DAVID REY

MARKET TIMING AND MODEL UNCERTAINTY: AN EXPLORATORY STUDY FOR THE SWISS STOCK MARKET

David Rey (david.rey@unibas.ch)
Department of Finance, Wirtschaftswissenschaftliches Zentrum (WWZ)
University of Basel, Holbeinstrasse 12, CH – 4051 Basel, Switzerland
Tel.: +41 (0) 61 267 33 07, Fax: +41 (0) 61 267 08 98

Abstract. We use statistical model selection criteria and Avramov's (2002) Bayesian model averaging approach to analyze the sample evidence of stock market predictability in the presence of model uncertainty. The empirical analysis for the Swiss stock market is based on a number of predictive variables found important in previous studies of return predictability. We find that it is difficult to discard any predictive variable as completely worthless, but that the posterior probabilities of the individual forecasting models as well as the cumulative posterior probabilities of the predictive variables are time-varying. Moreover, the estimates of the posterior probabilities are not robust to whether the predictive variables are stochastically detrended or not. The decomposition of the variance of predicted future returns into the components parameter uncertainty, model uncertainty, and the uncertainty attributed to forecast errors indicates that the respective contributions strongly depend 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 demonstrate the implications of model uncertainty for market timing strategies. In general, our 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. Again in contrast to AVRAMOV (2002), our analysis suggests that the out-of-sample performance of the Bayesian model averaging approach is not superior to the statistical model selection criteria. Consequently, model averaging does not seem to help improve the performance of the resulting short-term market timing strategies.

1. Introduction

Recent advances in asset pricing theory and the mounting empirical evidence of stock market predictability seem to have persuaded the majority of researchers to abandon the constant expected returns paradigm. The time variation and predictability of excess returns, labeled as a "new fact in finance" by COCHRANE (1999), is so widely accepted that it has generated a new wave of conditional asset pricing and portfolio choice models [see, e.g., BRENNAN et al. (1997), CAMPBELL and VICEIRA (1999, 2002), BARBERIS (2000), and XIA (2001)]. At the same time, however, there is less consensus on what drives this predictability. BEKAERT (2001) differentiates between three possibilities: it may reflect time-varying risk premiums (which he calls the "risk view"), it may reflect irrational behavior on the part of market participants (the "behavioral view"), or it may simply not be present in the data – a statistical fluke due to poor statistical inference (the "statistical view"). Of course, whether stock market predictability is consistent with market efficiency can only be interpreted in conjunction with an intertemporal equilibrium model of the economy. All theoretical attempts at interpretation of predictability will thus be model-dependent, and hence inconclusive. Nevertheless, recent advances in asset pricing theory seem to demonstrate that a certain degree of time-varying expected returns is necessary to reward investors for bearing certain dynamic risks associated with the business cycle. Loosely, it is claimed that the equity premium rises during an economic slow-down and falls during periods of economic growth, so that expected returns and business conditions move in opposite directions [see, e.g., FAMA and FRENCH (1989), CHEN (1991), FAMA (1991), and FERSON and HARVEY (1991)].

However, certain aspects of the empirical research on stock market predictability remain controversial. Specifically, considering the long list of authors criticizing the statistical methodologies in the literature about return predictability, it seems that the statistical view gains increasing credibility. Indeed, a number of recent contributions suggest that any evidence of return predictability may have more to do with poorly behaved test statistics than with stock market predictability.[1] Properly adjusting for small sample biases, near unit roots, and other statistical issues associated primarily with long-horizon regressions weakens and often reverses many of the standard conclusions.

Moreover, existing equilibrium pricing theories are not explicit about the predictive variables. The reported empirical evidence of return predictability is thus subject to data-over fitting concerns. For example, BOSSAERTS and HILLION (1999), NEELY and WELLER (1999), and particularly GOYAL and WELCH (2003a,b) 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. In addition, the multiplicity of potential predictive variables makes the empirical evidence difficult to interpret: we may find a predictive variable statistically significant based on a particular collection of predictive variables and sample period, but often not based on a competing specification or time period.

A recent surge of research has thus increased attention to parameter and model uncertainty, and their implications for optimal portfolio choice. In particular, the perspective of a Bayesian investor (who uses the sample evidence to update prior beliefs about the regression parameters and models) seems to be particularly suitable to deal with parameter and model uncertainty. We follow this literature and critically apply statistical model selection criteria as well as AVRAMOV's (2002) Bayesian model averaging approach to analyze stock market predictability, model uncertainty, and their implications for short-term market timing strategies.

Based on Swiss stock market data from 1975 to 2002, we show that it is difficult to discard any predictive variable as completely worthless, but that the posterior probabilities of the individual forecasting models as well as the cumulative posterior probabilities of the predictive variables are time-varying. Moreover, the estimates of the posterior probabilities are not robust to whether the predictive variables are stochastically detrended or not. The decomposition of the variance of predicted future returns into the components parameter uncertainty, model uncertainty, and the uncertainty attributed to forecast errors indicates that the respective contributions strongly depend 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. 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. Furthermore, the inclusion of more than one predictive variable rather deteriorates the out-of-sample performance of the forecasting models. Finally, again in contrast to AVRAMOV (2002), our analysis shows that the out-of-sample performance of the Bayesian model averaging approach is not superior to the statistical model selection criteria. Thus, model averaging does not seem to help improve the performance of the resulting short-term market timing strategies.

The remainder of the paper proceeds as follows. The next section summarizes 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.

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

$$e_t = \alpha_j + \boldsymbol{\beta}_j' \boldsymbol{x}_{j,t-1} + \xi_{j,t}, \tag{1}$$

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

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 constant risk premium. In contrast, when $n = M$, all M predictive variables are suspected relevant.

Given a set of M predictive variables, 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 values. However, the available time series is often limited, especially so when the ultimate purpose is to obtain the model with the best external 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 by construction. 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 over fitting, 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 to unit-root non-stationarities [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 as

$$\boldsymbol{e}' = [e_1 \cdots e_T] \quad \text{and}$$

$$\boldsymbol{X}_j = \begin{pmatrix} 1 & \boldsymbol{x}_{j,0}' \\ \vdots & \vdots \\ 1 & \boldsymbol{x}_{j,T-1}' \end{pmatrix}, \tag{2}$$

respectively, and the sum of squared regression errors by

$$SSE_j = \left(\boldsymbol{e} - \boldsymbol{X}_j\left(\boldsymbol{X}_j'\boldsymbol{X}_j\right)^{-1}\boldsymbol{X}_j'\boldsymbol{e}\right)^T \times\left(\boldsymbol{e} - \boldsymbol{X}_j\left(\boldsymbol{X}_j'\boldsymbol{X}_j\right)^{-1}\boldsymbol{X}_j'\boldsymbol{e}\right) \tag{3}$$

Akaike's information criterion is then given by

$$AIC_j = T\ln\left(SSE_j/T\right) + 2(n+1), \tag{4}$$

and Schwarz's criterion by

$$BIC_j = T\ln\left(SSE_j/T\right) + \ln(T)(n+1). \tag{5}$$

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

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

and the posterior information criterion

$$PIC_j = SSE_M\left(\frac{SSE_j}{SSE_M} - 1\right) + \frac{SSE_M}{T-(M+1)} \times \ln\left(\frac{\frac{|X_j'X_j|}{SSE_M}}{T-(M+1)}\right) \tag{7}$$

In each case, the regression specification is chosen that minimizes the respective criterion function (BOSSAERTS and HILLION [1999, Appendix A, equations (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 were the "true" regression specification. To 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 integrates model uncertainty in a more sensible way.

