International Sign Predictability of Stock Returns: The Role of the United States

Henri Nyberg and Harri Pönkä

CREATES Research Paper 2015-20

Department of Economics and Business
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

International Sign Predictability of Stock Returns: The Role of the United States*

Henri Nyberg†
University of Helsinki

Harri Pönkä‡
University of Helsinki and CREATES

Abstract

We study the directional predictability of monthly excess stock market returns in the U.S. and ten other markets using univariate and bivariate binary response models. Our main interest is on the potential benefits of predicting the signs of the returns jointly, focusing on the predictive power from the U.S. to foreign markets. We introduce a new bivariate probit model that allows for such a contemporaneous predictive linkage from one market to the other. Our in-sample and out-of-sample forecasting results indicate superior predictive performance of the new model over the competing models by statistical measures and market timing performance, suggesting gradual diffusion of predictive information from the U.S. to the other markets.

Keywords: Excess stock return, Directional predictability, Bivariate probit model, Market timing

JEL classification: C22, G12, G17

*The authors would like to thank Shamim Ahmed, Thanaset Chevapatrakul, Charlotte Christiansen, Markku Lanne, Matthijs Lof, Sabrina Struder, and seminar participants at Nottingham University Business School (April 2015), CREATES (Aarhus University, March 2015), and the Inaugural Royal Economic Society Symposium of Junior Researchers in Manchester (April 2015) for useful comments. Financial support from the Academy of Finland is gratefully acknowledged. The corresponding author (Pönkä) also acknowledges support from the Yrjö Jahnsson Foundation, the Foundation for the Advancement of Finnish Security Markets, and CREATES – Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

† Address: Department of Political and Economic Studies, University of Helsinki, P.O.Box 17 (Arkadiankatu 7), FIN–00014 University of Helsinki, Finland, E-mail: henri.nyberg@helsinki.fi.

‡ Corresponding author. Address: Department of Political and Economic Studies, University of Helsinki, P.O.Box 17 (Arkadiankatu 7), FIN–00014 University of Helsinki, Finland, E-mail: harri.ponka@helsinki.fi.

1 Introduction

There is a vast theoretical and empirical literature on asset return predictability. The main focus in the literature on stock returns has been the predictability of excess aggregate market returns (hereafter stock returns) by lagged financial and macroeconomic predictive variables. Although the majority of research has concentrated on the U.S., there is an increasing string of research focusing on lead-lag relationships in international asset markets. Rapach et al. (2005) examine the predictability of stock returns in 12 industrialized countries and find that interest rates are the most consistent and reliable predictors of stock returns. In the same vein, Ang and Bekaert (2007) show that the dividend yields and short-term interest rates are robust predictors for the stock returns in the U.S., U.K., France, and Germany. Hjalmarsson (2010) examines return predictability in a larger dataset comprising 40 developed international stock markets. Similarly as Rapach et al. (2005) and Ang and Bekaert (2007), he finds that the short-term interest rate as well as the term spread (the difference between the long-term and short-term interest rates), are generally superior predictors across countries.

In their recent article, Rapach et al. (2013) study the importance of the U.S. market movements in predicting international stock returns. Due to its major role in the world economy, investors are likely to focus on the U.S. markets, potentially creating spillovers of U.S. returns to other markets. The findings of Rapach et al. (2013) do in fact indicate that lagged U.S. returns predict stock returns in several other markets, which they link to the behavioral theory of Hong and Stein (1999) based on the idea of gradual diffusion of information. In this theory, limited investor attention and market participation causes information to diffuse slowly across markets generating cross-predictability between them. In the subsequent research, Hong et al. (2007) examine the gradual diffusion of information in monthly industry portfolio returns, while Menzly and Ozbas (2010) consider industries that have a supplier-customer relationship. Recently, Rizova (2013) have linked information diffusion between markets to international trade flows.

Similarly to Rapach et al. (2013), we examine the interdependencies between excess stock returns in the U.S. and ten other markets. Unlike them, however, we concentrate on the directional component of stock returns, i.e. we are interested in predicting the signs of the returns instead of the actual returns. In the previous finance literature, including the studies mentioned above, a vast amount of research effort has been put into the conventional predictive regression models and their extensions, such as regime switching models, containing various different predictors to examine whether there are statistically and economically significant (in- and out-of-sample) predictive patterns in stock returns (see the survey of Rapach and Zhou (2013)). A closely related and widely examined topic focuses on return and, in particular, volatility transmission and spillover effects between markets (see, e.g., the survey of Gagnon and Karolyi (2006)), where the role of the U.S. as a driver of movements in international stock markets has often been emphasized.

In contrast to these established approaches, the directional predictability of stock returns is, so far, a less covered topic although sign predictability is an important issue in various financial applications. Forecasting the signs of stock returns has often been motivated by its usefulness in market timing decisions (see, e.g., Pesaran and Timmermann (2002)). Already in Merton's (1981) classic market timing model, fund managers are interested in the sign rather than the actual value of the return when determining their asset allocations. A number of more recent empirical studies also highlight the potential usefulness of sign predictability in market timing, by showing that binary response models outperform the usual real-valued predictive regression models in forecasting return signs, by both statistical and economic goodness-of-fit measures (see, e.g., Leung et al. (2000), Nyberg (2011) and Pönkä (2014)).

In addition to the market timing perspective, Christoffersen and Diebold (2006) point out the presence of sign predictability in U.S. equity returns that may also exist in the absence of mean predictability. Their argument is based on the fact that predictable conditional volatility may be useful in forecasting the sign of the return (see also the related findings of Christoffersen et al. (2007) in an international setting and Chevapatrakul (2013) for the U.K.). Nyberg (2011) and Pönkä (2014) show that the return signs are indeed predictable and that there are even more useful predictors than the conditional volatility. In line with this evidence, Anatolyev and Gospodinov (2010) propose a multiplicative model for the U.S. stock returns decomposed into sign and absolute value components. They find economically and statistically significant gains in the forecasting performance of their model over and above conventional predictive regression models (see also Rydberg and Shephard (2003) for a related decomposition model).

Our study contributes to the existing literature on sign predictability in a number of ways. In particular, we examine an international dataset containing 11 industrialized countries, whereas the previous studies have concentrated almost exclusively on the U.S. stock market returns. Leung et al. (2000) consider an international dataset including the U.S., U.K. and Japan, but unlike us, they do not explore international linkages between the markets but concentrate purely on country-specific models. Furthermore, Anatolyev (2009) considers directional cross-predictability of daily returns from three European markets, three Baltic markets, and from two Chinese exchanges in a different multivariate model compared to ours.

A further contribution of our study is the proposal of a new bivariate (two-equation) probit model that facilitates studying the predictive role of the U.S. market for the other markets in a new way. With this model, we can also circumvent problematic econometric issues related to generated regressors (see, e.g., Pagan (1984)). Overall, the previous econometric literature on bivariate and multivariate binary response time series models is very scant. Our model has some similarities with Nyberg (2014) who studies business cycle linkages between the U.S. and Germany, and finds that joint modeling of recession probabilities in these two countries substantially increases predictive power compared to independent univariate models. Our new bivariate model differs from that of Nyberg (2014), as it allows for a contemporaneous predictive effect between the two markets.

Following the previous literature on examining the gradual diffusion of information across markets (see Hong et al. (2007), Menzly and Ozbas (2010) and Rapach et al. (2013)), we use monthly data in this study. Although we emphasize on the role of the U.S., we are not explicitly considering the speed of information diffusion between countries. Much of the relevant information is likely to be diffused more rapidly than in monthly frequency. Instead, we are interested in studying the role of the U.S. economic fundamentals (many of them not available in higher frequencies) in predicting signs of returns in non-U.S. countries.

Our in-sample results based on univariate (single-equation) probit models suggest that, in accordance with Rapach et al. (2013), the lagged excess U.S. stock return is a useful predictor of the sign of excess returns in a number of other markets, supporting the leading role of the U.S. However, our new bivariate model outperforms the univariate models in seven out of ten markets, suggesting that it is not only the lags of U.S. returns that have predictive power. In other words, it is advantageous to utilize the predictive power obtained for the U.S. market movements to predict signs of returns in other markets. Out-of-sample forecasting results generally confirm the in-sample findings: The new bivariate probit model yields the most accurate forecasts in the majority of markets in terms of statistical criteria. When examining the economic value of market timing decisions based on the sign predictions, the simple trading strategies yield higher returns than those based on the univariate probit models and the passive buy-and-hold strategy. This finding in turn complements the previous research on the economic value of volatility timing for short-horizon asset allocation strategies (cf., e.g., Fleming et al. (2001)).

The rest of the paper is organized in the following way. In Section 2, we introduce the econometric framework, i.e. the univariate and bivariate probit models. In Section 3, we describe the goodness-of-fit measures and statistical tests used in evaluating sign predictions. Section 4 introduces the dataset, including the predictive variables. In Sections 5 and 6, we report in-sample and out-of-sample forecasting results, respectively, where in the latter we also study the economic significance of out-of-sample forecasts in trading simulations. Finally, in Section 7 we conclude and discuss possible extensions of this study.

2 Sign Predictability

In the previous finance literature, a vast amount of research effort has been put into the conventional predictive regression model for excess stock returns, containing various different predictors (see, e.g., the survey of Rapach and Zhou (2013)). The directional predictability of excess stock returns is a less covered topic, but it holds potential for further research. As pointed out by Christoffersen and Diebold (2006), sign predictability may exist even in the absence of mean predictability, which can be particularly useful in terms of creating profitable investment strategies.

Throughout this paper, our focus is on the directional component of the excess stock market return. Let us denote a one-month excess market return for market j as $r_{jt} = r^n_{jt} - r^f_{jt}$, where r^n_{jt} is the nominal portfolio return and r^f_{jt} is the risk-free rate. When we use the word 'return' in the remainder of the paper, we refer to the excess stock return as defined here. The excess return can be transformed into binary time series

$$y_{jt} = \mathbf{1}(r_{jt} > \zeta), \tag{1}$$

where $\mathbf{1}(\cdot)$ is the indicator function and ζ is a user-determined constant. Following previous research (see, e.g., Leung et al. (2000), Christoffersen and Diebold (2006), Anatolyev and Gospodinov (2010) and Nyberg (2011)), we consider the leading case $\zeta = 0$, i.e., y_{jt} consists of the signs of the excess returns. Assuming $\zeta = 0$, expression (1) can be rewritten as

$$y_{jt} = \begin{cases} 1, \text{if the excess stock return } r_{jt} \text{ is positive,} \\ 0, \text{otherwise,} \end{cases} \tag{2}$$

for country j.

Let $E_{t-1}(\cdot)$ and $P_{t-1}(\cdot)$ denote the conditional expectation and probability, given information set Ω_{t-1}, respectively. The information set includes all relevant information, such as the past returns and the values of the predictive variables. As y_{jt} conditional on Ω_{t-1} follows a conditional Bernoulli distribution, the conditional probability of a positive excess return p_{jt} can be written as

$$p_{jt} = P_{t-1}(y_{jt} = 1) = E_{t-1}(y_{jt}). \tag{3}$$

The conditional probability of a negative return (i.e. $P_{t-1}(y_{jt} = 0)$) is then the complement probability $1 - p_{jt}$.

In order to study the predictability of the sign of the return y_{jt}, we need to specify a model for the probability of the positive return (3). In the previous literature, this has been carried out by examining univariate (single-equation) binary response models with different predictive variables. Hence, these models are briefly described next in Section 2.1 before turning to our main econometric contribution related to bivariate (two-equation) probit models (Section (2.2)).

2.1 Univariate Probit Model

Univariate binary response models, such as logit and probit models, have previously been used to examine the sign predictability of excess stock returns. Leung et al. (2000) find that classification-based models, including binary response models, outperform traditional predictive regressions in forecasting the direction of stock markets in terms of statistical goodness-of-fit tests and profitability of investment strategies built on their forecasts. Their study covers the U.S., U.K., and Japanese stock markets. Nyberg (2011) uses dynamic probit models to predict the direction of monthly U.S. excess returns and finds evidence in favor of sign predictability. Moreover, in line with Leung et al. (2000), his probit models yield superior forecasts over traditional predictive regressions. Pönkä (2014) examines the directional predictability of excess U.S. stock market returns by lagged excess returns on industry portfolios using dynamic probit models, and finds that a number of industries lead the stock market and that binary response models outperform conventional predictive regressions in forecasting the direction of the market return.

To determine the conditional probability of a positive excess stock return for country j (see (3)), a univariate probit model is specified as

$$p_{jt} = P_{t-1}(y_{jt} = 1) = \Phi(\pi_{jt}), \tag{4}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution and π_{jt} is a linear function of the variables in Ω_{t-1}. The most commonly used specification is the following

$$\pi_{jt} = \omega_j + \boldsymbol{x}'_{j,t-1}\boldsymbol{\beta}_j, \tag{5}$$

where $\boldsymbol{\beta}_j$ is the coefficient vector of the lagged predictive variables included in the vector $\boldsymbol{x}_{j,t-1}$ and ω_j is a constant term for country j. In the subsequent analysis, we also consider dynamic models where the lagged returns ($r_{j,t-1}$) and the lagged values of binary return indicators (1) are included in $\boldsymbol{x}_{j,t-1}$. The parameters of these models can be estimated using the method of maximum likelihood (ML). For more details on ML estimation and the computation of Newey-West type robust standard errors (see Newey and West (1987)), we refer to Kauppi and Saikkonen (2008) and de Jong and Woutersen (2011).

Overall, the presence of sign predictability culminates to whether we can find predictors in (5) that contain statistically significant predictive power over and above the constant term ω_j. In this study, we consider the same international dataset as Rapach et al. (2013), including various financial and macroeconomic predictive variables for a number of markets. The dataset will be described more detail in Section 4. As Rapach et al. (2013) do with conventional predictive regression models for the returns r_{jt}, we can examine the predictive power of the U.S. stock market for the signs of the excess returns in the other markets by including the lagged U.S. return in the univariate probit model along with the other (domestic) predictors in $\boldsymbol{x}_{j,t-1}$.

