Tactical Asset Allocation and Model Uncertainty: An Exploratory Study for the Swiss Stock Market

David Rey

First version: March 2004
Current version: October 2004

This research has been carried out within the NCCR FINRISK projects on “Asset Pricing and Portfolio Management”.
Tactical Asset Allocation and Model Uncertainty: An Exploratory Study for the Swiss Stock Market

David Rey

University of Basel (WWZ)

October 7, 2004

Abstract
This paper uses statistical model selection criteria and Avramov’s (2002) Bayesian model averaging approach to analyze the sample evidence on stock market predictability in the presence of model uncertainty. Based on Swiss stock market data, our posterior analysis finds that neither the cumulative posterior probabilities nor the posterior probabilities of the individual forecasting models are constant through time. We also decompose the variance of predicted future excess returns into three components: uncertainty attributed to forecast errors, parameter uncertainty, and model uncertainty. The empirical results indicate that the respective contributions are highly dependent on the time period under consideration and the initial values of the predictive variables. In contrast to Avramov (2002), model uncertainty is generally not more important than parameter uncertainty. Finally, we also demonstrate the implications of model uncertainty for tactical asset allocation strategies. Our results do not indicate any reliable out-of-sample return predictability. Moreover, in contrast to Avramov (2002), the out-of-sample performance of the Bayesian model averaging approach is not generally superior to the statistical model selection criteria. Consequently, even properly incorporating model uncertainty into a tactical asset allocation model does not seem to help improving the performance of the resulting short-term tactical asset allocation strategies.
We use a number of statistical model selection criteria and Avramov’s (2002) Bayesian model averaging approach to analyze the sample evidence on stock market predictability in the presence of model uncertainty. The empirical analysis for the Swiss stock market includes a number of predictive variables found important in previous studies of return predictability, as well as their stochastically detrended counterparts. The posterior analysis finds that neither the cumulative posterior probabilities nor the posterior probabilities of the individual forecasting models are constant through time. Which of the predictive variables receives the highest cumulative probability depends rather on the time period under consideration, and it seems difficult to discard any predictive variable as completely worthless. Moreover, the results are not robust to whether the predictive variables are stochastically detrended or not. We also decompose the variance of predicted future excess returns into three components: uncertainty attributed to forecast errors, parameter uncertainty, and model uncertainty. The empirical results indicate that the respective contributions are highly dependent on the time period under consideration and the initial values of the predictive variables. In contrast to Avramov (2002), model uncertainty is generally not more important than parameter uncertainty. Finally, we also demonstrate the implications of model uncertainty for tactical asset allocation strategies. The results do not indicate any reliable out-of-sample return predictability. Among the predictive variables, the dividend-price ratio exhibits the worst external validation on average. Moreover, in contrast to Avramov (2002), our analysis shows that the out-of-sample performance of the Bayesian model averaging approach is not generally superior to the statistical model selection criteria. Consequently, even properly incorporating model uncertainty into a tactical asset allocation model does not seem to help improving the performance of the resulting short-term tactical asset allocation strategies.

1 Introduction

Stock market predictability is now taken almost as a feature of the data. Indeed, it has been called “a new fact in finance” by Cochrane (1999) and the literature has identified numerous variables that potentially predict equity returns over time (see, e.g., Rey, 2003a,b). However, as recently pointed out by Avramov (2002), the “true” predictive regression specification is quite an open issue. Current equilibrium pricing theories are not explicit about which predictive variables should enter the predictive regression specification. Consequently, the reported empirical evidence on return predictability is subject to data-overfitting concerns. For example, Bossaerts and Hillion (1999), Neely and Weller (1999), and particularly Goyal and Welch (2003a,b) generally conclude that even their best prediction models have no out-of-sample forecasting power and fail to generate robust results that outperform simple unconditional benchmark models. Thus, given that the optimal set of predictive variables is basically unknown, we extend the empirical analysis of Rey (2004) and use a number of statistical model selection criteria and Avramov’s (2002) Bayesian model averaging approach to
analyze stock market predictability, model uncertainty, and their implications for tactical asset allocation strategies.

Based on Swiss stock market data from 1975 to 2002, we show that neither the cumulative posterior probabilities nor the posterior probabilities of the individual forecasting models are constant through time. Which of the predictive variables receives the highest cumulative probability depends rather on the time period under consideration, and it seems difficult to discard any predictive variable as completely worthless. Moreover, the results are not robust to whether the predictive variables are stochastically detrended or not. We also decompose the variance of predicted future excess returns into three components: uncertainty attributed to forecast errors, parameter uncertainty, and model uncertainty. The empirical results indicate that the respective contributions are highly dependent on the time period under consideration and the initial values of the predictive variables. In contrast to Avramov (2002), model uncertainty is generally not more important than parameter uncertainty.

Rey (2004) argues that the effect of dynamic rebalancing (i.e., the intertemporal hedging demand) is, in general, empirically negligible. In this paper, thus, we restrict the asset allocation problem to a (myopic) buy-and-hold framework. But in addition to parameter uncertainty, we also incorporate model uncertainty. The results do therefore no longer hinge on the validity of any single predictive regression specification. In particular, in the presence of model uncertainty, investment opportunities may be represented by a weighted Bayesian predictive distribution that integrates out the uncertainty about the true regression specification as well as the uncertainty about the respective regression parameters. Investigating the implications of model uncertainty from investment management perspectives, our results do not indicate any reliable out-of-sample return predictability. Among the predictive variables, the dividend-price ratio exhibits the worst out-of-sample forecasting ability on average. Moreover, the inclusion of more than one predictive variable rather detoriates the out-of-sample performance of the forecasting models. Finally, in contrast to Avramov (2002), our analysis shows that the out-of-sample performance of the Bayesian model averaging approach is not generally superior to the statistical model selection criteria.

Consequently, even properly incorporating model uncertainty into a tactical asset allocation model does not seem to help improving the performance of the resulting short-term tactical asset allocation strategies.

The remainder of the paper proceeds as follows. The next section summarizes both the statistical model selection criteria and Avramov’s (2002) Bayesian model averaging approach, including the Bayesian weighted predictive distribution and the corresponding variance decomposition. Section 3 contains the empirical results using data from the Swiss stock market. Section 4 concludes. A summary of the Swiss stock market data is provided in Appendix A.
2 Return Predictability and Model Uncertainty

Suppose that future excess returns on an equity portfolio are predictable using a simple linear regression specification. Given a set of $M$ predictive variables, there are $2^M$ competing predictive regression specifications. Each of these are then given by

$$e_t = \alpha_j + \beta_j^'x_{j,t-1} + \xi_{j,t},$$

(1)

where $e_t$ denotes the continuously compounded excess return over month $t$, $j$ is a model-specific indicator, and $x_{j,t-1}$ a model-unique subset of $n$ predictive variables. We may further assume that $\xi_{j,t}$ is normally distributed with conditional mean zero and variance $\sigma_{j,\xi}^2$.

The parameter $n$ ranges between zero and $M$. When $n = 0$, returns are assumed to be independently and identically distributed (i.i.d.), i.e., $e_t = \alpha_{iid} + \xi_{iid,t}$. In this case, the constant may be interpreted as a (fixed) risk premium. In contrast, when $n = M$, all $M$ predictive variables are suspected relevant.

Given a set of $M$ predictive variables, model risk or, equivalently, model uncertainty, corresponds to the uncertainty about the true predictive regression specification. Of course, in large samples, all $M$ predictive variables may be included in an all-inclusive regression specification. In this case, those predictive variables with no predictive power will have slope-coefficient estimates converging to zero, their true value. However, the available time series is often limited. Indeed, this may be especially true when the ultimate purpose is to obtain the model with the best external (i.e., out-of-sample) validity. A rolling scheme, for example, that fixes the size of the estimation window and therefore drops distant observations as recent ones are added, limits the available time series almost by definition. Consequently, the common predictive regression paradigm offers only little help in identifying the true set of predictive variables.

2.1 Statistical Model Selection Criteria

To start with, we apply a number of commonly adopted statistical model selection criteria to determine the best model among the set of all competing regression specifications. The ultimate purpose of these statistical model selection criteria is to avoid model overfitting, i.e., to retain only those models that have maximum external validity instead of minimum in-sample forecast errors. In our context of stock market predictability, this means that the preferred model should have the best out-of-sample forecasting performance.

