Explaining the declining ex-ante equity risk premium

Mark Freeman
School of Business and Economics
University of Exeter
Exeter EX4 4PU
England
Tel: +44-(0)1392-263155
Fax: +44-(0)1392-263242
email: Mark.C.Freeman@exeter.ac.uk

This paper describes the underlying cause of the widely reported sharp decline in the ex-ante equity risk premium during the second half of the Twentieth Century. By examining average income data for the rich, it is demonstrated that US stockholders experienced a significant reduction in economic risk around the time of World War II that was recognized by investors and incorporated into stock prices by the early 1960s. This was the fundamental factor that drove the bull market of that decade. It is argued that the ex-ante risk premium has been low since the mid 1960s. Despite the reduced cost of equity capital over the last forty years, the model continues to predict high ex-post stock returns into the 1990s, which is broadly consistent with observed market behavior. There is only very limited evidence of excess volatility in stock returns in either half of the 1900s.

Keywords : Equity Premium Puzzle, Market Segmentation
JEL Classification: G12, E24

1 Introduction

In a highly cited paper, Mehra & Prescott (1985) demonstrated that average real per-capita consumption growth is neither sufficiently volatile nor highly correlated with stock market returns to be able to explain, within a standard asset pricing framework, the high observed return to equities in the US during the Twentieth Century. This anomaly is known as the "equity premium puzzle". Many authors have attempted to resolve this problem but there is no widely accepted resolution — see, for example, Mehra (2003) for a recent literature review. One potential explanation, proposed initially by Blanchard (1993) and subsequently examined by Jagannathan, McGratten & Scherbina (2000), Claus & Thomas (2001), Bansal & Lundblad (2002) and Fama & French (2002), amongst others, is that the second half of the last century saw a significant reduction in the required rate of return for stocks that partially caused the high observed returns. While this is an appealing idea, there has been, to date, no formal, consumption-based, economic justification as to why the risk premium should have declined in this way. This paper fils this gap by providing evidence of a structural break in underlying risk to stockholders around the time of World War II. It is also shown that investors would have learned quickly of this change in regime. This, then, provides a clear underlying economic motivation to explain why the ex-ante risk premium declined significantly during the 1950s and early 1960s and why the stock market produced such strong performance in that decade.

This study is based on the idea of Mankiw & Zeldes (1991) that market segmentation lies at the heart of Mehra and Prescott's anomaly. Mankiw and Zeldes showed that the consumption of stockholders is both more volatile and more highly correlated with stock returns than the consumption of the economy-wide representative agent. These findings have been supported by recent studies by Brav, Constantinides & Geczy (2002) and Vissing-Jørgensen (2002). The limitation of existing result in this area is that they are limited to relatively short time horizons (post 1970) due to the paucity of high quality panel consumption data. With such a comparatively short time-frame, it is obviously not possible to examine a potential reduction in the equity premium in the second half of the Twentieth Century. To overcome this problem, this paper reverts to *income* data for the average high-income household going back to 1913. Such information has recently become available through Piketty & Saez (2003) and the associated working paper Piketty & Saez (2001). The key methodological assumption / innovation in this approach is that it provides a sufficiently lengthy time-series to consider changes in risk during the last hundred years.

It is asserted that the use of income rather than consumption data does not significantly bias the results of this study. There are several reasons for making this claim. First, all the tests that are conducted on income data of various earnings groups in this study are also run on economy-wide aggregate consumption data. The structural shift in risk around the time of WWII that is so clearly evident in the income data is just as clear from the consumption data[1]. The ratio of standard deviations before and after the structural shift is also similar between the series. The learning process for the shift in regime is highly similar whether income or consumption data is used as a basis for estimation. That is, the regime shift that drives this model is apparent even if this study is limited to aggregate consumption data. Second, it is possible in recent times to compare the consumption risk of stockholders with the volatility of income growth for high-earners. Mankiw & Zeldes (1991) find that, between 1970 and 1984, the standard deviation of food consumption to the stockholding representative agent was around 3.2%. The income of the top one percentile of earners ("P99-100") has a very similar standard deviation of 3.4% for these years. For the period 1982 - 1996, Vissing-Jørgensen (2002) finds a standard deviation of around 7.0% for average per-

[1] No attempt is made to try to explain why stockholder risk became so much lower at this time — this is left to economic historians such as Vatter (1963). Very briefly, though, a key short-term impetus which helped the US economy recover from the depression of 1949 was high government expenditure and low unemployment caused by the Korean War.

capita stockholder consumption growth[2]. The estimate based on income data is slightly higher at 9.7%, although this discrepancy disappears when adjustments are made for changes in mean growth. Therefore, it seems, once again, as if the average income risk of high-earning households is very similar to the consumption risk of stockholders. This is also broadly consistent with a consumption-smoothing argument within a segmented market. It is known that individual wealthy families receive income that is substantially more volatile than the aggregate data provided by Piketty & Saez (2003) (see Freeman (2003) for an analysis of individual risk in this context). Through stock and bond markets, investors cross-insure so that their idiosyncratic income risk is fully eliminated (see, in particular, Heaton & Lucas (1996)) and a representative agent emerges. In this context, because of market segmentation, this representative agent represents the most wealthy in the economy only. Under a budget constraint within a Lucas-style economy, the consumption and income of this rich representative agent must be equal in each period since there are no mechanisms by which to further consumption smooth. Therefore, the income risk of the rich representative agent can be

[2] She quotes semi-annual standard deviations, which have been multiplied by the square root of 2. This figure is for the middle and top stockholder layers based on the consumption of non-durables and services from the Consumer Expenditure Survey. For the bottom layer stockholders the risk is lower but similar at 6.2%. The consumption risk found by Mankiw & Zeldes (1991) using PSID data is also largely insensitive to the definition of "stockholder"

taken as a proxy for the consumption risk of individual stockholders[3].

If aggregate consumption risk and income risk to the wealthy both point to a structural change in risk in around the time of the Second World War, what is the advantage of basing a model on the latter series rather than the former? Wealthy income growth is more volatile than aggregate consumption growth, again capturing Mankiw and Zeldes' finding that stockholders' financial risks are greater than for the average citizen. While aggregate consumption and wealthy income data imply similar *patterns* for the equity premium since 1900, the *magnitude* of the effects are much greater for models based on Piketty and Saez's data. This paper is therefore able to capture both the shape and magnitude of the post-war bull market by parameterizing a simple discounted dividend model with risk premia estimated using income data.

While the focus of this paper is on the post-war period, the model is parameterized for the whole of the 1900s. It is able to broadly describe stock market behavior within a rational agent framework during this period. In

[3] Appendix A searches for direct evidence of stock selling in the early stages of the Great Depression by the rich to help consumption smooth. While it is difficult to draw any firm conclusions from this analysis, there is no compelling evidence of asset sales during this period by the most wealthy. This is not particularly surprising. The income of this top group is highly correlated with overall economic activity. When the wealthy are doing badly, so is everyone else. Therefore there are no other buyers in the market to help the rich consumption smooth by selling their financial assets. This problem is exacerbated by the Government issuing large amounts of new bonds into the market during difficult economic times.

particular, the model predicts (i) a higher average observed equity premium in the second half of the Twentieth Century than in the first half and (ii) highly volatile returns throughout the Century. Before the 1940s both income and dividend growth were highly volatile, leading to a high ex-ante and ex-post equity premium. During the 1970s and early 1980s, theoretical and observed returns were low as a consequence of poor corporate performance. The strong recovery in the 1990s was driven by the technology revolution. It is concluded that the ex-ante equity risk premium was low at the end of the Twentieth Century.

The paper proceeds as follows. In section 2, the ex-ante equity premium is estimated throughout the last one hundred years. This section first considers who the main stockholders were during this period and then examines the average income of this group to look for changes in underlying economic risk. A clear structural break in volatility emerges at approximately the time of the Second World War, and, using a Bayesian updating approach, it is demonstrated that investors would have learned very quickly about this change in risk. This finding is then incorporated into an asset pricing model to provide a conditional estimate of the required rate of return to stocks. Section 3 incorporates this cost of capital into a discounted dividend model, with dividend growth modelled as an MA(1), to provide an estimate of the level of the US stock market for each year from 1900–2000. Section 4 presents the results. The predicted stock market level and returns are compared against the observed values. It will be argued that the model is successful in explaining broad stock market behavior throughout the whole of the Twentieth Century. In particular, this model is able to explain the bull market of the post-War period. In addition, the predicted ex-post equity premium is higher since the structural break even though, by construction, the ex-ante equity premium is lower during the second half of the century. There is also only limited evidence of excess market volatility. Section 5 concludes.

2 Stockholder risk in the Twentieth Century

This paper is predicated upon the idea that the stock market is segmented. There are those who have sufficiently high income to participate and those who are precluded by virtue of their low income. It should be emphasised that this is *not* equivalent to segmenting the market on the basis of wealth. For example, Wolff (1994, p.161) states that "It is often believed that income and wealth are highly correlated. However ... suggest wide inequality of wealth among households with similar incomes". So, while it is accurate to say that, on average, high income households have high wealth, it cannot be safely assumed that dividing households by wealth percentile is the same as grouping them by income percentile. Figure 1, then presents an estimate of the division of stockholding by income groups since 1958. Details of the estimation process are provided in Appendix B.