Overall, due to high integration of the stock markets around the world, the excess returns and their signs are rather highly correlated between different countries (see the descriptive statistics in Section 4). Thus, it seems highly reasonable to consider the joint modeling of the direction of returns, which may well result in superior forecasts compared with country specific univariate models. Based on the results of Rapach et al. (2013), it is particularly interesting to include the U.S. market in such models. However, instead of the lagged return, we could consider the effect of the U.S. probability forecast for the positive stock return on the other markets, hence conditioning on a larger information set. This issue can be considered in a meaningful way with bivariate probit models, described in the following section.

2.2 Bivariate Probit Model

The main interest in this paper is on bivariate binary response models, where we examine pairwise directional predictability of stock returns in two markets. This will, in particular, allow us to consider the effect of the U.S. stock market to international markets concentrating on the directional component of the stock returns.

Let us now consider the random vector (y_{1t}, y_{2t}) containing the binary time series of the signs of the excess stock returns (2) in two markets of interest. Conditional on the information set Ω_{t-1}, the vector (y_{1t}, y_{2t}) follows a bivariate Bernoulli distribution,

$$(y_{1t}, y_{2t})|\Omega_{t-1} \sim B_2(p_{11,t}, p_{10,t}, p_{01,t}, p_{00,t}), \tag{6}$$

where the conditional probabilities of the different outcomes are

$$p_{kl,t} = P_{t-1}(y_{1t} = k, y_{2t} = l), \;\; k, l = 0, 1,$$

and they sum up to unity

$$p_{11,t} + p_{10,t} + p_{01,t} + p_{00,t} = 1.$$

Following the bivariate probit model originally proposed by Ashford and Sowden (1970) (see also, Greene (2012), 778–781), we assume the joint probabilities of the different outcomes of (y_{1t}, y_{2t}) to be determined as

$$\begin{aligned}
p_{11,t} &= P_{t-1}(y_{1t} = 1, y_{2t} = 1) = \Phi_2(\pi_{1t}, \pi_{2t}, \rho), \\
p_{10,t} &= P_{t-1}(y_{1t} = 1, y_{2t} = 0) = \Phi_2(\pi_{1t}, -\pi_{2t}, -\rho) \\
p_{00,t} &= P_{t-1}(y_{1t} = 0, y_{2t} = 0) = \Phi_2(-\pi_{1t}, -\pi_{2t}, \rho) \\
p_{01,t} &= P_{t-1}(y_{1t} = 0, y_{2t} = 1) = \Phi_2(-\pi_{1t}, \pi_{2t}, -\rho),
\end{aligned} \tag{7}$$

where $\Phi_2(\cdot)$ is the cumulative density function of the bivariate standard normal distribution with zero means, unit variances and correlation coefficient ρ, $|\rho| < 1$. Furthermore, similarly as in (5), π_{jt}, $j = 1, 2$, are assumed to be linear functions of the lagged stock returns (and their signs) and the other predictive variables included in the information set at time $t - 1$. The conditional probabilities of positive excess returns for markets $j = 1, 2$ are the marginal probabilities of the outcomes $y_{1t} = 1$ and $y_{2t} = 1$ equal to (cf. (4))

$$p_{1t} = P_{t-1}(y_{1t} = 1) = p_{11,t} + p_{10,t}, \tag{8}$$

and

$$p_{2t} = P_{t-1}(y_{2t} = 1) = p_{11,t} + p_{01,t}. \tag{9}$$

To complete the bivariate probit model, we need to determine the linear functions π_{jt}, $j = 1, 2$ (i.e. the dependence structures on the available predictive information). In the simplest case, introduced by Ashford and Sowden (1970), similar to univariate model (5),

$$\begin{bmatrix} \pi_{1t} \\ \pi_{2t} \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{x}'_{1,t-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{x}'_{2,t-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}, \tag{10}$$

where ω_1 and ω_2 are constant terms and $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are the coefficient vectors of the lagged predictive variables included in the vectors $\boldsymbol{x}_{1,t-1}$ and $\boldsymbol{x}_{2,t-1}$, respectively. In model (10), the explanatory variables have an immediate effect on the conditional probabilities (7) which, given the value of the correlation coefficient ρ, do not change unless the values of the explanatory variables change.

In this study, we are interested in the information transmission between stock markets in different countries and, especially, the possible leading role of the United States. Rizova (2013) point out that as the larger stock markets are more widely followed by investors, the cross-predictability caused by the gradual diffusion of information in other markets is likely to be weaker for the major markets. Although Rapach et al. (2013) find evidence that lagged U.S. returns significantly predict returns in nine out of ten countries in their study, it is likely that there are differences between the predictive role of the U.S. due to, e.g., the amount of investor attention and the relative importance of the U.S. as a trading partner. The literature on the influence of the U.S. on international markets via volatility spillovers across markets has also pointed out the leading role of the U.S. (see, e.g., the survey of Gagnon and Karolyi (2006)).

Hereafter the U.S. is the first country (i.e. $j = 1$) in model (10). Then, following Rapach et al. (2013), we include the lagged U.S. return in the vector $\boldsymbol{x}_{2,t-1}$ for the second country to examine whether the U.S. return predicts the sign of return in the other markets ($j = 2$). An alternative and more general approach that we consider is to allow the linear function π_{1t} related to the probability of the positive excess return to have an effect on π_{2t}. Specifically, we consider the following extension of model (10):

$$\begin{bmatrix} 1 & 0 \\ -c & 1 \end{bmatrix} \begin{bmatrix} \pi_{1t} \\ \pi_{2t} \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{x}'_{1,t-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{x}'_{2,t-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}, \tag{11}$$

where the coefficient c measures the contemporaneous effect from π_{1t} to π_{2t}. In the context of our application, this means that we study the effect of the U.S. on the other markets.[1] Note that although in (11) π_{1t} has a contemporaneous effect on π_{2t}, the predictive information in π_{1t} is actually coming from the lagged predictors in $\boldsymbol{x}_{1,t-1}$. In other words, the lagged U.S. return is not included as a predictor in $\boldsymbol{x}_{2,t-1}$, but it has only an indirect effect on π_{2t} via the coefficient c.

The linear function π_{2t} does not contemporaneously help to predict the sign of the return in market 1 (in the U.S.), while there is contemporaneous predictability in the opposite direction, when $c \neq 0$. That is, when $c \neq 0$, the predictive power obtained for the U.S. market is helpful in predicting the signs of the returns in other markets, but not vice versa.[2] Due to the nonlinear nature of model (11), we can also statistically check this identification assumption by comparing the log-likelihoods of two models where the matrix on the left hand side of (11) containing the contemporaneous linkage should be lower or upper-diagonal (when the ordering of the markets is given fixed).

In addition to the effect through π_{2t}, the lagged U.S. excess return may have an indirect effect on predictive power through the correlation coefficient ρ. The interpretation of the correlation coefficient is, however, somewhat complicated as it is related to the bivariate normal distribution used to obtain the response probabilities (7), based on the linear functions π_{jt}. Furthermore, as in Nyberg (2014), it turns out in our empirical analysis that the effect of ρ on the sign probability forecasts (7) is minor, although statistically significant. It is also worth noting that if $\rho = c = 0$, the bivariate model reduces to two univariate probit models without linkages between the markets.

In Appendix A, we will give details on the maximum likelihood estimation of the new bivariate probit model introduced above. In particular, we derive the formulae for the misspecification-robust standard errors of the bivariate probit model (11) to take the potential misspecification of the model into account when

[1]It is noteworthy that (11) bears resemblance to the structural vector autoregressive (SVAR) models commonly used in empirical macroeconomics and finance.

[2] To identify model (11), as long as $c \neq 0$, the predictive variables (and their lags) in $\boldsymbol{x}_{1,t-1}$ and $\boldsymbol{x}_{2,t-1}$ cannot be the same. This is not a problem in our application, because we use only domestic predictors for each country.

interpreting the estimation results. An important advantage of the joint model (11) is that it circumvents the well-known generated regressor problem (see, e.g., Pagan (1984)), as the effect of π_{1t} on π_{2t} is conveniently estimated within one model.

3 Goodness-of-Fit Measurement and Sign Predictability

We will employ a number of alternative measures to evaluate the in-sample and out-of-sample predictive performance of the models. The need for different types of measures to uncover different aspects of forecast quality of probability forecasts of binary outcomes is discussed in detail by Lahiri and Wang (2013). In addition, we need to modify some measures to suit our bivariate model. We also use some methods to evaluate directional predictability that have previously not been applied to sign forecasts of stock returns.

Following the usual practice in finance, one of our measures is a counterpart of the coefficient of determination (R^2) designed for binary response models. Estrella (1998) defined the pseudo-R^2 (for univariate models) as

$$psR^2 = 1 - \Big(\frac{logL_u}{logL_c}\Big)^{-(2/T)logL_c}, \tag{12}$$

where $logL_u$ and $logL_c$ are the maximum values of the constrained and unconstrained log-likelihood functions respectively, and T is the length of the time series. This measure takes on values between 0 and 1, and can be interpreted in the same way as the coefficient of determination in the usual linear predictive regression models. In Section 5, we also report its adjusted form (see Estrella (1998)) that takes into account the trade-off between the improvement in model fit and the number of estimated parameters.

Due to the form of (12), there is a linkage between to the pseudo-R^2 and the corresponding likelihood ratio test statistic testing the null hypothesis that the included predictive variables do not have predictive power. In other words, under the null hypothesis, the value of the log-likelihood function ($logL_c$) is obtained when only a constant term is included in the model. Hence, (12) measures the predictive power obtained with the predictors included in $\boldsymbol{x}_{j,t-1}$. In the bivariate probit model a nonzero correlation coefficient ρ poses a complication to this interpretation, as its nonzero value implies predictive power not accounted for by the predictors. Therefore, for bivariate models with $\rho \neq 0$, we propose a modification to (12)

$$psR^2_\rho = 1 - \Big(\frac{logL_u}{logL_c^\rho}\Big)^{-(2/T)logL_c^\rho}, \tag{13}$$

where $logL_c^\rho$ denotes the value of the restricted log-likelihood function of the bivariate probit model where $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \mathbf{0}$ (and $c = 0$ in model (11)). In other words, similarly as (12), expression (13) measures the predictive power of explanatory variables, but as the expressions (12) and (13) differ, they are not comparable.

The problems with the pseudo-R^2 statistics mean that we will also need to use some other statistics that allow us to do make comparisons between different univariate and bivariate probit models. Together with the pseudo-R^2, the Quadratic Probability Score

$$QPS = \frac{1}{T}\sum_{t=1}^{T} 2(y_{jt} - p_{jt})^2 \tag{14}$$

is also commonly used to evaluate probability forecasts, and it can be seen as a mean square error type of statistic for binary dependent variable models. The value of the QPS ranges between 0 and 2, with score 0 indicating perfect accuracy.

As previously, e.g., in Nyberg (2011) and Pönkä (2014), we also report the success ratio (SR), which is simply defined as the percentage of correct signal sign forecasts. A signal forecast for the sign of the return y_{jt} can be written as

$$\hat{y}_{jt} = \mathbf{1}(p_{jt} > \xi), \quad j = 1, 2, \tag{15}$$

where p_{jt} is the conditional probability of a positive excess return implied by a univariate or bivariate probit model. If p_{jt} is higher than the threshold ξ, the signal forecast $\hat{y}_{jt} = 1$ (i.e. positive excess return), while $\hat{y}_{jt} = 0$ if $p_{jt} \leq \xi$. This measure is useful in evaluating out-of-sample forecasts, but it can also be used in in-sample evaluation.

An unfortunate feature of the success ratio is that its effectiveness depends on the predefined probability threshold ξ. In this paper, following previous research, we report the success ratios implied by $\xi = 0.5$. This is also in line with the symmetric selection $\zeta = 0$ in (1) that the signal forecast (15) is the likeliest outcome (i.e. positive or negative return). Related to the SR statistic, Pesaran and Timmermann (1992) have proposed a statistical test of directional predictive accuracy that measures the distance of the value of SR from the success ratio obtained when the realized values y_{jt} and the forecasts $\hat{y}_{jt}$ are independent (see also Granger and Pesaran (2000)). As an extension of the Pesaran-Timmerman (1992) test statistic allowing for serial correlation in y_{kt}, Pesaran and Timmermann (2009) have suggested another predictability test (denoted PT). The traditional test neglecting autocorrelation suffer from the danger of finding spurious timing ability (see Chu et al. (2009)) and, therefore, we prefer the PT test.

Although $\xi = 0.5$ is a commonly used natural threshold in (15), it is not an innocent selection. It turns out that success ratios and market timing tests are rather highly dependent on threshold selection. Therefore, it is reasonable to look at an alternative approach to assess the accuracy of probability forecasts, namely the Receiver Operating Characteristic (ROC) curve. ROC analysis has long been used as a goodness-of-fit measure of classification accuracy in medical applications and biostatistics, but it has also recently been used in a small but growing number of economic applications (see, e.g., Berge and Jorda (2011); Lahiri and Wang (2013); Christiansen et al. (2014)). Following the idea of signal forecasts (15), we can define two widely used measures of classification accuracy, namely the true positive rate (TP) and the false positive rate (FP):

$$TP(\xi) = P_{t-1}(p_{jt} > \xi | y_{jt} = 1) \tag{16}$$

and

$$FP(\xi) = P_{t-1}(p_{jt} > \xi | y_{jt} = 0), \tag{17}$$

for any threshold $0 \leq \xi \leq 1$. The ROC curve is a mapping of the true positive rate (16) and the false positive rate (17) for all possible thresholds ξ described as an increasing function in the $[0, 1] \times [0, 1]$ space, with $TP(\xi)$ plotted on the Y-axis and $FP(\xi)$ on the X-axis. A ROC curve above the 45-degree line indicates forecast accuracy superior to a coin toss, whereas curves below it are considered 'perverse' forecasts for which the optimal response is exactly the opposite of what the forecast suggests.

