Available online at www.sciencedirect.com
SCIENCE DIRECT
North American Journal of Economics and Finance
xxx (2006) xxx–xxx
THE NORTH AMERICAN JOURNAL OF ECONOMICS AND FINANCE

A reexamination of the equity-premium puzzle: A robust non-parametric approach

G.C. Lim [a], Esfandiar Maasoumi [b], Vance L. Martin [a,*]

[a] *Economics Department, University of Melbourne, Vic. 3010, Australia*
[b] *Southern Methodist University, Australia*

Received 22 June 2004; received in revised form 23 January 2006; accepted 27 January 2006

Abstract

Recent tests of stochastic dominance of several orders, proposed by Linton, Maasoumi and Whang [Linton, O., Maasoumi, E., & Whang, Y. (2005). Consistent testing for stochastic dominance under general sampling schemes. *Review of Economic Studies*, *72*(3), 735–765], are applied to reexamine the equity-premium puzzle. An advantage of this non-parametric approach is that it provides a framework 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 the original Mehra–Prescott data and more recent data that include daily yields on Treasury bonds and daily returns on the S&P500 and the NASDAQ indexes. The empirical results show little evidence of stochastic dominance among the assets investigated. This suggests that the observed equity premium represents compensation for bearing higher risk, taking into account higher order moments such as skewness and kurtosis. There is some evidence of a reverse puzzle, whereby Treasury bonds stochastically dominate equities at the third order, a result which potentially reflects insufficient compensation that investors receive for having to bear the negative skewness associated with the S&P500 index.

Keywords: Equity-premium puzzle; Stochastic dominance; Non-parametric; Subsampling; Recentered bootstraps; Higher order moments

1. Introduction

If a risky asset or portfolio does not dominate a risk-free alternative, a premium will be demanded for holding it. An appropriate premium would depend on the agent's risk assessment which, in turn, depends on both the agent's utility function and the returns distribution.

* Corresponding author. Tel.: +61 3 834 5396; fax: +61 3 8344 6899.
E-mail address: vance@unimelb.edu.au (V.L. Martin).

1062-9408/$ – see front matter © 2006 Published by Elsevier Inc.
doi:10.1016/j.najef.2006.01.002

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 on 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, based on standard financial models. They label this phenomenon the equity-premium puzzle.[1] 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 many international financial markets (Mehra, 2003).[2]

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 seen 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 & Zin, 1991); habit formation (Constantinides, 1990); relative consumption (Abel, 1990); and subsistence consumption (Campbell & Cochrane, 1999).[3] The second class of models focuses on the specification of the probability distributions underlying the returns processes. The majority of the proposed models assume log-normality. 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 log-normality assumption as it is well documented that empirical returns distributions are highly non-normal and 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 allow for incomplete markets (Weil, 1992), trading costs through borrowing constraints (Heaton & Lucas, 1995), transaction costs (Aiyagari & Gertler, 1991), liquidity premia (Bansal & Coleman, 1996); and taxes (McGrattan & 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 non-parametric 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

[1] An associated puzzle is the risk-free rate puzzle (Weil, 1989), in which the implied risk-free rate predicted by theoretical models is too high relative to the observed rate. While 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.

[2] Campbell (1996) reports evidence of the equity-premium puzzle for both large and medium-sized markets.

[3] A related class of explanations would be those based on behavioral finance. For example Benartzi and Thaler (1995) suggest that the equity premium can be explained by recognizing that investors are more sensitive to losses than gains and that they evaluate their portfolios frequently.

kurtosis.[4] 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 equity-premium puzzle literature. If equities dominate bonds, especially at lower orders, there is indeed a puzzle whatever utility or other functionals within the associated class of 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 knowledge of the probability laws governing these returns. This suggests that evidence of an equity-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, and there are indeed *some* functionals, utility functions and probability distributions such as those adopted by Mehra and Prescott (1985), 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 the functionals that are inconsistent with premia of 6% or more are 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 equity-premium puzzle literature.

The non-parametric framework proposed is applied to two data sets. The first is the original Mehra–Prescott annual data set for the U.S. The second is daily observations on a risk-free bond and two risky-asset indices for the U.S., the S&P500 and NASDAQ indexes. 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 this is at a 1% nominal size of the test and not at the usual 5% level. 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 the higher risk associated with S&P500. Finally, there is no evidence of either first- or second-order stochastic dominance between the two “risky” indices, S&P500 and NASDAQ. However, there is some evidence that S&P500 third- and fourth-order stochastically dominates 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. Preliminary empirical evidence of the equity premium and estimates of the risk-aversion parameter using existing parametric models are reported in Section 2. The non-parametric 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

[4] Harvey and Siddique (2000) provide a recent discussion of the importance of skewness in asset pricing, while Lim, Martin, and Martin (2005) highlight the importance of skewness and kurtosis in the pricing of options.