Following Bossaerts and Hillion (1999), we use the following five statistical model selection criteria to select among the set of $2^M$ linear regression specifications: the adjusted $R^2$, Akaike’s information criterion (AIC; Akaike, 1974), Schwarz’s criterion (BIC; Schwarz, 1978), the Fisher information criterion (FIC; Wei, 1992), and the posterior information criterion (PIC; Phillips and Ploberger, 1996). While the first three model selection criteria are chosen on the basis of their popularity, the Fisher and posterior information criteria are chosen because of their robustness in the face of unit-root non-stationarities (see Bossaerts and Hillion, 1999, p. 409).
The adjusted \( R^2 \) is well-known. Formally, to define the other criteria, we write the sample of excess returns and predictive variables by

\[
e = [e_1 \cdots e_T]' \quad \text{and} \quad X_j = \begin{pmatrix} 1 & x_{j,0} \\ 1 & x_{j,1} \\ \vdots \\ 1 & x_{j,T-1} \end{pmatrix},\]

respectively, (2)

and the sum of squared regression errors by

\[
SSE_j = \left( e - X_j (X_j'X_j)^{-1} X_j' e \right)' \left( e - X_j (X_j'X_j)^{-1} X_j' e \right). \tag{3}
\]

Akaike’s (1974) information criteria is then given by

\[
AIC_j = T \ln \left( \frac{SSE_j}{T} \right) + 2(n + 1), \tag{4}
\]

and Schwarz’s (1978) criterion by

\[
BIC_j = T \ln \left( \frac{SSE_j}{T} \right) + (n + 1) \ln (T). \tag{5}
\]

For the Fisher and posterior information criteria, we may define model \( M \) to be the all-inclusive model. We then have for the Fisher information criterion

\[
FIC_j = SSE_j \frac{T}{T - (n + 1)} + SSE_M \frac{T - M}{T - (n + 1)} \ln \left( \frac{|X_j'X_j|}{SSE_j} \frac{SSE_j}{SSE_M} \frac{T - (n + 1)}{T - (M + 1)} \right), \tag{6}
\]

and the posterior information criterion

\[
PIC_j = SSE_M \left( \frac{SSE_j}{SSE_M} - 1 \right) + SSE_M \frac{T - (M + 1)}{T - (n + 1)} \ln \left( \frac{|X_j'X_j|}{SSE_j} \frac{SSE_j}{SSE_M} \frac{T - (n + 1)}{T - (M + 1)} \right). \tag{7}
\]

In each case, the regression specification is chosen that minimizes the respective criterion function (see Bossaerts and Hillion, 1999, Appendix A, eq. 5 to 8).

Overall, thus, statistical model selection criteria use a specific criterion to select a single regression specification. They then operate as if the chosen model corresponds to the “true” regression specification. Put it differently, implementing statistical model selection criteria is identical to the assumption that the selected regression specification is the “true” one with a unit probability and that all other competing models are completely worthless. In essence, thus, model uncertainty is actually ignored. In contrast, the following Bayesian model averaging approach recently proposed by Avramov (2002) averages over the dynamics implied by the set of all \( 2^M \) competing predictive regression specifications and therefore properly accounts for model uncertainty.
2.2 The Bayesian Model Averaging Approach

Specifically, the following Bayesian model averaging approach computes posterior probabilities for all \(2^M\) competing predictive regression specifications and then uses these probabilities as weights on the individual models to obtain a single composite weighted forecasting model.

We refer to Avramov (2002) for the full derivation and note that the posterior probability of model \(M_j\), denoted \(p(M_j)\), is given by

\[
p(M_j | z) = \frac{p(z | M_j) p(M_j)}{\sum_{i=1}^{2^M} p(z | M_i) p(M_i)}, \tag{8}
\]

where \(z\) is the data observed by the investor up until the start of his planning horizon, \(p(M_j)\) is the prior probability of \(M_j\) (which is at the discretion of the investor), and \(p(z | M_j)\) is the marginal likelihood of \(M_j\) given by

\[
p(z | M_j) = \frac{\ell(\theta_j, z, M_j) p(\theta_j | M_j)}{p(\theta_j | z, M_j)}, \tag{9}
\]

with \(p(\theta_j | M_j)\) and \(p(\theta_j | z, M_j)\) the joint prior and posterior distributions of the model-specific parameters, \(\ell(\theta_j, z, M_j)\) the likelihood function pertaining to \(M_j\), and, finally, \(\theta_j = (\alpha_j, \beta_j, \sigma^2_j)\) the set of regression parameters.

Avramov (2002, eq. 10) shows that the log marginal likelihood is given by

\[
\ln[p(z | M_j)] = -\frac{T}{2} \ln(\pi) + \frac{T_{j,0} - n - 1}{2} \ln\left(\frac{T_{j,0} \hat{\sigma}^2}{\hat{\sigma}^2 + \hat{\mu}^2} + \frac{T_{j,0} \hat{\mu}}{\hat{\mu}^2} + \frac{X_j^T X_j}{\hat{S}_j}\right) - \ln\left(\frac{T_{j,0} - n - 1}{2}\right) - \ln\left(\frac{T_{j,0}}{T_{j,0} - n - 1}\right) - \frac{n + 1}{2} \ln\left(\frac{T_{j,0}}{T_{j,0} - n - 1}\right), \tag{10}
\]

where

\[
\hat{S}_j = T_j \left(\hat{\sigma}^2 + \hat{\mu}^2\right) - \frac{T}{T_j} (T_{j,0} \hat{\mu} \hat{\mu}^T + \hat{\mu}^T \hat{x}_j^T + \hat{x}_j^T \hat{x}_j)^{-1} \left(T_{j,0} \hat{\mu} \hat{\mu}^T + \hat{\mu}^T \hat{x}_j^T + \hat{x}_j^T \hat{x}_j\right), \tag{11}
\]

and

\[
\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} e_t, \tag{12}
\]

\[
\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} (e_t - \hat{\mu})^2, \tag{13}
\]

\[
\hat{x}_j = \frac{1}{T} \sum_{t=0}^{T} x_{j,t}. \tag{14}
\]
Note that $\Gamma(y)$ stands for the Gamma function evaluated at $y$.\(^1\) $T$ is the actual sample size and $T^*_j = T + T_{j,0}$, where $T_{j,0}$ determines the strength of the informative prior. Kandel and Stambaugh (1996) and Avramov (2002) show that a reasonable value for the prior sample size increases as the model contains more predictive variables. The empirical analysis following below relies on this and takes 50 observations per parameter, i.e., $T_{j,0} = T_0 (n + 1)$ with $T_0 = 50$.\(^2\)

For the i.i.d. model ($n = 0$), we simply have $\hat{S}_{iid} = T^*_{iid} \hat{\sigma}^2_c$.

Given these posterior probabilities, Avramov (2002) proposes the cumulative posterior probabilities of the predictive variables as a statistic to summarize the weight of the respective predictive variables in the weighted forecasting model. Cumulative posterior probabilities are computed as

$$A \mathbb{P},$$

where $A$ is a $(2^M, M)$ matrix representing all forecasting models by zeros and ones, designating exclusions and inclusions of predictive variables, respectively, and the $(2^M, 1)$ vector $\mathbb{P}$ contains the posterior probabilities. Thus, cumulative posterior probabilities indicate the probabilities that each of the predictive variables appears in the weighted forecasting model.

### 2.2.1 The Bayesian Weighted Predictive Distribution

Let $z_{jt} = [e_t, x_{jt}]$ denote the data-generating process corresponding to model $j$ and assume that the evolution of $z_{jt}$ is given by

$$z_{jt} = a_j + B_j x_{j,t-1} + \xi_{jt},$$

with $\xi_{j,t} \sim \text{i.i.d. } N(0, \Sigma_j)$. The Bayesian weighted predictive distribution of cumulative excess returns averages over the model space and also integrates over the posterior distribution that summarizes parameter uncertainty about the VAR parameters $\theta_j = (a_j, B_j, \Sigma_j)$. According to Avramov (2002, eq. 19), it is given by

$$p\left(e_{T-T+\hat{T}} | \mathbb{Z}\right) = \sum_{j=1}^{2^M} p(M_j | \mathbb{Z})$$

$$\int p\left(e_{T-T+\hat{T}} | \theta_j, \mathbb{Z}, M_j\right) p(\theta_j | \mathbb{Z}, M_j) d\theta_j,$$

where $\hat{T}$ is the planning horizon in months and $e_{T-T+\hat{T}} \equiv e_{T+1} + e_{T+2} + \ldots + e_{T+\hat{T}}$. Since an analytical solution for the integral in equation (16) is not feasible when $\hat{T} > 1$, our empirical implementation following below is based on Monte Carlo integration. First, a model $M_j$ is drawn with probability $p(M_j | \mathbb{Z})$.