In our application, it is reasonable to think that different agents (investors) have their own risk profiles which can be interpreted in our framework as different selections of ξ. In other words, one (risk-averse) investor may require a higher probability of a positive return than another. The optimal threshold may also be time-varying, complicating our analysis further. As there obviously is no clear rule or reason to use a specific threshold, the ROC curve seems useful in assessing overall predictive ability of a given model.

The area under the ROC curve (AUC) is a convenient measure to summarize the predictive information contained in the ROC curve. The AUC is defined as the integral of the ROC curve between zero and one. Therefore, the AUC also gets values between 0 and 1, with the value of 0.5 corresponding a coin toss and the value 1 to perfect forecasts. The value of the AUC as such describes the overall level of sign predictability: A value of AUC above 0.5 indicates statistical predictability, i.e. successful market timing ability (with potential economic gains).

In a similar fashion as in the case of the traditional Pesaran-Timmermann market timing tests described above, it is reasonable to ask whether we can statistically test if the empirical AUC is greater than 0.5, implying sign predictability. However, in contrast to the testing procedures employed for cross-sectional datasets, corresponding reliable test in the time series setting does not exist yet (see the discussion in Hsu and Lieli (2014)). Thus, we use the AUC only as a device to summarize the goodness-of-fit and forecast performance of the probit models throughout this paper.

The complication of testing the statistical significance of the AUC at the moment partly points to the need for asset allocation experiments to examine the economic value of our sign forecasts. It is also rather common that forecasting results deemed statistically insignificant by statistical measures are still economically significant (see, e.g., Leitch and Tanner (1991) and Cenesizoglu and Timmermann (2012)), which also highlights the need for market timing tests. In Section 6.2, we thus motivate our forecasting models also in the trading strategy point of view to study their economic performance in more detail.

4 Data and Descriptive Statistics

In finance, a large number of potential predictors of excess stock returns have been considered in the linear predictive regression context (see the survey of Rapach and Zhou (2013) and the references therein). Typically very little out-of-sample predictive power is found, if any (see Goyal and Welch (2008) and Campbell and Thompson (2008)). In contrast to the usual predictive models, the previous research on (out-of-sample) sign predictability is rather scant and, to the best our knowledge, so far only Leung et al. (2000) and Anatolyev (2009) have examined international datasets (containing only a few countries).

By traditional predictive regressions, Ang and Bekaert (2007) studied stock return predictability in an international setting by three commonly used predictors; the short term interest rate, the dividend yield, and the earnings yield. Rapach et al. (2013) examined the effect of the U.S. stock market on international markets by including the lagged U.S. return as a predictor in linear regression models. In our analysis, we consider the same international dataset as Rapach et al. (2013)[3], which facilitates examining to what extent potential differences in results can be attributed to different forecasting methodologies. Rapach et al. (2013) examined the results of traditional predictive regression for excess stock returns, while in this paper we concentrate on sign predictability. The monthly dataset includes Australia (AUS), Canada (CAN), France (FRA), Germany (GER), Italy (ITA), Japan (JPN), the Netherlands (NED), Sweden (SWE), Switzerland (SUI), the United Kingdom (U.K.), and the United States (U.S.). The sample period ranges from February 1980 to December 2010.

In the dataset, the excess stock market returns (denoted by RM) are return indices that take dividends into account. These returns are transformed to binary return series (RMI) as in (1). In line with Rapach et al. (2013), our predictive variables include the three-month short-term interest rate (TB) and dividend yield (DY) for each market. We also consider additional predictive variables that Rapach et al. (2013) only used in their robustness checks. These variables include CPI inflation (INF), term spread (TS), the ten-year government bond yield ($10Y$), as well as the growth rates in the real exchange rate (REX), real oil price (OIL), and industrial production (IP).

The lagged values of RM and RMI are also included in the set of potential predictive variables. This allows us to study the relative usefulness of the actual lagged excess return RM and its sign component RMI. The use of the lagged RMI as a predictor has previously been considered previously by Anatolyev and Gospodinov (2010), Nyberg (2011) and Pönkä (2014) for U.S. data in different dynamic probit models.

As we concentrate on bivariate models with the U.S. market always as one of the two markets, it is useful to take a look at the correlations between the excess returns in U.S. and the other markets. These statistics along with the means and standard deviations of the excess returns are presented in Table 1. We also report the counterpart of the correlation coefficient for the signs of the excess returns, the so-called Φ-coefficient (mean square contingency coefficient) that is simply the Pearson correlation coefficient of two binary time series. The correlations of the returns with the U.S. return range from 0.417 (Italy) to 0.774 (Canada) for different countries. The correlation coefficients for the return signs are generally very similar although the values are lower than the correlations, due to properties

[3] We would like to thank the authors of Rapach et al. (2013) for making the dataset available at David Rapach's website: sites.slu.edu/rapachde/home/research.

of the statistic. The mean excess returns vary considerably across the markets; the lowest monthly average return is reported for Japan (0.219%) and the highest for Sweden (1.027%). The standard deviations range from 4.500 in the U.S. to 6.983 in Italy.

5 In-Sample Results

Before considering the out-of-sample predictive power of different models and predictive variables in Section 6, we first examine their in-sample performance in the full sample period from 1980 to 2010.[4] Following the typical convention in the previous similar studies, we consider only the one-month-ahead forecast horizon ($h = 1$) and the first lags of the predictors throughout in this section and Section 6.

In Section 5.1, we consider univariate models with the aim to find the best predictors in sample. In the same spirit as Rapach et al. (2013), in Section 5.2 we examine the potential predictive gains of including the lagged U.S. excess return in the model. In Section 5.3, we consider the bivariate probit models, introduced in Section 2.2, that facilitate examining the linkages between the U.S. and other markets in more detail.

5.1 Univariate Models

In this section, we study the predictive power of a number of domestic variables for the direction of the excess stock return separately in each of the eleven markets in the univariate probit model defined in (4) and (5). We initially consider models with the same two predictors, the dividend yield (DY) and the three-month T-bill (TB) rate, as Ang and Bekaert (2007) and Rapach et al. (2013) included in their main models. The results for these baseline models are presented in Table 2.

It turns out that DY and TB are statistically significant predictors of the direction of the U.S. return. The adjusted pseudo-R^2 equals 0.016, which is in line with a modest level of predictability typically found in previous studies. As far as the overall predictive power in the other markets is concerned, the results are rather similar for Canada and the Netherlands, although in the latter case the dividend yield is not statistically significant. However, for most of the other markets, these

[4] We also assessed the robustness of these results using a shorter in-sample period up to 1994M12, which is the endpoint before out-of-sample forecasting starts (see Section 6). The results turned out to be essentially similar as those in Sections 5.1–5.3.

two-predictor models have little or no predictive power, as the negative values of the adjusted pseudo-R^2 among other measures indicate.

The results on sign predictability presented in Table 2 are generally in line with those of Rapach et al. (2013) based on traditional, linear predictive regressions. In particular, the dividend yield does not seem to be a powerful predictor of stock returns in an international context. Similar findings have also been reported by Hjalmarsson (2010) who finds that while interest rate variables are rather robust predictors of stock returns in developed markets, the dividend-price ratio has very limited predictive ability in various international stock markets. The short-term interest rate has somewhat higher predictive power, and its negative estimated coefficient implies that higher interest rates decrease the probability of positive stock return.

Due to the relatively weak predictive power of dividend yield (DY) and short-term interest rate (TB) considered above, we next try to find the best predictors for each market by performing a model selection procedure that involves all the domestic variables in our dataset. We use the Akaike information criterion (AIC) as the model selection criterion. We first include each predictor separately in the model and select the one that minimizes the value of the AIC. In the next stage, we estimate all possible two-predictor models containing the selected variable and again choose the model with the smallest AIC. We continue this process sequentially and once we have found the optimal model (having the smallest value of the AIC), we perform some sensitivity checks against other model specifications (with the same number of predictors or less) to ensure robustness of our model selection procedure.

The selected univariate probit models for the different markets are presented in Table 3. For example, in the U.S. case the selected model contains five predictors, whereas for the other markets a model with fewer variables is typically selected (only one predictor for Australia and Japan). Also the model fit, measured by the adjusted pseudo-R^2, is higher for the U.S. than for the other countries (except for Switzerland). A similar pattern can also be seen in the QPS and SR statistics. In general, we obtain improvement in predictive power by allowing for a larger set of predictors compared with the case of including only TB and DY (see Table 2). The lagged domestic stock return (RM) and the real oil price (OIL) are the most commonly selected predictors. Interestingly, in line with the findings of Nyberg (2011), the lagged return (RM) is generally superior to the lagged sign of the return (RMI). Overall, the values of the adjusted pseudo-R^2 still remain rather modest, demonstrating statistically weak predictability, as is typical of predictive models for stock returns in general.

In Tables 2 and 3, we also present results for the Pesaran and Timmermann (2009) market timing test statistic (PT). In Table 2, we find a statistically significant value of the PT statistic in only two out of the eleven models. We also find that the values of the PT statistic are not all that well in line with the success ratio (SR); for example, for the case of Japan the PT statistic is statistically significant at the 10% level, while the success ratio is only as low as 0.524. It is also worth noting that the PT statistic for the U.K. is not applicable, because the model yields only positive signal forecasts ($\hat{y}_{jt} = 1$), i.e. the estimated probability of positive return is higher than 50% all the time). This finding highlights the need for other measures, such as the AUC, that is not dependent on only one specific threshold selection, which is $\xi = 0.5$ for the PT statistic and success ratio.[5] All in all, the results of the PT statistics are in line with other measures by generally indicating a higher level of predictability for the models presented in Table 3 than in Table 2.

Due to the difficulties with the success ratio and the PT test, we emphasize the AUC in describing the predictive ability of the probit models. The reported AUCs also lend support to including a wider selection of domestic predictive variables. In Table 2, the AUC values range from 0.524 for Japan to 0.589 for the Netherlands for the models which we contain the domestic dividend yield and the three-month interest rates as predictors. For the models in Table 3, the AUCs are actually higher for all the countries than in the previous case, and lie between 0.576 for Japan and 0.651 for Switzerland. This can be seen as further evidence in favor of going beyond the dividend yield and short-term interest rate as predictors when predicting the signs of the excess stock returns.

5.2 Univariate Models with the Lagged U.S. Return as a Predictor

As we are especially interested in the possible leading role of the U.S. in international stock markets, we next study univariate models presented in Table 3 augmented with the lagged U.S. excess return ($RM_{U.S.,t-1}$).[6] The results of these

[5]We report the results for the natural threshold of $\xi = 0.5$ in the tables, but we also experimented with alternative thresholds, which led to only minor changes compared to the results presented here.

[6] Hereafter we denote the market in the subscript.

models are reported in Table 4. For three out of ten markets, the lagged U.S. return is statistically significant (at least) at the 10% level, indicating improvement in predictive power. Interestingly, when we compare the AUC values between the univariate models in Tables 3 and 4, we find improvement in seven out of ten cases upon including $RM_{U.S.,t-1}$ in the model. In some cases the improvement is rather modest, but this finding is generally reconfirmed also by the adjusted pseudo-R^2, QPS, and the SR.[7]

Overall, our findings in the univariate probit models are in line with those of Rapach et al. (2013) for traditional linear predictive models. The lagged value of the U.S. excess return seems to contain useful additional predictive power to predict return directions internationally. However, in contrast to the results reported by Rapach et al. (2013), we have shown that the dividend yield and the lagged three-month interest rate are not the best predictors of the sign of the excess return in most of the markets considered. Instead, the lagged domestic excess stock return and the change in the real oil price are typically among the best predictors in sample.

5.3 Bivariate Models

In the previous section, we found that including the lagged U.S. return in the univariate models (marginally) improves the in-sample fit in some of the markets. To further explore the information diffusion from the U.S., in this section, we estimate bivariate probit models for the U.S. and the ten other markets. In particular, we want to examine whether including the combination of the U.S. predictors (i.e. π_{1t} in model (11)) can produce more accurate predictions for other markets over and above including the lagged U.S. return only.

In this section, we consider four different bivariate probit models. The most general model (Model 4) defined in equations (7) and (11) is based on the new bivariate model allowing for the contemporaneous predictive linkage from the U.S. to the other market. The examined models contain the following restrictions:

[7] As our aim is to test the predictive ability of the lagged U.S. return, we do not present detailed results on how returns in other markets help predict the sign of the U.S. return. However, we found that when we augment the model for the U.S. (see Table 3) with the lagged returns from each individual country separately, only the lagged Swedish and Italian returns turn out to be statistically significant predictors of the U.S. return. The finding that the foreign lagged returns do not predict the U.S. return sign is in line with the results of Rapach et al. (2013) obtained with the conventional predictive regression models for the actual return.

Model 1: $c = 0$, $\rho = 0$,
Model 2: $c = 0$,
Model 3: $\rho = 0$,
Model 4: unrestricted.

Model 1 is the most restricted version of the general bivariate model (Model 4), and it reduces to two univariate probit models considered already in Sections 5.1 and 5.2. Model 2 restricts c to zero, leaving out the contemporaneous linkage from the U.S. to the other market; nevertheless the correlation coefficient ρ still has an effect on the response probabilities (7). In Model 3, we restrict ρ to zero, but allow for the contemporaneous effect through c.

