



Helsinki
Center
of
Economic
Research

Discussion Papers

Forecasting the Direction of the U.S. Stock Market with Dynamic Binary Probit Models

Henri Nyberg
University of Helsinki and HECER

Discussion Paper No. 227
June 2008

ISSN 1795-0562

Forecasting the Direction of the U.S. Stock Market with Dynamic Binary Probit Models*

Abstract

Several empirical studies have documented that the signs of excess stock returns are, to some extent, predictable. In this paper, we consider the predictive ability of the binary dependent dynamic probit model in predicting the direction of monthly excess stock returns. The recession forecast obtained from the model for a binary recession indicator appears to be the most useful predictive variable and once it is employed, the sign of the excess return is predictable in-sample. A new dynamic "error correction" probit model proposed in the paper yields the best out-of-sample forecasts with the average trading strategy returns higher than in the buy-and-hold strategy or in the ARMAX models.

JEL Classification: C22, C25, E44, G11

Keywords: Probit model, directional predictability, stock return, recession forecast

Henri Nyberg

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

e-mail: henri.nyberg@helsinki.fi

* The author would like to thank Heikki Kauppi, Markku Lanne and Pentti Saikkonen for constructive comments. The author is responsible for remaining errors. Financial support from the Research Foundation of the University of Helsinki, Academy of Finland, the Okobank Group Research Foundation and the Finnish Foundation for Advancement of Securities Markets is gratefully acknowledged.

1 Introduction

In the financial econometric literature there is considerable evidence that excess stock market returns are, to some extent, predictable. The main objective has been to predict the overall level, the conditional mean, of excess stock returns. It is emphasized that even though the predictability is statistically small, it can be economically meaningful.

However, many studies have documented that only the direction of excess stock returns or other asset returns are predictable (see, among others, Breen, Glosten, and Jagannathan 1988, Hong and Chung 2003, and Cristoffersen and Diebold 2006). A possible explanation for this is that the noise in the observed returns is too high for the accurate forecasting of the overall return. Leitch and Tanner (1991) find that the direction of the change is the best criterion for predictability because traditional statistical summary statistics may not be closely related to the profits that investors are seeking in the financial market. Directional predictability is also important for market timing, which is crucial for asset allocation decisions between stock and risk-free interest rate investments.

The previous findings of directional predictability are mainly based on time series models for the excess stock return. For instance, Cristoffersen and Diebold (2006) and Cristoffersen et al. (2006) have considered the theoretical connection between asset return volatility and asset return sign forecastability and verified, that volatility and higher-order conditional moments of returns have statistically significant explanatory power in sign prediction. Even though there is not much previous research, binary dependent time series models provide an another way to forecast the direction of excess stock returns. Various classification-based qualitative models such as traditional static logit and probit models were considered by Leung, Daouk, and Chen (2000) whereas Hong and Chung (2003), Rydberg and Shephard (2003), and Anatolyev and Gospodinov (2007) used the so-called autologistic model to predict the return direction. In the last two papers the return is decomposed into a sign component and absolute value component which are modeled separately before the joint forecast is constructed.

In this study, various commonly used financial variables are considered as explana-

tory variables for forecasting the signs of the one-month U.S. excess stock returns in probit models. The paper also introduces a model in which the recession forecast constructed for a binary recession indicator is used as an explanatory variable in the predictive model. To the best of our knowledge, this kind of approach has not been applied earlier to forecast the stock return sign. As a motivation for this kind of model, for example, Fama and French (1989) and Chen (1991) propose that business conditions are important determinants of expected stock returns and therefore the recession forecast may be useful predictive variable in our model. Further, Chauvet and Potter (2000) have stressed that the stock market "cycle" leads the business cycle. This argument is based on the fact that the expectations about changes in the future economic activity could have important predictive power to predict excess stock returns. If there are expectations about a coming recession, excess stock returns are low and after a recession period stock returns should be positive.

In this paper, new dynamic probit models suggested by Kauppi and Saikkonen (2007) are employed and further extended. Since there is not much earlier evidence about suitable explanatory variables in sign prediction with probit models, various explanatory variables are first experimented and their in-sample forecasting performance is evaluated in the in-sample forecasting period. After that, the out-of-sample directional predictability for excess stock return sign is considered. It is not evident, however, how much the in-sample evidence should be emphasized in assessing overall return predictability because it does not guarantee out-of-sample predictability, as emphasized in many previous studies (see discussion, for example, in Goyal and Welch, 2004, and Campbell and Thompson, 2007).

The results show that the probit models have statistically significant in-sample predictive power for the signs of excess stock returns. A proposed new "error correction" model outperforms the other considered probit and "continuous" ARMAX models out-of-sample. The received excess investment returns over the buy-and-hold trading strategy are statistically and economically significant. Comparisons between different probit models indicate that the forecasting framework based on the constructed recession forecasts yields more accurate sign predictions than the models where only financial explanatory variables are employed.

The paper proceeds as follows. The employed forecasting model with recession forecasts, suggested dynamic probit models and, in particular, the new error correction model is presented in Section 2. In Section 3, the goodness-of-fit evaluation of sign forecasts and statistical tests for the sign predictability are introduced. The empirical evidence on the directional predictability of the U.S. excess stock returns is reported in Section 4. Section 5 concludes.

2 Forecasting Model

2.1 Dynamic Probit Models in Directional Forecasting

Let r_t be the excess stock return over the risk-free interest rate. In many studies, the directional predictability of excess stock returns is examined by using models for continuous dependent variables. For example, Cristoffersen et al. (2006) proposed a method of forecasting the direction of excess stock return, where they first model the conditional mean μ_t and the conditional variance σ_t^2 . Assuming that the data generating process of r_t is

$$r_t = \mu_t + \sigma_t \varepsilon_t,$$

where $\varepsilon_t \sim \text{IID}(0, 1)$, the conditional probability of a positive return is

$$\begin{aligned} P_{t-1}(r_t > 0) &= 1 - P_{t-1}(r_t \leq 0) \\ &= 1 - P_{t-1}\left(\varepsilon_t \leq \frac{-\mu_t}{\sigma_t}\right) \\ &= 1 - F_\varepsilon\left(\frac{-\mu_t}{\sigma_t}\right), \end{aligned}$$

where $F_\varepsilon(\cdot)$ is the cumulative distribution function of the error term ε_t . This shows that, if the conditional mean or the conditional variance is not constant, the sign of the excess stock return will be predictable.

In this paper, the main interest is to study the directional predictability using probit models where the dependent variable is the binary sign return indicator

$$I_t = \begin{cases} 1, & \text{if } r_t > 0, \\ 0, & \text{if } r_t \leq 0, \end{cases} \quad (1)$$

which takes the value one when the excess stock return is positive and zero otherwise. Thus, it is thought that I_t is a binary-valued stochastic process and, conditional the information set Ω_{t-1} which includes the predictive variables and lagged values of the stock indicator (1), it has a Bernoulli distribution with probability p_t^I , that is

$$I_t|\Omega_{t-1} \sim B(p_t^I).$$

The conditional probability p_t^I is modeled by specifying a model for the variable π_t^I , which is related to conditional probability via the equality $p_t^I = G(\pi_t^I)$, where the link function $G(\cdot)$ is typically the cumulative distribution function of the normal distribution or the logistic distribution.¹ In this study, the former one is used which results in the probit model. If E_{t-1} and P_{t-1} are denoted the conditional expectation and the conditional probability given the information set Ω_{t-1} , respectively, the conditional probability of positive excess stock return ($I_t = 1$) is

$$E_{t-1}(I_t) = P_{t-1}(I_t) = P_{t-1}(r_t > 0) = \Phi(\pi_t^I) = p_t^I, \quad (2)$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. The model built for π_t^I is supposed to depend on variables in the information set Ω_{t-1} that are considered useful in forecasting.

Previous literature indicates that there is not much autocorrelation between the two successive values of the stock indicator or the excess stock return (see also Figure 1). Thus, the benchmark forecasting model is the static model,

$$\pi_t^I = \omega + \mathbf{x}'_{t-1}\boldsymbol{\beta}, \quad (3)$$

where the employed explanatory variables, which are described in the Appendix, are collected in the vector \mathbf{x}_{t-1} . Because of the expected lack of correlation between I_{t-1} and I_t , this static model, without any dynamic structure, might be adequate. In order to investigate this, the value of the lagged return indicator I_{t-1} can be included in the model. This would yield the dynamic probit model,

$$\pi_t^I = \omega + \delta_1 I_{t-1} + \mathbf{x}'_{t-1}\boldsymbol{\beta}. \quad (4)$$

¹ The superscript "I" in π_t^I refers to excess return sign forecasting.