Second, we sample from the posterior distribution by first drawing from the marginal $p(\Sigma_j^{-1} | \mathbb{Z})$, a Wishart distribution, and then, given the $\Sigma_j$ draw,
from the conditional \( p(C_j | \Sigma_j, z) \), \( C_j = [a_j' B_j]' \), a multivariate Normal distribution. Repeating this many times gives an accurate representation of the posterior distribution. Third, for each draw of \( a_j, B_j \) and \( \Sigma_j \) from the posterior \( p(a_j, B_j, \Sigma_j | z) \), we sample from the Normal distribution with mean vector \( \mu_{j,T \rightarrow T+\hat{T}} \) and variance matrix \( \Sigma_{j,T \rightarrow T+\hat{T}} \), given by

\[
\mu_{j,T \rightarrow T+\hat{T}} = \hat{T} a_j + \left( \hat{T} - 1 \right) B_{j,0} a_j + \left( \hat{T} - 2 \right) B_{j,0}^2 a_j + \ldots + B_{j,0}^{\hat{T}-1} a_j \\
+ \left( B_{j,0} + B_{j,0}^2 + \ldots + B_{j,0}^{\hat{T}-1} \right) z_{j,T},
\]

and

\[
\Sigma_{j,T \rightarrow T+\hat{T}} = \Sigma_j + (I_j + B_{j,0}) \Sigma_j (I_j + B_{j,0})' + (I_j + B_{j,0} + B_{j,0}^2) \Sigma_j (I_j + B_{j,0} + B_{j,0}^2)' + \ldots + (I_j + B_{j,0} + B_{j,0}^2 + \ldots + B_{j,0}^{\hat{T}-1}) \Sigma_j (I_j + B_{j,0} + B_{j,0}^2 + \ldots + B_{j,0}^{\hat{T}-1})',
\]

with \( B_{j,0} = [0 B_j] \) and \( 0 \) an \( (n+1,1) \) vector of zeros. This gives a large sample of the predictive distribution \( p(e_{T \rightarrow T+\hat{T}} | z) \). The Bayesian weighted predictive distribution of cumulative excess returns may be used to compute the optimal allocation to equities, \( \omega \), when taking stock market predictability, parameter uncertainty, and model uncertainty into account, or, as below, to decompose the variance of the predicted excess returns into uncertainty attributed to forecast errors, parameter uncertainty, and model uncertainty.

### 2.2.2 Variance Decomposition

Based on the weighted predictive distribution given in equation (16), Avramov (2002) shows that predicted future excess returns are subject to three sources of uncertainty: (i) uncertainty attributed to forecast errors, (ii) parameter uncertainty, and (iii) model uncertainty. In particular, Avramov (2002, eq. 25) shows that the variance of the predicted excess returns can be decomposed as

\[
Var(e_{T \rightarrow T+\hat{T}} | z) = \sum_{j=1}^{M} p(M_j | z) \left[ E(T_j) + Var(\lambda_j) + \left( E(\lambda_j) - \hat{\lambda} \right)^2 \right],
\]

A more detailed description of this algorithm is originally given in Barberis (2000) or, using the same notation as above, in Rey (2004). An alternative algorithm, which samples directly from the distribution of returns and which is potentially more efficient when there is a large number of predictive variables, is proposed in Avramov (2000).
where $E(\Upsilon_j)$ and $\text{Var}(\lambda_j)$ are the two variance components attributed to forecast errors and parameter uncertainty, respectively, and $\Upsilon_j$ and $\lambda_j$ denote the first elements of the variance matrix $\Sigma_{j,T\rightarrow T+\hat{T}}$ and the mean vector $\mu_{j,T\rightarrow T+\hat{T}}$. The model uncertainty component is then given by

$$2M\frac{1}{p}(M_j|z) \left(E(\lambda_j) - \tilde{\lambda}\right)^2,$$  

(20)

where

$$\tilde{\lambda} = \sum_{j=1}^{2M} p(M_j|z) E(\lambda_j)$$

is the predicted mean of cumulative excess returns that averages across model-specific predicted means using posterior probabilities as weights (Avramov, 2002, eq. 26). The empirical section following below quantifies these three risk components for planning horizons of one month, $\hat{T} = 1$.

## 3 Empirical Results

Using data from the Swiss stock market, this section demonstrates the impact of return predictability, parameter uncertainty, and model uncertainty on tactical asset allocation strategies. In particular, we will start out with the calculation of the (cumulative) posterior probabilities and perform the variance decomposition suggested above. We then critically examine and compare the out-of-sample performance of both the statistical model selection criteria and the Bayesian model averaging approach.

### 3.1 The Data and Preliminary Evidence

Our investment universe consists of monthly observations on continuously compounded excess stock market returns over January 1975 through December 2002 (336 observations). The stock market data is summarized in Appendix A.

In deciding which predictive variables to include, attention was given to those variables found important in previous studies of return predictability. Of course, there is a natural concern about return predictability uncovered through collective “data-snooping” by a series of researchers (Lo and MacKinlay, 1990; Foster, Smith, and Whaley, 1997; Ferson, Sarkissian, and Simin, 2003, 2004). However, most of this research is based on U.S. data and, to our knowledge, there is no study for the Swiss stock market that uses data covering the period starting in 1975 and that includes the recent bear market.

Each of the $2^M$ competing predictive regression specifications considered retains a unique subset of the following $M = 7$ predictive variables:

(i) Dividend-price ratio, log (DPR),
(ii) Earnings-price ratio, log (EPR),
(iii) Term spread (TERM),
(iv) Nominal one-month Swiss interbank rate (IR),
(v) Realized stock market volatility, log (VOLA),
(vi) U.S. TED spread (TED), and, finally,
(vii) U.S. default risk spread (DEF).

The dividend-price ratio/earnings-price ratio is measured as the sum of dividends/earnings paid on the index over the previous year, divided by the current level of the index. The term spread is the difference between the (log) nominal yield on long-term government bonds provided by IMF and the (log) nominal three-month Swiss interbank rate. In the same way as Goyal and Santa-Clara (2003), we compute the monthly realized variance of the real stock market returns using within-month daily return data for each month as

\[ \text{Var}_{\text{Market}}^t = \sum_{d=1}^{D_t} r_{m,d}^2 + 2 \sum_{d=2}^{D_t} r_{m,d} r_{m,d-1}, \]

where \( D_t \) is the number of days in month \( t \) and \( r_{m,d} \) is the continuously compounded real stock market return on day \( d \). The second term on the right-hand side adjusts for the autocorrelation in daily returns using the approach proposed by French, Schwert, and Stambaugh (1987). The U.S. TED spread is calculated as the difference between (log) three-month Eurodollar rates and (log) three-month Treasury Bill rates, provided by the Federal Reserve Board of Governors. Finally, the U.S. default risk spread is formed as the difference in annualized (log) yields of Moody’s Baa and Aaa rated bonds.

Again, monthly data are used throughout, spanning 336/337 months from December 1974 to November/December 2002.

Motivated by the recent contributions of Ferson, Sarkissian, and Simin (2003, 2004), a second subset includes the same \( M = 7 \) predictive variables, but now transformed in the following simple way. We transform the predictive variables by subtracting off a trailing moving average of its own past values,

\[ x_{t-1}^* = x_{t-1} - \frac{1}{12} \sum_{\tau=1}^{12} x_{t-1-\tau}. \]

In words, we subtract a backward one-year moving average of past values from the prevailing value of the predictive variable to get a “stochastically detrended” time series that is equivalent to a triangularly weighted moving average of past changes in the predictive variable, where the weights decline as one moves back in time. Accordingly, the detrended time series is stationary if changes in the predictive variable are stationary. While this stochastic detrending method has already been used by Campbell (1991) and Hodrick (1992), only recently Ferson, Sarkissian, and Simin (2003, 2004) show that this is the most practically useful insurance against spurious regression bias (and therefore data mining). Since most of the above predictive variables are either manifestly non-stationary (realized stock market volatility is the exception), or, if not, their behavior is close enough to unit-root non-stationarity for small-sample statistics to be affected, it is interesting to compare the characteristics of the two data subsets.

