**Cardiff Economics
Working Papers**

GuangJie Li

*The Horizon Effect of Stock Return Predictability and Model
Uncertainty on Portfolio Choice: UK Evidence*

E2009/4

The Horizon Effect of Stock Return Predictability and Model Uncertainty on Portfolio Choice: UK Evidence

Corresponding address:

Guangjie LI

Aberconway Building

Cardiff Business School

Cardiff CF10 3EU

United Kingdom

Phone: +44-292-087-4833

Email: LIGJ@cf.ac.uk

Abstract   We study how stock return's predictability and model uncertainty affect a rational buy-and-hold investor's decision to allocate her wealth for different lengths of investment horizons in the UK market. We consider the FTSE All-Share Index as the risky asset, and the UK Treasury bill as the risk free asset in forming the investor's portfolio. We identify the most powerful predictors of the stock return by accounting for model uncertainty. We find that though stock return predictability is weak, it can still affect the invesor's optimal portfolio decision over different investment horizons.

Key words   stock return predictability, portfolio choice,Bayesian Model Averaging,SUR model

JEL classification   C11,G11,G15

1 Introduction

Finance advisors often tell people with long investment horizon to invest more into stocks than bonds. Fund managers will recommend different portfolios to investors with different investment horizons. For example, they may recommend some stock shares for long term investment and some others just for short term. Such ideas to allocate wealth among different financial products according to the length of investment horizon have been challenged by academics. Early work about horizon effect can be seen in Samuelson (1969) and Merton (1969). They proved that if the return of a risky asset is unpredictable, rational investors should choose the same portfolio regardless of the length of their investment. More recently, Samuelson (1989) and Samuelson (1990) readdressed the irrelevance of the length of investment horizon in portfolio management.

The absence of horizon effect primarily relies on the assumption that the return of the risky asset is unpredictable. However, there are also studies showing that return predictability can affect investor's optimal portfolio decision, see, for example, Kandel and Stambaugh (1996), Barberis (2000) and Xia (2001). To add more valuable insight into this debate, it is important to understand the nature of stock market inefficiencies, which is closely related to the question of whether stock return is predictable or not. Though most studies using daily or weekly data find very little evidence of predictability in terms of low R-squares or low p-values, many academic investigations into monthly data suggest some variables may have the ability to explain the movements in stock expected return. Fama and French (1988) reported that apart from dividend yields, past stock return in the US market can predict 40 percent of future return over the long run. Fama and French (1993) then identified five common risk factors in explaining the return of stocks and bonds. Consistent with Fama and French's results, Kothari and Shanken (1997) also found that book-to-market ratio (B/M) has predictive power. However, these studies have invited criticisms from other scholars. Hodrick (1992) and Goetzmann and Jorion (1993) argued that many findings based on long-

horizon return regressions may be inappropriate due to problems such as data snooping[1], nonrobustness of test statistics and poor small sample properties of the inference method.

Such controversy about stock return predictability can be better explained from two aspects. First, though there are many articles addressing the issue of stock return predictors, there is little consensus on what the important conditioning variables are. This issue can be regarded as model uncertainty, which here refers to the uncertainty about the true explanatory variables for the stock return, see, for example, Brennan and Xia (1999). Secondly, even if one believes to have found the correct set of predictors, the predictive relationship between stock return and the predictors cannot be estimated with certainty due to limited sample size. In other words, it is not possible for us to identify the true values of the parameters for our model in real life application. Parameter uncertainty or estimation risk can have an important effect on investor's optimal portfolio choice, see Bawa *et al.* (1979) and Barberis (2000). By taking into account both parameter and model uncertainty, one could better answer the question of whether stock return is predictable or not. Cremers (2002) and Avramov (2002) both used Bayesian model averaging (BMA) to consider such uncertainty and found that the BMA method, which averages the estimates from all potential models according to their posterior probabilities, can provide better forecasts of stock return than those selected based on certain criterion. The above studies are based on the US stock market. Relevant research on the UK market can be seen in Pesaran and Timmermann (1995), in which they employed recursive regression method to select a best single model based on certain information criterion to make out-of-sample forecasts. Though they acknowledged there is uncertainty about which model best forecast stock returns over time, they did not address this issue explicitly in their method.

In this paper we study the stock return predictability in the UK market by accounting for both parameter and model uncertainty. We then investigate the effect of such predictability on a rational investor's portfolio choice given different lengths of investment horizons. We find that the stock return predictability in the UK market is weak if we allow for model

uncertainty. Many explanatory variables are not as strong predictors as classical results suggest. Moreover, if we take account of the data generating processes (DGP) of the explanatory variables and allow them to be correlated with that of the stock return, the predicting power of these explanatory variables will decrease further. As for the horizon effect, we propose a computationally convenient statistic that can be used as a reference for how a rational buy-and-hold investor should adjust her optimal portfolio given different lengths of investment. We find that although the return predictability is weak, it still has a considerable effect on a rational buy-and-hold investor's portfolio choice as evidenced by different allocation proportion of wealth to risky asset over time given different initial information.

The paper proceeds as follows. Section 2 explains the asset allocation problem and the computation techniques used to solve it. Section 3 investigates the horizon effect when the risky asset's return is unpredictable. We look into the cases with and without parameter uncertainty and then propose a measure to capture the horizon effect. Section 4 studies the stock return predictability in the UK market by considering model uncertainty. Section 5 then examines the horizon effect of stock return predictability and model uncertainty. Finally Section 6 concludes.

2 The Asset Allocation Problem and the Calculation of the Optimal Portfolio

The basic economic model of the analysis consists of a risk averse investor, who allocates her wealth to either risk free (e.g. treasury bond) or risky asset (e.g. stock share) in order to maximize her utility function. This model has been studied by Kandel and Stambaugh (1996), Barberis (2000) and Avramov (2002) with a focus on the time horizon effects, i.e. how the investor will allocate her wealth given different lengths of investment horizons. Different from other studies, we will look into the horizon effect based on the UK data. Compared to Avramov (2002), we will take into account not only the effects of parameter and model

4

uncertainty, but also the interactions between the DGP of the return for the risky asset and those of its explanatory variables. Moreover, we will propose a computationally convenient statistic, which may shed some light on the behaviour of a rational investor when she has to choose between risky and risk free asset.

The investor's wealth preference is described by the constant relative risk-aversion power utility function (v) with the following form.

$$v(W) = \begin{cases} \frac{W^{1-A}}{1-A} & for \quad A > 0 \quad and \quad A \neq 1 \\ \ln W & for \quad A = 1 \end{cases} \tag{1}$$

where A is commonly referred to as the investor's coefficient of relative risk aversion and W denotes the investor's wealth. Without loss of generality, we assume the initial wealth of the investor is equal to one. Let us denote the rate of return of the risk free asset by r_f and the excess return of the risky asset over the risk free by r^2. For simplicity, we further assume that r_f is non-stochastic and only r is a random variable. Suppose the investor is going to hold the portfolio of the two assets from period T till period $T + \hat{T}$. At the end of her investment horizon, her cumulative excess return will be $R_{T+\hat{T}} = r_{T+1} + r_{T+2} + ... + r_{T+\hat{T}}$, which will also follow a certain distribution. If we assume the returns are continuously compounded and the investor allocates ω of her wealth to the risky asset, her total wealth at the end of the investment will be $(1 - \omega) \exp(\hat{T} r_f) + \omega \exp(\hat{T} r_f + R_{T+\hat{T}})$. The asset allocation problem for the investor is to solve

$$\max_{\omega} \int_{R_{T+\hat{T}}} \frac{\left[(1 - \omega) \exp(\hat{T} r_f) + \omega \exp(\hat{T} r_f + R_{T+\hat{T}}) \right]^{1-A}}{1 - A} p(R_{T+\hat{T}}) dR_{T+\hat{T}} \tag{2}$$

That is given a period of time, which is $\hat{T}$ periods long, the problem facing the investor is to choose ω to maximize her expected utility at the start of the investment, i.e. period T. Our study will focus on the investment horizon effect, i.e. the relationship between ω and $\hat{T}$. Note that it is generally impossible to obtain a closed form solution for (2) even if $p(R_{T+\hat{T}})$ is some standard density function. To solve the problem, Barberis (2000) restricted ω to

5

$[0, 1]$ and performed a grid search after simulating draws from $p(R_{T+\hat{T}})$ to integrate $R_{T+\hat{T}}$ out. Here we use a relatively convenient and possibly more efficient numerical method to tackle this problem. First we use Taylor expansion to approximate the power utility function around the mean of $R_{T+\hat{T}}$ to produce a polynomial of $R_{T+\hat{T}}$. We can choose the order of Taylor expansion to control the approximation accuracy. Then we obtain the moments of $R_{T+\hat{T}}$ analytically or by simulation and insert them into the polynomial to obtain a function of only ω. Finally we use a numerical routine to maximize the function.[3] In our application in later sections, we find that Taylor approximation with order around 10 could give us reasonably accurate results when $R_{T+\hat{T}}$ follows a normal or t distribution.

