Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Financial Economics 73 (2004) 465–496
JOURNAL OF Financial ECONOMICS
www.elsevier.com/locate/econbase

Estimating the market risk premium ☆

E. Scott Mayfield[a,*]

[a] *Charles River Associates, Inc., 200 Clarendon Street, Boston, MA 02116, USA*

Received 15 March 2002; accepted 19 March 2002

Available online 15 June 2004

Abstract

This paper provides a method for estimating the market risk premium that accounts for shifts in investment opportunities by explicitly modeling the underlying process governing the level of market volatility. I find that approximately 50% of the measured risk premium is related to the risk of future changes in investment opportunities. Evidence of a structural shift in the underlying volatility process suggests that the simple historical average of excess market returns may substantially overstate the magnitude of the market risk premium for the period since the Great Depression.

JEL classification: G10; G31

Keywords: Risk premium; Asset pricing; Structural breaks; Markov-switching models

1. Introduction

The market risk premium is one of the most important numbers in finance. Unfortunately, estimating and understanding its value has proven difficult.

☆ This paper was written while the author was an assistant professor at the Harvard Business School. I am grateful to Malcolm Baker, George Chacko, Ben Esty, Bob Merton, Andre Perold, Tom Piper, Rick Ruback, and seminar participants at Boston College, Charles River Associates, the Harvard Business School, the University of Connecticut, and the 27th Annual Meeting of the European Finance Association for their helpful comments. The paper has benefited greatly from the thoughtful comments of an anonymous referee. Any remaining errors or omissions are my own. Financial support from the Harvard Business School Division of Research is gratefully acknowledged.

*Corresponding author. Tel.: +1-617-425-3335; fax: +1-617-425-3132.
E-mail address: smayfield@crai.com (E. Scott Mayfield).

0304-405X/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jfineco.2002.03.001

Although a substantial body of research shows that expected returns vary over time, the static approach of estimating the risk premium as the simple average of historical excess stock returns remains the most commonly employed method in practice.[1] Merton (1980) suggests estimating the risk premium based on the theoretical relationship between expected returns and the contemporaneous variance of returns. Although this theoretical approach is appealing, empirical research has failed to document a significant positive relationship between expected returns and the level of market volatility.[2] Scruggs (1998) provides evidence suggesting the failure to find a positive relationship between excess returns and market volatility may result from not controlling for shifts in investment opportunities. Lettau and Ludvigson (2001) make a similar point, showing that rejections of the consumption capital asset pricing model may also be due to a failure to control for shifts in investment opportunities. In this paper, I develop a method for estimating the market risk premium based on the equilibrium relationship between volatility and expected returns when there are discrete shifts in investment opportunities—specifically, changes in the level of market volatility. I use this method to demonstrate the importance of accounting for the dynamic nature of market risk when estimating the risk premium from ex post market returns.

The volatility of market returns during the past century has varied significantly. Schwert (1989a, b) studies historical variations in market volatility and relates the fluctuations to changes in economic and financial market conditions. My results suggest that, as a result of changes in the level of market volatility, the simple historical average of excess market returns obscures significant variation in the market risk premium and that over half of the measured risk premium is associated with the risk of future changes in investment opportunities. My analysis also suggests that, as a result of a structural shift in the likelihood of future high-volatility periods, the simple historical average of excess market returns may substantially overstate the magnitude of the market risk premium for the period since the Great Depression.

In my model, market risk is characterized by periodic episodes of high market volatility followed by a return to a lower, more typical level. I assume that the evolution of these volatility states follows a Markov process, and I model the market risk premium as a function of the underlying process governing the evolution of the two volatility states.[3] The expression for the equilibrium risk premium in my model is a special case of the Merton (1973) intertemporal capital asset pricing model. Because individuals anticipate future changes in the volatility state and corresponding

[1] For examples of research showing that expected returns vary over time, see Fama and Schwert (1977), Shiller (1984), Campbell and Shiller (1988), Fama and French (1988, 1989), Campbell (1991), Hodrick (1992), and Lamont (1998). Bruner et al. (1998) survey a sample of 27 "highly regarded corporations" and find that the estimates of the risk premium are generally based on either the arithmetic or geometric average of historical excess market returns.

[2] See Campbell (1987), French et al. (1987), Baillie and DeGennaro (1990), Glosten et al. (1993).

[3] Many researchers, including Schwert (1989a), Turner et al. (1989), Cecchetti et al. (1990), Pagan and Schwert (1990), Hamilton and Susmel (1994), Hamilton and Lin (1996), Schaller and Van Norden (1997), and Kim et al. (2000) have used a two-state Markov-switching model to describe the time series properties of market returns.

changes in the level of stock prices, ex post measured returns are not equal to ex ante expected returns.[4] When individuals place a nonzero probability on the likelihood of a future change in volatility state, expected returns include the expected change in stock prices associated with a change in volatility state. While the economy remains in the low-volatility state, actual ex post returns are higher on average than expected returns. Conversely, while the economy remains in the high-volatility state, actual ex post returns will be lower on average than expected returns. Within each state, the difference between ex post returns and expected returns is similar to the peso-type problem discussed in Rietz (1988). My model generates periods of low-volatility and high ex post returns alternating with periods of high-volatility and low ex post returns, reconciling the empirical finding that returns are lower in periods of high volatility with the theoretical intuition that expected returns should be positively related to the level of market volatility.

My theoretical model maps directly into a standard empirical framework for estimating time variation in market volatility, providing a foundation for interpreting these earlier empirical results and a structural basis for estimating the market risk premium in a dynamic setting. Given the Markov structure of my model, its parameters can be estimated using the Hamilton (1989) Markov-switching model. Consistent with previous studies that use the Markov-switching model to describe the time series properties of stock market returns, my analysis shows that market returns can be described as having been drawn from two significantly different distributions: a low-volatility/high-return distribution, from which about 88% of the returns are drawn, and a high-volatility/low-return distribution, from which about 12% of the returns are drawn. In the low-volatility state, the annual standard deviation of returns is 13.0% and the mean annualized excess return is 12.4%. In contrast, the annual standard deviation of returns in the high-volatility state is 38.2% and the mean annualized excess return is -17.9%.[5]

My equilibrium expression for the risk premium allows the estimated moments of the two conditional return distributions to be mapped directly to preference parameters. Using this mapping, I decompose the unconditional risk premium into two state-dependent risk premia as well as into premia required for intrastate diffusion risk and interstate jump risk. My estimates for the annualized state-dependent risk premia in the low- and high-volatility states are 5.2% and 32.5%, respectively. Based on the estimated preference parameters, my analysis suggests that about 50% of the unconditional risk premium is related to the risk of future changes in the level of market volatility.

[4] The negative relationship between volatility and market prices, referred to as volatility-feedback, is examined in Malkiel (1979), Pindyck (1984), Poterba and Summers (1986), French et al. (1987), Campbell and Hentschel (1992), and Kim et al. (2000).

[5] When transitional months associated with changes in volatility states are excluded, the estimated standard deviation of returns in each volatility state remains essentially unchanged. The empirical method for identifying changes in volatility states tends to treat the jumps in stock prices associated with changes in volatility states as high-volatility returns, and the magnitude of the stock price changes during transitional months is comparable to the standard deviation of returns within the identified high-volatility periods.

Recent studies provide historical evidence of a structural shift in the market risk premium. Siegel (1992) documents that the market premium has not been constant over the past century and that excess stock returns during the mid-1900s are abnormally large. Pastor and Stambaugh (2001) use a Bayesian analysis to test for structural breaks in the distribution of historical returns and to relate those breaks to changes in the market risk premium. Fama and French (2002) provide evidence of a structural shift in the market risk premium by comparing the ex ante risk premium from a Gordon growth model with the ex post risk premium based on the historical average of excess market returns. Evidence of a structural shift in the volatility of market returns is also provided in earlier studies. Officer (1973) and Schwert (1989b) argue that market returns during the Great Depression era were unusually volatile, and Pagan and Schwert (1990) show that the volatility of market returns during the Great Depression was inconsistent with stationary models of conditional heteroskedastic returns. My model provides a structural basis for estimating the impact of such a structural shift on the market risk premium. Consistent with Pagan and Schwert (1990) and Pastor and Stambaugh (2001), I find evidence of a statistically significant shift in the underlying volatility process that governs the evolution of volatility states following the 1930s. Because of the structural shift in the Markov transition probabilities, the likelihood of entering into the high-volatility state falls from about 39% before 1940 to less than 5% after 1940. Given the lower likelihood of entering the high-volatility state, the risk premium falls from about 20.1% before 1940 to 7.1% after 1940.

Because of the structural shift in the underlying volatility process and the associated reduction in the market risk premium, ex post returns during the period following 1940 are not an unbiased estimate of ex ante expected returns. As investors learn that market risk has fallen because of the structural shift, stock prices will be bid up and ex post returns will be greater than ex ante expected returns. Elton (1999) stresses the importance of distinguishing between ex ante and ex post returns when average realized returns are used as a proxy for ex ante expected returns. Brown et al. (1995) make a related point, arguing that economies that survive ex post must have higher returns on average than the ex ante expected return of all economies. When I correct for this potential bias in my sample of ex post realized returns, my estimate of the market risk premium for the period after 1940 is 5.6%, suggesting that the simple historical average of excess market returns may substantially overstate the magnitude of the risk premium for the period since the Great Depression.

The remainder of the paper is structured as follows. Section 2 presents the analytical model of the risk premium with discrete volatility states. Section 3 describes the empirical framework used to identify and estimate the parameters of the model and reports the resulting decomposition of the unconditional risk premium. In Section 4, I test for a structural shift in the process governing the evolution of volatility states and show the impact on the market risk premium of such a shift. Section 5 summarizes the main findings of the paper.

2. A two-state model of the market risk premium

My analysis begins with the assumption that the variance of market returns follows a two-state Markov process. Defining $s_t \in (L, H)$ to represent the state of the economy at time t, the variance of returns at each instant is given by the equation

$$\sigma_t^2 = \begin{cases} \sigma_L^2, & \text{if } s_t = L, \\ \sigma_H^2, & \text{if } s_t = H, \end{cases} \tag{1}$$

where σ_L^2 is the variance of returns in the normal low-volatility state and σ_H^2 is the variance of returns in the abnormal high-volatility state. To focus on the risk of future changes in market volatility, I assume that investors know the current volatility state with certainty but face the possibility of a change in the volatility state at each point in time.[6] Because the variance process is Markov, the probability of a change in market volatility is a function of the current state only, such that

$$\pi_t = \begin{cases} \pi_L, & \text{if } s_t = L, \\ \pi_H, & \text{if } s_t = H. \end{cases} \tag{2}$$

In this environment, the risk premium must compensate investors for the current volatility of market returns as well as the risk associated with a change in volatility state.