\(^4\)A full list of references is provided in Rey (2003a,b).
Table 1: Multiple regressions of monthly excess returns on predictive variables: The all-inclusive regression specification.
The table exhibits the slope coefficients obtained by regressing continuously compounded monthly excess returns on a constant and all seven predictive variables (the all-inclusive model). The set of predictive variables includes: dividend-price ratio (DPR), earnings-price ratio (EPR), term spread (TERM), one-month Swiss interbank rate (IR), realized stock market volatility (VOLA), U.S. TED spread (TED), and U.S. default risk spread (DEF), as well as the corresponding stochastically detrended variables (SD). Estimates are given for three different time periods. The top two rows use data from January 1975 to December 2002. Estimates for the two subsamples are shown below. The first subsample uses data from January 1975 to December 1988; the second subsample is based on data from January 1989 to December 2002. The first row uses the full sample of monthly data from January 1975 to December 2002. Estimates for two subsamples are indicated below. The first subsample uses data from January 1975 to December 1988, covering the first half of the total time period, the second subsample is based on data from January 1989 to December 2002, covering the second half of the full sample.

<table>
<thead>
<tr>
<th></th>
<th>DPR</th>
<th>EPR</th>
<th>TERM</th>
<th>IR</th>
<th>VOLA</th>
<th>TED</th>
<th>DEF</th>
<th>ARSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>1975:01–2002:12</td>
<td>-0.003</td>
<td>0.002</td>
<td>*0.016</td>
<td>0.012</td>
<td>-0.003</td>
<td>0.006</td>
<td>-0.003</td>
<td>0.26%</td>
</tr>
<tr>
<td>SD</td>
<td>0.003</td>
<td>-0.003</td>
<td>-0.003</td>
<td>-0.005</td>
<td>-0.003</td>
<td>0.002</td>
<td>0.003</td>
<td>-0.50%</td>
</tr>
<tr>
<td>1975:01–1988:12</td>
<td>-0.002</td>
<td>***0.013</td>
<td>0.004</td>
<td>0.002</td>
<td>-0.006</td>
<td>0.001</td>
<td>0.003</td>
<td>6.07%</td>
</tr>
<tr>
<td>SD</td>
<td>0.000</td>
<td>0.004</td>
<td>0.004</td>
<td>0.002</td>
<td>-0.006</td>
<td>-0.001</td>
<td>0.005</td>
<td>0.05%</td>
</tr>
<tr>
<td>1989:01–2002:12</td>
<td>***0.038</td>
<td>-0.026</td>
<td>-0.002</td>
<td>-0.015</td>
<td>-0.000</td>
<td>***0.014</td>
<td>*-0.010</td>
<td>6.95%</td>
</tr>
<tr>
<td>SD</td>
<td>0.009</td>
<td>*-0.013</td>
<td>*-0.015</td>
<td>*-0.013</td>
<td>0.001</td>
<td>***0.013</td>
<td>-0.003</td>
<td>6.24%</td>
</tr>
</tbody>
</table>

To start with, Table 1 shows the slope coefficients obtained by regressing continuously compounded monthly excess returns on an intercept and all lagged predictive variables described above (the all-inclusive regression specification). The top row uses the full sample of monthly data from January 1975 to December 2002. Estimates for two subsamples are indicated below. The first subsample uses data from January 1975 to December 1988, covering the first half of the total time period, the second subsample is based on data from January 1989 to December 2002, covering the second half of the full sample.

Table 1 shows mixed evidence on return predictability. Over the full sample period, only the term spread is statistically significant and the adjusted $R^2$ is very low. When the predictive variables are stochastically detrended, the adjusted $R^2$ is even negative. From 1975 to 1988, the earnings-price ratio is highly
significant and the adjusted $R^2$ is 6.02%; but again, the evidence on return predictability is meager when the variables are stochastically detrended. Over the recent subsample, however, a number of predictive variables is statistically significant and the adjusted $R^2$s are somewhat more than 6%, irrespective of whether the predictive variables are stochastically detrended or not. Finally, the combined significance of the dividend-price ratio and the earnings-price ratio is difficult to judge. Depending on the time period under consideration and whether they are stochastically detrended or not, the respective estimated slope coefficients vary widely.

3.2 Posterior Probabilities of Forecasting Models

The consideration of all possible predictive regression specifications in the presence of the above seven predictive variables requires the comparison of $2^7 = 128$ models. Eq. (10) shows how to compute the marginal likelihood for every model, and eq. (8) weights the marginal likelihood by the model prior probability and normalizes the result to obtain the model posterior probability. As in Avramov (2002), prior probabilities, $p(M_j)$, are allocated equally across models.

Table 2 displays cumulative posterior probabilities, $A^\top P$. Recall that the cumulative posterior probabilities indicate the probabilities that each of the predictive variables appears in the weighted forecasting model.

Over the whole sample period, the cumulative posterior probabilities range from 47.32% for the U.S. default risk spread to 69.62% for the term spread, suggesting that the U.S. default risk spread and the term spread should appear in the weighted return-forecasting model with probabilities of 47.32% and 69.62%, respectively. However, when the predictive variables are stochastically detrended, the cumulative posterior probabilities are less dispersed and the term spread receives no longer the highest weight. In contrast, the highest cumulative posterior probability is associated with the U.S. default risk spread, and only the dividend-price ratio receives less weight than the term spread. Furthermore, it seems that posterior probabilities are not very stable over time. For example, from 1975 to 1988, the earnings-price ratio receives the highest cumulative posterior probability of 89.57%. The earnings-price ratio is thus much more important than the dividend-price ratio with a cumulative posterior probability of only 42.16%. Recently, it is the U.S. TED spread that exhibits the highest cumulative posterior probability of 77.47%. But again, these results change fundamentally when the predictive variables are stochastically detrended.

Figure 1 shows the posterior probabilities for each of the 128 predictive regression specifications, $M_j$. The graph on the left plots posterior probabilities for the regression specifications that retain the original predictive variables against posterior probabilities obtained for the set of stochastically detrended variables, using data from 1975 to 2002. The graph on the right plots posterior probabilities calculated over the period from 1975 to 1988 against posterior probabilities estimated with data from 1989 to 2002, using the original, i.e., not stochastically detrended, predictive variables. Both graphs reveal that the resulting posterior probabilities are very unstable, both regarding whether the predictive variables
Table 2: Cumulative posterior probabilities.
The table displays cumulative posterior probabilities for the seven predictive variables. The set of predictive variables includes: dividend-price ratio (DPR), earnings-price ratio (EPR), term spread (TERM), one-month Swiss interbank rate (IR), realized stock market volatility (VOLA), U.S. TED spread (TED), U.S. default risk spread (DEF), as well as the corresponding stochastically detrended variables (SD). Estimates are given for three different time periods. The top row uses data from January 1975 to December 2002. Estimates for the two subsamples are showed below. The first subsample uses data from January 1975 to December 1988; the second subsample is based on the time period from January 1989 to December 2002.

<table>
<thead>
<tr>
<th></th>
<th>DPR</th>
<th>EPR</th>
<th>TERM</th>
<th>IR</th>
<th>VOLA</th>
<th>TED</th>
<th>DEF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1975:01—2002:12</td>
<td>47.42%</td>
<td>47.40%</td>
<td>69.62%</td>
<td>57.16%</td>
<td>52.26%</td>
<td>52.91%</td>
<td>47.32%</td>
</tr>
<tr>
<td>SD</td>
<td>48.02%</td>
<td>51.56%</td>
<td>48.74%</td>
<td>52.60%</td>
<td>54.54%</td>
<td>49.86%</td>
<td>55.99%</td>
</tr>
<tr>
<td>1975:01—1988:12</td>
<td>42.16%</td>
<td>89.57%</td>
<td>41.63%</td>
<td>40.74%</td>
<td>49.58%</td>
<td>40.14%</td>
<td>41.60%</td>
</tr>
<tr>
<td>SD</td>
<td>49.42%</td>
<td>50.70%</td>
<td>50.53%</td>
<td>48.35%</td>
<td>61.34%</td>
<td>47.65%</td>
<td>57.48%</td>
</tr>
<tr>
<td>1989:01—2002:12</td>
<td>68.95%</td>
<td>57.65%</td>
<td>51.34%</td>
<td>50.98%</td>
<td>45.53%</td>
<td>77.47%</td>
<td>60.76%</td>
</tr>
<tr>
<td>SD</td>
<td>48.33%</td>
<td>66.99%</td>
<td>52.50%</td>
<td>50.81%</td>
<td>43.51%</td>
<td>83.42%</td>
<td>44.36%</td>
</tr>
</tbody>
</table>
Figure 1: Posterior probabilities.
The graph on the left plots posterior probabilities for the models that retain the original predictive variables against posterior probabilities obtained for the set of stochastically detrended predictive variables. Posterior probabilities are estimated over the full time period from 1975 to 2002. The graph on the right plots posterior probabilities calculated over the period from 1975 to 1988 against posterior probabilities using data from 1989 to 2002 (with original predictive variables).

are stochastically detrended or not and the time period under consideration. Compared to the original predictive variables, it seems that the posterior probabilities associated with the stochastically detrended variables are more equally spread across the regression specifications. Furthermore, the regression specifications with the highest (lowest) posterior probabilities over the first time period are often among the regression specifications with the lowest (highest) posterior probabilities over the second sample period.