In Section 5.1, we found that the fit of the univariate models is rather weak when including only DY and TB as predictors. Hence, instead of relying on these variables, we select the predictors for each market separately. The selection of predictors for Model 1 is straightforward, as no contemporaneous effects are allowed for between the two markets. Thus, we simply rely on the predictors selected for the univariate models in Table 3. However, it is not evident that these variables would be selected for Models 2–4. For example, when the parameter c is not restricted to zero, the joint effect of the U.S. predictors might affect the selection of explanatory variables for the other country. Nevertheless, for the sake of comparability, we have chosen to include the same explanatory variables as in Model 1 in all the models.

The main question of this study is whether there are benefits of using the bivariate model setup and, in particular, whether we can find evidence of gradual diffusion of predictive information from the U.S. to the other markets. We could follow the approach of Rapach et al. (2013), in which foreign (U.S.) variables are directly included in the univariate model for other markets. This is a good way to study the explanatory power of additional variables, such as the lagged U.S. market return, but our approach provides a more parsimonious way of studying the combined effect based on the constructed linear function π_{1t} behind the probability of positive excess return in the U.S. The results for the univariate models in which the lagged U.S. market return is included directly as a predictor for the direction of excess return in the other markets (see Table 4) can be compared with findings presented in this section.

As we have ten pairs of markets, we will not discuss the results for every pair in detail. Instead, we concentrate on three dissimilar cases that give a general overview of our results, and summarize the rest of the findings. The countries we focus on are the U.K., Canada, and Sweden. In addition to a few system-wide measures, we report goodness-of-fit statistics for the markets separately, as this allows us to compare the results with those of the univariate models and to evaluate the predictive power coming from the U.S. to the market of interest.

The results of the bivariate models for the pair of the U.S. and the U.K. are reported in Table 5. We first consider the case where we have two independent univariate probit models. This allows us to later compare the potential benefits of joint modeling of the markets. The parameter estimates of the bivariate independent probit model (Model 1) in Table 5 are reproduced from Table 3. The value of the log-likelihood function is the sum of those of the two univariate models (see Appendix A). Furthermore, as discussed in Section 3, we cannot directly compare pseudo-R^2s between different models because the benchmark model (i.e. restricted log-likelihood function) is different. In other words, the pseudo-R^2 measures for Model 2 and Model 4 (see (13)) are not directly comparable to those for Models 1 and 3 (see (12)). Similar argument applies also comparisons to the univariate probit models reported in Tables 2–4. Thus, we rely on other measures, mainly the AUC and the success ratio in comparing the different models.

[Table 5 here]

For $RMI_{U.S.}$, Models 1 and 2 in Table 5 (i.e. the models including the effect of a nonzero ρ) yield rather similar results, whereas for $RMI_{U.K.}$ the estimated parameter coefficients generally lose some of their statistical significance in Model 2. The parameter ρ is statistically highly significant, which suggest that there are some benefits of joint modeling, but on the other hand we find little or no improvement in predictive power measured by the success ratio and AUC.

With Model 3 (i.e. allowing for a nonzero parameter c) we do find that the adjusted pseudo-R^2 and the AUC clearly favor it over the independent model (Model 1). The success ratio and AUC are also higher for Model 3 than for Model 2. The estimated value of c is positive, as expected, but interestingly statistically insignificant at the 5% level even though the above-mentioned goodness-of-fit measures clearly demonstrate benefits when allowing for a contemporaneous predictive relationship from the U.S. to the U.K. stock market.[8]

[8] We also consider the possibility of effects running from other markets to the U.S., meaning that we reverse the order of the countries in Models 3 and 4. The estimated values of the log-

Overall, Model 3 appears the best according to the AUC and SR in spite of the statistically insignificant coefficient for parameter c. The results of the unrestricted bivariate model (Model 4) indicate that there is little or no benefit of allowing for both nonzero c and ρ compared with Model 3 in terms of the predictability of $RMI_{U.K.}$.

[Table 6 here]

In Table 6, we report the findings for the bivariate system of the U.S. and Sweden. The small Swedish markets are more likely to be affected by events in larger markets. The results indicate that the predictability of the direction of the Swedish markets is indeed improved by modeling it together with the U.S. market. In particular, the AUCs implied by Models 3 and 4 are greater than that implied by Model 1. Also, the parameter c (expressing the linkage between the markets) in Model 3 turns out statistically significant at the 10% level, and the improvement compared with Models 1 and 2 is evident in terms of all goodness-of-fit measures. This can be interpreted as clear evidence of gradual diffusion of information from the U.S. to the Swedish markets, which could indicate that the small Swedish markets that receive less investor attention are prone to be affected by the changes in larger markets.

[Table 7 here]

In Table 7, we present the results of the bivariate models for the U.S. and Canada. Interestingly, the transmission of stock returns and volatility between the U.S. and Canada has previously been studied by e.g. Karolyi (1995), but this is the first study focusing on the cross-predictability of the directional component of the returns. It is perhaps not that surprising that we also find a predictive effect from the U.S. to the Canadian market, as Canada is a relatively small economy with strong ties to its neighbor. We find c highly statistically significant in Model 3 and, in fact, it remains statistically significant for Canada also in Model 4, while for the other markets considered that is not the case. The differences in the AUCs are also rather large compared to the specifications where c is restricted to zero. Figure 1 illustrates the superior in-sample predictive ability presented in Table 7: The ROC curve of Model 3 is almost exclusively above the ROC curve of Model 1, implying thus also higher AUC. Both ROC curves are also above the 45-degree line implying useful predictive power.

---

likelihood functions show that the U.K. market hardly affects the U.S. market through parameter c. Similar findings hold for the other markets as well.

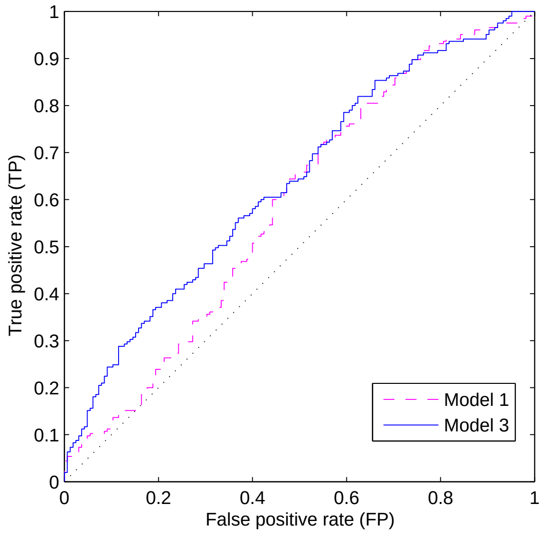

Figure 1: ROC curves of Models 1 and 3 for the Canadian stock return (see Table 7) .

For the remaining markets, the results are reported in Appendix B (see Tables 10–16). For five out of the ten markets we find c statistically significant at least at the 10% level in Model 3. It is worth noting that the domestic variables for both Australia and Japan turned out not rather poor predictors, and only a single domestic variable ended up in the models selected by the AIC. It is also noteworthy that for the large European markets, i.e. Germany, France, and the U.K., we find no significant effects through U.S. predictors. Also for two smaller European countries, namely the Netherlands and Switzerland, we have similar findings. Moreover, allowing ρ to be nonzero generally weakens the statistical significance of the other parameters, whereas it in itself is highly statistically significant in models for each of the markets. However, in line with Nyberg (2014), the statistical significance of ρ does not imply an improvement in overall predictability measured by, e.g., the AUC.

As a general finding among the bivariate models, the Model 3 specification is strongly supported by the AUC. In seven out of the ten markets, the AUC is highest for Model 3, and the independent model (Model 1) is preferred for the German and Swiss markets. Only for Italy, does Model 4 yield the highest AUC. As our findings indicate that Model 3 is generally the best bivariate model, in the following sections we will mostly focus on it.[9]

[9]Reversing the order of the equations in Model 3, i.e., allowing for predictive effects from each of the other markets on the direction of the U.S. return, we find the parameter c significant (at

Altogether, the results for the univariate models (Table 4) indicated that the lagged U.S. return ($RM_{U.S.,t-1}$) is a statistically significant predictor in only three out of the ten markets. In this section, we have found the parameter c statistically significant in five out of the ten bivariate models when ρ is restricted to zero (Model 3). This does not yet imply very strong support to the superiority of the bivariate models, but according to the AUC statistics the bivariate model (Model 3) outperforms the univariate models in Table 4 for eight out of the ten markets, with Australia and Switzerland being the only exceptions. The success ratio favors the bivariate model (Model 3) in seven out of the ten cases over the univariate model. Putting together all of this evidence we get relatively strong indication that the bivariate modeling is competitive in sample and, especially, Model 3 is found to work the best. In order to confirm these findings, we will examine the out-of-sample forecasting performance of these models in the following section.

6 Out-of-Sample Forecasting Results

It is a typical convention in time series forecasting to examine out-of-sample predictive performance, as the in-sample findings do not often hold out of sample. In particular, the commonly used in-sample goodness-of-fit measures are prone to favor overparametrized models, whereas in out-of-sample forecasting more parsimonious models often outperform more complicated ones. In Section 5.3, we found that the bivariate Model 3 (where $c \neq 0$ and $\rho = 0$) performed best. Thus, we will compare the out-of-sample performance of this model with that of the univariate model (i.e. Model 1) reported in Sections 5.1 and 5.2.

In line with the in-sample results, we consider one-month-ahead forecasts ($h = 1$) throughout this section for the forecasting period 1995M1–2010M12. Forecast performance is evaluated by means of statistical measures (Section 6.1) as well as simple asset allocation trading strategies to assess the economic value of the forecasts (Section 6.2). The forecasts are computed following a rolling window approach, where the estimation window is 15 years (i.e. 1980M01–1994M12 for the first forecasts). Several previous studies have shown that the predictive relations in asset markets may not be stable in time (see, e.g., Pesaran and Timmermann (2002)). Therefore, the rolling window approach is often preferred, as it is able to better take possible structural changes into account than the expanding window approach. We also performed robustness checks based the expanding window and

the 10% level) only in the model for the bivariate case of Italy and U.S.

a shorter 5-year rolling window, but the results remain essentially similar to those presented below (available upon request).

6.1 Statistical Forecast Evaluation

The out-of-sample forecasting results of the different models and markets are presented in Table 8. We focus on two measures of statistical forecasting performance that are easy to interpret and compare, i.e. the success ratio (SR) and the AUC. We also considered a number of alternative measures (the results are available upon request), including the Pesaran and Timmermann (2009) predictability test and the out-of-sample pseudo-R^2 and QPS. These measures indicate a low level of out-of-sample predictability, as has typically been found in previous studies as well. As discussed in Section 3, we believe that especially the AUC gives a better description of the predictive performance of the models than the above-mentioned measures and, hence, we emphasize that measure.

The results in Table 8 show that the out-of-sample success ratios and the AUCs are, as expected, generally lower than their in-sample counterparts. Moreover, we find that the univariate models where only the dividend yield and three-month interest rate are included as predictors (not presented), perform poorly also out of sample, and the AUCs for the models in which the model selection is done for each market separately (denoted by UNI in Table 8), are higher in all eleven cases. Furthermore, including the lagged U.S. return ($RM_{U.S.,t-1}$) as a predictor in the univariate model (model UNIRM) improves out-of-sample performance measured by the AUC in six out of the ten non-U.S. countries.

In Section 5.3, we found that the bivariate probit model with the contemporaneous linkage via the parameter c (Model 3) outperforms the univariate model containing the lagged U.S. return ($RM_{U.S.,t-1}$) in sample. The out-of-sample evidence is to some extent the same: The AUC is higher for six out of the markets. Similarly the success rate (SR) is higher for eight out of the ten markets. Therefore, we are able to conclude that the proposed bivariate model, where effects from the U.S. to the other market via c are allowed for, is useful in predicting the direction of excess stock returns in a number of international markets.

In Table 8, the bivariate model (Model 3) is also superior to the univariate model that includes the lagged U.S. excess return as a predictor. This finding gives further indication that the predictive power included in the U.S. economic fundamentals is not completely captured by the lagged U.S. return. There are some differences in results across the markets, but as a general finding, the role of the U.S. is, perhaps unsurprisingly, notable for countries such as Australia and Italy where the domestic predictors have little predictive ability.

6.2 Market Timing Tests

In addition to statistical measures, the out-of-sample performance of the models can also be assessed by the success of their implied trading strategies. This approach is partly motivated by Leitch and Tanner (1991), among others, who argue that the models performing well according to statistical criteria might not be profitable in market timing, and vice versa. As the central idea of this paper is to study the predictive role of information originating from the U.S. on the excess returns in other markets, it is also of interest to examine the economic significance of this predictive linkage.

Following the idea of decomposing a stock return to sign and absolute value (volatility) component (see, e.g., Anatolyev and Gospodinov (2010) and Rydberg and Shephard (2003)), it is of interest to examine whether these two components, and their potential predictability, have different consequences in terms of economic value in market timing experiments. In the previous literature, Fleming et al. (2001), among others, have considered the economic value of volatility timing where a variety of (generalized) autoregressive conditional heteroskedasticity (G)ARCH models have been employed. Below, we complement that literature concentrating on the economic value of sign predictability obtained above in terms of statistical criteria.

We consider simple trading strategies between stocks and bonds similar to those in Pesaran and Timmermann (1995), Leung et al. (2000), Guo (2006), and Nyberg (2011), among others, based on the out-of-sample forecasts of the models in Table 8 and explained more detail below. This facilitates a direct comparison of trading returns of different models and commonly used benchmarks, such as the buy-and-hold (B&H hereafter) strategy where the investor invests only in stocks during the whole out-of-sample period.

We assume that an investor makes a decision on asset allocation at the beginning of each month. The selection of assets consists of the stocks (risky assets) and the three-month T-bill rate (risk-free asset). The investment decision is based on the conditional probability of positive excess returns forecast by the models and the probability threshold ξ that we set at 0.5. If the signal forecast (15) is $\hat{y}_{jt} = 1$ (i.e. a positive return), the investor invests only in stocks. In our case this is the market portfolio, which is assumed tradable through a hypothetical index fund. If the forecast model predicts a downward movement in the stock market ($\hat{y}_{jt} = 0$), the investor allocates the whole portfolio value to the three-month T-bill. We assume zero transaction costs and no short sales for the sake of simplicity.