Next we discuss the force that may drive the horizon effect. Note that the demand for the risky asset in the investor's portfolio clearly hinges on how we set up the maximization problem and the constraints confronting the investor. However, it should be no surprise that the risky asset's return and its level of risk are the key factors. In other words, the first and the second moments of $R_{T+\hat{T}}$ should have an important role in determining the horizon effect. Note that the density function $p(R_{T+\hat{T}})$ will change with $\hat{T}$. Hence both the first and the second moments of $R_{T+\hat{T}}$ are functions of $\hat{T}$. We may be interested in knowing how fast the return changes relative to the change of risk. For example, if the risk of an asset increases with time, will the asset's expected return increase fast enough to counteract such effect so that the asset will still remain attractive to a rational investor? Here we propose the following expression which may help to answer this question:

$$MtoS = \frac{\frac{\partial \mu_{\hat{T}}}{\partial \hat{T}} \times \sigma_{\hat{T}}}{\frac{\partial \sigma_{\hat{T}}}{\partial \hat{T}} \times \mu_{\hat{T}}}, \tag{3}$$

where $\mu_{\hat{T}}$ and $\sigma_{\hat{T}}$ denote the mean and standard deviation of $R_{T+\hat{T}}$ respectively. The expression in (3) is no more than the ratio between the percentage rate of change of $R_{T+\hat{T}}$'s mean and standard deviation. It is similar in spirit to the Sharpe ratio and could provide a measure of the value of risk (in terms of the mean return) over time. In the following sec-

tions, we will illustrate how expression (3) is related to the investment horizon effect under different probability density functions of $R_{T+\hat{T}}$.

3 When the Excess Return is Unpredictable

Samuelson (1969) and Merton (1969) show that when stock return is not predictable, the optimal portfolio will be independent of wealth and all consumption-saving decisions in a multi-period portfolio rebalancing model. Different from Samuelson (1969), the distribution of the excess return will change with time in this paper. Barberis (2000) uses the US data and shows that the optimal portfolio is insensitive to investment time horizon if $R_{T+\hat{T}}$ is unpredictable and follows a normal distribution with mean and variance increasing linearly with time. In our empirical study, we will use the UK 3 month treasury bill rate as r_f. The excess return of the risky asset (r) is calculated as the return difference between the FTSE All-Share Index and r_f. Our data sample is from November 1978 up to September 2003, which includes 299 (T) observations of the FTSE All-Share Index. The mean of the excess rate of return (μ) of the FTSE index over T-bill in our sample is 0.4772%, while the sample standard deviation is $4.88\%(\sigma)$. We assume the excess return is unpredictable and follows a normal distribution as below,

$$r_t = \mu + \epsilon_t, \qquad \epsilon_t \sim IIDN\left(0, \sigma^2\right), \tag{4}$$

where μ is the mean of the stock excess return and σ^2 is the variance in the normal distribution. The cumulative excess return $R_{T+\hat{T}}$ will also be normal as the following,

$$R_{T+\hat{T}}|\mu, \sigma^2 \sim N(\mu_{\hat{T}}, \sigma_{\hat{T}}^2), \tag{5}$$

where $\mu_{\hat{T}} = \hat{T}\mu$, and $\sigma_{\hat{T}}^2 = \hat{T}\sigma^2$. As pointed out by Barberis (2000), if the investor ignores parameter uncertainty, i.e.taking the estimates of μ and σ from the past data as the true values of these parameters, the optimal holding proportion of the risky asset (ω) will not

change with time. It is easy to see that under such setup, $MtoS$ defined in (3) is equal to 2 and also independent of $\hat{T}$.

In our following studies we set $A = 5$ and $r_f = 0.3\%$, which is the last observation of the monthly rate of return of the 3-month T-bill. Here we study the horizon effect from one month to 5 years. By using the numerical method described in the previous section, the optimal holding proportion of the risky asset and the $MtoS$ defined in (3) are shown in the left column of Figure 1 for different investment lengths. We can see that ω is about 0.5 while $MtoS$ is 2. Both of them do not change with time. These results confirm the empirical findings of Barberis (2000) using the US data. We have just added $MtoS$ to analyze the relative change of the return and risk over time.

[Figure 1 here]

Next we turn our attention to the case when the investor no longer treats the estimates of μ and σ as their true values. In other words, the investor is now taking parameter uncertainty into account as termed by Barberis (2000). To model the parameters μ and σ in (5) as random variables, we adopt the Bayesian inference framework by assuming the joint distribution of μ and σ^2 follows a noninformative prior and hence their posterior follows a normal-gamma distribution.

$$p(\mu, \sigma^2) \propto \frac{1}{\sigma^2} \tag{6}$$

$$\mu|\sigma^2, D \sim N(\bar{r}, \frac{\sigma^2}{T}) \tag{7}$$

$$\sigma^2|D \sim IG\left((T-1)s^2, T-1\right) \tag{8}$$

where $\bar{r} = \frac{1}{T}\sum_1^T r_t = 0.4772\%$ and $s^2 = \frac{\sum_{t=1}^{T}(r_t - \bar{r})^2}{T-1} = 0.0023797$. Equation (7) and (8) show the posterior distributions of μ and σ^2, which are conditional on the data[4], denoted by D. Here σ^2 follows an inverted gamma distribution with degrees of freedom $T-1$. Its posterior mean

and variance are

$$E(\sigma^2|D) = \frac{\sum\limits_{t=1}^{T}(r_t - \bar{r})^2}{T-3} = 2.40 \times 10^{-3}, \tag{9}$$

$$Var(\sigma^2|D) = \frac{2\sum\limits_{t=1}^{T}(r_t - \bar{r})^2}{(T-3)^2(T-5)} = 5.51 \times 10^{-8}. \tag{10}$$

The conditional distribution of $R_{T+\hat{T}}$ is still given by (5) while the posterior distribution of $R_{T+\hat{T}}$ unconditional of μ and σ^2 now becomes

$$R_{T+\hat{T}}|D \sim t(\hat{\mu}_{\hat{T}}, \sigma^2_{\hat{T}}, T-1), \tag{11}$$

where $\hat{\mu}_{\hat{T}}$ denotes the mean parameter, which is equal to $\bar{r}\hat{T}$ and $\sigma^2_{\hat{T}}$ is the variance parameter equal to $s^2\hat{T}(1+\frac{\hat{T}}{T})$. With parameter uncertainty of μ and σ^2, it is equivalent as saying that $R_{T+\hat{T}}$ in (2) has a t density function with parameters described in (11). It can be seen that the mean and variance of $R_{T+\hat{T}}$ grow at different speeds as compared to (5). The ratio of the percentage rate of change between them ($MtoS$) is now the following,

$$MtoS = \frac{\frac{\partial \hat{\mu}_{\hat{T}}}{\partial \hat{T}} \times \hat{\sigma}_{\hat{T}}}{\frac{\partial \hat{\sigma}_{\hat{T}}}{\partial \hat{T}} \times \hat{\mu}_{\hat{T}}} = 2 - \frac{2\hat{T}}{T+2\hat{T}}. \tag{12}$$

Unlike the case of no parameter uncertainty, this ratio depends on $\hat{T}$ and is a decreasing function of $\hat{T}$, whose value is less than 2 unless $\hat{T}$ is 0. Barberis (2000) shows that the optimal holding proportion of the risky asset under parameter uncertainty will no longer be insensitive to the length of investment horizon. The interpretation could be that the longer is the investment horizon, the rational investor will become more doubtful about her initial estimation made at period T. Therefore her expected risk of the asset will grow faster with time than the case with no parameter uncertainty. The plots of the optimal portfolio and $MtoS$ are shown in the right column of Figure 1. We can see that both ω and $MtoS$ drop with the length of investment horizon. For ω, it falls by around 7% while $Mtos$ drops from 1.99 to 1.71. If we look at the expression of $MtoS$ in (12), we could find that it also involves

9

T, i.e. the original sample size. We may conjecture that the optimal portfolio may also be related to the sample size. Indeed, if we keep the values of all other parameters the same and just change T to its one tenth, the optimal ω will fall from 46% to 23% while $MtoS$ is from 1.93 to 1.2. This can be due to the fact that as the investor's estimation is based on smaller sample size, she will have less confidence in it and will feel that the asset is more risky in the long run. Hence the size of the drop of ω is much larger and the horizon effect is more pronounced. The interesting point here is that the ratio between the percentage rate of change of the excess return mean and standard deviation ($MtoS$) seems to be able to tell how a rational investor will behave given different lengths of investment horizons. However, in many applications, such as the one in the following sections, the first and second moments of $R_{T+\hat{T}}$ may not have closed forms, needless to say $MtoS$. Here we propose the following statistic, which can circumvent this problem and approximate the $MtoS$,

$$\hat{MtoS} = \frac{\ln(\mu_{\hat{T}}/\mu_{\hat{T}-1})}{\ln(\sigma_{\hat{T}}/\sigma_{\hat{T}-1})} \approx MtoS, \tag{13}$$

which is the ratio of the log differences between the contemporaneous expected mean and the one of one period earlier over its standard deviation counterpart[5]. The statistic approximates the instantaneous relative percentage change of expected return to that of risk. As we argued before, the $MtoS$ could be viewed as a measure of the economic value of risk in terms of return over time. While the optimal holding proportion of the risky asset is hard to calculate and depends on the setup of the maximization problem, such as what form the utility function takes and how risk averse is the investor, the statistic defined in (13) is easy to calculate and may provide a reference for the investor as to how attractive a particular portfolio is over time. We will apply this statistic to the subsequent sections where the analytical forms of the first and second moments of $R_{T+\hat{T}}$ are not available.

4 Whether Stock Return is Predictable or Not

4.1 Data and Some Statistical Results

All our data, except the Hoare Govett Smaller Companies Index, are from DataStream, covering the period from November 1978 to September 2003, altogether 299 observations (T). As before, we use r to denote the FTSE All-Share Index excess return, which is our dependent variable. The explanatory variables along with their short forms used in the analysis are shown in Appendix A.1. All of them are either business cycle variables or financial market variables suggested in the literature, which may possess explanatory power for excess return.