I derive the expression for the equilibrium risk premium in a continuous-time, representative agent model in which preferences are described by power utility. The mathematical derivation of the equilibrium risk premium is provided in the appendix.[7] The equilibrium risk premium is given by the expression

$$\mathrm{E}[R_t] - R_t^{\mathrm{f}} = \gamma\sigma_t^2 + \pi_t J_t\left[1 - (1 + K_t^*)^{-\gamma}\right], \tag{3}$$

where $\mathrm{E}[R_t]$ is the expected return on the market at time t, R_t^{f} is the contemporaneous risk-free rate of return, γ is the coefficient of relative risk aversion, π_t is the instantaneous probability of a change in volatility state, J_t is the percentage change in wealth associated with a change in volatility state, and K_t^* is the percentage change in the optimal level of consumption resulting from a change in volatility state. Using Eq. (3), I decompose the risk premium into two components. The first term on the right-hand side of Eq. (3) is the component that accounts for current volatility risk, which I refer to as the intrastate risk premium. The second term is the component that accounts for changes in the level of market volatility, which I refer to as the interstate risk premium. Because there are only two volatility states, no uncertainty exists over the magnitude of the future change in volatility. Instead, uncertainty exists only over the time at which the level of volatility will change. Eq. (3) is a special case of Merton's (1973) intertemporal capital asset pricing

[6] Turner et al. (1989) study the inference problem faced by investors when the current state is not known and must, instead, be learned. My model is more in the spirit of the Merton (1980) model, in which agents have access to continuous return data over a discrete interval of time such that they are able to estimate the variance of the underlying data generating process to any degree of precision required.

[7] George Chacko provided helpful insights for formulating the state-dependent structure of the programming problem.

model in which changes in investment opportunities are restricted to unpredictable, state-dependent changes in the level of market volatility.[8]

In my formulation of the investor's problem, I allow for constraints on consumption that may limit the degree to which individuals are able to adjust their consumption when the economy switches volatility state. In the appendix, I show that the interstate component of the risk premium is a function of the optimal change in the level of consumption associated with the change in volatility state, even when the ability of investors to adjust their consumption is constrained. The intuition behind this result is that, around the optimum, the loss in utility from being constrained away from the optimum is equal to the loss in utility associated with the optimal change in consumption resulting from a change in volatility state. Assuming that the constraint binds only in the high-volatility state, the distortion in consumption is summarized by the value of the Lagrange multiplier λ_H and is given by the expression

$$\lambda_H = 1 - \left(\frac{1 + K_L^*}{1 + \tilde{K}_L}\right), \tag{4}$$

where $\tilde{K}_L$ is the actual change in consumption associated with a switch to the high-volatility state. Using Eq. (4) and the estimated value of K_L^*, the value of the Lagrange multiplier λ_H can be inferred from the actual change in consumption $\tilde{K}_L$ observed during periods when the economy enters the high-volatility state.

Because volatility levels are discrete, wealth and optimal consumption levels change in a discontinuous fashion when the economy changes state. However, given that there are only two volatility states, the wealth and consumption effects of a change in state are negated after every two changes in state, such that

$$W_t'' = (1 + J_t')(1 + J_t)W_t = W_t \tag{5}$$

and

$$C_t^{*''} = (1 + K_t^{*'})(1 + K_t^*)C_t^* = C_t^*, \tag{6}$$

where W_t'' and $C_t^{*''}$ are the wealth and optimal consumption levels after two state changes and J_t' and $K_t^{*'}$ are the changes in wealth and optimal consumption associated with switching out of the alternate volatility state. For this reason, the change in the levels of wealth and optimal consumption associated with the alternate volatility state can be written in terms of the changes associated with the current volatility state, such that

$$J_t' = \frac{1}{1 + J_t} - 1 \tag{7}$$

and

$$K_t^{*'} = \frac{1}{1 + K_t^*} - 1. \tag{8}$$

[8] Schwert (1989a, b) documents that changes in market volatility are correlated with changes in economic and financial market conditions.

From Eqs. (7) and (8), the magnitude of the jumps in wealth and consumption associated with changes in state are summarized by the two parameters J_t and K_t^*.

The percentage change in the optimal level of consumption K_t^* is determined by the change in the optimal consumption–wealth ratio together with the percentage change in wealth associated with a change in state J_t. The equilibrium consumption–wealth ratio in each state is given by the expression

$$\frac{C_t^*}{W_t} = \frac{\rho + (\gamma - 1)\mu_t - \frac{1}{2}\gamma(\gamma - 1)\sigma_t^2}{\gamma} + \frac{\pi_t}{\gamma}\left[1 - \left(\frac{1 + J_t}{1 + K_t^*}\right)^{\gamma}\right], \tag{9}$$

where C_t^* is optimal consumption at time t, W_t is wealth at time t, ρ is the investor's subjective discount rate, and μ_t is the expected return conditional on remaining in the current state. Consistent with my terminology for the two components of the risk premium, I refer to μ_t as the expected intrastate return. Because the optimal consumption–wealth ratio is itself a nonlinear function of K_t^*, when the model parameters are estimated, I solve numerically for the value of K_t^* that solves Eq. (9). In the appendix, I show that Eq. (9) collapses to the formula for the consumption–wealth ratio derived in Merton (1969) for the infinite horizon lifetime portfolio selection problem under uncertainty when a single volatility state is assumed.

Because wealth changes when the economy changes state, the expected return on the market is not equal to the expected intrastate return. The expected return on the market is given by the equation

$$\mathrm{E}[R_t] = \mu_t + \pi_t J_t. \tag{10}$$

When the economy is in the low-volatility state, investors expect a reduction in wealth when the economy enters the high-volatility state. For this reason, in the low-volatility state, the expected return on the market is less than the expected intrastate return. Similarly, when the economy is in the high-volatility state, investors expect an increase in wealth when the economy reenters the low-volatility state and the expected return on the market is greater than the expected intrastate return.

Fig. 1 depicts the distinction between state-dependent risk premia and expected intrastate excess returns. For each state, the slope of the line labeled "Expected market return" shows required returns and the slope of the line labeled "Expected intrastate return" shows expected returns conditional on the economy remaining in the current state. The vertical line segments at the boundary of low- and high-volatility states represent the jump in wealth associated with a change in volatility state. The figure is drawn such that expected intrastate returns are constant while required returns vary with changes in volatility state. Because of expected changes in wealth associated with changes in volatility state, expected intrastate returns vary by less than state-dependent expected returns. In the low-volatility state, expected intrastate returns are greater than required returns, and in the high-volatility state, expected intrastate returns are less than required returns. If the expected increase in wealth associated with a return to the low-volatility state is sufficiently large, then expected intrastate returns in the high-volatility state can be negative even though the risk premium is positive. My model provides a plausible explanation for reconciling the empirical observation that returns are lower in periods of high

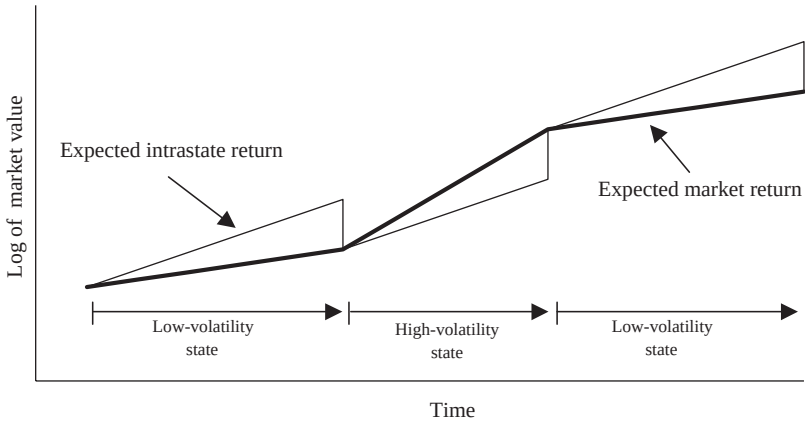

Fig. 1. Expected return on the market versus expected intrastate returns. The vertical axis depicts the log of market value and the horizontal axis represents time. The economy is initially in the low-volatility state, switches into the high-volatility state, and returns to the low-volatility state. The slope of the bold line labeled "Expected market return" is equal to the required return in each volatility state. The slope of the thin line labeled "Expected intrastate return" is equal to the expected return conditional on the economy remaining in each state. The vertical line segment at the boundary of low- and high-volatility states represents the jump in wealth associated with a change in state.

volatility with the theoretical intuition that expected returns should be positively related to the level of market volatility.

3. Model estimation

This section presents the results from estimating the theoretical model.

3.1. Data

The model described in Section 2 is estimated using data from the Center for Research in Security Prices (CRSP). I use monthly value-weighted returns including dividends for NYSE, Amex, and Nasdaq stocks (VWRETD) over the period from 1926 through 2000 as my proxy for market returns. Excess returns are calculated using the contemporaneous yield on one-month Treasury bills from the risk-free rate file provided with the CRSP government bond data.

Table 1 reports summary statistics for monthly excess returns. The average annualized excess return over the sample period is 8.3%, and the annualized standard deviation of returns is 19.0%. The largest and smallest one-month returns are 38.2% and −29.0%, respectively. The reported skewness measure is negative and statistically significant, indicating that large negative returns are more frequent than large positive returns. Finally, the reported measure of excess kurtosis indicates that large returns occur more frequently than would be the case if returns were normally

Table 1
Summary statistics for monthly excess returns, 1926–2000
Excess returns are constructed as the monthly value-weighted return including dividends for NYSE, Amex, and Nasdaq stocks in excess of the contemporaneous yield on one-month Treasury bills. Data were obtained from the Center for Research in Security Prices stock and government bond files. The first column reports the sample statistics, and the second column shows the associated p-value for a test that the true value of the statistic equals zero.

Statistic	Estimate	p-value
Mean (annualized)	8.3%	0.0039
Standard deviation (annualized)	19.0%	
Maximum	38.2%	
Minimum	−29.0%	
Skewness (ln returns)	−0.512	<0.0001
Excess kurtosis (ln returns)	7.043	<0.0001
Number of observations	900	

distributed. As Fama (1965) points out, time variation in market volatility will produce excess kurtosis in stock returns.