In Table 9, we report the annualized average returns as well as the Sharpe ratios that can take the riskiness of the portfolio into account. The latter is a convenient tool in ranking portfolio performance, although its numerical value is difficult to interpret. In Table 9, we compare the performance of the probit models to the buy-and-hold strategy. The B&H strategy yields very different returns in the different markets; whereas the annual return was 12.12% in Sweden, the return in the Japanese stock market was actually negative (-1.92%) for the out-of-sample period 1995M01-2010M12.

We find that the return implied by the strategy based on the forecasts of the bivariate model (BIV, Model 3) is higher than that of the competing strategies in eight out of the ten markets, and in the remaining two cases (Canada and Sweden), the model augmented with the lagged U.S. excess return (UNIRM) performs the best.[10] The values of the Sharpe ratio confirm these findings for all the markets except for Italy, where the Sharpe ratio is slightly higher for the univariate model (UNIRM) despite the higher average return implied by the bivariate model. The findings between the other strategies are less ambiguous; the buy-and-hold strategy yields the lowest returns in six out of the ten cases, but in four cases the UNIRM strategy performs the worst. Overall, the superiority of the bivariate model also in the trading strategies lend further support to the prominent role of the U.S. stock market in predicting the direction of returns in other markets.

7 Conclusions

We study the interrelationships between excess stock market returns in the U.S. and ten other markets. In contrast to the usual predictive regression models for actual returns, we focus on predicting the sign component of excess returns. The previous research on the sign predictability in stock returns is rather limited, although it is an important issue in various financial applications, such as market timing decisions. In the spirit of the gradual diffusion of predictive information

[10] We study the robustness of the results by considering an alternative strategy, where the threshold ξ is set equal to the rolling average of realized past values of y_{jt}. Findings in favor of the bivariate model are weaker than those presented in Table 9, but the bivariate model (BIV, Model 3) still performs the best.

across markets (see Hong and Stein (1999), Hong et al. (2007), Menzly and Ozbas (2010) and Rizova (2013)), we explore whether the combined effect of the U.S. market fundamentals (i.e. the predictive power obtained for the U.S. market) is useful in predicting the signs of returns in a number of international markets. To examine this potential leading role of the U.S., we introduce a new bivariate probit model, which adds to the previous scant econometric research on bivariate and multivariate binary time series models.

Our results show that in the univariate probit model the lagged U.S. excess stock return is a useful predictor of the sign of the excess return in a number of other markets. This finding is consistent with the previous results of Rapach et al. (2013), who study actual return predictability with conventional predictive regressions. We also find that the lagged domestic stock return and the real oil price are generally the best predictors of the sign of the return. In any case, the new bivariate probit model, allowing for a contemporaneous predictive linkage from the U.S. to the other market, outperforms the above-mentioned univariate models containing the lagged U.S. return as a predictor in eight out of ten markets supporting the gradual diffusion of directional predictive information from the U.S. to the other markets. In particular, this suggest that the predictive power is not restricted to just the lagged U.S. return. Instead, it is beneficial to use the obtained predictive power of sign forecast for the U.S. in other countries. Our out-of-sample forecasting results generally confirm these in-sample findings. Specifically, we find that the bivariate model produces the best out-of-sample sign forecasts for the majority of markets and, importantly, utilizing these forecasts result in higher trading returns in simple asset allocation experiments than a number of competing models.

This study could be extended in a number of ways. The possible time variation in the parameters of binary response models has not been studied in the context of sign predictability of returns although, e.g., Pesaran and Timmermann (2002) have pointed out issues related to model instability. In terms of testing the gradual diffusion of information across markets, the use of higher frequency data (such as daily data) could also be considered with the new bivariate model suggested in this study. Furthermore, more complicated (out-of-sample) trading strategies might also be of interest, but this requires first a closer examination of the linkage between the binary response models and portfolio optimization decisions, which lies outside of the scope of this paper.

References

S. Anatolyev. Multi-market direction-of-change modeling using dependence ratios. *Studies in Nonlinear Dynamics and Econometrics*, 13:1–24, 2009.

S. Anatolyev and N. Gospodinov. Modeling financial return dynamics via decomposition. *Journal of Business and Economic Statistics*, 28:232–245, 2010.

A. Ang and G. Bekaert. Return predictability: Is it there? *Review of Financial Studies*, 20:651–707, 2007.

J.R. Ashford and R.R. Sowden. Multi-variate probit analysis. *Biometrics*, 26: 535–546, 1970.

T.J. Berge and O. Jorda. Evaluating the classification of economic activity into recessions and expansions. *American Economic Journal: Macroeconomics*, 3: 246–277, 2011.

J.Y. Campbell and S.B. Thompson. Predicting excess returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, 21:1509–1531, 2008.

T. Cenesizoglu and A. Timmermann. Do return prediction models add economic value. *Journal of Banking and Finance*, 36:2974–2987, 2012.

T. Chevapatrakul. Return sign forecasts based on conditional risk: Evidence from the UK stock market index. *Journal of Banking and Finance*, 37:2342–2353, 2013.

C. Christiansen, J.N. Eriksen, and S.T. Moller. Forecasting US recessions: The role of sentiment. *Journal of Banking and Finance*, 49:459–468, 2014.

P.F. Christoffersen and F.X. Diebold. Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science*, 52:1273–1288, 2006.

P.F. Christoffersen, F.X. Diebold, R.S. Mariano, A.S. Tay, and Y.K. Tse. Direction-of-change forecasts based on conditional variance, skewness and kurtosis dynamics: international evidence. *Journal of Financial Forecasting*, 1:1–22, 2007.

C.-S. Chu, L. Lu, and Z. Shi. Pitfalls in market timing tests. *Economics Letters*, 103:123–126, 2009.

J. Davidson. *Econometric Theory*. Wiley-Blackwell, Oxford, 2000.

R.M. de Jong and T. Woutersen. Dynamic time series binary choice. *Econometric Theory*, 27:673–702, 2011.

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

J. Fleming, C. Kirby, and B. Ostdiek. The economic value of volatility timing. *Journal of Finance*, 56:329–352, 2001.

L. Gagnon and G.A. Karolyi. Price and volatility transmission across borders. *Financial Markets, Institutions and Instruments*, 15:107–158, 2006.

A. Goyal and I. Welch. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21:1455–1508, 2008.

C.W.J. Granger and M.H. Pesaran. Economic and statistical measures of forecast accuracy. *Journal of Forecasting*, 19:537–600, 2000.

W. Greene. *Econometric Analysis*. Prentice Hall International, London, 7th edition, 2012.

H. Guo. On the out-of-sample predictability of stock returns. *Journal of Business*, 79:645–670, 2006.

E. Hjalmarsson. Predicting global stock returns. *Journal of Financial and Quantitative Analysis*, 45:49–80, 2010.

H. Hong and J.C. Stein. A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance*, 54:2143–2184, 1999.

H. Hong, W. Torous, and R. Valkanov. Do industries lead stock markets? *Journal of Financial Economics*, 83:367–396, 2007.

Y.-C. Hsu and R. Lieli. Inference for ROC curves based on estimated predictive indices: A note on testing AUC = 0.5. *Unpublished manuscript*, 2014.

G.A. Karolyi. A multivariate GARCH model of international transmissions of stock returns and volatility: The case of the United States and Canada. *Journal of Business and Economic Statistics*, 13:11–25, 1995.

H. Kauppi and P. Saikkonen. Predicting U.S. recessions with dynamic binary response models. *Review of Economics and Statistics*, 90:777–791, 2008.

K. Lahiri and J.G. Wang. Evaluating probability forecasts for GDP declines using alternative methodologies. *International Journal of Forecasting*, 29:175–190, 2013.

G. Leitch and J.E. Tanner. Economic forecast evaluation: Profit versus the conventional error measures. *American Economic Review*, 81:580–590, 1991.

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

L. Menzly and O. Ozbas. Market segmentation and cross-predictability of returns. *Journal of Finance*, 65:1555–1580, 2010.

R. Merton. On market timing and investment performance: An equilibrium theory of value for market forecasters. *Journal of Business*, 54:363–406, 1981.

W. Newey and K. West. A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708, 1987.

H. Nyberg. Forecasting the direction of the US stock market with binary probit models. *International Journal of Forecasting*, 27:561–578, 2011.

H. Nyberg. A bivariate autoregressive probit model: business cycle linkages and transmission of recession probabilities. *Macroeconomic Dynamics*, 18:838–862, 2014.

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

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

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

M.H. Pesaran and A. Timmermann. Market timing and return prediction under model instability. *Journal of Empirical Finance*, 9:495–510, 2002.

M.H. Pesaran and A. Timmermann. Testing dependence among serially correlated multi-category variables. *Journal of the American Statistical Association*, 485: 325–337, 2009.

H. Pönkä. Predicting the direction of US stock markets using industry returns. *HECER Discussion Paper*, 385, 2014.

D.E. Rapach and G. Zhou. Forecasting stock returns. In G. Elliott and A. Timmermann, editors, *Handbook of Economic Forecasting*, volume 2A, pages 329–383. North-Holland, 2013.

D.E. Rapach, M.E. Wohar, and J. Rangvid. Macro variables and international stock return predictability. *International Journal of Forecasting*, 21:137–166, 2005.

D.E. Rapach, J.K. Strauss, and G. Zhou. International stock return predictability: What is the role of the United States. *Journal of Finance*, 68:1633–1662, 2013.

S. Rizova. Trade momentum. *Journal of International Financial Markets, Institutions and Money*, 24:258–293, 2013.

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

Appendix A: Maximum likelihood estimation

This appendix shows how the log-likelihood function of the new bivariate probit model (Model 4) are determined by equations (7) and (11). The restricted models (Models 1–3) can be obtained by imposing suitable restrictions on Model 4. Special attention below will be paid to the derivation of the robust standard errors of the estimates of the parameters.

The notation closely follows Greene (2012), pp. 778–781 (see also Nyberg (2014)). We start with the construction of the log-likelihood function. Suppose we have observed a binary time series y_{jt}, $j = 1, 2$, such as (2). Define $q_{jt} = 2y_{jt} - 1$ and $\mu_{jt} = q_{jt}\pi_{jt}$, $j = 1, 2$, so that

$$q_{jt} = \begin{cases} 1 & \text{if } y_{jt} = 1, \\ -1 & \text{if } y_{jt} = 0, \end{cases}$$

and

$$\mu_{jt} = \begin{cases} \pi_{jt} & \text{if } y_{jt} = 1, \\ -\pi_{jt} & \text{if } y_{jt} = 0. \end{cases}$$

Furthermore, set

$$\rho_t^* = q_{1t} q_{2t} \rho.$$

The conditional probabilities of the different outcomes of (y_{1t}, y_{2t}) given in (7) can thus be expressed as

$$p_{ij,t} = \Phi_2(\mu_{1t}, \mu_{2t}, \rho_t^*), \quad i, j = 0, 1,$$

where ρ is the correlation coefficient in the bivariate normal distribution function.

Let $\boldsymbol{\theta} = \begin{bmatrix} \omega_1 & \boldsymbol{\beta}_1 & \omega_2 & \boldsymbol{\beta}_2 & c & \rho \end{bmatrix}'$ denote the vector of the parameters of the bivariate probit model (11). The conditional log-likelihood function, conditional on the initial values, is the sum of the individual log-likelihoods $l_t(\boldsymbol{\theta})$,

$$\begin{aligned} l(\boldsymbol{\theta}) &= \sum_{t=1}^{T} l_t(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log\Big(\Phi_2(\mu_{1t}, \mu_{2t}, \rho_t^*)\Big) \\ &= \sum_{t=1}^{T} \Big(y_{1t} y_{2t} \log(p_{11,t}) + y_{1t}(1 - y_{2t}) \log(p_{10,t}) + (1 - y_{1t}) y_{2t} \log(p_{01,t}) \\ &\quad + (1 - y_{1t})(1 - y_{2t}) \log(p_{00,t}) \Big). \end{aligned}$$

The maximization of $l(\boldsymbol{\theta})$ is clearly a highly nonlinear problem, but it can be straightforwardly carried out by standard numerical methods.

To obtain robust standard errors for the parameter coefficients, we need the score of the log-likelihood function. The score vector is defined as

$$s(\boldsymbol{\theta}) = \sum_{t=1}^{T} s_t(\boldsymbol{\theta}) = \sum_{t=1}^{T} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}},$$

where

$$s_t(\boldsymbol{\theta}) = \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{\Phi_2(\mu_{1t}, \mu_{2t}, \rho_t^*)} \frac{\partial \Phi_2(\mu_{1t}, \mu_{2t}, \rho_t^*)}{\partial \boldsymbol{\theta}}.$$

Split the parameter vector into three disjoint components, namely $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^{'} \quad \boldsymbol{\theta}_2^{'} \quad \rho]^{'}$, where the parameters in $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are related to the specifications of π_{1t} and π_{2t}. Note, however, that in contrast to the usual bivariate specification (Model 2), the parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are not separable in Model 4 (and Model 3) as the linear function π_{2t} is dependent on π_{1t} via the coefficient c and, thus, the estimates of $\boldsymbol{\theta}_1$ are not necessarily the same as obtained with the univariate independent models (Model 1).