To summarize, in contrast to Avramov (2002), who considers a subset of 14 predictive variables and uses U.S. data from 1953 to 1998, our results show smaller differences between the cumulative posterior probabilities. Thus, we do not conclude that only one or at most two predictive variables are retained as useful in the highest-probability models, and that the other predictive variables are discarded as worthless. Moreover, which of the predictive variables receives the highest cumulative probability is highly dependent on the time period under consideration and whether they are stochastically detrended or not.

3.3 Variance Decomposition

As described in Section 2.2.2, we perform the variance decomposition of predicted future excess returns into three components: uncertainty attributed to forecast errors, parameter uncertainty, and model uncertainty. In contrast to Avramov (2002), where the variance decomposition is solely based on the full sample and a single set of predictive variables, $x_{j,T}$, equal to actual end-of-sample realizations, our approach, based on the two following schemes, is more dynamic.

The first, the rolling scheme (see, e.g., Akhtar, 1989), fixes the estimation
window size and drops distant observations as recent ones are added. The model
cparameters are thus first estimated with data from 1 to k, next with data from
2 to k + 1,..., and finally with data from T − k to T − 1. In our case with k = 60
months, the variance decomposition is thus performed 276 times and each is
based on realizations of the predictive variables at the end of the respective
rolling sample.

The second scheme, the recursive (see, e.g., Fair and Shiller, 1990), uses all
available data in the sense that the variance decomposition is first estimated
based on data from 1 to k, next with data from 1 to k + 1,..., and finally from
1 to T − 1. This again gives a total of 276 variance decompositions with both
parameter estimates and initial values of the predictive variables changing over
time.

The consideration of a number of initial values (and parameter estimates)
seems to be more important than varying the strength of the informative prior,
T0. As in Avramov (2002), the results (not reported) indicate that the variance
decomposition is not highly sensitive to different values of T0. In addition, our
dynamic approach corresponds to the out-of-sample analysis following below
and is thus much more appropriate for investors ultimately concerned with the
out-of-sample performance of a corresponding tactical asset allocation strategy
than Avramov’s (2002) static approach.

Figure 2 shows the resulting time series of the contributions of the three
components to the overall uncertainty about predicted returns. The planning
horizon is one month ( T = 1). Based on our Swiss stock market data, the results
indicate that the variance decomposition is highly dependent on the time period
under consideration and the initial values of the predictive variables. Still, the
contributions of the uncertainty attributed to forecast errors are by far the most
important. This is especially true in the case of the recursive scheme, where
both the parameter and model uncertainty components practically disappear
over time (i.e., with the increasing sample size). In contrast to Avramov (2002),
however, model uncertainty is generally not more important than parameter
uncertainty. On average, the respective contributions are 4.85% and 11.48% for
the rolling scheme, and 1.82% and 3.96% for the recursive scheme. Whether the
predictive variables are stochastically detrended or not does not seem to make
any significant difference to the variance decomposition. The average values are
basically the same: 3.12% and 9.24% for the rolling scheme, and 0.87% and
2.86% for the recursive scheme, respectively. We would thus not generally claim
that model uncertainty is larger than parameter uncertainty.

An interesting sidestep is to consider the evolution of the variance of the pre-
dicted excess returns over time, Var ( et → eT+1 | z). Figure 3 shows the resulting
time series of these conditional volatilities. The graph on the left is based on the
rolling scheme, the graph on the right on the recursive scheme. In both cases
and irrespective of whether the predictive variables are stochastically detrended
or not, the graphs exhibit clear positive time trends. Conditional volatilities
thus seem to have considerably increased over the last 20 years.

The next section explores the out-of-sample predictive ability of the Bayesian
model averaging approach and compares it to the forecasting power of the sta-
Figure 2:
Variance decomposition: Rolling and recursive scheme.
The graphs show the resulting time series of the contributions of the three components to the overall uncertainty about predicted returns: Uncertainty attributed to forecast errors, parameter uncertainty, and model uncertainty (bold). The top two graphs are based on the original predictive variables. The graph on the left uses the rolling scheme with $k = 60$ months, the graph on the right the recursive scheme. The first estimates are thus available for January 1980. The graphs on the bottom are based on the stochastically detrended variables. For each sample period, the number of simulations is 50 per regression specification. The planning horizon is one month.
Figure 3: Conditional volatilities.

Based on the rolling scheme with $k = 60$ months, the graph on the left shows the evolution of the variance of the predicted excess returns over time. The graph on the right is based on the recursive scheme. The first estimates are available for January 1980. The bold lines are for the original predictive variables, the fine lines correspond to the stochastically detrended variables. For each sample period, the number of simulations is 50 per regression specification. The planning horizon is one month.

3.4 External Validation: Out-of-Sample Evidence

As already argued above, formal model selection criteria try to determine the linear regression specification with the best external validation. To verify whether they indeed pick models with external validity, we test their out-of-sample forecasting power and compare it to the corresponding out-of-sample performance of the Bayesian model averaging approach. After all, even a sophisticated trader could only have used prevailing information to estimate his models, not the entire sample period.

In particular, we consider the following predictive regression specifications. We start out with the i.i.d. model (historical mean as forecast, $n = 0$), the seven models that include only one of the seven predictive variables to the forecasting model ($n = 1$), and the all-inclusive model ($n = M = 7$). We then consider the external validity of the five statistical model selection criteria discussed above (adjusted $R^2$, AIC, BIC, FIC, and PIC), the Bayesian weighted model, a model that equally weights all possible regression specifications, and, finally, a model suggested by Engstrom (2003), which we combine with the Bayesian model averaging approach.

In brief, while pointing out the conditional relationship between the equity premium and the dividend-price ratio, Engstrom (2003) argues that within a very broad class of economic models of risk, “unconditional” predictive regression specifications may be misspecified and have almost no power against the specific form of predictability suggested by reasonable treatments of risk. He shows that a very general model of risk implies an intrinsically time-varying re-
relationship between the dividend-price ratio and the conditional equity premium and that the coefficient on the dividend-price ratio represents a conditional covariance between the stochastic discount factor and future pricing kernels and dividend growth. Thus, as a quick and easy first check for state dependence of this quantity, he suggests to model the time-varying coefficient on the dividend-price ratio as a non-stochastic, affine function of a set of predictive variables such as

$$
e_t = \alpha + \beta_{t-1} DPR_{t-1} + \xi_t$$

$$= \alpha + (\beta_0 + \beta_j x_{j,t-1}) DPR_{t-1} + \xi_t.$$  (21)

where $x_{j,t-1}$ represents the set of predictive variables that are expected to drive conditional expectations in the economy (but does no longer include the dividend-price ratio). Again, however, the “true” set of the predictive variables is a very open issue. Therefore, the combination of Engstrom’s (2003) contributions with the Bayesian model averaging approach is only straightforward.

Overall, thus, we examine the out-of-sample performance of 17 different forecasting models. While we focus on monthly observations, our analysis is both based on the rolling and recursive schemes described above. A first set of results is based on $k = 60$ months. This gives a total of 276 monthly out-of-sample observations, from January 1980 to December 2002.

Table 3 and 4 display the following statistics to analyze the properties of the monthly out-of-sample return forecasts and the respective forecast errors: the information coefficient, the regression coefficients of a Mincer-Zarnowitz regression, the root mean squared error (RMSE), the number of negative return forecasts, and the number of months where a statistical model selection criterion retains the i.i.d. no predictability model. The information coefficient is simply the correlation coefficient between the predicted one-period-ahead excess returns and the subsequently realized excess returns (see, e.g., Grinold and Kahn, 2000).

The Mincer-Zarnowitz regression (Mincer and Zarnowitz, 1969) is a regression of the realization on the forecast:

$$e_{t+1} = \kappa + \vartheta E_t (e_{t+1}) + \nu_t.$$  (22)

If the forecast is optimal with respect to the information used to construct it, we would expect a zero intercept ($\kappa = 0$) and unit slope ($\vartheta = 1$).\(^5\)

To save space, we only report the results of the original predictive variables. The corresponding results of the stochastically detrended variables are qualitatively the same and do not affect our overall conclusions in any regard.