[Insert figure 1 around here]

Stockholding is divided into four main categories: "Direct", "Indirect", "Non-profit" and "Other". The "Direct" and "Indirect" categories, which refer to the levels of household ownership of stock, are then further subdivided into categories "P0-90", "P90-99" and "P99-100", which refer to the income percentile of that household. For example, "P90-99" refers to households that had income in the top 10%, but not top 1%, of all tax units in that year. Direct stockholdings include bank personal trusts. Indirect stockholdings are limited to portfolios where the household ultimately takes the market risk. This category, therefore, includes mutual funds, variable annuities, 401k plans, etc. The "Non-profit" category is for non-profit organizations and, as argued in appendix B, the estimation of this portion is particularly problematic. The "Other" category includes foreign and corporate stock owners and defined benefit pension plans. As appendix B makes clear, there are a number of estimation problems that arise when constructing this graph. For this reason, figure 1 should be taken as indicative only. Nevertheless, it is possible to draw some important broad conclusions:

- In the period up to the mid 1960s, almost four-fifths of stocks were held either directly or indirectly by households. Direct stockholding was much more important than indirect stockholding. The ownership of equity was heavily concentrated amongst higher income households – certainly in the top decile but also extensively within the top percentile.

- Since 1970, the importance of household stockownership overall, direct stock ownership over indirect ownership, and ownership by the highest income groups, have all significantly declined. It appears to be very difficult to socially characterise the representative agent stock holder in recent times on the basis of this figure.

For the period before 1958, it is easier to characterise equity holders by income category since direct holdings by households was the overwhelming method of corporate ownership. Given this, the breakdown of stock ownership in the first half of the 1900s can be proxied by the proportion of dividend income directly received by each group. This information is given in Kuznets (1953, Table 124). In 1948, the top decile (percentile) received approximately three-quarters (half) of all dividends declared by households. These percentages seem high, but are substantially down on the concentration of stockownership earlier in the century, when the top decile (percentile) received approximately 90% (70%) of dividends. That is, the concentration of stock ownership in those families with the highest income has declined steadily since the start of the Century. In conclusion, it appears that by concentrating on households with the highest income, the majority of stock ownership is captured through to about 1970. Throughout the whole Century equity market participation broadened and since the early 1980s it has been less easy to identify stockholders through simple income characteristics. As the key features of the model presented below occurred in the 1940s to 1960s, proxying "stockholders" with "high income households" appears reasonable. It should, though, be noted that segmenting the market by income is not as successful as segmenting by wealth. In 1983 (1992), the top decile by income owned 77% (52%) of all household stock held either directly or indirectly compared to 90% (81%) owned by the top decile by wealth (Wolff (2000, Table 6)). Similarly, in 1958, the top percentile by income held 52% of all directly owned stock, while the top percentile by wealth held 75% (Smith & Franklin (1974)).

The income data upon which this study is based is described in detail in Piketty & Saez (2003) and covers the period from 1913 to 1998. It is taken from annual tax returns statistic compiled by the IRS and therefore refers to tax units rather than individuals. Throughout this paper, the term "household" is taken to be synonymous with "tax unit". With the exception of figure 2, all the income figures used in this study exclude capital gains. Piketty and Saez use "a gross income definition including all items reported on tax returns before all deductions. ... Income ... is computed before individual income taxes and individual payroll taxes and after corporate taxes". All figures are given in real $1998. To illustrate the key characteristics of this data for high income households, figure 2 shows the income (including capital gain but before personal taxation) of the top one percentile of tax units, broken down by category.

[Insert figure 2 about here]

There are two particularly important characteristics that emerge from this graph. First, dividend income made a much greater contribution to income in the early part of the Century. This suggests that the correlation between stockholder consumption and dividends was much higher in the pre-WWII period. Second, this figures indicates that income risk was more volatile in the first half of the Twentieth Century than it has been in recent times. This is examined in more detail in the next subsection.

2.1 Ex-post evidence of a change in risk

In this subsection, it will be formally demonstrated that the volatility of income growth of stockholders was higher in the first part of last century than the second. Income date is taken from Piketty & Saez (2001, Table A7) for a large number of income groups within the top decile of earners. The volatility of the logarithmic growth in total income of all these series was tested for structural breaks using the method of Chen & Gupta (1997). This method was also used to examine for structural breaks in volatility for real aggregate consumption growth and real market returns[4]. The reason for examining aggregate consumption is that the market segmentation story of this paper can then be readily compared with the implications of a complete market, representative agent, economy. Tests for structural breaks in the volatility of market returns are undertaken in light of recent work by Pástor & Stambaugh (2001) and Kim, Morley & Nelson (2003). They argue that, if the market price of risk is at least reasonably constant, then shifts in stock

[4]Robert Shiller's website provides real per-capita consumption since the 1880s for the total US population. Market returns are taken from Global Financial Data (GFD: www.globalfindata.com), which is used throughout for financial market data. The GFD series for prices, dividends and the risk-free rate are compared with figures on Robert Shiller's website. The two sets of series are very highly correlated and much of the difference is rounding error in the early years. The GFD data is preferred since there is less approximation error and the data can be taken up to 2000. Real prices, dividends and the real risk-free rates are calculated by adjusting the nominal series for the US CPI index as again provided by GFD.

market volatility should indicate structural breaks in the equity premium.

The Chen and Gupta test provides a diagnostic to indicate the statistical significance of the most prominent structural break in any given time series. It also indicates the point in time when this structural shift was most likely to occur. To identify more than one structural shift in volatility, the process is run iteratively on sub-samples of the data that are separated by previously identified break points. Table 1 presents information on the first structural change in variance found by the Chen and Gupta test. The pre-break period will be referred to in the remainder of the paper as "state 1" and the post-break period as "state 2".

[Insert Table 1 around here]

As can be seen, for all series an important reduction in volatility is observed that is most likely to have occurred between 1938 and 1947. In most cases, the significance is very much better than 1% and, in all cases, the break is significant at 5%. In the case of market returns, this confirms the findings of both Pástor & Stambaugh (2001) and Kim et al. (2003) that the equity premium should have declined around this period. The table presents evidence on the mean of income growth before and after the regime-shift. In the analysis below these figures will be referred to as μ_1 and μ_2 respectively.

It also presents evidence on the standard deviation of income growth before and after the break in variance (σ_1, σ_2 respectively). The sharp change in risk is clearly demonstrated. For example, for P99-100 — those in the top percentile of income in that year — $\sigma_1 = 11.9\%$ and $\sigma_2 = 6.1\%$. Table 1 also provides a theoretical estimation of the equity premium in the two sub-samples. From the fundamental theorem of asset pricing, the equity premium in terms of simple returns is given as:

$$E_t[r_{mt+1} - r_{ft+1}] = \frac{-Cov_t[r_{mt+1} - r_{ft+1}, \pi_{t+1}]}{E_t[\pi_{t+1}]} \tag{1}$$

where π_{t+1} is the pricing kernel and r_{ft+1} the risk-free rate from time t to $t+1$ which is known with certainty at t. If investors have time-separable power utility:

$$U(c_t, t) = \begin{cases} e^{-\rho t}\dfrac{c_t^{1-\gamma} - 1}{1-\gamma} & \gamma \neq 1 \\ e^{-\rho t}\ln(c_t) & \gamma = 1 \end{cases} \tag{2}$$

then the equity premium is given as:

$$E_t[r_{mt+1} - r_{ft+1}] = \frac{-Cov_t[r_{mt+1} - r_{ft+1}, R_{ct+1}^{-\gamma}]}{E_t[R_{ct+1}{}^{-\gamma}]} \tag{3}$$

where $R_{ct+1} = c_{t+1}/c_t$ is the simple growth in consumption. This model has been parameterized for $\gamma = 4$. In almost all cases, the predicted equity premium in state 1 is greater than state 2. As the market becomes more segmented by income, the stronger this pattern becomes. This is the main reason for concentrating on income, rather than consumption, data in this study. There is a clear break in aggregate consumption risk and the pre-break estimated equity premium is higher than the post-break premium in this case. However, the two values are close in absolute terms and well below the empirically observed values: 0.08% and -0.09% respectively. For income group P99-100, the two predicted values for the equity premium are 5.2% and -0.2% respectively.

The data is then tested for further structural breaks. In state 1 there is no evidence of a further structural break[5]. However, in state 2, for the higher income groups, there is evidence of a second structural break in 1986 for a number of the series. The reasons for this are clear when the growth rate of average household income is examined for the years 1987–1988.

[5]For two series starting in 1913, the test suggests a strucutral break in 1915, which appears spurious given the extremely small size of this first sub-period. For this reason, all income data is started at 1917 in this case. There is then no evidence of a further structural break between 1917 and the breakpoint identified in table 1.