2.2 The Bayesian Model Averaging Approach

Specifically, the 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 j, denoted M_j, is given by

$$p\left(M_j|z\right) = \frac{p\left(z|M_j\right)p\left(M_j\right)}{\sum_{i=1}^{2^M} p\left(z|M_j\right)p\left(M_j\right)}, \tag{8}$$

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 $p(z \mid M_j)$ is the marginal likelihood of M_j given by

$$p(z|M_j) = \frac{\ell\left(\boldsymbol{\theta}_j,\ z,\ M_j\right)p\left(\boldsymbol{\theta}_j|M_j\right)}{p\left(\boldsymbol{\theta}_j|z,\ M_j\right)}, \tag{9}$$

with $p(\boldsymbol{\theta}_j \mid M_j)$ and $p(\boldsymbol{\theta}_j \mid z, M_j)$ the joint prior and posterior distributions of the model-specific parameters, $\ell(\boldsymbol{\theta}_j, \boldsymbol{z}, M_j)$ the likelihood function pertaining to M_j, and, finally, $\boldsymbol{\theta}_j = (\alpha_j, \boldsymbol{\beta}_j, \sigma_{j,\xi})$ the set of regression parameters.

AVRAMOV (2002, equation 10) shows that the log marginal likelihood is given by

$$\ln p(z|M_j) = -\frac{T}{2}\ln(\pi) + \frac{T_{j,0}-n-1}{2} \times \ln\left(T_{j,0}s^2\right) - \frac{T_j^*-n-1}{2} \times \ln\left(S_j\right) - \ln\left(\Gamma\left(\frac{T_{j,0}-n-1}{2}\right)\right) + \ln\left(\Gamma\left(\frac{T_j^*-n-1}{2}\right)\right) - \frac{n+1}{2}\ln\left(\frac{T_j^*}{T_{j,0}}\right), \tag{10}$$

where

$$S_j = T_j^*\left(s^2+\widehat{\mu}^2\right) - \frac{T}{T_j^*}\left(T_{j,0}\left[\widehat{\mu}\ \widehat{\mu}\ \widehat{\boldsymbol{x}}_j'\right] + \boldsymbol{e}'\boldsymbol{X}_j\right) \times \left(\boldsymbol{X}_j'\boldsymbol{X}_j\right)^{-1}\left(T_{j,0}\left[\widehat{\mu}\ \widehat{\mu}\ \widehat{\boldsymbol{x}}_j'\right]^T + \boldsymbol{X}_j'\boldsymbol{e}\right), \tag{11}$$

and

$$\widehat{\mu} = \frac{1}{T}\sum_{t=1}^{T} e_t,\ s^2 = \frac{1}{T}\sum_{t=1}^{T}\left(e_t-\widehat{\mu}\right)^2,\quad \widehat{\boldsymbol{x}}_j = \frac{1}{T}\sum_{t=0}^{T-1}\boldsymbol{x}_{j,t}. \tag{12}$$

Note that $\Gamma(y)$ stands for the Gamma function evaluated at y. T is the actual sample size and $T^*_j = T + T_{j,0}$, where $T_{j,0}$ determines the strength of the informative prior.[2]

For the i.i.d. model ($n = 0$), we simply have $S_{iid} = T^*_{iid}s^2$.

Given these posterior probabilities, AVRAMOV (2002) proposes the cumulative posterior probabilities of the predictive variables to summarize the weight of the respective predictive variables in the weighted forecasting model. Cumulative posterior probabilities are computed as $\boldsymbol{A}'\boldsymbol{P}$, where $\boldsymbol{A}$ 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 $\boldsymbol{P}$ 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 $z'_{j,t} = [e_t \boldsymbol{x}'_{j,t}]$ denote the data-generating process corresponding to model j and assume that the evolution of $\boldsymbol{z}'_{j,t}$ is given by

$$z_{j,t} = \boldsymbol{a}_j + \boldsymbol{B}_j\boldsymbol{x}_{j,t-1} + \xi_{j,t}, \tag{13}$$

with $\xi_{j,t} \sim$ i.i.d. $N(\boldsymbol{0}, \boldsymbol{V}_j)$. 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 $\boldsymbol{\theta}_j = (\alpha_j, \boldsymbol{\beta}_j, \sigma_{j,\xi})$. According to AVRAMOV (2002, equation 19), it is given by

$$p(e_{T\to T+\tau}|\boldsymbol{z}) = \sum_{j=1}^{2^M} p\left(M_j|\boldsymbol{z}\right) \int p\left(e_{T\to T+\tau}|\boldsymbol{\theta}_j,\ z,\ M_j\right) \times p\left(\boldsymbol{\theta}_j|z,\ M_j\right)d\boldsymbol{\theta}_j, \tag{14}$$

where τ is the planning horizon in months and $e_{T\to T+\tau} \equiv e_{T+1} + e_{T+2} + \ldots + e_{T+\tau}$. Since an analytical solution for the integral in equation (14) does not exist when $\tau > 1$, the empirical implementation is based on Monte Carlo integration.

First, a model M_j is drawn with probability $p(M_j \mid z)$. Second, we sample from the posterior distribution by first drawing from the marginal $p(\boldsymbol{V}^{-1}{}_j \mid z)$, a Wishart distribution. Then, given $\boldsymbol{V}_j$, we draw $\boldsymbol{C}_j = [\boldsymbol{a}_j' \boldsymbol{B}_j']$ from the conditional $p(\boldsymbol{C}_j \mid \boldsymbol{V}_j, z)$, a multivariate Normal distribution. Repeating this many times gives an accurate representation of the posterior distribution. Third, for each draw of $\boldsymbol{a}_j$, $\boldsymbol{B}_j$ and $\boldsymbol{V}_j$ from the posterior $p(\boldsymbol{a}_j, \boldsymbol{B}_j, \boldsymbol{V}_j \mid z)$, we sample from the Normal distribution with mean vector

$$\begin{aligned}\boldsymbol{\mu}_{j,T\to T+\tau} = {} & \tau \boldsymbol{a}_j + (\tau-1)\boldsymbol{B}_{j,0}\boldsymbol{a}_j \\ & + (\tau-2)\boldsymbol{B}_{j,0}^2\boldsymbol{a}_j + \ldots + \boldsymbol{B}_{j,0}^{\tau-1}\boldsymbol{a}_j \\ & + \left(\boldsymbol{B}_{j,0}^1 + \boldsymbol{B}_{j,0}^2 + \ldots + \boldsymbol{B}_{j,0}^{\tau}\right)\boldsymbol{z}_{j,T},\end{aligned} \tag{15}$$