Let us partition the score vector accordingly as

$$s_t(\boldsymbol{\theta}) = \Big[s_{1t}(\boldsymbol{\theta}_1)^{'} \quad s_{2t}(\boldsymbol{\theta}_2)^{'} \quad s_{3t}(\rho) \Big]^{'}.$$

The components of $s_t(\boldsymbol{\theta}_j)$ with respect of $\boldsymbol{\theta}_j$, $j = 1, 2,,$ can be written as

$$\begin{aligned} s_{jt}(\boldsymbol{\theta}_j) &= \frac{1}{\Phi_2(\mu_{1t}, \mu_{2t}, \rho_t^*)} \frac{\partial \Phi_2(\mu_{1t}, \mu_{2t}, \rho_t^*)}{\partial \boldsymbol{\theta}_j} \\ &= \frac{1}{\Phi_2(\mu_{1t}, \mu_{2t}, \rho_t^*)} \Big[\frac{\partial \Phi_2(\mu_{1t}, \mu_{2t}, \rho_t^*)}{\partial \mu_{1t}} \frac{\partial \mu_{1t}}{\partial \pi_{1t}} \frac{\partial \pi_{1t}}{\partial \boldsymbol{\theta}_j} + \frac{\partial \Phi_2(\mu_{1t}, \mu_{2t}, \rho_t^*)}{\partial \mu_{2t}} \frac{\partial \mu_{2t}}{\partial \pi_{2t}} \frac{\partial \pi_{2t}}{\partial \boldsymbol{\theta}_j} \Big]. \end{aligned}$$

For Model 4, we obtain

$$\frac{\partial \pi_{2t}}{\partial \boldsymbol{\theta}_1} = \Big[\frac{\partial \pi_{2t}}{\partial \omega_1} \quad \frac{\partial \pi_{2t}}{\partial \boldsymbol{\beta}_1} \Big]^{'} = \Big[c \quad \boldsymbol{x}_{1,t-1} c \Big]^{'},$$

and

$$\frac{\partial \pi_{1t}}{\partial \boldsymbol{\theta}_2} = \Big[\frac{\partial \pi_{1t}}{\partial \omega_2} \quad \frac{\partial \pi_{1t}}{\partial \boldsymbol{\beta}_2} \quad \frac{\partial \pi_{1t}}{\partial c} \Big]^{'} = \mathbf{0},$$

while for Model 2 the first derivative is also zero (when the contemporaneous link does not exist ($c = 0$)).

Therefore, the first component, $s_t(\boldsymbol{\theta}_1)$, is

$$s_{1t}(\boldsymbol{\theta}_1) = \frac{1}{\Phi_2(\mu_{1t}, \mu_{2t}, \rho_t^*)} \Big[\phi(\mu_{1t}) \Phi\Big(\frac{\mu_{2t} - \mu_{1t}\rho_t^*}{\sqrt{1 - \rho_t^{*2}}} \Big) q_{1t} \frac{\partial \pi_{1t}}{\partial \boldsymbol{\theta}_1} + \phi(\mu_{2t}) \Phi\Big(\frac{\mu_{1t} - \mu_{2t}\rho_t^*}{\sqrt{1 - \rho_t^{*2}}} \Big) q_{2t} \frac{\partial \pi_{2t}}{\partial \boldsymbol{\theta}_1} \Big],$$

and the second component is

$$s_{2t}(\boldsymbol{\theta}_2) = \frac{1}{\Phi_2(\mu_{1t}, \mu_{2t}, \rho_t^*)} \phi(\mu_{2t}) \Phi\Big(\frac{\mu_{1t} - \mu_{2t}\rho_t^*}{\sqrt{1 - \rho_t^{*2}}} \Big) q_{2t} \frac{\partial \pi_{2t}}{\partial \boldsymbol{\theta}_2},$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and cumulative distribution functions of the standard normal distribution, respectively. In Model 4, the derivatives $\partial\pi_{1t}/\partial\boldsymbol{\theta}_1$ and $\partial\pi_{2t}/\partial\boldsymbol{\theta}_2$ equal

$$\frac{\partial\pi_{1t}}{\partial\boldsymbol{\theta}_1} = \left[\begin{array}{cc} \frac{\partial\pi_{1t}}{\partial\omega_1} & \frac{\partial\pi_{1t}}{\partial\boldsymbol{\beta}_1} \end{array} \right]' = \left[\begin{array}{cc} 1 & \boldsymbol{x}_{1,t-1} \end{array} \right]',$$

and

$$\frac{\partial\pi_{2t}}{\partial\boldsymbol{\theta}_2} = \left[\begin{array}{ccc} \frac{\partial\pi_{2t}}{\partial\omega_2} & \frac{\partial\pi_{2t}}{\partial\boldsymbol{\beta}_2} & \frac{\partial\pi_{2t}}{\partial c} \end{array} \right]' = \left[\begin{array}{ccc} 1 & \boldsymbol{x}_{2,t-1} & \pi_{1t} \end{array} \right]'.$$

The values of $s_{jt}(\boldsymbol{\theta}_1)$ depend on the realized values of y_{1t} and y_{2t}. For instance, if $y_{1t} = 1$ and $y_{2t} = 1$, then by the definitions of μ_{jt} and q_{1t}, we get

$$s_{1t}(\boldsymbol{\theta}_1) = \frac{1}{\Phi_2(\pi_{1t}, \pi_{2t}, \rho)} \left[\phi(\pi_{1t})\Phi\Big(\frac{\pi_{2t} - \pi_{1t}\rho}{\sqrt{1-\rho}}\Big)\frac{\partial\pi_{1t}}{\partial\boldsymbol{\theta}_1} + \phi(\pi_{2t})\Phi\Big(\frac{\pi_{1t} - \pi_{2t}\rho}{\sqrt{1-\rho}}\Big)\frac{\partial\pi_{2t}}{\partial\boldsymbol{\theta}_1} \right],$$

and

$$s_{2t}(\boldsymbol{\theta}_1) = \frac{1}{\Phi_2(\pi_{1t}, \pi_{2t}, \rho)}\phi(\pi_{2t})\Phi\Big(\frac{\pi_{1t} - \pi_{2t}\rho}{\sqrt{1-\rho}}\Big)\frac{\partial\pi_{2t}}{\partial\boldsymbol{\theta}_2}.$$

Following Greene (2012, pp. 780), the score with respect of the correlation coefficient ρ becomes

$$s_{3t}(\rho) = \frac{\partial l_t(\boldsymbol{\theta})}{\partial\rho} = \frac{1}{\Phi_2(\mu_{1t}, \mu_{2t}, \rho_t^*)}\frac{\partial\Phi_2(\mu_{1t}, \mu_{2t}, \rho_t^*)}{\partial\rho_t^*}\frac{\partial\rho_t^*}{\partial\rho} = \frac{\phi_2(\mu_{1t}, \mu_{2t}, \rho_t^*)}{\Phi_2(\mu_{1t}, \mu_{2t}, \rho_t^*)}q_{1t}q_{2t}.$$

As above, the value of $s_{3t}(\rho)$ depends on the realized values of the dependent variables. For example, if $y_{1t} = 1$ and $y_{2t} = 1$, then we get

$$s_{3t}(\rho) = \frac{\phi_2(\pi_{1t}, \pi_{2t}, \rho)}{\Phi_2(\pi_{1t}, \pi_{2t}, \rho)},$$

and if $y_{1t} = 1$ and $y_{2t} = 0$,

$$s_{3t}(\rho) = -\frac{\phi_2(\pi_{1t}, -\pi_{2t}, -\rho)}{\Phi_2(\pi_{1t}, -\pi_{2t}, -\rho)}.$$

Maximization of the log-likelihood function yields the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$, which solves the first-order condition $s(\widehat{\boldsymbol{\theta}}) = \mathbf{0}$, where the score vector is obtained above. At the moment there is no formal proof of the asymptotic distribution of the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}$. However, under appropriate regularity conditions, including the stationarity of explanatory variables ($\boldsymbol{x}_{j,t-1}$) and the correctness of the probit model specification, it is reasonable to assume that the ML estimator $\widehat{\boldsymbol{\theta}}$ is consistent and asymptotically normal. This facilitates the use of the conventional tests for the components of the parameter vector $\boldsymbol{\theta}$ in the usual way.

Throughout this paper, the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}$ is interpreted as a quasi-maximum likelihood estimator (QMLE). Therefore, we consider the following asymptotic distribution of $\widehat{\boldsymbol{\theta}}$

$$T^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_*) \xrightarrow{d} N\Big(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}_*)^{-1}\mathcal{J}(\boldsymbol{\theta}_*)\mathcal{I}(\boldsymbol{\theta}_*)^{-1}\Big),$$

where the asymptotic covariance matrix consists of $\mathcal{I}(\boldsymbol{\theta}) = \operatorname{plim} T^{-1}\sum_{t=1}^{T}(\partial^2 l_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{'})$ and $\mathcal{J}(\boldsymbol{\theta}) = \operatorname{plim} T^{-1}\sum_{t=1}^{T} s_t(\boldsymbol{\theta})s_t(\boldsymbol{\theta})^{'}$. In this expression, $\boldsymbol{\theta}_*$ is the value in the parameter space of $\boldsymbol{\theta}$ assumed to maximize the probability limit of $T^{-1}l(\boldsymbol{\theta})$ (see, e.g., Davidson (2000, Section 9.3) for details). If the model is correctly specified, then $\mathcal{I}(\boldsymbol{\theta}) = \mathcal{J}(\boldsymbol{\theta})$.

Robust standard errors based on the QMLE (reported in the estimation results in Sections 5 and 6) are obtained from the diagonal elements of the asymptotic covariance matrix, where $\mathcal{I}(\boldsymbol{\theta})$ and $\mathcal{J}(\boldsymbol{\theta})$ are replaced by their sample analogues. That is, we compute the diagonal elements of

$$\widehat{\mathcal{I}}(\widehat{\boldsymbol{\theta}})^{-1}\widehat{\mathcal{J}}(\widehat{\boldsymbol{\theta}})\widehat{\mathcal{I}}(\widehat{\boldsymbol{\theta}})^{-1}.$$

A consistent estimator of the matrix $\mathcal{I}(\boldsymbol{\theta}_*)$ is obtained as

$$\widehat{\mathcal{I}}(\widehat{\boldsymbol{\theta}}) = T^{-1}\sum_{t=1}^{T}(\partial^2 l_t(\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{'}),$$

but the estimation of the matrix $\mathcal{J}(\boldsymbol{\theta})$ is more complicated. Following the procedure proposed by Kauppi and Saikkonen (2008), applied to univariate probit models in this paper, we use a general estimator given by

$$\widehat{\mathcal{J}}(\widehat{\boldsymbol{\theta}}) = T^{-1}\left(\sum_{t=1}^{T} s_t(\widehat{\boldsymbol{\theta}})s_t(\widehat{\boldsymbol{\theta}})^{'} + \sum_{j=1}^{T-1} w_{T_j}\sum_{t=j+1}^{T}\Big(s_t(\widehat{\boldsymbol{\theta}})s_{t-j}(\widehat{\boldsymbol{\theta}})^{'} + s_{t-j}(\widehat{\boldsymbol{\theta}})s_t(\widehat{\boldsymbol{\theta}})^{'}\Big),\right)$$

where $w_{T_j} = k(j/mT)$ for an appropriate kernel function $k(x)$. In our empirical application, we use the Parzen kernel function (see Davidson (2000), p. 227) and, similarly as Kauppi and Saikkonen (2008), the bandwidth m_T is selected according to the rule $m_T = floor(4(T/100)^{2/9})$, where the function $floor(x)$ rounds x to the nearest integer less than or equal to x.

Tables

Table 1: Descriptive statistics for excess stock returns and their sign components.

	U.S.	AUS	CAN	FRA	GER	ITA	JPN	NED	SWE	SUI	U.K.
Mean	0.547	0.353	0.303	0.495	0.511	0.422	0.219	0.677	1.027	0.551	0.497
St.dev	4.500	5.070	4.725	5.727	5.710	6.983	5.386	5.381	6.727	4.634	4.682
Correl	1.000	0.589	0.774	0.644	0.630	0.428	0.417	0.704	0.566	0.688	0.730
Φ-coeff.	1.000	0.458	0.594	0.480	0.402	0.368	0.263	0.547	0.372	0.452	0.490

Notes: This table reports descriptive statistics for monthly excess stock market returns (in percentages) for the full sample period 1980M2–2010M12. The reported statistics include the means, standard deviations and correlations with the U.S. excess return series. We also report the Φ-coefficients for the binary time series extracted from the excess stock returns (see (2)).

Table 2: Univariate probit models with the dividend yield and three-month interest rate as predictors.

	AUS	CAN	FRA	GER	ITA	JPN	NED	SWE	SUI	U.K.	U.S.
$CONST$	0.573**	0.253	0.307*	0.271	0.252	-0.110	0.137	0.314*	0.490**	-0.041	0.214
	(0.288)	(0.232)	(0.160)	(0.213)	(0.229)	(0.213)	(0.163)	(0.180)	(0.226)	(0.265)	(0.173)
TB_{t-1}	-0.026	-0.062**	-0.020	-0.053*	-0.012	0.017	-0.076**	-0.017	-0.066*	-0.022	-0.090***
	(0.019)	(0.029)	(0.021)	(0.032)	(0.012)	(0.029)	(0.032)	(0.015)	(0.035)	(0.023)	(0.032)
DY_{t-1}	-0.42	0.111	-0.002	0.047	-0.034	0.122	0.125**	-0.012	-0.012	0.117	0.181**
	(0.085)	(0.136)	(0.053)	(0.079)	(0.060)	(0.160)	(0.058)	(0.058)	(0.129)	(0.088)	(0.086)
$logL$	-250.673	-249.488	-251.738	-251.551	-255.669	-255.621	-246.125	-252.824	-245.850	-248.122	-245.390
AIC	253.673	252.488	254.738	254.551	258.669	258.621	249.125	255.824	248.850	251.122	248.390
QPS	0.484	0.481	0.487	0.487	0.498	0.498	0.474	0.490	0.472	0.478	0.470
psR^2	0.011	0.026	0.004	0.007	0.003	0.002	0.018	0.004	0.017	0.005	0.024
$adj.psR^2$	0.003	0.018	-0.004	-0.001	-0.005	-0.006	0.010	-0.004	0.009	-0.003	0.016
SR	0.592	0.614	0.576	0.584	0.535	0.524	0.608	0.554	0.622	0.600	0.616
AUC	0.557	0.582	0.545	0.540	0.534	0.524	0.589	0.552	0.565	0.540	0.580
PT	1.164	12.782**	0.217	0.377	1.202	2.854*	1.323	0.174	0.286	NaN	0.016

Notes: We present the in-sample results of the univariate probit models (see (4) and (5) for eleven markets. This table presents the predictive power of the dividend yield (DY) and three-month interest rate (TB). Robust standard errors of the estimated coefficients are reported in brackets (see Kauppi and Saikkonen (2008)). The goodness-of-fit measures are described more detail in Section 3. In the table, *, **, and *** denote the statistical significance of the estimated coefficients and the Pesaran and Timmermann (2009) (PT) predictability test at 10%, 5% and 1% significance levels, respectively.