If the coefficient δ_1 is statistically significant, then the lagged direction of the stock return is a useful predictor of the future direction of excess stock returns.

In the last few years, new binary time series models have been introduced. In this paper, we concentrate on the model variants suggested by Kauppi and Saikkonen (2007). They add the lagged value of π_{t-1}^I , referred to as autoregressive structure, to the model equation. Thus, the static model (3) and the dynamic model (4) are extended to the autoregressive model²

$$\pi_t^I = \omega + \alpha_1 \pi_{t-1}^I + \mathbf{x}'_{t-1} \boldsymbol{\beta} \quad (5)$$

and to the dynamic autoregressive model

$$\pi_t^I = \omega + \alpha_1 \pi_{t-1}^I + \delta_1 I_{t-1} + \mathbf{x}'_{t-1} \boldsymbol{\beta}, \quad (6)$$

respectively. By recursive substitution, and assuming $|\alpha_1| < 1$, the latter model can be written as

$$\pi_t^I = \sum_{i=1}^{\infty} \alpha_1^{i-1} c + \delta_1 \sum_{i=1}^{\infty} \alpha_1^{i-1} I_{t-i} + \sum_{i=1}^{\infty} \alpha_1^{i-1} \mathbf{x}'_{t-i} \boldsymbol{\beta}. \quad (7)$$

Therefore, if several lagged values of the stock indicator (1) or explanatory variables \mathbf{x}_t are useful in forecasting, autoregressive specifications could be useful parsimonious models.

Parameters of the probit models (4)–(6), as well as those of the case of a new model presented in the next section, can be estimated by the method of maximum likelihood as described in de Jong and Woutersen (2007) and Kauppi and Saikkonen (2007).

2.2 An Error Correction Model

Based on, for example, the principles of efficient market theory, the lagged values of the stock indicator (1) should not have predictive power to predict future market directions. This indicates that the estimated coefficient of the lagged return indicator δ_1 may be zero or close to zero. Therefore, in the dynamic autoregressive model (6), if $\delta_1 = 0$ and there are no explanatory variables in the model, that is $\boldsymbol{\beta} = \mathbf{0}$, then the

² In this paper, the same model attributes as Kauppi and Saikkonen (2007) are used.

autoregressive parameter α_1 is not identified as seen from equation (7) by imposing the above mentioned restrictions.³ Even if the coefficient δ_1 is just close to zero, there is a potential identification problem that can affect parameter estimation and have implications on the forecasting accuracy of excess return sign predictions.

Imposing the restriction $\delta_1 = 1 - \alpha_1$ in the unrestricted dynamic autoregressive model (6) and assuming $|\alpha_1| < 1$, a new "restricted" dynamic autoregressive model can be formulated as

$$\pi_t^I = \omega + \alpha_1 \pi_{t-1}^I + I_{t-1}(1 - \alpha_1) + \mathbf{x}'_{t-1} \boldsymbol{\beta}. \quad (8)$$

Because of the assumption $|\alpha_1| < 1$, the coefficient for the lagged return indicator I_{t-1} , $1 - \alpha_1$, is always positive. In fact, it is expected that in our application α_1 will be positive and quite high.

For simplicity, in this paper we refer to the model (8) as an "error correction" model (ecm). The reason is that adding $-\pi_{t-1}^I$ to both sides of equation (8) we obtain the error correction form

$$\Delta \pi_t^I = \omega + (1 - \alpha_1)(I_{t-1} - \pi_{t-1}^I) + \mathbf{x}'_{t-1} \boldsymbol{\beta}, \quad (9)$$

where $\Delta \pi_t^I = \pi_t^I - \pi_{t-1}^I$. Thus, the difference between I_{t-1} and π_{t-1}^I measures the long-run relationship between the value of the stock return indicator and the transformed probability $\pi_{t-1}^I = G^{-1}(p_{t-1}^I)$ in probit model. Rewriting model (9) as

$$\pi_t^I = \omega + \pi_{t-1}^I + (1 - \alpha_1)(I_{t-1} - \pi_{t-1}^I) + \mathbf{x}'_{t-1} \boldsymbol{\beta},$$

it can be seen that, for α_1 close to one, the model can be expected to exhibit "near unit root" behavior, implying rather strong persistence in the model equation π_t^I . This means that the conditional probability of positive excess stock return does not change much between successive time periods.

As seen from equation (8), the parameter $1 - \alpha_1$ is always positive and it can be interpreted as the proportion of the disequilibrium between I_{t-1} and π_{t-1}^I in period $t-1$. A positive value of the error correction term $(I_{t-1} - \pi_{t-1}^I)$ increases the probability

³ Note that when explanatory variables \mathbf{x}_t are included in the model, then there is no such identification problem, even though $\delta_1 = 0$.

of positive excess stock return in the following period and, of course, vice versa if the error correction term is negative. Later on in this paper, we will see that the conditional probability of positive excess stock return p_t^I is typically close to 0.50 in most models which means that π_t is close to zero.

It is worth noting that the error correction model (8) is somewhat the same as the autoregressive conditional multinomial model (ACM) suggested by Russell and Engle (2005). In their model, the term $(I_{t-1} - \pi_{t-1}^I)$ is replaced by $(I_{t-1} - \Phi(\pi_{t-1}))$. The model (8) without the term $\mathbf{x}'_{t-1}\boldsymbol{\beta}$ is also similar to the IGARCH model, suggested by Engle and Bollerslev (1986), for conditional heteroskedasticity in models for continuous variables.

2.3 Recession Forecast as an Explanatory Variable

A novel idea of this paper is to study whether recession forecasts have explanatory power to forecast the direction of excess stock returns. In empirical finance literature, it is shown that as forward-looking variables, lagged stock returns should provide information about the future evolution of economic activity and potential recession periods (e.g. Pesaran and Timmermann, 1995, Estrella and Mishkin, 1998 and Nyberg, 2008). Therefore, if the expectations of future economic activity are correct, the movements of the stock market should lead movements in economic activity (e.g. Fama 1990). Theoretically, this relation can be justified by present value or discounted-cash-flow models, where the price of a stock is equal to expected future dividends that are related to economic activity.

Our main goal is to forecast recession periods and use the potential explanatory power of the obtained recession forecasts to make better forecasts for the sign of the excess stock returns. This is done by using the binary recession indicator

$$y_t = \begin{cases} 1, & \text{if economy is in a recessionary state at month } t, \\ 0, & \text{if economy is in an expansionary state at month } t. \end{cases} \quad (10)$$

In this study, recession dates defined by the NBER are used. As Chauvet and Potter (2000) argue, one feature, but also a potential problem, with the NBER recession dates is that they do not reflect short-lived contraction periods in the economy, which could have notable explanatory power for predicting excess stock returns. Further,

Chauvet and Potter (2000) construct the transition probabilities of "bear" and "bull" stock market using Markov chain methods.⁴ They find that bear markets generally start a couple of months before an economic slowdown or recession period and end some months before a recession period ends. Thus, it seems evident that movements in the stock market should lead the business cycle. Evidence of this kind can be seen in Figure 2. The U.S excess stock returns are often negative before a recession period begins. On the other hand, it seems that the returns are positive in the last few recession months, indicating expectations about recovery in economic activity.

According to this idea, for example, in the general dynamic autoregressive model (6), the estimated recession forecast p_{t+5}^y may be included in the vector $\mathbf{x}_{t-1} = (p_{t+5}^y \quad \tilde{\mathbf{x}}_{t-1})$, where $\tilde{\mathbf{x}}_{t-1}$ contains other financial explanatory variables. Therefore, a predictive probit model contains the fitted values of the binary explanatory recession indicator (10) (cf. Maddala 1983, 122–123). Parameter estimation and forecasting is carried out with a two-step procedure, where the recession and the stock return sign prediction models are estimated separately. It is worth noting, however, that in this kind of model, the asymptotic distribution of the maximum likelihood estimate could be non-standard because the recession probability forecast included in the model is based on the estimated model (cf. Pagan, 1984).

In this study, forecast horizon in recession forecasting is assumed to be six months. A six-month recession forecast for the value of the recession indicator (10) at time $t + 5$, based on the information set Ω_{t-1} at time $t - 1$, is the conditional probability

$$E_{t-1}(y_{t+5}) = P_{t-1}(y_{t+5} = 1) = \Phi(\pi_{t+5}^y) = p_{t+5}^y.$$

In recession forecasting, an autoregressive probit model

$$\pi_{t+5}^y = c + \phi\pi_{t+4}^y + \mathbf{z}'_{t-1}\mathbf{b} \tag{11}$$

is employed where, according to the findings in recession forecasting literature, domestic and foreign term spread and lagged stock return are used as predictors. The values of these variables are included in the vector \mathbf{z}_{t-1} . The usefulness of the domestic term spread, defined as the spread between the long-term and the short-term

⁴ "Bull" ("Bear") market corresponds to a period when stock market prices are increasing (decreasing).

interest rates, to predict recession periods is demonstrated in many studies (see, among others, Estrella and Mishkin, 1998 and Estrella, 2005). Using dynamic probit models, Nyberg (2008) also suggests that the foreign term spread (the term spread of Germany) and stock market returns can be used to forecast coming recession periods.⁵ In that study, the obtained recession forecasts were quite accurate at least six months ahead. Therefore, the results are reported only for the six-month-ahead forecast horizon.

