# Discounting The Equity Premium Puzzle

G.C. Lim
University of Melbourne

Esfandiar Maasoumi
Southern Methodist University

Vance L. Martin*
University of Melbourne

June 2004

**Abstract**

Recent tests of stochastic dominance of several orders proposed by Linton, Maasoumi and Whang (2003) are applied to reexamine the equity premium puzzle. An advantage of this nonparametric framework is that it provides a means to assess whether the existence of a premium is due to particular cardinal choices of either the utility function or the underlying returns distribution, or both. The approach is applied to a number of data sets including the original Mehra-Prescott data and more recent data that includes daily yields on Treaury bonds and commercial paper, and daily returns on the S&P500 and the NASDAQ indexes. The empirical results show little evidence of stochastic dominance amongst the assets investigated. This suggests that there is no puzzle and that the observed equity premium indeed represents the price for bearing higher risk, taking into account higher order moments such as skewness and kurtosis.

Key words: Equity premium puzzle, stochastic dominance, nonparametric, subsampling, recentered bootstraps.

*The authors thank participants at the Econometrics seminar at Monash University for helpful comments. Corresponding author: Vance L. Martin, Economics Department, University of Melbourne, Victoria, 3010, Australia. Ph: +61 3 8344 5396. Fax: +61 3 8344 6899. Email: vance@unimelb.edu.au.

# 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 observed premium between real returns on investments in equity and the real yields from investing in 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.[1] What makes the puzzle so important is that it is empirically robust as it arises in different sample periods, occurs for a broad selection of assets and is characteristic of many international financial markets (Mehra (2003)). The empirical observation of the premia is, therefore, a robust fact!

The equity premium puzzle can be viewed as the manifestation of misspecification error on the 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 sole aim to derive empirically sensible estimates of the risk aversion parameter. This specification search of theoretical models can be categorized into three broad groups. The first class of models focuses on preferences. This class of models looks at extending

[1] 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.

existing parametric utility functions by allowing for generalized expected utility (Epstein and Zin (1989, 1991)); habit formation (Constantinides (1990)); relative consumption (Abel (1990)); and subsistence consumption (Campbell and Cochrane (1999)). 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 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)).

An important characteristic of the proposed theoretical models to explain the equity premium puzzle 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 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.[2] The approach consists of couching the equity

[2] 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

premium puzzle in terms of testing for various levels of stochastic dominance between the returns on equities and bonds. The non-existence of any stochastic dominance ranking, especially of first and second order, means that for agents with Von Neumann-Morgenstern concave utility functions, investment in equity, for example, is not sufficiently attractive to invest in without a substantial premium. The expected utility paradigm suggests that. To quantify what is a reasonable 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 any evidence of a "premium puzzle" is necessarily an artifact of the specific functionals chosen if there is no stochastic dominance. Non-dominance, or "maximality", implies that there is no uniform (weak) ranking over the risk free asset, and there are indeed *some* functionals, utility functions and probability distributions, that would result in any "strong" ranking one may desire! In fact, according to some functionals, the 6 % differential initially observed by Mehra and Prescott (1985) may be too little, and almost surely so for some risk averse individuals. It is believed that Stochastic Dominance testing provides an alternative approach which overcomes the twin and intertwined obstacles of cardinal utility identification and heterogeneity in asset returns.

The rest of the paper proceeds as follows. Empirical evidence of the equity premium and the risk aversion parameter 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.

---

and kurtosis in the pricing of options.

# 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. A description of the definitions of the variables and the sources is given in Appendix A.

## 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

$$PREMIUM = 6.980 - 1.036 = 5.944\%,$$

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. This is supported by the statistics on the Sharpe ratio (mean divided by the standard deviation) which show that the mean return per unit of risk of equities is 42.196%, which is greater than 18.076%, the corresponding Sharp ratio for bonds. 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 $-0.107$).

## 2.2 Estimates of Relative Risk Aversion

The second form of the equity premium puzzle that is commonly presented is in terms of estimates of the relative risk aversion parameter, $\gamma$. Formally, this parameter is identified by specifying the stochastic discount factor which forms the basis of pricing financial assets. Let $R_t$ be a vector of $N$ asset returns. The pricing equation is (see Campbell, Lo and MacKinlay (1997))

$$1 = E_t\left[\left(1 + R_{t+1}\right) M_{t+1}\right], \tag{1}$$

where $M_t$ is the stochastic discount factor and $E_t\left[\cdot\right]$ is the conditional expectation operator. This model is used to price all financial assets with $i = 1, 2, \cdots, N$, representing the number of assets. In the case of the consumption based capital asset pricing model (CCAPM), $M_{t+1}$ is the ratio of the (discounted) future and present marginal utilities, with utility expressed as a function of consumption. The pricing equation becomes

$$1 = \int \cdots \int \left(1 + R_{t+1}\right) M_{t+1}\left(C_{t+1}, C_t; \delta, \gamma, \Psi\right) f\left(C_{t+1}, R_{t+1} \middle| \Omega_t\right) dC_{t+1} R_{t+1}, \tag{2}$$

where $\delta$ is the parameter used to discount future utility, $\gamma$ is the relative risk aversion parameter, $\Psi$ represents an additional set of parameters that characterize the risk aversion of agents, and $f\left(C_{t+1}, R_{t+1} \middle| \Omega_t\right)$ is the multivariate conditional distribution that assigns probabilities to the states of nature based on the information set $\Omega_t$. This expression contains the basis of the models commonly proposed to explain the equity premium puzzle. As noted in the introduction, most of the effort has been devoted to modelling preferences. This is represented by adopting different functional forms for the stochastic discount factor $M_{t+1}(C_{t+1}, C_t; \delta, \gamma, \Psi)$. The second line of research has focussed on the specification of the conditional distribution $f(C_{t+1}, R_{t+1} | \Omega_t)$. For many of the models this distribution is commonly chosen to be multivariate lognormal.

By specifying a power utility function, the stochastic discount factor in (2) is simply parameterized in terms of the discount parameter ($\delta$) and the relative risk aversion parameter ($\gamma$), with $\Psi = 0$.[3] This model leads to a range of alternative expressions which have been used to estimate $\gamma$. Estimates of $\gamma$ from some of these approaches are given in Table 3 using the Mehra-Prescott data. See Appendix B for the details of these calculations. The first observation to make is that the estimates of this parameter are not robust, ranging from as high as 46.926 to a low of 1.799! Psychologists and experimentalists have found similarly disconcerting wide ranges for this parameter. Second, the equity premium puzzle is predicated on an important decision in Mehra and Prescott (1985); namely, that estimates of $\gamma$ in excess of 10 constitute excessive risk aversion which are inconsistent with empirical studies documented at that time.

# 3 Stochastic Dominance Testing

This section outlines the framework for conducting stochastic dominance tests in the context of the equity premium puzzle. The approach is based on the work of Linton, Maasoumi and Whang (2003) who propose nonparametric tests of stochastic dominance using Kolmogorov-Smirnov type tests and the McFadden (1989) maximality test. Inference is performed by using subsampling to construct p-values as well as bootstrapping methods. A re-

[3] The power utility function is

$$u(C_t) = \frac{C_t^{1-\gamma} - 1}{1-\gamma}.$$

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 $(100 \times R_{s,t})$ | Bonds $(100 \times R_{b,t})$ | Consump. $(100 \times R_{c,t})$ |
|---|---|---|---|
| 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 |
| Sharpe ratio$^{(a)}$ | 42.196 | 18.076 | 50.922 |

(a) Computed as the sample mean divided by the standard deviation and expressed in percentage terms.

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 $R_t^s$ | Bonds $R_t^b$ | Consump. $R_t^c$ |
|---|---|---|---|
| $R_t^s$ | 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.[(a)]

| Model | Method and source | $\gamma$ |
|---|---|---|
| 1 | Mehra (2003, equation 15) | 26.085 |
| 2 | Mehra (2003, equation 16) | 46.926 |
| 3 | CLM (1997, equation 8.2.9): no instruments[(b)] | 1.799 |
| 4 | CLM (1997, equation 8.2.10): no instruments | 11.062 |
| 5 | CLM (1997, equation 8.2.9): with instruments | 1.823 |
| 6 | CLM (1997, equation 8.2.10): with instruments | 3.351 |
| 7 | Hansen and Singleton (1983): GMM | 15.397 |
| 8 | Grossman, Melino and Shiller (1987) | 24.755 |

(a) See Appendix B for details of the calculations.

(b) CLM is an abbreviation for Campbell, Lo and MacKinlay.