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 1889–1978 (Mehra–Prescott data)[a]

Statistic	Equity ($R_{s,t}$)	Bonds ($R_{b,t}$)	Consump. ($R_{c,t}$)
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
S.D.	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
Covariances (lower triangle) and correlations (upper triangle)			
Equity ($R_{s,t}$)	270.576	0.113	0.375
Bonds ($R_{b,t}$)	10.577	32.468	−0.107
Consump. ($R_{c,t}$)	22.011	−2.166	12.722

[a] Equity returns and consumption growth are computed as arithmetic returns. See Mehra and Prescott (1985) and Kocherlakota (1996) for details of constructing the data.

data set, as well as a more recent data set that uses daily equity returns and bond yields. The main empirical results point to a lack of stochastic dominance among the financial returns series investigated. Section 5 provides some concluding comments and suggestions for future research.

2. Background 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. These data consist of annual U.S. data on real asset prices and aggregate real consumption expenditures beginning in 1889 and ending in 1979, a total of 91 observations.[5]

Descriptive statistics on real equity returns ($R_{s,t}$), real bond yields ($R_{b,t}$) and the real consumption growth rate ($R_{c,t}$), are given in Table 1. Equity returns and consumption growth are computed as arithmetic returns, thereby reducing the effective sample size to $T = 90$. All variables are expressed in percentages per annum. The size of the equity premium between equities and bonds is approximately 6% p.a. (6.980–1.036%). 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, that is, 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 Fig. 1 which plots the two series over the sample period, 1889–1978.

The strength of the contemporaneous linear relationships among the three series is highlighted by the covariances (lower triangle) and correlations (upper triangle) in Table 1. Consumption and

[5] Understanding the time-series properties of the data is also important in designing appropriate bootstrap procedures to undertake stochastic dominance tests. This connection is elaborated upon in Sections 3 and 4.

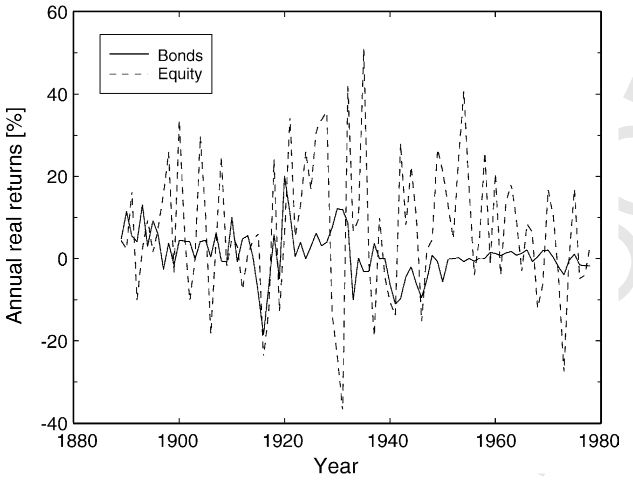

Fig. 1. Bond yields and equity returns: real, percentage per annum, 1889–1978.

equities have a positive association (correlation of 0.375), as do equities and bonds (correlation of 0.113), while consumption and bonds have a negative association (correlation of −0.107).

Estimates of the relative risk-aversion parameter γ are presented in Table 2 for the Mehra–Prescott data using the descriptive statistics in Table 1. Details of the calculations are given in the footnote of this table. All of these estimates are based on parametric representations

Table 2
Alternative estimates of the relative risk-aversion parameter, γ: 1889–1978, Mehra–Prescott data

Model	Method and source	γ
1	Mehra (2003, Eq. (15))[a]	26.085
2	Mehra (2003, Eq. (16))[b]	46.926
3	Campbell et al. (1997, Eq. (8.2.9))[c]	1.799
4	Campbell et al. (1997, Eq. (8.2.10))[d]	11.062
5	Campbell et al. (1997, Eq. (8.2.9))[e]	1.823
6	Campbell et al. (1997, Eq. (8.2.10))[f]	3.351
7	Hansen and Singleton (1983): GMM[g]	15.397
8	Grossman, Melino, and Shiller (1987)[h]	24.755

The following definitions are used. Let $r_{s,t} = \ln(1 + R_{s,t})$, $r_{b,t} = \ln(1 + R_{b,t})$ and $r_{c,t} = \ln(1 + R_{c,t})$, represent log returns: $\hat{\mu}_s$ and $\hat{\mu}_b$ are the respective sample means of $r_{s,t}$ and $r_{b,t}$, $\hat{\sigma}_s^2$ is the sample variance of $r_{s,t}$ and $\hat{\sigma}_{s,c}$ is the sample covariance of $r_{s,t}$ and $r_{c,t}$. For arithmetic 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 $\hat{\gamma}_2 = (\hat{\mu}_s - \hat{\mu}_b + 0.5\hat{\sigma}_s^2)\hat{\sigma}_c^{-2}$.