	Income growth 1987-1988
$P0-90$	0.82%
$P90-95$	4.7%
$P95-99$	9.6%
$P95-99.5$	22.5%
$P99.5-99.9$	37.4%
$P99-99.99$	61.6%
$P99.99-100$	76.4%

This period saw a huge relative re-distribution of wealth with those with highest income doing exceptionally well. For categories P95 and above, these two years are certainly extreme outliers. There are at least three ways in which investors could rationally have interpreted this event. The first possibility is that, under the belief that volatility is fixed within regimes, this extreme event would be seen as a pre-cursor of high future uncertainty. In this case the ex-ante equity risk premium should have risen drastically during this period, which would predict a severe fall in the market. The second interpretation is that investors saw this sharp growth as an indication of significantly higher future earnings. The third interpretation is that investors saw this event as a permanent change in the *level* of income rather than its rate of change. In this case, the event does not alter investors' perceptions of either μ or σ. Under both the second and third interpretations, investors would have become less risk-averse over these two years (due to decreasing absolute risk aversion) and so the market should have performed well over this period. We believe that the balance of evidence lies most strongly with the third interpretation for three reasons: (i) it seems unlikely that a source of economic risk should emerge that would only affect the rich but have no impact on the poor, (ii) since 1989, income growth of the rich has indeed been relatively smooth but not dramatic and (iii) 1986–87 were indeed very strong years for the market. For this reason, table 1 had been re-estimated excluding the years 1987–1998. The revised figures are presented in table 2

[Insert table 2 around here]

The inferences to be drawn in this case are very similar to those that came from table 1. The main differences are that, in some cases, the structural shift in risk for the higher income groups is estimated to be about one decade later than table 1 would suggest. It is also apparent that their is greater statistical significance for the shift in risk for the higher income groups once the 1987–8 "jump" is excluded from the sample.

While this Chen and Gupta test is interesting in identifying the change in risk ex-post, there are three problems when using it as the basis for an asset pricing model: (i) it cannot be used contemporaneously since it requires the entire sample of data to estimate the break point, (ii) not all of the technical requirements of the test (constant mean, zero autocovariance) are satisfied by this data and (iii) while one given year is identified as the most likely time for the shift in regime, the test does not definitively identify this as the actual year of change. Therefore, the results of this paper are only directly dependent on the results of this test in the sense that the variables μ_1, μ_2, σ_1 and σ_2 will be taken from table 1. Instead, to better identify the timing of the change in risk, the paper follows the Bayesian learning approach of Moore & Schaller (2002).

2.2 Contemporaneous identification of risk

Was the reduction in risk in the middle of the last Century a one-off with no possibility of being reversed? Or, alternatively, was it a single switch of regime within a Markov-type environment? If the transition probability is low then these two hypotheses are observationally equivalent with less than one hundred years available. This paper assumes a switching process and investors learn about current underlying state through Bayes Theorem.

It is assumed that investors know that there are two states with indicator dummy variable $\delta_{st} \in \{0, 1\}$ for state $s \in \{1, 2\}$ so that $\delta_{2t} = 1 - \delta_{1t}$ for all t. The logarithmic growth rate of any given income / consumption series, r_{ct}, follows

$$r_{ct} = (\mu_1 + \varepsilon_{1t})\delta_{1t} + (\mu_2 + \varepsilon_{2t})\delta_{2t} \tag{4}$$

where μ_1, μ_2 denote the expected log growth of income in the two states and $\varepsilon_{1t}\tilde{\ }N(0, \sigma_1^2)$, $\varepsilon_{2t}\tilde{\ }N(0, \sigma_2^2)$ are noise terms. Parameters μ_1, μ_2, σ_1 and σ_2 are known to investors and, for estimation purposes, their values have been taken from table 1. Investors also know that the transition probability matrix, T, is given by

$$T = \begin{bmatrix} \zeta & 1-\zeta \\ 1-\zeta & \zeta \end{bmatrix} \tag{5}$$

for some large ζ. In the base parameterizations, $\zeta = 98\%$, which is within the 90% posterior bands of Kim et al. (2003, Table 2), although slightly higher than their mean estimate. Our results are unlikely to be highly sensitive to small changes in this parameter.

The variables that investors cannot observe are δ_{1t} and δ_{2t}. Therefore they have to draw inferences about these variables. If Ω_{t-1} denotes the information that investors have available to them at that time then $P_{t-1}(1)$ =Prob($\delta_{1t} = 1|\Omega_{t-1}$) and $P_{t-1}(2)$ =Prob($\delta_{2t} = 1|\Omega_{t-1}$). That is, $P_{t-1}(1)$ is the probability that an investor assigns at time $t-1$ to being in state 1 at time t. Given the conditional probability density functions for r_{ct} are

$$
\begin{aligned}
f(r_{ct}|\delta_{1t}=1) &= \frac{1}{\sqrt{2\pi\sigma_1}}\exp\left[\frac{-1}{2\sigma_1^2}(r_{ct}-\mu_1)^2\right] \\
f(r_{ct}|\delta_{2t}=1) &= \frac{1}{\sqrt{2\pi\sigma_2}}\exp\left[\frac{-1}{2\sigma_2^2}(r_{ct}-\mu_2)^2\right]
\end{aligned} \tag{6}
$$

then, according to Bayes Theorem:

$$
\text{Prob}(\delta_{st}=1|\Omega_t) = \frac{P_{t-1}(s)f(r_{ct}|\delta_{st}=1)}{P_{t-1}(1)f(r_{ct}|\delta_{1t}=1)+P_{t-1}(2)f(r_{ct}|\delta_{2t}=1)} \tag{7}
$$

and

$$
P_t(s) = \zeta\text{Prob}\left[\delta_{st}=1|\Omega_t\right] + (1-\zeta)\text{Prob}\left[\delta_{st}=0|\Omega_t\right] \tag{8}
$$

As observed by Moore & Schaller (2002, footnote 11), this is the same as the likelihood-based procedure used in Hamilton (1989). This places a maximum (minimum) possible value on $P_t(s)$ of ζ $(1-\zeta)$.

To run a Bayesian process, it is necessary to make assumptions about the prior probabilities of the underlying variables. Throughout, it is assumed that the starting probability of being in the high risk state $P_t(1)=\zeta$. Again, the results presented are highly insensitive to the choice of prior probabilities since income and consumption are so volatile early in the sample that $P_t(1)$ reaches its highest permissible value, almost immediately whatever the prior. The probabilities of being in the high-risk state during the Twentieth Century, $P_t(1)$, are presented in figure 3

[Insert figure 3 about here]

The crucial characteristic of this graph is that all income, consumption and market series identify the reduction in risk over the same period between the 1940s and mid 1960s. Given tables 1 and 2 above, this should, perhaps, not be surprising. This implies that, rather than declining slowly over the second half of the Century, the *ex-ante* risk premium would have reached a new, lower, level by the mid-1960s at the latest.

While the lines in figure 3 are all broadly similar, it is necessary to choose specific probabilities $P_t(s)$ for use in the asset pricing model below. Before 1918, estimates are taken from figure 3 for aggregate consumption. Since 1918, the probabilities are taken for the series "P0-90". Taking estimates from the low income group may appear inconsistent with the market segmentation story of this paper. However, this helps overcome the 1987–8 effect and what appears to be a spurious reduction in risk during the Second World War which is apparent in some series[6]. In other years, the probabilities from this series are similar to those from higher income groups. Figure 4 shows

[6] For example, $P_t(1)$, when estimated from aggregate consumption for the entire sample fell from 96.7% in 1940 to 13.3% in 1945 but had then rose to 97.8% by 1947. Similar patterns are observed in the high-income data when the Bayesian method is taken for the period up to 1986. To characterise a war period as "low risk" does not seem economically reasonable (although no adjustments are made here for the Korean and Vietnam wars). In an earlier draft of the paper, figure 4 was estimated for P99-100 for the period up to 1986. $P_t(1)$ was then set equal to 0.98 for all Second World War years. The results are economically similar to those reported here.

the probabilities that are used in the parameterizations given later in the paper.

[Insert figure 4 about here]

2.3 The ex-ante equity premium

One of the unusual features of this parameterization is that, for many series in Table 1, the equity premium has been forecast to be very small or even negative since the regime shift in risk. This seems extremely intuitively unlikely. Therefore, it would appear that the model is missing a fixed element of the equity premium. Given the number of potential explanations for the puzzle highlighted by Mehra (2003), it is unsurprising that this single explanation cannot explain the whole effect. That is, this paper is concentrating on the *change* in risk premium over the Twentieth Century rather than its absolute level. For this reason, approximately 3.5% of risk premium is added to the estimated values for both states. Even if this adjustment is not made, graph 6 below is largely unaltered — the equity premium can still be broadly explained. However, since the cost of capital is uniformly too low, the predicted level of the market becomes too high throughout the entire sample. In this case, the theoretical level of the market, presented in graph 5 below, becomes significantly less realistic. In other words, within the context of this paper, the "equity premium puzzle" can be rephrased in terms of questioning why it is necessary to add on this fixed 3.5% throughout the last one hundred years. Given this adjustment, though, use $\rho(1)$, $\rho(2)$ to respectively denote the *logarithmic, real* equity premium conditional on the high risk / low risk state. Set $\rho(1) = 8\%$ and $\rho(2) = 3.5\%$.