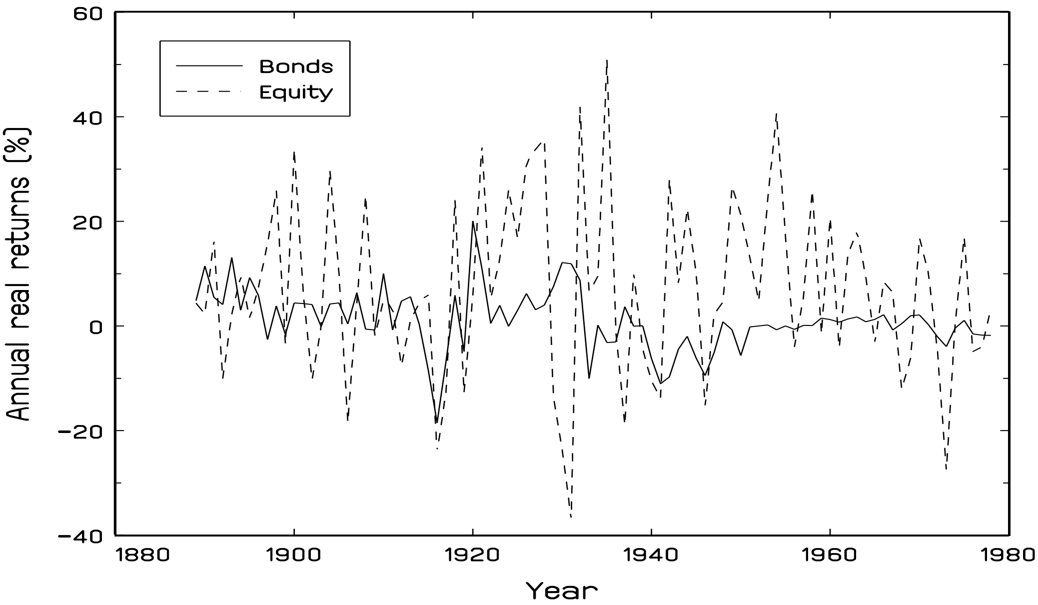


Figure 1: Bond yields and equity returns: real, percentage per annum, 1889 to 1978.

lated 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 testing framework.[4]

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

### 3.1 Definitions

Consider two stationary time series of returns, $R_{i,t}$ and $R_{j,t}$, $t = 1, 2, \cdots, T$, with respective cumulative distribution functions, $F_i(r)$ and $F_j(r)$, over the support $r$. The returns are not expected to be *iid*, but can exhibit some dependency structures in the moments of the distribution.[5] The null hypotheses that $R_{i,t}$ stochastically dominates $R_{j,t}$, for various orders are defined as follows:

$$\begin{array}{lll} H_0: & \text{(First order)} & F_i(r) \leq F_j(r) \\ H_0: & \text{(Second order)} & \int_0^r F_i(t)\,dt \leq \int_0^r F_j(t)\,dt \\ H_0: & \text{(Third order)} & \int_0^r \int_0^t F_i(s)\,dsdt \leq \int_0^r \int_0^t F_j(s)\,dsdt \\ H_0: & \text{(Fourth order)} & \int_0^r \int_0^t \int_0^s F_i(u)\,dudsdt \leq \int_0^r \int_0^t \int_0^s F_j(u)\,dudsdt. \end{array} \quad (3)$$

The alternative hypothesis is that there is no stochastic dominance. From the definitions of first, second, third and fourth order stochastic dominance, if $R_{i,t}$ first order stochastically dominates $R_{j,t}$, then it stochastically dominates $R_{j,t}$ at all orders, and so on. 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

$$\Pr(R_{i,t} > r) \geq \Pr(R_{j,t} > r). \quad (4)$$

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

[5] Formally, the returns processes are assumed to be strictly stationary and $\alpha - mixing$ with $\alpha(j) = O\left(j^{-\delta}\right)$, for some $\delta > 1$.

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 non-parametric 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

$$E\left[u\left(R_{i,t}\right)\right] \geq E\left[u\left(R_{j,t}\right)\right], \text{where } u' \geq 0. \tag{5}$$

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))

$$\ln E_t\left[\frac{(1+R_{s,t+1})}{(1+R_{b,t+1})}\right] = \gamma\sigma_{s,c}, \tag{6}$$

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

$$E\left[u\left(R_{i,t}\right)\right] \geq E\left[u\left(R_{j,t}\right)\right], \text{where } u' \geq 0, u'' \leq 0. \tag{7}$$

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 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

$$E\left[u\left(R_{i,t}\right)\right] \geq E\left[u\left(R_{j,t}\right)\right], \text{where } u' \geq 0, u'' \leq 0, u''' \geq 0. \tag{8}$$

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.

Fourth order stochastic dominance relates to 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 $j$, within the class of utility functions with positive first and third derivatives and negative second and fourth derivatives

$$E\left[u\left(R_{i,t}\right)\right] \geq E\left[u\left(R_{j,t}\right)\right], \text{where } u' \geq 0, u'' \leq 0, u''' \geq 0, u'''' \leq 0. \tag{9}$$

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.

Figures 2 to 7 highlight the stochastic dominance features of a number of hypothetical asset return distributions. In Figure 2 the returns distributions are both normal with common volatility, $\sigma_1 = \sigma_2 = 6$, but with different means $\mu_1 = 1$ and $\mu_2 = 6$. Here $F_2$ first order stochastically dominates $F_1$ as asset 2 yields a higher mean return than asset 1 $(\mu_2 > \mu_1)$ for the same level of risk $(\sigma_2 = \sigma_1)$. This dominance continues for higher orders. The equity premium of $\mu_2 - \mu_1 = 5$, in this case represents a puzzle as the relatively higher return earnt from investing in asset 2 comes without any additional risk. Within the class of utility functions that exhibit nonsatiation, asset 2 stochastically dominates asset 1.

In Figure 3, the returns distribution are both normal with common mean, but with differing volatilities. Unlike in Figure 2, there is no evidence of first order stochastic dominance. In contrast, however, $F_1$ second order stochastically dominates $F_2$, as asset 1 has lower risk than asset 2 ($\sigma_2 < \sigma_1$) whilst the mean returns are the same ($\mu_2 = \mu_1$). Within the class of concave utility functions, asset 2 stochastically dominates asset 1. The expected return on asset 2 is too low relative to the higher risk associated with this asset. This is demonstrated in Figure 4 where now asset 2 exhibits a higher average return to compensate for the higher risk (compare the distribution of asset 2 in Figures 3 and 4). As Figure 4 shows no evidence of stochastic dominance of any order between the two assets, this suggests that 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. A similar result occurs in Figure 5 where asset 2 exhibits relatively fatter tails (Student t with $v = 2.5$ degrees of freedom), but is compensated by a relatively higher mean return than asset 1. In this and the previous example, the two assets are unrankable (maximal) as rational agents are indifferent between the two assets.

The effects of skewness as well as kurtosis in the returns distribution are highlighted in Figures 6 and 7, where the distribution of asset 2 is based on the generalized Student t distribution of Lye and Martin (1993), whilst the distribution of asset 1 is still normal.[6] In Figure 6 $F_1$ fourth order stochastically dominates $F_2$, whereas in Figure 7 there is no stochastic dominance of any order.

[6]The generalised Student t distribution $GST(\mu, \sigma, \nu, \theta)$ is given by

$$f(r) = \exp\left[\theta \tan^{-1}\left(z/\sqrt{\nu}\right) - 0.5\left(1+\nu\right)\ln\left(z^2+v\right) - 0.5z^2 - \eta\right],$$

with $z = (r-\mu)/\sigma$, and $\eta$ is the normalising constant to ensure that the distribution integrates to unity.

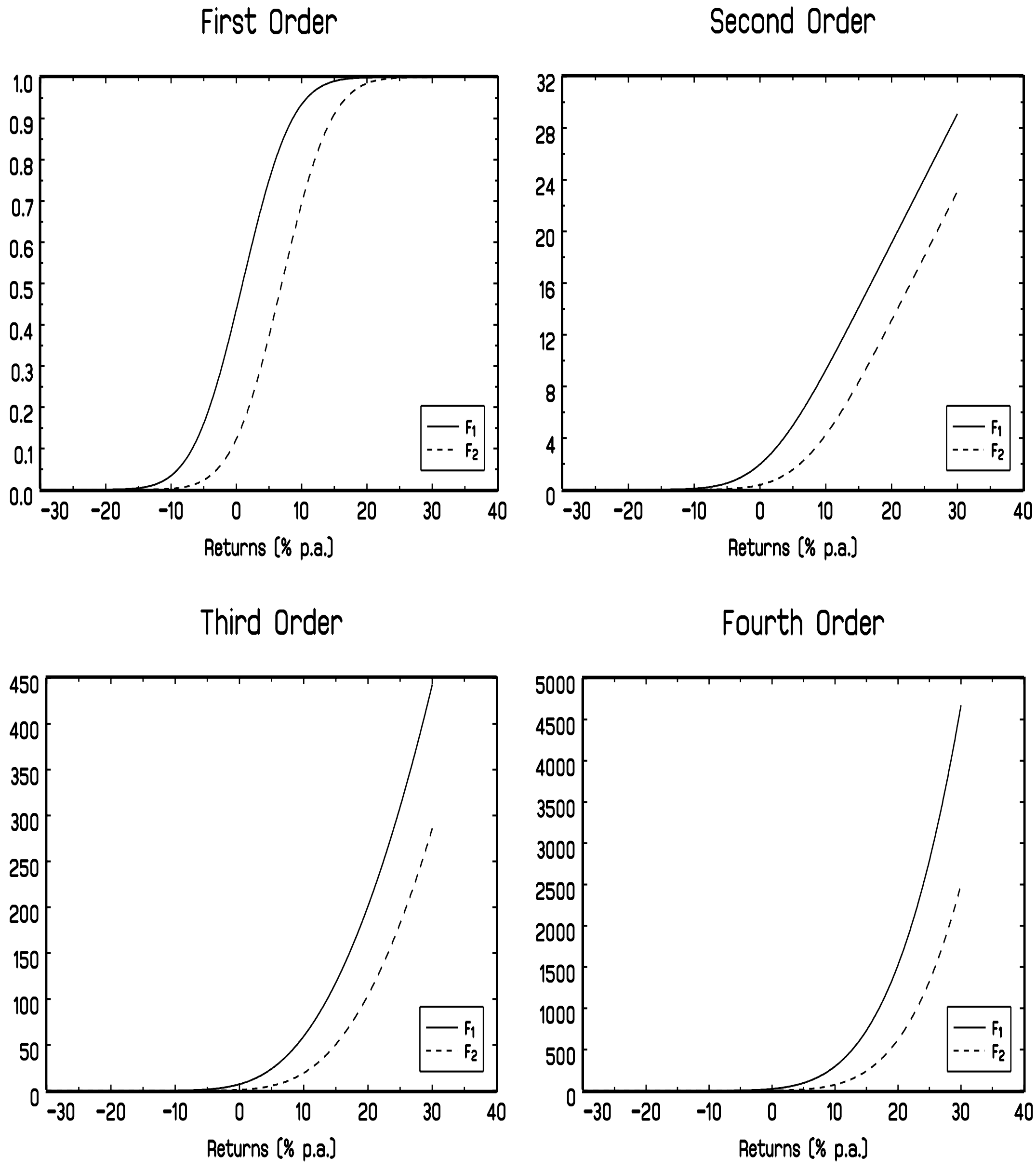


Figure 2: Hypothetical asset returns distributions, first to fourth order stochastic dominance as defined in (3): $F_1 = N(1, 6^2)$, $F_2 = N(7, 6^2)$.