[c] Computed as $\hat{\gamma}_3 = \hat{\sigma}_{s,c}\hat{\sigma}_c^{-2}$, by regressing $r_{s,t}$ on a constant and $r_{c,t}$.

[d] Computed as $\hat{\gamma}_4 = \hat{\sigma}_s^2\hat{\sigma}_{s,c}^{-1}$, by regressing $r_{c,t}$ on a constant and $r_{s,t}$.

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

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

[g] The GMM estimate is based on the two moment conditions $E[\delta(1 + R_{c,t})^{-\gamma}(1 + R_{b,t}) - 1]$, $E[\delta(1 + R_{c,t})^{-\gamma}(1 + R_{s,t}) - 1]$, with instruments as in (e) or (f).

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

using power utility preferences and log-normal returns (see Campbell, Lo, & MacKinlay (1997)). The first observation to make is that the estimates of the relative risk-aversion parameter are not robust, with estimates ranging from a high of 46.926 to a low of 1.799, despite the same underlying model. Psychologists and experimentalists have found similarly disconcertingly wide ranges for this parameter. This variation in the estimates of γ suggest that either the preference function, or the distribution of returns, or both, are inappropriate. These results also highlight the need for adopting a non-parametric approach in modeling the equity premium to avoid basing inferences on incorrect parametric specifications.

3. Stochastic-dominance testing

This section provides a non-parametric approach based on stochastic-dominance testing to re-evaluate the equity-premium puzzle. This has the advantage of testing if the observed equity premium represents adequate compensation for risk preferences based on second and even higher moments of the underlying returns distribution, while imposing a minimalist set of conditions on preferences. This contrasts with the existing literature which tends to focus on tight parametric representations of the utility and distribution functionals. A lack of stochastic dominance between asset returns is evidence that the premium is adequate compensation for bearing risk, whereas evidence of stochastic dominance suggests a puzzle as equity returns are too high, or even potentially too low, to be consistent with the risk preferences of investors.

3.1. Definition of stochastic dominance

Consider two stationary time series of returns, $R_{i,t}$ and $R_{j,t}$, $t=1, 2, \ldots, T$, with respective cumulative distribution functions (CDFs), $F_i(r)$ 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. The null hypotheses that $R_{i,t}$ stochastically dominates $R_{j,t}$, for various orders are as follows:

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

The null hypotheses in this paper are unambiguous as the test for stochastic dominance combines the test that $R_{i,t}$ stochastically dominates $R_{j,t}$ with the 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 ith asset are in excess of r, say, is higher than the corresponding probability associated with the jth asset

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

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 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

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

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 log-normality, the relationship between the returns on equity ($R_{s,t}$) and bond yields ($R_{b,t}$) is (Campbell et al. (1997))

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

where γ 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 $\mu' \geq 0$, $\mu'' \leq 0$. This class of agents is characterized by risk aversion, whereby a risk premium is needed to compensate investors from holding assets whose 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, $\mu' \geq 0$, $\mu'' \leq 0$, $\mu''' \geq 0$. 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 Ingersoll (1987) and 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 $\mu' \geq 0$, $\mu'' \leq 0$, $\mu''' \geq 0$, $\mu'''' \leq 0$. 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.

Fig. 2 highlights the stochastic-dominance features of four hypothetical asset return distributions. All distributions are assumed to be normal, $N(\mu,\sigma^2)$ with mean μ and volatility σ^2,

$$F_1 = N(1, 6^2), \quad F_2 = N(7, 6^2), \quad F_3 = N(1, 12^2), \quad F_4 = N(6, 12^2).$$