Table 3: In-sample estimation results of the selected univariate probit models for different markets.

	AUS	CAN	FRA	GER	ITA	JPN	NED	SWE	SUI	U.K.	U.S.
$CONST$	0.546***	0.428***	0.177***	0.016	0.054	0.056	0.175	0.070	0.349***	-0.149	0.600***
	(0.193)	(0.126)	(0.065)	(0.111)	(0.070)	(0.066)	(0.166)	(0.082)	(0.125)	(0.266)	(0.214)
TB_{t-1}		-0.044***					-0.079**				
		(0.017)					(0.033)				
DY_{t-1}							0.122**			0.134*	0.449***
							(0.060)			(0.069)	(0.120)
RM_{t-1}			0.027**	0.023*	0.015*	0.030**			0.068***		0.049**
			(0.011)	(0.012)	(0.009)	(0.013)			(0.021)		(0.024)
RMI_{t-1}									-0.379**		-0.320
									(0.190)		(0.220)
IP_{t-1}			0.127**								0.161
			(0.062)								(0.106)
$10Y_{t-1}$	-0.040**										-0.206***
	(0.019)										(0.050)
TS_{t-1}				0.107*				0.116**	0.135***		
				(0.059)				(0.044)	(0.051)		
INF_{t-1}										-0.391***	
										(0.148)	
REX_{t-1}		0.051									
		(0.033)									
OIL_{t-1}			-0.015*	-0.011	-0.024***		-0.015*	-0.015**	-0.013		
			(0.008)	(0.007)	(0.008)		(0.008)	(0.007)	(0.009)		
$logL$	-250.401	-248.924	-245.414	-248.208	-249.676	-252.916	-244.054	-248.586	-238.778	-244.902	-240.258
AIC	252.401	251.924	249.414	252.208	252.676	254.916	248.054	251.586	253.562	247.902	246.258
QPS	0.484	0.480	0.471	0.478	0.482	0.490	0.467	0.479	0.454	0.469	0.458
psR^2	0.013	0.029	0.038	0.025	0.035	0.017	0.029	0.027	0.055	0.022	0.051
$adj.psR^2$	0.008	0.021	0.028	0.014	0.027	0.011	0.018	0.019	0.042	0.014	0.036
SR	0.559	0.597	0.595	0.614	0.565	0.543	0.619	0.565	0.627	0.614	0.638
AUC	0.578	0.588	0.603	0.592	0.597	0.576	0.611	0.605	0.651	0.581	0.620
PT	0.018	2.45	4.111**	13.116***	5.776**	0.276	3.641*	1.520	5.015**	4.335**	8.441***

Notes: The table presents the in-sample estimation results for the best univariate models based on the model selection procedure described in Section 5.1 for each country. See also the notes to Table 2.

Table 4: In-sample estimation results for univariate probit models including the lagged U.S. return.

	AUS	CAN	FRA	GER	ITA	JPN	NED	SWE	SUI	U.K.
$CONST$	0.543***	0.414***	0.171**	0.016	0.045	0.043	0.142	0.057	0.347	-0.156
	(0.182)	(0.124)	(0.067)	(0.113)	(0.066)	(0.066)	(0.188)	(0.076)	(0.138)	(0.267)
TB_{t-1}		-0.045***					-0.079**			
		(0.016)					(0.032)			
DY_{t-1}							0.127**			0.135*
							(0.057)			(0.070)
RM_{t-1}			0.017	0.008	0.009	0.022*			0.063**	
			(0.015)	(0.015)	(0.011)	(0.013)			(0.027)	
RMI_{t-1}									-0.379*	
									(0.215)	
IP_{t-1}			0.128							
			(0.055)							
$10Y_{t-1}$	-0.041**									
	(0.019)									
TS_{t-1}				0.101				0.112**	0.135***	
				(0.063)				(0.046)	(0.050)	
INF_{t-1}										-0.390***
										(0.148)
REX_{t-1}		0.017								
		(0.037)								
OIL_{t-1}			-0.015**	-0.010	-0.024***		-0.015**	-0.015**	-0.013*	
			(0.007)	(0.007)	(0.007)		(0.007)	(0.007)	(0.007)	
$RM_{u.s.,t-1}$	0.026*	0.038**	0.020	0.031	0.020	0.025	0.024	0.034**	0.008	0.008
	(0.015)	(0.016)	(0.019)	(0.019)	(0.016)	(0.016)	(0.015)	(0.015)	(0.020)	(0.015)
$logL$	-248.804	-246.139	-244.852	-246.897	-248.881	-251.759	-242.769	-245.947	-238.701	-244.749
AIC	251.804	250.139	249.852	251.897	252.881	254.759	247.769	249.947	244.701	248.749
QPS	0.480	0.474	0.470	0.474	0.481	0.488	0.465	0.472	0.454	0.469
psR^2	0.022	0.044	0.041	0.032	0.039	0.023	0.036	0.041	0.055	0.023
$adj.psR^2$	0.014	0.033	0.028	0.018	0.029	0.015	0.023	0.031	0.041	0.012
SR	0.592	0.589	0.600	0.616	0.573	0.549	0.614	0.616	0.619	0.611
AUC	0.584	0.600	0.607	0.597	0.601	0.573	0.609	0.619	0.651	0.583
PT	2.646	3.623*	5.399**	10.929***	6.720***	1.094	4.785**	15.577***	3.351*	3.815*

Notes: The table presents the in-sample estimation results for the univariate probit models for ten non-U.S. countries including the lagged U.S. return as a predictor. Other predictors are the same as in Table 3. See also the notes to Table 3.

Table 5: In-sample estimation results for bivariate Models 1–4 for the U.S. and the U.K. markets.

Dep.	Exp.	Model 1	Model 2	Model 3	Model 4
$RMI_{U.S.}$	$CONST$	0.600***	0.552***	0.612***	0.620**
		(0.214)	(0.239)	(0.219)	(0.283)
	$DY_{U.S.,t-1}$	0.449***	0.402***	0.430***	0.466***
		(0.120)	(0.116)	(0.136)	(0.136)
	$RM_{U.S.,t-1}$	0.049**	0.027	0.057**	0.041
		(0.024)	(0.023)	(0.025)	(0.037)
	$RMI_{U.S.,t-1}$	-0.320	-0.133	-0.417*	-0.253
		(0.220)	(0.193)	(0.245)	(0.338)
	$IP_{U.S.,t-1}$	0.161	0.108	0.184	0.152
		(0.106)	(0.084)	(0.116)	(0.129)
	$10Y_{U.S.,t-1}$	-0.206***	-0.193***	-0.193***	-0.220***
		(0.050)	(0.046)	(0.061)	(0.056)
$RMI_{U.K.}$	$CONST$	-0.149	-0.062	-0.174	-0.112
		(0.266)	(0.342)	(0.243)	(0.315)
	$DY_{U.K.,t-1}$	0.134*	0.111	0.113*	0.105
		(0.069)	(0.097)	(0.067)	(0.086)
	$INF_{U.K.,t-1}$	-0.391***	-0.384**	-0.361**	-0.364**
		(0.148)	(0.152)	(0.154)	(0.157)
	ρ		0.721***		0.719***
			(0.014)		(0.015)
	c			0.405	0.284
				(0.294)	(0.289)
	$logL$	-485.160	-438.313	-483.855	-437.652
	AIC	494.160	448.313	493.855	448.652
	$QPS_{U.S.}$	0.458	0.459	0.459	0.458
	$QPS_{U.K.}$	0.469	0.469	0.465	0.467
	psR^2	$0.072^{\dagger}$	$0.074^{\ddagger}$	$0.079^{\dagger}$	$0.078^{\ddagger}$
	$adj.psR^2$	$0.049^{\dagger}$	$0.048^{\ddagger}$	$0.053^{\dagger}$	$0.049^{\ddagger}$
	$SR_{U.S.}$	0.638	0.659	0.632	0.635
	$SR_{U.K.}$	0.614	0.614	0.622	0.597
	$AUC_{U.S.}$	0.620	0.624	0.615	0.623
	$AUC_{U.K.}$	0.581	0.584	0.601	0.598
	$PT_{U.S.}$	8.441***	14.071***	8.481***	6.810***
	$PT_{U.K.}$	4.335**	4.470**	7.495***	0.918

Notes: The table presents the in-sample estimation results for the different bivariate probit models for the U.S. and the U.K. markets. Robust standard errors are reported in brackets. In the table, *, **, and *** denote the statistical significance at the 10, 5 and 1% level, respectively. Note that the psR^2 and $adj.psR^2$ values are only comparable between Models 1 and 3 (denoted by †), and Models 2 and 4 (denoted by ‡).

Table 6: In-sample estimation results for bivariate Models 1–4 for the U.S. and Swedish markets.

Dep.	Exp.	Model 1	Model 2	Model 3	Model 4
$RMI_{U.S.}$	$CONST$	0.600***	0.603**	0.501**	0.578**
		(0.214)	(0.236)	(0.238)	(0.284)
	$DY_{U.S.,t-1}$	0.449***	0.396***	0.401***	0.436***
		(0.120)	(0.153)	(0.138)	(0.162)
	$RM_{U.S.,t-1}$	0.049**	0.032	0.057**	0.053
		(0.024)	(0.026)	(0.022)	(0.035)
	$RMI_{U.S.,t-1}$	-0.320	-0.274	-0.294	-0.317
		(0.220)	(0.203)	(0.215)	(0.235)
	$IP_{U.S.,t-1}$	0.161	0.093	0.204**	0.183
		(0.106)	(0.129)	(0.096)	(0.162)
	$10Y_{U.S.,t-1}$	-0.206***	-0.187***	-0.177***	-0.199**
		(0.050)	(0.071)	(0.063)	(0.079)
RMI_{SWE}	$CONST$	0.070*	0.092	0.068	-0.027
		(0.082)	(0.082)	(0.118)	(0.144)
	$TS_{SWE,t-1}$	0.116**	0.091**	0.109**	0.086*
		(0.044)	(0.045)	(0.047)	(0.048)
	$OIL_{SWE,t-1}$	-0.015**	-0.013	-0.014*	-0.012
		(0.007)	(0.009)	(0.007)	(0.008)
	ρ		0.539***		0.528***
			(0.019)		(0.022)
	c			0.587*	0.500
				(0.347)	(0.477)
	$logL$	-488.845	-466.083	-486.266	-464.429
	AIC	497.845	476.083	496.266	475.429
	$QPS_{U.S.}$	0.458	0.459	0.459	0.458
	QPS_{SWE}	0.479	0.480	0.470	0.473
	psR^2	$0.077^{\dagger}$	$0.062^{\ddagger}$	$0.090^{\dagger}$	$0.070^{\ddagger}$
	$adj.psR^2$	$0.054^{\dagger}$	$0.035^{\ddagger}$	$0.065^{\dagger}$	$0.042^{\ddagger}$
	$SR_{U.S.}$	0.638	0.646	0.641	0.641
	SR_{SWE}	0.565	0.570	0.608	0.605
	$AUC_{U.S.}$	0.620	0.621	0.619	0.621
	AUC_{SWE}	0.605	0.605	0.639	0.634
	$PT_{U.S.}$	8.441***	6.735***	12.410***	11.093***
	PT_{SWE}	1.520	1.794	13.443***	11.070***

Notes: See the notes to Table 5.

Table 7: In-sample estimation results for bivariate Models 1–4 for the U.S. and Canadian markets.

Dep.	Exp.	Model 1	Model 2	Model 3	Model 4
$RMI_{U.S.}$	$CONST$	0.600***	0.587	0.546**	0.576
		(0.214)	(0.357)	(0.216)	(0.384)
	$DY_{U.S.,t-1}$	0.449***	0.266**	0.410***	0.425*
		(0.120)	(0.135)	(0.146)	(0.218)
	$RM_{U.S.,t-1}$	0.049**	0.017	0.054**	0.051
		(0.024)	(0.029)	(0.022)	(0.043)
	$RMI_{U.S.,t-1}$	-0.320	-0.219	-0.255	-0.266
		(0.220)	(0.224)	(0.219)	(0.358)
	$IP_{U.S.,t-1}$	0.161	0.102	0.151	0.152
		(0.106)	(0.099)	(0.114)	(0.168)
	$10Y_{U.S.,t-1}$	-0.206***	-0.137**	-0.189***	-0.197**
		(0.050)	(0.070)	(0.060)	(0.099)
RMI_{CAN}	$CONST$	0.428**	0.433***	0.095	0.099
		(0.126)	(0.183)	(0.171)	(0.196)
	$TB_{CAN,t-1}$	-0.44***	-0.44**	-0.026	-0.024
		(0.017)	(0.024)	(0.016)	(0.019)
	$REX_{CAN,t-1}$	0.051	0.039	0.028	0.033
		(0.033)	(0.041)	(0.039)	(0.039)
	ρ		0.806***		0.799***
			(0.020)		(0.010)
	c			0.883***	0.835**
				(0.329)	(0.385)
	$logL$	-489.183	-424.031	-483.512	-419.470
	AIC	498.183	434.031	493.512	430.470
	$QPS_{U.S.}$	0.458	0.463	0.459	0.458
	QPS_{CAN}	0.480	0.480	0.466	0.466
	psR^2	$0.079^{\dagger}$	$0.051^{\ddagger}$	$0.107^{\dagger}$	$0.074^{\ddagger}$
	$adj.psR^2$	$0.056^{\dagger}$	$0.024^{\ddagger}$	$0.083^{\dagger}$	$0.046^{\ddagger}$
	$SR_{U.S.}$	0.638	0.641	0.641	0.643
	SR_{CAN}	0.597	0.608	0.600	0.603
	$AUC_{U.S.}$	0.620	0.614	0.623	0.623
	AUC_{CAN}	0.588	0.583	0.634	0.634
	$PT_{U.S.}$	8.441***	6.875***	10.236***	10.568***
	PT_{CAN}	2.450	6.687***	5.508**	4.304**

Notes: See the notes to Table 5.

