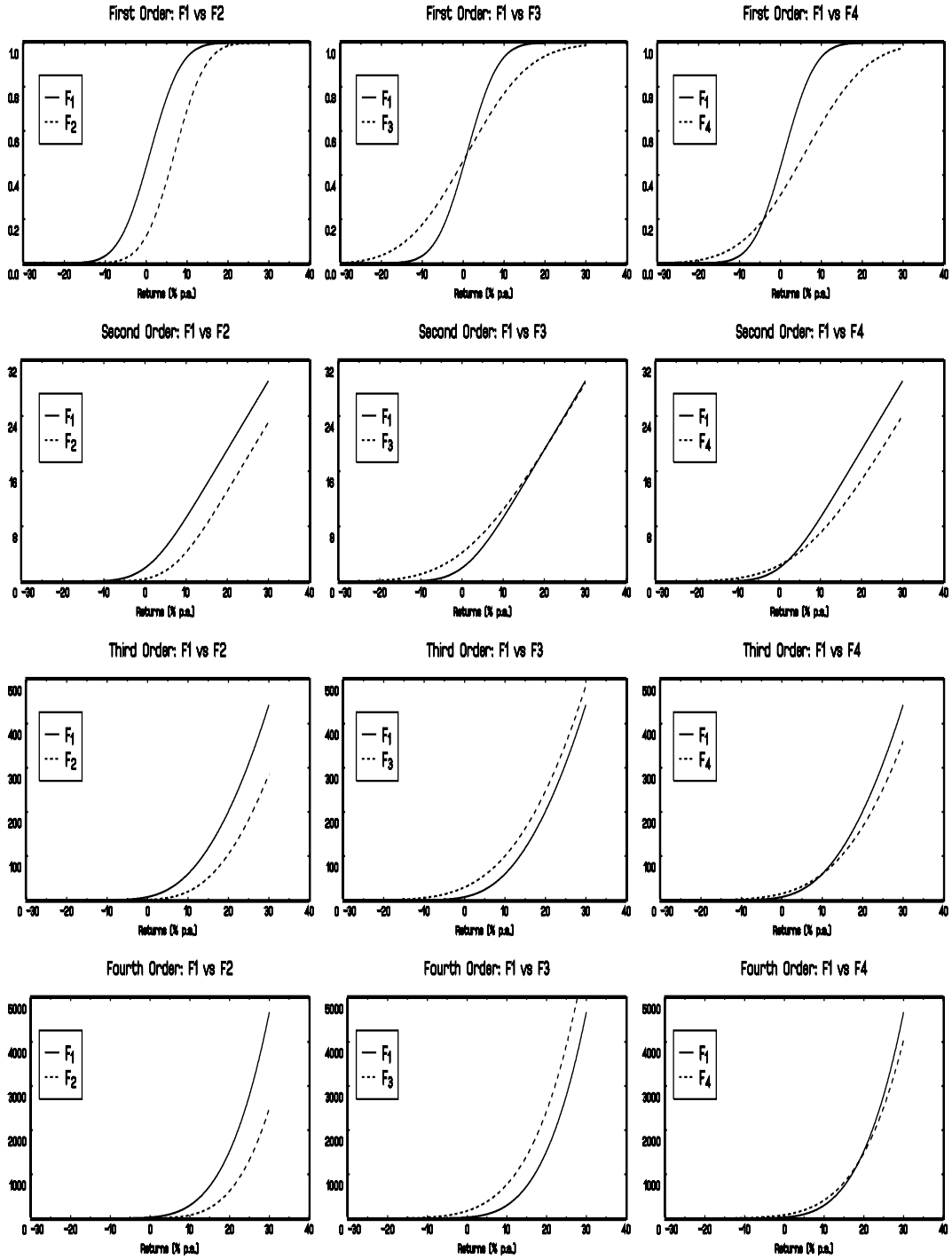


Figure 2: Hypothetical asset returns distributions, first to fourth order stochastic dominance as defined in (1):  $F_1 = N(1, 6^2)$ ,  $F_2 = N(7, 6^2)$ ,  $F_3 = N(1, 12^2)$ ,  $F_4 = N(6, 12^2)$ .