3.2. Methodology

To estimate the components of the market risk premium in each volatility state, I map the fundamental parameters of the model to the expected intrastate excess returns by combining Eqs. (3) and (10). This yields the expression

$$\mu_t - R_t^{\mathrm{f}} = \gamma\sigma_t^2 - \pi_t J_t(1 + K_t^*)^{-\gamma}. \tag{11}$$

Because the model is estimated using holding-period returns, the instantaneous transition probabilities π_t are converted to their discrete time equivalents. To do this, I write the instantaneous expected change in wealth associated with a change in volatility state in terms of the equivalent holding-period expected change in wealth, such that

$$\pi_t J_t = \pi_t' \ln(1 + J_t), \tag{12}$$

where π_t' is the discrete time transition probability. Eq. (12) requires that, over the expected duration of each volatility state, the continuously compounded expected change in wealth is equal to the actual change in wealth associated with a change in state.[9] Combining Eqs. (11) and (12) yields

$$\mu_t - R_t^{\mathrm{f}} = \gamma\sigma_t^2 - \pi_t' \ln(1 + J_t)(1 + K_t^*)^{-\gamma}. \tag{13}$$

Eq. (13) is the basis for my estimation method, which has three steps. In the first step, I use the Hamilton (1989) Markov-switching model to estimate the moments of the two state-dependent return distributions μ_t and σ_t as well as the transition

[9] The mathematical derivation of Eq. (11) comes from the requirement that $e^{(\pi_t J_t)D_t} - 1 = J_t$, where the expected duration of each volatility state D_t is given by the formula $D_t = 1/\pi_t'$.

probabilities π'_t that govern the dynamics of the underlying volatility process. In the second step, I use Eq. (13) together with Eqs. (7)–(9) to find the corresponding values of γ, J_t, and K_t^* that are consistent with the estimated moments of the two state-dependent return distributions.[10] Because there are only two free parameters, γ and J_L, available to match the two state-dependent means, μ_L and μ_H, the model is exactly identified. In the third step, I use the expression for the risk premium given by Eq. (3) together with the estimated model parameters to calculate the intrastate and interstate components of the risk premium in each volatility state.

3.3. Results

Table 2 presents the empirical results from my three-step method. Panel A provides the results from applying the Markov-switching model to my sample of returns. I assume that each monthly return is drawn from one of two state-dependent distributions and that returns are log-normally distributed in each state. Parameter estimates are obtained via maximum likelihood using the method described in Berndt et al. (1974). Standard errors are reported in parentheses. Panel B reports the estimated values of the preference parameters γ, J_t, and K_t^* that are consistent with the estimated time series model presented in Panel A. Finally, Panel C reports the implied decomposition of the market risk premium. Because of the nonlinear nature of the model, the standard errors of the coefficients reported in Panels B and C are simulated based on 500 random draws of the time series model parameters from a multivariate normal distribution with mean-vector and variance-covariance matrix equal to those reported in Panel A.

Panel A reports the time series model parameter estimates. The return distributions in the two volatility states are significantly different. The estimated annualized standard deviation of returns varies from 13.0% in the low-volatility state to approximately 38.2% in the high-volatility state. The annualized mean return in the low-volatility state is 12.4% and is significantly different from zero. The annualized mean return in the high-volatility state is −17.9% but is not significantly different from zero. The two volatility states are persistent. The point estimates of the transition probabilities π'_L and π'_H indicate a 0.017 and 0.119 probability of switching out of the low- and high-volatility states, respectively. Both estimated transition probabilities are significantly less than 0.5, indicating that both volatility states tend to persist over time. Based on the estimated transition probabilities, the expected durations of the low- and high-volatility states are approximately 59.2 and 8.4 months, respectively. These results are consistent with previous studies that use the Markov-switching model to describe the time series properties of returns, including Schwert (1989a), Turner et al. (1989), Pagan and Schwert (1990), and Schaller and Van Norden (1997).

[10] Eq. (9) also requires that the subjective discount rate ρ be specified. I set the value of ρ equal to the value estimated in Campbell and Cochrane (1999) of 0.1165. I also test a variety of alternative values for ρ and find that my results are not sensitive to the specific value of ρ chosen.

Table 2
Parameter estimates and implied risk premium decomposition
Estimates are based on 900 monthly excess returns from January 1926 through December 2000. Panel A reports the parameter estimates for the two-state Markov switching model based on Eq. (13). Panel B reports the estimated values of the preference parameters γ, J_t, and K_t^* from Eqs. (7)–(9) that are consistent with the estimated time series model. Panel C shows the implied decomposition of the market risk premium based on Equation (3) and the estimated model parameters. Standard errors are reported in parentheses. Because of the nonlinear nature of the model, the standard errors reported in Panels B and C are simulated.

Volatility state	Time series parameters			Preference parameters			Risk premium decomposition			
							State probability	State-dependent premium		
	$\mu_t - r_t$	σ_t	π_t	γ	J_t	K_t^*		Intrastate	Interstate	Total
	Panel A			*Panel B*				*Panel C*		
Low volatility	0.124	0.130	0.017	*1.129*	*−0.296*	−0.2488	0.876	0.019	0.033	0.052
($s_t = L$)	(0.017)	(0.004)	(0.007)	(0.565)	(0.088)	(0.108)	(0.037)	(0.010)	(0.011)	(0.016)
High volatility	−0.179	0.382	0.119	1.129	0.421	0.404	0.124	0.165	0.160	0.325
($s_t = H$)	(0.140)	(0.022)	(0.038)	(0.565)	(0.218)	(0.289)	(0.037)	(0.078)	(0.077)	(0.116)
Unconditional mean	0.086							0.037	0.049	0.086
	(0.023)							(0.018)	(0.016)	(0.023)
Log-likelihood value	1,491.0									
Number of observations	900									

Panel B reports the preference parameter estimates. The estimated values of the two free parameters γ and J_L are presented in italics. The other parameters are simultaneously determined using Eqs. (7)–(9) but are not independently estimated. The point estimate for γ equals 1.129 and is significantly different from zero at the 5% level based on a one-tailed test. The point estimate for the jump parameter J_L equals -29.6% and is significantly different from zero. The corresponding value of J_H is 42.1%. The implied values for the optimal percent change in consumption K_t^* in the low- and high-volatility states are -28.8% and 40.4%, respectively. Although the estimate of K_t^* for the low-volatility state is significant, given the high volatility of returns in the high-volatility state, the estimate of K_t^* for the high-volatility state is not significantly different from zero.

Panel C reports the implied decomposition of the market risk premium. The first column of the Panel C reports the unconditional probability of each volatility state based on the estimated transition probabilities presented in Panel A. The second and third columns of Panel C show the intrastate and interstate components of the two state-dependent risk premia. The fourth column of Panel C reports the state-dependent risk premium for each volatility state. For each component of the risk premium, the unconditional estimate is calculated as the probability weighted average of the two state-dependent estimates. The estimated values of the unconditional components of the risk premia are reported in the fourth row of the panel. Based on the estimated transition probabilities, the unconditional probability of the economy being in the low- and high-volatility states is 0.876 and 0.124, respectively. The point estimate of the risk premium in the low-volatility state is 5.2%. About 330 basis points, or 64% of the low-volatility state risk premium, are associated with the risk of a change in state. The point estimate of the risk premium in the high-volatility state is 32.5%. About 1,600 basis points, or 49% of the high-volatility state risk premium, are associated with the risk of a change in state. The unconditional risk premium is equal to 8.6%. About 490 basis points, or 57% of the unconditional risk premium, are associated with the risk of changes in state. These results suggest that more than half of the measured market risk premium is related to the risk of future changes in the level of market volatility.

3.4. Statistical tests

I perform a series of statistical tests of the estimated model reported in Table 2. My statistical analysis is presented in two parts: tests of the time series model and tests of the theoretical model. In my analysis of the time series model, I test whether the two volatility states are statistically different as well as whether the assumption of only two volatility states is reasonable. I also test the assumption that returns are independently, log-normally distributed within each state. In my analysis of the theoretical model, I use the low- and high-volatility episodes identified in the time series analysis to test the predictions of the theoretical model, including the statistical properties of returns in each identified state and the extent to which market prices jump when the economy switches between states.

The two volatility states are statistically different. I test the estimated model against the null hypothesis that both the mean and variance of returns is constant.

The likelihood ratio statistic for the test is 155.4 and the corresponding p-value is less than 0.0001, indicating that the null hypothesis can be rejected at any reasonable level of confidence. I also test the extent to which the explanatory power of the model is improved by the inclusion of a third volatility state. Although the inclusion of a third state increases the value of the estimated likelihood function, the increase is not statistically significant. The likelihood ratio statistic for a test of three states against a null hypothesis of two states is 8.82. The corresponding p-value of 0.1816 indicates that the null hypothesis of two states cannot be rejected at standard levels of significance.

The assumption that returns are independent within each volatility state is reasonable. I augment the time series model to allow for first-order serial correlation in returns within each volatility state. The point estimates for the serial correlation coefficients in the low- and high-volatility states are 0.28 and 1.26, respectively. Neither estimated coefficient is statistically significant. The likelihood ratio statistic for a test of the null hypothesis that both coefficients are zero is 0.82 and the corresponding p-value is 0.9915, indicating that the null hypothesis cannot be rejected at any reasonable level of confidence.

The assumption that returns are log-normally distributed within each volatility state is reasonable. Fig. 2 compares the cumulative distribution function (CDF) for the estimated model with the sampled cumulative distribution of returns. I also show the CDF for the assumption that the data are unconditionally log-normal. The top panel of the figure shows each of the cumulative distribution functions, and the bottom panel shows the difference between the estimated and sampled CDFs. To assess the reasonableness of the distributional assumptions, I perform a Kolmogorov-Smirnov test of the difference between the estimated and sample distributions.[11] Consistent with the two volatility states being statistically different, the null hypothesis that the data are unconditionally log-normal can be rejected at the 1% level. In contrast, the null hypothesis that the data are log-normally distributed within each volatility state cannot be rejected at the 5% level.

The results of these statistical tests of the estimated time series model suggest that a simple two-state model provides a reasonable description of monthly market returns. Based on the high-volatility periods identified by the two-state time series model, I perform statistical tests of the main predictions from the theoretical model. I define high-volatility periods as those months for which the implied probability of being in the high-volatility state is greater than 0.5. Based on this criteria, there are 21 high-volatility periods during the period from 1926 through 2000. Of the 900 months in the sample, 804 months are categorized as low volatility and 96 months are categorized as high volatility. Descriptive statistics for these low- and high-volatility periods are provided in Table 3.

[11] The Kolmogorov-Smirnov (K-S) statistic for a test of the null hypothesis that the data are unconditionally log-normal is 0.0708. The critical value of the K-S statistic for a 1% test with 900 observations is 0.0543, indicating that the null hypothesis can be rejected. In contrast, the K-S statistic for a test of the null hypothesis that the data are log-normally distributed within each volatility state is 0.0211. The critical value of the K-S statistic for a 5% test with 900 observations is 0.0453, indicating that the null hypothesis cannot be rejected.