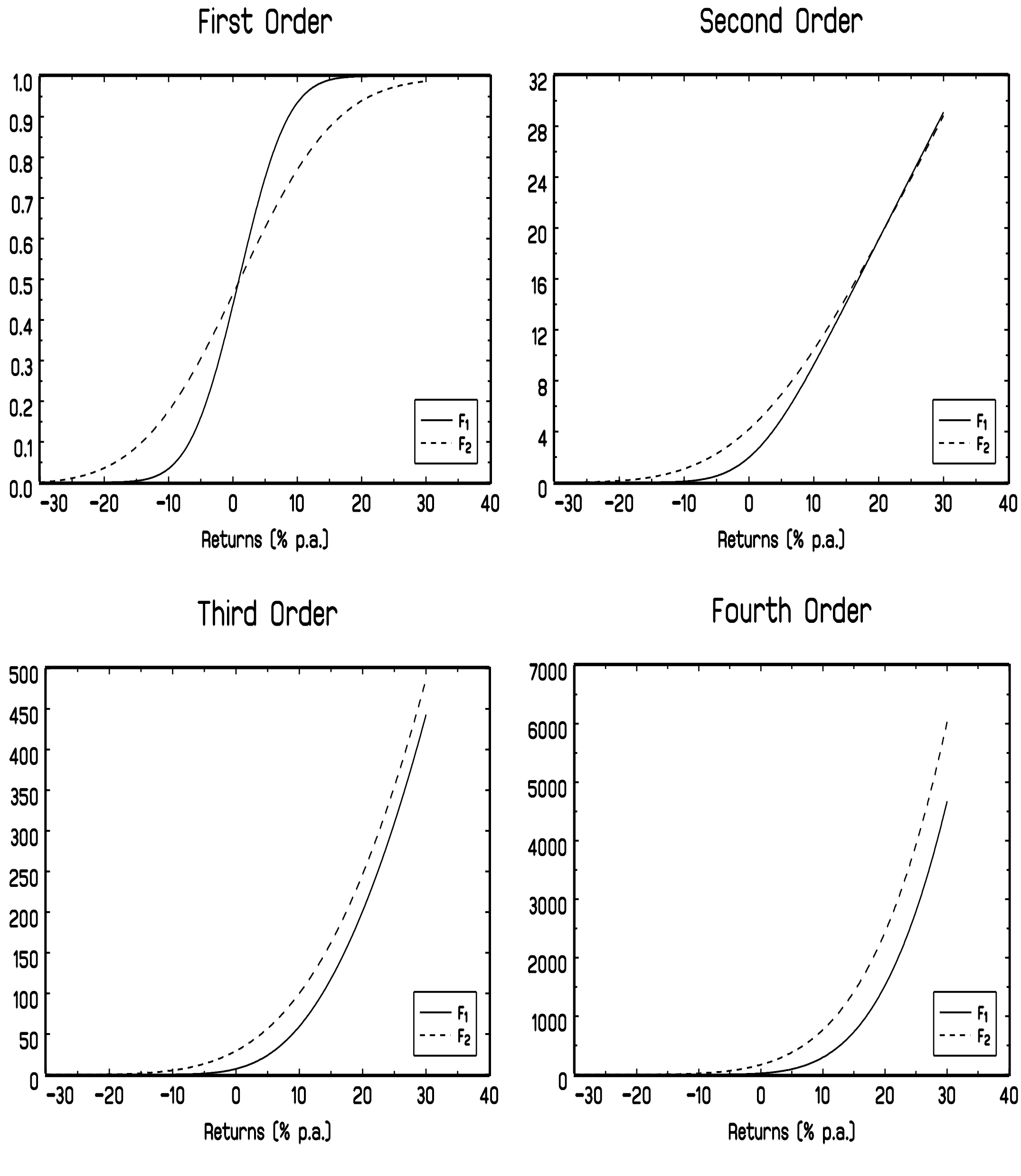


Figure 3: Hypothetical asset returns distributions, first to fourth order stochastic dominance as defined in (3): $F_1 = N(1, 6^2)$, $F_2 = N(1, 12^2)$.

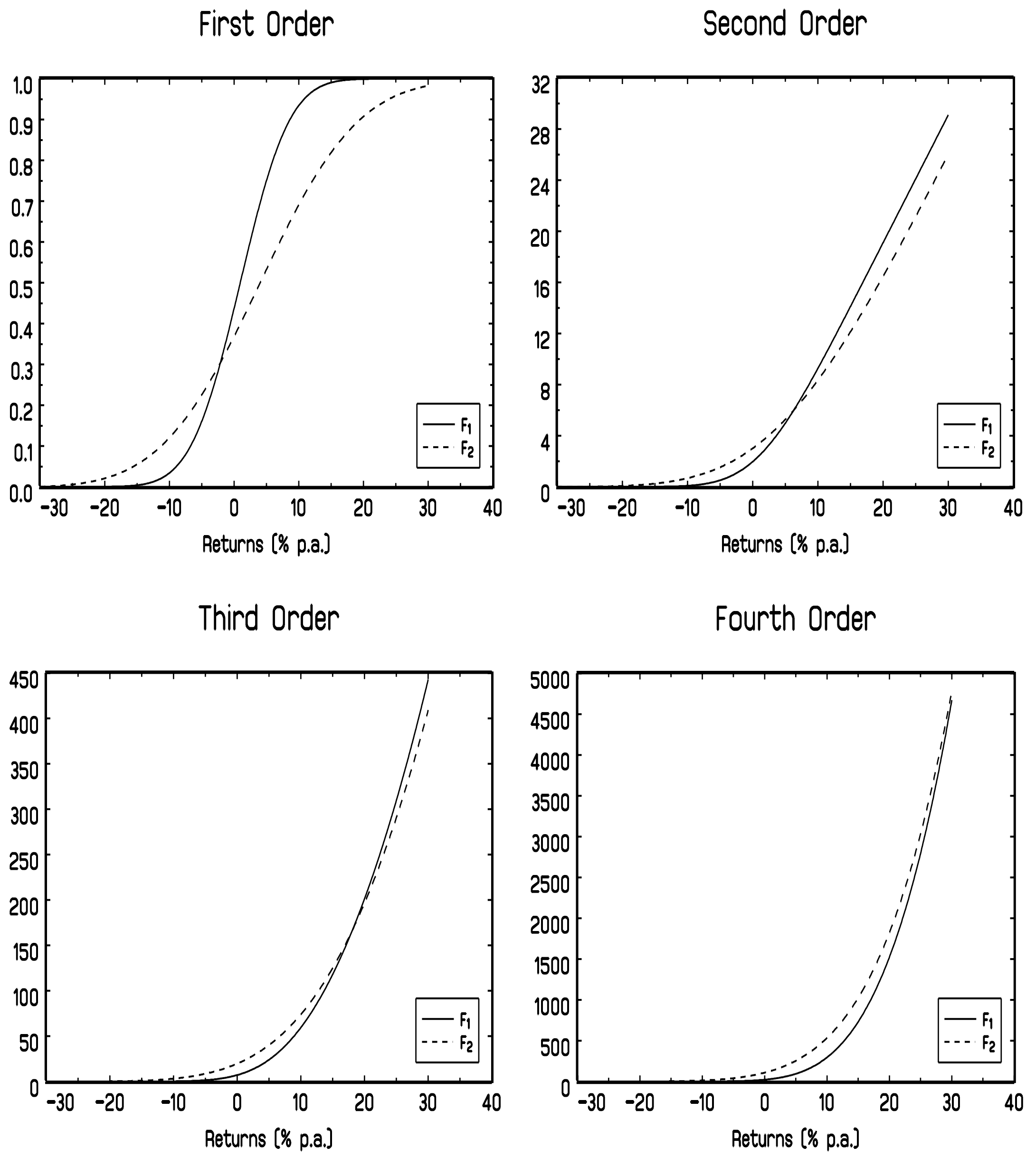


Figure 4: Hypothetical asset returns distributions, first to fourth order stochastic dominance as defined in (3): $F_1 = N(1, 6^2)$, $F_2 = N(6, 12^2)$.