is the indicator function. Each statistic is an extension of the Kolmogorov-Smirnov test which equals the maximum distance between the two empirical CDFs,  $\widehat{F}_i(r)$  and  $\widehat{F}_j(r)$ . Following McFadden (1989), the statistics (5) and (6) are combined to provide an unambiguous overall test of first order SD:

$$MF_1 = \min_{i \neq j} (SD_{1,i,j}, SD_{1,j,i}). \quad (9)$$

Suppose that the null is true so the distribution function of  $R_{i,t}$  lies to the right of the distribution function of  $R_{j,t}$ , apart from at the tails where it is zero, as is the case in the first column of Figure 2. Now  $F_i(r) < F_j(r)$ , yielding a negative value for the support of the distribution under the null, whilst at the tails the difference is zero. Taking the sup in (5) results in a value of the test statistic of  $SD_{1,i,j} = 0$ . If the null is false then either there is no SD, in which case the two CDFs cross, or  $R_{i,t}$  is first order stochastically dominated by  $R_{j,t}$ . In either case the test statistic is positive,  $SD_{1,i,j} > 0$ . Under the null of "dominance", it must be that  $MF_1 \leq 0$ . Under the alternative the empirical CDFs must cross, resulting in  $MF_1 > 0$ . In this case the assets are "maximal", that is, they are unrankable. In the context of the equity premium puzzle both assets would be appropriately priced by the market and any premium simply reflects the price of bearing higher risk.<sup>5</sup>

In the case of *iid* data, the sampling distributions of (5) and (8) under the null was originally derived by Kolmogorov (1933), whilst McFadden (1989) derived the sampling distribution of (9). For the case where the data exhibit some dependence, the form of the (asymptotic) sampling distribution is generally unknown and depends on the unknown, underlying distributions.<sup>6</sup> To circumvent this problem the sampling distribution of the test statistics are approximated using a resampling scheme based on "subsampling"; see Politis, Romano and Wolf (1999) for a review of this approach. An important advantage of resampling is that it can accommodate dependence in asset returns over both time and contemporaneously, as demonstrated in Table 2. For a more detailed description of the "block" sampling approach used in this paper, see Linton, Maasoumi and Whang (2004).

### 3.2.2 Higher Order

To test for higher orders of SD, the CDFs are replaced by the pertinent integrated CDFs. To perform this calculation in practice, the approach adopted is to compute the  $m^{th}$  order

---

<sup>5</sup>The maximality test statistic in (9) can be extended to testing for maximality amongst more than two assets.

<sup>6</sup>Note that "pivotal" statistics are therefore not available.

CDF of asset return  $R_{i,t}$ , by<sup>7</sup>

$$\hat{F}_{m,i}(r) = \frac{1}{T(m-1)!} \sum_{t=1}^T I(R_{i,t} \leq r) (r - R_{i,t})^m. \quad (10)$$

Alternatively, the higher order CDF can be computed by cumulative sums of the lower order CDFs. The corresponding test statistics of higher order SD are denoted as  $SD_{m,i,j}$ ,  $SD_{m,j,i}$  and  $MF_m$ , in the case of subsampling, and with a superscript  $c$  in the case of bootstrapping. It is worth noting that a statistical finding of a given rank order does not imply a statistical ranking at higher orders at the same significance level. While the mathematical (probability one) rankings are ordered, sampling variation does allow for small probabilities of "apparent contradictions".

## 4 Applications

### 4.1 Mehra-Prescott Annual Data

In this section tests of SD between real Treasury bond yields ( $R_{b,t}$ ) and real equity returns ( $R_{s,t}$ ) over the period 1889-1978,  $T = 90$ , for the Mehra and Prescott data, are presented. Figure 3 gives the empirical distribution functions and various cumulative empirical distribution functions for the two series.<sup>8</sup> Inspection of the graphs suggests no evidence of any SD as the two empirical distribution functions cross for all orders of SD. The following tests provide "degrees of statistical confidence" one may attach to inferences.

First, second, third and fourth order SD tests based on  $MF_m$  as well as the individual SD tests ( $SD_{m,i,j}$ ,  $SD_{m,j,i}$ ), are reported in Table 4. The first column gives the order of SD being tested, with the null hypothesis given in the second column. The calculated, sample value of the test statistic is reported in the third column. The last three columns provide information on the sampling distribution of the test statistic with the p-values reported in the last column. The bootstraps are based on "recentered paired bootstraps" with overlapping blocks. The block sizes are set at  $B = 9$  using the rule  $B = \alpha \lceil \sqrt{T} \rceil$  with  $\alpha = 1$ . This represents a string of 10 years of data in each block. For a sample of size  $T = 90$ , this yields 82 overlapping blocks. For each bootstrap, 9 blocks are randomly drawn and stacked

<sup>7</sup>Expression (10) is motivated by integrating  $\int_0^r F_i(t) dt$  in (1) by parts and replacing it by its empirical analogue. Repeating the integrations for the higher order integrals yields equation (10).