Table 3 presents the results for the rolling scheme. In general, the results are disappointing and display quite undesirable properties. The information coefficients are generally small, often even negative (but never significantly different from zero at conventional significance levels). This is particularly true for the all-inclusive model and the adjusted $R^2$, FIC and PIC model selection criteria.

\(^5\)Notice that we do not adjust the $t$-statistics for error in the estimation of the parameters of the prediction model (see, e.g., West (1996) and Bossaerts and Hillion, 1999).
### Table 3: Bayesian model averaging: External validity based on the rolling scheme.

The table displays several statistics examining the properties of the out-of-sample monthly return forecasts and the respective forecast errors generated by a number of different predictive regression specifications. They include the i.i.d. model (IID), the 7 forecasting models that include only one of the following predictive variables: dividend-price ratio (DPR), earnings-price ratio (EPR), term spread (TERM), one-month Swiss interbank rate (IR), realized stock market volatility (VOLA), U.S. TED spread (TED), U.S. default risk spread (DEF), and the all-inclusive model (ALL). In addition, the table also shows the results for the five statistical model selection criteria (ARSq, AIC, BIC, FIC, and PIC), the Bayesian model averaging approach (BAYES), the model that equally weights all possible regression specifications (EQ), and, finally, the model suggested by Engstrom (2003), enhanced with the Bayesian model averaging approach (ENG-BAYES). The information coefficient (IC), the Mincer-Zarnowitz regression, and the root mean squared error (RMSE) are described in the text. NoNF denotes the number of negative forecasts (in percentages) and NoIID denotes the number of months where a statistical selection criterion retains the i.i.d. no predictability model (in percentages). The rolling scheme fixes the estimation window size (k = 60 months) and drops distant observations as recent ones are added. Results are based on monthly observations from January 1980 to December 2002 (276 monthly observations). */**/*** indicate p-values less than 0.1/0.05/0.01.

<table>
<thead>
<tr>
<th></th>
<th>IC</th>
<th>Mincer-Zarnowitz</th>
<th>RMSE</th>
<th>NoNF</th>
<th>NoIID</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Const</td>
<td>Slope coeff.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IID</td>
<td>-0.0015</td>
<td>0.0043</td>
<td>**-0.0120</td>
<td>0.0489</td>
<td>0.1667</td>
</tr>
<tr>
<td>DPR</td>
<td>-0.0088</td>
<td>0.0046</td>
<td>***-0.0430</td>
<td>***0.0496</td>
<td>0.2899</td>
</tr>
<tr>
<td>EPR</td>
<td>0.0374</td>
<td>0.0031</td>
<td>***0.1643</td>
<td>***0.0494</td>
<td>0.2246</td>
</tr>
<tr>
<td>TERM</td>
<td>0.0520</td>
<td>0.0027</td>
<td>***0.2270</td>
<td>***0.0492</td>
<td>0.2572</td>
</tr>
<tr>
<td>IR</td>
<td>0.0503</td>
<td>0.0025</td>
<td>***0.2088</td>
<td>***0.0495</td>
<td>0.2283</td>
</tr>
<tr>
<td>VOLA</td>
<td>-0.0467</td>
<td>0.0062</td>
<td>***-0.2755</td>
<td>***0.0496</td>
<td>0.2283</td>
</tr>
<tr>
<td>TED</td>
<td>0.0200</td>
<td>0.0038</td>
<td>***0.0923</td>
<td>***0.0494</td>
<td>0.2862</td>
</tr>
<tr>
<td>DEF</td>
<td>-0.0026</td>
<td>0.0043</td>
<td>***-0.0146</td>
<td>***0.0493</td>
<td>0.3732</td>
</tr>
<tr>
<td>ALL</td>
<td>-0.0671</td>
<td>*0.0057</td>
<td>***-0.1547</td>
<td>***0.0543</td>
<td>0.3370</td>
</tr>
<tr>
<td>ARSq</td>
<td>-0.0698</td>
<td>*0.0063</td>
<td>***-0.1791</td>
<td>***0.0537</td>
<td>0.2862</td>
</tr>
<tr>
<td>AIC</td>
<td>-0.0203</td>
<td>0.0050</td>
<td>***-0.0566</td>
<td>***0.0524</td>
<td>0.2138</td>
</tr>
<tr>
<td>BIC</td>
<td>-0.0337</td>
<td>0.0055</td>
<td>***-0.1207</td>
<td>***0.0511</td>
<td>0.2065</td>
</tr>
<tr>
<td>FIC</td>
<td>-0.0667</td>
<td>*0.0058</td>
<td>***-0.1608</td>
<td>***0.0539</td>
<td>0.3225</td>
</tr>
<tr>
<td>PIC</td>
<td>-0.0667</td>
<td>*0.0058</td>
<td>***-0.1608</td>
<td>***0.0539</td>
<td>0.3225</td>
</tr>
<tr>
<td>BAYES</td>
<td>-0.0414</td>
<td>0.0057</td>
<td>***-0.1678</td>
<td>***0.0506</td>
<td>0.2391</td>
</tr>
<tr>
<td>EQ</td>
<td>-0.0288</td>
<td>0.0053</td>
<td>***-0.1276</td>
<td>***0.0501</td>
<td>0.2282</td>
</tr>
<tr>
<td>ENG-BAYES</td>
<td>-0.0479</td>
<td>*0.0059</td>
<td>***-0.1955</td>
<td>***0.0506</td>
<td>0.2318</td>
</tr>
</tbody>
</table>
(the latter two obviously retain the same models). The Bayesian weighted model is somewhat better, but still worse than AIC, BIC, and the models that include only one predictive variable (with the realized stock market volatility as exception). The slope coefficients of the Mincer-Zarnowitz regressions are far from \( \vartheta = 1 \); estimates are generally close to zero and even negative in a lot of cases. The RMSE are lowest for the unconditional i.i.d. model and generally increase with the number of predictive variables included in the regression specification. The Diebold and Mariano (1995) statistics indicate that all of the reported out-of-sample RMSE performances are statistically significantly different from the i.i.d no predictability model.\(^6\) They thus all significantly underperform the prevailing mean model. With respect to the number of negative return forecasts (in percentages), usually more than 20% of the predicted excess returns are negative. While this may be expected for linear regression specifications, it should nevertheless be of some concern, as the expected market risk premium should actually be positive (Merton, 1980, and particularly Boudoukh, Richardson, and Smith, 1993). A small number of months (in percentages) where a statistical model selection criterion retains the i.i.d. no predictability model may indicate the existence of return predictability. While AIC, FIC and PIC never decide against predictability, both the adjusted \( R^2 \) criterion and BIC retain in more than 30% of the months the prevailing mean model. Finally, the performance of Engstrom’s (2003) model, enhanced with the Bayesian model averaging approach, is generally no better than all the other forecasting models.

The results given in Table 4 for the recursive scheme are of a similar magnitude. In brief, none of the forecasting models detect reliable out-of-sample predictability, they all display a rather poor out-of-sample performance. While the RMSE statistics are somewhat better compared to the rolling scheme, the information coefficients are generally much worse.

In addition, Table 5 also reports the average values for each of the 7 predictive variables. Average values are computed as \( A'P / (2^M/2) \), where the vector \( A \) is defined as previously and the \( (2^M, 1) \) vector \( P \) contains the respective statistic.

The most interesting result in Table 5 is perhaps the poor average performance of the dividend-price ratio. In both cases of the rolling and the recursive scheme, it exhibits the worst out-of-sample predictive ability.