Consistent with the study by Pesaran and Timmermann (1995), we do not include the observation in October 1987, which is an outlier[6]. Figure 2 displays the monthly excess returns of the FTSE All-Share Index over our sample range. First sight suggests there do not seem to be any obvious patterns, such as autocorrelation. This can be confirmed in Figure 3[7]. The two parallel horizontal lines indicate the 95% confidence interval. We can see that all the autocorrelation coefficients up to twenty lags are well within the 95% confidence lines.

[Figure 2 here]

[Figure 3 here]

A rough idea about the extent to which the excess return can be predicted using different variables can be seen in the OLS regression results in Table 1 obtained by regressing the excess return on all other variables. The numbers in bold indicate they are significant at 10% level of significance. Such practice by regressing the excess return on all other variables could be subject to criticism such as data snooping and model misspecification. In the next subsection, we will use the BMA method to look into this issue.

[Table 1 here]

4.2 Bayesian Model Averaging in a Univariate Linear Model

The Efficient Market Hypothesis (EMH, e.g. Fama (1970)) states that in an efficient capital market, stock return is not predictable. Numerous empirical work has shown that the capital market is not efficient. While traditional asset pricing models, like CAPM, precludes the use of predictors in determining return, the literature of style investment, which studies investment return based on certain economic or accounting variables, has prospered in the past decades. Banz (1981) documented that small-cap stocks have historically outperformed large-cap stocks in the US by a margin that could not be explained by conventional measures of risk. Hence the capital size of a stock may help predict its return. Later influential work can be seen in Fama and French (1993), who documented five common risk factors in the returns of stocks and bonds (the whole market returns, firm size, book-to-market ratio, maturity risk and default risk). For the UK market, Pesaran and Timmermann (1995) found that in addition to dividend yield, several business cycle variables help to predict the excess return. Different variables can be seen in predicting returns in numerous other papers. The variables presented in Appendix A.1 are based primarily on these studies. Though there are many articles mentioning possible predictors, there is little consensus on what the most important conditioning predictors are.

Here we apply BMA techniques to a linear model to identify the most important predictors using the UK data. We assume the predictors and the dependent variable have a linear relationship and the disturbance term has no serial correlation and heteroscedasticity:

$$r_t = \underline{a}_p + B'_p x_{t-1,p} + \epsilon_{t,p}, \ \epsilon_{t,p} \sim i.i.d.N(0, \sigma_p^2), \tag{14}$$

where r stands for the excess return, and x stands for the set of predictors used, which do not include any lag terms of r. The subscript p is a model specific parameter, which implies that the parameters are different for different models. There are altogether 15 (K) possible predictors which may enter the regression to explain the excess return. The total number

of different models with different regressors, is $P = 2^{15}$. Each model, M_p, is described by a $K \times 1$ binary vector $\gamma = (\gamma_1, ..., \gamma_K)'$, where a one (zero) indicates the inclusion (exclusion) of a variable. We denote the sum of all elements in γ by k_p, which is the dimension of the column vector $x_{t-1,p}$. If we stack up all the observations for equation (14), then it can be written as

$$\mathbf{r} = \underline{a}_p \iota + \mathbf{X}_p B_p + \epsilon_p \qquad \epsilon_p \sim N(0, \sigma_p^2 \mathbf{I}) \tag{15}$$

where ι is a vector of ones, $\mathbf{X}_p = [x_{0,p}, x_{1,p}, x_{2,p}, ..., x_{T-1,p}]'$, and $\mathbf{r} = [r_1, r_2, r_3, ... r_T]'$.

The following analysis relies heavily on the bench mark prior developed by Fernandez *et al.* (2001). To implement their approach, we first reparameterize the intercept term $(\underline{a}_p)$ in the regression such that the new intercept term (a_p) is orthogonal to the slope (B_p) in the likelihood function, i.e. $\underline{a}_p = a_p - \frac{\iota' X_p B_p}{T}$. In doing so, we have changed (15) into

$$\mathbf{r} = a_p \iota + H \mathbf{X}_p B_p + \epsilon_p, \ \epsilon_p \sim N(0, \sigma_p^2 \mathbf{I}), \tag{16}$$

where $H = I_T - \frac{\iota \iota'}{T}$ is the demean matrix. The bench mark prior proposed by Fernandez *et al.* (2001) looks like the following.

$$p(a_p, \sigma_p^2) \propto \frac{1}{\sigma_p^2} \tag{17}$$

$$B_p | \sigma_p^2, a_p \sim N(0, \sigma_p^2 (g \mathbf{X}_p' H \mathbf{X}_p)^{-1}) \tag{18}$$

Here we use flat prior for the equation variance and the constant. For the slope vector B_p, we use the g prior designed by Zellner (1986). It substantially reduces the trouble of eliciting the values for too many hyperparameters by using the explanatory variables to specify the prior variance. The strength of the prior only depends on g. After extensive Monte Carlo

experiments, Fernandez *et al.* (2001) recommended choosing

$$g = \begin{cases} \frac{1}{T} & \text{if } T > K^2 \\ \\ \frac{1}{K^2} & \text{if } T \leq K^2 \end{cases} \tag{19}$$

where T stands for the sample size and K stands for the number of potential predictors. Note that g appears in the prior variance of the slope vector, which controls our confidence in the prior. The choice of g in (19) means we always prefer a more noninformative prior such that the variances for the slopes in (18) are bigger than the alternative. It can be shown that the posterior of B_p follows a multivariate t distribution with mean:

$$E(B_p|D, M_p) = \bar{B}_p = \frac{1}{g+1}(\mathbf{X}'_p H \mathbf{X}_p)^{-1}\mathbf{X}'_p H \mathbf{r}, \tag{20}$$

and covariance matrix:

$$Var(B_p|D, M_p) = \frac{\bar{v}\bar{s}_p^2}{\bar{v}-2}\bar{V}_p, \tag{21}$$

where $\bar{v} = T$ is the degrees of freedom and $\bar{v}\bar{s}_p^2 = \mathbf{r}'H\mathbf{r} - \frac{1}{1+g}\mathbf{r}'H\mathbf{X}_p(\mathbf{X}'_p H \mathbf{X}_p)^{-1}\mathbf{X}'_p H \mathbf{r}$. The marginal likelihood takes the following form:

$$p(D|M_p) \propto (\frac{g}{1+g})^{\frac{k_p}{2}}\left(\bar{v}\bar{s}_p^2\right)^{-\frac{T-1}{2}} \tag{22}$$

We can see that the marginal likelihood penalizes the models with a large number of regressors (k_p) since $\frac{g}{1+g}$ is less than 1. For our case there are $P = 2^{15}$ models. Given this model space, there is uncertainty about what is the correct model. Hence it makes sense to consider the parameters unconditional of the model space. This requires us to calculate the posterior model probability as shown in the following,

$$p(M_p|D) = \frac{p(D|M_p)p(M_p)}{p(D)} = \frac{p(D|M_p)p(M_p)}{\sum\limits_{p=1}^{P} p(D|M_p)p(M_p)}. \tag{23}$$

To specify the model prior, $p(M_p)$, it is possible to use a flat prior, which gives every model the same prior probability($\frac{1}{2^{15}}$ in our case). However, George (1999) notes that when many of the regressors in the regression are highly correlated, large subsets of models are essentially

14

equivalent. In other words, highly correlated regressors can be viewed as proxies for each other and they capture the same theory. If a flat prior is adopted on the model space, then excessive prior probability will be allocated to such similar models at the expense of some unique models. George (2001) suggests the following dilution model prior,

$$p(M_p) \propto |\mathbf{R}|\prod_{i=1}^{K}\pi^{\gamma_i}(1-\pi)^{\gamma_i} \tag{24}$$

where $\mathbf{R}$ is the correlation matrix of regressors included and π is the probability of including a variable. Note that if $\pi > \frac{1}{2}(\pi < \frac{1}{2})$, we prefer models with more (less) regressors. Here we set $\pi = \frac{1}{2}$. The determinant of the correlation matrix in the model prior serves to penalize the models with redundant regressors. We can see this by noting that $|\mathbf{R}| = \mathbf{1}$ when the regressors are orthogonal and $|\mathbf{R}|$ approaches 0 when the regressors become more collinear. Some explanatory variables exhibit high degree of correlation such as oil price, industrial production, monetary supply, treasury bill rate and dividend yield.[8]

The underlying logic of the BMA technique is that we should mix our estimates from different models based on their posterior model probabilities calculated from (23). Such practice can well account for model uncertainty. Let β denote the parameter of interest, such as the slope parameter or the predicted excess return. Leamer (1978) showed that unconditional on the model space, the posterior mean and variance of β can be calculated as:

$$E(\beta_i|D) = \sum_{p=1}^{P}I(\gamma_i = 1|M_p, D)p(M_p|D)E(\beta_i|M_p, D), \tag{25}$$

$$Var(\beta_i|D) = E(\beta_i^2|D) - E^2(\beta_i|D), \tag{26}$$

where $E(\beta_i^2|D) = \sum_{p=1}^{P}I(\gamma_i = 1|M_p, D)p(M_p|D)E(\beta_i^2|M_p, D)$. An investor may be interested in knowing how important the variables are in explaining the excess return. We therefore need to have a measure of the importance of the included regressor i unconditional of the

model space. The following posterior inclusion probability of variable i serves this purpose.

$$p(\gamma_i = 1|D) = \sum_{p=1}^{P} I(\gamma_i = 1|D, M_p)p(M_p|D) \qquad (27)$$

Table 2 shows the estimation results for the slope parameters. For comparison, we also report the prior inclusion probability for each regressor as implied by (24). We have put the variables with more than 10% inclusion probabilities in bold. These are the relatively powerful explanatory variables in our results. The variables with the highest inclusion probabilities in descending order are Dy, Inf, Smb, Jan and $M0$. If we compare the results to those in Table 1, we can see that only Dy, Jan and $M0$ are robust for both Bayesian and classical approaches while the variables of oil price, Tb and $Tbchng$ have relatively lower inclusion probability in contrast to their significant results without model averaging. Hence one should be more cautious of the significance of the latter set of explanatory variables.