<sup>8</sup>The support of the cumulative distribution function is based on the range of the data with the number of intermediate points set equal to the sample size,  $T$ .

producing a bootstrap sample equal to  $T$  observations.<sup>9</sup> The number of replications is set at 10000.

The reported value of the first order SD in Table 4 is 1.160, a positive value and a traditional p-value of .03. A one sided test is appropriate, however, and would be handily rejected at the 95% level (the "bottom 5%"; 0.105), and even the lower tail 99% critical value is positive. In fact, even the 99% confidence interval does not contain "zero" as revealed from inspection of the individual first order SD tests, but there may be a somewhat larger set of utility functionals that favor equities over bonds, than the other way round. This is because there are no subsamples in which the CDFs do not cross. It is worth noting that an implied critical value of "zero" may correspond to a conventionally low test size in some cases. As Linton et al (2004) have shown, our tests are "consistent" and their distribution converges to  $-\infty$  under the strict null of dominance ( $MF_1 \ll 0$ ). The asymptotic distribution is Gaussian on the boundary of the null ( $MF_1 = 0$ ). A "zero" would appear to be the appropriate critical value to choose in a setting where economists would find it lacking in credibility to conclude dominance when the sample CDFs cross and would choose to maximize test power.

The test value for second order SD in Table 4 is also positive with a p-value of 0.000. This implies that agents with preferences characterized by monotonically increasing and concave utility functions are indifferent between bonds and equities, as the higher premium on equities provides sufficient compensation for bearing the higher risk in equities. The one-sided test of SSD is rejected at the 95% level (.000). We note that there is .05 probability of negative values for the statistic, suggesting that a 99% confidence interval for SSD includes "zero". Thus "equal ranking" is not rejected at this level of confidence.

The results of the third and fourth order SD tests also show that neither security dominates the other, with the dominant test values in both cases being positive and yielding p-values less than 1%. This suggests that bonds and equities are unrankable in terms of skewness and kurtosis and that agents who have a preference for positive skewness and an aversion for kurtosis, are indifferent between holding the two assets. Again we note that there is 0.05 probability of negative values for the statistics, suggesting that a 99% confi-

---

<sup>9</sup>An important input into the subsampling approach is the size of the blocks,  $B$ . Politis, Romano and Wolf (1999) discuss various methods for determining the block size. In determining  $B$  it is important that it grows at a slower rate than the sample size  $T$ . Given this property the approach adopted here is to choose  $B$  using the formula  $B = \alpha \lceil \sqrt{T} \rceil$ , where  $\lceil \sqrt{T} \rceil$  denotes the largest integer that is less than or equal to  $\sqrt{T}$ , and  $\alpha$  is a constant. Following Linton, Maasoumi and Whang (2003) sensitivity analysis were performed on the block size to establish the robustness properties of the subsampling procedure. These results were reported in an earlier version of the paper and are available from the authors upon request.