and variance matrix

$$\begin{aligned}\boldsymbol{V}_{j,T\to T+\tau} = {} & \boldsymbol{V}_j + \left(\boldsymbol{I}_j + \boldsymbol{B}_{j,0}\right)\boldsymbol{V}_j\left(\boldsymbol{I}_j + \boldsymbol{B}_{j,0}\right)^T \\ & + \left(\boldsymbol{I}_j + \boldsymbol{B}_{j,0} + \boldsymbol{B}_{j,0}^2\right)\boldsymbol{V}_j \\ & \times\left(\boldsymbol{I}_j + \boldsymbol{B}_{j,0} + \boldsymbol{B}_{j,0}^2\right)^T + \ldots + \\ & + \left(\boldsymbol{I}_j + \boldsymbol{B}_{j,0} + \boldsymbol{B}_{j,0}^2 + \ldots + \boldsymbol{B}_{j,0}^{\tau-1}\right) \\ & \times\boldsymbol{V}_j\left(\boldsymbol{I}_j + \boldsymbol{B}_{j,0} + \boldsymbol{B}_{j,0}^2 + \ldots + \boldsymbol{B}_{j,0}^{\tau-1}\right)^T,\end{aligned} \tag{16}$$

with $\boldsymbol{B}_{j,0} = [\boldsymbol{0}\ \boldsymbol{B}_j]$ and $\boldsymbol{0}$ a $(n + 1,1)$ vector of zeros. This gives a large sample of the predictive distribution $p(e_{T\to T+\tau}|z)$.[3] The Bayesian weighted predictive distribution of cumulative excess returns may be used to compute the optimal allocation to equities when taking stock market predictability, parameter uncertainty, and model uncertainty into account, or, as below, to decompose the variance of the predicted returns into parameter uncertainty, model uncertainty, and the uncertainty attributed to forecast errors.

2.2.2 Variance Decomposition

Based on the weighted predictive distribution given in equation (14), AVRAMOV (2002) shows that predicted future returns are subject to three sources of uncertainty: (i) parameter uncertainty, (ii) model uncertainty, and (iii) the uncertainty attributed to forecast errors. In particular, AVRAMOV [2002, equation (25)] shows that the variance of the predicted excess returns can be decomposed as

$$\begin{aligned}Var(e_{T\to T+\tau}|z) = {} & \sum_{j=1}^{2^M} p(M_j|z)\Big(E(\varphi_j) \\ & + Var(\lambda_j) + \left(E(\varphi_j) - \widetilde{\lambda}\right)^2\Big),\end{aligned} \tag{17}$$

where $E(\varphi_j)$ and $\mathrm{Var}(\lambda_j)$ are the two variance components attributed to forecast errors and parameter uncertainty, respectively, and φ_j and λ_j denote the first elements of the variance matrix and the mean vector given in equations (15) and (16), respectively. The model uncertainty component is then given by

$$\sum_{j=1}^{2^M} p(M_j|z)\left(E(\varphi_j) - \widetilde{\lambda}\right)^2, \tag{18}$$

where

$$\widetilde{\lambda} = \sum_{j=1}^{2^M} p(M_j|z)E(\lambda_j) \tag{19}$$

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

3. Empirical Results

The following empirical examination analyzes stock market predictability, model uncertainty, and their implications for the corresponding market timing strategies. It also compares the out-of-sample performance of the statistical model selection criteria with 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 (including dividends) over January 1975 through December 2002.

In deciding which predictive variables to include, attention was given to those variables found important in previous studies of return predictability. CAMPBELL and SHILLER (1988b) and FAMA and FRENCH (1988, 1989), for example, are among the first who document that the dividend yield and particularly the dividend–price ratio on aggregate stock portfolios predict future (long-horizon) (stock market) returns. Other examples of predictive variables include short-term interest rates [e.g., CAMPBELL (1991)], yield spreads between long-term and short-term interest rates and between low- and high-quality bond yields [e.g., KEIM and STAMBAUGH (1986) and FAMA and FRENCH (1989)], stock market volatility [e.g., FRENCH et al. (1987) and GOYAL and SANTA-CLARA (2003)], Eurodollar-U.S. Treasury (TED) spread [e.g., FERSON and HARVEY (1993)], book-to-market ratios [e.g., KOTHARI and SHANKEN (1997) and PONTIFF and SCHALL (1998)], dividend–payout and price–earnings ratios [e.g., LAMONT (1998) and CAMPBELL and SHILLER (1988a)], and more complex measures based on analysts' forecasts [LEE et al. (1999)]. Recently, BAKER and WURGLER (2000) have shown that the share of equity in new finance is a negative predictor of future equity returns. LETTAU and LUDVIGSON (2001) find evidence of predictability using a consumption-wealth ratio, the level of consumption relative to income and wealth.

Of course, there is a natural concern about collective "data-snooping" by a series of researchers [LO and MACKINLAY (1990), FOSTER et al. (1997), and FERSON et al. (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 up to and including the recent bear market.

In what follows, each of the 2^M competing predictive regression specifications considered thus 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).[4]

Motivated by the recent contributions by FERSON et al. (2003, 2004), a second subset includes the same $M = 7$ predictive variables, transformed as

$$\boldsymbol{x}_{t-1}^{*} = \boldsymbol{x}_{t-1} - \frac{1}{12}\sum_{\tau=1}^{12} \boldsymbol{x}_{t-1-\tau} \tag{20}$$

We thus 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. While this stochastic detrending method has already been used by CAMPBELL (1991) and HODRICK (1992), FERSON et al. (2003, 2004) show that this is the most practically useful insurance against spurious regression bias. 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 these two data subsets.

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 of 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 of return predictability is modest when the variables are stochastically detrended. Over the recent subsample, however, a number of predictive variables is statistically significant and the adjusted R^2s 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. Equation (10) shows how to

Table 1: Multiple Regressions of Monthly Excess Returns on Predictive Variables: The All-Inclusive Regression Specifications

	DPR	*EPR*	*TERM*	*IR*	*VOLA*	*TED*	*DEF*	*Adj. R^2*
				1975:01 – 2002:12				
SD	−0.003	0.002	*0.016	0.012	−0.003	0.006	−0.003	0.26%
	0.003	−0.003	−0.003	−0.005	−0.003	0.002	0.003	−0.50%
				1975:01 – 1988:12				
SD	−0.002	***0.013	0.004	0.002	−0.006	0.001	0.003	6.07%
	0.000	0.004	0.004	0.002	−0.006	−0.001	0.005	0.05%
				1989:01 – 2002:12				
SD	***0.038	−0.026	−0.002	−0.015	−0.000	***0.014	*−0.010	6.95%
	0.009	*−0.013	*−0.015	*−0.013	0.001	***0.013	−0.003	6.24%

Note:
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 the time period from January 1989 to December 2002. *,**,*** indicate *p*-values less than 10%, 5%, 1% (using White standard errors). All predictive variables are standardized with mean zero and variance one.

compute the marginal likelihood for every model, and equation (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.[5]

Table 2 displays cumulative posterior probabilities, $\boldsymbol{A}'\boldsymbol{P}$.

