Estimating the Market Risk Premium: A Revisit

Jinho Bae †

Yeungnam University

April 25, 2007

Abstract

Mayfield (2004) estimated the market risk premium in the ICAPM framework that accounts for changes in investment opportunities. We claim that these estimates were inconsistent due to an endogeneity problem associated with the assumption that investors have perfect knowledge about the volatility states. We estimate the market risk premium controlling for the endogeneity. Our empirical results show that imposing this perfect knowledge assumption understates the total market risk premium and overstates the relative importance of the risk premium for shifts in investment opportunities.

† Correspondence author: 214-1, Dae-dong, School of Economics and Finance, Yeungnam University, Gyongsan, Gyongbuk 712-749, Korea; jinhby@yumail.ac.kr; +82-53-810-2715

1. Introduction

It has long been of interest to both practitioners and academics how much of a reward investors require for bearing market risk. Various methods have been proposed to estimate the market risk premium, ranging from simply averaging the historical excess market returns to many theoretical models. One of the appealing models is Merton's (1973) intertemporal capital asset pricing model (hereafter ICAPM). The model predicts that the market risk premium is the sum of a compensation for conditional market variance and a compensation for future shifts in investment opportunities. Mayfield (2004) provides a method for estimating the market risk premium in the ICAPM framework that accounts for changes in investment opportunities. Mayfield finds that shifts in investment opportunities that are specified as being changes in volatility states account for approximately 50% of the measured market risk premium. [1]

In our work, described in this paper, we call into question the risk premium estimates of Mayfield (2004) since these estimates seem incorrectly estimated due to an endogeneity problem. The endogeneity emanates from an assumption about the investors' information set in that investors are able to observe the current volatility state with certainty. Mayfield (2004) justified the assumption by documenting that the assumption was in the spirit of the Merton's (1980) model. However, we have not clearly understood how investors perceived the volatility state. A similar notion was that economic agents knew that there were business cycles but they were not able to correctly identify business conditions to any degree of precision. Furthermore, even the work of Merton (1980) made it explicit that the variance rate on the market return, in reality, was not observable, and therefore must be estimated (see page 330). Hence, it was reasonable to believe that investors drew inference about volatility states, rather than observing the states. Then, the volatility state that investors were assumed to observe was a proxy for investors' estimate of the true state. In the estimation from the Markov regime-switching model that Mayfield

[1] The importance of controlling for shifts in investment opportunities is also stressed by Scruggs (1998) and Guo and Whitelaw (2006). These studies find that when shifts in investment opportunities are controlled for, a significant positive risk-return tradeoff is obtained while earlier empirical evidence on the relationship between risk and return is mixed.

(2004) used to specify the dynamics of the volatility states, he created an econometric problem. The state variable was not uncorrelated to the disturbance terms, unlike in typical Markov-switching models. Kim et al. (2005) showed that in the presence of endogenous switching, maximum likelihood estimates from the Markov regime-switching models using the Hamilton's (1989) filter were inconsistent.

Here in our work, we re-estimated the market risk premium by dealing with the endogeneity problem. In addition, in our work, we statistically tested whether or not the true volatility state was observable by investors by testing for the endogeneity. For these purposes, we used a highly useful strategy, recently developed by Kim et al. (2005). In Section 2, we review Mayfield (2004) and discuss the endogeneity issue. In Section 3 we present empirical results. We conclude in Section 4.

2. Model Specification

2.1 The Basic Model of Mayfield (2004)

Here we review the basic model of Mayfield (2004) used to estimate the market risk premium. Mayfield (2004) solved an investor's utility maximization problem to derive the following expression for the equilibrium market risk premium:

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

where $E(R_t)$ is the expected market return, R_t^f is the risk-free rate, and γ is the coefficient of relative risk aversion. The market volatility that measures the market risk is specified as

$$\sigma_t^2 = \sigma_0^2(1 - S_t) + \sigma_1^2 S_t, \tag{2}$$

where S_t is a state variable that takes on values of 0 or 1. This indicates that σ_t^2 is either σ_0^2 or σ_1^2 depending on the state variable S_t. We restrict $\sigma_1^2 > \sigma_0^2$, meaning that $S_t = 1$ is the high volatility state. The two states make recurrent switches between each other according to a first-order Markov process with transition probabilities

$$\pi_t = \begin{cases} \pi_0 = Pr(S_{t+1} = 1|S_t = 0) \\ \pi_1 = Pr(S_{t+1} = 0|S_t = 1), \end{cases} \tag{3}$$

where π_0 is the probability of a shift to the high volatility state from the low volatility state and π_1 vice versa. A key assumption of Mayfield (2004) was that investors had perfect knowledge about the current volatility state. Observing the current risk, they required a compensation for it. The first term of the right-hand side of Equation (1) indicates this compensation. If they were in the low volatility state, they required $\gamma\sigma_0^2$. Otherwise, $\gamma\sigma_1^2$ was required. We refer to the first term as the intrastate risk premium, following Mayfield (2004).

Each period, investors faced a possible state change for the next period. Changes in the state can be seen as discrete shifts in the investment opportunities of Merton (1973). [2] Investors desired to hedge these shifts. The second term of the right hand side of Equation (1) indicates this hedging. J_t is the percentage change in wealth associated with a change in the volatility state, and K_t^* is the percentage change in the optimal level of consumption resulting from a change in the volatility state. It is worthwhile noting that the changes in wealth and consumption associated with a shift in state, J_t and K_t, are state-dependent and have the following relations resulting from the existence of only two states

$$J_1 = \frac{1}{1 + J_0} - 1, \tag{4}$$

$$K_1 = \frac{1}{1 + K_0} - 1, \tag{5}$$

Unlike the current state, investors have no perfect knowledge about future states. Investors require a compensation that is in proportion to the probability of a state shift. For example, if the economy is in a low volatility state, i.e., $S_t = 0$, investors require a premium for a possible state shift as much as $\pi_0 ln(1 + J_0)[1 - (1 + K_0^*)^{-\gamma}]$. Mayfield (2004) refers to the hedging component as the interstate risk premium.

[2] Scruggs (1998) and Guo and Whitelaw (2006) instrument shifts in investment opportunities using macro variables, excess returns on a long-term government bond, and the consumption-wealth ratio, respectively.

There remains one difficulty with treating Equation (1) as the mean equation and applying Hamilton's (1989) filter to estimate the risk premium. The parameters γ, J_t, and K_t that are eventually used to estimate the two components of the risk premium cannot be identified. For identification, Mayfield (2004) adopted the following strategy. Initially, he recognized that the expected return on the market can be expressed alternatively as

$$E(R_t) = U_t + \pi_t ln(1 + J_t), \tag{6}$$

where U_t is the expected intrastate return and $\pi_t ln(1 + J_t)$ is the expected change in wealth from a shift in states. Equation (6), together with Equation (1), yields

$$\mu_t = \gamma\sigma_t^2 - \pi_t ln(1 + J_t)(1 + K_t^*)^{-\gamma}, \tag{7}$$

where $\mu_t = U_t - R_t^f$. This expression provides the basis on which parameters of our interest can be estimated in two steps. The first step is to estimate the two-state Markov-switching mean and variance model

$$R_t - R_t^f = \mu_t + \sigma_t\epsilon_t, \quad \epsilon_t \sim i.i.d.N(0, 1), \tag{8}$$

$$\mu_t = \mu_0(1 - S_t) + \mu_1 S_t, \tag{9}$$

$$\sigma_t^2 = \sigma_0^2(1 - S_t) + \sigma_1^2 S_t, \tag{10}$$

to obtain the estimates $\hat{\mu}_0$, $\hat{\mu}_1$, $\hat{\sigma}_0^2$, $\hat{\sigma}_1^2$, $\hat{\pi}_0$, and $\hat{\pi}_1$. In the second step, the six time series estimates from the first step are mapped into the parameters γ, J_0, K_0, σ_0^2, σ_1^2, π_0, and π_1. Notice that once J_0 and K_0^* are estimated, it is straightforward to estimate J_1 and K_1^* using the relations given in Equations (4) and (5). Mapping yields

$$\hat{\mu}_0 = \gamma\hat{\sigma}_0^2 - \hat{\pi}_0 ln(1 + J_0)(1 + K_0^*)^{-\gamma}, \tag{11}$$

$$\hat{\mu}_1 = \gamma\hat{\sigma}_1^2 - \hat{\pi}_1 ln(1 + J_1)(1 + K_1^*)^{-\gamma}. \tag{12}$$