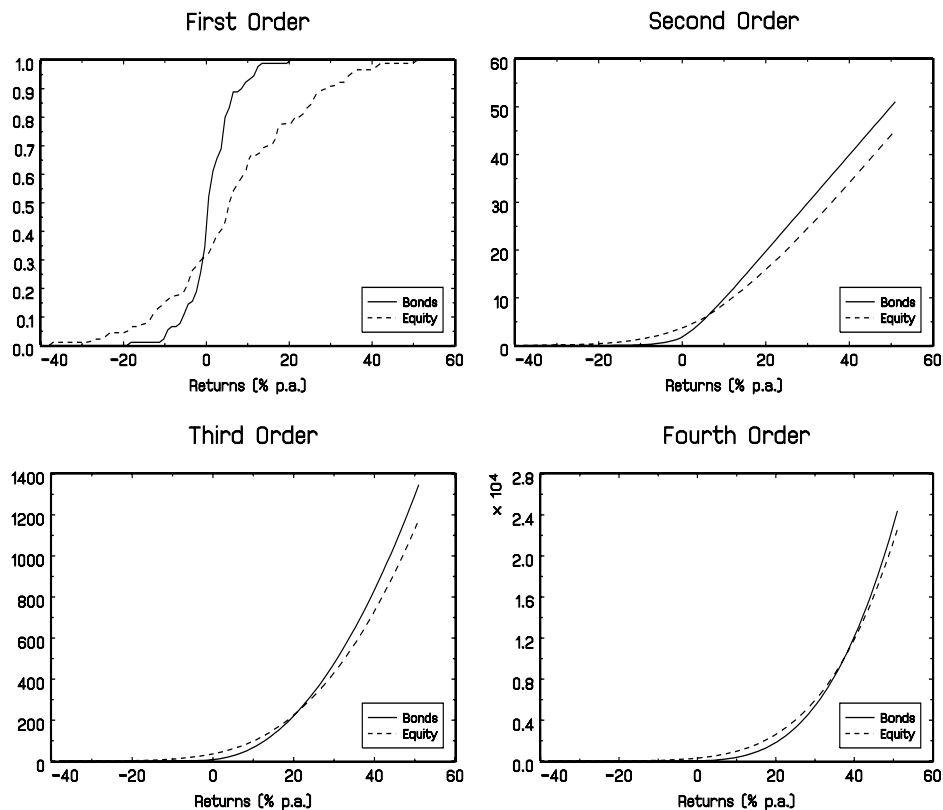


Figure 3: First to fourth order empirical cumulative distribution functions for real bond yields and real equity returns: percentage per annum, 1889 to 1978.

dence interval for SSD includes "zero". Thus "equal ranking" is not rejected at this level of confidence, and higher order moments matter, albeit only slightly.

Overall the results show that there is no clear SD between bond yields and equity returns for the Mehra-Prescott data. This is especially true for risk preferences characterized by second, third and fourth order moments. Within the context of the equity premium puzzle, this result implies that the equity premium between equities and bonds reported in Table 1 simply reflects the risk preferences of agents. There is just one case where there is evidence of an equity premium puzzle. This occurs where utility functions are simply characterized by preferences that do not exhibit non-satiation and the size of the test is chosen to be 1%. However, adopting a 5% level for the test reveals no first order SD and hence no puzzle.

Table 4:

SD tests of real bond yields ( $R_{b,t}$ ) and equity returns ( $R_{s,t}$ ): Mehra-Prescott data, 1889 to 1978. Bootstraps based on recentered paired bootstraps with overlapping blocks.

| Stochastic<br>Dominance | Null Hypothesis               | Statistic | Bottom<br>5% | Top<br>5% | pv    |
|-------------------------|-------------------------------|-----------|--------------|-----------|-------|
| First:                  | Non-maximal                   | 1.160     | 0.105        | 1.054     | 0.030 |
|                         | $R_{b,t}$ <i>SD</i> $R_{s,t}$ | 3.479     | 0.316        | 2.214     | 0.002 |
|                         | $R_{s,t}$ <i>SD</i> $R_{b,t}$ | 1.160     | 0.211        | 1.687     | 0.222 |
| Second:                 | Non-maximal                   | 18.974    | 0.000        | 7.695     | 0.000 |
|                         | $R_{b,t}$ <i>SD</i> $R_{s,t}$ | 56.710    | 0.000        | 35.101    | 0.002 |
|                         | $R_{s,t}$ <i>SD</i> $R_{b,t}$ | 18.974    | 0.000        | 24.244    | 0.103 |
| Third:                  | Non-maximal                   | 316.439   | 0.000        | 104.355   | 0.000 |
|                         | $R_{b,t}$ <i>SD</i> $R_{s,t}$ | 1600.640  | 0.000        | 1531.280  | 0.042 |
|                         | $R_{s,t}$ <i>SD</i> $R_{b,t}$ | 316.439   | 0.000        | 1134.520  | 0.300 |
| Fourth:                 | Non-maximal                   | 7345.971  | 0.000        | 1380.440  | 0.000 |
|                         | $R_{b,t}$ <i>SD</i> $R_{s,t}$ | 16774.407 | 0.000        | 39940.516 | 0.265 |
|                         | $R_{s,t}$ <i>SD</i> $R_{b,t}$ | 7345.971  | 0.000        | 37645.651 | 0.357 |

## 4.2 Daily Financial Data

Tests of SD are now applied to daily data on three financial assets consisting of a risk free asset (3 month Treasury bonds), and two risky assets (S&P500 and NASDAQ prices).<sup>10</sup> The data begin after July 4th, 1989, and end on July 14th, 2003, a total of 3661 observations. Computing daily continuously compounded equity returns results in a sample of size  $T = 3660$ . The equity returns are scaled by 252 to annualize the daily returns and by 100 to express the returns as a percentage.