An advantage in the recession prediction model (11) is that it does not contain a lagged value of the recession indicator (10). It is important to take into account that it takes several months before, for example NBER, can be sure what the state of economy really is. Hence, the values of the recession indicator are known with a considerable delay. Nevertheless, in model (11), no multiperiod iterative forecasts (see Kauppi and Saikkonen, 2007) for the recession indicator are made because all predictive power comes from the employed explanatory variables z_t . Thus, it is not needed to specify the assumed "publication" lag in the known values of y_t exactly.

3 Evaluation of Forecasts

3.1 Statistical and Economic Goodness-of-Fit Measures

Both in-sample and out-of-sample performance of predictive models is evaluated with frequently used goodness-of-fit measures. One is Estrella's (1998) pseudo- R^2 measure

$$psR^2 = 1 - (\hat{l}_u / \hat{l}_c)^{-(2/T)\hat{l}_c}, \quad (12)$$

where \hat{l}_u is the maximum value of the estimated unconstrained log-likelihood function and \hat{l}_c is its constrained counterpart in a model which only contains a constant term. This measure takes on values between 0 and 1 and it can be interpreted in the same way as the coefficient of determination in linear models. The value of the maximized log-likelihood function also enables to compare model performance using model selection criteria such as the Schwarz information criterion BIC (Schwarz, 1978).

⁵ Further information of explanatory variables is in the Appendix.

The binary nature of the dependent variable leads to ask, what is the percentage of correct "matches" of the realized values and the forecasts of the stock indicator. This ratio is denoted by CR . By the hypothesis of non-predictability of stock return signs, the estimated value of the CR should be close to 0.50, which means that the employed model is unable to forecast the future market directions correctly. Thus, it is desirable to specify a threshold value that translates the probability forecasts into forecasting signals. The most commonly used and natural threshold choice is 0.50, which is also used in this paper.

For financial analysts and investors, the most important model evaluation criterion is the return on their investment. There are many different kinds of trading strategies that can be applied. Here a trading simulation similar to that in Leung et al. (2000) is used. At the beginning of each month, the investor makes an asset allocation decision. She can shift her assets either into stocks or in the risk-free Treasury Bills and the money that has been invested either of these alternatives remains there until the next decision date. In this trading strategy, the mentioned 50 percent threshold value is used. Then, the portfolio consist of interest rate investment in Treasury Bills ($I_t^f = 0$), if $p_t^f \leq 0.50$, and stocks ($I_t^f = 1$), if $p_t^f > 0.50$. Here the superscript f refers to forecast.

In this trading simulation, also transaction costs are taken into account. Following Granger and Pesaran (2000), the marginal cost of transaction for asset allocation changes between stocks and interest rates will be denoted by ζ_s and ζ_b , respectively. This means that, every time the asset allocation changes, the amount of transaction cost is subtracted from the final investment return. In this paper, the "low cost scenario" suggested by Pesaran and Timmermann (1995), where $\zeta_s = 0.005$ and $\zeta_b = 0.001$, is applied. For example, when the risk-free interest rate investments are switched to stocks, 0.50 % of the whole amount of the portfolio value is lost. As Granger and Pesaran (2000) have shown, it is possible to form non-constant "payoff" probability ratios of switches between stocks and interest rates as an alternative for this 50 % threshold. However, in this study, these payoff ratios are not very useful because, according to these threshold ratios, the asset allocation decision is to stick to stocks almost all the time. Therefore, probability forecasts, even if accurate, have little economic value.

When considering the predictability of excess stock return signs with different trading rules, one important evaluation criterion is the overall portfolio return, denoted by RET . Nevertheless, as for example, Hong and Chung (2003) emphasize, it is also worth considering risk-adjusted returns because different trading rules involve different levels of risk. In this evaluation, one commonly used measure is the Sharpe ratio (Sharpe 1966 and 1994)

$$SR = \frac{\overline{RET}^k - \overline{RET}^{rf}}{\hat{\sigma}^k}, \quad (13)$$

where \overline{RET}^k is the average portfolio return based on the model and trading rule k , \overline{RET}^{rf} is the average risk-free portfolio return (bond investment strategy) and $\hat{\sigma}^k$ is the sample standard deviation of portfolio returns RET_t^k . The higher the Sharpe ratio is, the higher the return and the lower the volatility. Portfolios with a high Sharpe ratio are preferable to those with a low Sharpe ratio.

3.2 Testing the Statistical Predictability

For the evaluation of the directional forecasting performance and market timing, a test proposed by Pesaran and Timmermann (1992) is available. The null hypothesis of this test is that the value of the correct prediction ratio CR does not differ statistically significantly from the ratio that would be obtained in the case of non-predictability, where the forecasts and the realized values of the return indicator I_t are independent. Granger and Pesaran (2000) show that the test statistic can be expressed as

$$PT = \frac{\sqrt{m} KS}{\left(\frac{\hat{P}_I(1-\hat{P}_I)}{I(1-I)}\right)^{1/2}}. \quad (14)$$

Here KS is the Kuipers score $KS = HR - FR$ between the "hit rate"

$$HR = \frac{\hat{I}^{uu}}{\hat{I}^{uu} + \hat{I}^{du}}$$

and the "false rate"

$$FR = \frac{\hat{I}^{ud}}{\hat{I}^{ud} + \hat{I}^{dd}}.$$

Forecast classification is denoted by

$$\begin{aligned}\hat{I}^{uu} &= \sum_{t=1}^m \mathbf{1}(I_t^f = 1, I_t = 1), \quad \hat{I}^{ud} = \sum_{t=1}^m \mathbf{1}(I_t^f = 1, I_t = 0), \\ \hat{I}^{du} &= \sum_{t=1}^m \mathbf{1}(I_t^f = 0, I_t = 1), \quad \hat{I}^{dd} = \sum_{t=1}^m \mathbf{1}(I_t^f = 0, I_t = 0),\end{aligned}$$

where f refers to forecast, u to an "up" signal ($I_t = 1$) and d to a "down" signal ($I_t = 0$) and $\mathbf{1}(\cdot)$ is an indicator function. Furthermore, in the test statistic (14), \bar{I} is the sample average of the sign indicator I_t values in the m -month sample period and $\bar{P}_I = \bar{I}HR + (1 - \bar{I})FR$. Under the null hypothesis of non-predictability, PT test statistic has an asymptotic standard normal distribution.

The directional predictability of an underlying data generation process is, however, not the same thing as a successful trading strategy. To evaluate the forecasts of the best forecasting models, we test the significance of the differences between the investment returns on the best models and trading strategies. This is tested by means of the Diebold-Mariano test (1995). Because the forecast horizon is one month, $h = 1$, the test statistic is

$$DM = \frac{\sqrt{m} \bar{d}}{\sqrt{\text{var}(\bar{d})}}, \quad (15)$$

where \bar{d} is the average difference between the predicted excess returns of the considered models. As in the PT test, under the null hypothesis of equal forecast accuracy, DM statistic also has an asymptotic $N(0, 1)$ -distribution.

4 Empirical Results

4.1 Data and Previous Findings

The monthly data set contains financial variables which have been used to predict the overall and the direction of excess stock returns in the previous literature. The data set covers the period from January 1968 to December 2006 and it is obtained from different sources mentioned in Appendix. The first 12 observations are used as initial values. The total number of observations, T , is 468. In out-of-sample forecasting, the data set is divided into two subsamples: the estimation and the forecasting sample.

The first out-of-sample forecasts will be made for January 1989 and last for December 2006.

In out-of-sample forecasting, parameters are estimated recursively using an expanding window of observations, where models are estimated using data from the start of data set through to present time to obtain a new one period forecast. This procedure is repeated until the end of the forecasting sample. In this study, the use of an alternative rolling estimation window is problematic, because there are not many recession periods in the post-1970 time and there would be estimation samples with no recession period at all.

The one-month excess stock return is defined as the continuously compounded return of the price index P_t minus the risk-free interest rate rf_t

$$r_t = 100 \log\left(\frac{P_t}{P_{t-1}}\right) - rf_t. \quad (16)$$

Here P_t is the value of the *S&P500* stock index and the one-month risk-free return rf_t is approximated by the three-month U.S. Treasury Bill rate i_t . With excess stock returns r_t , the values of the binary stock return indicator described in equation (1) can be formulated.