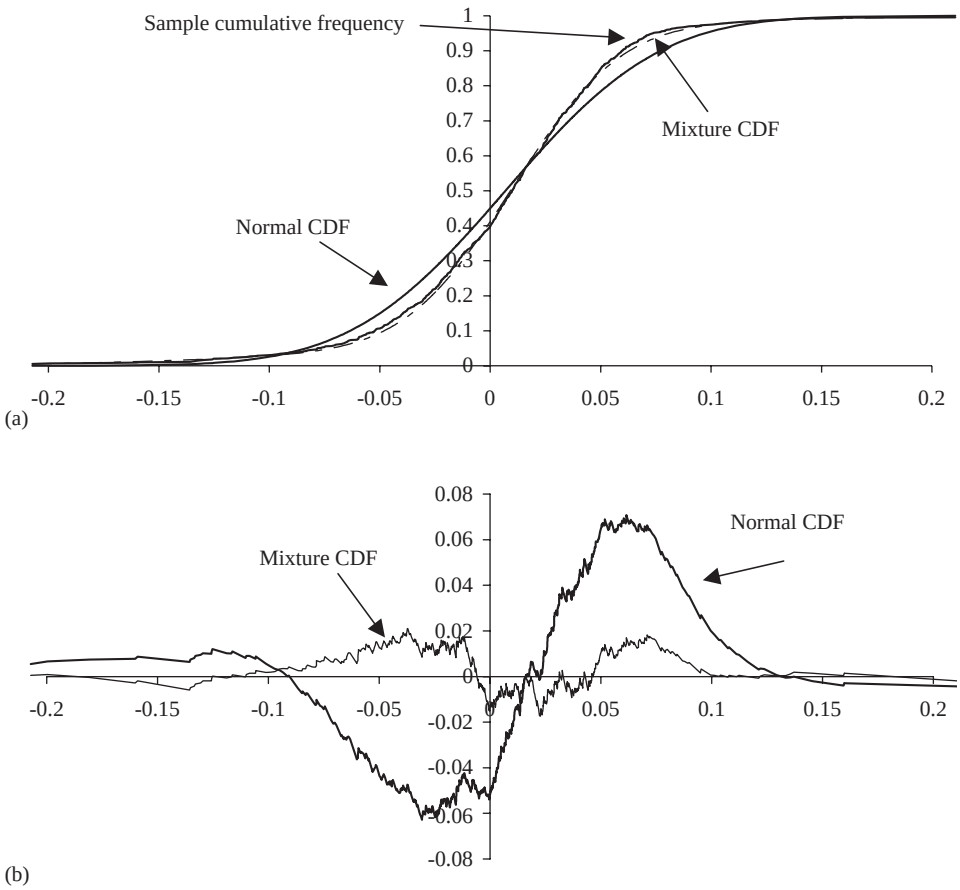

Fig. 2. Sample cumulative frequency distribution versus cumulative distribution functions (CDFs) for estimated mixture distribution and normal models. The mixture distribution is the implied distribution from the estimated two-state model presented in Table 2. The normal distribution is for the comparable static model with constant mean and variance. Panel A shows the cumulative distribution functions, and Panel B shows the corresponding errors between the actual and predicted CDFs.

The top panel of Table 3 groups returns into four categories: the first month of high-volatility periods, subsequent high-volatility months, the first month of low-volatility periods, and subsequent low-volatility months. For each category, I report the mean excess return and the associated *p*-value for a test of the null hypothesis that the true mean is zero. In addition, I report the standard deviation of returns, the average probability of being in the high-volatility state, and the number of observations for each category. The bottom panel of the table reports the results of hypothesis tests related to the predictions of the theoretical model.

Market returns are substantially more volatile during the identified high-volatility periods than low-volatility periods. Excluding the first month of each episode, the

Table 3
Statistical tests of categorized excess returns
Each monthly excess return is categorized as having been from one of two major categories: low- and high-volatility periods. A high-volatility period is defined as a continuous series of months for which the inferred probability of being in the high-volatility state is greater than 0.5. All other months are categorized as low volatility. Over the historical period, 21 high-volatility periods are identified. To test the predictions from the theoretical model regarding the transition between volatility states, returns are further categorized as having been from the first month or subsequent months of either a low- or high-volatility period. The top panel reports descriptive statistics for each category, and the bottom panel reports the results of a series of hypothesis tests.

Category	Monthly returns				
	Mean	*p*-value	Standard deviation	$Pr(s_t = H)$	*N* obs
Categorized returns					
All months	0.0069	0.0002	0.0549	0.1300	900
High-volatility periods					
First month	−0.1262	0.0000	0.0707	0.8844	21
Subsequent months	0.0114	0.4164	0.1212	0.8485	75
Low-volatility periods					
First month	0.0221	0.0004	0.0246	0.3694	22
Subsequent months	0.0096	0.0000	0.0379	0.0346	782
Hypothesis tests[a]	*t*-statistic	*p*-value			
First month of high-volatility periods equal to subsequent months of high-volatility periods	6.6075	<0.0001			
First month of low-volatility periods equal to subsequent months of low-volatility periods	2.3113	0.0301			
First month of high-volatility periods (ln returns) equal to negative of first month of low-volatility periods (ln returns)	6.5194	<0.0001			
Subsequent months of high-volatility periods equal to subsequent months of low-volatility periods	0.1295	0.8973			

[a] Based on the Smith-Satterhwaite test for difference in population means with unequal variances, Miller and Freund (1977).

annualized standard deviation of returns during the identified low- and high-volatility periods is 13.1% and 42.0%, respectively. Although the level of volatility in the two states is significantly different, the average excess return is not. Excluding the first month of each episode, the annualized average excess return during low- and high-volatility episodes is 13.7% and 11.5%, respectively. The *p*-value for a test of the null hypothesis that average excess returns in the low- and high-volatility periods are equal is 0.8973, indicating that the null hypothesis cannot be rejected at any reasonable level of

confidence. This result is consistent with the time path of expected returns depicted by Fig. 1 in the theoretical discussion of the model. In addition, returns during the transition between volatility states are also generally consistent with those depicted in Fig. 1.

The average first month of low- and high-volatility episodes is significantly different from subsequent months. High-volatility periods start with a substantial loss in market value. The average excess return during the first month of the high-volatility periods equals −12.6% and is significantly different from zero. In contrast, the average excess return during subsequent high-volatility months is positive 1.1% but is not significantly different from zero. The *p*-value for a test of the null hypothesis that the mean of the first month of high-volatility periods equals the mean of subsequent high-volatility months is less than 0.0001, indicating that the null hypothesis can be rejected at any reasonable level of confidence. Low-volatility periods start with a significant increase in market value. The average excess return during the first month of the low-volatility periods is 2.2% and is significantly different from zero. The average excess return during subsequent low-volatility months equals 0.96% and is also significantly different from zero. Although the difference between the first-month and subsequent months of low-volatility periods is less pronounced than that of high-volatility periods, the average return during the first month of each low-volatility period is more than twice that of subsequent months and the difference in the mean returns is statistically significant. The *p*-value for a test of the null hypothesis that the mean of the first month of low-volatility periods equals the mean of subsequent low-volatility months is 0.0301, indicating that the null hypothesis can be rejected at the 5% level.

One aspect of the theoretical model is not supported by the data. Because the theoretical model assumes that there are only two states and that investors always correctly know the current state, the magnitude of the jump in log market value when the economy switches from the low-volatility state to the high-volatility state equals the magnitude of the jump in log market value when the economy returns to the low-volatility state. Although the point estimates of the average excess monthly returns low- and high-volatility periods are of the correct sign, the magnitude of the loss in market value when the economy enters the high-volatility state is significantly greater than the magnitude of the increase in market value when the economy returns to the low-volatility state. The *p*-value for a test of the null hypothesis that the magnitude of the mean excess log return during the first month of high-volatility periods is equal to the magnitude of the mean excess log return during the first month of low-volatility periods is less than 0.0001, indicating that the null hypothesis can be rejected at any reasonable level of confidence.

One explanation for the difference in first-month returns is that investors do not have perfect knowledge of the current state and so they must infer the volatility state from the returns they observe.[12] In this case, investors' ability to infer the current state is asymmetric. When the economy is in the low-volatility state, the standard deviation of returns is small and determining whether the economy has switched to

[12] Turner et al. (1989) explicitly incorporate learning into a Markov-switching model in which investors are uncertain of the true state.

the high-volatility state is easy. Large returns are unlikely to occur in the low-volatility state, so their occurrence quickly reveals to investors that the economy is in the high-volatility state. However, the inference problem is more difficult when the economy is in the high-volatility state. In the high-volatility state, small returns do not immediately reveal that the economy has switched states because a reasonable chance of getting a small return exists even though the standard deviation of returns is high. Instead, investors learn that the economy has returned to the low-volatility state over time by failing to observe enough large returns—or, in other words, by observing more small returns than are likely to occur in the high-volatility state. When investors have to learn whether the economy has switched states, the increase in market value associated with a return to the low-volatility state likely will occur over a longer period of time than the decrease in market value associated with a switch to the high-volatility state. In addition to the assumption that investors have perfect knowledge of the true volatility state, another important issue regarding the estimated model presented in Table 3 is whether the process governing the evolution of volatility states is constant over the estimation period.

Fig. 3 plots the historical returns on which the model is estimated along with the identified high-volatility periods represented by the shaded areas. Visual inspection of the figure suggests that the average duration of high-volatility periods is shorter during the later part of the sample than during the first part. The average duration of high-volatility periods is 7.2 months for the period from 1926 to 1940 versus only 2.6 months for the period after 1940. In addition, the average duration of low-volatility periods appears longer during the later part of the sample than during the first part of the sample. The average duration of low-volatility periods is only 11.3 months for the period from 1926 to 1940 versus 58.4 months for the period after 1940. The

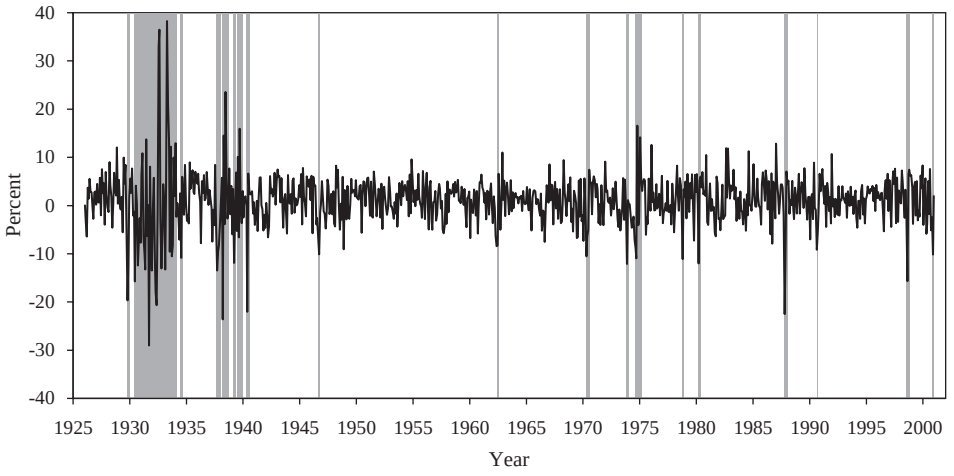

Fig. 3. Monthly excess returns and high-volatility state probability. The solid line plots the monthly excess returns for the period 1926 through 2000. The shaded areas correspond to the high-volatility episodes identified in Table 3. A high-volatility period is defined as a continuous series of months for which the inferred probability of being in the high-volatility state is greater than 0.5.

differences in the average durations of low- and high-volatility periods suggest that the transition probabilities governing the evolution of volatility states may not be constant over the historical period. A shift in the underlying volatility process is consistent with previous studies by Schwert (1989b), Pagan and Schwert (1990), and Pastor and Stambaugh (2001) that find evidence of structural shifts in the volatility of market returns. In my two-state model of the market risk premium, a shift in the transition probabilities governing the underlying volatility process would result in a change in the likelihood of the low- and high-volatility states and lead to a change in the unconditional market risk premium.