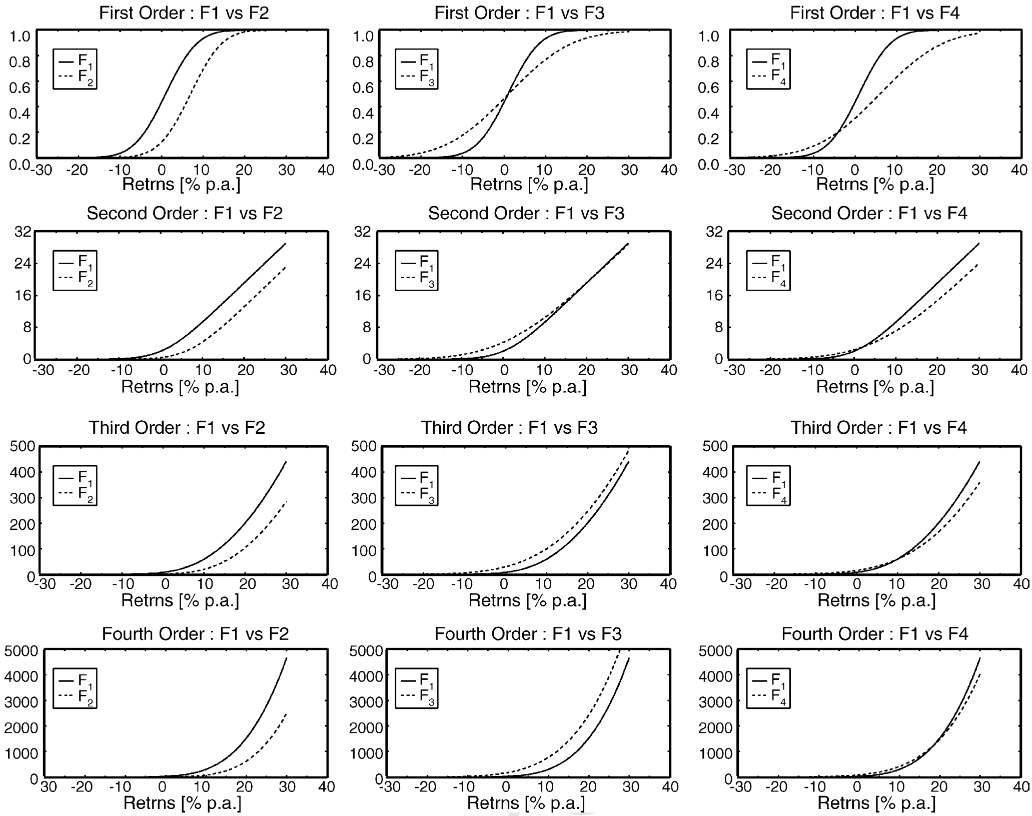

Fig. 2. Hypothetical asset returns distributions, first- to fourth-order stochastic dominance as defined in (1): $F_1 = N(1, 6^2)$, $F_2 = N(7, 6^2)$, $F_3 = N(1, 12^2)$, $F_4 = N(6, 12^2)$.

The first column of Fig. 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 Fig. 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 ($\sigma_1 < \sigma_3$), while 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 Fig. 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 Fig. 2 with the distribution of asset 4 in the third column). There is no SD of any order between the two assets in this case. The higher expected return of F_4 relative to F_1 is indeed appropriate compensation for bearing the higher risk. The equity premium of $\mu_2 - \mu_1 = 5$, now does not represent a puzzle.

3.2. Testing

The approach for conducting stochastic-dominance tests is based on the approach by Linton, Maasoumi, and Whang (2005), who propose non-parametric tests of stochastic dominance by extending the Kolmogorov–Smirnov statistics of McFadden (1989). Inference is performed by using subsampling to construct p-values as well as recentered bootstrap methods. An important advantage of this approach is that it can accommodate the general dependence structures observed in returns that arise from conditional volatility (Bollerslev, Chou, & Kroner (1992)) and higher order moments (Harvey and Siddique, 2000), as well as the observed contemporaneous correlations among assets.[6]

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 $R_{j,t}$

$$\mathrm{SD}_{1,i,j} = \sqrt{T}\sup_{r}(\hat{F}_i(r) - \hat{F}_j(r)), \tag{5}$$

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

$$SD_{1,j,i} = \sqrt{T}\sup_{r}(\hat{F}_j(r) - \hat{F}_i(r)). \tag{6}$$

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

$$\hat{F}_k(r) = \frac{1}{T}\sum_{t=1}^{T} I(R_{k,t} \le r), \tag{7}$$

where

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

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

$$\mathrm{MF}_1 = \min_{i \ne j}(\mathrm{SD}_{1,i,j}, \mathrm{SD}_{1,j,i}). \tag{9}$$

Suppose that the null is true, so that the distribution function of $R_{i,t}$ lies to the right of the distribution function of $R_{j,t}$, except for the tails where it is zero, as in the first column of Fig. 2. Now $F_i(r) < F_j(r)$, yielding a negative value for the support of the distribution under the null, while at the tails the difference is zero. Taking the sup in (5) results in a value of the test statistic of

[6] A related approach is by Barrett and Donald (2003). However, this approach assumes *iid* returns as well as returns being contemporaneously uncorrelated. See Abhyankar and Ho (2003) for a comparison of the Linton et al. (2005) and Barrett and Donald (2003) approaches in the case of financial data.