Some descriptive statistics of the three series are given in Table 5. The sample means show that the equity premium between the risk free asset and the two equity assets is between 4 and 8, which is similar to the premium reported in Table 1 for the Mehra-Prescott data. Inspection of the standard deviations show that the higher mean returns are associated with higher volatility.

Table 5 also reveals a sizeable premium of just over 4% between the two risky assets, S&P500 and the NASDAQ. This is presumably compensation for the relatively higher risk associated with investing in the NASDAQ, where the sample standard deviation is nearly twice as large as the sample standard deviation of the S&P500. A further component of this premium could be the result of the marginally higher kurtosis estimate of the NASDAQ over the S&P500 leading investors to demand an even higher premium for investing in the NASDAQ. Interestingly, the skewness estimate of the S&P500 is negative compared to the positive estimate of the NASDAQ. If agents prefer positive skewness to negative skewness, this would suggest that the observed premium between the two equities could be even higher if the two returns exhibited similar skewness characteristics. In general, all of the daily yields and returns exhibit significant nonnormalities, as revealed by the Bera-Jarque normality test. This result raises the possibility that higher order moments are important in identifying the SD properties of the assets. This is in contrast to the results of the normality test using annual data reported in Table 1 which showed very little evidence of non-normalities.

Tables 6 to 7 respectively provide SD tests for two pairs of assets: Treasury bond yields and the return on S&P500 ( $r_{tb,t}, r_{sp,t}$ ); and the returns on the two risky assets, S&P500 and

---

<sup>10</sup>The fact that the stochastic dominance tests are based on just asset returns and not consumption data is an important advantage of the approach. This is especially true when testing on daily data as consumption data is measured at a lower frequency. This result is akin to the approach of Campbell (1993) who evaluates the CCAPM having substituted out consumption. Also note that the asset returns used in this example are in nominal terms in contrast to the asset returns defined in the previous example using the Mehra-Prescott data, which are in real terms.

NASDAQ ( $r_{sp,t}, r_{nd,t}$ ). The p-values are based on subsampling with the size of the blocks given by  $B = \alpha \lceil \sqrt{T} \rceil$  with  $\alpha = 4$ . This yields blocks of size  $B = 240$  resulting in 3421 replications to construct the sampling distributions of the test statistics.<sup>11</sup>

Table 6 shows that there is no first or second order SD between Treasury bonds ( $R_{tb,t}$ ) and S&P500 ( $R_{sp,t}$ ). This implies that there is no puzzle as the observed premium between the two assets of just under 4% reported in Table 5 represents an appropriate amount of compensation for agents bearing higher risk who have concave utility functions. Interestingly, there is some evidence of second and higher order SD of Treasury bonds over S&P500 for a nominal size less than 5%; see the 0.00 entries for the "bottom 5%"). This would suggest that there is a puzzle, but in reverse! This dominance possibly reflects the negative skewness in S&P500 (Table 5) whereby agents are not receiving sufficient compensation for bearing negative skewness when they prefer positive skewness.

The results in Table 7 reveal evidence at the 1% level that S&P500 ( $R_{sp,t}$ ) dominates NASDAQ ( $R_{nd,t}$ ) at the third order. There is a lot of "kissing" points between the two curves for low return levels. This last result suggests that, in spite of slight negative skewness in S&P 500, agents with an aversion to higher order volatility and kurtosis in the NASDAQ do not find the premium of just over 4% between the two assets as sufficient compensation. Indeed, this premium would be even larger if the two assets exhibited similar (positive) skewness characteristics.

Overall the SD tests reveal no strong evidence of dominance at the first order in any of the cases investigated. There is some evidence of third order SD of Treasury bills over S&P500, and S&P500 over NASDAQ. This last result reveals the importance of higher order moments, particularly skewness and kurtosis, in determining the risk preferences of agents and the subsequent risk premium observed in the mean. This explains the greater success of those authors (e.g., Epstein and Zin (1991)) who have chosen functionals that allow a role for higher order moments than the mean and the variance.

## 5 Conclusions

This paper has provided a flexible procedure to test for equity premia without the need to specify the underlying utility function or the probability distribution governing returns. The

---

<sup>11</sup>The support of the cumulative distribution functions is based on the range of the data in each block with the number of intermediate points set equal to  $B$ , the size of the blocks.

