



Helsinki  
Center  
of  
Economic  
Research

Discussion Papers

# Robustness of the Risk-Return Relationship in the U.S. Stock Market

Markku Lanne  
University of Helsinki, RUESEG and HECER

and

Jani Luoto  
University of Jyväskylä

Discussion Paper No. 168  
May 2007

ISSN 1795-0562

HECER – Helsinki Center of Economic Research, P.O. Box 17 (Arkadiankatu 7), FI-00014  
University of Helsinki, FINLAND, Tel +358-9-191-28780, Fax +358-9-191-28781,  
E-mail [info-hecer@helsinki.fi](mailto:info-hecer@helsinki.fi), Internet [www.hecer.fi](http://www.hecer.fi)

# Robustness of the Risk-Return Relationship in the U.S. Stock Market\*

## Abstract

In this paper, we study the risk-return relationship in monthly U.S. stock returns (1928:1—2004:12) using GARCH-in-Mean models. In particular, we consider the robustness of the relationship with respect to the omission of the intercept term in the equation for the expected excess return recently recommended by Lanne and Saikkonen (2006). The existence of the relationship is quite robust, but its estimated strength is dependent on the prior belief concerning the intercept. This is the case in particular in the first half of the sample period, where also the coefficient of the relative risk aversion is found to be smaller and the equity premium greater than in the latter half.

**JEL Classification:** C11, C22, G12.

**Keywords:** ICAPM model; relative risk aversion; GARCH-in-Mean model; Bayesian analysis.

Markku Lanne

Department of Economics,  
P.O. Box 17 (Arkadiankatu 7)  
FI-00014 University of Helsinki  
FINLAND

e-mail: [markku.lanne@helsinki.fi](mailto:markku.lanne@helsinki.fi)

Jani Luoto

School of Business and Economics  
P.O. Box 35  
FI-40014 University of Jyväskylä  
FINLAND

e-mail: [jani.luoto@econ.jyu.fi](mailto:jani.luoto@econ.jyu.fi)

\* Markku Lanne thanks the Research Unit on Economic Structures and Growth (RUESG) in the University of Helsinki funded jointly by the Academy of Finland, University of Helsinki, Bank of Finland, Nokia Group, and the Yrjö Jahnsen Foundation.

## 1. Introduction

A voluminous literature is devoted to empirically studying the relationship between the conditional mean and variance of stock returns. The starting point in most of the previous research is the following linear equation implied by Merton's (1973) Intertemporal Capital Asset Pricing Model (ICAPM),

$$E_{t-1}(r_t) = \mu_0 + \mu_1 \text{Var}_{t-1}(r_t),$$

where  $r_t$  is the expected excess market return and  $\mu_0$  and  $\mu_1$  are parameters. The slope coefficient  $\mu_1$  is expected to be positive. Theoretically the intercept term  $\mu_0$  should be zero, but in most previous empirical studies this restriction has not been imposed. The empirical evidence is mixed, with relatively little support for the ICAPM (see Ghysels et al., 2005 and the references therein). Typically  $\mu_1$  has been found insignificant, and in some cases, its estimate is even negative.

Most of the previous empirical work on the ICAPM is based on the GARCH-in-Mean (GARCH-M) model. Within this model, Lanne and Saikkonen (2006) recently showed that omitting the zero restriction in the conditional mean equation may lead to a considerable loss of power in the standard Wald test. This may explain why the slope coefficient has most often been found insignificant. Constraining the intercept term to zero, Lanne and Saikkonen (2006) were able to show a positive risk-return relationship in the U.S. stock market.

While restricting the intercept term in the equation for the conditional mean to zero produces reasonable empirical results in addition to being well motivated from the viewpoint of both economic theory and econometrics, it may be criticized on the grounds that it forces the expected stock return to be equal to the risk-free interest rate under the null hypothesis of zero impact of the conditional variance on the expected return. This is in contrast with the voluminous stock premium puzzle literature put forth by Mehra and Prescott (1985), according to which the average stock return, in the U.S. market in particular, has historically been excessively high in relation to the risk. Therefore, the results of Lanne and Saikkonen (2006) concerning the risk-return relationship should be checked for robustness to confirm that they are not driven by restricting the intercept term to zero, which seems to be in conflict with some previous empirical findings. To this end we make use of Bayesian methods that easily facilitate checking the effects of allowing the intercept parameter to deviate from zero on the detected risk-return relationship. Specifically, we estimate the model assuming a number of different prior beliefs to see how sensitive the estimate of the slope coefficient is. In the U.S. data, the finding of a positive risk-return relationship seems to be

relatively robust with respect to the prior. Furthermore, we find estimates of the coefficient of the relative risk aversion that accord with most of the previous literature.