[Table 2 here]

[Table 3 here]

Table 3 lists the top 10 models with the highest posterior model probabilities. The column headed by "model" list the regressors included for the particular model. The explanation of the variables can be found in Appendix A.1. We can see that the model with the highest posterior probability is the one without any explanatory variables. Moreover, the top 10 models are all parsimonious models with at most 3 regressors. Their posterior probabilities sum up to 0.49, while for the top 100 models out of 32768, the sum is 84%. All of the top 100 models have no more than 4 regressors, with 82 of them with less than 3. A point to note is that although the model without any explanatory variables has the highest posterior model probability, its posterior probability is not much higher than those of other top models. All other models except the top model can be viewed as evidence supporting stock return predictability. Their model probabilities sum up to around 87%. Another point to note is that the variable with the highest inclusion probability (around 40%) is the dividend yield.

16

This is the only regressor whose posterior inclusion probability is higher than its prior.[9] The inclusion probabilities of other variables are at most between 10% and 15%. Our inclusion priors do not seemed to be confirmed by the data. This reveals a substantial amount of model uncertainty. It seems that apart from dividend yield, none of the explanatory variables are overwhelmingly strong predictors of the stock return, although the models supporting predictability have higher posterior probability than the model supporting no predictability.

4.3 Bayesian Model Averaging in an SUR Model

The previous subsection reveals that the excess return seems to be predictable and some explanatory variables have relatively high posterior inclusion probabilities. However, as Holmes *et al.* (2001) point out, if one can incorporate the data generating processes of the explanatory variables into estimation, the true model for the dependent variable may receive higher posterior model probability since different DGPs can borrow strength from each other. In this subsection, we will implement this idea in a seemingly unrelated regression (SUR) model to investigate more closely the predictability of excess return.

We assume that the explanatory variables have their own data generating processes and that such processes could be correlated with each other and with that of the excess return.

$$r_t = a_{0p} + B'_{0p}x_{t-1,p} + F'_{0p}y_{t,p} + \epsilon_{0t,p} \tag{28}$$

$$x_{i,t} = a_i + B'_i w_{i,t-1} + \epsilon_{it} \tag{29}$$

The disturbance terms in (28) and (29) are assumed to be correlated with each other but have no heteroscedasticity and no serial correlation. As before p is the model specific subscript. Here we separate the explanatory variables into dummy variables y and non-dummy variables x, which have their own generating processes described in equation (29). The regressors for the predictor equations (w) may include the lag of the excess stock return and those of the predictors. To ease the computational burden in estimation[10], we wish to reduce the number

of equations and the parameters to be estimated. We only pick up the five variables with the highest inclusion probabilities calculated in the previous subsection, which consist of one dummy variable, Jan, two financial variables, Smb and Dy, and two business cycle variables, Inf and $M0$. Therefore there are altogether 5 equations in our system. All the equations in (29) ($i = 1, 2, 3, 4$) have an intercept term. Holmes *et al.* (2001) suggests a full search of potential regressors for each equation under SUR framework. To make it simpler, here we just use all possible explanatory variables[11] for each predictor equation. Since our focus is still on the excess return, for different models we assume only the regressors in equation (28) will change and the predictor equations will stay the same for different models. The marginal likelihood for a particular model in our case should be based on all equations. In this sense our work differs from the previous researchers such as Avramov (2002). Unlike the univariate case, an SUR model like ours has no closed form for the marginal likelihood. We use Savage-Dickey density ratio (see Verdinelli and Wasserman, 1995) to calculate the posterior model probability. Estimation details are discussed in Appendix A.2.

Table 4 lists all the models along with their posterior probabilities in descending order. We obtain the results after 1 million draws in the Gibbs sampler. Remember we only change the regressors of the excess return equation in (28) to form different models, while the regressors for other equations of (29) remain the same in the BMA exercise. Different from the univariate BMA case, the model without any regressor has much higher probability while the posterior model probabilities of most of the other top ten models in the univariate framework fall substantially. The sum of the model probabilities of all the models supporting stock return predictability is now only around 30%. This indicates under the SUR model we find less favourable evidence for stock return predictability. A point to note is that the posterior model probability of the one with only January effect jumps from 0.02 to about 0.135, which accounts for more than one third of the posterior model probability of the models supporting stock return predictability. It seems that if we incorporate the DGPs of the explanatory variables for the excess return and allow such DGPs to be correlated with

each other, we find much weaker support for stock return predictability compared with the univariate case. Further analysis, though, needs to be carried out to see whether such weak predictability has an impact on the investor's portfolio strategy.

[Table 4 here]

Table 5 shows the estimation results of all the parameters for the excess return equation (28) after model averaging, where numerical standard error (NSE) is equal to $\frac{\text{standard deviation}}{\sqrt{\text{number of draws}}}$, which is a measure of accuracy for the mean estimates. When the true posterior mean has no closed form, the numerical method we use implies that it should lie in the region of (estimated mean-1.96NSE, estimated mean+1.96NSE) with about 95% probability. Compared with Table 2, the slopes of Smb, Dy and Inf have decreased in scale (in absolute value). We can also see a huge drop in inclusion probability for most of them except the January dummy. The estimates for the variance matrix of the disturbance terms are shown in the lower triangle of Table 6 with standard deviations in brackets. The correlation coefficients of different equations are in the upper triangle. Note that the correlation between the excess return equation and other equations do exist (ranging from 1% to 14% in absolute value), which could justify the use of the SUR model.

[Table 5 here]

[Table 6 here]

5 The Horizon Effect of Stock Return Predictability and Model Uncertainty

Equation (28) and (29) provide us a framework to make forecasts of more than one period ahead based on the information of current period. To simplify the illustration, we need to write equation (28) and (29) into the form of vector autoregression (VAR). First let us define

the following[12],

$$z_t = \begin{bmatrix} r_t \\ x_{1t} \\ ... \\ x_{4t} \end{bmatrix}, \quad B = \begin{bmatrix} 0 & B'_{0,p} \\ & \tilde{B} \\ & {}_{4\times5} \end{bmatrix}$$

$$A = \begin{bmatrix} a_{0p} & a_1 & a_2 & a_3 & a_4 \end{bmatrix}', \quad F = \begin{bmatrix} F'_{0,p} \\ 0 \end{bmatrix}$$

Equation (28) and equation (29) can now be written as

$$z_t = B \cdot z_{t-1} + A + H \cdot y_{t,p} + \epsilon_t, \qquad \epsilon_t \sim iidN(0, \Sigma), \tag{30}$$

where H is the demean matrix resulting from reparameterizing the intercept term in the excess return equation. Denote $C = (B'_{0p}, F'_{0p}, a_{0p}, B'_1, a_1...B'_4, a_4)'$. Now we can use the following to estimate the mean and variance of z_t h periods ahead conditional on a particular model and the parameters in the excess return and the predictor equations.

$$E(z_{T+h}|C, \Sigma, D, M_p) = B^h \cdot z_T + \sum_{i=0}^{h-1} B^i \cdot A + \sum_{i=0}^{h-1} B^i \cdot H \cdot y_{T+h-i,p} \tag{31}$$

$$Var(z_{T+h}|C, \Sigma, D, M_p) = \Sigma + B\Sigma B' + ... + B^{h-1}\Sigma(B^{h-1})' \tag{32}$$

Note that y is the dummy variable, i.e. Jan in our case, which captures the periodic phenomenon. In our evaluation of the moments of the cumulative excess return (i.e. $R_{T+\hat{T}} = r_{T+1} + r_{T+2} + ... + r_{T+\hat{T}}$), we set $y = 0$ since we are more interested in the relationship between the stock excess return and the economic fundamentals over time. The cumulative excess return $R_{T+\hat{T}}$ is the first element in the vector $\sum_{h=1}^{\hat{T}} z_{T+h}$, whose mean and variance can be calculated as

$$\mu_{cum} = B(B^{\hat{T}} - I)(B - I)^{-1} z_T + [B(B^{\hat{T}} - I)(B - I)^{-2} - \hat{T}(B - I)^{-1}]A \tag{33}$$

$$Var_{cum} = \sum_{h=1}^{\hat{T}} \delta(h)\Sigma\delta(h)', \ \delta(h) = (B^h - I)(B - I)^{-1} \tag{34}$$

The predictive distribution of the cumulative excess return conditional on C, Σ and the model is

$$R_{T+\hat{T}}|C, \Sigma, D, M_p \sim N(\mu_{cum}^{(1)}, Var_{cum}^{(1,1)}) \tag{35}$$

where $\mu_{cum}^{(1)}$ stands for the first element in μ_{cum} and $Var_{cum}^{(1,1)}$ is the (1,1) element of the variance matrix.