The relevant equity premium for asset pricing depends not only on the current value, but projections of future values. It will be shown below that, at time t, to discount the dividend at time $t+i$, the relevant risk premium is $E_t \exp\left[\sum_{j=1}^{i} -\rho_{t+j}\right]$ where ρ_{t+j} is the one-period logarithmic equity premium j periods ahead. If (i) there was never a change in state and (ii) if the current state was perfectly known, then this is easily calculated; $E_t \exp\left[\sum_{j=1}^{i} -\rho_{t+j}\right] = \exp(-0.08i)$ or $\exp(-0.035i)$ depending on the current state. However, uncertainty about the current state, together with potential future changes in state, makes estimation more complex. Define $\rho_i(s) = E_t\left(\exp\left[\sum_{j=1}^{i} -\rho_{t+j}\right] | \delta_{st+1} = 1\right)$; that is the relevant equity premium for discounting the $t+i^{th}$ cashflow conditional on being in state $s \in \{1,2\}$ at time $t+1$. Then:

$$E_t \exp\left[\sum_{j=1}^{i} -\rho_{t+j}\right] \quad = \quad P_t(1)\rho_i(1) + P_t(2)\rho_i(2) \tag{9}$$

To calculate $\rho_i(1)$ and $\rho_i(2)$, an iterative computational approach is taken.

For any given i, the probability of being in state s exactly k times is calculated for all $k \in \{0, i\}$. Since the ordering of states is irrelevant, this is sufficient to calculate the equity premium

$$\rho_i(s) = \sum_{k=0}^{i} \text{Prob}\left(\sum_{j=1}^{i} \delta_{1t+j} = k | \delta_{st+1} = 1\right) \exp[-0.08k - 0.035(i-k)] \tag{10}$$

where (surpressing the dependence on $\delta_{st+1} = 1$ for notational simplicity):

$$\begin{aligned}
\text{Prob}\left(\sum_{j=1}^{i} \delta_{1t+j} = k | \delta_{1t+i} = 1\right) &= \zeta \text{Prob}\left(\sum_{j=1}^{i-1} \delta_{1t+j} = k-1 | \delta_{1t+i-1} = 1\right) + \\
&\quad (1-\zeta)\text{Prob}\left(\sum_{j=1}^{i-1} \delta_{1t+j} = k-1 | \delta_{1t+i-1} = 0\right) \\
\text{Prob}\left(\sum_{j=1}^{i} \delta_{1t+j} = k | \delta_{1t+i} = 0\right) &\quad (1-\zeta)\text{Prob}\left(\sum_{j=1}^{i-1} \delta_{1t+j} = k | \delta_{1t+i-1} = 1\right) + \\
&\quad \zeta \text{Prob}\left(\sum_{j=1}^{i-1} \delta_{1t+j} = k | \delta_{1t+i-1} = 0\right)
\end{aligned} \tag{11}$$

and

$$\text{Prob}\left(\sum_{j=1}^{i} \delta_{1t+j} = k\right) = \sum_{s=0}^{1} \text{Prob}\left(\sum_{j=1}^{i} \delta_{1t+j} = k | \delta_{1t+i} = s\right) \tag{12}$$

3 Simulating market prices

Consider the discounted dividend model, where p_t, d_t and R_t respectively represent the market price, dividend and gross one-period cost of capital. Then:

$$
\begin{aligned}
p_t &= \frac{E_t[d_{t+1} + p_{t+1}]}{R_{t+1}} \\
&= \frac{1}{R_{t+1}} E_t \left[d_{t+1} + \frac{E_{t+1}\left\{d_{t+2} + p_{t+2}\right\}}{R_{t+2}} \right]
\end{aligned}
\tag{13}
$$

For any two time-series, $\tilde{x}_t, \tilde{y}_t$, positive integers j, l and non-negative integer k, $E_t[\tilde{x}_{t+j} E_{t+j+k}(\tilde{y}_{t+j+k+l})] = E_t[\tilde{x}_{t+j}\tilde{y}_{t+j+k+l}]$. To see this, $\tilde{y}_{t+j+k+l} = E_{t+j+k}[\tilde{y}_{t+j+k+l}] + \tilde{\varepsilon}_{t+j+k+l}$ where $\tilde{\varepsilon}_{t+j+k+l}$ is unforecastable at $t+j+k$. Given that k is non-negative, this random variable is also unforecastable at $t+j$. $E_t[\tilde{x}_{t+j} E_{t+j+k}(\tilde{y}_{t+j+k+l})] = E_t[\tilde{x}_{t+j}(\tilde{y}_{t+j+k+l} - \tilde{\varepsilon}_{t+j+k+l})] = E_t[\tilde{x}_{t+j}\tilde{y}_{t+j+k+l}] - E_t[\tilde{x}_{t+j}\tilde{\varepsilon}_{t+j+k+l}]$. Due to the independence of $\tilde{x}_{t+j}$ and $\tilde{\varepsilon}_{t+j+k+l}$, the second term is zero. The discounted dividend model can therefore be re-written as

$$
p_t = \frac{1}{R_{t+1}} E_t \left[d_{t+1} + \frac{d_{t+2}}{R_{t+2}} + \frac{E_{t+2}\left\{d_{t+3} + p_{t+3}\right\}}{R_{t+2}R_{t+3}} \right] \tag{14}
$$

and, by iteration,

$$
p_t = d_t \sum_{i=1}^{\infty} E_t \left[\prod_{j=1}^{i} G_{t+j} R_{t+j}^{-1} \right] \tag{15}
$$

where $G_{t+j} = d_{t+j}/d_{t+j-1}$ is the gross dividend growth rate. Let $g_{t+j} = \ln G_{t+j}$ and $r_{t+j} = \ln R_{t+j}$. Then

$$p_t = d_t \sum_{i=1}^{\infty} E_t \exp\left[\sum_{j=1}^{i} g_{t+j} - r_{t+j}\right] \tag{16}$$

The discount rate can be divided into a risk-free component r_{ft+j} and a risk premium component ρ_{t+j}. Then

$$p_t = d_t \sum_{i=1}^{\infty} E_t \exp\left[\sum_{j=1}^{i} g_{t+j} - r_{ft+j} - \rho_{t+j}\right] \tag{17}$$

It is next assumed that the risk premium is independent of both dividend growth rates and the risk-free rate. This is consistent with the principle that risk premia depend on the second, rather than first, moment. It will also be assumed that the differences between logarithmic dividend growth rates and risk-free rates are normally distributed. Under these two assumptions:

$$p_t = d_t \sum_{i=1}^{\infty} \exp\left[\mathbf{1_i'} \mathbf{m_{t,i}} + \frac{1}{2}\mathbf{1_i}'\mathbf{S_{t,i}}\mathbf{1_i}\right] E_t \exp\left[\sum_{j=1}^{i} -\rho_{t+j}\right] \tag{18}$$

Here $\mathbf{1_i}$ is a $i \times 1$ vector of 1s. $\mathbf{m_{t,i}}$ is an $i \times 1$ vector with elements $E_t[g_{t+j} - r_{ft+j}]$ and $\mathbf{S_{t,i}}$ is an $i \times i$ variance/covariance matrix with elements $\mathrm{Cov}_t[g_{t+j} - r_{ft+j}, g_{t+k} - r_{ft+k}]$ respectively.

3.1 ARIMA modelling dividend growth

In order to estimate the model, the time series $g_t - r_{ft}$ (which is called the "dividend growth process") is ARIMA modelled in-sample[7]. Real dividends and the real risk-free rate are again taken from GFD. One important adjustment is made to these series. In more recent times, particularly the 1990s, stock repurchases were an important component of cash repayments to investors. Therefore, from 1972 onwards, stock repurchases are included as a form of dividend payment. Data from 1972 — 2000 are taken from Grullon & Michaely (2002). In each year, the multiplier$_t$ = (dividends$_t$ + repurchases$_t$) / dividends$_t$ is calculated. Then in each year from 1972 onwards, the GFD dividend series is multiplied by multiplier$_t$ / multiplier$_{1972}$. This then splices the data onto the original series for 1972 but allows for changes in stock repurchases thereafter[8]. If stock repurchases are excluded from dividend data, then the model predicts lower stock returns in the 1990s in particular.

The first finding is that there is no evidence of a unit root in the dividend growth process. In an earlier, related, study Barsky & De Long (1993) argue

[7]In earlier versions of the paper g_t and r_{ft} were ARIMA modelled separately. However, as the volaltility of real dividend growth is much greater than the volatility of the real risk-free rate, the g_t effect dominates the r_{ft} effect. It is therefore simpler to model $g_t - r_{ft}$ as one series and it make almost no difference to the presented results.