The plan of the paper is as follows. In Section 2, the GARCH-M model and the implementation of the employed Bayesian estimation methods are discussed. In Section 3, the empirical results are reported. Finally, Section 4 concludes.

## 2. Econometric Methodology

Let us consider the following GARCH-M model for the excess stock return  $r_t$ ,

$$r_t = \mu_0 + \mu_1 h_t + h_t^{1/2} \varepsilon_t, \quad (1)$$

where  $\mu_0$  and  $\mu_1$  are parameters,  $\varepsilon_t$  is an independently and identically distributed random variable with mean zero and variance unity and  $h_t$  is the conditional variance. We assume that  $h_t$  follows the GARCH(1,1) process of Bollerslev (1986),

$$h_t = \beta_0 + \beta_1 h_{t-1} + \beta_2 h_{t-1} \varepsilon_{t-1}^2, \quad (2)$$

where the parameters  $\beta_0$ ,  $\beta_1$  and  $\beta_2$  satisfy the conditions  $\beta_0 \geq 0$ ,  $\beta_1 \geq 0$ ,  $\beta_2 > 0$  and  $\beta_1 + \beta_2 < 1$ . The GARCH (1,1) specification has almost invariably been found an adequate description of the dynamics of the conditional variance of stock returns (see, e.g., Hansen and Lunde (2005)). The model can be estimated by the method of maximum likelihood (ML). Various distributional assumptions concerning  $\varepsilon_t$  have been entertained in the previous literature. Because the returns typically exhibit high excess kurtosis, normality has been found inadequate, while the standardized Student's t distribution with a relatively small degrees-of-freedom parameter  $\nu$  has turned out to be quite satisfactory, and that is the assumption made in our empirical analysis. Assuming that  $\nu > 2$ , the likelihood function for a sample of  $T$  observations  $r = (r_1, r_2, \dots, r_T)'$  can be written as (Bollerslev, 1986)

$$l(\eta; r) = \prod_{t=1}^T \frac{c(\nu)}{\sqrt{h_t}} \left( 1 + \frac{u_t^2}{(\nu - 2)h_t} \right)^{-\frac{\nu+1}{2}}, \quad (3)$$

where  $u_t = h_t^{1/2} \varepsilon_t$ ,  $\eta = (\beta_0, \beta_1, \beta_2, \mu_0, \mu_1, \nu)'$  is the vector consisting of all the parameters, and

$$c(\nu) = \Gamma\left(\frac{\nu+1}{2}\right) / \left[ \Gamma\left(\frac{\nu}{2}\right) \sqrt{\pi(\nu-2)} \right]$$

with  $\Gamma(\cdot)$  the gamma function.

As discussed by Engle et al. (1987), the classical maximum likelihood estimator of  $\eta$  is consistent and asymptotically normally distributed if the model is correctly specified. This facilitates standard asymptotic inference, but as pointed out by Lanne and Saikkonen (2006), the standard Wald test has very low power if the model is misspecified in that the intercept term  $\mu_0$  is, in fact, equal to zero. Therefore, Lanne and Saikkonen (2006) recommended imposing this constraint if it is true. However, even though this restriction is implied by the theoretical model in our application of interest, it may not be consistent with the data, and the robustness of the results with respect to it needs to be checked. To this end, we suggest using Bayesian methods that easily allow for incorporating prior beliefs about the parameters. While restricting the intercept term to be equal to zero represents a very strong prior belief, other kinds of beliefs cannot easily be considered in the classical framework.

Most of the empirical work concerning (G)ARCH models is based on classical inference. However, at least Geweke (1989), Kleibergen and van Dijk (1993), Bauwens and Lubrano (1998, 2002), Nakatsuma (2000), Kaufmann and Fruhwirth-Schnatter (2002), Vrontos et al. (2000, 2003), and Bauwens et al. (2006) have employed the Bayesian framework.

The starting point of the Bayesian analysis of the GARCH-M model is the likelihood function (3) multiplied by the prior density function of the parameters,  $p(\eta)$ , yielding the posterior density:

$$q(\eta|r) \propto p(\eta)l(\eta;r) \tag{4}$$

The prior density reflects the researcher's prior beliefs concerning plausible parameter values. Our goal is to find out how the prior beliefs on  $\mu_0$  affect the estimate of  $\mu_1$  in the GARCH-M model, and this can be accomplished by comparing the posterior distributions based on a number of different priors.