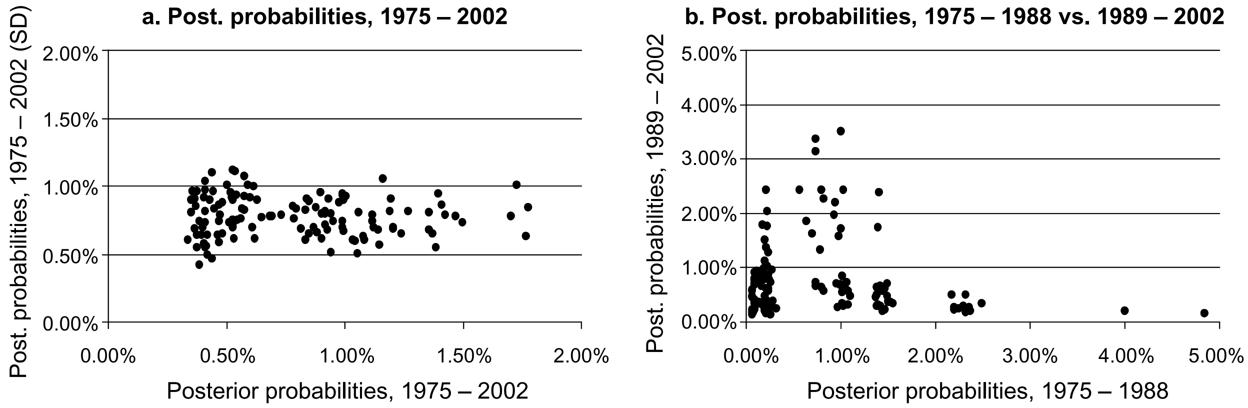
Over the whole sample period, the cumulative posterior probabilities range from 47.32% for the U.S. default risk spread to 69.62% for the term spread, suggesting that the U.S. default risk spread and the term spread should appear in the weighted return-forecasting model with probabilities of 47.32% and 69.62%, respectively. However, when the predictive variables are stochastically detrended, the cumulative posterior probabilities are less dispersed and the term spread no longer receives 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%, which is significantly above 50% at the 10% significance level. 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, mirroring the findings summarized in Table 1, 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. 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

Table 2: Cumulative Posterior Probabilities

	DPR	*EPR*	*TERM*	*IR*	*VOLA*	*TED*	*DEF*
			1975:01 – 2002:12				
	47.42%	47.40%	69.62%	57.16%	52.26%	52.91%	47.32%
SD	48.02%	51.56%	48.74%	52.60%	54.54%	49.86%	55.99%
			1975:01 – 1988:12				
	42.16%	*89.57%	41.63%	40.74%	49.58%	40.14%	41.60%
SD	49.42%	50.70%	50.53%	48.35%	61.34%	47.65%	57.48%
			1989:01 – 2002:12				
	68.95%	57.65%	51.34%	50.98%	45.53%	77.47%	60.76%
SD	48.33%	66.99%	52.50%	50.81%	43.51%	83.42%	44.36%

Note:
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 monthly data from January 1975 to December 2002. Estimates for the two subsamples are showed below. The first subsample uses monthly data from January 1975 to December 1988; the second subsample is based on the time period from January 1989 to December 2002. A Monte Carlo analysis tests whether the Bayesian approach is able to recover the data-generating process correctly and whether the cumulative posterior probabilities are significantly different from 0.5. *,**,*** indicate *p*-values less than 10%, 5%, 1%.

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 monthly data from 1989 to 2002 (with original predictive variables).

detrended, predictive variables. Both graphs reveal that the resulting posterior probabilities are very unstable, both regarding whether the predictive variables are stochastically detrended or not and the time period under consideration. Compared to the original predictive variables, 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 for our set of predictive variables. 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

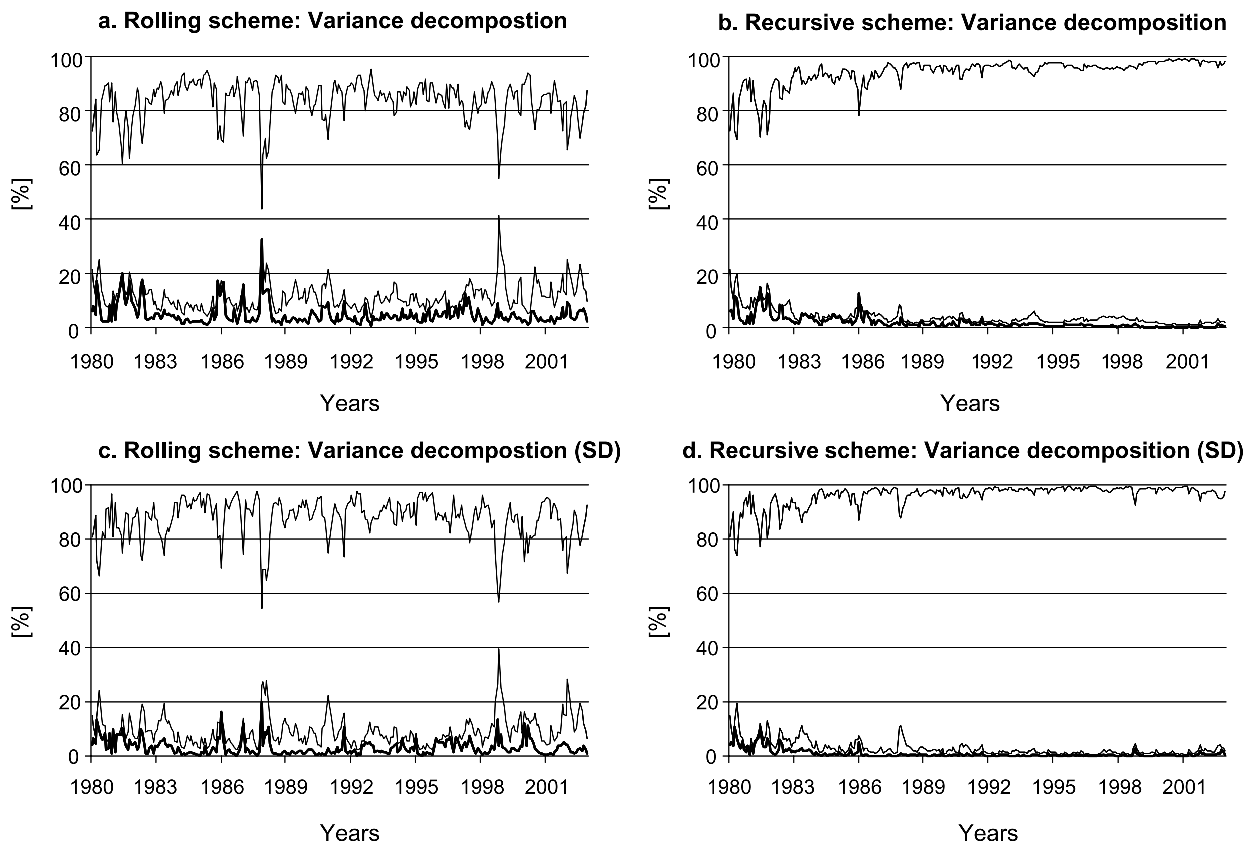
As described above, we perform the variance decomposition of predicted future returns into the components parameter uncertainty, model uncertainty, and the uncertainty attributed to forecast errors. In contrast to AVRAMOV (2002), where the variance decomposition is solely based on the full sample and a single set of predictive variables, $\boldsymbol{x}_{j,T}$, 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., AKGIRAY (1989)], fixes the estimation window size and drops distant observations as recent ones are added. The model parameters are thus first estimated with data from 1 to k, next with data from 2 to

$k + 1, \ldots$, 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 $k+1, \ldots$, 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. Our dynamic approach thus corresponds to the out-of-sample analysis following below and is much more appropriate for investors ultimately concerned with the out-of-sample performance of corresponding market timing strategies than AVRAMOV's (2002) static approach.[6]

Figure 2 shows the resulting time series of the contributions of the three components to the overall uncertainty about predicted returns. The plan-

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 graphs on the left use 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.

ning horizon is one month ($\tau = 1$). 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 contribution of the uncertainty attributed to forecast errors is 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.