[8]A similar multiplier can be constructed from the data in Dittmar (2000) for the period 1977–1996. While the absolute values in Dittmar (2000) differ from those in Grullon & Michaely (2002), the multiplier in each year is similar.

that observed changes in the US stock market are broadly consistent with the discounted dividend model. However, their dividend growth process was integrated, meaning that change in dividend were highly informative about future growth. This assumption has been largely questioned in subsequent literature. For example, Pagès (1999) argues that this dividend growth process is inconsistent with using the Gordon Growth model. This paper agrees with the findings that shocks in dividend growth have very little predictive ability for future growth rates. This paper therefore follows Bansal & Lundblad (2002) in ARMA modelling $g_t - r_{ft}$. They use an ARMA(1,1) process:

$$g_t - r_{ft} = \alpha_0 + \alpha_1 \left[g_{t-1} - r_{ft-1}\right] + \eta_t + \beta\eta_{t-1}$$

This model is estimated on the data for this paper (t-statistics in brackets):

Model	α_0	α_1	β_1	AIC
MA(1)	0.271% (0.18)	0	0.250 (2.78)	7.929
AR(1)	0.261% (0.17)	0.227 (2.46)	0	7.936
ARMA(1,1)	0.271% (0.19)	-0.005 (-0.02)	0.255 (0.94)	7.947

The ARMA(1,1) is thus rejected and the MA(1) marginally chosen over the AR(1). Given the low estimate of α_1 in the AR(1) model, both will result in similar forecasts of future dividends. In particular, dividend shocks have noticeable forecasting power for future growth rates only one period ahead.

In the case of the MA(1) process, $E_t[g_{t+1} - r_{ft+1}] = \alpha_0 + \beta_1\eta_t$ and, for all $j > 1$, $E_t[g_{t+j} - r_{ft+j}] = \alpha_0$. Given this, $\mathbf{1_i'm_{t,i}} = i\alpha_0 + \beta_1\eta_t$. Use $\sigma^2 = \text{Var}(\eta_t)$, which is taken to be homoskedastic. Since dividend growth is taken to be MA(1), $\text{Var}_t[g_{t+j} - r_{ft+j}] = \sigma^2(1+\beta_1^2)$ for $j > 1$. The conditional first order auto-covariance $\text{Cov}_t[g_{t+j} - r_{ft+j}, g_{t+k} - r_{ft+k}] = \beta_1\sigma^2$ for all $j \geq 1$. The autocorrelation is zero for all higher orders. So, $\mathbf{1_i}'\mathbf{S_{t,i}1_i} = i\sigma^2(1+\beta_1^2)$ $+ 2(i-1)\beta_1\sigma^2$. Defining:

$$\Gamma_i = \exp[i\alpha_0 + 0.5i\sigma^2(1+\beta_1^2) + (i-1)\beta_1\sigma^2]$$

and equation 18 becomes:

$$p_t = d_t \sum_{i=1}^{\infty} \Gamma_i \exp\left[\beta_1\eta_t\right] E_t \exp\left[\sum_{j=1}^{i} -\rho_{t+j}\right] \tag{19}$$

and combining with equation 9

$$p_t = d_t \sum_{i=1}^{\infty} \Gamma_i \exp\left[\beta_1\eta_t\right] \{P_t(1)\rho_i(1) + P_t(2)\rho_i(2)\} \tag{20}$$

This is the asset pricing model that will be used to simulate the US stock market over the past Century.

4 Results

The summation on the right-hand side of equation 20 runs to infinity. However, as the risk premium is being generated by an iterative computational technique, it is clearly not possible to generate the series that far. For this reason, it is assumed that all cashflows after the 1,000th year have zero present value. Again, the presented results are largely insensitive to the point of truncation. The errors, η_t, are estimated in-sample from the MA(1) modelling of the total sample while d_t is observable contemporaneously. Given these values, it is possible to estimate the price for each year, p_t. When the parameter estimates are included for Γ_i, $\rho_i(1)$, and $\rho_i(2)$;

$$p_t = d_t \exp\left[0.250\eta_t\right]\left[19.11P_t(1) + 34.11P_t(2)\right] \tag{21}$$

The values 19.11, 34.11 can be broadly interpreted as price / dividend ratios in the two states. This is fitted for all years from 1890–2000.

Figure 5 presents the model's prediction of the market value against the actual observation for the period 1900–2000. Figure 6 converts this into ten-year rolling average equity premia:

[Insert figures 5 and 6 around here]

and the static properties of the real ex-post excess returns to the market are:

	Mean		Standard deviation	
	Theory	Observed	Theory	Observed
1890–1946	5.1%	4.8%	20.0%	21.2%
1947–2000	6.9%	8.1%	13.1%	15.9%

It is known, by construction in the theoretical case, that the (logarithmic) ex-ante equity premium is 4.5% higher before 1946 than after this date. However, the predicted ex-post premia are actually higher since the structural shift in risk than before. That is, an econometrician, observing the equity premium as a time-series, would not be able to directly observe the change in state in 1946 (although, as shown by Pástor & Stambaugh (2001) and Kim et al. (2003), they can infer this from the decrease in the volatility of returns). While the predicted returns after 1947 are slightly lower than the observed returns, the equity premium puzzle no longer remains as an order-of-magnitude. In addition, the second moment of the equity premium is similar to the observed values. That is, while there is slight excess volatility, this no longer remains a serious issue.

Under the explanation presented in this paper, the Twentieth Century can be divided into four periods. From 1900–1946, two sources of risk were high. The income of the rich was very volatile and highly correlated with stock returns. This resulted in a high ex-ante equity premium which was close to the observed actual values. In addition, dividends were volatile, leading to a high second moment of stock returns. From 1947 until around 1961, investors quickly recognized the volatility shift of their income. This led to a significant reduction in the ex-ante equity premium and the very strong bull market of the 1950s. One of the main successes of this paper is in its ability to explain why the stock market should have performed so strongly in this decade. As noted elsewhere (see, for example, Mankiw, Romer & Shapiro (1985)), previous models of this type have found this decade anomalous. After the mid 1960s until the early 1980s the market had low ex-ante and ex-post returns. Poor dividend performance caused a depressed market. The model in this paper is over-optimistic for this period. Since 1980, the ex-ante equity premium has remained low, but the stock market has been strongly driven by improved corporate performance. In contrast to the 1970s, the model of this paper is somewhat over-pessimistic for this period, although this may just be a re-correction; the predicted level of the index in 2000 is close to its observed values.

5 Conclusion

Fama & French (2002, p.658) conclude their paper with the observation that:

> "... our main message is that the unconditional expected eq-

> uity premium of the last 50 years is probably far below the realized premium"

This study presents an underlying economic interpretation of why this should be so. Income data of the rich clearly demonstrates a highly significant change in risk for stockholders during the middle of the last Century. Incorporating this finding into a simple discounted dividend model helps explain the observed returns to stocks over the whole of the last one hundred years, and particularly the post World War II period, as demonstrated by figures 5 and 6 above. It can therefore be concluded that the *ex-ante* equity premium at the start of the Twenty-First century is significantly below the average historical values.

Three limitations with this study suggest ways forward for future research. First, concentrating on the pre-tax income of the average wealthy household is not an ideal proxy for the consumption of individual stockholders. It has been argued that this is unlikely to bias the results and the benefits of looking at long-horizon data in this context outweigh the costs of using income numbers. This does, though, obviously place limitations on the accuracy of the presented theoretical prices and returns. One extension of this study would be to examine alternative measures of stockholder risk throughout the last one hundred years. Second, if consumption risk did decline significantly in the 1950s, then this should have been associated with a sharp rise in the real risk-free rate as the precautionary savings demand fell. There is some evidence of this; see, for example Blanchard (1993, figure 10). Nevertheless, the real risk-free rate has been smoother than this model would suggest unless there is notable market segmentation between stockholders and bondholders. Further work is needed on the risk-free rate puzzle in light of the findings in this study. Finally, it has been necessary to add an extra equity premium of around 3.5% to the forecast value throughout the sample in order to explain the level of the US stock market. Since this paper is concentrating on *changes* in risk premium, this has not been of overriding concern here. The author believes that the "equity premium puzzle" can now be re-expressed as trying to explain this 3.5%.

Appendix A

In this appendix, it is examined whether the rich representative agent sold stocks and bonds during the early stages of the Great Depression to help smooth consumption.

If investors are selling assets to compensate for transitory low income, then it might be expected that income *including* capital gain would be smoother than income excluding capital gain. This is not the case. Between 1917 and 1951, the standard deviation of the growth in income including (excluding) capital gain was 12.4% (10.6%). In 1928, average capital gain income in 1998 dollars was over $42,000. By 1932 this has declined to under $1,500. That is, trading stocks and bonds appears to increase, not reduce, the financial risk to this representative agent. Looking at income from capital *gain*, though, may be misleading. It is possible that in the Great Depression, the rich sold multiple assets to consumption smooth but at a capital loss due to the fall in stock prices. Similarly, in 1928, investors may have been realizing capital gains, but then immediately re-investing the money into the market rather than consuming it.