$SD_{1,i,j} = 0$. 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 stochastic 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.[7]

In the case of *iid* data, the sampling distributions of (5) and (8) under the null were originally derived by Kolmogorov (1933), while 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.[8] To circumvent this problem, the sampling distribution of the test statistics is approximated using a resampling scheme based on subsampling and bootstrap methods. (See Politis, Romano, & Wolf (1999), and Linton et al. (2005) for a review of this approach.) The approach is to sample pairs of overlapping sub-periods of the data. By sampling the data in blocks, this captures the dependence structure in the data, while sampling the data in paired blocks preserves its contemporaneous structure. The sampling distribution is constructed by computing the test statistics for each sampled block and constructing the *p*-values from the empirical distributions. In the case where unique blocks are sampled, the approach is called sub-sampling, whereas the approach is called bootstrapping where non-unique blocks are sampled and stacked to reconstruct a sample of size T.

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 CDF of asset return $R_{i,t}$, by[9]

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

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. 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 can result in apparent contradictions with a small probability.

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, $T = 90$, are presented for the Mehra and Prescott data. Fig. 3 gives the

[7] The maximality test statistic in (9) can be extended to testing for maximality among more than two assets.

[8] This problem is akin to performing a test of the population mean, where the test statistic is a function of the unknown population variance.

[9] Expression (10) is motivated by integrating $\int_0^r F_i(t)\mathrm{d}t$ in (1) by parts and replacing it by its empirical analogue. Repeating the integrations for the higher order integrals yields Eq. (10).

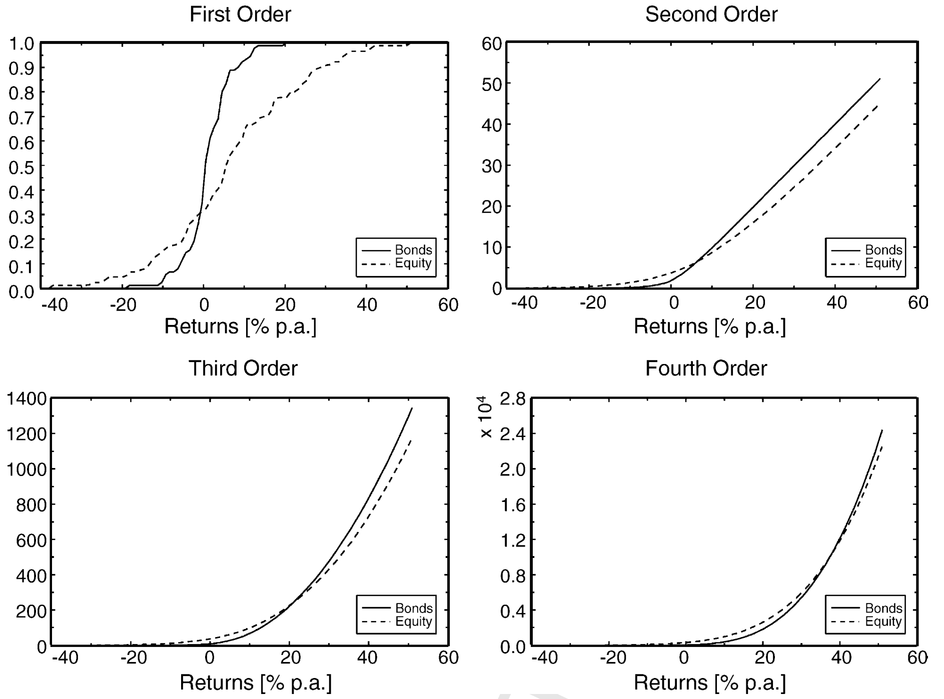

Fig. 3. First- to fourth-order empirical cumulative distribution functions for real bond yields and real equity returns: percentage per annum, 1889–1978.

empirical distribution functions and higher order cumulative empirical distribution functions for the two series.[10] Inspection of the graphs suggests no evidence of any SD, as the two empirical distribution functions cross for all orders of SD.

The SD tests based on MF_m, $m = 1, 2, 3, 4$ as well as the individual SD tests ($SD_{m,i,j}$, $SD_{m,j,i}$), are reported in Table 3. The first column gives the order of SD being tested, with the null hypothesis given in the second column. The test statistics are given in the third column, with the calculated values reported in the fourth column. The next three columns provide information on the sampling distribution of the test statistics with the p-values reported in the last column. The sampling distribution is based on "recentered paired bootstraps" with overlapping blocks. The block sizes are set at $B = 9$, using the rule $B = \alpha\left[\sqrt{T}\right]$ with $\alpha = 1$. This represents a string of 10 years of data in each block. For a sample of size $T = 90$, there are 82 overlapping blocks. For each bootstrap, nine blocks are randomly drawn and stacked producing a bootstrap sample equal to T observations. The number of replications is set at 10,000.[11]

The reported value of the test of first-order SD using MF_1 in Table 3 is 1.160. Comparing this value with the critical value associated with the top 5% of values, 1.054, provides evidence of no first-order SD between Treasury bonds and equities.