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 statistical model selection criteria.

3.4 External Validation: Out-of-Sample Evidence

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 the most 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: 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 weights all possible regression specifications equally, 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 "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 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

$$\begin{aligned} e_t &= \alpha + \beta_{t-1} DPR_{t-1} + \xi_t \\ &= \alpha + \left(\beta_0 + \boldsymbol{\beta}_j' \boldsymbol{x}_{j,t-1}\right) DPR_{t-1} + \xi_t, \end{aligned} \tag{21}$$

where $\boldsymbol{x}_{j,t-1}$ represents the set of predictive variables that are expected to drive conditional expectations in the economy. Again, however, the "true" set of the predictive variables is unknown. Therefore, the combination of ENGSTROM's (2003) contribution with the Bayesian model averaging approach is only straightforward.

Overall, we thus 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.

Tables 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

$$e_{t+1} = \kappa + \nu E_t(e_{t+1}) + \phi_t. \tag{22}$$

If the forecast is optimal with respect to the information used to construct it, the null hypothesis is $\kappa = 0$ and $\nu = 1$.[7]

To save space, we only report the results of the original predictive variables. The corresponding

Table 3: Bayesian Model Averaging: External Validity Based on the Rolling Scheme

		Mincer–Zarnowitz				
	IC	*Constant κ*	*Slope ν*	*RMSE*	*NoNF*	*NoIID*
IID	−0.0015	0.0043	**−0.0120	0.0489	0.1667	–
DPR	−0.0088	0.0046	***−0.0430	***0.0496	0.2899	–
EPR	0.0374	0.0031	***0.1643	***0.0494	0.2246	–
TERM	0.0520	0.0027	***0.2270	***0.0492	0.2572	–
IR	0.0503	0.0025	***0.2088	***0.0495	0.2283	–
VOLA	−0.0467	0.0062	***−0.2755	***0.0496	0.2283	–
TED	0.0200	0.0038	***0.0923	***0.0494	0.2862	–
DEF	−0.0026	0.0043	***−0.0146	***0.0493	0.3732	–
ALL	−0.0671	*0.0057	***−0.1547	***0.0543	0.3370	–
Adj. R^2	−0.0698	*0.0063	***−0.1791	***0.0537	0.2862	0.0109
AIC	−0.0203	0.0050	***−0.0566	***0.0524	0.2138	0.0725
BIC	−0.0337	0.0055	***−0.1207	***0.0511	0.2065	0.3297
FIC	−0.0667	*0.0058	***−0.1608	***0.0539	0.3225	0.0000
PIC	−0.0667	*0.0058	***−0.1608	***0.0539	0.3225	0.0000
BAYES	−0.0414	0.0057	***−0.1678	***0.0506	0.2391	–
EQ	−0.0288	0.0053	***−0.1276	***0.0501	0.2282	–
ENG-BAYES	−0.0479	*0.0059	***−0.1955	***0.0506	0.2318	–

Note:
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 (Adj. R^2, AIC, BIC, FIC, and PIC), the Bayesian model averaging approach (BAYES), the model that weights all possible regression specifications equally (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 10%, 5%, 1%.

Table 4: Bayesian Model Averaging: External Validity Based on the Recursive Scheme

	IC	*Mincer–Zarnowitz* Constant κ	Slope ν	*RMSE*	*NoNF*	*NoIID*
IID	*−0.1029	**0.0275	**−4.0596	0.0486	0.0000	–
DPR	−0.0745	*0.0051	***−0.4207	***0.0499	0.5181	–
EPR	0.0022	0.0042	***0.0114	***0.0494	0.3986	–
TERM	−0.0092	0.0048	**−0.0819	***0.0489	0.0906	–
IR	−0.0495	0.0073	***−0.4041	***0.0492	0.0290	–
VOLA	−0.0812	*0.0107	***−1.2399	***0.0488	0.0652	–
TED	−0.0768	*0.0089	***−0.7574	***0.0491	0.0000	–
DEF	−0.0625	0.0084	**−0.8664	***0.0488	0.0543	–
ALL	−0.0965	*0.0054	***−0.4152	***0.0508	0.3913	–
Adj. R^2	−0.0465	*0.0054	***−0.2068	***0.0502	0.2790	0.3116
AIC	−0.0505	*0.0053	***−0.2109	***0.0504	0.3406	0.0000
BIC	−0.0239	0.0049	***−0.1173	***0.0497	0.2717	0.3297
FIC	−0.0855	*0.0058	***−0.4050	***0.0504	0.3188	0.0000
PIC	−0.0855	*0.0058	***−0.4050	***0.0504	0.3188	0.0000
BAYES	−0.0777	*0.0059	***−0.4302	***0.0499	0.3261	–
EQ	−0.0945	**0.0067	***−0.6641	***0.0596	0.2862	–
ENG−BAYES	−0.0896	*0.0060	***−0.5002	***0.0500	0.3478	–

Note:
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 (Adj. R^2, AIC, BIC, FIC, and PIC), the Bayesian model averaging approach (BAYES), the model that weights all possible regression specifications equally (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 10%, 5%, 1%.

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 (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). Estimates of the slope coefficients of the Mincer–Zarnowitz regressions are far from $\nu = 1$; 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. 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

Table 5: Average Values for the Predictive Variables

	Rolling Scheme			*Recursive Scheme*		
	IC	*RMSE*	*NoNF*	*IC*	*RMSE*	*NoNF*
DPR	−0.0473	0.0519	0.2810	−0.0964	0.0505	0.3941
EPR	−0.0289	0.0518	0.2679	−0.0625	0.0503	0.3720
TERM	−0.0243	0.0518	0.2907	−0.0654	0.0500	0.3418
IR	−0.0258	0.0518	0.2955	−0.0728	0.0501	0.3312
VOLA	−0.0262	0.0520	0.3036	−0.0741	0.0502	0.3555
TED	−0.0168	0.0517	0.3160	−0.0701	0.0502	0.3514
DEF	−0.0222	0.0516	0.3111	−0.0729	0.0501	0.3547

Note:
The table displays average values of the information coefficient (IC), the root mean squared error (RMSE), and the number of negative return forecasts (in percentages) 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).

market risk premium should actually be positive [(BOUDOUKH et al. (1993), and CAMPBELL and THOMPSON (2005)]. A small number of months where a statistical model selection criterion retains the i.i.d. no predictability model may indicate the existence of return predictability. While FIC and PIC never decide against predictability, the adjusted R^2 criterion, AIC and especially BIC retain in up to 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 $\boldsymbol{A}'\boldsymbol{P}/(2^M/2)$, where the vector $\boldsymbol{A}$ is defined as previously and the $(2^M,1)$ vector $\boldsymbol{P}$ contains the respective statistic. The most interesting result of Table 5 is 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.