4. The effect of a structural shift in the volatility process

In this section of the paper, I augment the model to allow for a structural shift in the transition probabilities governing the evolution of the two volatility states. I assume there is a single structural break during the estimation period and test the estimated model against the null hypothesis of no structural break. To determine the most likely date for a structural shift in the volatility process, I estimate the augmented model for all possible annual breakpoints from 1927 through 1999 and select the breakpoint that maximizes the value of the estimated likelihood function. The analysis is then structured around the two subperiods defined by the most likely date for the structural shift in the volatility process.

Consistent with the approach presented in Section 3, the estimation method has three steps. In the first step, I estimate the time series model parameters allowing for a structural shift in the transition probabilities π_t and the means of the two state-dependent distributions μ_t.[13] I assume that the volatility of returns in each state remains constant over the estimation period. In the second step, I use Eq. (13) together with Eqs. (7)–(9) to find the corresponding values of γ, J_t, and K_t^* for each subperiod. I assume the value of γ is constant over the estimation period, but that the parameters J_t and K_t^* shift to correspond to the new transition probabilities. In the state-dependent model with a structural break, there are three free parameters, γ, $J_{L,\text{pre}}$, and $J_{L,\text{post}}$, available to match the four state-dependent means, $\mu_{L,\text{pre}}$, $\mu_{H,\text{pre}}$, $\mu_{L,\text{post}}$, and $\mu_{H,\text{post}}$. In contrast to the model presented in Section 3, the augmented model is no longer exactly identified. To find the values of the preference parameters that are consistent with the estimated moments of the two state-dependent distribution functions, I solve for the values of γ, $J_{L,\text{pre}}$, and $J_{L,\text{post}}$ that minimize the probability-weighted sum of the squared standardized errors over the entire estimation period. In the third step, I use the expression for the risk premium given by Eq. (3) together with the estimated model parameters to decompose the risk premium for each subperiod. These results are reported in Table 4.

[13] Diebold et al. (1994) discuss the estimation of time-varying transition probabilities in Markov-switching models.

Table 4
Parameter estimates and implied risk premium decomposition allowing for a structural shift in the underlying volatility process
Estimates are based on 900 monthly excess returns from January 1926 through December 2000. The most likely date for the structural shift in the volatility process is 1940. Panel A reports the parameter estimates for the augmented time series model based on Eq. (13). The augmented model allows for a shift in the transition probabilities π_t and the means μ_t of the two state-dependent distributions. The risk aversion coefficient γ and the standard deviation of returns within each state σ_t are assumed to remain constant, such that the intrastate risk premia are constant over the entire estimation period. For each of the two subperiods defined by the date for the structural shift, Panel B reports the estimated values of the preference parameters γ, J_t, and K_t^* from Eqs. (7)–(9) that are consistent with the estimated time series model. Panel C shows the implied decomposition of the market risk premium based on Eq. (3) and the estimated model parameters. Standard errors are reported in parentheses. Because of the nonlinear nature of the model, the standard errors reported in Panels B and C are simulated.

Volatility state										
	Time series parameters			Preference parameters			State probability	Risk premium decomposition		
								State-dependent risk		
	$\mu_t - r_t$	σ_t	π_t	γ	J_t	K_t^*		Intrastate	Interstate	Total
Pre-1940 (1926–1939)		*Panel A*			*Panel B*				*Panel C*	
Low volatility	0.243	0.127	0.033	*1.703*	*−0.265*	−0.289	0.611	0.028	0.097	0.124
($s_t = L$)	(0.052)	(0.004)	(0.029)	(0.762)	(0.133)	(0.151)	(0.075)	(0.011)	(0.053)	(0.057)
High volatility	−0.076	0.373	0.052	1.703	0.360	0.407	0.389	0.238	0.085	0.322
($s_t = H$)	(0.138)	(0.020)	(0.035)	(0.762)	(0.439)	(0.496)	(0.075)	(0.099)	(0.037)	(0.103)
Unconditional mean	0.119							0.109	0.092	0.201
	(0.0270)							(0.045)	(0.044)	(0.064)
Post-1940 (1940–2000)										
Low volatility	0.118	0.127	0.027	1.703	*−0.175*	−0.152	0.955	0.028	0.020	0.048
($s_t = L$)	(0.018)	(0.004)	(0.016)	(0.762)	(0.111)	(0.127)	(0.022)	(0.011)	(0.013)	(0.014)
High volatility	−0.574	0.373	0.571	1.703	0.213	0.179	0.045	0.238	0.322	0.560
($s_t = H$)	(0.487)	(0.020)	(0.158)	(0.762)	(0.268)	(0.361)	(0.022)	(0.099)	(0.294)	(0.273)
Unconditional mean	0.087							0.037	0.034	0.071
	(0.025)							(0.017)	(0.020)	(0.019)
Log-likelihood value	1,505.3									
Number of observations	900									

Panel A of Table 4 reports the results of the augmented time series model. After testing all possible annual breakpoints from 1927 to 1999, the date of the most likely breakpoint is 1940. The structural shift in the volatility process is statistically significant. The p-value for a likelihood ratio test of the null hypothesis of no structural shift is 0.0064, indicating that the null hypothesis can be rejected at standard levels of significance.[14] I also perform a test for structural change, which does not rely on the assumption that a structural shift has taken place. Based on the Andrews (1993) Lagrange multiplier test for regime changes, the null hypothesis that market returns during the 1930s were drawn from the same regime as the other returns can be rejected at the 1% level.[15] These results are consistent with results in Pagan and Schwert (1990) and Pastor and Stambaugh (2001) showing that the 1930s were a period of unusually high market volatility that cannot be explained by a single process over the complete historical period.

As a result of the structural shift in the volatility process, the expected duration of the high-volatility state falls dramatically after 1940. Before 1940, the point estimates of the transition probabilities π_t indicate that both volatility states are persistent. After 1940, however, only the low-volatility state is persistent. The expected duration of the low-volatility state increases marginally from 30.2 months for the period before 1940 to 37.2 months for the period after 1940. In contrast, the expected duration of the high-volatility state falls significantly from 19.2 months for the period before 1940 to only 1.8 months for the period after 1940.[16] The reduction in the length of time the economy is expected to remain in the high-volatility state dramatically reduces the unconditional probability of the economy being in the high-volatility state. As a result of the shift in the volatility process, the probability of the economy being in the high-volatility state falls from 38.9% for the period before 1940 to only 4.5% for the period after 1940.

Panel B reports the preference parameter estimates consistent with the augmented time series model. The point estimate of γ equals 1.703 and is larger than the estimate in the model with no structural shift. The point estimate of J_L equals -26.5% for the period before 1940 and -17.5% for the period after 1940. Because the higher discount rates associated with the high-volatility state are expected to be applied for a shorter period of time during the period after 1940, the point estimates for the expected change in market value when the economy enters the high-volatility state are consistent with the shortening of the expected duration of the high-volatility state.

[14] The likelihood ratio statistic for the null hypothesis of no structural shift equals 14.3 and is distributed as a chi-square with 4 degrees of freedom.

[15] The sup(LM) equals 29.62. The 1930s period corresponds to $\pi \in (0.0544, 0.1878)$ and a critical value of 22.54 for a 1% test.

[16] The reduction in the persistence of the high-volatility state is consistent with the results in Poterba and Summers (1986) showing that volatility is not persistent enough for volatility-feedback to be the sole cause of the changes in market value that are observed. However, my results suggest that volatility-feedback may have played a much larger role during the period before 1940.

Panel C reports the implied risk premium decomposition for the periods before and after the 1940 structural shift. Because of the dramatic reduction in the likelihood of being in the high-volatility state, the unconditional risk premium falls significantly after 1940. For the period before 1940, the point estimate of the unconditional risk premium is 20.1%. In contrast, for the period after 1940, the point estimate of the unconditional risk premium is only 7.1%. Although the magnitude of the individual components of the risk premium changes as a result of the structural shift, the proportion of the risk premium associated with the risk of future changes in volatility state remains relatively constant at about 45%.

Given the estimated reduction in the market risk premium, the average of ex post returns during the period following 1940 is likely to be a biased proxy of the ex ante expected return during the period since 1940. As investors learn that market risk has fallen because of the structural shift in the volatility process, stock prices will be bid up and ex post realized returns will be greater than ex ante expected returns. Assuming a real risk-free rate of 1%, a reduction in the market risk premium from 20% to 7% would cause the value of a perpetuity growing at a real rate of 2% per year to increase by approximately 213%. However, it is unlikely that investors would instantaneously realize that the transition probabilities governing the evolution of the two volatility states had changed. Given the expected duration of the low- and high-volatility periods, learning the values of the new transition probabilities would not be a trivial exercise and could easily take many years to uncover. For example, if this learning process took place over a period of 20 years, ex post returns would exceed ex ante expected returns during this period by approximately 5.9%. For this reason, I test for evidence of positive abnormal returns during the period following the 1940 structural shift in the underlying volatility process. Table 5 reports these results.

Table 5 presents actual excess returns for alternative subperiods from 1940 to 2000. I group the data by decade and report the average excess return for two periods: the decades immediately following the 1940 structural shift and the subsequent decades. The estimates in Table 5 show that the average excess return during the period from 1940 to 1959 is significantly greater than that during the subsequent 41-year period from 1960 through 2000. Consistent with the hypothesis of a structural shift in the volatility process following the 1930s, the p-value for a one-tailed test of the null hypothesis that the mean excess returns during these two periods are equal is 0.0458, indicating that the null hypothesis can be rejected at the 5% level. The magnitude of the excess return from 1940 to 1959 is also consistent with change in the market risk premium reported in Table 4. The average excess return during the 20-year period following the structural shift of 6.5% is comparable to the amortized percentage change in the value of a growing perpetuity implied by the reduction in the market risk premium of 5.9%. These results are consistent with the hypothesis that investors may have updated their beliefs regarding the level of market risk at some point during the period from 1940 to 1960. Given the evidence of abnormal returns after 1940, I re-estimate the model presented in Table 4 allowing for an abnormal return during the period following the structural shift.