To examine this, compare the overall change in interest payments between 1928 and 1934 with those experienced by the rich. In 1928, the total amount of Public Sector debt[9] was approximately $33,393m, corporate debt outstanding around $26,476m and bank deposits $61,480m . The amount of interest paid by the public sector was $1,426m. I estimate that the amount of interest paid by corporate debt was $1,430m (details available upon request). Averaging the short-term commercial paper rate over 1928 gives an interest rate of 4.84%. Using this as a proxy for bank rates over the period, suggests that banks paid interest of $2,975m. This provides an estimate of total interest payments of $5,831m.

By 1934, Public Sector interest payments had only risen to $1,571m, even though the level had debt had risen quite sharply to $45,982m. This suggests that, if the size of the public sector bond market had not changed, interest payments would have dropped to $1,141. By contrast, corporate bond interest payments probably did not change very much over this period since there was only limited refinancing by the corporate bond sector. However, we estimate that approximately 17% of corporate bond payments were

[9]Figures for public sector debt are taken from series Y531 and Y528 in "Historical Statistics of the United States; Colonial Times to 1970", U.S. Department of Commerce Bureau of the Census. Figures are only available quadrennially, so this refers to 1927, not 1928. Corporate debt data is taken from Hickman (1960) and for bank deposits from Wicker (1996, Table 1.1). Commercial paper rates are provided by Macaulay (1938, Appendix A10)

lost as a consequence of financial distress (details available upon request). Therefore, on a like-for-like basis, it is estimated that corporate bond interest payments equalled \$1,167m in 1934. Finally, the short-term interest rate averaged 0.89% in 1934. In addition, approximately 7.5% of deposits were lost as a consequence of bank failure. On the basis of the market size in 1928, this would suggest bank interest payments of \$506m. So, on the same-size market, this suggests total interest payments of \$2,814m. This is a nominal decline of 52% over this period. Therefore, if the rich had neither bought nor sold bonds and working under the assumption that they invested approximately in proportion to the overall market weights in 1928, it might be expected that their nominal interest payments fell by 52% nominally during the early stages of the Great Depression. In fact, their nominal interest payments fell by somewhat more than this - around 63%. This suggests that there may have been some bond selling by the rich representative agent during this period. However, the 1917 to 1951 period appears, in general, to have been a time when the rich sold bonds irrespective of short-term consumption smoothing arguments. In 1917 real interest payments were \$26,491 against \$5,152 in 1951. Once this time-trend is taken into account, the incremental fall in nominal interest payments to the rich is estimated to be 56% between 1928 and 1934: very close to the estimated aggregated number. This suggests very limited consumption smoothing by selling bonds during the Great Depression.

A similar argument can be applied to dividend payments. Between 1928 and 1934 nominal dividend payments to the rich representative agent fell by 60%. Aggregate nominal dividends over the same period fell by 54%. Again, this suggests there may have been limited selling of stocks, but the difference may well just be caused by measurement error. To test this, an OLS regression of the percentage change in real dividend income of the rich (Δd_{rt}) against aggregate percentage changes in dividend (Δd_{at}) from 1917 to 1951 was then run:

$$\begin{array}{rcccc} \Delta d_{rt} = & -0.0156 & + & 0.90\Delta d_{at} & +\epsilon_t \\ & [0.016] & & [0.09] & \end{array} \tag{22}$$

As can be seen, the wealthy, on average, sold stock over this period; the intercept is less than zero. However, there is no evidence of more stock selling (buying) in bad (good) years. If this were the case, then the gradient of the previous regression would have been greater than one. The cross-plot indicates that this result is not influenced by non-linearity.

Appendix B

This appendix describes the estimation process for figure 1. This is a two-stage process. First, stock holdings are divided into four broad categories, "Direct", "Indirect", "Other" and "Non-profit" for the total sample. Then "Direct" and "Indirect" are further divided into income percentiles "P0-90", "P90-99" and "P99-100"

Stage 1

To break equity ownership into broad categories, data was taken from table 6 in Poterba & Samwick (1995). "Direct" ownership is defined as "Flow of Funds household ownership" - "Nonprofits" - "Mutual Funds" - "Defined contribution pension" - "Variable annuities". Notice that this still includes "Band personal trusts" as a direct form of stock ownership. This is because this category was only identified in 1969 and therefore this is the most consistent method of handling the data. "Indirect" stock holding is therefore the sum of "Mutual funds" + "Defined contribution pensions" + "Variable annuities". Notice that this does not include holdings in defined benefits plans and therefore all the financial market risk in this category is ultimately borne by the underlying household. The "Non-profit" is taken directly from this table. The "Other " category is then just the residual. A further breakdown of this category is given in Poterba & Samwick (1995, Table 5), but this is not of direct relevance here.

There appear to be two fairly serious problems with using this data. The first involves using the Federal Reserve Board's flow of funds. The definition of "equity" here is very broad and includes closely-held company stock. It is known that the ownership characteristic of such companies is different to that of publicly quoted firms. In particular, the level of indirect holding is much lower for such companies. Since this paper is concentrating on the equity premium, a more narrow definition of equity would be preferable and this suggests that figure 1 overestimates the level of direct stockholding for the purposes of this study. There is a further problem. The flow of funds does not distinguish between "household" and "non-profit" sectors and this division is estimated by Poterba and Samwick. They describe this as follows (p.314)

> Experimental data presented in the Flow of Funds accounts show that the equity holdings of non-profit institutions averaged 15.7 percent of the household sector's equity holdings during the period 1987–92. Therefore we multiply the Flow of Funds household sector equity value by 0.843 for each year between 1952–1994 in order to remove these holdings.

While figure 1 makes no adjustment to Poterba and Samwick's number, this assumption leads to a major uncertainty about the accuracy of figure 1 in the early part of the sample.

Stage 2

To investigate household holdings by income percentile, two sources are used. For 1983 and 1992, information is taken from Poterba & Samwick (1995, Table 10). It should be noticed that this is not completely consistent with Stage 1 since this table concentrated on publicly traded stock only. Since this table is for income levels, not percentiles, it is necessary to interpolate from this information so a linear spline technique is used. This method implies that the top income decile owned 82.5% of publicly traded stock in 1983, which is very similar to the estimate of Avery & Elliehausen (1986, Table 6) of 85%, suggesting that this interpolation technique is reasonable. The "Percentage of stock owned; publicly traded stock" is used to divide "Direct" into income percentiles in figure 1[10]. By appropriately weighting "Percentage of stock owned; publicly traded, mutual funds, IRA/Keogh, and defined contribution plans" and "Percentage of stock owned; publicly traded stock" by the relative size of the "Direct" and "Indirect" categories, it is possible to deduce the proportions of "Indirect" owned by each income category.

Before 1983, the "Indirect" category was very small. For this reason, the 1983 weightings for indirect stock ownership by income category are applied to all early years. This is unlikely to introduce a serious problem. For the "Direct" category, information is taken from Blume, Crockett & Friend (1974, Table 4)

[10] There is a potential problem here as the "Direct" category includes "Bank personal trusts". No information is given on the breakdown of trusts ownership by income group so no adjustment is made for this.

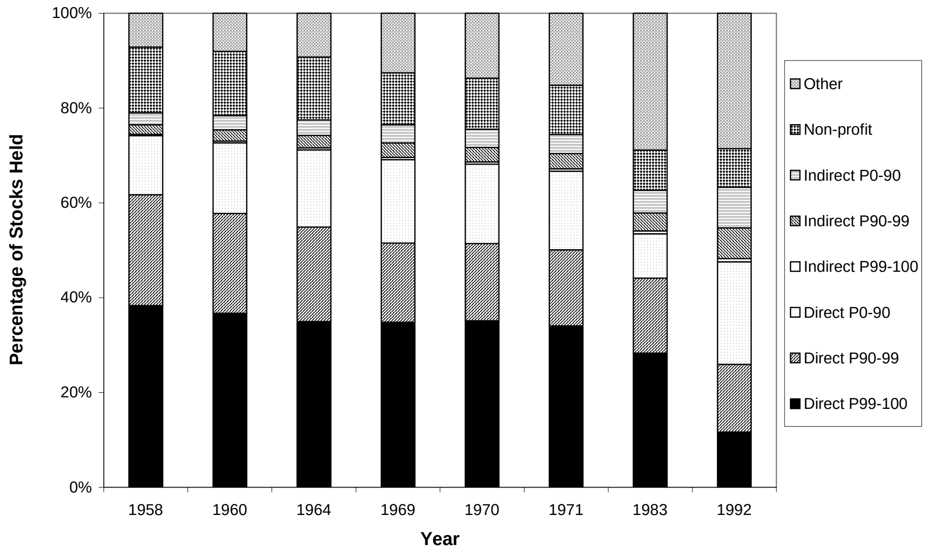

Figure 1: This graph estimates the proportion of the stock market held directly and indirectly but in defined contribution schemes (mutual funds, non-defined benefit pension plans, etc.) by those in the US population with the highest income. For example, P90-99 are those families with income in the highest 10%, but not highest 1%, in that year. These figures are the author's estimates based on Poterba and Samwick (1995) tables 5, 6 and 10 and Blume *et. al.* (1974) table 4. Since 1969, "bank personal trusts" holdings are counted as a direct form of stockholding to be consistent with earlier periods. It should be emphasised that there is substantial estimation error in arriving at this graph and it should therefore be interpreted as indicative only. See the text for more details.