Goyal and Welch (2003a,b) suggest still another way to look at the results. They suggest a simple, recursive residuals (out-of-sample) graphical approach to evaluating the forecasting ability of the predictive regression specifications. Their simple graphical diagnostic plots the cumulative sum-squared forecast error from the unconditional i.i.d. model minus the cumulative sum-squared forecast error from the respective predictive regression specification:

\[
Net - SSE(\tau) = \sum_{t=1}^{\tau} SE_{t}^{i.i.d.} - SE_{t},
\]

where \( SE_t \) is the squared out-of-sample forecast error in month \( t \). Thus, Figure

\(^6\)Harvey, Leybourne, and Newbold (1997, 1998) modify this statistic for overlapping observations.
<table>
<thead>
<tr>
<th></th>
<th>IC</th>
<th>Mincer-Zarnowitz</th>
<th>RMSE</th>
<th>NoNF</th>
<th>NoIID</th>
</tr>
</thead>
<tbody>
<tr>
<td>IID</td>
<td>*-0.1029</td>
<td>**0.0275</td>
<td>**4.0596</td>
<td>0.0486</td>
<td>0.0000</td>
</tr>
<tr>
<td>DPR</td>
<td>-0.0745</td>
<td>*0.0051</td>
<td>**0.4207</td>
<td>***0.0499</td>
<td>0.5181</td>
</tr>
<tr>
<td>EPR</td>
<td>0.0022</td>
<td>0.0042</td>
<td>***0.0114</td>
<td>***0.0494</td>
<td>0.3986</td>
</tr>
<tr>
<td>TERM</td>
<td>-0.0092</td>
<td>0.0048</td>
<td>***0.0819</td>
<td>***0.0493</td>
<td>0.0906</td>
</tr>
<tr>
<td>IR</td>
<td>-0.0495</td>
<td>0.0073</td>
<td>***0.4041</td>
<td>***0.0492</td>
<td>0.0290</td>
</tr>
<tr>
<td>VOLA</td>
<td>-0.0812</td>
<td>*0.0107</td>
<td>***1.2399</td>
<td>***0.0488</td>
<td>0.0652</td>
</tr>
<tr>
<td>TED</td>
<td>-0.0768</td>
<td>*0.0089</td>
<td>***0.7574</td>
<td>***0.0491</td>
<td>0.0000</td>
</tr>
<tr>
<td>DEF</td>
<td>-0.0625</td>
<td>0.0084</td>
<td>***0.8664</td>
<td>***0.0488</td>
<td>0.0543</td>
</tr>
<tr>
<td>ALL</td>
<td>-0.0965</td>
<td>*0.0054</td>
<td>***0.4152</td>
<td>***0.0508</td>
<td>0.3913</td>
</tr>
<tr>
<td>ARS$^q$</td>
<td>-0.0465</td>
<td>*0.0054</td>
<td>***0.2068</td>
<td>***0.0502</td>
<td>0.2790</td>
</tr>
<tr>
<td>AIC</td>
<td>-0.0505</td>
<td>*0.0053</td>
<td>***0.2109</td>
<td>***0.0504</td>
<td>0.3406</td>
</tr>
<tr>
<td>BIC</td>
<td>-0.0239</td>
<td>0.0049</td>
<td>***0.1173</td>
<td>***0.0497</td>
<td>0.2717</td>
</tr>
<tr>
<td>FIC</td>
<td>-0.0855</td>
<td>*0.0058</td>
<td>***0.4500</td>
<td>***0.0504</td>
<td>0.3188</td>
</tr>
<tr>
<td>PIC</td>
<td>-0.0855</td>
<td>*0.0058</td>
<td>***0.4500</td>
<td>***0.0504</td>
<td>0.3188</td>
</tr>
<tr>
<td>BAYES</td>
<td>-0.0777</td>
<td>*0.0059</td>
<td>***0.4302</td>
<td>***0.0499</td>
<td>0.3261</td>
</tr>
<tr>
<td>EQ</td>
<td>-0.0945</td>
<td>***0.0067</td>
<td>***0.6641</td>
<td>***0.0496</td>
<td>0.2862</td>
</tr>
<tr>
<td>ENG-BAYES</td>
<td>-0.0896</td>
<td>*0.0060</td>
<td>***0.5002</td>
<td>***0.0500</td>
<td>0.3478</td>
</tr>
</tbody>
</table>

Table 4: Bayesian model averaging: External validity based on the recursive scheme.
The table displays several statistics examining the properties of the out-of-sample monthly return forecasts and the respective forecast errors generated by a number of different predictive regression specifications. They include the i.i.d. model (IID), the 7 forecasting models that include only one of the following predictive variables: dividend-price ratio (DPR), earnings-price ratio (EPR), term spread (TERM), one-month Swiss interbank rate (IR), realized stock market volatility (VOLA), U.S. TED spread (TED), U.S. default risk spread (DEF), and the all-inclusive model (ALL). In addition, the table also shows the results for the five statistical model selection criteria (ARS$^q$, AIC, BIC, FIC, and PIC), the Bayesian model averaging approach (BAYES), the model that equally weights all possible regression specifications (EQ), and, finally, the model suggested by Engstrom (2003), enhanced with the Bayesian model averaging approach (ENG-BAYES). The information coefficient (IC), the Mincer-Zarnowitz regression, and the root mean squared error (RMSE) are described in the text. NoNF denotes the number of negative forecasts (in percentages) and NoIID denotes the number of months where a statistical selection criterion retains the i.i.d. no predictability model (in percentages). The recursive scheme uses all available data. Results are based on monthly observations from January 1980 to December 2002 (276 monthly observations). */**/*** indicate p-values less than 0.1/0.05/0.01.
Table 5: Average values for the predictive variables.
The table displays average values of the information coefficient (IC), the root mean squared error (RMSE), and the number of negative return forecasts for the following set of predictive variables: dividend-price ratio (DPR), earnings-price ratio (EPR), term spread (TERM), one-month Swiss interbank rate (IR), realized stock market volatility (VOLA), U.S. TED spread (TED), and U.S. default risk spread (DEF). The results are based on the rolling (with k = 60 months) and the recursive scheme and include monthly observations from January 1980 to December 2002 (276 observations).

<table>
<thead>
<tr>
<th></th>
<th>Rolling Scheme</th>
<th></th>
<th>Recursive Scheme</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IC</td>
<td>RMSE</td>
<td>NoNF</td>
<td>IC</td>
</tr>
<tr>
<td>DPR</td>
<td>-0.0473</td>
<td>0.0519</td>
<td>0.2810</td>
<td>-0.0964</td>
</tr>
<tr>
<td>EPR</td>
<td>-0.0289</td>
<td>0.0518</td>
<td>0.2679</td>
<td>-0.0625</td>
</tr>
<tr>
<td>TERM</td>
<td>-0.0243</td>
<td>0.0518</td>
<td>0.2907</td>
<td>-0.0654</td>
</tr>
<tr>
<td>IR</td>
<td>-0.0258</td>
<td>0.0518</td>
<td>0.2955</td>
<td>-0.0728</td>
</tr>
<tr>
<td>VOLA</td>
<td>-0.0262</td>
<td>0.0520</td>
<td>0.3036</td>
<td>-0.0741</td>
</tr>
<tr>
<td>TED</td>
<td>-0.0168</td>
<td>0.0517</td>
<td>0.3160</td>
<td>-0.0701</td>
</tr>
<tr>
<td>DEF</td>
<td>-0.0222</td>
<td>0.0516</td>
<td>0.3111</td>
<td>-0.0729</td>
</tr>
</tbody>
</table>

4 makes it easy to understand the relative performance of the different forecasting models. A positive value indicates that the regression specification has outperformed the prevailing mean model so far: its forecast error is lower than the one of the unconditional moving average equity premium in a given month.

Figure 4 confirms the results in Table 3 and 4 and shows that all regression specifications, including both the statistical model selection criteria and the Bayesian model averaging approach, practically never outperform the prevailing mean model.

Finally, we also explored whether our conclusions are robust to different values of k, the length of the rolling scheme. Figure 5 shows the information coefficients and the RMSE for different values of k, starting from k = 24 to k = 60 (the above benchmark case). The two graphs on the left plot the average values for each of the seven predictive variables. The graphs on the right show the average values for different values of n, the number of predictive variables retained in the regression specifications. It seems that our original choice of k = 60 is probably not optimal, at least not with respect to the information coefficient. Smaller values of k, say between 30 and 45, may promise somewhat better results, but the information coefficients remain rather modest even then. With respect to the RMSE, however, the specification of the rolling scheme with k = 60 is quite optimal. After all, since it is not a priori clear whether the information coefficient or the RMSE is a more important criterion for the performance of a tactical asset allocation strategy, we may just conclude that investors should mind the dividend-price ratio as predictive variable and should
Figure 4:
Cumulative relative out-of-sample, sum-squared forecast error performance.
This figure plots the cumulative relative out-of-sample, sum-squared error performance, as described in the text. The graphs on the left show the results for the forecasting models that retain only one predictive variable and the all-inclusive specification (ALL, bold). The graphs on the right display the results for the five model selection criteria (ARSq, AIC, BIC, FIC, and PIC) and the Bayesian model averaging approach (BAYES, bold). Both the rolling (with $k = 60$) and recursive scheme include monthly observations from January 1980 to December 2002 (276 observations).
保留仅一个或少数的预测变量。