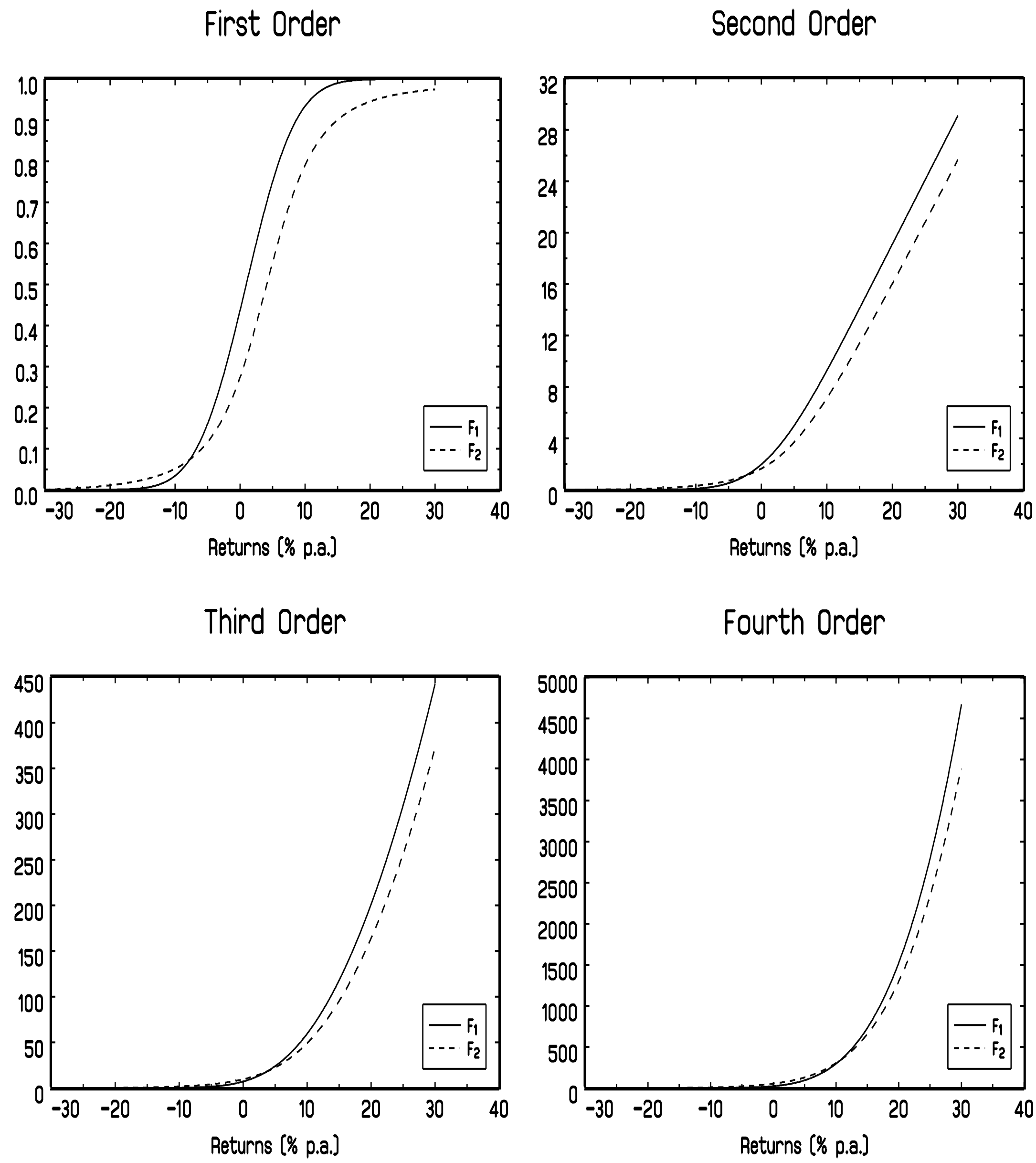


Figure 5: Hypothetical asset returns distributions, first to fourth order stochastic dominance as defined in (3): $F_1 = N\left(1, 6^2\right)$, $F_2 = St\left(3, 6^2, 2.5\right)$.

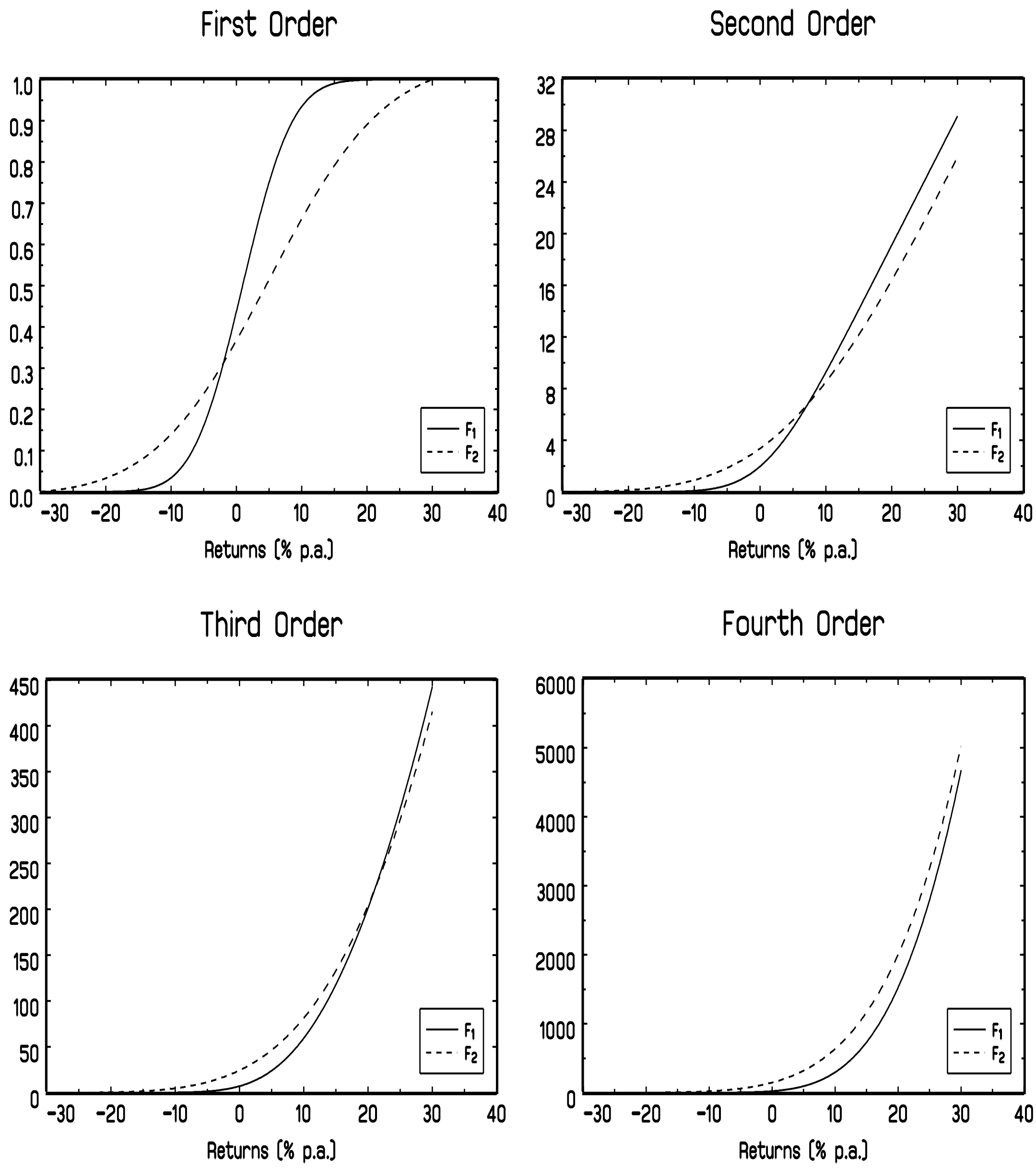


Figure 6: Hypothetical asset returns distributions, first to fourth order stochastic dominance as defined in (3): $F_1 = N\left(1, 6^2\right)$, $F_2 = GST\left(1, 20^2, 5, 1\right)$.