Note that for the models with explanatory variables other than the dummy in the excess return equation, our results are sensitive to the values of the initial variables, i.e. z_T. Predictability in the context of equation (31) and (32) means that investors use the dynamic model to predict the future based on the current information. The estimated mean and variance of $R_{T+\hat{T}}$ from (35) should be viewed as the investor's belief of the cumulative stock excess return and risk accordingly. Another reason for incorporating the DGPs for the explanatory variables of the excess return equation is that we want to make forecast of excess return $\hat{T}$ periods ahead. Barberis (2000) showed that when the excess return can be predicted only by dividend yield, the optimal stock holding proportion will be very sensitive to the initial value of dividend yield while less sensitive if the investor takes into account parameter uncertainty[13]. Note that in Table 4 the top two models receiving large amount of posterior model probabilities include no explanatory variables and only the dummy variable respectively[14]. These regressors do not appear in our forecast exercise in equation (33). In the situations like these, we are virtually saying that the stock return is fairly unpredictable. However, our BMA results are based on the average of all the potential models. Whether the weak predictability will lead to any conspicuous horizon effect requires further analysis.

First we will use the sample mean of all the explanatory variables concerned to form the investor's initial condition. The final results shown in Figure 4 are obtained after 8 million draws. The solid line represents the forecast path of the mean and standard deviation of

$R_{T+\hat{T}}$ over time. We can see that the mean of the excess return is positive throughout our investment horizon and like the standard deviation, it rises in scale as the investment horizon lengthens. In addition to our forecast, we have also included the evolution paths of the mean and standard deviation of $R_{T+\hat{T}}$ when the excess return is unpredictable with and without parameter uncertainty (in dashed and dotted line respectively), as indicated in (5) and (11). We can see that $R_{T+\hat{T}}$'s forecast mean is not as high as the one under no predictability in the long run, while its standard deviation is above the one without parameter uncertainty and slightly below the one with parameter uncertainty. Given z_T is the mean of the predictors, the evolution paths of the mean and standard deviation of the excess return from the BMA results are very similar to those under no predictability and with parameter uncertainty. Therefore we may conjecture that the optimal holding proportion of stock should decrease with time in the long run. This should make intuitive sense since our framework does not only take into account parameter uncertainty but also model uncertainty. When the initial condition for the investor is formed by taking the sample mean of the predictors, it is close to the case with no predictability since in our sample we find little evidence supporting predictability.

[Figure 4 here]

To confirm our guess, we can calculate the $\hat{MtoS}$ statistic defined in (13) and the optimal holding proportion of stock. We use the algorithm mentioned in Section 2 to search for the optimal ω. Figure 5 shows the results in solid lines. Except for the initial tiny rise, the optimal holding proportion of stock falls consistently. As for the $\hat{MtoS}$, although it has some zigzag movements[15], it clearly demonstrates a downward sloping trend over the long run. Hence we have reason to believe that the $\hat{MtoS}$ statistic captures investor's willingness to hold a risky asset over time to some degree. Under our initial condition, the weak predictability and model uncertainty lead to relatively slow increase of the mean of the excess return compared to the risk, which makes the FTSE rather unattractive in the long run. A rational investor under our utility maximization setting hence should decrease

her holding of the FTSE index asset over time. While it is difficult to calculate the optimal holding proportion of the risky asset, it is very convenient to calculate the $M\hat{to}S$ statistic as long as we can simulate draws from the predictive distribution of the excess return. Although the statistic does not depend on how the utility maximization problem is set up, it may still provide a reference for the investor in regard to how attractive an asset is over time.

[Figure 5 here]

Next we will turn to the question of whether the weak predictability of stock return will induce any horizon effect. As mentioned before, predictability should imply that the investor use the present information to predict the future. If there is no horizon effect caused by predictability, the investor should be insensitive to different values of z_T (the initial condition). Here we try two more values in addition to the mean of the predictors: zero and twice the mean of the predictors. The results are also shown in Figure 5. The dashed line is obtained from the initial value zero while the dashed-star line is from twice the predictors' mean. We can see that three paths of ω from three initial values look quite different, though all of them are downward sloping over time. The dashed-star line (from twice the predictors' mean) falls faster than the other two and its $MtoS$ line is below those of the other two. If we set the initial condition to a zero vector, the starting ω is much less than the other two cases. Over time, its optimal holding proportion seems to be parallel to the one obtained by setting z_T to the mean of the predictors. We can see its $MtoS$ line is initially below the z_T-mean line and the two get intertwined over time. To summarize, it seems that although the stock return predictability is weak, it still has a considerable effect on the investor's optimal portfolio decision over time.

6 Conclusion

In this paper, we study the horizon effect of stock return predictability, that is, for different lengths of investment horizons how a rational investor should allocate between risky and

risk free asset. We show that the investor's portfolio choice for different investment horizons can be linked to the relative time variation of stock expected return and its expected risk. We propose a computationally convenient statistic to capture such horizon effect and show that it could be related to an investors' optimal holding proportion of a risky asset. We also study the stock return predictability for the UK market, i.e. what variables may be useful in predicting stock excess return. We argue that Bayesian model averaging is more preferable than simply focusing on a particular model in terms of picking up the variables truly useful in predicting the return. By using BMA, we can avoid the problem of data snooping and take into account parameter and model uncertainty. We have studied the potential useful predictors under both univariate and multivariate frameworks. Our univariate results show that for the UK market, the most powerful predictors are dividend yield, January effect, monetary supply, inflation rate and company size effect. However, if we allow the data generating processes of stock excess return to be correlated with those of its explanatory variables, the predicting power decreases for most variables. Only January effect still remains relatively robust. Though the evidence for stock return predictability is rather weak, it can still lead to considerable horizon effect. In this paper, we only consider one risky asset (stock index). In the future, we could extend our framework to consider several risky assets. With regard to stock predictability, we have just considered the predictability in return. It could be fruitful to study the case when the same set of explanatory variables can predict stock volatility.

Acknowledgement

Notes

1. It implies that such patterns in the data may happen by chance.

2. That is the difference of the rate of return between the two assets.

3. All these procedures could be easily implemented in Maple once we obtain the moments of $R_{T+\hat{T}}$.

4. Here the data are the observed excess returns.

5. For some situation, we may need to replace the formula by $M\hat{t}oS = \frac{(\mu_{\hat{T}} - \mu_{T\hat{}1})/\mu_{T\hat{}1}}{(\sigma_{\hat{T}} - \sigma_{T\hat{}1})/\sigma_{T\hat{}1}}$

6. In that month, there was a stock market crash. The index dropped by around 27%.

7. The results are obtained from the MatLab routine autocor.m.

8. Any two of these variables have correlation of more than 70% in absolute value. Detailed results are available upon request from the author.

9. In other words, our data strengthen this prior.

10. The author agrees that model uncertainty should be considered for all equations at the same time. However, the current computation technology does not allow such practice. Moreover we should place our focus on the first equation about the stock return.

11. The potential explanatory variables for each equation are the predictor variables of the stock return and the stock return itself, i.e. $r, Smb, Dy, Infl, M0$ and Jan. All explanatory variables except Jan enter the predictor equation in the form of one period lag.

12. Here $\tilde{B}$ denote the collection of the slope parameters in equation (29)

13. The marginal effect of dividend yield on stock return in most applications is positive. Therefore given that a rational investor's initial value of dividend yield is positive, she should have more position in stock if she has longer investment horizon.

14. The sum of their model probabilities is around 83%.

15. Such movements could be due to the numerical error during the simulation. As the number of draws increases in the Gibbs sampler, the range of oscillation should be reduced.

References

AVRAMOV, D. (2002). Stock-return predictability and model uncertainty. *Journal of Financial Economics*, **64** (3), 423–458.

BANZ, R. (1981). The relation between return and market value of common stocks. *Journal of Financial Economics*, **9**, 3–18.

BARBERIS, N. (2000). Investing for the long run when returns are predictable. *Journal of Finance*, **55** (1), 225–264.

BAWA, V. S., BROWN, S. J. and KLEIN, R. W. (1979). *Estimation Risk and Optimal Portfolio Choice*. North-Holland, New York.

BRENNAN, M. J. and XIA, Y. (1999). *Assessing Asset Pricing Anomalies*. working paper, University of California, Los Angeles.

CREMERS, K. J. M. (2002). Stock return predictability: A bayesian model selection perspective. *The Review of Financial Studies*, **15** (4), 1223–1249.

FAMA, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, **25**, 383–417.

— and FRENCH, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, **23**, 3–25.

— and — (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, **33** (1), 3–56.

FERNANDEZ, C., LEY, E. and STEEL, M. F. (2001). Benchmark priors for bayesian model averaging. *Journal of Econometrics*, **100**, 381–427.

GEORGE, E. I. (1999). Discussion of "model averaging and model search strategies" by m. clyde. In J. Bernardo, J. Berger, A. Dawid and A. Smith (eds.), *Bayesian Statistics 6*, Oxford University Press.

— (2001). *Dilution Priors for Model Uncertainty*. Notes for msri workshop on nonlinear estimation and classification, Unicersity of California, Berkeley.

GOETZMANN, W. N. and JORION, P. (1993). Testing the predictive power of dividend yields. *Journal of Finance*, **48**, 663–679.

GRANGER, C. W. J. (1992). Forecasting stock market prices: Lessons for forecasters. *International Journal of Forecasting*, **8**, 3–13.

HODRICK, R. (1992). Dividend yields and expected stock returns. *Review of Financial Studies*, **5**, 357–386.

HOLMES, C., MALLICK, B. and DENISON, D. (2001). *Bayesian model order determination and basis selection for seemingly unrelated regressions.* Technical report, Oxford Centre for Gene Function.

KANDEL, S. and STAMBAUGH, R. F. (1996). On the predictability of stock returns: An asset-allocation perspective. *Journal of Finance*, **51** (2), 385–424.