The use of improper priors, i.e. priors that are not well-defined density functions, causes marginal likelihoods (which we use in model comparison) to be indeterminate, since the normalizing constants of these priors are not defined. To confront this problem and in order to avoid possible non-integrability of the posterior distribution we will use proper priors. Specifically, using the standard assumption on prior independence,  $p(\eta) = p(\beta_0) \cdot p(\beta_1) \cdot p(\beta_2) \cdot p(\mu_0) \cdot p(\mu_1) \cdot p(\nu)$  (see e.g. Zellner, 1971), we set  $\beta_0 \sim \text{logN}(-3, 4)$ ,  $\beta_1 \sim \text{logN}(-0.2, 4)$ ,  $\beta_2 \sim \text{logN}(-2.3, 4)$ ,  $\mu_0 \sim \text{N}(0, \sigma_\mu)$ ,  $\mu_1 \sim \text{N}(0, 10)$ , and  $\nu \sim \text{Exp}(0.1)$ , where  $\text{N}(m, s)$  and  $\text{logN}(m, s)$  denote the normal and log normal distributions with mean  $m$  and standard deviation  $s$ , respectively, and  $\text{Exp}(\tau)$  is an exponential distribution with rate  $\tau$ . The joint prior density function  $p(\eta)$  in Equation (4) is, hence, a product of univariate normal, log normal and exponential distributions. This choice of prior reflects the fact that the variance parameters  $\beta_0$ ,  $\beta_1$  and  $\beta_2$  should be nonnegative. Moreover, as was shown by Bauwens and Lubrano (1998), sufficient prior information is needed on the Student-t degree of freedom parameter  $\nu$  to force the posterior to tend to zero quickly enough at the tail, in order to be integrable. We decided to follow Geweke (1993) in using an exponential density for the prior of  $\nu$ . Note that the prior means of the parameters are chosen on the basis of previous empirical studies, especially that of Lanne and Saikkonen (2006). In addition, with these values of prior standard deviations, marginal priors turned out to be practically noninformative.

Since the posterior density function (p.d.f.) in Equation (4) and none of its full conditionals are in the form of any standard p.d.f, we used a version of the random walk Metropolis algorithm for Markov Chain Monte Carlo (MCMC) to generate a Monte Carlo sample from it. The algorithm uses a multivariate normal distribution for the jump distribution on changes in the transformed parameters  $\theta = (\ln(\beta_0), \ln(\beta_1), \ln(\beta_2), \mu_0, \mu_1, \nu)'$ . We apply logarithmic transformations to the parameters  $\beta_0$ ,  $\beta_1$  and  $\beta_2$  to obtain approximate normality of their marginal posteriors, which makes the posterior simulations more efficient. The inverse of the Hessian of the log posterior density at the posterior mode, scaled by the factor  $2.4^2/6$ , is used to obtain an optimal covariance matrix of the multivariate normal jump distribution; see e.g. Gelman et al. (2004).

As we are interested in the effect of the prior distribution, several models based on different priors are considered. These models are compared by means of Bayes factors and estimated average discrepancies. A Bayes factor is defined to be the ratio of the marginal likelihoods of the two competing models, while the estimated average discrepancy approximates the posterior expectation of the deviance between the data and the model. The latter is used as an alternative approach, since

the use of Bayes factors is based on the assumption that one of the competing models is true, which may not be valid; see Gelman et al. (2004).

The Bayes factor is based on the marginal likelihood

$$q(r|M_k) = \int L(\eta_k; r, M_k) p(\eta_k|M_k) d\eta_k$$

of each of the models  $M_k$  ( $k = 1, 2, \dots, K$ ). It is estimated from the simulated independent posterior sample using the reciprocal importance estimator with a multivariate normal importance density; see Gelfand and Dey (1994). Given these marginal likelihoods, the Bayes factors can be computed using the formula

$$\hat{B}_{ij} = \frac{\hat{q}(r|M_i)}{\hat{q}(r|M_j)}, \quad (5)$$

where the hat indicates an estimator. As Miazhyńska and Dorffner (2004) show, for GARCH-type models the accuracy of the reciprocal importance estimator is statistically equal to more complex estimators, such as the bridge sampling estimator, Chib's candidate's estimator and the reversible jump MCMC estimator,. A slight modification of Jeffreys' (1961) scale of evidence for interpreting Bayes factors is given in Table 1; see Wasserman (1997).