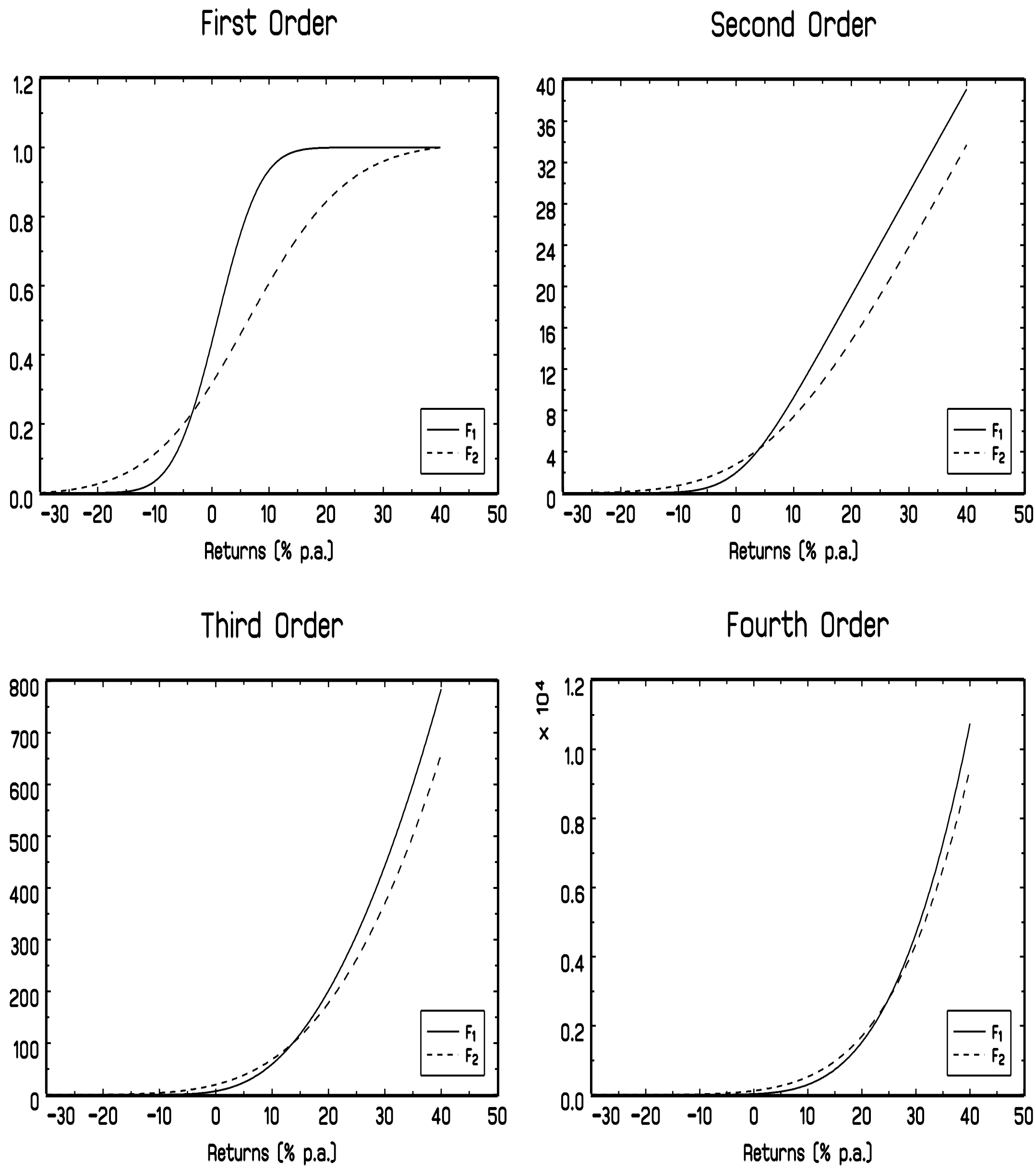


Figure 7: Hypothetical asset returns distributions, first to fourth order stochastic dominance as defined in (3): $F_1 = N(1, 6^2)$, $F_2 = GST(1, 20^2, 2.55, 1)$

## 3.2 Testing

### 3.2.1 First Order

Consider testing the null hypotheses that $R_{i,t}$ first order stochastically dominates $R_{j,t}$, using the approach of Linton, Maasoumi and Whang (2003). The test statistic is

$$SD_{1,i,j} = \sqrt{T} \sup_{r} \left( \widehat{F}_i(r) - \widehat{F}_j(r) \right), \tag{10}$$

where $T$ is the sample size, and $\widehat{F}_i(r)$ and $\widehat{F}_j(r)$ are the respective empirical cumulative distribution functions of $R_{i,t}$ and $R_{j,t}$

$$\begin{aligned} \widehat{F}_i(r) &= \frac{1}{T} \sum_{t=1}^{T} I(R_{i,t} \leq r) \qquad (11) \\ \widehat{F}_j(r) &= \frac{1}{T} \sum_{t=1}^{T} I(R_{j,t} \leq r), \end{aligned}$$

and

$$I(R_{i,t} \leq r) = \begin{cases} 1: & R_{i,t} \leq r \\ 0: & R_{i,t} > r \end{cases}, \tag{12}$$

is the indicator function. The test statistic is based on the Kolmogorov-Smirnov test which equals the maximum distance between the two empirical cumulative distributions, $\widehat{F}_i(r)$ and $\widehat{F}_j(r)$.

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 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 (10) results in a value of the test statistic of $SD_{1,i,j} = 0$. If the null is false then either there is no stochastic dominance, in which case the two cumulative distribution functions cross, or $R_{i,t}$ is first order stochastically dominated by $R_{j,t}$. In either case the test statistic is positive, $SD_{1,i,j} > 0$. To test that $R_{j,t}$ indeed first order stochastically dominates $R_{i,t}$, (10) is reversed resulting in the statistic

$$SD_{1,j,i} = \sqrt{T} \sup_{r} \left( \widehat{F}_j(r) - \widehat{F}_i(r) \right). \tag{13}$$

The test statistics (10) and (13), can be combined to provide an overall maximality test of first order stochastic dominance following McFadden (1989)

$$MF_1 = \min_{i \neq j} \left( SD_{1,i,j}, SD_{1,j,i} \right). \tag{14}$$

Under the null, one of the assets is stochastically dominant, whereby the value of the test statistic is $MF_1 \leq 0$. Under the alternative hypothesis there is no stochastic dominance. As the empirical cumulative distribution functions must cross under the alternative, the test statistic produces a positive value, $MF_1 > 0$. In this case the assets are maximal, that is, they are unrankable. In the context of the equity premium puzzle both assets are appropriately priced by the market and any premium simply reflects the price of bearing higher risk.

The maximality test statistic in (14) can be extended to testing for maximality amongst more than two assets to provide an initial test of maximality.[7] If the null is rejected, no stochastic dominance exists amongst the assets. If the null is not rejected, then there is evidence of stochastic dominance. To identify the nature of the stochastic dominance it is necessary to perform the individual stochastic dominance tests in (10) and (13).

In the case of *iid* data, the sampling distributions of (10) and (13) under the null was originally derived by Kolmogorov (1933), whilst McFadden (1989) derived the sampling distribution of (14). 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[8]. 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

[7] Care needs to be made in implementing this testing strategy as the support of the cumulative distribution functions needs to be chosen to cover the full range of the full data set.

[8] Note that "pivotal" statistics are therefore not available.

important advantage of resampling is that it can accommodate dependence in asset returns over both time and contemporaneously, as demonstarted in Table 2. The approach consists of dividing the data into $T-B+1$ overlapping blocks of size $B$, to be determined below. The first (paired) block for the asset returns $R_{1,t}$ and $R_{2,t}$, is given by

$$\begin{bmatrix} R_{1,1}, & R_{1,2}, & \cdots & R_{1,B} \\ R_{2,1}, & R_{2,2}, & \cdots & R_{2,B} \end{bmatrix}.$$

The second block is

$$\begin{bmatrix} R_{1,2}, & R_{1,3}, & \cdots & R_{1,B+1} \\ R_{2,2}, & R_{2,3}, & \cdots & R_{2,B+1} \end{bmatrix},$$

while the last block is

$$\begin{bmatrix} R_{1,T-B+1}, & R_{1,T-B+2}, & \cdots & R_{1,T} \\ R_{2,T-B+1}, & R_{2,T-B+2}, & \cdots & R_{2,T} \end{bmatrix}.$$

For each block, the test statistics $SD_{1,i,j}, SD_{1,j,i}$ and $MF_1$, are computed. Let the corresponding test statistics based on the $k^{th}$ subsample value be respectively denoted as

$$\begin{aligned} SD_{1,i,j,k} &= \sqrt{B}\sup_r \left(\widehat{F}_{i,k}(r) - \widehat{F}_{j,k}(r)\right) \\ SD_{1,j,i,k} &= \sqrt{B}\sup_r \left(\widehat{F}_{j,k}(r) - \widehat{F}_{i,k}(r)\right) \\ MF_{1,k} &= \min_{i\neq j}\left(SD_{1,i,j,k}, SD_{1,j,i,k}\right), \end{aligned} \tag{15}$$

where $\widehat{F}_{i,k}(r)$ and $\widehat{F}_{j,k}(r)$ are the empirical distribution functions based on the $k^{th}$ block of asset returns $R_{i,t}$ and $R_{j,t}$ respectively. The pertinent p-values for the three tests of first order stochastic dominance are computed as

$$pv_{1,i,j} = \frac{1}{T-B+1}\sum_{k=1}^{T-B+1} I\left(SD_{1,i,j}(k) \leq SD_{1,i,j}\right) \tag{16}$$

$$pv_{1,j,i} = \frac{1}{T-B+1}\sum_{k=1}^{T-B+1} I\left(SD_{1,j,i}(k) \leq SD_{1,j,i}\right) \tag{17}$$

$$pv_{1,mp} = \frac{1}{T-B+1}\sum_{k=1}^{T-B+1} I\left(MF_1(k) \leq MF_1\right). \tag{18}$$

A p-value less than a nominal size of $\alpha$ leads to rejection of the null hypothesis. In performing the subsampling procedure to compute the p-values, the time dependence structure in returns is captured by extracting time series runs of the data, whilst the contemporaneous dependence is modelled by matching the same time period for each return series across the simulation runs.

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. Linton, Maasoumi and Whang (2003) approach the problem by performing a sensitivity analysis on the block size to establish the robustness properties of the subsampling procedure. 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

$$B = \alpha \left[ \sqrt{T} \right], \tag{19}$$

where $\left[ \sqrt{T} \right]$ denotes the largest integer that is less than or equal to $\sqrt{T}$, and $\alpha$ is a constant.