There are three unknown parameters γ, J_0, K_0^* while the number of equations is two. Hence, an additional equation is required. Mayfield (2004) used the expression for the equilibrium consumption-wealth ratio, $\frac{C_t^*}{W_t}$:

$$\frac{C_t^*}{W_t} = \frac{\rho + (\gamma - 1)\mu_t - \frac{1}{2}\gamma(\gamma - 1)\sigma_t^2}{\gamma} + \frac{\pi_t}{\gamma}[1 - (\frac{1 + J_t}{1 + K_t^*})^{\gamma}]. \tag{13}$$

Equation (13) is numerically solved for K_t^* due to its nonlinearity. Once the preference parameters are estimated, it is straightforward to calculate the total market risk premium and evaluate the relative importance of the intrastate and interstate risk premiums to the total risk premium based on Equation (1).

2.2 The Issue of Endogenous Switching

This subsection discusses the potential econometric problem, endogenous switching, of the model described earlier and how to correctly estimate the Markov-switching model in the presence of endogenous switching via a brief summary of Kim et al. (2005). To start with, Kim et al. (2005) employed the following probit specification for the state variable S_t

$$S_t = \begin{cases} 0 & \text{if} \quad \eta_t < a_{S_{t-1}} \\ 1 & \text{if} \quad \eta_t \geq a_{S_{t-1}}, \end{cases} \tag{14}$$

where $\eta_t \sim i.i.d.N(0,1)$. Then, the transition probabilities given in Equation (3) can be expressed as

$$\begin{aligned} 1 - \pi_0 &= Pr(\eta_t < a_0) = \Phi(a_0), \\ \pi_0 &= Pr(\eta_t \geq a_0) = 1 - \Phi(a_0), \\ \pi_1 &= Pr(\eta_t < a_1) = \Phi(a_1), \\ 1 - \pi_1 &= Pr(\eta_t \geq a_1) = 1 - \Phi(a_1), \end{aligned} \tag{15}$$

where Φ is the standard normal cumulative distribution function. Assuming a bi-variate normal for the joint density function for ϵ_t and η_t, we have

$$\begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix} \sim N(0, \Sigma), \quad \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \tag{16}$$

When ϵ_t is uncorrelated with η_t, i.e., $\rho = 0$, we have ϵ_t independent of $\{S_t, \cdots, S_1\}$. This case corresponds to Mayfield's model (2004). As assumed by Mayfield (2004), if investors observe the states with certainty, the state indicator variable S_t contained in the regressor μ_t is uncorrelated with the disturbance term ϵ_t. It is well established that when the state variable is exogenous to the disturbance terms, the maximum likelihood estimates based on Hamilton's(1989) filter are consistent.

However, if ϵ_t is correlated with η_t, i.e., $\rho \neq 0$, ϵ_t is no longer independent of $\{S_t, \cdots, S_1\}$ and the maximum likelihood estimates, based on Hamilton's filter are no longer consistent. An obvious example is when we relax the assumption made by Mayfield (2004). If in fact investors cannot observe the volatility state with certainty, investors should draw an inference on the latent state variable conditional on the available information. Then, the time series model, Equation (8), to be used to estimate the risk premium by Mayfield (2004) changes to

$$R_t - R_t^f = \mu_0(1 - E(S_t|I_{t-1})) + \mu_1 E(S_t|I_{t-1}) + \sigma_t v_t, \quad v_t \sim i.i.d.N(0,1), \tag{17}$$

where I_{t-1} is the information set available to investors at time $t-1$. This indicates that S_t in Equation (9) proxies $E(S_t|I_{t-1})$, and

$$\epsilon_t = v_t + (\mu_1 - \mu_0)(S_t - E(S_t|I_t)). \tag{18}$$

It is now clear that ϵ_t is correlated with the state variable S_t, i.e., $Cov(S_t, \epsilon_t) \neq 0$. This points to the existence of endogenous switching. Hence, if Equation (8) is estimated when Equation (17) is correct, the resulting estimates are no longer consistent.