Table 1. Jeffreys' scale of evidence for Bayes factors

Bayes Factor	Interpretation
$B_{ij} < 1/10$	Strong evidence for $M_j$
$1/10 < B_{ij} < 1/3$	Moderate evidence for $M_j$
$1/3 < B_{ij} < 1$	Weak evidence for $M_j$
$1 < B_{ij} < 3$	Weak evidence for $M_i$
$3 < B_{ij} < 10$	Moderate evidence for $M_i$
$B_{ij} > 10$	Strong evidence for $M_i$

Finally, the estimated average discrepancy is defined as

$$\hat{D}_{avg}(r) = \frac{1}{N} \sum_{i=1}^N D(r, \eta^i), \quad (6)$$

where  $r = (r_1, r_2, \dots, r_T)'$  are the data,  $D(r, \eta) = -2 \log l(r; \eta)$  is the ‘deviance’, and  $\{\eta^i\}_{i=1}^N$  a sample from the posterior ( $N$  is the sample size). In the limit, as the sample size tends to infinity, the model with the lowest expected deviance will have the highest posterior probability.

### 3. Empirical Results

In this section, we apply the approach of Section 2 to U.S. stock returns. The emphasis is on demonstrating the effects of the prior distribution of the intercept term  $\mu_0$  in the conditional mean equation (1) of the GARCH-M model on the conclusions. In particular, as discussed in Section 2, we assume a zero mean normal prior distribution for  $\mu_0$  and consider several alternative plausible values for the prior variance of  $\mu_0$  to see how the tightness of the prior assumption  $\mu_0 = 0$  affects the estimation results. Specifically, to study the effect of the prior of  $\mu_0$  on the estimate of  $\mu_1$ , we let the prior standard deviation  $\sigma_\mu$  of  $\mu_0$  vary from  $6^{-1} \times 0.003$  to  $10 \times 0.003$ , where 0.003 is the standard error of the maximum likelihood estimate of  $\mu_0$  reported by Lanne and Saikkonen (2006). In practice,  $\sigma_\mu = 10 \times 0.003$  yields a noninformative prior distribution for  $\mu_0$ , while  $\sigma_\mu = 6^{-1} \times 0.003$  is small enough to force the posterior mean of  $\mu_0$  to be close to zero. Moreover, as explained in Section 2, the model fits of these competing models are compared by means of Bayes factors and estimated average discrepancies.

The data set consists of the monthly excess return on the value-weighted CRSP index (in excess of the 3-month Treasury bill rate) from 1928:1 to 2004:12.<sup>1</sup> The same data from the period 1928:1 – 2000:12 were used by Lanne and Saikkonen (2006). In addition to the entire sample, models are estimated for the 1928:1 – 1966:6 and 1966:7 – 2004:12 subsample periods of equal length. This serves as a check of robustness and parameter constancy. Lanne and Saikkonen (2006) found the estimate of the slope coefficient  $\mu_1$  to be relatively stable once the intercept term  $\mu_0$  is restricted to zero, while the estimates from the different sample periods deviated considerable when  $\mu_0$  was unrestricted. Moreover, results in the empirical finance literature suggest that the coefficient of relative risk aversion has increased, while the equity premium puzzle has diminished since the 1960’s (see, e.g., Fama and French, 2002 and the references therein). Therefore, it is interesting to see whether our modeling approach leads to similar conclusions.

---

<sup>1</sup> The data are downloaded from Kenneth French’s homepage.



In estimating the models by the MMCMC algorithm, we use 400,000 draws, discarding the first 40,000 as a burn-in period. As a convergence check, three chains with different starting values are simulated. For each chain, we pick out every 90<sup>th</sup> draw from those included, to achieve a nearly independent sample. Thus, our final results are based on 12,000 draws. The potential scale reduction factor of Gelman and Rubin (1992) was between 1 and 1.01 for each parameter. The multivariate version of Gelman and Rubin's diagnostic, proposed by Brooks and Gelman (1997), was 1 in all cases. Finally, the frequency of accepted jumps was between 0.18 and 0.26.<sup>2</sup>