An alternative approach to subsampling for deriving the sampling distribution of the test statistics is to use a recentered bootstrap with overlapping blocks. The procedure consists of randomly drawing with replacement from the set of paired blocks used in the subsampling scheme. These blocks are then stacked to form the sample used in the resampling scheme. The number of blocks chosen is based on constructing a bootstrap sample size comparable to the sample size of the data, $T$. In performing the bootstrapping procedure the test statistics at each bootstrap sample are recentered using the empirical distribution function corresponding to each order of dominance being tested following the approach of Linton, Maasoumi and Whang (2003). Formally, the test statistics in (15) are reexpressed as

$$\begin{aligned} SD^{c}_{1,i,j,k} &= \sqrt{T} \sup_{r} \left( \widehat{F}_{i,k}(r) - \widehat{F}_{j,k}(r) - \widehat{F}_{i}(r) + \widehat{F}_{j}(r) \right) \\ SD^{c}_{1,j,i,k} &= \sqrt{T} \sup_{r} \left( \widehat{F}_{j,k}(r) - \widehat{F}_{i,k}(r) - \widehat{F}_{j}(r) + \widehat{F}_{i}(r) \right) \\ MF^{c}_{1,k} &= \min_{i \neq j} \left( SD_{1,i,j,k}, SD_{1,j,i,k} \right), \end{aligned} \tag{20}$$

where the superscript $c$ represents the recentered bootstraped test statistics to distinguish these test statistics from tests in (15) which are based on the subsampling scheme.[9]

### 3.2.2 Higher Order

The discussion so far has focussed on first order stochastic dominance testing. To test for higher orders of stochastic dominance, the cumulative distribution functions are replaced by the pertinent integrated cumulative distribution functions. To perform this calculation in practice, the approach adopted is to compute the $m^{th}$ order empirical cumulative distribution function of asset return $R_{i,t}$, by[10]

$$\widehat{F}_{m,i}(r) = \frac{1}{T(m-1)!} \sum_{t=1}^{T} I(R_{i,t} \leq r)(r - R_{i,t})^{m}. \tag{21}$$

Alternatively, the higher order cumulative distribution functions can be computed by cumulative sums of the lower order cumulative distribution functions. The corresponding test statistics of higher order stochastic dominance

[9] As the blocks are overlapping, it is necessary to weight the data before computing the empirical distribution functions $\widehat{F}_i(r)$ and $\widehat{F}_j(r)$ in (20), to reflect the frequency each data point is used in the bootstrap samples. The weighting function is

$$w_t = \begin{cases} t/B & : \quad t < B \\ 1 & : \quad B \leq t \leq T - B + 1 \\ (T - t + 1)/B & : \quad T - B + 2 \leq t \leq T \end{cases}.$$

[10] Expression (21) is motivated by integrating $\int_0^r F_i(t)\,dt$ in (3) by parts and replacing it by its empirical analogue. Repeating the integrations for the higher order integrals yields equation (21).

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.

# 4 Applications

## 4.1 Mehra-Prescott Annual Data

In this section tests of stochastic dominance 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 8 gives the empirical distribution functions and various cumulative empirical distribution functions for the two series.[11] Inspection of the graphs suggests no evidence of any stochastic dominance as the two empirical distribution functions cross for all orders of stochastic dominance. The following tests provide degrees of statistical significance one may attach to inferences.

First, second, third and fourth order stochastic dominance tests based on McFadden's maximality test ($MF_m$) as well as the individual stochastic dominance tests ($SD_{m,i,j}, SD_{m,j,i}$), are reported in Table 4. The first column gives the order of stochastic dominance 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 in (19) 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 producing a bootstrap sample equal to $T$ observations. The total number of replications is set at

[11] The support of the cumulative distribution function is based on the range of the data with the number of intermediate points set equal to $T$.

10000.[12]

The reported value of McFadden's maximality test for first order stochastic dominance in Table 4 is 1.160, with a p-value of 0.030. Following the traditional usage of p-values, we note that only choosing a nominal size of 1% results in a non-rejection of the null hypothesis. This implies that the two assets are first order unrankable at the nominal size of 5% or higher. Inspection of the individual first order stochastic dominance tests reveals that the null in the case of $R_{b,t}$ dominating $R_{s,t}$ is rejected even at the 1% level ($pv = 0.002$), but not the reverse ($pv = 0.222$). This may suggest that there may be a somewhat larger set of utility functionals that favor equities over bonds, than the other way round. A careful inspection of the test distributions makes clear, however, that the probability of negative values for the statistics are zero or close to zero. This means that there are practically no subsamples in which the CDFs do not cross. While a critical value of "zero" may correspond to a conventionally high test size, it would appear to be the appropriate conservative value to choose in this setting. Economists would find it lacking in credibility to conclude dominance when the sample CDFs cross and would choose to maximize test power.

The McFadden maximality test for second order stochastic dominance in Table 4 yields 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 a higher risk from investing in equities.

The results of the third and fourth order stochastic dominance tests also show that neither bond yields nor equity returns dominate each other, with the McFadden maximality test in both cases yielding p-values less than even 1%. This suggests that bonds and equities are unrankable in terms of skew-

[12] Sensitivity of the results to different block sizes are reported in Appendix C.

ness and kurtosis and that agents who have a preference for positive skewness and an aversion for kurtosis, are indifferent between holding the two assets.

Overall the results show that there is no clear stochastic dominance 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 stochastic dominance and hence no puzzle.

## 4.2 Daily Financial Data

Tests of stochastic dominance are now applied to daily data on four financial assets consisting of two risk free assets (3 month Treasury bonds and 6 month Commercial paper yields), and two risky assets (S&P500 and NASDAQ prices).[13] The data begin on 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. See Appendix A for sources and definitions. Some descriptive statistics of the four series are given in Table 5. The sample means show that the equity premium between the two risk free assets and the two equity assets

[13] 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.

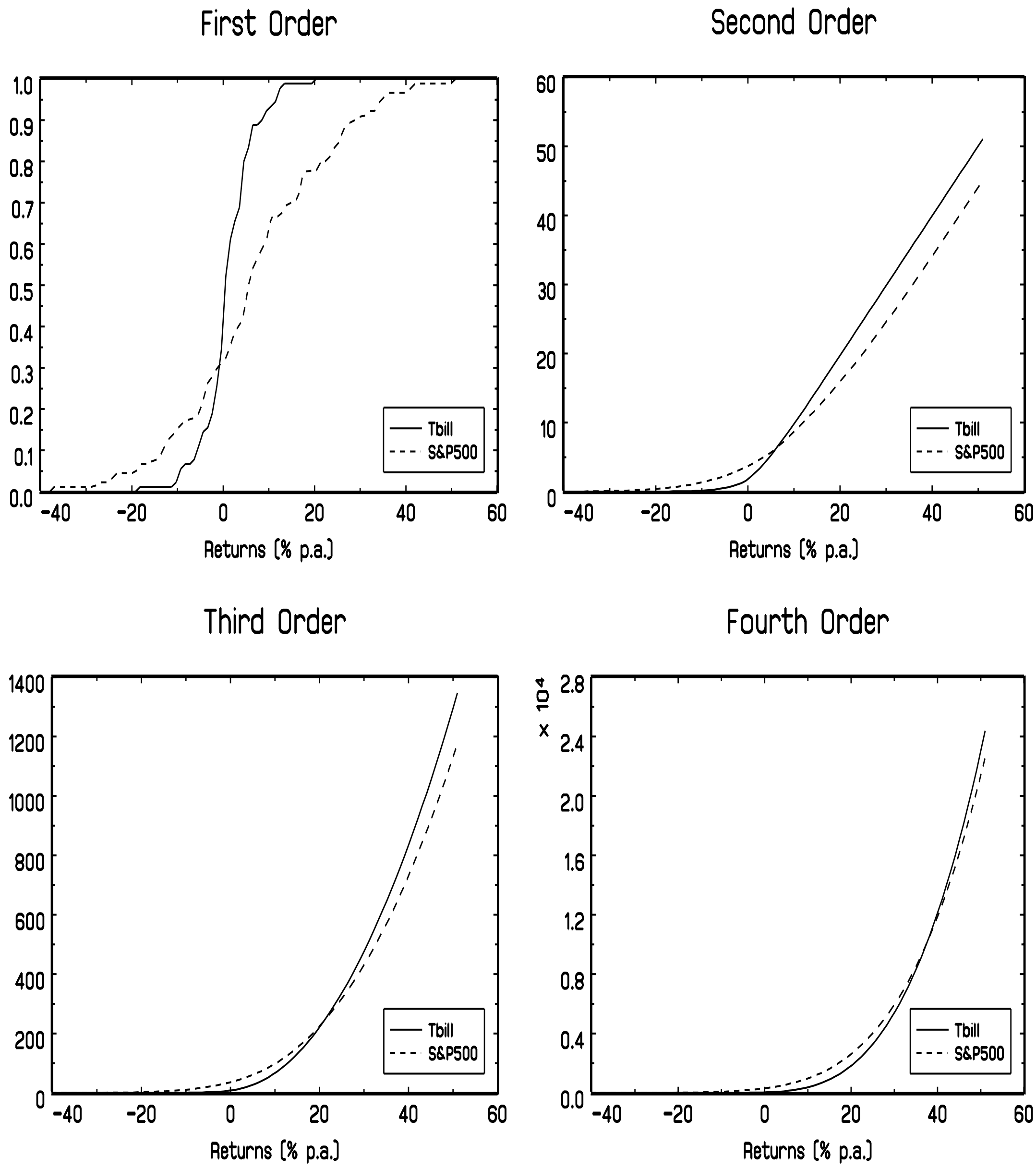


Figure 8: First to fourth order empirical cumulative distribution functions for real bond yields and real equity returns: percentage per annum, 1889 to 1978.

Table 4:

Stochastic dominance 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. The block size is $B = 9$, the sample size of the bootstraps is 90 and the number of replications is 10000.

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

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. However, as the Sharp ratios reveal that the mean return per unit of risk is much higher for the two risk free assets than the two risky assets, this suggests that the equity premia are in fact too low!

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 from 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 all 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 stochastic dominance properties of the assets. This is in contrast to the annual data which showed very little evidence of non-normalities in the data; see Table 1.

Tables 6 to 8 provide stochastic dominance tests for three pairs of assets: $(r_{tb,t}, r_{sp,t})$, $(r_{tb,t}, r_{cp,t})$ and $(r_{sp,t}, r_{nd,t})$. The p-values are based on subsampling with the size of the blocks given by $\alpha = 4$ in (19). This yields blocks of size $B = 240$ resulting in 3421 replications to construct the sampling distributions of the test statistics.[14]

[14] 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 6 shows that there is no first or second order stochastic dominance 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 third order stochastic dominance of Treasury bonds over S&P500 for a nominal size less than 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.