[10] 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.

[11] Sensitivity analyses with the block sizes varying from 6 to 12 yield similar p-values as reported in Table 3. These results were presented in an earlier version of this paper and are available from the authors upon request.

Table 3
SD tests of real bond yields $R_{b,t}$ and equity returns $R_{s,t}$: Mehra–Prescott data, 1889–1978

Stochastic dominance	Null hypothesis	Statistic	Value	Bottom 5%	Top 5%	pv
First	Non-maximal	MF_1	1.16	0.105	1.054	0.03
	$R_{b,t}$ SD $R_{s,t}$	$SD_{1,b,s}$	3.479	0.316	2.214	0.002
	$R_{s,t}$ SD $R_{b,t}$	$SD_{1,s,b}$	1.16	0.211	1.687	0.222
Second	Non-maximal	MF_2	18.974	0	7.695	0
	$R_{b,t}$ SD $R_{s,t}$	$SD_{2,b,s}$	56.71	0	35.101	0.002
	$R_{s,t}$ SD $R_{b,t}$	$SD_{2,s,b}$	18.974	0	24.244	0.103
Third	Non-maximal	MF_3	316.44	0	104.36	0
	$R_{b,t}$ SD $R_{s,t}$	$SD_{3,b,s}$	1600.6	0	1531.3	0.042
	$R_{s,t}$ SD $R_{b,t}$	$SD_{3,s,b}$	316.44	0	1134.5	0.3
Fourth	Non-maximal	MF_4	7346	0	1380.4	0
	$R_{b,t}$ SD $R_{s,t}$	$SD_{4,b,s}$	16774	0	39941	0.265
	$R_{s,t}$ SD $R_{b,t}$	$SD_{4,s,b}$	7346	0	37646	0.357

Bootstraps based on recentered paired bootstraps with overlapping blocks.

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. (2005) 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. This situation arises in the test of second- and higher order SD in Table 3.

The test value of MF_2 for testing second-order SD in Table 3 has a value of 18.974, 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. However, the critical value of the bottom 5% of values is zero, showing that there is a 0.05 probability of negative values for the statistic, and that a 95% confidence interval for second-order SD includes zero. Thus, “equal ranking” is not rejected at this level of confidence.

The results of the third- and fourth-order SD tests using MF_3 and MF_4 also show that neither security dominates the other, with the SD test values in both cases being positive and yielding p-values of 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 a 0.05 probability of negative values for the statistics, suggesting that a 95% confidence interval for SD includes zero. Thus, “equal ranking” of assets is not rejected at this level of confidence, and higher order moments matter, albeit only slightly.

Table 4
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 4 July 1989 and ending 14 July 2003[a]

Statistic	Treas. bills ($R_{tb,t}$)	S&P500 ($R_{sp,t}$)	NASDAQ ($R_{nd,t}$)
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
S.D.	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.

Overall, the results show that there is no clear SD between bond yields and equity returns for the Mehra–Prescott data. This is also true for risk preferences characterized by second- and higher 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.

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).[12] The data begin after 4 July 1989, and end on 14 July 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 4. The sample means show that the equity premia between the risk-free asset and the two equity assets are between 4 and 8, which encompasses the premium estimate 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 4 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

[12] 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 result is similar to the approach of Campbell (1993), who evaluates the CCAPM having substituted out consumption. Also note that as price data on goods markets are not available daily, the asset returns used in this example are expressed in nominal terms in contrast to the asset returns defined in the previous example, which are expressed in real terms.

Table 5
SD tests of Treasury yields ($R_{tb,t}$) and S&P500 equity returns ($R_{sp,t}$): 4 July 1989 and ends 14 July 2003

Stochastic dominance	Null hypothesis	Statistic	Value	Bottom 5%	Top 5%	pv
First	Non-maximal	MF_1	29.373	6.520	7.552	0.000
	$R_{b,t}$ SD $R_{s,t}$	$SD_{1,b,s}$	29.373	6.713	8.391	0.000
	$R_{s,t}$ SD $R_{b,t}$	$SD_{1,s,b}$	30.117	6.520	8.456	0.000
Second	Non-maximal	MF_2	249.298	0.000	70.166	0.000
	$R_{b,t}$ SD $R_{s,t}$	$SD_{2,b,s}$	249.298	0.000	70.166	0.000
	$R_{s,t}$ SD $R_{b,t}$	$SD_{2,s,b}$	6267.950	116.448	260.006	0.000
Third	Non-maximal	MF_3	0.000	0.000	0.000	0.050
	$R_{b,t}$ SD $R_{s,t}$	$SD_{3,b,s}$	0.000	0.000	0.000	0.050
	$R_{s,t}$ SD $R_{b,t}$	$SD_{3,s,b}$	2.553×10^6	3162.678	1.686×10^4	0.000
Fourth	Non-maximal	MF_4	0.000	0.000	0.000	0.000
	$R_{b,t}$ SD $R_{s,t}$	$SD_{4,b,s}$	0.000	0.000	0.000	0.000
	$R_{s,t}$ SD $R_{b,t}$	$SD_{4,s,b}$	4.111×10^9	3.129×10^5	1.937×10^6	0.000