The estimation results are presented in Table 2, in the topmost panel for the entire sample and in the lower panels for the two subsample periods. Each column corresponds to a different prior distribution of the intercept term  $\mu_0$ , with the standard deviation ranging from  $10 \times 0.003$  to  $6^1 \times 0.003$ . As far as the GARCH parameters  $\beta_0$ ,  $\beta_1$  and  $\beta_2$  and the degrees-of-freedom parameter  $\nu$  are concerned, their posterior medians are insensitive to the prior. The parameters  $\beta_1$  and  $\beta_2$  are also accurately estimated, and the persistence in the volatility of the excess returns is high in all samples although the persistence tends to be somewhat lower in the latter subsample period. The estimated degrees of freedom indicate excess kurtosis, as expected. These findings are in accordance with the previous empirical results.

The models are compared by means of the Bayes factors and estimated discrepancy criteria. The Bayes factors reported in Table 1 are based on the benchmark of the model with the largest prior variance of the intercept term ( $10 \times 0.003$ ). In the entire sample, the estimated average discrepancy criterion and Bayes factors seem to favor models with relatively large prior variance of the intercept term  $\mu_0$ . In other words, the models allowing the intercept to deviate from zero, fit the data better. The Bayes factors lend the strongest support to the model with the prior standard deviation of the intercept term equal to  $2 \times 0.003$ . This model gives moderate evidence in favor of a positive risk-return relationship. The posterior probability of  $\mu_1$  being positive is above 90%. Note that, the posterior medians of  $\mu_1$  vary considerably between the models. In the preferred model, the median

---

<sup>2</sup> In the sub-sample 1966/7-2004/12 we use an adaptive Metropolis algorithm because the covariance matrix estimate based on local behaviour of the posterior at its highest peak turned out to give too optimistic a view of precision and thus failed to yield an efficient covariance matrix for the normal jump distribution. Specifically, we use a uniform prior between 0 and 1 for  $\beta_0$ ,  $\beta_1$  and  $\beta_2$  and the following adaptive random walk Metropolis algorithm: We first simulate 10,000 draws using a diagonal covariance matrix with diagonal entries 0.00001 in the jump distribution. We then use these draws to estimate the posterior covariance matrix of  $\eta$  and scale it by the factor  $2.4^2/6$ . We continue by simulating 10,000 draws and calculate a more accurate covariance matrix for  $\eta$ . We repeat this 2 – 5 times. Finally, we run 400,000 draws for three simulated chains using separate starting values and pick out every 90<sup>th</sup> draw after excluding the first 40,000 as a burn-in period. The reader should note that this choice of the priors of the  $\beta$ 's has no practical effect on the results; it only appears to be computationally more convenient to work with uniform rather than log-normal priors here.

equals 1.206 which is clearly smaller than the estimate obtained by Lanne and Saikkonen (2006). However, such a small value is better in accordance with the results that Ghysels et al. (2005) obtained using the far more sophisticated and data-demanding MIDAS method, as well as estimates of the coefficient of the relative risk aversion in the previous literature. Interestingly, the models with prior standard deviation of  $\mu_0$  above  $2^{-1} \times 0.003$  imply marginal posterior distributions of  $\mu_0$  with high probability of positive values, which is consistent with the equity premium puzzle.

In the first subsample period, the preferred model is the one with the prior variance of  $\mu_0$  equal to  $4 \times 0.003$  according to the estimated discrepancy criteria and Bayes factors. The posterior median of  $\mu_1$  is as low as 0.506. In this period, the results lend only weak support to the ICAPM, since only about 70% of the posterior mass of  $\mu_1$  lies above zero.

The results for the latter subsample period suggest that the intercept term is zero in accordance with the ICAPM. The most preferred model is the one with standard deviation of  $\mu_0$  equal to  $6^{-1} \times 0.003$ , and, in this model, the coefficient of relative risk aversion is positive with high probability. All the models imply a considerably larger coefficient of relative risk aversion than in the first subsample period. There is also surprisingly little variation in the posterior medians of  $\mu_1$  between the models. The differences between the subsample periods are consistent with the findings in the previous finance literature. In particular, the estimates of Ghysels et al. (2005) exhibit a similar change in the value of the coefficient of relative risk aversion. The fact that the intercept term was positive in the first but not in the second subsample period, probably reflects the moderation of the equity premium puzzle reported in the recent literature.