Several explanatory variables to forecast the direction of excess stock returns will be considered. As confirmed by Leung et al. (2000), the majority of useful information for forecasting stock returns is contained in interest rates and lagged stock returns. Hence, the financial explanatory variables, that are considered in the predictive models, are the short-term and long-term interest rates and their first differences, the U.S term spread, earnings/price and dividends/price-variables, and the realized volatility.

In previous studies, both the lagged excess stock returns and the lagged values of the return indicator are used as predictors. Leung et al. (2000) use first differences of interest rates and lagged excess stock returns in their comparison between the sign and the overall return forecasting models, and concluded that in probit and logit models, several past returns should be included in the model. If the explanatory power is distributed among many lags of past returns, then the autoregressive models (5) and (6) could be useful in forecasting. On the other hand, Anatolyev and Gospodinov (2007) use the lagged sign return indicator I_{t-1} in their dynamic logit model for the

direction of the future excess stock return. The corresponding estimated regression coefficient was positive but statistically insignificant.

Interest rate spreads between different maturities may offer information about future expectations in financial markets (see, for instance, Fama and French, 1989). In recession forecasting, the term spread (SP^{US}) is expected to transmit the expectations for future monetary policy. The lower is the difference between the long-term and short-term interest rates, the more restrictive is the current monetary policy. The term spread could also have its own implication on the stock market, not only on the real economic activity.

Dividends (D_t) and earnings (E_t) divided by the value of the price index P_t have been among the most commonly used explanatory variables (e.g. Campbell and Shiller, 1988 and Cochrane, 1997). In this paper, the dividend-price (earnings-price) ratio is computed by dividends (earnings) of *S&P500* stock index companies over the past year. Since monthly data of dividends and earnings are not available, DP_t and EP_t are constructed as sums of dividends and earnings over the past year divided by the current, monthly price level P_t .

Numerous studies have also documented a notable dependence of stock return and stock return volatility with important implications on asset pricing. In this paper, the realized monthly volatility σ_t , based on the sum of squared daily observations within one month (Cristoffersen et al. 2006), is examined.

4.2 In-Sample Results

Even though our main interest lies in the out-of-sample predictions of the direction of future excess stock returns, at first, the in-sample performance of different probit models and combinations of explanatory variables are experimented using the sample period from January 1968 until to December 1988. Explanatory variables are included one by one in the model.⁶

The main results/findings are as follows. According to the statistical significance of predictors, psR^2 -values, and the returns of trading strategies in the chosen in-sample period, the recession forecast and the first difference of the short-term interest rate are

⁶ All estimations and forecasts are computed with MATLAB 7.4.0.

the best predictive variables. Employing these variables, there seems to be evidence that the excess stock return signs are predictable in-sample. The first difference of the long-term interest rate and the realized volatility also have some predictive power. Interestingly, the corporate earning (EP_t) and dividend (DP_t) variables, used in many previous studies, are not particularly useful predictors.⁷ When the recession forecast is employed with different financial explanatory variables in the model equation, the evidence is very much the same as in the above. The first difference of the short-term interest rate appears to be the best predictor with recession forecast in-sample.

Table 1 presents details of parameter estimates in different probit models when the recession forecast and the first difference of the short-term interest rate Δi_{t-1} are used as explanatory variables. The robust standard errors suggested by Kauppi and Saikkonen (2007) are presented. However, it should be noted that these standard errors may be inaccurate because the estimated recession forecast is employed in the model.⁸

It seems that the lagged stock indicator I_{t-1} has not statistically significant predictive ability for the sign of the stock return. In the error correction model, the autoregressive coefficient α_1 is clearly statistically significant but in other dynamic models, it is not. As expected, the estimated coefficients of the first difference of the short-term interest rate and recession forecast are negative. In the error correction model, the recession forecast is not statistically significant but the first difference of the short-term interest rate is. In other models, both of these predictors are statistically significant predictors according to presented robust standard errors.

Figure 3 depicts the realized values of the sign indicator I_t and the estimated probability of a positive excess stock return in the static (3) model (left panel) and in the error correction model (8) (right panel), whose estimation results are shown in Table 1 (the first and the fifth model). Both models seems to give rather the same in-sample predictions. In recessionary periods, both models suggest to investing in risk-free interest rate. More or less the only significant difference between the models is approximately the time period from 1976 to 1979. At that time, the probability

⁷ Further information on in-sample performance of the different models is available upon request.

⁸ In addition, there are no formal proofs of the asymptotic distributions of the maximum likelihood estimator in models (5), (6) and (8)

forecast in the static model is above 0.50 threshold value while in the error correction model it is below the threshold.

As the psR^2 -values indicate, the statistical predictive power for the sign of the excess stock return is, as expected, quite low. Although the statistical predictability is small, the portfolio investment performance yields evidence about a useful sign predictability in excess stock returns. The average rates of returns on different models and trading strategies are higher than in the "passive" buy-and-hold trading strategy (hereafter B&H strategy), where one is investing only in stocks. This annualized benchmark return is 6.07 %. In the best models, the returns, including the transaction costs are between 9.50 % to 10.50 %. The error correction probit model seems to yield smaller in-sample returns than its counterparts.

The statistical significance of return differences between an examined model and B&H returns is tested. Table 1 presents the values of the statistical test statistics introduced in Section 3.2. Since we are only interested in cases where the proportion of correctly predicted signs and the portfolio returns in estimated models are higher than under the null hypothesis of unpredictability, only the positive and statistically significant values of the PT and DM test statistics (see (14) and (15)) provide evidence of predictability. The values of the market timing test statistic PT are statistically significant at the 1 % level under all experimented models in Table 1. Thus, the null hypothesis of non-predictability is rejected providing in-sample evidence that excess return signs are predictable. In the DM tests, the null hypothesis of equal performance between the returns in the considered model and the B&H strategy are rejected in all models at 5 % level except in the error correction model where the p -value of the test statistic is 0.123. When risk-adjusted returns are considered, the DM_{ra} test statistics are statistically significant in all models providing evidence of profitable trading strategies based on the forecasts from the probit models.

Although the unrestricted dynamic autoregressive probit model (6) gives a better in-sample fit than the error correction model in models presented in Table 1, in some other models with different explanatory variables, the error correction model gives higher pseudo- R^2 and CR values. In contrast to the error correction model (8), in many other unrestricted dynamic autoregressive models, the autoregressive

coefficient α_1 is typically negative. Thus, if the probability of a positive excess stock return has been high in some period, it tends to be lower in the next period. As an example, consider a model in which the first difference of the short-term interest rate Δi_{t-1} is the only explanatory variable. Figure 4 shows the estimated probabilities of positive excess stock returns in the unrestricted dynamic autoregressive model (6) (left panel) and in the error correction model (8) (right panel) in this example case. As can be seen from the left panel, when the estimate for the autoregressive coefficient α_1 is negative, the probability of excess stock return is fluctuating heavily around the threshold value 0.50. Respectively, on the other hand, in the right panel with a positive and high estimate of α_1 , the probability forecasts follow a relatively persistent swing. Thus, it seems that the error correction model yields less transactions between stocks and bonds, and consequently also less transaction costs, than the unrestricted dynamic autoregressive model. This is particularly striking in models presented in Figure 4 and could be an important property in out-of-sample forecasting.

4.3 Out-of-Sample Results

When forecasting the signs of excess stock returns, it is important to compare the different models out-of-sample. Previous results on the predictive models for the overall excess stock return suggest that the in-sample predictability does not necessarily imply out-of-sample predictability. For example, Han (2007) finds that a statistically superior predictive VAR-GARCH model in-sample that does not consistently outperform its competitors in terms of portfolio investments returns out-of-sample. Goyal and Welch (2004) argue that traditional predictive models for the excess return cannot beat the historical average return out-of-sample and there is no single variable that has theoretically meaningful and robust explanatory power. On the other hand, Campbell and Thompson (2007) show that some predictors perform better than the historical average when restrictions on regression coefficients are imposed. They and, for instance, Anatolyev and Gospodinov (2007) have stressed that while the out-of-sample predictive power is small it can be utilized in market timing decisions to earn economically higher excess returns than the B&H strategy even out-of-sample.