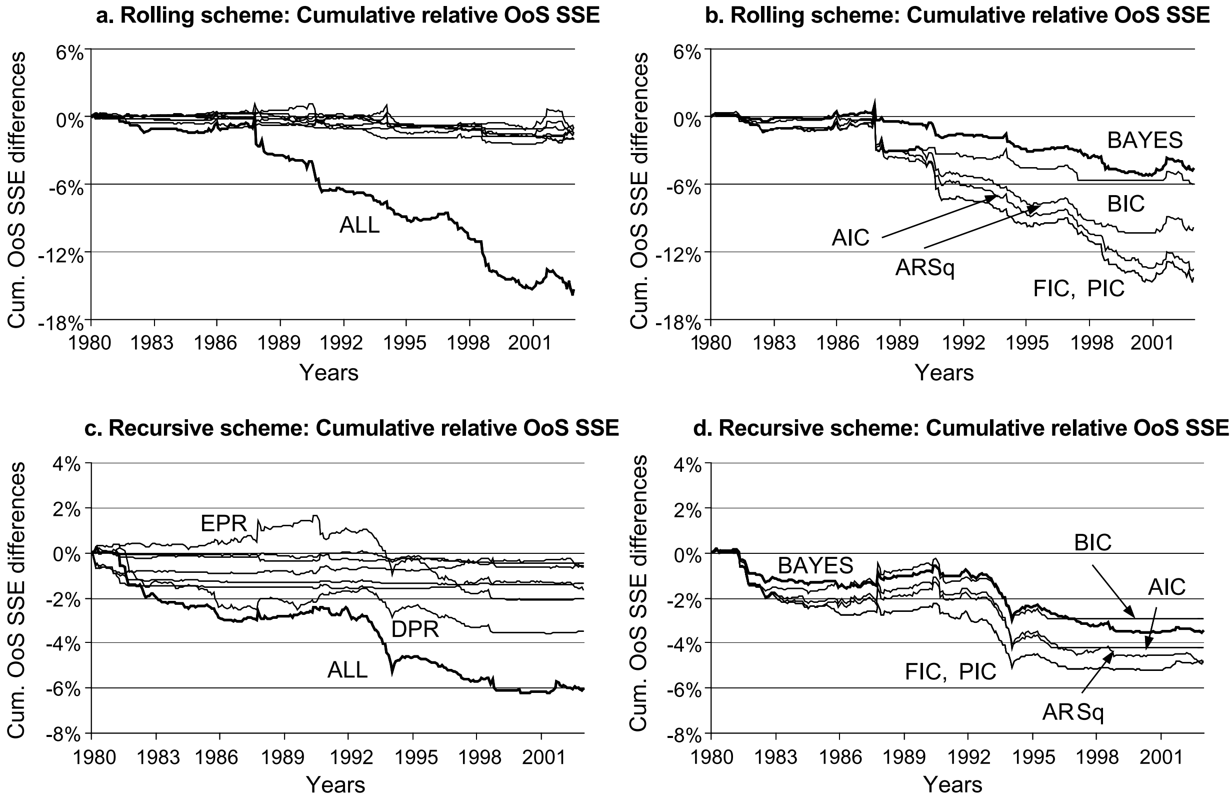
GOYAL and WELCH (2003a,b) suggest another way to look at the results. They suggest a simple, recursive residuals (out-of-sample) graphical approach to evaluating the forecasting ability of the predictive regression specifications. Their simple graphical diagnostic plots the cumulative sum-squared forecast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

$$Net - SSE(\tau) = \sum\nolimits_{t=1}^{\tau} SE_t^{iid} - SE_t, \tag{23}$$

where SE_t is the squared out-of-sample forecast error in month t. Thus, Figure 3 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 3 confirms the results of Tables 3 and 4, and shows that all regression specifications, including the statistical model selection criteria and the Bayesian model averaging approach, practically never outperform the prevailing mean model.

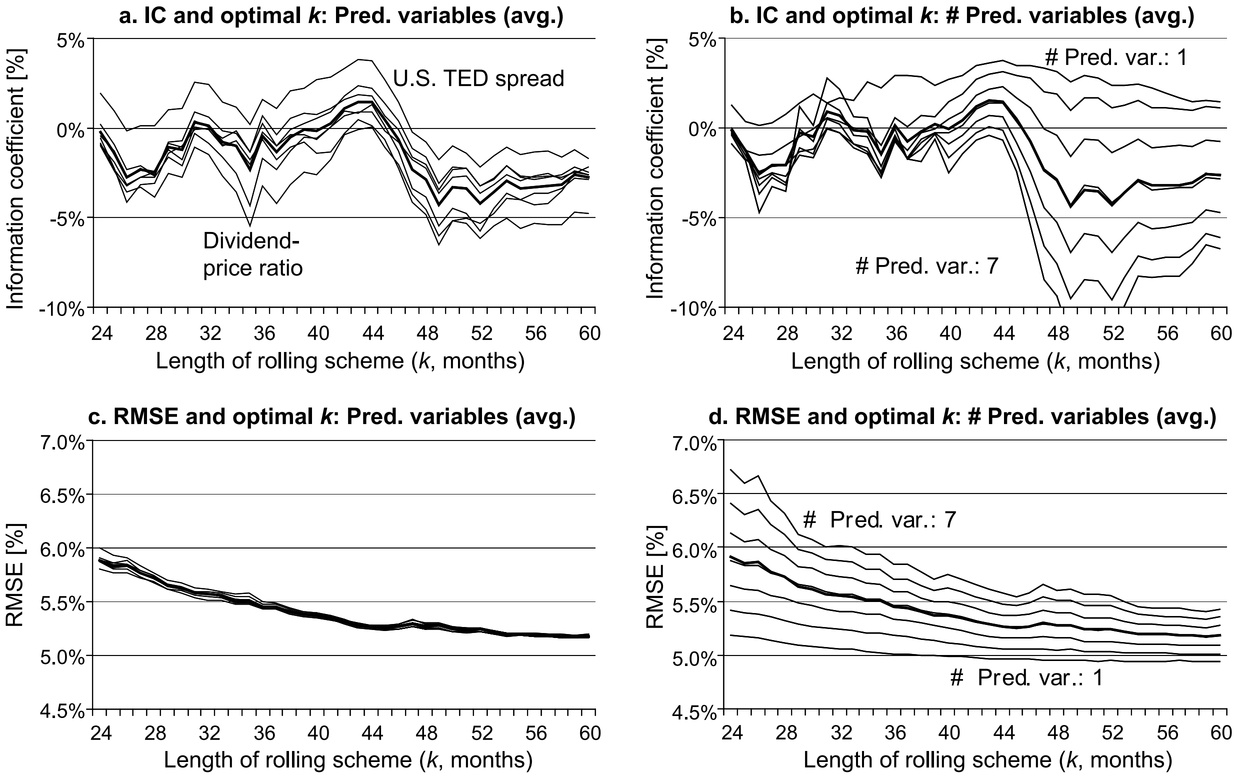
Figure 3: Cumulative Relative Out-of-Sample, Sum-Squared Forecast Error Performance



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 (Adj. R^2, 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).