In sum, our results lend support to the robustness of the positive risk-return relationship in the U.S. stock market data, albeit the evidence is somewhat weaker in the first subsample period. In other words, this conclusion is shown not to be affected by (falsely) restricting the intercept term in the equation for the expected excess return equal to zero, when its true value is close to zero. The estimates of the key parameter, however, were shown to be affected by imposing this restriction if untrue, as was the case in the full sample and the first subsample periods. As shown by Lanne and Saikkonen (2006), for power considerations, the restriction is needed in classical inference, and our analysis can be seen as a robustness check of the validity of their empirical results.

Table 2. Estimation results.

Estimated model									
$r_t = \mu_0 + \mu_1 h_t + h_t^{1/2} \varepsilon_t; h_t = \beta_0 + \beta_1 h_{t-1} + \beta_2 h_{t-1} \varepsilon_{t-1}^2$									
$E(\sigma_{\mu_0})$	10×0.003	8×0.003	6×0.003	4×0.003	2×0.003	0.003	2 <sup>-1</sup> ×0.003	4 <sup>-1</sup> ×0.003	6 <sup>-1</sup> ×0.003
<b>1928:1-2004:12</b>									
$\beta_0$	1e-4 (4×1e-5)	1e-4 (4×1e-5)	1e-4 (4×1e-5)	1e-4 (4×1e-5)	1e-4 (4×1e-5)	1e-4 (5×1e-5)	1e-4 (5×1e-5)	1e-4 (5×1e-5)	1e-4 (5×1e-5)
$\beta_1$	0.830 (0.033)	0.830 (0.033)	0.830 (0.033)	0.830 (0.033)	0.830 (0.033)	0.828 (0.033)	0.826 (0.035)	0.823 (0.035)	0.823 (0.035)
$\beta_2$	0.127 (0.028)	0.126 (0.029)	0.127 (0.028)	0.127 (0.029)	0.127 (0.028)	0.125 (0.027)	0.121 (0.027)	0.118 (0.026)	0.117 (0.026)
$\mu_0$	0.007 (0.002)	0.007 (0.002)	0.007 (0.002)	0.007 (0.002)	0.006 (0.002)	0.005 (0.002)	0.002 (0.001)	6×1e-4 (7×1e-4)	3×1e-4 (5×1e-4)
$\mu_1$	0.886 (0.977)	0.880 (0.975)	0.895 (0.960)	0.951 (0.953)	1.206 (0.940)	1.838 (0.865)	2.713 (0.752)	3.205 (0.664)	3.328 (0.638)
$\nu$	8.124 (2.188)	8.135 (2.206)	8.141 (2.149)	8.177 (2.178)	8.261 (2.310)	8.475 (2.364)	8.740 (2.440)	8.805 (2.510)	8.862 (2.532)
$P(\mu_0 > 0)$	99.9%	99.9%	99.9%	99.9%	99.7%	99.1%	93.3%	79.3%	71.0%
$P(\mu_1 > 0)$	82.4%	82.1%	82.0%	84.0%	90.3%	98.6%	100%	100%	100%
$\hat{B}_{1j}$	-	1.65	0.64	0.49	0.41	0.82	2.82	5.64	6.89
$\hat{D}_{avg}(Y)$	-3044	-3044	-3044	-3044	-3044	-3043	-3039	-3037	-3036