In this study, the out-of-sample period consists of 216 months from January 1989

to December 2006. The out-of-sample recession forecast p_{t+5}^y is constructed before making any sign forecasts for excess return signs. As described in Section 4.1, the parameters in the sign and in the recession prediction models are estimated recursively. Table 2 shows the out-of-sample performance of the best in-sample predictive models and also some other probit models. The idea is to compare the predictive performance of different models when the same combinations of explanatory variables, that turned out to be the best out-of-sample predictive variables, are examined.⁹

According to commonly used statistical model evaluation measures there is not much out-of-sample predictability in excess stock return signs. The values of the out-of-sample psR^2 -measures (see equation (12)) are, even in the best models, small or even negative¹⁰. The percentage of correct forecasts, CR , vary between 0.51 and 0.61. Contrary to the employed statistical measures, the results of portfolio returns RET and Sharpe ratios SR exhibit evidence of useful predictability for asset allocation decisions even though average portfolio returns vary strongly between different models. As in the in-sample evidence, the models with recession forecasts generate better sign forecasts than models without these forecasts. It is worth noting that the sign prediction models containing the constructed recession forecast outperform the models containing the variables used in recession forecasting, especially out-of-sample.

It is interesting that the error correction model (8) clearly outperforms the corresponding unrestricted dynamic autoregressive model (6) out-of-sample. As seen in the best in-sample models in Table 1, the autoregressive model (5) and the dynamic autoregressive model (6) outperform the error correction model (8) but the out-of-sample evidence seems to be very different. In Table 2, the psR^2 -values of the error correction models are clearly positive and the ratios of correct predictions, CR , are higher than in other probit models considered. Above all, error correction models can generate more profitable trading strategies than the other probit models. Perhaps the most striking finding is the performance of a model, with no explanatory variables or the recession forecast ("—" in Table 2). The psR^2 -values, CR -ratio, average excess returns and Sharpe ratios are clearly higher in the error correction model.

⁹ The models with other financial variables, presented in Section 4.1, are also considered and the results on those models are available upon request.

¹⁰ Negative psR^2 means very poor out-of-sample forecasting performance (Estrella, 1998).

Overall, compared with the dynamic models (4)–(6), the static probit model (3), without the autoregressive term π_{t-1}^I or the lagged I_{t-1} , seems to be an adequate model for the excess stock return sign. The error correction model (8) appears to be the only dynamic model which yields better forecasts in this data set than the static model. A potential identification problem in the dynamic autoregressive model (6) discussed in Section 2.2 is a possible explanation for this superior out-of-sample performance of the error correction model.

The recession forecast is the main predictive variable in different models. The first differences of the short-term and the long-term interest rates are also fairly good predictors almost in all probit models and perform consistently better than other financial explanatory variables examined. For instance, the realized volatility was quite a good predictive variable in-sample but its out-of-sample performance is very poor.

As a robustness check in the comparison between different models, we apply a model confidence set approach suggested by Hansen, Lunde and Nason (2003, 2005) to determine a set of the best forecasting models. Construction of the MCS involves a sequence of tests for equal predictive ability. At each step the null hypothesis of equal investment return performance between a forecasting model and the realized return is tested. After the first non-rejection, the $1 - \alpha$ model confidence set is constructed at given significance level α .¹¹ In this study, using different proposed test statistics (see details Hansen et.al. 2003 and 2005), a model set based on the 75 % significance level is constructed since the return differences between models are relatively small and thus the traditional significance levels α , such as 5 % level, are not very informative because in that case, almost all models are selected in the MCS.

According to the in-sample evidence, only the models in which the recession forecast is used as a predictive variable are considered. Thus, there are nine different combinations of explanatory variables and five different probit models. Given this confidence level of 25 % the MCS contains only two error correction models presented in Table (4). Therefore, it gives further evidence that the error correction

¹¹ The MCS results are computed using Ox version 4.1 (see Doornik 2006) and the MULCOM package version 1.00 (Hansen, Lunde and Nason, 2005).

probit model yields the best forecasts in terms of investment returns out-of-sample.

Figure 5 depicts the out-of-sample probability forecasts for the positive excess stock returns in two models which are highlighted in Table 2. The most notable difference is that in 2002, the error correction model gives a signal to invest in interest rate when the monthly stock returns are most of the time negative. Further, in Table 4 the values of the PT and DM test statistics in these best error correction and static model are shown. In the error correction model, the p -value of the PT test statistic is 0.053 and the p -value of the DM test statistic is 0.121. The risk adjusted returns are statistically significantly higher than the returns in the B&H strategy (p -value 0.00). On the other hand, the p -values of test statistics in the best static probit model reflects that the excess stock return signs are not predictable with this model.

It can be seen that all models mostly suggest investing in stocks. For instance, in Figure 5, the conditional probabilities of the positive excess stock return are above the 0.50 threshold. Thus, the return differences between probit models and the B&H-trading strategy are zero in most months. In addition, because the probability of recession is principally almost zero when the economy is in the expansionary state, the recession forecast should be a particularly useful predictor of negative excess stock return months when the economic activity is declining. Hence, the values of the DM test statistics are also calculated based on only those months when the excess stock returns have been non-positive (that is $I_t = 0$). There are 86 months with negative excess return in out-of-sample period. In Table 4, the values of test statistics $DM^{I_t=0}$, and $DM_{ra}^{I_t=0}$ in the case of risk-adjusted returns, are strongly statistically significant in the best models. As seen, the in-sample evidence in Table 1 is similar.

When the investment returns on the best error correction model and the best static model are compared, the p -value of the DM test statistic is 0.111 and 0.048, respectively, when the risk-adjusted returns are considered. Thus, the error correction model yields higher returns but the statistical significance between return differences is relatively weak in the considered out-of-sample period. According to "asymmetric" DM test statistics discussed above, the best error correction model outperform the best static model in all traditional statistical significance levels.

4.4 Comparison Between ARMAX and Probit Models

It is interesting to make some comparisons between the forecasting performance of sign predictions in probit models and in ARMAX models. In fact, there is not many previous studies that compare the predictive performance of these models. Leung et al. (2000) find some evidence that qualitative response models, including logit and probit models, outperform continuous models in their out-of-sample forecasting. They considered a sample of U.S., U.K. and Japanese stock indices from January 1991 to December 1995. In their study, the ratios of correct sign predictions and the investment returns from qualitative dependent models are higher than in models for continuous variables.

In ARMAX models the same explanatory variables as in probit models are considered. The dependent variable is the excess stock return and it is assumed that a positive forecast gives the signal to buy stocks (i.e. $I_t^f = 1$). This is consistent with the definition of the stock return indicator (1). As in probit models, the in-sample predictive performance of different ARMAX models is first analyzed.¹² The estimated values of the BIC model selection criterion (Schwarz 1978) suggested an ARMAX(2,0)¹³ model with the first difference of the short-term interest rate Δi_{t-1} and the recession forecast p_{t+5}^y as explanatory variables. An ARMAX(1,0) model with the recession forecast and the U.S. term spread SP^{US} generates the highest in-sample investment return.

Out-of-sample forecasting performance of the best in-sample models and some other ARMAX models are shown in Table 3. The values of the out-of-sample coefficient of determination R_{OS}^2 , suggested by Campbell and Thompson (2007), are reported although they are not fully comparable to the out-of-sample psR^2 -values used in probit models. As a whole, the percentage of the correct forecasts among the best probit models appears to be somewhat higher than in the best ARMAX models but, in particular, the investment return performance is clearly better among in the best probit models. Only the best two ARMAX models, highlighted in Table 3 and

¹² In the estimation of ARMAX models, UCSD_GARCH toolbox package (Kevin Sheppard) for MATLAB is used.

¹³ For instance, ARMAX(2,0) is the same as the AR(2)-model with explanatory variables.

depicted in Figure 6, yield considerable higher returns than other ARMAX models. When the return differences between the best error correction model and the best ARMAX models are tested, the return differences are also statistically significant at 5 % level based on the all considered *DM* test statistics shown in Table 5. Therefore, the error correction model seems to superior predictive model also against the alternative ARMAX models.

5 Conclusions

In this paper, we examine the predictability of the U.S. excess stock return signs by using dynamic binary probit models. The proposed forecasting method, where the six-month recession forecast for the recession indicator is used as an explanatory variable seems to outperform other predictive models. The direction of the excess stock return is predictable and it is possible to earn statistically significantly higher investment returns compared with the buy-and-hold trading strategy in-sample. However, out-of-sample predictability turns out to be weaker. This is in line with previous findings related to stock return forecasting. In fact, in out-of-sample forecasting, the best dynamic probit model appears to be the error correction model proposed in the paper. Using this model, the average investment returns and the number of correct sign predictions are higher than in other probit models or in the buy-and-hold trading strategy and also compared with the best ARMAX models out-of-sample. The return differences between the best probit and ARMAX models are also statistically significant.

The analysis of the paper can be extended in various ways. A system analysis in which the recession and sign forecasts are determined endogenously in the same model is of particular interest. In this paper the six-month ahead recession forecast is taken as a given, but this selection need not be optimal in terms of predictive power in sign predictions. It could also be interesting to compare the directional predictability in other countries and employing some other continuous time series models, such as GARCH models.