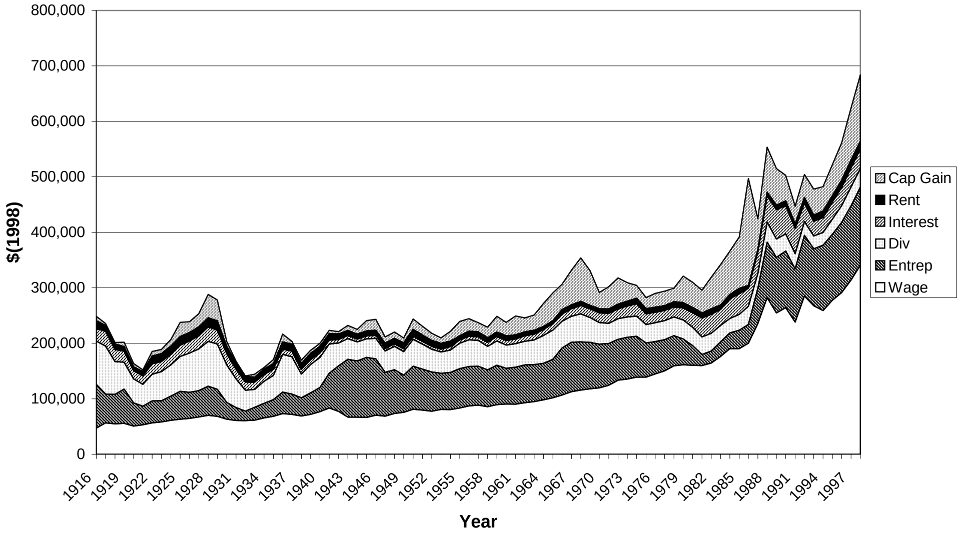

Figure 2: This figure shows the breakdown of pre-tax average income in 1998 dollars by category. This is for households in category P99-100: that is those, in the top 1% of pre-tax income in any given year. Notice that, uniquely for this paper, income here includes capital gains. Source: Piketty & Saez (2001)

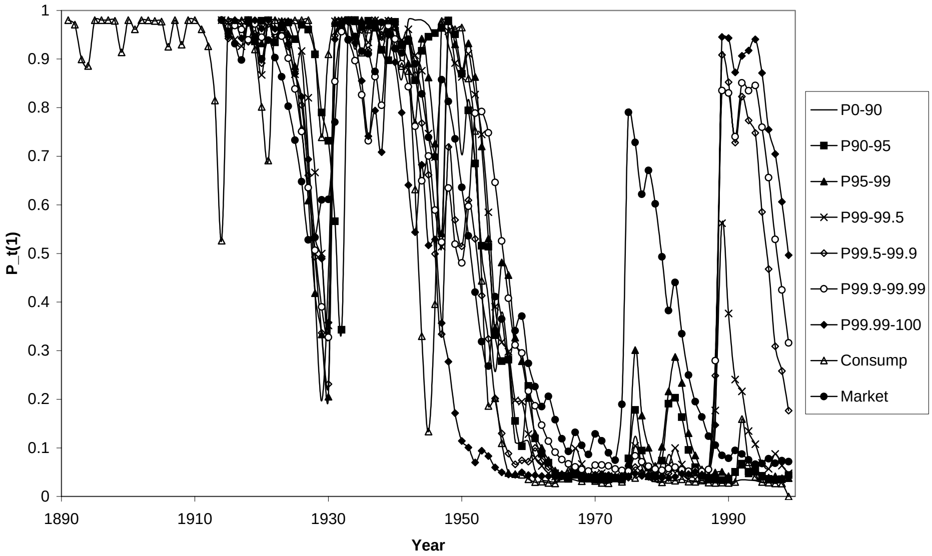

Figure 3: This graph presents estimates of the probabilities that investors assign at time $t-1$ of being in the high risk state at time t. This is based on the income data of Piketty and Saez (2001), the aggregate real consumption data given by Robert Shiller's website and stock market returns from Global Financial Data. The estimates are based upon the Bayesian learning model of Moore and Schaller (2002).

Figure 4: This graph gives the probabilities of being in the high-risk state, $P_t(1)$, used in equation 20 to generate figures 4 and 5 below. $P_t(2) = 1 - P_t(1)$. Before 1918, the probability is taken from aggregate consumption estimates. Since 1918, figures are taken from the estimates for "P0-90" to avoid the effects of the second world war and the 1986–7 period.

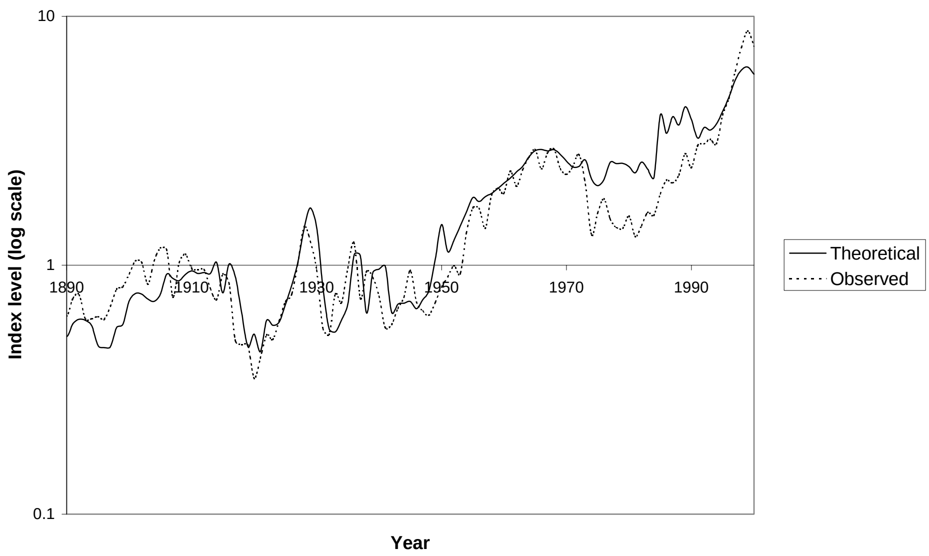

Figure 5: This graph presents the observed real level of the US stock market and the simulated real level of the market as calculated from the model presented in this paper in equation 20

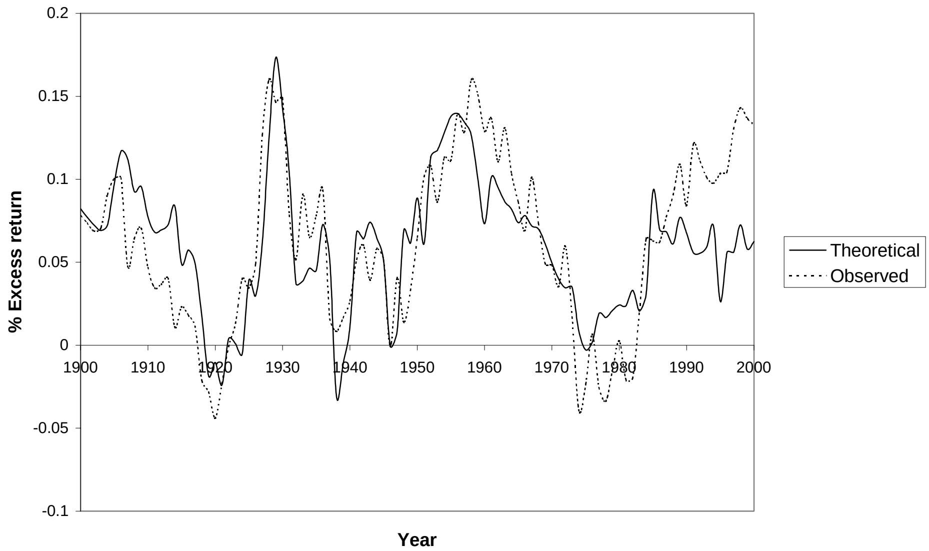

Figure 6: This graph presents the observed 10-year rolling average real equity premium from 1900-2000 for the United States and the simulated value from the model presented in this paper.