<b>1928:1-1966:6</b>									
$\beta_0$	1e-4 (5×1e-5)	1e-4 (5×1e-5)	1e-4 (5×1e-5)	1e-4 (5×1e-5)	1e-4 (5×1e-5)	1e-4 (5×1e-5)	1e-4 (6×1e-5)	1e-4 (6×1e-5)	1e-4 (6×1e-5)
$\beta_1$	0.817 (0.041)	0.817 (0.041)	0.817 (0.041)	0.817 (0.040)	0.819 (0.041)	0.817 (0.042)	0.816 (0.043)	0.815 (0.044)	0.814 (0.045)
$\beta_2$	0.157 (0.045)	0.158 (0.044)	0.159 (0.044)	0.158 (0.044)	0.157 (0.044)	0.155 (0.043)	0.148 (0.041)	0.145 (0.040)	0.144 (0.040)
$\mu_0$	0.011 (0.003)	0.010 (0.003)	0.010 (0.003)	0.010 (0.003)	0.009 (0.003)	0.006 (0.002)	0.002 (0.001)	0.001 (0.001)	3e-4 (5×1e-4)
$\mu_1$	0.392 (0.949)	0.412 (0.940)	0.429 (0.930)	0.506 (0.935)	0.820 (0.916)	1.562 (0.872)	2.400 (0.814)	2.823 (0.773)	2.906 (0.772)
$\nu$	8.7838 (4.275)	8.744 (4.353)	8.690 (4.293)	8.907 (4.511)	9.227 (4.769)	10.240 (5.568)	11.043 (6.080)	11.355 (6.108)	11.409 (6.657)
$P(\mu_0 > 0)$	100.0%	100.0%	100.0%	100.0%	100.0%	99.7%	95.6%	81.8%	73.0%
$P(\mu_1 > 0)$	66.5%	67.5%	67.8%	70.4%	82.6%	96.8%	100.0%	100.0%	100.0%
$\hat{B}_{1j}$	-	0.82	0.66	0.56	0.71	4.56	31	78	78
$\hat{D}_{avg}(Y)$	-1481	-1481	-1481	-1481	-1481	-1478	-1473	-1469	-1468
<b>1966:7-2004:12</b>									
$\beta_0$	5e-4 (4×1e-4)	5e-4 (4×1e-4)	5e-4 (4×1e-4)	5e-4 (4×1e-4)	4e-4 (4×1e-4)	4e-4 (4×1e-4)	4e-4 (4×1e-4)	4e-4 (4×1e-4)	4e-4 (4×1e-4)
$\beta_1$	0.677 (0.199)	0.679 (0.199)	0.684 (0.195)	0.687 (0.194)	0.693 (0.190)	0.696 (0.187)	0.701 (0.183)	0.702 (0.185)	0.700 (0.186)
$\beta_2$	0.094 (0.043)	0.094 (0.044)	0.096 (0.044)	0.098 (0.044)	0.102 (0.044)	0.105 (0.045)	0.105 (0.044)	0.106 (0.044)	0.106 (0.045)
$\mu_0$	-0.003 (0.009)	-0.003 (0.009)	-0.002 (0.008)	-0.002 (0.007)	-9e-4 (0.005)	-2e-4 (0.003)	-5e-5 (0.001)	-2e-5 (0.001)	-4e-6 (0.001)
$\mu_1$	4.734 (4.440)	4.741 (4.247)	4.507 (3.876)	4.335 (3.412)	3.883 (2.416)	3.615 (1.599)	3.522 (1.196)	3.495 (1.031)	3.515 (1.012)
$N$	8.321 (3.599)	8.313 (3.410)	8.264 (3.435)	8.302 (3.373)	8.314 (3.412)	8.267 (3.320)	8.326 (3.315)	8.315 (3.399)	8.346 (3.337)
$P(\mu_0 > 0)$	36.6%	36.6%	38.1%	38.6%	43.2%	46.5%	48.2%	48.6%	49.4%
$P(\mu_1 > 0)$	90.0%	90.0%	90.7%	91.7%	95.4%	99.9%	99.9%	100.0%	100.0%
$\hat{B}_{1j}$	-	0.58	0.54	0.35	0.25	0.22	0.21	0.21	0.19
$\hat{D}_{avg}(Y)$	-1559	-1559	-1559	-1559	-1559	-1560	-1560	-1560	-1560