KOOP, G. (2003). *Bayesian Econometrics.* Chichester, England: John Wiley & Sons.

KOTHARI, S. P. and SHANKEN, J. (1997). Book-to-market, dividend yield, and expected market returns: A time-series analysis. *Journal of Financial Economics*, **44** (2), 168–203.

LEAMER, E. (1978). *Specification Searches: Ad Hoc Inference with Non-Experimental Data.* New York: Wiley.

MADIGAN, D. and YORK, J. (1995). Bayesian graphical models for discrete data. *International Statistical Review*, **63**, 215–232.

MERTON, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*, **51** (3), 247–57.

PESARAN, M. H. and TIMMERMANN, A. (1995). Predictability of stock returns: Robustness and economic significance. *Journal of Finance*, **50** (4), 1201–28.

POIRIER, D. (1985). Bayesian hypothesis tesing in linear models with continuously induced conjugate priors across hypotheses. In J. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith (eds.), *Bayesian Statistics 2*, New York: Elsevier, pp. 711–722.

SAMUELSON, P. A. (1969). Lifetime portfolio selection by dynamic stochastic programming. *The Review of Economics and Statistics*, **51** (3), 239–46.

— (1989). The judgment of economic science on rational portfolio management: Indexing, timing, and long-horizon effects. *Journal of Portfolio Management*, pp. 4–12.

— (1990). Asset allocation could be dangerous to your health. *Journal of Portfolio Management*, pp. 5–8.

VERDINELLI, I. and WASSERMAN, L. (1995). Computing bayes factors using a generalization of the savage-dickey density ratio. *Journal of the American Statistical Association*, **90** (430), 614–618.

XIA, Y. (2001). Variable selection for portfolio choice. *Journal of Finance*, **56** (1), 205–246.

ZELLNER, A. (1986). On assessing prior distributions and bayesian regression analysis with g-prior distribution. In P. K. Goel and A. Zellner (eds.), *Bayesian Inference and Decision Techniques: Essays in Honour of Bruno de Finetti*, Amsterdam: North-Holland, pp. 233–243.

A Appendix

A.1 The Explanatory Variables for the Excess Return Equation

1. January Dummy (Jan), which captures the January effect in the stock market

2. monthly return of the three-month Treasury bill (Tb)

3. the first difference of Treasury bill ($Tbchng$), which is calculated as $Tb(t) - Tb(t-1)$

4. the difference of return between small market capitalization companies and big ones (Smb), which is the difference between the total returns of Hoare Govett Smaller Companies index (HGSC) and FTSE 100 Index

5. dividend yield, the ratio of dividend over stock price(Dy)

6. the difference between monthly returns of 20 year UK government gilt and the 3 month T-bill ($TERM$)

7. monthly industrial production ($Indp$)

8. money supply, seasonally adjusted ($M0$)

9. monthly percentage change of industrial production ($Indp\%ch$)

10. monthly percentage change of monetary supply ($M0\%ch$)

11. Monthly inflation (Inf)

12. monthly oil price ($Oilp$)

13. monthly percentage change of oil price ($Oil\%ch$)

14. the difference between returns of high book-to-market ratio company index and low ones (HML), which is calculated as the difference between the total returns of MSCI value index and growth index

15. monthly change of inflation rate($Infch$), which is calculated as $Inf(t) - Inf(t-1)$

A.2 Estimation of the SUR Model

Let us define $\epsilon_t = [\epsilon_{0t}, \epsilon_{1t}, \epsilon_{2t}, \epsilon_{3t}, \epsilon_{4t}]'$ and assume

$$\epsilon_t \sim N(0, \Sigma) \text{ and } E(\epsilon_j \epsilon_k') = 0 \text{ for } j \neq k. \tag{36}$$

We will estimate equation (28), (29) and (36) in an SUR framework. Koop (2003) illustrates how to estimate SUR model in a Bayesian way. Our analysis partly relies on it. First we need to write equation (28) and (29) into matrix form by defining the following notations.

$$\underset{(m \times 1)}{z_t} = \begin{bmatrix} r_t \\ x_{1t} \\ \ldots \\ x_{4t} \end{bmatrix}, \quad \underset{(m \times k_p)}{\tilde{X}_t} = \begin{bmatrix} x'_{t-1,p} & y'_t & 1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & w'_{1,t-1} & 1 & 0 & \ldots & 0 & 0 \\ \ldots & & & \ldots & & \ldots & & \ldots & \\ 0 & 0 & 0 & 0 & 0 & 0 & \ldots & w'_{4,t-1} & 1 \end{bmatrix}$$

$$\underset{(k_p \times 1)}{C} = \begin{bmatrix} B'_{0p} & F'_{0p} & a_{0p} & B'_1 & a_1 & \ldots & B'_4 & a_4 \end{bmatrix}'$$

29

So equation (28) and (29) can be rewritten as

$$z_t = \tilde{X}_t C + \epsilon_t \tag{37}$$

We adopt the independent Normal Wishart prior for C and Σ, which looks like

$$p(C, \Sigma) = p(C)p(\Sigma) = f_N(C|\underline{C}, \underline{V})f_{IW}(\Sigma) \tag{38}$$

where the prior parameters $\underline{C}$ and $\underline{V}$ denote the mean and variance in the normal distribution. Although we have tried to limit the number of our parameters, we still end up with 49 parameters to estimate when we include all regressors into equation (28), which means the specification of the hyperparameters could be a huge task. Here we try to be as least subjective as possible. Koop (2003) recommends a general rule of thumb for doing BMA: it is acceptable to use a noninformative improper prior for parameters which are common to all models and informative proper priors for parameters changing over models. Since for different models we only change the regressors in equation (28), only the dimension of B_{0p} and F_{0p} will change across models. For these parameters, we use a proper prior. We will use a noninformative prior for the other parameters. The prior for Σ looks like the following.

$$f_{IW}(\Sigma) \propto |\Sigma|^{-\frac{1}{2}(5+1)} \tag{39}$$

Let us denote $\mathbf{c}_p = \begin{bmatrix} B'_{0p} & F'_{0p} \end{bmatrix}'$. For parameters $[a_{0p} \ B'_1 \ a_1...B'_4 \ a_4]'$, we set their covariance elements in $\underline{V}$ and the diagonal elements in $\underline{V}^{-1}$ to zero so that the corresponding values of the hyperparameters in $\underline{C}$ are irrelevant. We will leave the prior for $\mathbf{c}_p$ to later discussion. For the moment we just assume we have a proper prior for it.

The posterior distributions for C and Σ have no analytical forms since the stock return equation and the predictor equations have different regressors. We have to use Gibbs sampler to evaluate them.

$$C|D, \Sigma \sim N(\bar{V}(\underline{V}^{-1}\underline{C} + \sum_{t=2}^{T}\tilde{X}_t\Sigma^{-1}z_t, \bar{V}) \tag{40}$$

$$\Sigma|D, C \sim IW(\sum_{t=2}^{T}(z_t - \tilde{X}_t C)(z_t - \tilde{X}_t C)', T) \tag{41}$$

where $\bar{V} = \underline{(V}^{-1} + \sum_{t=2}^{T} \tilde{X}_t' \Sigma^{-1} \tilde{X}_t)^{-1}$.

We first choose some arbitrary values for C and draw Σ from equation (41) and then plug the draw of Σ into (40) to make a new draw of C. Repeating this process will give us a chain of draws. We discard a certain number of the initial draws as burn-in and only retain the remaining draws. The sample average of such draws can give us the estimates of the posterior means for C and Σ.

Next we discuss the estimation details of the posterior model probability. Savage-Dickey density ratio (see Verdinelli and Wasserman, 1995) is used to calculate the Bayes factors of all restricted models relative to the model with all regressors included in equation (28). We denote the model with all regressors included by subscript all. We can view different models as fixing different parts of the elements in $\mathbf{c}_{all}$, which we call η, to 0 with probability 1. Again we attach a model specific subscript p to η for all restricted models[16]. Then the Savage-Dickey density ratio (Bayes factor) could be evaluated as

$$BF_{p,all} = \frac{p(D|M_p)}{p(D|M_{all})} = \frac{p(\eta_p = 0|D, M_{all})}{p(\eta_p = 0|M_{all})} \tag{42}$$

Though it is straightforward to evaluate the denominator from the marginal prior distribution, there is no direct way to evaluate the numerator since we do not know the analytical form of the posterior distribution for η_p. What we know is the posterior distribution of η_p conditional on Σ. We can have posterior draws of C and Σ from the Gibbs sampler. If we denote the number of draws from the Gibbs sampler by N, we can evaluate the numerator in (42) as

$$p(\eta_p = 0|D, M_{all}) = \frac{1}{N}\sum_{i=1}^{N}p(\eta_p = 0|\Sigma_i, D, M_{all}) \tag{43}$$

For us to use the Savage-Dickey density ratio to calculate the Bayes factor, the following condition must hold, see Verdinelli and Wasserman (1995).

$$p(\mathbf{c}_p|\eta_p = 0, M_{all}) = p(\mathbf{c}_p|M_p) \tag{44}$$

To guarantee the above condition to hold, we must choose the prior for $\mathbf{c}_p$ carefully. We first specify the prior for $\mathbf{c}_{all}$ using the g prior like that in equation (18) without the term σ_p^2. We choose g as in (19). Here we use Ω to denote the variance hyperparameter for $\mathbf{c}_{all}$ and break it into blocks corresponding to $\mathbf{c}_p$ and η_p,

$$\mathbf{c}_{all} = \begin{bmatrix} \mathbf{c}_p \\ \eta_p \end{bmatrix} \middle| M_{all} \sim N\left(0, \Omega = \begin{bmatrix} \Omega_{11,p} & \Omega_{12,p} \\ \Omega_{21,p} & \Omega_{22,p} \end{bmatrix}\right), \tag{45}$$

where Ω takes the form of a g prior in (18). It can be proved that the prior for $\mathbf{c}_p$ should have the following form for condition (44) to be satisfied,

$$p(\mathbf{c}_p|M_p) \sim N(0, \Omega_{11,p} - \Omega_{12,p}\Omega_{22,p}^{-1}\Omega_{21,p}), \tag{46}$$

which means for models with restriction $\eta_p = 0$, we have more confidence in $\mathbf{c}_p = 0$ a priori compared to the all inclusive model.