McFadden's maximality test reported in Table 7 for first order stochastic dominance reveals some evidence of dominance amongst the risk free assets, Treasury bonds ($R_{tb,t}$) and Commercial paper ($R_{cp,t}$), as the null is not rejected even at the 10% level. Inspection of the individual stochastic dominance tests shows that neither null is rejected with p-values of 0.156 and 0.136. Closer inspection of the sampling distributions reveals that at least 5% of the tail of the distribution of the test statistic that $R_{cp,t}$ first order stochastically dominates $R_{tb,t}$, is equal to zero, thereby providing weak evidence that dominance is from Commercial paper to Treasury bonds. The evidence is stronger for the second order stochastic dominance tests where Commercial paper stochastically dominates Treasury bonds at the 5% level. This dominance continues for higher orders which is consistent with the properties of stochastic dominance.

The results in Table 8 reveal evidence at the 1% level that S&P500 ($R_{sp,t}$) stochastically dominates NASDAQ ($R_{nd,t}$) at the third order. This last result suggests that agents with a preference for positive skewness prefer S&P500 to NASDAQ. However, as noted already, Table 5 shows that S&P500 exhibits negative skewness whilst NASDAQ exhibits positive skewness. This would suggest that the premium of just over 4% between the two assets would be even larger if the two assets exhibited similar skewness characteristics.

Overall the stochastic dominance tests reveal no strong evidence of dominance at the first order in any of the cases investigated. There is some evidence of second order stochastic dominance of Commercial paper over Treasury bonds. There is also some evidence of third order stochastic dominance of Treasury bills over S&P500, and S&P500 over NASDAQ. This last result reveals the importance of higher order moments, particularly skewness, in determining the risk preferences of agents and the subsequent risk premium observed in the mean.

# 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 approach is nonparametric, being based on testing for stochastic dominance. The tests for various orders of stochastic dominance 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 approach was applied to two data sets. The first was based on the original Mehra-Prescott data which is annual data for the U.S.. The second data consisted of daily observations on two risk-free and two risky assets for the U.S.. The empirical results found little evidence of stochastic dominance in both data sets. There was some evidence of stochastic dominance of equities over bonds in the Mehra and Prescott annual data, but just for utility functions characterized by non-satiation with the nominal size of the test chosen as 1%, but not at 5%. Expanding this class of utility functions to concave functions revealed no evidence of stochastic dominance. The empirical results using daily data revealed no first or second order stochastic dominance between Treasury bills and S&P500. There was some evidence of third order

Table 5:

Descriptive statistics on 3 month Treasury bond yields $(R_{tb,t})$ , 6 month Commercial paper yields $(R_{cp,t})$ , 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.[(a)]

| Statistic | Treas. Bills $(R_{tb,t})$ | Comm. Paper $(R_{cp,t})$ | S&P500 $(R_{sp,t})$ | NASDAQ $(R_{nd,t})$ |
|---|---|---|---|---|
| Mean | 4.666 | 4.963 | 8.446 | 12.636 |
| Median | 5.070 | 5.410 | 1.235 | 20.483 |
| Maximum | 8.390 | 9.050 | 1433.898 | 4335.149 |
| Minimum | 0.790 | 0.900 | -1894.149 | -2615.187 |
| Std. Dev. | 1.762 | 1.854 | 276.316 | 500.497 |
| Skewness | -0.159 | -0.243 | -0.144 | 0.117 |
| Kurtosis | 2.739 | 2.700 | 7.013 | 7.515 |
| BJ (p.v.) | 0.000 | 0.000 | 0.000 | 0.000 |
| Sharp[(b)] | 264.791 | 267.749 | 3.057 | 2.525 |

(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.

(b) Computed as the sample mean divided by the standard deviation and expressed as a percentage by multiplying by 100.

Table 6:

Stochastic dominance 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 Dominance | Null Hypothesis | Statistic | Bottom 5% | Top 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 |
| | $R_{sp,t}$ $SD$ $R_{tb,t}$ | 2553508.478 | 3162.678 | 16869.941 | 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 |
| | $R_{sp,t}$ $SD$ $R_{tb,t}$ | 4111155096.118 | 312977.269 | 1937374.132 | 0.000 |

Table 7:

Stochastic dominance tests of 3 month Treasury yields ($R_{tb,t}$) and 6 month Commercial paper yields ($R_{cp,t}$): July 4th, 1989 and ends July 14th 2003. Bootstraps based on subsampling with $B = 240$ block sizes and 3421 replications.

| Stochastic Dominance | Null Hypothesis | Statistic | Bottom 5% | Top 5% | pv |
|---|---|---|---|---|---|
| First: | Non-maximal | 0.017 | 0.000 | 0.129 | 0.136 |
| | $R_{tb,t}$ $SD$ $R_{cp,t}$ | 13.257 | 2.969 | 15.234 | 0.156 |
| | $R_{cp,t}$ $SD$ $R_{tb,t}$ | 0.017 | 0.000 | 0.129 | 0.136 |
| Second: | Non-maximal | 0.000 | 0.000 | 0.129 | 0.051 |
| | $R_{tb,t}$ $SD$ $R_{cp,t}$ | 6596.672 | 113.220 | 730.445 | 0.000 |
| | $R_{cp,t}$ $SD$ $R_{tb,t}$ | 0.000 | 0.000 | 0.129 | 0.051 |
| Third: | Non-maximal | 0.000 | 0.000 | 0.323 | 0.051 |
| | $R_{tb,t}$ $SD$ $R_{cp,t}$ | 11182432.522 | 13281.364 | 79427.594 | 0.000 |
| | $R_{cp,t}$ $SD$ $R_{tb,t}$ | 0.000 | 0.000 | 0.323 | 0.051 |
| Fourth: | Non-maximal | 0.000 | 0.000 | 0.775 | 0.051 |
| | $R_{tb,t}$ $SD$ $R_{cp,t}$ | 10458963600.875 | 825222.565 | 5877529.146 | 0.000 |
| | $R_{cp,t}$ $SD$ $R_{tb,t}$ | 0.000 | 0.000 | 0.775 | 0.051 |

Table 8:

Stochastic dominance 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 Dominance | Null Hypothesis | Statistic | Bottom 5% | Top 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 |
| | $R_{nd,t}$ $SD$ $R_{sp,t}$ | 1317716.588 | 2310.493 | 11953.189 | 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 |
| | $R_{nd,t}$ $SD$ $R_{sp,t}$ | 2950998064.688 | 228159.063 | 1455097.530 | 0.000 |

stochastic dominance of Treasury bills over S&P500, suggesting that 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. Amongst the risk free assets, Treasury bonds and Commercial paper, there was some evidence that the latter stochastically dominated the former at all orders investigated, especially for 2nd and higher orders. Finally, there was no evidence of either first or second order stochastic dominance 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 stochastic dominance is that it confirms that existing models have indeed misspecified either the utility function, or the returns distribution, or both. It also suggests that there exists a utility function when combined with an appropriate probability distribution that will generate "acceptable" risk aversion parameter estimates. That is, the search could be fruitful! The results also 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 stochastic dominance 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 (2001a,b).

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, different phases of the business cycle etc. 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 stochastic dominance tests. Second, the assumption of expected utility theory can be partially relaxed by performing Prospect Dominance tests following the approach of Linton, Maasoumi and Whang (2003). 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, if required, of the assets. 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 and the home equity bias puzzle (Lewis (1999)).

# A Appendix: Data Definitions and Sources

## A.1 Mehra-Prescott Data

The data consists of annual observations for the period 1889-1978. The variables are:

1. Series $R_{s,t}$ : the annual real returns on equity computed as

$$R_{s,t} = \left( \frac{S_{t+1} + D_t - S_t}{S_t} \right),$$

   where $S_t$ is the real annual average Standard & Poor's Composite Stock Price Index and $D_t$ is the real dividends. The price index is the consumption price deflator.

2. Series $R_{b,t}$ : the annual real return on bonds computed as[15]

$$\begin{aligned} R_{b,t} &= \left( \frac{1 + R_{n,t}}{1 + \Pi_t} - 1 \right) \\ &= \left( (1 + R_{n,t}) \left( \frac{P_t}{P_{t+1}} \right) - 1 \right), \end{aligned}$$

   where $R_{n,t}$ is the nominal yield on relatively riskless short-term securities, and $\Pi_t = (P_{t+1} - P_t) / P_t$ is the inflation rate where $P_t$ is the consumption price deflator.

3. Series $R_{c,t}$ : the annual growth rate of real consumption computed as

$$R_{c,t} = \left( \frac{C_{t+1} - C_t}{C_t} \right),$$

   where $C_t$ is real per capita consumption on durables.

Because the formulae use future values, the effective sample of real returns is 1889 to 1978.

[15] This formula is based on Kocherlakota (1996) which differs from the formula used by Mehra and Prescott (1985) who use a discrete time approximation to compute the real return on bonds. The formula presented in Kocherlakota is for the gross return and not the net return, as is reported here.

## A.2 Daily Data

Based on daily US data for the period July 4th 1989 to July 14th 2003, a total of 3661 observations. the variables are

1. Series $R_{tb,t}$ : US Treasury Bond yield, 3mth, percentage p.a.

2. Series $R_{cp,t}$ : Commercial paper yield 6mth, percentage p.a.

3. Series $P_{sp,t}$ : S&P100 equity index

4. Series $P_{nd,t}$ : NASDAQ100 equity index

The equity returns are computed as

$$\begin{aligned} R_{sp,t} &= 25200(\ln P_{sp,t} - \ln P_{sp,t-1}) \\ R_{nd,t} &= 25200(\ln P_{nd,t} - \ln P_{nd,t-1}), \end{aligned}$$

where the factor 25200, converts daily equity returns into annualized percentages. The total number of effective observations is then $T = 3660$.