Hamilton's (1989) filter allows us to estimate Markov-switching models by enabling us to construct a conditional likelihood function. For example, the conditional likelihood function $I_T = [r_T, \cdots, r_1]'$ where r_t is the excess return on the market, i.e., $r_t = R_t - R_t^f$, and can be constructed as

$$L(\theta) = \Pi_{t=1}^{T} f(r_t|I_{t-1}; \theta), \tag{19}$$

where

$$f(r_t|I_{t-1}; \theta) = \sum_{i=1}^{2} \sum_{j=1}^{2} f(r_t|S_t = i, S_{t-1} = j; \theta) Pr(S_t = i, S_{t-1} = j|I_{t-1}; \theta). \tag{20}$$

In the absence of endogenous switching, it holds that

$$E(\epsilon_t|S_t = i, S_{t-1} = j) = 0, \tag{21}$$

and the conditional densities r_t given S_t and S_{t-1}, $f(r_t|S_t = i, S_{t-1} = j; \theta)$, follow normal densities such that

$$f(r_t|S_t = i, S_{t-1} = j; \theta) = \frac{1}{\sigma_i} \phi(r_{i,t}^*), \tag{22}$$

where ϕ is the standard normal probability density function and $r^*_{i,t} = \frac{r_t - \mu_t}{\sigma_i}$.

However, in the presence of endogenous switching, the original Hamilton's (1989) filter is not applicable and must be modified. Now the expectation of ϵ_t, conditional on S_t and S_{t-1}, is no longer zero. Rather, it is true that

$$\begin{aligned} E(\epsilon_t|S_t = 0, S_{t-1} = j) &= E(\epsilon_t|\eta_t < a_j) = -\rho\frac{\phi(a_j)}{\Phi(a_j)}, \\ E(\epsilon_t|S_t = 1, S_{t-1} = j) &= E(\epsilon_t|\eta_t \geq a_j) = \rho\frac{\phi(a_j)}{1-\Phi(a_j)}. \end{aligned} \tag{23}$$

Accordingly, the conditional densities of r_t, given S_t and S_{t-1}, must be modified to

$$\begin{aligned} f(r_t|S_t = 0, S_{t-1} = j; \theta) &= \frac{\phi(r^*_{0,t})\Phi((a_j - \rho r^*_{0,t})/\sqrt{1-\rho^2})}{\sigma_0 p_{0j}}, \\ f(r_t|S_t = 1, S_{t-1} = j; \theta) &= \frac{\phi(r^*_{1,t})\Phi((a_j - \rho r^*_{1,t})/\sqrt{1-\rho^2})}{\sigma_1 p_{1j}}, \end{aligned} \tag{24}$$

where $p_{00} = 1 - \pi_0$, $p_{01} = \pi_0$, $p_{10} = \pi_1$, and $p_{11} = 1 - \pi_1$. It is worthwhile noting that Equations (23) and (24) collapse to Equations (21) and (22) when switching is not endogenous, i.e., $\rho = 0$. Hence, testing for endogenous switching can be achieved by testing if the correlation between ϵ_t and η_t, ρ, is zero or not, using such tests as a Wald or likelihood ratio test.

3. Empirical Results

Data examined in this paper was data that was the continuously compounded monthly returns on the value-weighted portfolio of the NYSE, Amex, and the Nasdaq stocks in excess of the one-month Treasury bill yield over the period of 1926-2005. [3] This data was taken from the Center for Research's Security Prices database.