With the Bayes factor we are able to calculate the posterior odds ratio as

$$PO_{p,all} = \frac{p(M_p|D)}{p(M_{all}|D)} = \frac{p(D|M_p)p(M_p)}{p(D|M_{all})p(M_{all})} \tag{47}$$

The prior model probability is calculated as before in (24) with $\pi = \frac{1}{2}$. Finally we can calculate the posterior model probability for model p using the following,

$$p(M_p|D) = \frac{PO_{p,all}}{\sum\limits_{p=1}^{P} PO_{p,all}}. \tag{48}$$

The mean and variance estimates of C and Σ unconditional on the model space can be obtained in a similar way as in equation (25).

Tables

Table 1 OLS estimates from the excess return equation including all regressors

Regressors	Coefficients	T-statistic	P-value
$Intercept$	0.007	0.088	0.465
Jan	**0.0219**	**1.614**	**0.054**
Tb	**-9.1271**	**-1.7349**	**0.041**
$Tbchng$	**7.45**	**1.3325**	**0.092**
Smb	-0.0864	-1.0134	0.156
Dy	**2.265**	**3.1529**	**8.96e-4**
$TERM$	**-9.998**	**-2.018**	**0.0223**
$Indp\%ch$	0.04985	0.225	0.411
$M0\%ch$	-0.106	-0.795	0.214
Inf	-0.174	-1.093	0.138
$Oil\%ch$	-0.0881	-0.734	0.232
HML	-0.0689	-0.746	0.228
$1987Oct$	**-0.2496**	**-5.33**	**9.9e-8**
$Oilp$	**0.00033**	**1.303**	**0.0968**
$Indprd$	-1.41e-5	-0.019	0.492
$M0$	**-0.0031**	**-2.04**	**0.0213**
$Infch$	0.0605	0.104	0.4588
R^2	0.17778		
Adjusted R^2	0.12804		

Table 2 Univariate Posterior Estimates of the Slope Parameters

	slope	standard deviation	incl prob	prior incl
Jan	**1.62e-3**	**5.455e-3**	**0.1257**	0.498
Tb	-0.01926	0.43324	5.1e-2	0.213
$Tbchng$	0.223	1.6047	0.06596	0.492
Smb	**-0.0175**	**0.0522**	**0.14752**	0.491
Dy	**0.311**	**0.489**	**0.391**	0.254
$TERM$	0.0512	0.5828	0.06325	0.427
$Indp\%ch$	2.47e-3	5.17e-2	0.05564	0.493
$M0\%ch$	-0.0289	0.1649	0.07454	0.492
Inf	**-0.0362**	**0.0998**	**0.1644**	0.304
$Oil\%ch$	-0.0098	0.044	0.090	0.47
HML	-0.0059	0.0325	0.076	0.495
$Oilp$	-6.1e-6	3.01e-5	0.081	0.229
$Indp$	-3.33e-5	1.56e-4	0.085	0.26
$M0$	**-6.21e-5**	**2.3e-4**	**0.11**	0.239
$Infch$	-0.0197	0.1466	0.065	0.468

Table 3 Univariate Posterior Model Probabilities

ranking	model	model prob
1	0	0.129
2	Dy	0.098
3	Dy,Inf	0.056
4	M0	0.036
5	Indp	0.028
6	Oilp	0.027
7	Smb	0.025
8	Jan	0.019
9	Smb,Dy	0.014
10	Jan,Dy	0.0137

Table 4 Posterior model probabilities under the SUR framework

Ranking	Model Probability	Regressors in (28)	Ranking	Model Probability	Regressors in (28)
1	0.697	0	17	7.975e-5	Smb Dy Jan
2	0.1346	Jan	18	5.59e-5	Smb M0 Jan
3	0.0981	Dy	19	4.65e-5	Inf M0 Jan
4	0.0332	M0	20	4.04e-5	Dy M0 Jan
5	0.014	Dy Jan	21	2.62e-5	Smb Infl
6	0.0076	M0 Jan	22	1.71e-5	Smb Dy Infl
7	0.0059	Smb	23	6.39e-6	Dy Infl M0
8	0.0032	Infl	24	4.47e-6	Smb Infl Jan
9	0.0029	Dy Infl	25	1.687e-6	Smb Dy Infl Jan
10	0.001	Smb Jan	26	1.68e-6	Smb Infl M0
11	6.068e-4	Infl Jan	27	1.596e-6	Smb Dy M0
12	5.996e-4	Smb Dy	28	7.2e-7	Dy Infl M0 Jan
13	3.06e-4	Dy Infl Jan	29	3.98e-7	Smb Infl M0 Jan
14	2.7e-4	Smb M0	30	2.46e-7	Smb Dy M0 Jan
15	2.56e-4	Dy M0	31	3.79e-8	Smb Dy Infl M0
16	1.77e-4	Infl M0	32	4.19e-9	Smb Dy Infl M0 Jan

Table 5 BMA estimation results for the excess return equation under the SUR framework

return equation	Smb	Dy	$Infl$	$M0$	Jan	$const$
slope	-0.00075	0.043	-0.00079	-0.0022	0.0052	0.0038
std	0.0112	0.163	0.01611	0.01315	0.013	0.0077
NSE	3.53e-5	0.0005	5.09e-5	4.16e-5	4.14e-5	2.44e-5
incl prob	0.00683	0.0843	0.00697	0.03492	0.1527	1
prior incl	0.498	0.339	0.386	0.385	0.5	1

**Table 6** BMA estimates for the error variance matrix under the SUR framework

| Equation of	r	Smb	Dy	$Infl$	$M0$
r	0.0021775 (0.00018281)	*-0.14287*	*0.013314*	*-0.095545*	*0.077462*
Smb	-0.00020629 (8.645e−5)	0.0009575 (7.9883e−5)	*0.16883*	*0.090462*	*-0.063564*
Dy	$4.9834e - 7$ (2.252e−6)	$4.1904e - 6$ (1.4901e−6)	$6.4337e - 7$ (5.3803e−8)	*0.32002*	*0.16179*
$Infl$	$-2.4844e - 5$ (1.5398e−5)	$1.5598e - 5$ (1.0253e−5)	$1.4303e - 6$ (2.7633e−7)	$3.1049e - 5$ (2.5901e−6)	*0.015963*
$M0$	$2.2598e - 5$ (2.2496e−5)	$-1.2296e - 5$ (1.1387e−5)	$8.1132e - 7$ (2.9989e−7)	$5.561e - 7$ (2.048e−6)	$3.9085e - 5$ (3.2603e−6)