## B Appendix: Alternative Models of Risk Aversion

This appendix provides the details for estimating the relative risk aversion parameter $\gamma$, for the various models reported in Table 3. The series are: $R_{s,t}$ (real annual return on equity), $R_{b,t}$ (real annual yield on bonds), $R_{c,t}$ (real annual consumption growth). The models are based on power utility and lognormal returns.

Model 1 (Mehra, 2003, equation 15), is given by

$$\widehat{\gamma}_1 = \frac{\widehat{\mu}_s - \widehat{\mu}_b + 0.5\widehat{\sigma}_s^2}{\widehat{\sigma}_{s,c}}, \tag{22}$$

where $\widehat{\mu}_s$ and $\widehat{\mu}_b$ are the respective sample means of $\ln(1 + R_{s,t})$ and $\ln(1 + R_{b,t})$, $\widehat{\sigma}_s^2$ is the sample variance of $\ln(1 + R_{s,t})$ and $\widehat{\sigma}_{s,c}$ is the covariance of $\ln(1 + R_{s,t})$ and $\ln(1 + R_{c,t})$.

Model 2 (Mehra, 2003, equation 16), is the same as Model 1 with $\widehat{\sigma}_{s,c}$ replaced by $\widehat{\sigma}_c^2$, the sample variance of $\ln(1 + R_{c,t})$

$$\widehat{\gamma}_2 = \frac{\widehat{\mu}_s - \widehat{\mu}_b + 0.5\widehat{\sigma}_s^2}{\widehat{\sigma}_c^2}. \tag{23}$$

Model 3 (Campbell, Lo and MacKinlay, 1997, equation 8.2.9) is based on the regression equation

$$\ln(1 + R_{s,t}) = \alpha + \gamma \ln(1 + R_{c,t}) + u_t, \tag{24}$$

where $u_t$ is a disturbance term and $\alpha$ is an intercept parameter. Estimating this equation by OLS gives

$$\widehat{\gamma}_3 = \frac{\widehat{\sigma}_{s,c}}{\widehat{\sigma}_c^2}. \tag{25}$$

Model 4 (Campbell, Lo and MacKinlay, 1997, equation 8.2.10), is based on the reverse regression equation

$$\ln(1 + R_{c,t}) = \phi + \gamma^{-1} \ln(1 + R_{s,t}) + v_t, \tag{26}$$

where $v_t$ is a disturbance term and $\phi$ is an intercept parameter. Estimating this equation by OLS gives

$$\widehat{\gamma}_4 = \frac{\widehat{\sigma}_s^2}{\widehat{\sigma}_{s,c}}. \tag{27}$$

Model 5 (Campbell, Lo and MacKinlay, 1997, equation 8.2.9) is based on (24) but uses an instrumental variable estimator to correct for dependence between $R_{c,t}$ and $u_t$. The set of instruments used are $\{const, R_{s,t-1}, R_{b,t-1}, R_{c,t-1}\}$.

Model 6 (Campbell, Lo and MacKinlay, 1997, equation 8.2.10) is based on (26) but uses an instrumental variable estimator to correct for dependence between $R_{s,t}$ and $v_t$. The same set of instruments used as in Model 5.

Model 7 (Hansen and Singleton, 1983) is based on estimating $\gamma$ by GMM using the following Euler equations

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

where $\delta$ is the discount factor. The set of instruments used is as in Models 5 and 6; namely, $\{const, R_{c,t-1}, R_{b,t-1}, R_{s,t-1}\}$.

Model 8 (Grossman, Melino and Shiller, 1987) is computed as

$$\widehat{\gamma}_8 = \frac{\left(\widetilde{\mu}_s - \widetilde{\mu}_b\right)\left(1+\widetilde{\mu}_c\right)}{\widetilde{\sigma}_{s,c} - \widetilde{\sigma}_{b,c}},$$

where $\widetilde{\mu}_s, \widetilde{\mu}_b$ and $\widetilde{\mu}_c$ are respectively the sample means of $R_{s,t}, R_{b,t}$ and $R_{c,t}$; $\widetilde{\sigma}_{s,c}$ is the sample covariance of $R_{s,t}$ and $R_{c,t}$, and $\widetilde{\sigma}_{b,c}$ is the sample covariance of $R_{b,t}$ and $R_{c,t}$. Here $\tilde{}$ represents a sample estimate based on the return, which is distinguished from $\hat{}$ which is used to compute sample estimates of log returns.

# C Appendix: Sensitivity Analysis

Additional sensitivity results of the maximality test conducted in Table 4.

| Stochastic Dominance | Maximality Statistics | Block size | Bottom 5% | Top 5% | pv |
|---|---|---|---|---|---|
| | | 6 | 0.105 | 0.949 | 0.015 |
| | | 7 | 0.114 | 1.020 | 0.019 |
| | | 8 | 0.150 | 1.061 | 0.030 |
| First | 1.160 | 9 | 0.105 | 1.054 | 0.030 |
| | | 10 | 0.105 | 1.160 | 0.044 |
| | | 11 | 0.161 | 1.136 | 0.049 |
| | | 12 | 0.170 | 1.164 | 0.050 |
| | | 6 | 0.000 | 6.219 | 0.000 |
| | | 7 | 0.000 | 6.800 | 0.000 |
| | | 8 | 0.000 | 7.491 | 0.000 |
| Second | 18.974 | 9 | 0.000 | 7.695 | 0.000 |
| | | 10 | 0.000 | 8.011 | 0.000 |
| | | 11 | 0.000 | 8.513 | 0.000 |
| | | 12 | 0.000 | 8.947 | 0.001 |
| | | 6 | 0.000 | 89.598 | 0.000 |
| | | 7 | 0.000 | 93.153 | 0.000 |
| | | 8 | 0.000 | 101.810 | 0.000 |
| Third | 316.439 | 9 | 0.000 | 104.355 | 0.000 |
| | | 10 | 0.000 | 111.628 | 0.001 |
| | | 11 | 0.000 | 122.032 | 0.001 |
| | | 12 | 0.000 | 129.551 | 0.001 |
| | | 6 | 0.000 | 1149.172 | 0.000 |
| | | 7 | 0.000 | 1261.683 | 0.000 |
| | | 8 | 0.000 | 1334.019 | 0.000 |
| Fourth | 7345.971 | 9 | 0.000 | 1380.440 | 0.000 |
| | | 10 | 0.000 | 1483.635 | 0.000 |
| | | 11 | 0.000 | 1610.020 | 0.000 |
| | | 12 | 0.000 | 1747.024 | 0.000 |

# 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] 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.

[7] Campbell, J.Y. (1993), "Intertemporal Asset Pricing Without Consumption", *American Economic Review*, 83, 487-512.

[8] 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.

[9] Campbell, J.Y. and Cochrane, J.H. (2000), "Explaining the Poor Performance of Consumption-based Asset Pricing Models", *Journal of Finance*, LV, 2863-2878.

[10] Campbell, J.Y., Lo, A.W., and MacKinlay, A.C. (1997), *The Econometrics of Financial Markets*, Princeton University Press.

[11] 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.

[12] Fama, E.F. and French, F. (2002), "The Equity Premium", Journal of Finance, LVII, 637-659.

[13] Grant, S. and Quiggin, J. (2001a), "The Risk Premium for Equity: Explanations and Implications", mimeo.

[14] Grant, S. and Quiggin, J. (2001b), "Public Investment and the Risk Premium for Equity", 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. Shiryayev (ed.), Kluwer, Dordrecht.

[21] Lewis, K.K. (1999), "Trying to Explain Home Bias in Equities and Consumption", *Journal of Economic Literature*, 37, 571-608.

[22] 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.

[23] Linton, O., Maasoumi, E. and Whang, Y. (2003), "Consistent Testing for Stochastic Dominance Under General Sampling Schemes, Working Paper, London School of Economics.

[24] Lye, J.N. and Martin, V.L. (1993), "Robust Estimation, Non-normalities and Generalized Exponential Distributions", *Journal of the American Statistical Association*, 88, 253-259.

[25] 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.

[26] McGrattan, E. and Prescott, E.C. (2001), "taxes, Regulations, and Asset Prices", Working Paper 610, Federal reserve Bank of Minneapolis.

[27] Mehra R. (2003), "The Equity Premium: Why is it a Puzzle", *Financial Analysts Journal*, 59, 54-69.

[28] Mehra, R. and Prescott, E.C. (1985), "The Equity Premium: A Puzzle", *Journal of Monetary Economics*, 15, 145-161.

[29] Mehra, R. and Prescott, E.C. (2003), "The Equity Premium in Retrospect", NBER Working Paper Series, 9525.

[30] Politis, D.N., Romano, J.P. and Wolf, M. (1999), *Subsampling*, Springer-Verlag, New York.

[31] Rietz, T.A. (1988), "The Equity Risk Premium: A Solution", *Journal of Monetary Economics*, 22, 117-131.

[32] Siegel, J.J. and Thaler, R.H. (1997), "Anomalies: The Equity Premium Puzzle", *Journal of Economic Perspectives*, 11, 191-2000.

[33] Weil, P. (1992), "Equilibrium Asset Prices with Undiversifiable Labor Income Risk", Journal of Economic Dynamics and Control, 16, 769-790.