Table 5
Analysis of excess returns during the period following the 1940 structural shift in the volatility process
Excess returns are grouped by decade into two subperiods following the structural shift: the period immediately following 1940 structural shift and the subsequent period. For each subperiod, the annualized mean excess return is reported along with the annualized standard deviation in returns and the difference in the means of the two subperiods. The last column reports the *p*-value for a one-tailed test of the null hypothesis of equal mean excess returns in the two subperiods.

Post-1940 subperiod	Mean	Standard deviation	Difference in means	*p*-value[a]
1: 1940–1949	10.0%	15.4%		
2: 1950–2000	8.2	14.5	1.8%	0.3662
1: 1940–1959	12.8	13.4		
2: 1960–2000	6.4	15.2	6.5	0.0458
1: 1940–1969	10.3	13.2		
2: 1970–2000	6.8	15.9	3.5	0.1775
1: 1940–1979	8.1	14.2		
2: 1980–2000	9.3	15.4	−1.2	0.6185
1: 1940–1989	8.2	14.7		
2: 1990–2000	9.9	14.2	−1.8	0.6438

[a] Based on Smith-Satterhwaite test for difference in population means with unequal variances, Miller and Freund (1977).

Table 6 reports the results from re-estimating the augmented model, allowing for abnormal returns during the 20-year period subsequent to the 1940 structural shift. The model is identical to that reported in Table 4 except for the inclusion of a dummy variable in the equations for the mean of each state-dependent distribution. The dummy variable equals one during the period from 1940 through 1959 and zero otherwise. The coefficient on the dummy variable provides an estimate of the mean abnormal return during the period following the structural shift. The point estimate of the average abnormal return during this period equals 5%, indicating that realized returns following the structural shift exceeded those required based on the underlying volatility process. The *p*-value for a one-tailed test that the estimated coefficient equals zero is 0.0941, indicating that the null hypothesis that there were no abnormal returns during this period can be rejected at the 10% level.

The estimated value of the market risk premium is substantially lower as a result of controlling for the presence of abnormal returns subsequent to the shift in the underlying volatility process. The point estimate of the unconditional risk premium for the period since 1940 is 5.6%, about 270 basis points lower than the historical average of excess market returns. Consistent with Brown et al. (1995) and Elton (1999), these results suggest that the simple historical average of excess market returns may substantially overstate the market risk premium for the period after the Great Depression. In addition, my results are consistent with the empirical finding in Fama and French (2002) that actual returns during the past 50 years have been much higher than expected. However, my method provides a structural basis for controlling for the extent of this bias and, as a result, provides an unbiased estimate of the market risk premium.

Table 6
Parameter estimates and implied risk premium decomposition allowing for a structural shift in the underlying volatility process and subsequent abnormal returns

Estimates are based on 900 monthly excess returns from January 1926 through December 2000. The date for the structural shift in the volatility process is 1940. The time series model adjusts for abnormal return during the 20-year period from 1940 through 1959. Panel A reports the parameter estimates for the augmented time series model based on Eq. (13). The augmented model allows for a shift in the transition probabilities π_t and the means μ_t of the two state-dependent distributions. The risk aversion parameter γ and the standard deviation of returns within each state σ_t are assumed to remain constant, such that the intrastate risk premia are constant over the entire estimation period. For each of the two subperiods defined by the date for the structural shift, Panel B reports the estimated values of the preference parameters γ, J_t , and K_t^* from Eqs. (7)–(9) that are consistent with the estimated time series model. Panel C shows the implied decomposition of the market risk premium based on Eq. (3) and the estimated model parameters. Simulated standard errors are reported in parentheses.

Volatility state	Time series parameters			Preference parameters			Risk premium decomposition			
							State probaility	State-dependent risk		
	$\mu_t - r_t$	σ_t	π_t	γ	J_t	K_t^*		Intrastate	Interstate	Total
Pre-1940 (1926–1939)		*Panel A*			*Panel B*				*Panel C*	
Low volatility	0.243	0.127	0.033	*1.491*	*−0.282*	−0.294	0.611	0.024	0.090	0.114
($s_t = L$)	(0.052)	(0.004)	(0.033)	(0.693)	(0.137)	(0.144)	(0.075)	(0.011)	(0.045)	(0.051)
High volatility	−0.075	0.375	0.052	1.491	0.393	0.416	0.389	0.209	0.084	0.293
($s_t = H$)	(0.138)	(0.020)	(0.038)	(0.693)	(0.365)	(0.460)	(0.075)	(0.096)	(0.042)	(0.107)
Unconditional mean	0.119							0.096	0.087	0.184
	(0.027)							(0.047)	(0.041)	(0.066)
Post-1940 (1940–2000)										
Low volatility	0.100	0.127	0.027	1.491	*−0.156*	−0.141	0.956	0.024	0.014	0.038
($s_t = L$)	(0.021)	(0.004)	(0.017)	(0.693)	(0.102)	(0.111)	(0.021)	(0.011)	(0.010)	(0.014)
High volatility	−0.576	0.375	0.579	1.491	0.184	0.164	0.044	0.209	0.237	0.447
($s_t = H$)	(0.492)	(0.020)	(0.157)	(0.693)	(0.186)	(0.214)	(0.021)	(0.096)	(0.261)	(0.267)
Unconditional mean	0.070							0.032	0.024	0.056
	(0.027)							(0.017)	(0.016)	(0.020)
Abnormal return: 1940–1959	0.050									
	(0.038)									
Log-likelihood value	1,506.3									
Number of observations	900									

5. Summary

This paper presents a method for estimating the market risk premium that incorporates shifts in investment opportunities and demonstrates the importance of accounting for the dynamic nature of market risk. Because of peso-type problems similar to that discussed in Rietz (1988), when investors anticipate changes in market value associated with future changes in the level of market risk, the ex post observed relationship between volatility and excess returns may severely distort the true ex ante relationship between risk and expected returns. My results suggest that the simple historical average of excess market volatility obscures significant variation in the market risk premium and that about half of the measured risk premium is associated with the risk of future changes in the level of market volatility.

The results presented in this paper also highlight the importance of distinguishing between ex post realized and ex ante expected returns as emphasized in Elton (1999). My analysis suggests that because of a structural shift in the volatility process underlying market returns and a reduction in the market risk premium, ex post returns during the period following the 1930s are not an unbiased estimate of ex ante expected returns. The bias in ex post returns is closely related to the survival bias discussed in Brown et al. (1995). My method provides a structural basis for controlling for the extent of this bias and allows for an unbiased estimate of the market risk premium. My corrected estimates suggest that the simple historical average of excess market returns substantially overstates the magnitude of the market risk premium for the period since the Great Depression.

Appendix A

Here, I derive the expression for the equilibrium risk premium given by Eq. (3) in Section 2. In the first section, I lay out the details of the investor's utility maximization problem and define the model parameters and assumptions. In the second section, I outline the steps involved in finding the equilibrium solution to this stochastic programming problem. And in the third section, I show that my solution collapses to the Merton (1969) solution to optimal lifetime portfolio selection under uncertainty when there are no changes in volatility states.

A.1. Model parameters and assumptions

I solve the utility maximization problem for a representative investor in an infinite horizon, continuous-time model with discrete volatility states. I assume that preferences are described by a power utility function parameterized by γ, the coefficient of relative risk aversion. I also assume that there are only two assets in which the investor can invest: a risk-free asset yielding a certain rate of return equal to r_t and a risky asset denoted S_t with an uncertain rate of return equal to $\mathrm{d}S_t/S_t$. The standard deviation σ_t of the returns on the risky asset varies over time and is assumed to take on only two values, σ_L and σ_H. The simple average of the two

volatility levels is denoted by the parameter $\bar{\sigma}$. Correspondingly, the expected drift in the price of the risky asset μ_t varies with state and takes on two values, μ_L and μ_H. The simple average of the two means is denoted by the parameter $\bar{\mu}$. In each volatility state, the probability that the economy will switch to the alternative volatility state is determined by the parameter π_t. Because the evolution of volatility states is assumed to follow a Markov process, π_t takes on two values, π_L and π_H. The simple average of the two values for π_t is denoted by the parameter $\bar{\pi}$. At each instant, the investor chooses an amount of consumption C_t and a fraction ω_t of his wealth W_t to invest in the risky asset. The investor's problem is given as

$$\max_{C_t,\omega_t} \quad \mathrm{E}_v \int_v^{\infty} \mathrm{e}^{-\rho t} \frac{C_t^{1-\gamma}}{1-\gamma} \mathrm{d}t, \tag{A.1}$$

$$\text{s.t. } \mathrm{d}W_t = \omega_t W_t \frac{\mathrm{d}S_t}{S_t} + (1-\omega_t) r_t W_t \,\mathrm{d}t - C_t \,\mathrm{d}t, \tag{A.2}$$

$$\mathrm{d}S_t = \mu_t S_t \,\mathrm{d}t + \sigma_t S_t \,\mathrm{d}Z + J_t S_t \,\mathrm{d}N(\pi_t), \tag{A.3}$$

$$\mathrm{d}\mu_t = 2(\bar{\mu} - \mu_t) \,\mathrm{d}N(\pi_t), \tag{A.4}$$

$$\mathrm{d}\sigma_t = 2(\bar{\sigma} - \sigma_t) \,\mathrm{d}N(\pi_t), \tag{A.5}$$

$$\mathrm{d}\pi_t = 2(\bar{\pi} - \pi_t) \,\mathrm{d}N(\pi_t), \tag{A.6}$$

$$\mathrm{d}J_t = 2\left(\bar{J} - J_t\right) \mathrm{d}N(\pi_t), \tag{A.7}$$

$$\mathrm{d}\hat{\lambda}_t = 2\left(\overline{\hat{\lambda}} - \hat{\lambda}_t\right) \mathrm{d}N(\pi_t), \tag{A.8}$$

and

$$C_t > \bar{C}_t, \tag{A.9}$$

where $\mathrm{d}Z$ is a standard Weiner process and $\mathrm{d}N(\pi_t)$ is a Poisson process that is equal to either zero or one. When $\mathrm{d}N(\pi_t) = 1$, Eqs. (A.4)–(A.6) cause the drift, volatility, and transition parameters to jump to the alternative state. Given the discrete jumps in these state variables, the equation describing the evolution of the stock price S_t includes the term $J_t S_t \,\mathrm{d}N(\pi_t)$, which allows the stock price to jump when the economy switches between volatility states. The parameter J_t is the magnitude of the jump in stock price that occurs when the economy switches state. The value of the jump parameter J_t takes on two values, J_L and J_H. The simple average of the two jump values is denoted by the parameter $\bar{J}$. Finally, Eqs. (A.8) and (A.9) allow for the possibility that consumption may be constrained in one of the volatility states. The value of the Lagrange multiplier associated with this constraint is given by the parameter $\hat{\lambda}_t$, which takes on two values, $\hat{\lambda}_L$ and $\hat{\lambda}_H$. The simple average of the two Lagrange multipliers is denoted by the parameter $\overline{\hat{\lambda}}$.