We also explored the robustness of our conclusions to different values of k, the length of the rolling scheme. Figure 4 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 may promise 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 corresponding market timing strategies, we may just conclude that investors should avoid the dividend–price ratio as predictive variable and should retain only a small number of predictive variables.

Figure 4: Optimal Length of the Rolling Scheme



The graphs plot the information coefficients and the root mean squared error (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. The bold lines represent overall averages.

In sum, thus, consistent with the results of BOSSAERTS and HILLION (1999), NEELY and WELLER (1999), GOYAL and WELCH (2003a,b), and SCHWERT (2003), who conclude that the out-of-sample predictive ability of the dividend price ratio and the other predictive variables is abysmal, and, in the words of Schwert, disastrous, our results in Tables 3, 4, and 5 do not show 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 deteriorates 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.

4. Conclusion

We implement statistical model selection criteria and AVRAMOV's (2002) Bayesian model averaging approach to analyze the sample evidence of stock market predictability. Based on Swiss stock market data, we obtain the following general results. First, the posterior probabilities of the individual forecasting models and the cumulative posterior probabilities are not constant through time. Second, the estimates of the posterior probabilities are not robust to whether the predictive variables are stochastically detrended or not. Third, the contributions of parameter uncertainty, model uncertainty, and the uncertainty attributed to forecast errors are dependent on the time period under consideration and the initial values of the predictive variables. Thus, model uncertainty is not more important than parameter uncertainty. Fourth, from an investment management perspective, 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.[8]

Moreover, the inclusion of more than one predictive variable rather deteriorates the out-of-sample performance of the forecasting models. Finally, 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 and the use of quarterly and half-yearly data instead of the monthly data.[9]

The poor external validity of all the predictive regression specifications may indicate model non-stationarity: the parameters of the best prediction model change over time. It is still an open question why this might be. One potential explanation is that the correct regression specification is actually nonlinear, while statistical model selection criteria and the Bayesian model averaging approach chose exclusively among linear models.[10]

Still, 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 these possible regression specifications. So the poor out-of-sample performance of the predictive regression models really raise questions about the predictive variables' role in these models. Consequently, attempting to fit more complicated models such as multiple-beta, conditional APT-type models, might seem a futile exercise, especially when parameter and model risk are taken into account. Overall, it thus seems very questionable whether a (business cycle-related) time-varying equity premium can be predicted using simple regression techniques and whether the respective results should be considered as a serious input for corresponding short-term market timing strategies.

Acknowledgements

The valuable comments of Manuel Ammann, Markus Schmid, and Heinz Zimmermann are gratefully acknowledged. The author appreciates the helpful suggestions from an anonymous referee.

ENDNOTES

[1] See, e.g., AMIHUD and HURVICH (2004), FERSON et al. (2003), TOROUS et al. (2004), and VALKANOV (2003). REY (2003a,b) provides an overview.

[2] 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. We follow them and take 50 observations per parameter, i.e., $T_{j,0} = T_0(n+1)$ with $T_0 = 50$. Our conclusions are robust to different specifications of the hypothetical prior size ($T_0 = 25$ or $T_0 = 100$).

[3] A more detailed description of this algorithm is given in BARBERIS (2000). An alternative algorithm, which is potentially more efficient when there is a large number of assets and predictive variables, is proposed in AVRAMOV (2002).

[4] 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. Realized stock market volatility is calculated as suggested by GOYAL and SANTA-CLARA (2003), using within-month daily return data for each month. The U.S. TED spread is 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.

[5] See BRANDT (2004) for a critical comment about evenly distributed prior probabilities across all models.

[6] The consideration of a number of initial values (and parameter estimates) is more important than varying the strength of the informative prior, T_0. As in AVRAMOV (2002), the results (not reported) indicate that the variance decomposition is not highly sensitive to different values of T_0.

[7] We do not adjust the *t*-statistics for error in the estimation of the parameters of the prediction model [see, e.g., BOSSAERTS and HILLION (1999)].

[8] BOUDOUKH et al. (2004) show that the dividend price ratio process changed remarkably during the 1980's and 1990's, but that the total payout ratio (dividends plus repurchases over price) changed very little. Hence, they conclude that the decline in the predictive power of the dividend price ratio in recent U.S. data is vastly overstated. The lack of data makes the respective analysis impossible for the Swiss stock market, however.

[9] The corresponding results are available from the author upon request.

[10] A recent contribution by CAMPBELL and THOMPSON (2005) shows that the imposition of sensible restrictions with respect to the signs of the coefficients and the return forecasts improve the out-of-sample predictive power. A discussion of those arguments as well as an empirical verification for the Swiss stock market are postponed for future research.

REFERENCES

AKAIKE, H. (1974): "A New Look at the Statistical Model Identification", IEEE Transactions on Automatic Control, AC-19, pp. 716–723.

AKGIRAY, V. (1989): "Conditional Heteroskedasticity in Time Series of Stock Returns: Evidence and Forecasts", Journal of Business 62, pp. 55–80.

AMIHUD, Y. and C. M. HURVICH (2004): "Predictive Regressions: A Reduced-Bias Estimation Method", Journal of Financial and Quantitative Analysis 39, pp. 813–841.

AVRAMOV, D. (2002): "Stock Return Predictability and Model Uncertainty", Journal of Financial Economics 64, pp. 423–458.

BAKER, M. and J. WURGLER (2000): "The Equity Share in New Issues and Aggregate Stock Returns", Journal of Finance 55, pp. 2219–2257.

BARBERIS, N. (2000): "Investing for the Long Run When Returns Are Predictable", Journal of Finance 55, pp. 225–264.

BEKAERT, G. (2001): "Editor's Foreword to the Special Issue: 'On the Predictability of Asset Returns'", Journal of Empirical Finance 8, pp. 451–457.

BOSSAERTS, P. and P. HILLION (1999): "Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?", Review of Financial Studies 12, pp. 405–428.

BOUDOUKH, J., M. RICHARDSON and T. SMITH (1993): "Is the Ex-Ante Risk Premium Always Positive?", Journal of Financial Economics 34, pp. 387–408.

BOUDOUKH, J., R. MICHAELY, M. RICHARDSON and M. ROBERTS (2004): "On the Importance of Measuring Payout Yield: Implications for Empirical Asset Pricing", Working Paper 10651 (NBER).

BRANDT, M. W. (2004): "Portfolio Choice Problems", to appear in Y. Ait-Sahalia and L. P. Hansen, eds., Handbook of Financial Econometrics, Elsevier Science.

BRENNAN, M. J., E. S. SCHWARTZ, and R. LAGNADO (1997): "Strategic Asset Allocation", Journal of Economic Dynamics and Control 21, pp. 1377–1403.

CAMPBELL, J. Y. (1991): "A Variance Decomposition for Stock Returns", Economic Journal 101, pp. 157–179.

CAMPBELL, J.Y. and R. J. SHILLER (1988a): "Stock Prices, Earnings, and Expected Dividends", Journal of Finance 43, pp. 661–676.