Bootstraps based on subsampling with $B = 240$ block sizes and 3421 replications.

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 non-normalities, as revealed by the Bera–Jarque normality test. This feature of the data 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 no strong evidence of non-normalities.

Tables 5 and 6, 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 NASDAQ ($r_{sp,t}$, $r_{nd,t}$). The p-values are based on subsampling, with the size of the blocks given

Table 6
SD tests of S&P500 equity returns ($R_{sp,t}$) and NASDAQ equity returns ($R_{nd,t}$): 4 July 1989 and ends 14 July 2003

Stochastic dominance	Null hypothesis	Statistic	Value	Bottom 5%	Top 5%	pv
First	Non-maximal	MF_1	6.496	0.968	3.098	0.000
	$R_{b,t}$ SD $R_{s,t}$	$SD_{1,b,s}$	7.124	1.226	3.357	0.000
	$R_{s,t}$ SD $R_{b,t}$	$SD_{1,s,b}$	6.496	0.968	3.938	0.000
Second	Non-maximal	MF_2	133.433	0.000	43.442	0.000
	$R_{b,t}$ SD $R_{s,t}$	$SD_{2,b,s}$	133.433	0.000	45.185	0.000
	$R_{s,t}$ SD $R_{b,t}$	$SD_{2,s,b}$	2425.769	38.407	136.781	0.000
Third	Non-maximal	MF_3	0.000	0.000	0.000	0.046
	$R_{b,t}$ SD $R_{s,t}$	$SD_{3,b,s}$	0.000	0.000	0.000	0.048
	$R_{s,t}$ SD $R_{b,t}$	$SD_{3,s,b}$	1.317×10^6	2310.493	1.195×10^4	0.000
Fourth	Non-maximal	MF_4	0.000	0.000	0.000	0.011
	$R_{b,t}$ SD $R_{s,t}$	$SD_{4,b,s}$	0.000	0.000	0.000	0.022
	$R_{s,t}$ SD $R_{b,t}$	$SD_{4,s,b}$	2.950×10^9	2.281×10^5	1.455×10^6	0.000

Bootstraps based on subsampling with $B = 240$ block sizes and 3421 replications.

by $B = \alpha\left[\sqrt{T}\right]$ with $\alpha = 4$. This yields blocks of size $B = 240$, resulting in 3421 replications to construct the sampling distributions of the test statistics.[13]

The reported value of 29.373 for MF_1 in Table 5 and its p-value of 0.000 show that there is no evidence of first-order SD between Treasury bonds ($R_{tb,t}$) and S&P500 ($R_{sp,t}$). The reported value for MF_2 has a p-value of 0.000, showing that there is also no evidence of second-order SD between the two assets, although the critical value of the bottom 5% is zero. These results imply that there is no puzzle, as the observed premium between the two assets of just under 4% reported in Table 4 represents an appropriate amount of compensation for agents bearing higher risk who have concave utility functions.

Interestingly, there is some evidence of third- and higher order SD of Treasury bonds over S&P500 for a nominal size marginally below 5%. This would suggest that there is a puzzle, but in reverse! This dominance possibly reflects the negative skewness in S&P500 (Table 4), whereby agents are not receiving sufficient compensation for bearing negative skewness when they prefer positive skewness.

The main result of the SD tests between S&P500 ($R_{sp,t}$) and NASDAQ ($R_{nd,t}$), presented in Table 6, is that there is evidence at the 1% level that S&P500 dominates NASDAQ at the third order. There are 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 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. Both of these results reveal 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 partly explains the greater success of studies (e.g., Epstein and Zin, 1991) which have chosen functionals that allow a role for higher order moments than the mean and the variance.

5. Conclusions

This paper has provided a non-parametric approach based on stochastic-dominance testing to reexamine the equity-premium debate without the need to specify the underlying utility and probability functionals. 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 empirical results found little evidence of SD in the data sets investigated. 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% levels. 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 was interpreted to imply that the observed equity premium might in fact be too small to compensate agents adequately for bearing

[13] 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.