A.2. Derivation of the equilibrium solution

Given the problem described above, the indirect utility function at time v is defined as a function of the state variables at time v, such that

$$I_v = \max\ \mathrm{E}_v \int_v^{\infty} \mathrm{e}^{-\rho t} \frac{C_t^{1-\gamma}}{1-\gamma}\,\mathrm{d}t, \tag{A.10}$$

where $I_v = I(W_v, \mu_v, \sigma_v, \pi_v, J_v, \hat{\lambda}_v)$. From the principle of optimality,

$$\begin{aligned} 0 = &\frac{C_t^{1-\gamma}}{1-\gamma} - \rho I + [(\omega_t(\mu_t - r_t) + r_t)W_t - C_t]\frac{\partial I}{\partial W} \\ &+ \frac{1}{2}\omega_t^2\sigma_t^2 W_t^2 \frac{\partial^2 I}{\partial W^2} + \pi_t \mathrm{E}_t[I_t' - I_t] + \hat{\lambda}_t C_t, \end{aligned} \tag{A.11}$$

where I_t' is the value of the indirect utility function subsequent to the next change of state and is equal to

$$I_t' = I\begin{pmatrix} W_t + \omega_t J_t W_t, \mu_t + 2(\bar{\mu} - \mu_t), \sigma_t + 2(\overline{\sigma} - \sigma_t), \\ \pi_t + 2(\bar{\pi} - \pi_t), J_t + 2(\bar{J} - J_t), \hat{\lambda}_t + 2(\overline{\hat{\lambda}} - \hat{\lambda}_t) \end{pmatrix}. \tag{A.12}$$

The first-order conditions for the investor's problem with respect to C_t and ω_t are given by the expressions

$$0 = C_t^{-\gamma} - \frac{\partial I}{\partial W} + \hat{\lambda}_t \tag{A.13}$$

and

$$0 = (\mu_t - r_t)W_t \frac{\partial I}{\partial W} + \omega_t\sigma_t^2 W_t^2 \frac{\partial^2 I}{\partial W^2} + \pi_t \mathrm{E}_t\left[J_t W_t \frac{\partial I'}{\partial W}\right]. \tag{A.14}$$

Defining $\hat{\lambda}_t$ in terms of the marginal utility of wealth, such that

$$\hat{\lambda}_t = \lambda_t \frac{\partial I}{\partial W}, \tag{A.15}$$

consumption at each instant is given by the expression

$$C_t = \left[(1-\lambda_t)\frac{\partial I}{\partial W}\right]^{-1/\gamma}. \tag{A.16}$$

Because the net supply of the risk-free asset must equal zero in general equilibrium, the risk-free rate adjusts such that $\omega_t = 1$. Substituting Eq. (A.16) into Eq. (A.11), setting $\omega_t = 1$, and simplifying yields

$$\begin{aligned} 0 = &\frac{1}{1-\gamma}(1-\lambda_t)^{(\gamma-1)/\gamma}\left(\frac{\partial I}{\partial W}\right)^{(\gamma-1)/\gamma} - \rho I + \mu_t W_t \frac{\partial I}{\partial W} \\ &- (1-\lambda_t)^{(\gamma-1)/\gamma}\left(\frac{\partial I}{\partial W}\right)^{(\gamma-1)/\gamma} + \frac{1}{2}\sigma_t^2 W_t^2 \frac{\partial^2 I}{\partial W^2} + \pi_t\,\mathrm{E}_t[I_t' - I_t]. \end{aligned} \tag{A.17}$$

To solve Eq. (A.17), I guess the solution to be of the form

$$I_t = f_t \frac{W_t^{1-\gamma}}{1-\gamma}, \tag{A.18}$$

where $f_t = f(\mu_t, \sigma_t, \pi_t, J_t, \lambda_t)$. Because Eq. (A.18) must hold in each volatility state, the solution for the indirect utility function subsequent to the next change of state, I'_t, is given by the expression

$$I'_t = f'_t \frac{(W'_t)^{1-\gamma}}{1-\gamma}, \tag{A.19}$$

where f'_t and W'_t equal the values of f_t and W_t, respectively, in the subsequent volatility state. Given this solution, the first and second partial derivatives of I_t with respect to wealth are

$$\frac{\partial I}{\partial W} = f_t W_t^{-\gamma} \tag{A.20}$$

and

$$\frac{\partial^2 I}{\partial W^2} = -\gamma f_t W_t^{-(1+\gamma)}. \tag{A.21}$$

Substituting Eqs. (A.20) and (A.21) into Eq. (A.17), yields

$$\begin{aligned} 0 = {} & \frac{1}{1-\gamma}(1-\lambda_t)^{(\gamma-1)/\gamma}\left[f_t W_t^{-\gamma}\right]^{(\gamma-1)/\gamma} - \rho\left[f_t \frac{W_t^{1-\gamma}}{1-\gamma}\right] \\ & + \mu_t f_t W_t^{1-\gamma} - (1-\lambda_t)^{(\gamma-1)/\gamma}\left[f_t W_t^{-\gamma}\right]^{(\gamma-1)/\gamma} \\ & + \frac{1}{2}\sigma_t^2 W_t^2\left[-\gamma f_t W_t^{-(1+\gamma)}\right] + \pi_t \mathrm{E}_t\left[f'_t \frac{(W'_t)^{1-\gamma}}{1-\gamma} - f_t \frac{W_t^{1-\gamma}}{1-\gamma}\right]. \end{aligned} \tag{A.22}$$

In general equilibrium, $\omega_t = 1$ such that all wealth is held in the form of the risky asset. For this reason, the expression Eq. (A.22) can be simplified by substituting the expression $W'_t = (1+J_t)W_t$. This yields the expression

$$\begin{aligned} 0 = {} & f_t^{-1/\gamma}\gamma(1-\lambda_t)^{(\gamma-1)/\gamma} - \rho + (1-\gamma)\mu_t \\ & - \frac{1}{2}\gamma(1-\gamma)\sigma_t^2 + \pi_t \mathrm{E}_t\left[(1+\varepsilon_t)(1+J_t)^{1-\gamma} - 1\right], \end{aligned} \tag{A.23}$$

where $1+\varepsilon_t = f'_t/f_t$. From Eqs. (A.16) and (A.20), $(1+\varepsilon_t)$ is given by the expression

$$(1+\varepsilon_t) = \frac{(1-\lambda_t)(1+J_t)^{\gamma}}{(1-\lambda'_t)(1+K_t)^{\gamma}}. \tag{A.24}$$

Substituting Eq. (A.24) into Eq. (A.23) and solving for $f(\mu_t, \sigma_t, \pi_t, J_t, \lambda_t)$ yields

$$\begin{aligned} f_t = {} & \left[\frac{\rho + (\gamma-1)\mu_t - \frac{1}{2}\gamma(\gamma-1)\sigma_t^2}{\gamma(1-\lambda_t)^{1-\gamma}}\right. \\ & \left. + \frac{\pi_t}{\gamma(1-\lambda_t)^{1-\gamma}}\left(1 - \frac{(1-\lambda_t)(1+J_t)^{\gamma}}{(1-\lambda'_t)(1+K_t)^{\gamma}}\right)\right]^{-\gamma}, \end{aligned} \tag{A.25}$$

where K_t is the jump in consumption that is expected conditional on switching state. Because λ_t' can be expressed in terms of λ_t using Eqs. (A.8), (A.25) verifies that Eq. (A.18) is the solution to Eq. (A.17).

Using Eqs. (A.16), (A.20), and (A.25), the equilibrium consumption–wealth ratio in the model is given by

$$\frac{C_t}{W_t} = \frac{\rho + (\gamma - 1)\mu_t - \frac{1}{2}\gamma(\gamma - 1)\sigma_t^2}{\gamma(1 - \lambda_t)} + \frac{\pi_t}{\gamma(1 - \lambda_t)}\left(1 - \frac{(1 - \lambda_t)(1 + J_t)^{\gamma}}{(1 - \lambda_t')(1 + K_t)^{\gamma}}\right). \quad \text{(A.26)}$$

In Section A.3, I show that, when there are no changes in volatility states, the second term of Eq. (A.26) equals zero and the first term is equivalent to the Merton (1969) solution to the infinite horizon lifetime portfolio selection problem under uncertainty.

The expression for the equilibrium risk premium is found by taking the mathematical expectation of $\mathrm{d}S_t/S_t$ and substituting the equilibrium within-state excess return implied by the first-order condition for ω_t. From Eq. (A.3), the expected excess return on the risky asset is given by the expression

$$\mathrm{E}_t\left[\frac{\mathrm{d}S_t}{S_t}\right] - r_t = \mu_t + \pi_t J_t - r_t. \quad \text{(A.27)}$$

The expression for the within-state excess return $\mu_t - r_t$ is derived by substituting Eqs. (A.20) and (A.21) into Eq. (A.14), setting $\omega_t = 1$, and simplifying, such that

$$\mu_t - r_t = \gamma\sigma_t^2 - \pi_t J_t(1 + \varepsilon_t)(1 + J_t)^{-\gamma}. \quad \text{(A.28)}$$

Combining Eqs. (A.27) and (A.28), substituting Eq. (A.24), and simplifying yields the expression for the equilibrium risk premium

$$\mathrm{E}_t\left[\frac{\mathrm{d}S_t}{S_t}\right] - r_t = \gamma\sigma_t^2 + \pi_t J_t\left(1 - \frac{(1 - \lambda_t)}{(1 - \lambda_t')(1 + K_t)^{\gamma}}\right). \quad \text{(A.29)}$$

If the constraint on consumption does not bind in either state, then Eq. (A.29) can be simplified as

$$\mathrm{E}_t\left[\frac{\mathrm{d}S_t}{S_t}\right] - r_t = \gamma\sigma_t^2 + \pi_t J_t[1 - (1 + K_t)^{-\gamma}]. \quad \text{(A.30)}$$

Eq. (A.30) is the expression for the market risk premium provided in the text as Eq. (3). Eq. (A.30) shows that the equilibrium risk premium in each state can be decomposed into two state-dependent risk premia, an intrastate risk premium and an interstate risk premium. The first term, $\gamma\sigma_t^2$, describes the required intrastate risk premium required to compensate for diffusion risk within the current state. The second term, $\pi_t J_t[1 - (1 + K_t)^{-\gamma}]$, describes the required interstate risk premium required to compensate for potential jump risk arising from a change in volatility state.