Percentile	P0-90	P90-95	P95-99	P99-99.5	P99.5-99.9	P99.9-99.99	P99.99-100
Year of Break	1946	1947	1947	1947	1941	1939	1938
Significance	<1%	<1%	<1%	<1%	<1%	<5%	<1%
Mean before break	1.560%	1.001%	0.977%	1.227%	0.492%	-1.845%	-2.353%
Std dev before break	11.092%	6.266%	6.176%	8.691%	11.519%	13.999%	21.950%
Simple EP before break	0.792%	0.017%	1.283%	2.252%	1.853%	4.796%	12.050%
Mean after break	1.083%	1.893%	1.835%	1.590%	1.486%	1.898%	2.810%
Std dev after break	3.083%	2.463%	2.613%	3.835%	5.585%	8.773%	10.840%
Simple EP after break	-0.011%	0.027%	0.152%	0.007%	0.135%	-0.704%	-1.135%

Percentile	P90-100	P95-100	P99-100	P99.5-100	P99.9-100	Consump	Market
Year of Break	1947	1939	1939	1939	1938	1949	1941
Significance	<1%	<1%	<1%	<2%	<5%	<1%	<5%
Mean before break	0.4168%	-0.3659%	-0.6784%	-1.0816%	-2.3965%	1.668%	3.992%
Std dev before break	7.2951%	9.3191%	11.8682%	12.9877%	15.4420%	4.276%	26.148%
Simple EP before break	2.1527%	4.2592%	5.1858%	5.8191%	8.3895%	0.081%	
Mean after break	1.9206%	1.9086%	1.9215%	1.9793%	2.3491%	1.841%	8.591%
Std dev after break	3.1172%	4.0133%	6.0824%	7.0838%	9.0228%	1.068%	16.282%
Simple EP after break	0.0017%	0.0127%	-0.2375%	-0.3714%	-0.7733%	-0.085%	

Table1: This table runs the Chen and Gupta (1997) test for structural breaks in volatility on Piketty and Saez (2001) average household income growth, aggregate consumption growth ("consump") and market returns ("market"). "P90-95", for example, refers to the average income of households in the top 10%, but not 5%, of taxable income in any given year. The table identifies the year when the most prominent structural shift in volatility was most likely to occur, together with the significance of the structural break. Characteristics of the data in the two subsamples (the mean and standard deviation of income growth and a theoretically estimated equity premium (EP)) is also provided.

Percentile	P0-90	P90-95	P95-99	P99-99.5	P99.5-99.9	P99.9-99.99	P99.99-100
Year of Break	1946	1947	1947	1950	1950	1952	1938
Significance	<1%	<1%	<1%	<1%	<1%	<1%	<1%
Mean before break	1.560%	1.001%	0.977%	1.330%	0.572%	-1.318%	-2.353%
Std dev before break	11.092%	6.266%	6.176%	8.459%	10.469%	12.750%	21.950%
Simple EP before break	0.792%	0.017%	1.283%	2.150%	1.828%	2.983%	12.050%
Mean after break	1.312%	2.082%	1.767%	1.005%	0.697%	1.142%	1.225%
Std dev after break	3.321%	2.575%	2.731%	3.200%	3.628%	4.479%	6.887%
Simple EP after break	-0.125%	-0.048%	0.110%	-0.005%	-0.086%	-0.584%	-1.165%

Percentile	P90-100	P95-100	P99-100	P99.5-100	P99.9-100	Consump	Market
Year of Break	1950	1950	1952	1941	1939	1949	1938
Significance	<1%	<1%	<1%	<1%	<1%	<1%	<10%
Mean before break	0.7036%	0.5032%	-0.1384%	-0.3520%	-1.8463%	1.668%	5.885%
Std dev before break	7.0967%	8.1984%	10.3303%	12.7925%	15.3879%	4.276%	26.931%
Simple EP before break	2.0507%	2.8233%	3.3669%	4.8103%	7.9919%	0.081%	
Mean after break	1.5046%	1.2945%	1.1573%	0.6184%	0.8425%	1.965%	6.199%
Std dev after break	2.5042%	2.7427%	3.3446%	4.5549%	5.6968%	1.085%	17.276%
Simple EP after break	-0.0573%	-0.0530%	-0.1062%	-0.0285%	-0.6903%	-0.053%	

As table 1, except in this case the data is truncated in 1986. This is to avoid the difficulties that arise from the sharp growth in income for the highest income percentiles in 1987–8.

References

Avery, R. B. & Elliehausen, G. E. (1986), 'Financial characteristics of high-income families', *Federal Reserve Bulletin* **March**, 163–177.

Bansal, R. & Lundblad, C. (2002), 'Market efficiency, asset returns and the size of the risk premium in global equity markets', *Journal of Econometrics* **109**, 195–237.

Barsky, R. B. & De Long, J. B. (1993), 'Why does the stock market fluctuate?', *Quarterly Journal of Economics* **108**, 291–311.

Blanchard, O. J. (1993), 'Movements in the equity premium', *Brookings Papers on Economic Activity* **2**, 75–118.

Blume, M. E., Crockett, J. & Friend, I. (1974), 'Stockownership in the United States: Characteristics and trends', *Survey of Current Business* **November**, 16–40.

Brav, A., Constantinides, G. M. & Geczy, C. C. (2002), 'Asset pricing with heterogeneous consumers and limited participation: empirical evidence', *Journal of Political Economy* **110**, 793–824.

Chen, J. & Gupta, A. (1997), 'Testing and locating variance changepoints with application to stock prices', *Journal of the American Statistical Association* **92**, 739 – 747.

Claus, J. & Thomas, J. (2001), 'Equity premiua as low as three percent? Evidence from analysts' earnings forecasts for domestic and international stock markets', *Journal of Finance* **56**, 1629–1666.

Dittmar, A. K. (2000), 'Why do firms repurchase stocks?', *Journal of Business* **73**, 331–355.

Fama, E. F. & French, K. R. (2002), 'The equity premium', *Journal of Finance* **57**, 637–659.

Freeman, M. C. (2003), 'Can market incompleteness resolve asset pricing puzzles?', *Journal of Business Finance & Accounting* . Forthcoming.

Grullon, G. & Michaely, R. (2002), 'Dividends, share repurchases and the substitution hypothesis', *Journal of Finance* **57**, 1649–1684.

Hamilton, J. D. (1989), 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica* **57**, 357–84.

Heaton, J. & Lucas, D. (1996), 'Evaluating the effects of incomplete markets on risk sharing and asset pricing', *Journal of Political Economy* **104**, 443–87.

Hickman, W. B. (1960), *Statistical measures of corporate bond financing since 1900*, Princeton University Press.

Jagannathan, R., McGratten, E. R. & Scherbina, A. (2000), 'The declining US equity premium', *Quarterly Review* **Fall**, 3–16. Federal Reserve Bank of Minneapolis.

Kim, C., Morley, J. C. & Nelson, C. (2003), The structural break in the equity premium. Working paper: Washington University.

Kuznets, S. (1953), *Shares of upper income groups in income and savings*, National Bureau of Economic Research.

Macaulay, F. R. (1938), *Some theoretical problems suggested by the movement of interest rates, bond yields and stock prices in the United States since 1856*, Reprinted: Risk Classics Library.

Mankiw, N. G., Romer, D. & Shapiro, M. D. (1985), 'An unbiased reexamination of stock market volatility', *Journal of Finance* **40**, 677–687.

Mankiw, N. G. & Zeldes, S. P. (1991), 'The consumption of stockholders and nonstockholders', *Journal of Financial Economics* **29**, 97–112.

Mehra, R. (2003), 'The equity premium: why is it a puzzle?', *Financial Analysts Journal* **59**, 54–69. Jan/Feb.

Mehra, R. & Prescott, E. C. (1985), 'The equity premium: a puzzle', *Journal of Monetary Economics* **15**, 145–61.

Moore, B. & Schaller, H. (2002), 'Persistent and transitory shocks, learning and investment dynamics', *Journal of Money, Credit & Banking* **34**, 650 – 677.

Pagès, H. (1999), A note on the Gordon growth model with nonstationary dividend growth. Bank for International Settlements, Working Paper 75.

Pástor, L. & Stambaugh, R. F. (2001), 'The equity premium and structural breaks', *Journal of Finance* **56**, 1207–1239.

Piketty, T. & Saez, E. (2001), Income inequality in the United States, 1913–1998. NBER working paper 8467.

Piketty, T. & Saez, E. (2003), 'Income inequality in the United States, 1913–1998', *Quarterly Journal of Economics* **108**, 1–39.

Poterba, J. M. & Samwick, A. A. (1995), 'Stock ownership patterns, stock market fluctuations, and consumption', *Brookings Papers on Economic Activity* **2**, 295–372.

Smith, J. D. & Franklin, S. D. (1974), 'The concentration of personal wealth, 1922–1969', *American Economic Review* **64**, 162–167.

Vatter, H. G. (1963), *The U.S. economy in the 1950s; an economic history*, W. W. Norton.

Vissing-Jørgensen, A. (2002), 'Limited asset market participation and the elasticity of intertemporal substitution', *Journal of Political Economy* **110**, 825–853.

Wicker, E. (1996), *The banking panics of the Great Depression*, Cambridge University Press.

Wolff, E. N. (1994), 'Trendsin household wealth in the United States, 1962–83 and 1983–89', *The Review of Income and Wealth* **40**, 143–176.

Wolff, E. N. (2000), Recent trends in wealth ownership, 1983–1998. Jerome Levy Economics Institute Working Paper #300.