References

Anatolyev, S., and Gospodinov, N. (2007): Modeling financial return dynamics by decomposition. Centre for Economic and Financial Research at New Economic School. Working Paper, 95. <http://www.cefir.ru/papers/WP95Anatolyev.pdf>.

Breen, W., Glosten, L. R., and Jagannathan, R. (1988): Economic significance of predictable variations in stock index returns. *Journal of Finance*, 44, 1177–1189.

Campbell, J. Y., and Shiller, R. J. (1988): Stock Prices, earnings, and expected dividends. *Journal of Finance*, 43(3), 661–676.

Campbell, J. Y., and Thompson, S. B. (2007): Predicting excess stock returns out of sample: Can anything beat the historical average. *Review of Financial Studies*, forthcoming.

Chauvet, M., and Potter, S. (2000): Coincident and leading indicators of the stock market. *Journal of Empirical Finance*, 7, 87–111.

Chen, N. (1991): Financial investment opportunities and the macroeconomy. *Journal of Finance*, 46, 529–554.

Cochrane, J. H. (1997): Where is the market going? Uncertain facts and novel theories. *Economic Perspectives*, 21(6), Federal Reserve Bank of Chicago, 3–37.

Cristoffersen, P. F., and Diebold, F. X. (2006): Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science*, 52, 1273–1287.

Cristoffersen, P. F., Diebold, F. X., Mariano, R. S., Tay, A. S., and Tse, Y. K. (2006): Direction-of-change forecasts based on conditional variance, skewness and kurtosis dynamics: International evidence. *Journal of Financial Forecasting*, 1, 3–24.

de Jong, R. M., and Woutersen, T. M. (2007): Dynamic time series binary choice. Economics Working Paper Archive, 538. The Johns Hopkins University, Department of Economics. Available at <http://www.econ.jhu.edu/pdf/papers/WP538.pdf>.

Doornik, J. A. (2006): *Ox: An Object-Orientated Matrix Programming Language*. Timberlake Consultants Ltd., London, 4.1 edition.

Engle, R. F., and Bollerslev, T. (1986): Modeling the persistence of conditional variances. *Econometric Reviews*, 5, 1–50.

Estrella, A. (1998): A new measure of fit for equations with dichotomous dependent variables. *Journal of Business and Economic Statistics*, 16, 198–205.

Estrella, A. (2005): The yield curve as a leading indicator: Frequently asked questions. Federal Reserve Bank of New York. http://www.newyorkfed.org/research/capital_markets/ycfaq.pdf.

Estrella, A., and Mishkin, F. S. (1998): Predicting U.S. recessions: Financial variables as leading indicators. *Review of Economics and Statistics*, 80(1), 45–61.

Fama, E. F. (1990): Stock returns, expected returns, and real activity. *Journal of Finance*, 45, 1089–1108.

Fama, E. F., and French, K. R. (1989): Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, 25, 23–49.

Goyal, A., and Welch, I. (2004): A comprehensive look at the empirical performance of equity premium prediction. NBER Working Paper, 10483.

Granger, C. W. J., and Pesaran, M. H. (2000): Economic and statistical measures of forecast accuracy. *Journal of Forecasting*, 19, 537–560.

Han, Y. (2007): Return predictability, economic profits, and model mis-specification: How important are the better specified models? Tulane University. <http://ssrn.com/abstract=967564>.

Hansen, P. H., Lunde, A., and Nason, J M. (2003): Choosing the best volatility models: The model confidence set approach. *Oxford Bulletin of Economics and Statistics*, 65, 839–861.

Hansen, P. H., Lunde, A., and Nason, J M. (2005): Model confidence sets for forecasting models. Federal Reserve Bank of Atlanta, working paper 2005-7.

Hong, Y., and Chung, J. (2003): Are the directions of stock price changes predictable? Statistical theory and evidence. Manuscript, Cornell University.

Kauppi, H., and Saikkonen, P. (2007): Predicting U.S.recessions with dynamic binary response models. *Review of Economics and Statistics*, forthcoming.

Leitch, G., and Tanner, J. (1991): Economic forecasts evaluation: Profits versus the conventional error measures. *American Economic Review*, 81, 580–590.

Leung, M. T., Daouk H., and Chen, A. S. (2000): Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Forecasting*, 16, 173–190.

Maddala, G. S. (1983): *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, New York.

Nyberg, H. (2008): Dynamic Probit Models and Financial Variables in Recession Forecasting. HECER Discussion Paper, 225. Helsinki Center of Economic Research.

Pagan, A. (1984): Econometric issues in the analysis of regressions with generated regressors. *International Economic Review*, 25(1), 221–247.

Pesaran, M. H., and Timmermann A. (1992): A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics*, 10, 461–465.

Pesaran, M. H., and Timmermann A. (1995): Predictability of stock returns: Robustness and economic significance. *Journal of Finance*, 50(4), 1201–1228.

Rydberg, T., and Shephard, N. (2003): Dynamics of trade-by-trade price movements: Decomposition and models. *Journal of Financial Econometrics*, 1, 2–25.

Russell, J. R., and Engle, R. F. (2005): A discrete-state continuous-time model of financial transactions prices and times: The autoregressive conditional multinomial-autoregressive conditional duration model. *Journal of Business and Economic Statistics*, 23, 166-180.

Schwarz, G. (1978): Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464.

Sharpe, W. F. (1966): Mutual fund performance. *Journal of Business*, 39, 119–138.

Sharpe, W. F. (1994): The Sharpe ratio. *Journal of Portfolio Management* Fall, 49–58.

Appendix: Data Set

Stock index P_t : Standard&Poors 500 U.S. stock index. <http://finance.yahoo.com> and <http://www.econstats.com>.

Recession periods y_t : NBER chronology of turning points of business cycles. <http://www.nber.org/cycles.html> and <http://www.nber.org/cycles/recessions.html>.

Short-term interest rate i_t : Three month U.S. Treasury Bill rate, secondary market. <http://www.federalreserve.gov/releases/h15/data.htm>.

Long-term interest rate R_t : Ten year U.S. Treasury Bond rate, constant maturity. <http://www.federalreserve.gov/releases/h15/data.htm>.

First differences of the short and the long-term interest rates $\Delta i_t = i_t - i_{t-1}$ and $\Delta R_t = R_t - R_{t-1}$.

U.S. term spread $SP_t^{US} = R_t - i_t$.

One month realized variance σ_t : Sum of squared daily stock returns in S&P500-index within one month. <http://finance.yahoo.com> and <http://www.econstats.com>.

Dividend-price ratio $DP_t = D_t/P_t$: (log) Dividends over the past year divided by the current stock index value.

Earnings-price ratio $EP_t = E_t/P_t$: (log) Earnings over the past year divided by the current stock index value.

Data for dividends and earnings (S&P500 index firms) is obtained from the homepage of Robert Shiller's book *Irrational exuberance* <http://www.irrationalexuberance.com> and Standard&Poor's dividends and earnings report <http://www2.standardandpoors.com>.

Tables and Figures

Table 1: In-sample estimation results.

	static	dynamic	auto.	dyn.auto.	ecm
	model(3)	model(4)	model(5)	model(6)	model(8)
ω	0.07 (0.10)	0.02 (0.12)	0.07 (0.09)	0.03 (0.13)	-0.07 (0.03)
π_{t-1}^I			0.04 (0.24)	-0.03 (0.28)	0.85 (0.05)
I_{t-1}		0.08 (0.11)		0.08 (0.19)	
Δi_{t-1}	-0.24 (0.04)	-0.30 (0.12)	-0.30 (0.12)	-0.29 (0.12)	-0.16 (0.07)
p_{t+5}^y	-0.50 (0.25)	-0.47 (0.25)	-0.49 (0.24)	-0.49 (0.25)	-0.02 (0.05)
log-L	-161.33	-161.22	-160.72	-160.64	-162.10
psR^2	0.041	0.041	0.046	0.046	0.034
CR	0.580	0.591	0.579	0.579	0.579
RET	10.50	9.98	9.88	9.57	8.12
SR	0.95	0.86	0.79	0.72	0.25
PT	0.003	0.002	0.006	0.006	0.008
DM	0.009	0.017	0.017	0.023	0.123
DM_{ra}	0.000	0.000	0.000	0.000	0.003
$DM^{I_t=0}$	0.000	0.000	0.000	0.000	0.000
$DM_{ra}^{I_t=0}$	0.000	0.000	0.000	0.000	0.000

Notes: The models are estimated using the in-sample data from 1969 M1 to 1988 M12. Robust standard errors, given in parentheses, are computed with procedures suggested by Kauppi and Saikkonen (2007). The error correction model (8) is denoted by "ecm". RET is the average annualized in-sample portfolio return in considered model and SR is the corresponding Sharpe ratio (13). The p -values of the Pesaran and Timmermann (1992) test (14) and Diebold and Mariano (1995) test (15) are reported. In DM tests, the buy-and-hold trading strategy is the benchmark. Further, ra means the risk-adjusted returns. The values of test statistics $DM^{I_t=0}$ and $DM_{ra}^{I_t=0}$ are obtained when only months with negative excess return ($I_t = 0$) are considered.