Following the estimation strategy of Mayfield (2004), we estimated the model described in Section 2, first ignoring the endogeneity problem and then accounting for it. Table 1 reports the parameter estimates of the two-state Markov-switching mean and variance model in Panel A and those of the preference parameters of the equilibrium risk

[3] Mayfield (2004) examined a sample period from 1926-2000. However, the results did not differ qualitatively.

premium in Panel B. The column labeled Ignoring Endogeneity corresponds to the assumption that investors have perfect knowledge about the current volatility state. The column labeled Accounting for Endogeneity corresponds to relaxing the assumption in a way that investors must infer the current volatility state.

First of all, we find that the null hypothesis of no endogeneity, i.e., $\rho = 0$ is rejected. The ρ estimate is significantly smaller than zero: -0.4274 with standard error of 0.1074. In addition, the log likelihood ratio statistic for $\rho = 0$ is 11.4 while the 1% critical value is 6.64. Hence, it is reasonable to believe that investors have no ability to observe the current volatility state with certainty. This suggests that the Markov-switching model is subject to endogenous switching and the parameter estimates are inconsistent when the endogeneity is ignored. In fact, the estimate of μ_1 is highly sensitive to whether or not the endogeneity is accounted for. When the endogeneity is not accounted for, the estimate of μ_1 is -25.73%. In contrast, the estimate is only -4.65% when the endogeneity is accounted for. This indicates that the restrictive assumption about the investors' information set severely understates the mean excess returns during the high volatility state. Some preference parameter estimates are also sensitive to whether or not the endogeneity is accounted for. The estimate of γ, is 0.8067 when the endogeneity is not accounted for and is 1.3304 when accounted for. Hence, the assumption understates investors' attitude toward risk. The point estimate for the jump parameter J_0 is also understated by the assumption: -0.2936 for the exogenous switching case vs. -0.2171 for the endogenous switching case.

Table 2 shows the risk premium estimates based on the estimates given in Table 1. Panel A corresponds to the assumption that investors are able to observe the true volatility states. Panel B corresponds to relaxing this assumption. Two non-negligible differences are found between the two cases. First, the assumption understates the total risk premium. The risk premium is 6.05% when switching is assumed to be exogenous, and 8.30% when switching is endogenous. Second, the relative importance of the interstate risk premium to the total risk premium is overstated by the assumption. When endogeneity is ignored, the interstate risk premium accounts for 54.5% of the total unconditional mean risk premium. However, when endogeneity is controlled for, it is only 45% of the total unconditional mean risk premium. This difference emanates from the magnitude of the interstate risk premium relative to the intrastate risk premium for $S_t = 1$. When endogeneity is not accounted for, the intra- and interstate risk premiums are of similar magnitudes: 0.1037 vs. 0.0929. When it is controlled for, the intrastate risk premium, 0.1753, is almost twice as large as the interstate risk premium 0.1009.

4. Conclusion

Mayfield (2004) proposed a method for estimating the market risk premium in the ICAPM framework that accounted for changes in investment opportunities. He estimated a Markov-switching model. The market risk was characterized as a return volatility which made recurrent shifts between the low- and high volatility state. Shifts in investment opportunities were modeled as state changes. Investors were assumed to have perfect knowledge about the current state. They required a premium for the volatility they observed and also a premium to hedge a possible state change next period.

The risk premium estimates of Mayfield (2004) were invalid. The Markov-switching model that he estimated was subject to endogenous switching that was directly associated with the assumption about the investors' information set. We performed a statistical test for the endogeneity and re-estimated the market risk premium, dealing with the endogeneity problem. The strategy recently developed by Kim et al. (2005) was employed. For the analysis of the value-weighted portfolio of the NYSE, Amex, and the Nasdaq stocks in excess of the one-month Treasury bill yield over the period of 1926-2005, the hypothesis of no endogeneity was rejected. Failure to account for the endogeneity lead to an understatement of the total market risk premium and overstatement of the relative importance of the risk premium associated with shifts in investment opportunities.