Table 5:

Descriptive statistics on 3 month Treasury bond yields ( $R_{tb,t}$ ), returns on S&P500 ( $R_{sp,t}$ ) and returns on the NASDAQ ( $R_{nd,t}$ ): expressed as percentage per annum, beginning July 4th, 1989 and ending July 14th 2003.<sup>(a)</sup>

| Statistic | Treas. Bills<br>( $R_{tb,t}$ ) | S&P500<br>( $R_{sp,t}$ ) | NASDAQ<br>( $R_{nd,t}$ ) |
|-----------|--------------------------------|--------------------------|--------------------------|
| Mean      | 4.666                          | 8.446                    | 12.636                   |
| Median    | 5.070                          | 1.235                    | 20.483                   |
| Maximum   | 8.390                          | 1433.898                 | 4335.149                 |
| Minimum   | 0.790                          | -1894.149                | -2615.187                |
| Std. Dev. | 1.762                          | 276.316                  | 500.497                  |
| Skewness  | -0.159                         | -0.144                   | 0.117                    |
| Kurtosis  | 2.739                          | 7.013                    | 7.515                    |
| BJ (p.v.) | 0.000                          | 0.000                    | 0.000                    |

(a) S&P500 and NASDAQ returns computed as the daily difference of the natural logarithms of daily prices, multiplied by 252 to convert daily returns into annualized values, and by 100 to express the returns as a percentage.

Table 6:

SD tests of Treasury yields ( $R_{tb,t}$ ) and S&P500 equity returns ( $R_{sp,t}$ ): July 4th, 1989 and ends July 14th 2003. Bootstraps based on subsampling with  $B = 240$  block sizes and 3421 replications.

| Stochastic<br>Dominance | Null Hypothesis            | Statistic           | Bottom<br>5%        | Top<br>5%           | pv    |
|-------------------------|----------------------------|---------------------|---------------------|---------------------|-------|
| First:                  | Non-maximal                | 29.373              | 6.520               | 7.552               | 0.000 |
|                         | $R_{tb,t}$ $SD$ $R_{sp,t}$ | 29.373              | 6.713               | 8.391               | 0.000 |
|                         | $R_{sp,t}$ $SD$ $R_{tb,t}$ | 30.117              | 6.520               | 8.456               | 0.000 |
| Second:                 | Non-maximal                | 249.298             | 0.000               | 70.166              | 0.000 |
|                         | $R_{tb,t}$ $SD$ $R_{sp,t}$ | 249.298             | 0.000               | 70.166              | 0.000 |
|                         | $R_{sp,t}$ $SD$ $R_{tb,t}$ | 6267.950            | 116.448             | 260.006             | 0.000 |
| Third:                  | Non-maximal                | 0.000               | 0.000               | 0.000               | 0.050 |
|                         | $R_{tb,t}$ $SD$ $R_{sp,t}$ | 0.000               | 0.000               | 0.000               | 0.050 |
|                         | $R_{sp,t}$ $SD$ $R_{tb,t}$ | $2.553 \times 10^6$ | 3162.678            | $1.686 \times 10^4$ | 0.000 |
| Fourth:                 | Non-maximal                | 0.000               | 0.000               | 0.000               | 0.000 |
|                         | $R_{tb,t}$ $SD$ $R_{sp,t}$ | 0.000               | 0.000               | 0.000               | 0.000 |
|                         | $R_{sp,t}$ $SD$ $R_{tb,t}$ | $4.111 \times 10^9$ | $3.129 \times 10^5$ | $1.937 \times 10^6$ | 0.000 |

Table 7:

SD tests of S&P500 equity returns ( $R_{sp,t}$ ) and NASDAQ equity returns ( $R_{nd,t}$ ): July 4th, 1989 and ends July 14th 2003. Bootstraps based on subsampling with  $B = 240$  block sizes and 3421 replications.

| Stochastic<br>Dominance | Null Hypothesis |      |            | Statistic           | Bottom<br>5%        | Top<br>5%           | pv    |
|-------------------------|-----------------|------|------------|---------------------|---------------------|---------------------|-------|
| First:                  | Non-maximal     |      |            | 6.496               | 0.968               | 3.098               | 0.000 |
|                         | $R_{sp,t}$      | $SD$ | $R_{nd,t}$ | 7.124               | 1.226               | 3.357               | 0.000 |
|                         | $R_{nd,t}$      | $SD$ | $R_{sp,t}$ | 6.496               | 0.968               | 3.938               | 0.000 |
| Second:                 | Non-maximal     |      |            | 133.343             | 0.000               | 43.442              | 0.000 |
|                         | $R_{sp,t}$      | $SD$ | $R_{nd,t}$ | 133.343             | 0.000               | 45.185              | 0.000 |
|                         | $R_{nd,t}$      | $SD$ | $R_{sp,t}$ | 2425.769            | 38.407              | 136.781             | 0.000 |
| Third:                  | Non-maximal     |      |            | 0.000               | 0.000               | 0.000               | 0.046 |
|                         | $R_{sp,t}$      | $SD$ | $R_{nd,t}$ | 0.000               | 0.000               | 0.000               | 0.048 |
|                         | $R_{nd,t}$      | $SD$ | $R_{sp,t}$ | $1.317 \times 10^6$ | 2310.493            | $1.195 \times 10^4$ | 0.000 |
| Fourth:                 | Non-maximal     |      |            | 0.000               | 0.000               | 0.000               | 0.011 |
|                         | $R_{sp,t}$      | $SD$ | $R_{nd,t}$ | 0.000               | 0.000               | 0.000               | 0.022 |
|                         | $R_{nd,t}$      | $SD$ | $R_{sp,t}$ | $2.950 \times 10^9$ | $2.281 \times 10^5$ | $1.455 \times 10^6$ | 0.000 |