Table 2: Out-of-sample performance of different probit models.

model	\mathbf{x}_{t-1}	psR^2	CR	RET	SR
static (3)	p_{t+5}^y	0.015	0.59	7.63	0.86
static (3)	$\Delta R_{t-1}, p_{t+5}^y$	0.013	0.59	8.02	0.97
static (3)	$\Delta i_{t-1}, p_{t+5}^y$	0.014	0.59	7.02	0.72
dynamic (4)	–	neg.	0.53	3.92	0.05
dynamic (4)	p_{t+5}^y	0.005	0.57	6.16	0.57
dynamic (4)	$\Delta R_{t-1}, p_{t+5}^y$	0.002	0.57	7.32	0.88
dynamic (4)	$\Delta i_{t-1}, p_{t+5}^y$	0.004	0.58	7.46	0.92
auto (5)	p_{t+5}^y	neg.	0.51	2.30	-0.40
auto (5)	$\Delta R_{t-1}, p_{t+5}^y$	0.015	0.58	7.20	0.79
auto (5)	$\Delta i_{t-1}, p_{t+5}^y$	0.014	0.58	6.53	0.61
dyn.auto (6)	–	neg.	0.51	4.11	0.10
dyn.auto (6)	p_{t+5}^y	0.008	0.57	5.42	0.40
dyn.auto (6)	$\Delta R_{t-1}, p_{t+5}^y$	0.006	0.56	7.03	0.75
dyn.auto (6)	$\Delta i_{t-1}, p_{t+5}^y$	0.011	0.56	5.46	0.38
ecm (8)	–	0.014	0.59	8.62	1.08
ecm (8)	p_{t+5}^y	0.018	0.61	10.33	1.58
ecm (8)	$\Delta R_{t-1}, p_{t+5}^y$	0.016	0.59	9.09	1.23
ecm (8)	$\Delta i_{t-1}, p_{t+5}^y$	0.017	0.59	9.78	1.39

Notes: See also notes to Table 1. The average return of the buy-and-hold trading strategy is 7.52 % (annual) with the corresponding Sharpe ratio $SR=0.79$. The risk free return or return of bond holding strategy is 3.71 %. A note "neg" means negative psR^2 -value.

Table 3: Out-of-sample performance of ARMAX models.

model	\mathbf{x}_{t-1}	R_{OS}^2	CR	RET	SR
ARMAX(1,0)	p_{t+5}^y	0.031	0.56	7.12	0.78
ARMAX(1,0)	p_{t+5}^y, SP_{t-1}^{US}	0.055	0.54	6.07	0.57
ARMAX(1,0)	$p_{t+5}^y, \Delta R_{t-1}$	0.084	0.56	4.98	0.31
ARMAX(1,0)	$p_{t+5}^y, \Delta i_{t-1}$	0.077	0.57	6.16	0.57
ARMAX(2,0)	p_{t+5}^y	0.034	0.58	5.83	0.48
ARMAX(2,0)	p_{t+5}^y, SP_{t-1}^{US}	0.056	0.51	4.36	0.16
ARMAX(2,0)	$p_{t+5}^y, \Delta R_{t-1}$	0.092	0.53	4.76	0.26
ARMAX(2,0)	$p_{t+5}^y, \Delta i_{t-1}$	0.082	0.59	6.28	0.59

Notes: The ARMAX($p, 0$) model for r_t is: $r_t = a + \sum_{i=1}^p b_i r_{t-i} + \mathbf{x}_{t-1} \mathbf{d}$. See also notes to Table (2).

Table 4: Statistical tests for the best error correction models and the best static probit model.

model	\mathbf{x}_{t-1}	CR	RET	SR	PT	DM	DM_{ra}	$DM^{I_t=0}$	$DM_{ra}^{I_t=0}$
ecm	p_{t+5}^y	0.61	10.33	1.58	0.053	0.121	0.011	0.000	0.000
ecm	$p_{t+5}^y, \Delta R_{t-1}$	0.59	9.78	1.45	0.278	0.274	0.028	0.000	0.000
static	$p_{t+5}^y, \Delta R_{t-1}$	0.59	8.02	0.97	0.326	0.348	0.189	0.000	0.000

Notes: The best error correction model (8) (*ecm*) and the best static model (3), highlighted in Table 2, are presented. The another error correction model is the model included in the estimated model confidence set (MCS). The p -values of the PT and DM tests are presented. In the DM tests the B&H trading strategy is the alternative asset allocation strategy. In Table, ra means the risk-adjusted returns and the test statistics $DM^{I_t=0}$ and $DM_{ra}^{I_t=0}$ are obtained when only months with negative excess return ($I_t = 0$) are taken into account.

Table 5: Diebold-Mariano tests between the best error-correction model and the best ARMAX models.

model	\mathbf{x}_{t-1}	DM	DM_{ra}	$DM^{I_t=0}$	$DM_{ra}^{I_t=0}$
ARMAX(1,0)	p_{t+5}^y	0.046	0.019	0.000	0.000
ARMAX(2,0)	$p_{t+5}^y, \Delta i_{t-1}$	0.015	0.006	0.000	0.000

Notes: The p -values of Diebold and Mariano tests between the returns from the error-correction model presented in the first row in Table 4 and the ARMAX models mentioned in the first column.

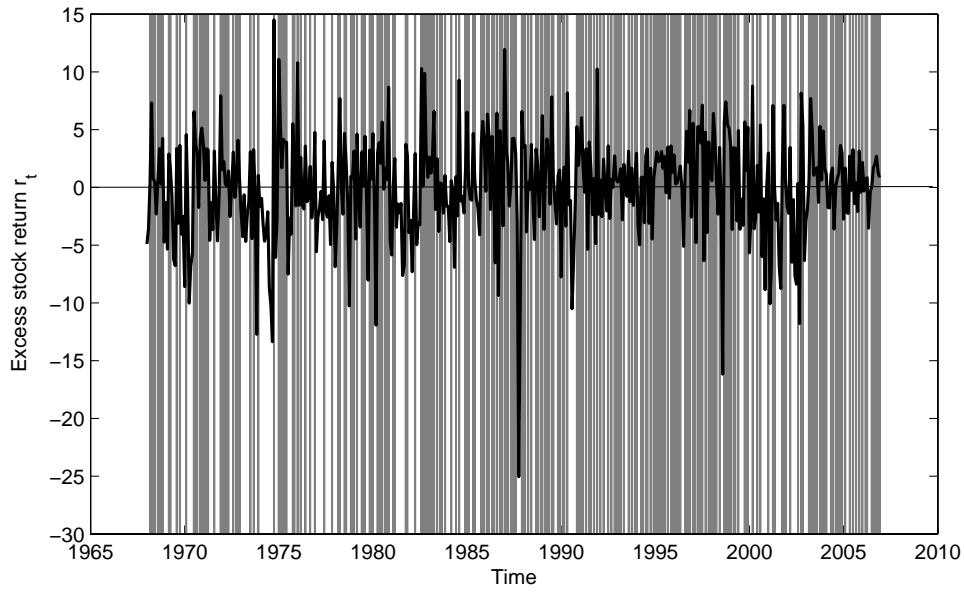


Figure 1: Excess stock returns and the values of the indicator variable I_t (shaded area when $I_t = 1$).