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 inappropriate 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 research has focused 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. 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 et al. (2005). 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, the framework presented here can also be applied to testing the validity of other puzzles such as the risk-free puzzle.

Acknowledgements

We would like to thank two anonymous referees for insightful comments and suggestions on an earlier version of the paper.

References

Abel, A. B. (1990). Asset prices under habit formation and catching up with the Joneses. *American Economic Review*, *80*, 38–42.

Abhyankar, A., & Ho, K. (2003). Exploring long-run abnormal performance using stochastic dominance criteria: additional evidence from IPOs. Mimeo.

Aiyagari, S. R., & Gertler, M. (1991). Asset returns with transaction costs and uninsured individual risk. *Journal of Monetary Economics*, *27*, 311–331.

Bansal, R., & 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.

Barrett, G., & Donald, S. (2003). Consistent tests for stochastic dominance. *Econometrica*, *71*(1), 71–104.

Benartzi, S., & Thaler, R. H. (1995). Myopic loss aversion and the equity premium puzzle. *Quarterly Journal of Economics*, *110*, 73–92.

Bollerslev, T., Chou, R. Y., & Kroner, K. F. (1992). ARCH modeling in finance: a review of the theory and empirical evidence. *Journal of Econometrics*, *52*, 5–59.

Campbell, J. Y. (1993). Intertemporal asset pricing without consumption. *American Economic Review*, *83*, 487–512.

Campbell, J.Y. (1996). Consumption and the stock market. NBER Working Paper, No. 5619.

Campbell, J. Y., & Cochrane, J. H. (1999). By force of habit: a consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy*, *107*, 205–251.
Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). *The econometrics of financial markets*. Princeton: Princeton University Press.
Epstein, L. G., & Zin, S. E. (1991). Substitution, risk aversion and the temporal behavior of consumption and asset returns: an empirical analysis. *Journal of Political Economy*, *99*, 263–286.
Grant, S., & Quiggin, J. (2001). The risk premium for equity: explanations and implications. Mimeo.
Grossman, S., Melino, A., & Shiller, R. (1987). Estimating the continuous time consumption based asset pricing model. *Journal of Business and Economic Statistics*, *5*, 315–328.
Hansen, L., & Singleton, K. (1983). Stochastic consumption, risk aversion and the temporal behavior of asset returns. *Journal of Political Economy*, *91*, 249–268.
Harvey, C. R., & Siddique, A. (2000). Conditional skewness in asset pricing tests. *Journal of Finance, LV*, 1263–1295.
Heaton, J., & Lucas, D. (1995). The importance of investor heterogeneity and financial market imperfections for the behavior of asset prices. *Carnegie-Rochester Series in Public Economics*.
Ingersoll, J. E., Jr. (1987). *Theory and financial decision making*. Totowa, New Jersey, USA: Rowman and Littlefield Studies in Financial Economics.
Kocherlakota, N. R. (1996). The equity premium: it's still a puzzle. *Journal of Economic Literature*, *34*, 42–71.
Kolmogorov, A. N. (1933). On the empirical determination of a distribution law. Reprinted in: In A. N. Shiryayev (Ed.), "*Selected works of A. N. Kolmogorov ", Probability theory and mathematical statistics* (Vol. II). Dordrecht: Kluwer.
Lim, G. C., Martin, G. M., & Martin, V. L. (2005). Parametric pricing of higher order moments in S&P500 options. *Journal of Applied Econometrics*, *20*, 377–404.
Linton, O., Maasoumi, E., & Whang, Y. (2005). Consistent testing for stochastic dominance under general sampling schemes. *Review of Economic Studies*, *72*(3), 735–765.
McFadden, D. (1989). Testing for stochastic dominance. In T. Fomby, & T. K. Seo (Eds.), *Studies in the economics of uncertainty* (Part II), (in honor of J. Hadar) New York: Springer-Verlag.
McGrattan, E., & Prescott, E.C. (2001). Taxes, regulations, and asset prices. Working Paper 610, Federal Reserve Bank of Minneapolis.
Mehra, R. (2003). The equity premium: Why is it a puzzle. *Financial Analysts Journal*, *59*, 54–69.
Mehra, R., & Prescott, E. C. (1985). The equity premium: a puzzle. *Journal of Monetary Economics*, *15*, 145–161.
Politis, D. N., Romano, J. P., & Wolf, M. (1999). *Subsampling*. New York: Springer-Verlag.
Rietz, T. A. (1988). The equity risk premium: a solution. *Journal of Monetary Economics*, *22*, 117–131.
Weil, P. (1989). The equity premium puzzle and the risk-free rate puzzle. NBER Working Paper 2829.
Weil, P. (1992). Equilibrium asset prices with undiversifiable labor income risk. *Journal of Economic Dynamics and Control*, *16*, 769–790.