approach is nonparametric, being based on testing for SD. The tests for various orders of SD helped to reveal how higher order moments are priced and, in turn, whether the observed premium in equities was sufficient compensation for bearing risk.

The empirical results found little evidence of SD in both of our data sets. There was some weak evidence of third and higher order SD of equities over bonds in the Mehra and Prescott annual data, but only at 1%, and not at 5%. The empirical results using daily data revealed no first or second order dominance between Treasury bills and S&P500. There was weak evidence of third order SD of Treasury bills over S&P500, suggesting that some agents ranked the risk free asset over the risky asset when pricing skewness. This result also suggested that the observed equity premium might in fact be too small to compensate agents adequately for bearing higher risk associated with S&P500. Finally, there was no evidence of either first or second order SD between the risky assets, S&P500 and NASDAQ. However, there was some evidence that S&P500 third and fourth order stochastically dominated NASDAQ. Given that S&P500 exhibited negative skewness and NASDAQ positive skewness, this suggested that the observed premium between the two assets would be even higher if they exhibited the same skewness characteristics.

One implication of the lack of SD is that many of the existing models may be based on either inadequate utility functions, or incorrect returns distributions, or both. It also suggests that there exist utility functions and appropriate probability distributions that will generate “acceptable” risk aversion parameter estimates. That is, the search could be fruitful! The results point to the need to search over probability distributions that capture higher order moments in preferences, such as skewness and kurtosis. This result is interesting given that most of the specifications have focussed on respecifying the preference function. Furthermore, the lack of SD results suggest that research that has been devoted to formulating models that depart from the assumptions of complete and frictionless markets may be useful in so far as they are informative about the nature of preferences and about higher order moments in the probability distributions of the assets; see also the work of Grant and Quiggin (2001).

The empirical results presented can be extended in a number of ways. First, the returns can be conditioned on a set of factors representing the state of the economy, for example the different phases of the business cycle. The approach would be to run an auxiliary regression of each of the returns series on a set of factors, including a constant term, and use the residuals from this regression in the SD tests. Second, the assumption of expected utility

theory can be partially relaxed by considering S-shaped utility functions and performing Prospect Dominance tests following the approach of Linton, Maasoumi and Whang (2004). Third, the daily data results can be extended to computing the McFadden maximality test over the full set of assets investigated so as to provide an overall ranking. Fourth, a number of robustness checks on the empirical results could be carried out, including sensitivity to the design of the resampling procedures. Finally, the framework presented here can also be applied to testing the validity of other puzzles such as the risk free puzzle.

## References

- [1] Abel, A.B. (1990), "Asset Prices Under Habit Formation And Catching Up With The Joneses", *American Economic Review*, 80, 38-42.
- [2] Abhyankar, A. and Ho, K. (2003), "Exploring Long-Run Abnormal Performance Using Stochastic Dominance Criteria: Additional Evidence From IPOs", mimeo.
- [3] Aiyagari, S.R. and Gertler, M. (1991), "Asset Returns with Transaction Costs and Uninsured Individual Risk", *Journal of Monetary Economics*, 27, 311-331.
- [4] Bansal, R. and Coleman, J.W. (1996), "A Monetary Explanation of the Equity Premium, Term Premium and Risk Free rate Puzzles", *Journal of Political Economy*, 104, 1135-1171.
- [5] Barrett, G. and Donald, S. (2003), "Consistent Tests for Stochastic Dominance", *Econometrica*, 71 (1), 71-104.
- [6] Benartzi, S. and Thaler, R.H. (1995), "Myopic Loss Aversion and the Equity Premium Puzzle", *Quarterly Journal of Economics*, 110, 73-92
- [7] Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992), "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence", *Journal of Econometrics*, 52, 5-59.
- [8] Campbell, J.Y. (1993), "Intertemporal Asset Pricing Without Consumption", *American Economic Review*, 83, 487-512.
- [9] Campbell, J.Y. and Cochrane, J.H. (1999), "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behaviour", *Journal of Political Economy*, 107, 205-251.