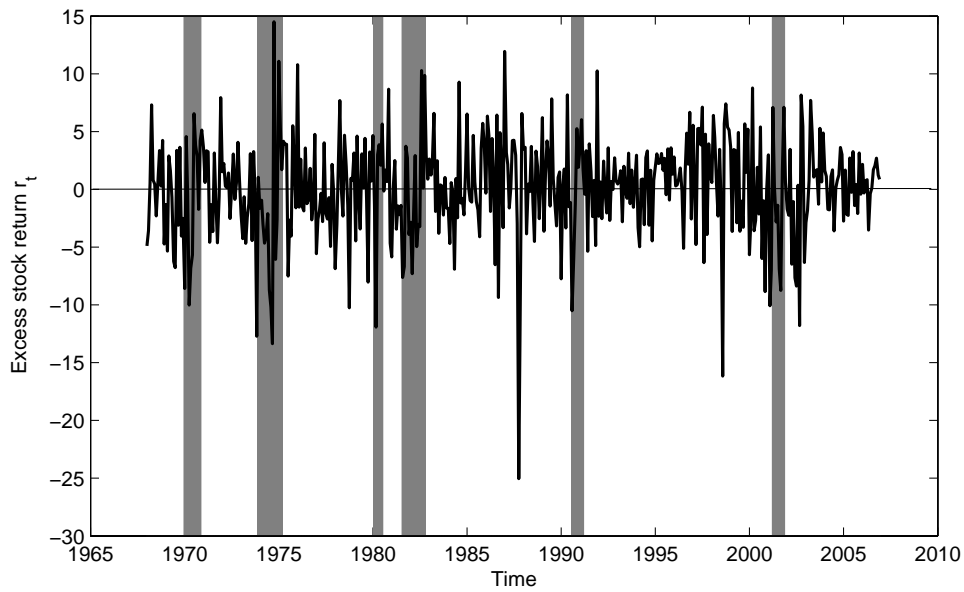


Figure 2: Excess stock returns and the NBER recessions y_t (shaded areas).

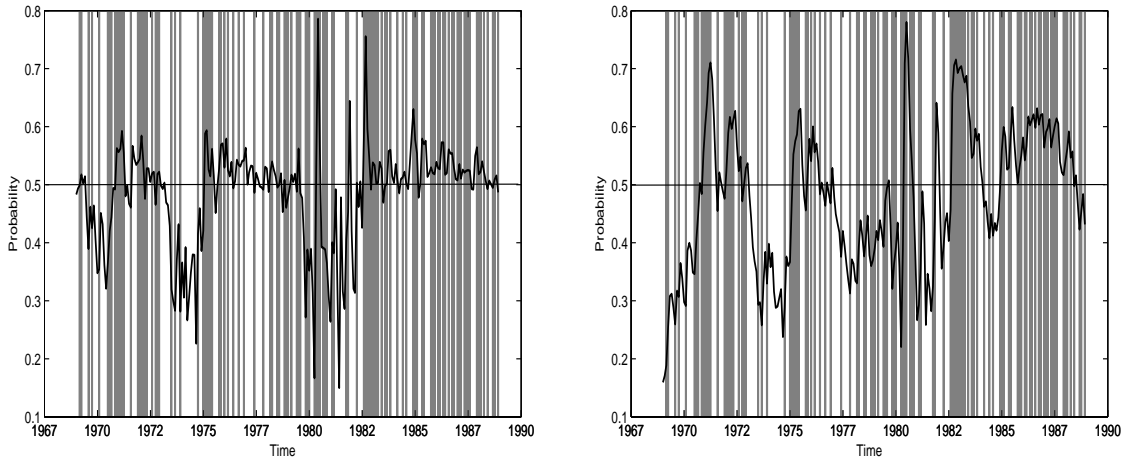


Figure 3: In-sample predictions of the static model (3) and the error correction model (8). The shaded area reflects the months with positive excess stock returns. The black line is the probability forecast p_t^I and the 50 percent threshold is also depicted.

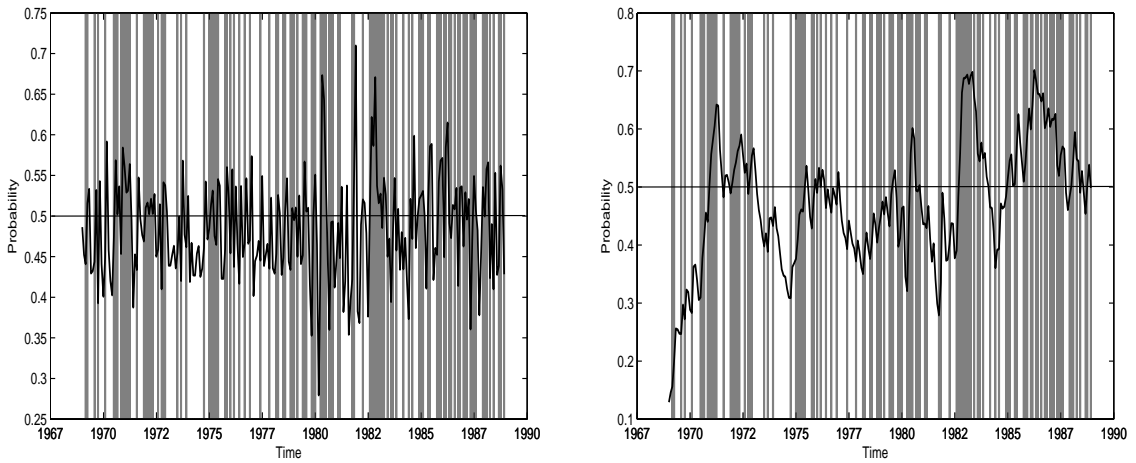


Figure 4: In-sample predictions of the dynamic autoregressive (6) and the error correction model (8) when Δi_{t-1} is the only predictive variable.

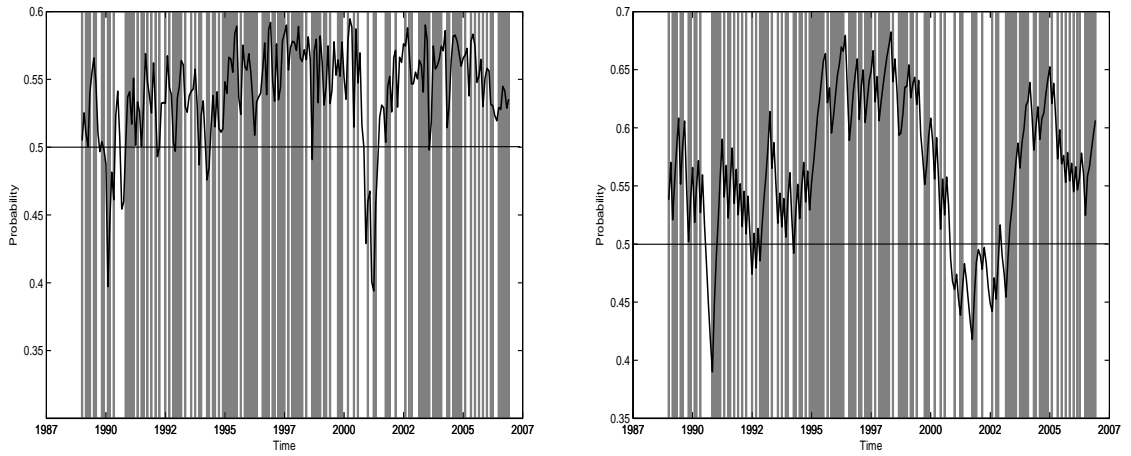


Figure 5: Out-of-sample predictions of the static (3) model with ΔR_{t-1} and p_{t+5}^y (left panel). In the right panel, the error correction model (8) with p_{t+5}^y .

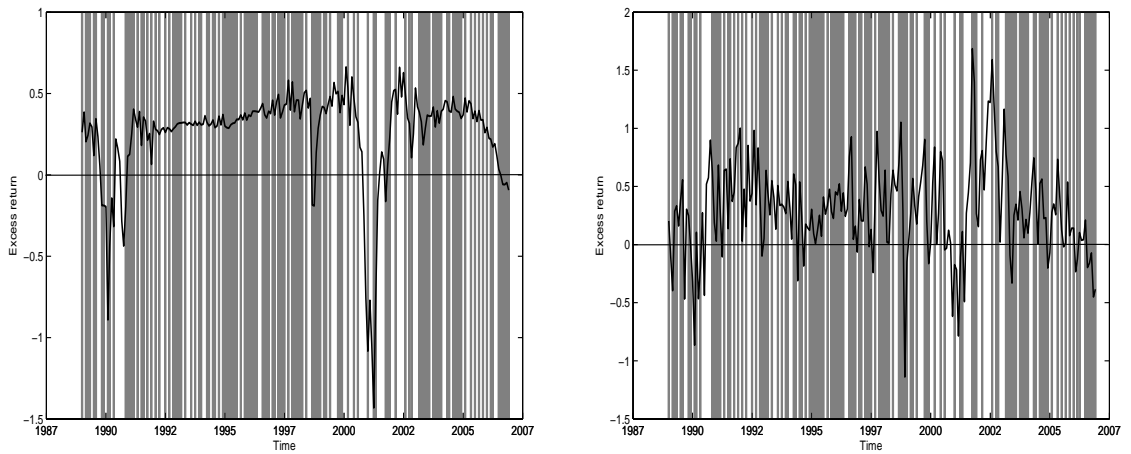


Figure 6: Out-of-sample predictions of the ARMAX(1,0) model with p_{t+5}^y (left panel). In the right panel, ARMAX(2,0) model with p_{t+5}^y and Δi_{t-1} is presented.