CAMPBELL, J.Y. and R. J. SHILLER (1988b): "The Dividend–Price Ratio and Expectations of Future Dividends and Discount Factors", Review of Financial Studies 1, pp. 195–228.

CAMPBELL, J. Y. and S. B. THOMPSON (2005): "Predicting the Equity Premium Out of Sample: Can Anything Beta the Historical Average?", Working Paper (Harvard University), April 2005.

CAMPBELL, J. Y. and L. M. VICEIRA (1999): "Consumption and Portfolio Decisions When Expected Returns Are Time-Varying", Quarterly Journal of Economics 144, pp. 433–495.

CAMPBELL, J. Y. and L. M. VICEIRA (2002): Strategic Asset Allocation: Portfolio Choice for Long-Term Investors: Oxford University Press, Oxford.

CHEN, N. (1991): "Financial Investment Opportunities and the Macroeconomy", Journal of Finance 46, pp. 529–554.

COCHRANE, J. H. (1999): "New Facts in Finance", Economic Perspectives 23, pp. 36–58.

DIEBOLD, F. X. and R. S. MARIANO (1995): "Comparing Predictive Accuracy", Journal of Business and Economics Statistics 13, pp. 253–263.

ENGSTROM, E. (2003): "The Conditional Relationship Between the Equity Risk Premium and the Dividend–Price Ratio", Working Paper (Columbia University), November 2002.

FAIR, R. C. and R. J. SHILLER (1990): "Comparing Information in Forecasts from Econometric Models", American Economic Review 80, pp. 375–389.

FAMA, E. F. (1991): "Efficient Capital Markets: II", Journal of Finance 46, pp. 1575–1617.

FAMA, E. F. and K. R. FRENCH (1988): "Dividend Yields and Expected Stock Returns", Journal of Financial Economics 22, pp. 3–25.

FAMA, E. F. and K. R. FRENCH (1989): "Business Conditions and Expected Returns on Stocks and Bonds", Journal of Financial Economics 25, pp. 23–49.

FERSON, W. E. and C. R. HARVEY (1991): “The Variation of Economic Risk Premiums”, Journal of Political Economy 99, pp. 385–415.

FERSON, W. E. and C. R. HARVEY (1993): “The Risk and Predictability of International Equity Returns”, Review of Financial Studies 6, pp. 527–566.

FERSON, W. E., S. SARKISSIAN, and T. SIMIN (2003): “Spurious Regression in Financial Economics?”, Journal of Finance 58, pp. 1393–1413.

FERSON, W. E., S. SARKISSIAN, and T. SIMIN (2004): “Is Stock Return Predictability Spurious?”, Journal of Investment Management 3, pp. 1–10.

FOSTER, F. D., T. SMITH, and R. E. WHALEY (1997): “Assessing Goodness-of-Fit of Asset Pricing Models: The Distribution of the Maximal R^2”, Journal of Finance 52, pp. 591–607.

FRENCH, K. R., G. W. SCHWERT and R. F. STAMBAUGH (1987): “Expected Stock Returns and Volatility”, Journal of Financial Economics 19, pp. 3–29.

GOYAL, A. and P. SANTA-CLARA (2003): “Idiosyncratic Risk Matters!”, Journal of Finance 58, pp. 975–1007.

GOYAL, A. and I. WELCH (2003a): “Predicting the Equity Premium with Dividend Ratios”, Management Science 49, pp. 639–654.

GOYAL, A. and I. WELCH (2003b): “A Note on ‘Predicting Returns with Financial Ratios’”, Working Paper (Yale International Center for Finance), December 2003.

GRINOLD, R. C. and R. N. KAHN (2000): Active Portfolio Management: McGraw-Hill

HODRICK, R. J. (1992): “Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement”, Review of Financial Studies 5, pp. 357–386.

KANDEL, S. and R. F. STAMBAUGH (1996): “On the Predictability of Stock Returns: An Asset-Allocation Perspective”, Journal of Finance 51, pp. 385–424.

KEIM, D. B. and R. F. STAMBAUGH (1986): “Predicting Returns in the Stock and Bond Markets”, Journal of Financial Economics 17, pp. 357–390.

KOTHARI, S. P. and J. SHANKEN (1997): “Book-to-Market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis”, Journal of Financial Economics 44, pp. 169–203.

LAMONT, O. (1998): “Earnings and Expected Returns”, Journal of Finance 53, pp. 1563–1587.

LEE, C., J. MYERS, and B. SWAMINATHAN (1999): “What Is The Intrinsic Value of the Dow?”, Journal of Finance 54, pp. 1693–1742.

LETTAU, M. and S. LUDVIGSON (2001): “Consumption, Aggregate Wealth, and Expected Stock Returns’, Journal of Finance 56, pp. 815–849.

LO, A. W. and A. C. MACKINLAY (1990): “Data-Snooping Biases in Tests of Financial Asset Pricing Models”, Review of Financial Studies 3, pp. 431–467.

MINCER, J. and V. ZARNOWITZ (1969): “The Evaluation of Economic Forecasts”, in J. Mincer, (ed.), Economic Forecasts and Expectations, NBER.

NEELY, C. J. and P. WELLER (1999): “Predictability in International Asset Returns: A Reexamination”, Working Paper (Federal Reserve Bank of St. Louis), February 1999.

PHILLIPS, P. C. B. and W. PLOBERGER (1996): “Posterior Odds Testing for a Unit Root with Data-Based Model Selection”, Econometrica 64, pp. 381–412.

PONTIFF, J. and L. SCHALL (1998): “Book-to-Market as a Predictor of Market Returns”, Journal of Financial Economics 49, pp. 141–160.

REY, D. (2003a): “Stock Market Predictability: Is it There? A Critical Review”, Working Paper (Department of Finance, WWZ University of Basel), December 2003.

REY, D. (2003b): “Current Research Topics – Stock Market Predictability: Is it There?”, Financial Markets and Portfolio Management 17, pp. 379–387.

SCHWARZ, G. (1978): “Estimating the Dimension of a Model”, Annals of Statistics 6, pp. 416–464.

SCHWERT, G. W. (2003): “Anomalies and Market Efficiency”, in: Constantinides, G. M., M. Harris, and R. Stulz, (eds.), Handbook of the Economics of Finance, Elsevier Science, North-Holland, Amsterdam.

TOROUS, W., R. VALKANOV, and S. YAN (2004): "On Predicting Stock Returns with Nearly Integrated Explanatory Variables", Journal of Business 77, pp. 937–966.

VALKANOV, R. (2003): "Long-horizon Regressions: Theoretical Results and Applications", Journal of Financial Economics 68, pp. 201–232.

WEI, C. (1992): "On Predictive Least Squares Principles", Annals of Statistics 20, pp. 1–42.

XIA, Y. (2001): "Learning About Predictability: The Effects of Parameter Uncertainty on Dynamic Asset Allocation", Journal of Finance 56, pp. 205–246.

David Rey studied economics with an emphasis in finance at the Universities of Bern, Lancaster (UK), and St. Gallen (HSG), where he also received his doctorate in finance. He works as a research assistant at the Wirtschaftswissenschaftliches Zentrum (WWZ) of the University of Basel. His main research interests are in the areas of asset pricing and asset management.