References

[1] Guo, H. and Whitelaw, R., 2006. Uncovering the Risk-Return Relation in the Stock Market. Journal of Finance 61, 1433-1463.

[2] Hamilton, J.D., 1989. A New Approach to the Economic Analysis of Nonstationary Time Series Models. Journal of Econometrics 70, 127-157.

[3] Kim, C.J., Piger, J.M., and Startz, R., 2005. Estimation of Markov Regime-Switching Regression Models with Endogenous Switching. Federal Reserve Bank of St. Louis Working Paper.

[4] Mayfield, E.S., 2004. Estimating the Market Risk Premium. Journal of Financial Economics 73, 465-496.

[5] Merton, R.C., 1973. An Intertemporal Asset Pricing Model. Econometrica 41, 867-888.

[6] Merton, R.C., 1980. On Estimating the Expected Return on the Market: An Exploratory Investigation. Journal of Financial Economics 8, 323-361.

[7] 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.

Table 1. Parameter Estimates: 1926-2005

	Ignoring Endogeneity	Accounting for Endogeneity
Panel A: Time Series Parameters		
μ_0	0.1191	0.1064
	(0.0178)	(0.0173)
μ_1	-0.2573	-0.0465
	(0.1095)	(0.1248)
σ_0	0.1280	0.1286
	(0.0044)	(0.0042)
σ_1	0.3586	0.3630
	(0.0275)	(0.0267)
$1-\pi_0$	0.9797	0.9803
	(0.1578)	(0.1470)
$1-\pi_1$	0.8911	0.8902
	(0.2216)	(0.2091)
ρ		-0.4274
		(0.1074)
ll	1586.43	1592.13
Panel B: Preference Parameters		
γ	0.8067	1.3304
	(0.2889)	(0.5891)
J_0	-0.2936	-0.2171
	(0.0850)	(0.0811)
J_1	0.4156	0.2774
	(0.2057)	(0.1730)
K_0	-0.2470	-0.2458
	(0.1823)	(0.0895)
K_1	0.3280	0.3259
	(0.2044)	(0.1450)

Estimates of parameters of the time series model are in Panel A and those of the preference parameters in Panel B. The time series model is a two-state Markov-switching mean and variance model. Preference parameters are those from the equilibrium risk premium expression derived as a solution to a utility maximization problem. The column labeled Ignoring Endogeneity corresponds to the assumption that investors have perfect knowledge about the current volatility state. The column labeled Accounting for Endogeneity corresponds to relaxing this assumption. Continuously compounded monthly returns on the value-weighted portfolios of the NYSE, Amex, and Nasdaq stocks in excess of the one-month Treasury bill rate over the period of 1926-2005 were examined. ll indicates the log likelihood value and parentheses are standard errors.

Table 2. Risk Premium Estimates: 1926-2005

	Intrastate	Interstate	Total
Panel A: Ignoring Endogeneity			
$S_t = 0$	0.0132	0.0218	0.0350
	(0.0049)	(0.0146)	(0.0185)
$S_t = 1$	0.1037	0.0929	0.1857
	(0.0396)	(0.0753)	(0.1069)
Uncon. mean	0.0275	0.0330	0.0605
	(0.0083)	(0.0211)	(0.0269)
Panel B: Accounting for Endogeneity			
$S_t = 0$	0.0220	0.0264	0.0484
	(0.0098)	(0.0114)	(0.0170)
$S_t = 1$	0.1753	0.1009	0.2018
	(0.0847)	(0.0548)	(0.1162)
Uncon. mean	0.0453	0.0377	0.0830
	(0.0183)	(0.0152)	(0.0256)

This table reports the risk premium estimates based on Equation (1) and the parameter estimates given in Table 1. Panel A is when endogeneity is not accounted for and Panel B is when it is accounted for. The column labeled Intrastate shows within-state risk premiums. The heading, Interstate, indicates the risk premium required by investors when the economy enters the alternate volatility state. The Uncon. mean is the weighted average of risk premiums for two states with weights being steady-state probabilities.