总结，一致地有Bossaerts和Hillion (1999), Neely和Weller (1999), Goyal和Welch (2003a,b), Schwert (2003), 和 Drobetz和Hoechle (2004), 谁结论出的样外的预测能力的股息-价格比率和其他预测变量是极差的,而且,在 Schwert (2003), 失败,我们的结果在Table 3, 4, 和5也显示没有任何可靠的样外的收益率的预测能力。预测变量中，股息-价格比率显示样外的预测能力在平均最差。此外，超过一个预测变量的包括损害了样外的预测模型的性能。最后，与Avramov（2002）,我们的分析显示样外的预测性能的Bayesian模型 averaging方法不是优于统计的模型选择标准。因此，正确地纳入模型的不确定性一个战术的资产分配模型给人看来也不帮助提高样外的性能的结果的短期的战术资产分配策略。
4 Conclusion

In this paper, we extend the empirical results of Rey (2004) and implement a number of statistical model selection criteria and Avramov’s (2002) Bayesian model averaging approach to analyze the sample evidence on return predictability. Based on Swiss stock market data, we also point out the implications of model uncertainty for tactical asset allocation strategies. We obtain the following general results. First, neither the cumulative posterior probabilities nor the posterior probabilities of the individual forecasting models are constant through time. Which of the predictive variables receives the highest cumulative probability depends rather on the time period under consideration, and it seems difficult to discard any predictive variable as completely worthless. Moreover, the results are not robust to whether the predictive variables are stochastically detrended or not. Second, we conclude that the contributions of the uncertainty attributed to forecast errors, parameter uncertainty, and model uncertainty are highly dependent on the time period under consideration and the initial values of the predictive variables. In contrast to Avramov (2002), model uncertainty is generally not more important than parameter uncertainty. Third, conditional volatilities seem to have considerably increased over the last 20 years. Fourth, investigating the implications of model uncertainty from investment management perspectives, our results do not indicate any reliable out-of-sample return predictability. Among the predictive variables, despite good theoretical reasons, the dividend-price ratio exhibits the worst out-of-sample forecasting ability on average. Moreover, the inclusion of more than one predictive variable rather detoriates the out-of-sample performance of the forecasting models. Finally, in contrast to Avramov (2002), our analysis shows that the out-of-sample performance of the Bayesian model averaging approach is not generally superior to the statistical model selection criteria. These results are robust with respect to the length of the rolling window.

The poor external validity of all the predictive regression specifications probably indicates model non-stationarity: the parameters of the best prediction model change over time. It is still an open question why this is. One potential explanation is that the correct regression specification is actually nonlinear, while both our statistical model selection criteria and the Bayesian model averaging approach chose exclusively among linear models. Still, the statistical model selection criteria pick the best linear prediction model, and the Bayesian model averaging approach averages over the dynamics implied by the set of all possible regression specifications and therefore properly accounts for model uncertainty. It is indeed surprising that none of these approaches seem to work out-of-sample. Consequently, even properly incorporating model uncertainty into a tactical asset allocation model does not seem to help improving the performance of the resulting short-term tactical asset allocation strategies.
A  Description of Data

Appendix A describes the Swiss stock market data used throughout the paper.

A.1  The Stock Market Index

The empirical results in this paper are based on Swiss stock market data. Monthly data are used throughout, spanning 336 months from January 1975 to December 2002. We also investigate two subsamples of equal length (each 168 months). The first subsample uses data from January 1975 to December 1988, covering the first half of the total time period, the second subsample is based on data from January 1989 to December 2002, covering the second half of the full sample.

The Swiss stock market index is a value-weighted aggregate of the following industry sector indices:

- Airlines and Transportation (Datastream Mnemonic: AIRLNBV),
- Financials (BANKSBV, INSURBV),
- Food (FOODSBV, BREWSB),
- Industrials (GENINBV),
- Pharma (PHARMBV),
- Retailers (MULTIBV, FDRETBV),
- Utilities (UTILSBV), and
- Other Businesses (OTHBUBV, LESURBV).

All return series include dividends (total returns). To obtain continuously compounded real returns, total returns are deflated using monthly rates of change in the Consumer Price Index (CPI), provided by the Swiss National Bank. Continuously compounded excess returns are less the prevailing one-month Swiss interbank rate (SWIBK1M) at the beginning of the month.

Figure 6 shows the nominal price index (January 1975 = 100) and the respective continuously compounded return series.

Table 6 presents summary statistics for the continuously compounded excess and real returns.

Mean Reversion in Stock Market Returns  An interesting and important sidestep is to consider whether the stock market returns exhibit mean reversion over the respective sample periods. In particular, we calculate the $q$-period variance ratio statistic $VR(q)$ as

$$VR(q) \equiv \frac{Var(e_{t-q})}{qVar(e_t)} = 1 + 2 \sum_{j=1}^{q-1} \left( 1 - \frac{j}{q} \right) Corr(e_t, e_{t+j}),$$

(24)
Figure 6:
The Swiss stock market.
The graph on the left shows the time series of the aggregated nominal price index (January 1975 = 100). The graph on the right plots the respective continuously compounded return series. Monthly data are used from January 1975 to December 2002.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>6.02%</td>
<td>7.40%</td>
<td>6.65%</td>
</tr>
<tr>
<td>Median</td>
<td>0.86%</td>
<td>0.86%</td>
<td>0.51%</td>
</tr>
<tr>
<td>Max.</td>
<td>16.81%</td>
<td>16.95%</td>
<td>16.81%</td>
</tr>
<tr>
<td>Min.</td>
<td>-27.57%</td>
<td>-27.72%</td>
<td>-27.57%</td>
</tr>
<tr>
<td>Volatility</td>
<td>16.14%</td>
<td>16.19%</td>
<td>14.75%</td>
</tr>
<tr>
<td>Skewness</td>
<td>-1.13</td>
<td>-1.14</td>
<td>-1.42</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>8.25</td>
<td>8.36</td>
<td>14.44</td>
</tr>
<tr>
<td>J.-Bera</td>
<td>457.6</td>
<td>476.0</td>
<td>971.8</td>
</tr>
<tr>
<td>Prob.</td>
<td>***0.00</td>
<td>***0.00</td>
<td>***0.00</td>
</tr>
<tr>
<td>Obs.</td>
<td>336</td>
<td>336</td>
<td>168</td>
</tr>
</tbody>
</table>

Table 6:
Summary statistics of the Swiss stock market data.
The table presents summary statistics of continuously compounded real and excess monthly Swiss market stock returns. These include mean (ann.), median, maximum and minimum value, volatility (ann.), skewness, and kurtosis. The table includes the Jarque and Bera (1980) test of normality. Estimates are given for three different time periods: 1975 to 2002 (full sample), 1975 to 1988, and 1989 to 2002. **/***/*** indicate p-values less than 0.1/0.05/0.01.
Figure 7: 
Mean reversion and variance ratio statistics.
The graphs show variance ratio statistics for time horizons varying from one to 60 months. A variance ratio below one indicates mean reversion, a ratio above one mean aversion. The graph on the left is based on data from 1975 to 1988. The graph on the right uses monthly data from 1989 to 2002. The fine lines plot variance ratios calculated based on the full sample (1975 to 2002). The dotted lines are based on the standard approach given in equation (29). Bold lines correspond to coefficient and bias-adjusted variance ratio statistics (Lo and MacKinlay, 1988, 1989).

with $e_{t-q} = e_t + e_{t+1} + ... + e_{t+q-1}$ and $Corr(e_t, e_{t+j})$ is the jth-order autocorrelation coefficient of $\{e_t\}$. If returns are positively autocorrelated, variances grow faster than linearly and the variance ratio is above one for $q > 1$, $VR(q) > 1$. Alternatively, in the presence of negative autocorrelation, the variance of the sum of one-month returns is smaller than the sum of the one-month return’s variances; hence $VR(q) < 1$, variances grow slower than linearly. This is the case of mean reversion. If returns are i.i.d., $VR(q) = 1$.

Figure 7 shows variance ratio statistics for time horizons, $q$, varying from one to 60 months. The results reveal an interesting pattern. Based on the data covering the first half of the total time period (1975 to 1988), continuously compounded excess returns seem to mean revert. The variance ratio statistic is well below one for long horizons. Recently, however, the mean reversion pattern has completely disappeared. Rather than mean reversion, the Swiss stock market exhibits strong mean aversion since 1989. However, if the estimates of the variance ratio statistics are based on overlapping $q$-period returns and corrected for the bias in the variance estimators (Lo and MacKinlay, 1988, 1989), the evidence on mean reversion over the first subsample vanishes. Over the recent subsample, the tendency of mean aversion is even amplified.

28
References