- [10] Campbell, J.Y. and Cochrane, J.H. (2000), “Explaining the Poor Performance of Consumption-based Asset Pricing Models”, *Journal of Finance*, LV, 2863-2878.
- [11] Campbell, J.Y., Lo, A.W., and MacKinlay, A.C. (1997), *The Econometrics of Financial Markets*, Princeton University Press.
- [12] Epstein, L.G. and Zin, S.E. (1991), “Substitution, Risk Aversion and the Temporal Behaviour of Consumption and Asset Returns: An Empirical Analysis”, *Journal of Political Economy*, 99, 263-286.
- [13] Fama, E.F. and French, K. (2002), “The Equity Premium”, *Journal of Finance*, LVII, 637-659.
- [14] Grant, S. and Quiggin, J. (2001a), “The Risk Premium for Equity: Explanations and Implications”, mimeo.
- [15] Grossman, S., Melino, A. and Shiller, R. (1987), “Estimating the Continuous Time Consumption Based Asset Pricing Model”, *Journal of Business and Economic Statistics*, 5, 315-328.
- [16] Hansen, L. and Singleton, K. (1983), “Stochastic Consumption, Risk Aversion and the Temporal Behavior of Asset Returns”, *Journal of Political Economy*, 91, 249-268.
- [17] Harvey, C.R. and Siddique, A. (2000), “Conditional Skewness in Asset Pricing Tests”, *Journal of Finance*, LV, 1263-1295.
- [18] Heaton, J. and Lucas, D. (1995), “The Importance of Investor Heterogeneity and Financial Market Imperfections for the Behaviour of Asset Prices”, *Carnegie-Rochester Series in Public Economics*.
- [19] Kocherlakota, N.R. (1996), “The Equity Premium: It’s Still a Puzzle”, *Journal of Economic Literature*, 34, 42-71.
- [20] Kolmogorov, A.N. (1933), “On the Empirical Determination of a Distribution Law”, reprinted in “Selected Works of A.N. Kolmogorov”, *Probability Theory and Mathematical Statistics*, Vol.II, A.N. Shiriyayev (ed.), Kluwer, Dordrecht.
- [21] Lim, G.C., Martin, G.M. and Martin, V.L. (2004), “Parametric Pricing of Higher Order Moments in S&P500 Options”, *Journal of Applied Econometrics*, forthcoming.

- [22] Linton, O., Maasoumi, E. and Whang, Y. (2004), "Consistent Testing for Stochastic Dominance Under General Sampling Schemes", *Review of Economic Statistics*, forthcoming.
- [23] McFadden, D. (1989), "Testing for Stochastic Dominance", in Part II of T. Fomby and T.K. Seo (eds.), *Studies in the Economics of Uncertainty*, (in honor of J. Hadar), Springer-Verlag.
- [24] McGrattan, E. and Prescott, E.C. (2001), "Taxes, Regulations, and Asset Prices", Working Paper 610, Federal Reserve Bank of Minneapolis.
- [25] Mehra R. (2003), "The Equity Premium: Why is it a Puzzle", *Financial Analysts Journal*, 59, 54-69.
- [26] Mehra, R. and Prescott, E.C. (1985), "The Equity Premium: A Puzzle", *Journal of Monetary Economics*, 15, 145-161.
- [27] Mehra, R. and Prescott, E.C. (2003), "The Equity Premium in Retrospect", NBER Working Paper Series, 9525.
- [28] Politis, D.N., Romano, J.P. and Wolf, M. (1999), *Subsampling*, Springer-Verlag, New York.
- [29] Rietz, T.A. (1988), "The Equity Risk Premium: A Solution", *Journal of Monetary Economics*, 22, 117-131.
- [30] Siegel, J.J. and Thaler, R.H. (1997), "Anomalies: The Equity Premium Puzzle", *Journal of Economic Perspectives*, 11, 191-2000.
- [31] Weil, P. (1992), "Equilibrium Asset Prices with Undiversifiable Labor Income Risk", *Journal of Economic Dynamics and Control*, 16, 769-790.