Eq. (A.29) can also be used to show that the equilibrium risk premium is invariant to the actual jumps in consumption that occur when the economy changes state. For example, if the constraint on consumption does not bind in either state, such that $\lambda_L = \lambda_H = 0$, then the risk premium in the low-volatility state is given by the

expression

$$\mathrm{E}_t[R_L] - r_L = \gamma\sigma_L^2 + \pi_L J_L\left[1 - (1 + K_L^*)^{-\gamma}\right], \tag{A.31}$$

where K_L is the optimal change in the level of consumption when the economy switches from the low- to the high-volatility state. Alternatively, if consumption is unable to adjust when the economy enters the high-volatility state, then the constraint on consumption will bind in the high-volatility state, such that $\lambda_H > \lambda_L = 0$. In this case, the expression for the risk premium in the low-volatility state is given by the expression

$$\mathrm{E}_t[R_L] - r_L = \gamma\sigma_L^2 + \pi_L J_L\left[1 - (1 - \lambda_H)^{-1}(1 + \tilde{K}_L)^{-\gamma}\right], \tag{A.32}$$

where $\tilde{K}_L$ is the constrained change in the level of consumption when the economy switches from the low- to the high-volatility state. As a result of the constraint on consumption, the shadow price increases to reflect the fact that the actual level of consumption is no longer equal to the optimal level. The shadow price on the consumption constraint in the high-volatility state is given by the expression

$$\lambda_H = 1 - \left(\frac{1 + K_L^*}{1 + \tilde{K}_L}\right)^{\gamma}. \tag{A.33}$$

Eq. (A.33) is the expression for the Lagrange multiplier on the consumption constraint in the high-volatility state provided in the text as Eq. (4).

A.3. The special case of no changes in volatility state

This section shows that, when there are no changes in volatility state, my solution collapses to the Merton (1969) solution to the lifetime portfolio selection problem under uncertainty. Eqs. (A.26) and (A.30) summarize my solution to the investor's utility maximization problem when there are two discrete volatility states. Eq. (A.26) describes the optimal consumption–wealth ratio and Eq. (A.30) describes the equilibrium risk premium. If, instead, a single volatility state is assumed, then the dynamics associated with changes in volatility states can be turned off by setting $\pi_t = 0$ and $\lambda_t = 0$. By setting $\pi_t = 0$, only one volatility state is possible. With only one volatility state, there are no wealth jumps associated with changes in state and $\mathrm{E}_t[\mathrm{d}S_t/S_t] = \mu_t$. Also, because there are no jumps in wealth, there are no jumps in optimal consumption, so that $\lambda_t = 0$. Thus, for the special case of a single volatility state, Eqs. (A.26) and (A.30) can be rewritten as

$$\frac{C_t}{W_t} = \frac{\rho}{\gamma} + (\gamma - 1)\left[\frac{\mu_t}{\gamma} - \frac{\sigma_t^2}{2}\right] \tag{A.34}$$

and

$$\mu_t - r_t = \gamma\sigma_t^2. \tag{A.35}$$

Rearranging Eq. (A.34) yields

$$\frac{C_t}{W_t} = \frac{\rho}{\gamma} - (1 - \gamma)\left[\frac{\sigma_t^2}{2} + \frac{\mu_t - \gamma\sigma_t^2}{\gamma}\right]. \tag{A.36}$$

Using Eq. (A.35) to simplify the term $\mu_t - \gamma\sigma_t^2$, Eq. (A.36) can be rewritten as

$$\frac{C_t}{W_t} = \frac{\rho}{\gamma} - (1-\gamma)\left[\frac{\sigma_t^2}{2} + \frac{r_t}{\gamma}\right]. \tag{A.37}$$

Eq. (A.35) can also be used to express σ_t^2 in terms of excess returns, such that

$$\frac{C_t}{W_t} = \frac{\rho}{\gamma} - (1-\gamma)\left[\frac{\mu_t - r_t}{2\gamma} + \frac{r_t}{\gamma}\right]. \tag{A.38}$$

Finally, Eq. (A.35) can be used to rewrite the first term in brackets in a manner similar to that in Merton (1969)

$$\begin{aligned}\frac{C_t}{W_t} &= \frac{\rho}{\gamma} - (1-\gamma)\left[\frac{\mu_t - r_t}{2\gamma}\left(\frac{\mu_t - r_t}{\gamma\sigma_t^2}\right) + \frac{r_t}{\gamma}\right] \\ &= \frac{\rho}{\gamma} - (1-\gamma)\left[\frac{(\mu_t - r_t)^2}{2\gamma\sigma_t^2} + \frac{r_t}{\gamma}\right].\end{aligned} \tag{A.39}$$

Eq. (A.39) is equivalent to the Merton (1969) expression for the optimal consumption–wealth ratio in the infinite horizon lifetime portfolio selection problem.[17] This demonstrates that my model solution contains the Merton (1969) solution as a special case when there are no changes in volatility state.

References

Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821–856.

Baillie, R.T., DeGennaro, R.P., 1990. Stock returns and volatility. Journal of Financial and Quantitative Analysis 25, 203–214.

Berndt, E., Hall, B., Hall, R., Hausman, J., 1974. Estimation and inference in nonlinear structural models. Annals of Economic and Social Measurement 3/4, 653–665.

Brown, S.J., Goetzmann, W.N., Ross, S.A., 1995. Survival. The Journal of Finance 50, 853–872.

Bruner, R.F., Eades, K.M., Harris, R.S., Higgins, R.C., 1998. Best practices in estimating the cost of capital, survey and synthesis. Financial Practice and Education, 13–28.

Campbell, J.Y., 1987. Stock returns and the term structure. Journal of Financial Economics 18, 373–399.

Campbell, J.Y., 1991. A variance decomposition for stock returns. Economic Journal 101, 157–179.

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., Hentschel, L., 1992. No news is good news. Journal of Financial Economics 31, 281–318.

Campbell, J.Y., Shiller, R.J., 1988. The dividend-price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1, 195–227.

Cecchetti, S., Lam, P., Mark, N.C., 1990. Mean reversion in equilibrium asset prices. American Economic Review 80, 398–418.

[17] The optimal consumption–wealth ratio for the infinite horizon problem is provided as Eq. (42) in the original Merton (1969) article.

Diebold, F.X., Lee, J., Weinbach, G.C., 1994. Regime switching with time-varying transition probabilities. Nonstationary Time Series Analysis and Cointegration, 283–302.
Elton, E.J., 1999. Presidential address: expected return, realized return, and asset pricing tests. Journal of Finance 54, 1199–1220.
Fama, E.F., 1965. The behavior of stock-market prices. Journal of Business 38, 34–105.
Fama, E.F., French, K.R., 1988. Dividend yields and expected stock returns. Journal of Financial Economics 22, 3–27.
Fama, E.F., French, K.R., 1989. Business conditions and expected returns on stocks and bonds. Journal of Financial Economics 25, 23–49.
Fama, E.F., French, K.R., 2002. The equity premium. Journal of Finance 57, 637–659.
Fama, E.F., Schwert, G.W., 1977. Asset returns and inflation. Journal of Financial Economics 5, 115–146.
French, K.R., Schwert, G.W., Stambaugh, R.F., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3–29.
Glosten, L.R., Jagannathan, R., Runkle, D.E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779–1801.
Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series models. Journal of Econometrics 70, 127–157.
Hamilton, J.D., Lin, G., 1996. Stock market volatility and the business cycle. Journal of Applied Econometrics 11, 573–593.
Hamilton, J.D., Susmel, R., 1994. Autoregressive conditional heteroskedasticity and changes in regime. Journal of Econometrics 64, 307–333.
Hodrick, R., 1992. Dividend yields and expected stock returns: alternative procedures for inference and measurement. Review of Financial Studies 5, 357–386.
Kim, C-J, Morley, J.C., Nelson, C.R., 2000. Is there a significant positive relationship between stock market volatility and the equity premium? Mimeo. Washington University.
Lamont, O., 1998. Earnings and expected returns. Journal of Finance 53, 1563–1587.
Lettau, M., Ludvigson, S., 2001. Resurrecting the (C)CAPM: a cross-sectional test when risk premia are time-varying. Journal of Political Economy 109, 1238–1287.
Malkiel, B.G., 1979. The capital formation problem in the United States. The Journal of Finance 34, 291–306.
Merton, R.C., 1969. Lifetime portfolio selection under uncertainty: the continuous-time case. Review of Economics and Statistics 51, 247–257.
Merton, R.C., 1973. An intertemporal asset pricing model. Econometrica 41, 867–888.
Merton, R.C., 1980. On estimating the expected return on the market: an exploratory investigation. Journal of Financial Economics 8, 323–361.
Miller, I., Freund, J.E., 1977. Probability and Statistics for Engineers. Prentice-Hall Inc., Englewood Cliffs, NJ.
Officer, R.R., 1973. The variability of the market factor of the New York Stock Exchange. Journal of Business 46, 434–453.
Pagan, A.R., Schwert, G.W., 1990. Alternative models for conditional stock volatility. Journal of Econometrics 45, 267–290.
Pastor, L., Stambaugh, R.F., 2001. The equity premium and structural breaks. Journal of Finance 56, 1207–1239.
Pindyck, R.S., 1984. Risk, inflation, and the stock market. American Economic Review 74, 335–351.
Poterba, J.M., Summers, L.H., 1986. The persistence of volatility and stock market fluctuations. The American Economic Review 76, 1142–1151.
Rietz, T.A., 1988. The equity premium: a solution. Journal of Monetary Economics 22, 117–131.
Schaller, H., Van Norden, S., 1997. Regime-switching in stock market returns. Applied Financial Economics 7, 177–191.
Schwert, G.W., 1989a. Business cycles, financial crises, and stock volatility. Carnegie-Rochester Conference Series on Public Policy 31, 83–126.
Schwert, G.W., 1989b. Why does stock market volatility change over time? Journal of Finance 44, 1115–1153.

Scruggs, J.T., 1998. Resolving the puzzling intertemporal relation between the market risk premium and conditional market variance: a two-factor approach. Journal of Finance 53, 575–603.

Shiller, R.J., 1984. Stock prices and social dynamics. Brookings Papers on Economic Activity 2, 457–498.

Siegel, J.J., 1992. The equity premium: stock and bond returns since 1802. Financial Analysts Journal 48, 28–38.

Turner, C.M., Startz, R., Nelson, C.R., 1989. A Markov model of heteroskedasticity, risk, and learning in the stock market. Journal of Financial Economics 25, 3–22.