Figures
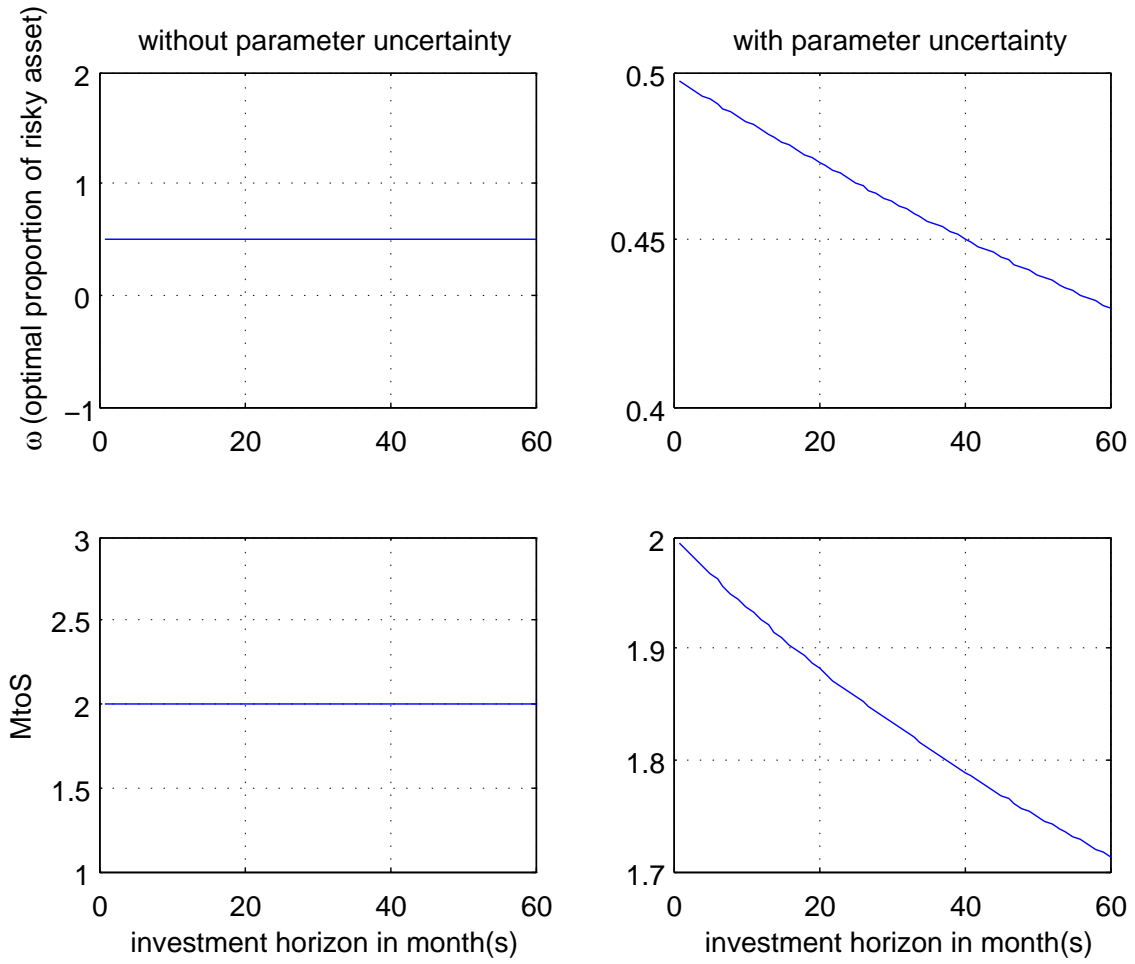
Figure 1 Optimal holding of stock and *MtoS* with respect to time horizon when excess return is unpredictable
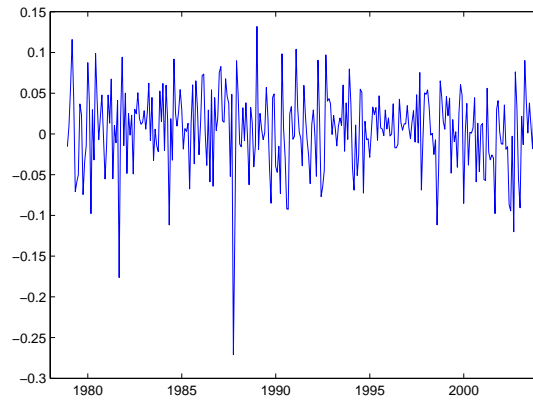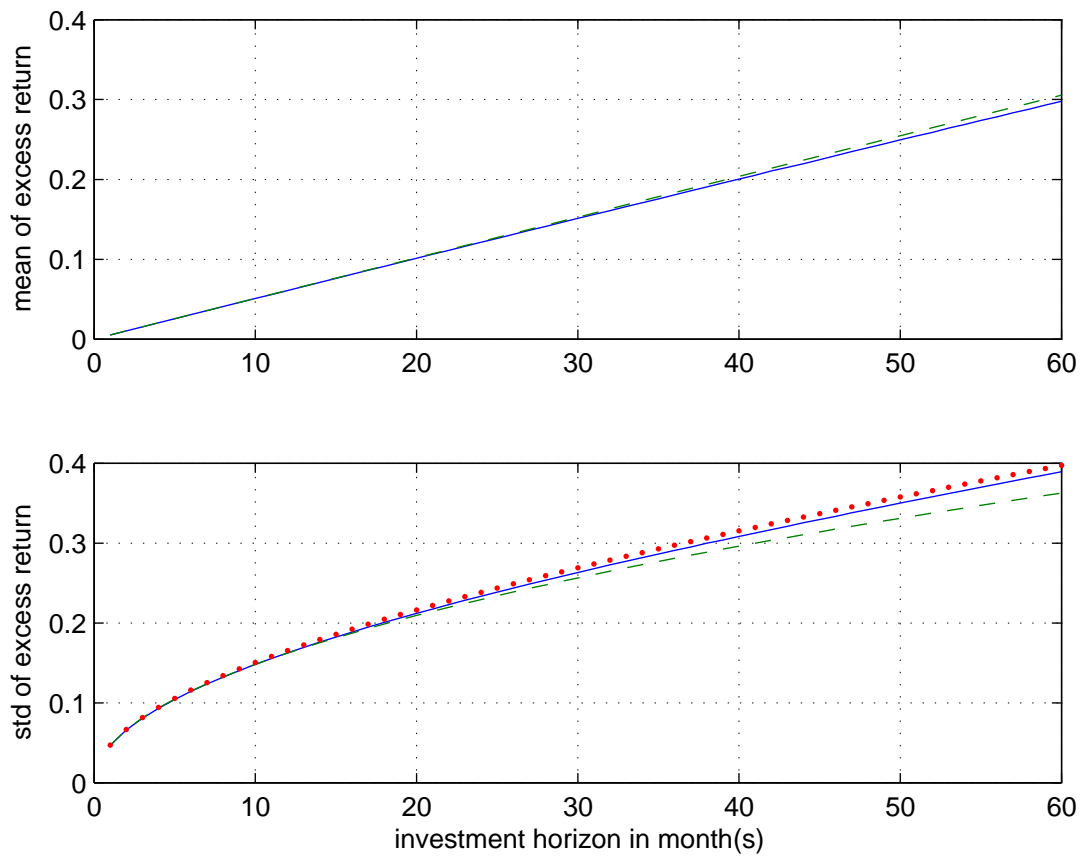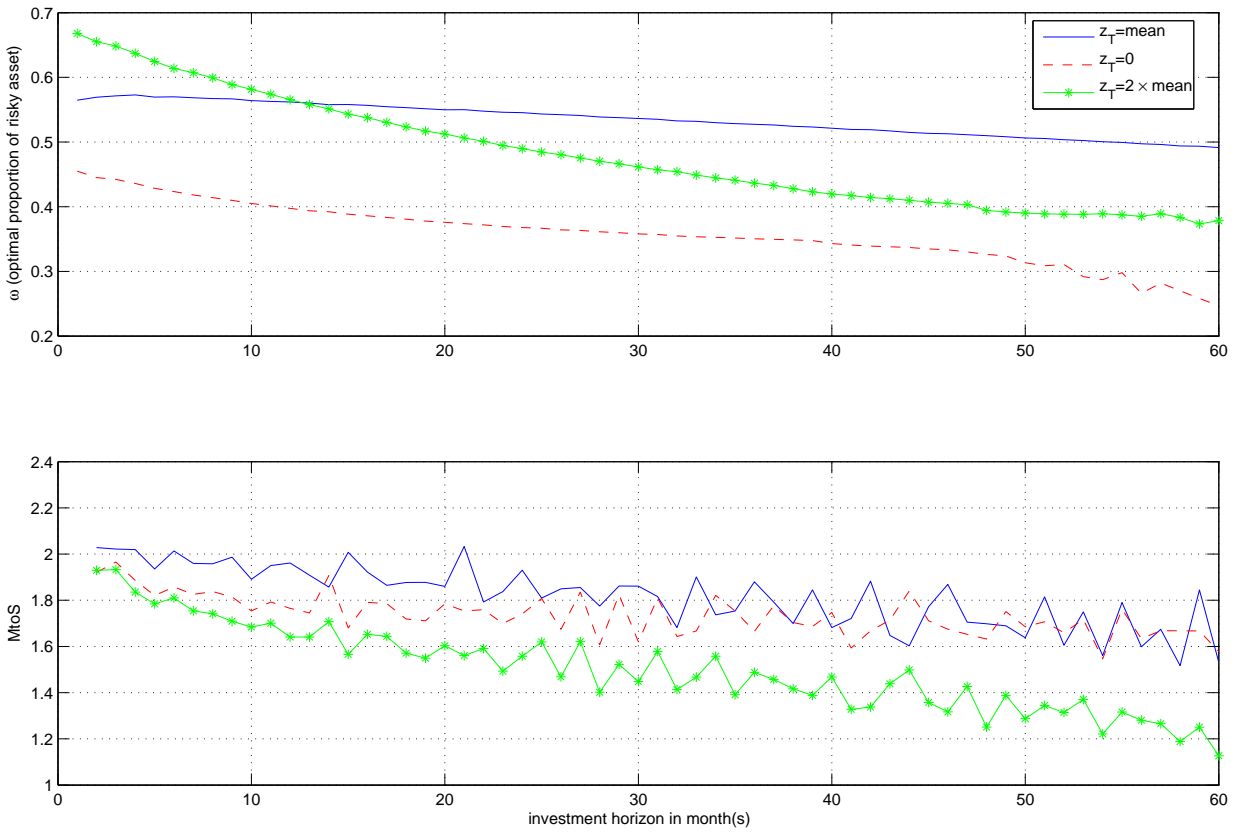
Figure 2 Monthly excess return on Financial Times All-Share Index
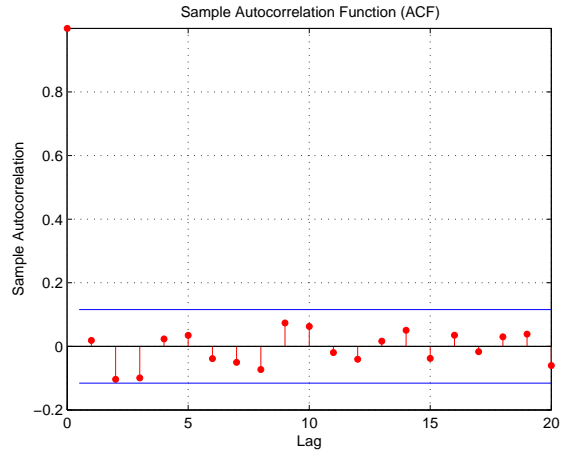
Figure 3 Sample autocorrelation of FTSE All-Share Index excess return

Figure 4 The mean and standard deviation of $R_{T+\hat{T}}$ (solid line)

Figure 5 The optimal holding proportion of stock and the MtoS statistic