The posterior median and standard deviation (in the parenthesis) are shown in the table.

## 4. Conclusion

In this paper, we have studied the robustness of the risk-return relationship in the U.S. stock market with respect to the inclusion of an intercept term in the equation for the expected return in a GARCH-M model typically used in this context. Interpreting this model as the empirical counterpart of Merton's (1973) ICAPM model, there should be no intercept, while the well-documented presence of the so-called equity premium puzzle in U.S. stock returns suggests that omitting it might seriously deteriorate the fit of the model and affect the conclusions concerning the risk-return relationship. Moreover, according to the econometric results of Lanne and Saikkonen (2006), the unnecessary inclusion of an intercept leads to considerably loss of power in the significance tests of this relationship. The recommendation of these authors is thus to adopt the prior belief of no intercept implied by Merton's (1973) theoretical model. However, in their classical framework it is not easy to assess the impact of imposing this belief, whereas we are able to do that following the Bayesian approach. Our general idea is to estimate the model based on a number of different prior beliefs concerning the intercept term and compare the properties of the resulting estimates of the parameter measuring the strength of the risk-return relationship.

Three major conclusions stand out. First, the existence of a risk-return relationship is fairly robust in that it does not strongly depend on the choice of the prior distribution of the intercept, although the evidence is somewhat stronger in the latter half of the sample period. Second, there is some evidence in favor of a positive intercept term in the first half of the sample period, while it can quite safely be assumed to be zero in the latter half. This is in accordance with the findings in the previous literature of a diminishing equity premium. Third, although the existence of the relationship is robust, the size of the estimated parameter measuring its magnitude, which can be interpreted as the coefficient of relative risk aversion of the representative investor, is, to some extent, dependent on the specified prior of the intercept. This is particularly the case in the first half of the data. The estimates of the coefficient of relative risk aversion are overall smaller than those obtained by Lanne and Saikkonen (2006), but well in accordance with the previous literature in general. Furthermore, risk aversion is found to be greater in the latter half of the sample period, as also documented by the previous research.

## References

- Bauwens, L., and M. Lubrano (1998), Bayesian inference on GARCH models using Gibbs sampler, *Econometrics Journal*, 1, 23 – 46.
- Bauwens, L., and M. Lubrano (1998), Bayesian option pricing using asymmetric GARCH models, *Journal of Empirical Finance*, 9/3, 321-344.
- Bauwens, L., A. Preminger, and J. Rombouts (2006), Regime Switching GARCH Models, Core Discussion Paper 2006/11.
- Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307 – 327.
- Brooks, S.P., and A. Gelman (1997), General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7, 434 – 455.
- Engle, R.F., D.M. Lilien, and R.P. Robins (1987), Estimating time-varying risk premia in the term structure: the ARCH-M model. *Econometrica* 55, 391 – 407.
- Fama, E.F., and K.R. French (2002), The equity premium. *Journal of Finance* 57, 637 – 659.
- Gelfand, A., and D. Dey (1994), Bayesian model choice: Asymptotic and exact calculations, *Journal of Royal Statistical Society, Ser. B* 56, 501 – 514.
- Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin (2004), *Bayesian Data Analysis*, 2<sup>nd</sup> edition, Chapman & Hall/CRC.
- Gelman, A., and D.B. Rubin (1992), Inference from iterative simulation using multiple sequences, *Statistical Science*, 7, 457 – 511.
- Geweke, J. (1989), Exact predictive densities in linear models with ARCH disturbances, *Journal of Econometrics*, 40, 63-86.
- Geweke, J. (1993), Bayesian treatment of the independent student-t linear model, *Journal of Applied Econometrics*, 8, 19 – 40.
- Ghysels, E., P. Santa-Clara, and R. Valkanov (2005), There is a risk-return tradeoff after all. *Journal of Financial Economics* 19, 3 – 29.
- Hansen, P.R., and A. Lunde (2005), A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics* 20, 873 – 889.
- Jeffreys, H. (1961). *Theory of Probability*, 3rd edition, Oxford University Press, Oxford.
- Kaufmann, S. and S. Fruhwirth-Schnatter (2002). Bayesian analysis of switching ARCH models, *Journal of Time Series Analysis*, 23, 425–458.
- Kleibergen, F. and H. van Dijk (1993). Non-stationarity in GARCH models: a bayesian analysis, *Journal of Applied Econometrics*, 8, 41–61.

Lanne M., and P. Saikkonen (2006), Why is it so difficult to uncover the risk-return tradeoff in stock returns? *Economic Letters*, 92, 118 – 125.

Mehra, R., and E.C. Prescott (1985), The equity premium: A puzzle. *Journal of Monetary Economics* 15, 145 – 161.

Merton, R.C. (1973), An intertemporal capital asset pricing model. *Econometrica* 41, 867 – 887.

Miazynskaia, T. S., and G. Dorffner (2004), A comparison of Bayesian model selection based on MCMC with an application to GARCH-type models, *Statistical Papers*, 47, 525-549.

Nakatsuma, T. (2000), Bayesian analysis of ARMA-GARCH models: a Markov chain sampling approach, *Journal of Econometrics*, 95, 57–69.

Vrontos, D., P. Dellaportas and D.N. Politis (2000), Full Bayesian inference for GARCH and EGARCH models, *Journal of Business and Economics Statistics*, 18, 187-198.

Vrontos, D., P. Dellaportas and D.N. Politis (2003), A full-factor multivariate GARCH model, *The Econometrics Journal*, 6, 312-334.

Wasserman, L. (1997). Bayesian model selection and model averaging, *Technical report*, Statistics Department, Carnegie Mellon University.

Zellner, A., (1971), *An Introduction to Bayesian Inference in Econometrics*, J. Wiley and Sons, Inc., New York.