Table 8: Out-of-sample forecasting results.

Model	Statistic	AUS	CAN	FRA	GER	ITA	JPN	NED	SWE	SUI	U.K.
UNI	*SR*	0.615	0.620	0.589	0.547	0.542	0.505	0.630	0.599	0.620	0.599
	AUC	0.510	0.494	0.572	0.572	0.529	0.533	0.573	0.576	0.592	0.529
UNIRM	*SR*	0.609	0.635	0.599	0.536	0.547	0.510	0.630	0.615	0.615	0.597
	AUC	0.526	0.572	0.577	0.580	0.540	0.516	0.570	0.592	0.590	0.501
BIV	*SR*	0.635	0.594	0.609	0.547	0.557	0.542	0.656	0.589	0.620	0.615
	AUC	0.546	0.551	0.591	0.567	0.575	0.543	0.573	0.581	0.572	0.557

Notes: This table displays the out-of-sample forecasting results for the period 1995M01–2010M12. The forecasts are based on the rolling estimation window of 15 years. Model UNI refers to the univariate probit models that are selected separately for each country, $UNIRM$ refers to UNI models augmented with the U.S. lagged return ($RM_{U.S.,t-1}$), and BIV refers to the bivariate model with the contemporaneous linkage via the parameter c (Model 3).

Table 9: Out-of-sample market timing tests.

Model	Statistic	AUS	CAN	FRA	GER	ITA	JPN	NED	SWE	SUI	U.K.
B&H	*RETURN*	10.30%	9.77%	8.19%	7.04%	6.27%	-1.92%	7.84%	12.12%	7.86%	7.95%
	SHARPE	1.24	1.34	0.92	0.67	0.35	-0.42	0.81	1.37	1.39	0.76
UNI	*RETURN*	10.30%	12.47%	8.72%	7.26%	7.99%	-0.29%	11.37%	15.87%	7.77%	7.95%
	SHARPE	1.24	2.13	1.18	0.73	0.96	-0.16	1.48	2.19	1.45	0.76
UNIRM	*RETURN*	10.05%	14.00%	10.48%	6.57%	9.84%	-0.03%	11.02%	17.13%	6.81%	7.74%
	SHARPE	1.17	2.63	1.56	0.63	1.33	-0.09	1.41	2.63	1.21	0.71
BIV	*RETURN*	11.30%	12.37%	10.74%	8.39%	10.03%	1.62%	14.18%	14.93%	8.61%	9.82%
	SHARPE	1.60	2.38	1.67	0.94	1.32	0.41	2.16	2.01	1.66	1.29

Notes: This table displays average annual returns and Sharpe ratios of investment strategies based on different forecasting models for the period 1995M01-2010M12. See also the notes to Table 8.

Appendix B: Estimation results of bivariate models for other countries

Table 10: In-sample estimation results for bivariate Models 1–4 for the U.S. and Australia

Dep.	Exp.	Model 1	Model 2	Model 3	Model 4
$RMI_{U.S.}$	$CONST$	0.600***	0.652***	0.521**	0.645***
		(0.214)	(0.222)	(0.228)	(0.239)
	$DY_{U.S.,t-1}$	0.449***	0.392***	0.419***	0.465***
		(0.12)	(0.116)	(0.146)	(0.148)
	$RM_{U.S.,t-1}$	0.049*	0.041**	0.044*	0.048*
		(0.024)	(0.024)	(0.026)	(0.029)
	$RMI_{U.S.,t-1}$	-0.320	-0.412**	-0.181	-0.371
		(0.220)	(0.203)	(0.270)	(0.281)
	$IP_{U.S.,t-1}$	0.161	0.105	0.178	0.153
		(0.106)	(0.100)	(0.115)	(0.139)
	$10Y_{U.S.,t-1}$	-0.206***	-0.181***	-0.195***	-0.214***
		(0.050)	(0.050)	(0.059)	(0.066)
RMI_{AUS}	$CONST$	0.546***	0.545***	0.351*	0.423**
		(0.193)	(0.222)	(0.189)	(0.183)
	$10Y_{AUS,t-1}$	-0.040***	-0.040**	-0.32*	-0.036**
		(0.019)	(0.019)	(0.017)	(0.016)
	ρ		0.676***		0.670***
			(0.015)		(0.017)
	c			0.495*	0.343
				(0.281)	(0.304)
	$logL$	-490.660	-450.781	-488.798	-449.911
	AIC	498.660	459.781	497.798	459.911
	$QPS_{U.S.}$	0.458	0.460	0.458	0.458
	QPS_{AUS}	0.484	0.484	0.478	0.481
	psR^2	$0.063^{\dagger}$	$0.060^{\ddagger}$	$0.073^{\dagger}$	$0.064^{\ddagger}$
	$adj.psR^2$	$0.042^{\dagger}$	$0.036^{\ddagger}$	$0.050^{\dagger}$	$0.038^{\ddagger}$
	$SR_{U.S.}$	0.638	0.627	0.643	0.643
	SR_{AUS}	0.559	0.562	0.597	0.584
	$AUC_{U.S.}$	0.620	0.611	0.621	0.618
	AUC_{AUS}	0.578	0.578	0.581	0.576
	$PT_{U.S.}$	8.441***	4.489**	9.143***	11.880***
	PT_{AUS}	0.018	0.315	1.858	0.036

Notes: See the notes to Table 5.

Table 11: In-sample estimation results for bivariate Models 1–4 for the U.S. and France

Dep.	Exp.	Model 1	Model 2	Model 3	Model 4
$RMI_{U.S.}$	$CONST$	0.600***	0.441**	0.662***	0.495
		(0.214)	(0.224)	(0.232)	(0.324)
	$DY_{U.S.,t-1}$	0.449***	0.468***	0.406***	0.490***
		(0.120)	(0.122)	(0.154)	(0.135)
	$RM_{U.S.,t-1}$	0.049**	0.028	0.056**	0.034
		(0.024)	(0.028)	(0.026)	(0.042)
	$RMI_{U.S.,t-1}$	-0.320	-0.162	-0.375	-0.210
		(0.220)	(0.201)	(0.244)	(0.300)
	$IP_{U.S.,t-1}$	0.161	0.107	0.183	0.129
		(0.106)	(0.104)	(0.115)	(0.148)
	$10Y_{U.S.,t-1}$	-0.206***	-0.201***	-0.194***	-0.214***
		(0.050)	(0.056)	(0.061)	(0.069)
RMI_{FRA}	$CONST$	0.177***	0.170**	0.098	0.136
		(0.065)	(0.078)	(0.104)	(0.121)
	$RM_{FRA,t-1}$	0.027**	0.029*	0.024**	0.029*
		(0.011)	(0.016)	(0.011)	(0.015)
	$IP_{FRA,t-1}$	0.127**	0.077	0.134**	0.080
		(0.062)	(0.061)	(0.052)	(0.062)
	$OIL_{FRA,t-1}$	-0.015*	-0.010	-0.014*	-0.009
		(0.008)	(0.018)	(0.008)	(0.017)
	ρ		0.711***		0.709***
			(0.022)		(0.023)
	c			0.321	0.134
				(0.335)	(0.387)
	$logL$	-485.672	-442.062	-484.871	-441.929
	AIC	495.672	453.062	495.871	453.929
	$QPS_{U.S.}$	0.458	0.459	0.458	0.459
	QPS_{FRA}	0.471	0.472	0.469	0.472
	psR^2	$0.087^{\dagger}$	$0.088^{\ddagger}$	$0.091^{\dagger}$	$0.089^{\ddagger}$
	$adj.psR^2$	$0.062^{\dagger}$	$0.060^{\ddagger}$	$0.064^{\dagger}$	$0.058^{\ddagger}$
	$SR_{U.S.}$	0.638	0.630	0.643	0.635
	SR_{FRA}	0.595	0.608	0.597	0.597
	$AUC_{U.S.}$	0.620	0.621	0.617	0.620
	AUC_{FRA}	0.603	0.598	0.612	0.600
	$PT_{U.S.}$	8.441***	2.312	13.078***	5.306**
	PT_{FRA}	4.111**	7.290***	5.492**	4.596**

Notes: See the notes to Table 5.

Table 12: In-sample estimation results for bivariate Models 1–4 for the U.S. and Germany

Dep.	Exp.	Model 1	Model 2	Model 3	Model 4
$RMI_{U.S.}$	$CONST$	0.600***	0.635***	0.555**	0.642**
		(0.214)	(0.241)	(0.233)	(0.260)
	$DY_{U.S.,t-1}$	0.449***	0.419***	0.440***	0.463***
		(0.120)	(0.116)	(0.124)	(0.130)
	$RM_{U.S.,t-1}$	0.049**	0.040	0.051**	0.046
		(0.024)	(0.027)	(0.024)	(0.030)
	$RMI_{U.S.,t-1}$	-0.320	-0.315	-0.286	-0.323
		(0.220)	(0.217)	(0.234)	(0.238)
	$IP_{U.S.,t-1}$	0.161	0.111	0.177	0.145
		(0.106)	(0.096)	(0.116)	(0.125)
	$10Y_{U.S.,t-1}$	-0.206***	-0.197***	-0.199***	-0.216***
		(0.050)	(0.053)	(0.053)	(0.058)
RMI_{GER}	$CONST$	0.016	0.076	-0.047	0.018
		(0.111)	(0.108)	(0.118)	(0.116)
	$RM_{GER,t-1}$	0.023*	0.024	0.020*	0.024
		(0.012)	(0.017)	(0.011)	(0.016)
	$TS_{GER,t-1}$	0.107*	0.063	0.099*	0.060
		(0.059)	(0.059)	(0.060)	(0.059)
	$OIL_{GER,t-1}$	-0.011*	-0.007	-0.010	-0.006
		(0.007)	(0.013)	(0.007)	(0.012)
	ρ		0.605***		0.603***
			(0.019)		(0.016)
	c			0.304	0.254
				(0.235)	(0.231)
	$logL$	-488.466	-458.652	-487.673	-458.056
	AIC	498.466	469.652	498.673	470.056
	$QPS_{U.S.}$	0.458	0.458	0.458	0.458
	QPS_{GER}	0.478	0.479	0.476	0.478
	psR^2	$0.075^{\dagger}$	$0.068^{\ddagger}$	$0.079^{\dagger}$	$0.071^{\ddagger}$
	$adj.psR^2$	$0.049^{\dagger}$	$0.039^{\ddagger}$	$0.050^{\dagger}$	$0.040^{\ddagger}$
	$SR_{U.S.}$	0.638	0.641	0.641	0.635
	SR_{GER}	0.614	0.608	0.595	0.586
	$AUC_{U.S.}$	0.620	0.620	0.620	0.621
	AUC_{GER}	0.605	0.582	0.598	0.588
	$PT_{U.S.}$	8.441***	9.839***	10.126***	7.688***
	PT_{GER}	13.116***	11.737***	5.589**	3.258*

Notes: See the notes to Table 5.

Table 13: In-sample estimation results for bivariate Models 1–4 for the U.S. and Italy

Dep.	Exp.	Model 1	Model 2	Model 3	Model 4
$RMI_{U.S.}$	$CONST$	0.600***	0.514**	0.643***	0.610**
		(0.214)	(0.216)	(0.229)	(0.250)
	$DY_{U.S.,t-1}$	0.449***	0.427***	0.419***	0.458***
		(0.120)	(0.108)	(0.139)	(0.133)
	$RM_{U.S.,t-1}$	0.049**	0.030	0.058***	0.048
		(0.024)	(0.025)	(0.025)	(0.034)
	$RMI_{U.S.,t-1}$	-0.320	-0.192	-0.384*	-0.311
		(0.220)	(0.216)	(0.229)	(0.268)
	$IP_{U.S.,t-1}$	0.161	0.138	0.158	0.161
		(0.106)	(0.098)	(0.114)	(0.121)
	$10Y_{U.S.,t-1}$	-0.206***	-0.194***	-0.196***	-0.211***
		(0.050)	(0.050)	(0.058)	(0.060)
RMI_{ITA}	$CONST$	0.054	0.053	-0.044	-0.036
		(0.070)	(0.075)	(0.087)	(0.094)
	$RM_{ITA,t-1}$	0.015*	0.008	0.013	0.007
		(0.009)	(0.010)	(0.009)	(0.009)
	$OIL_{ITA,t-1}$	-0.024***	-0.022***	-0.023***	-0.021***
		(0.008)	(0.007)	(0.008)	(0.010)
	ρ		0.552***		0.547***
			(0.017)		(0.015)
	c			0.393*	0.357
				(0.226)	(0.245)
	$logL$	-489.934	-466.167	-488.506	-465.014
	AIC	498.934	476.167	498.506	476.014
	$QPS_{U.S.}$	0.458	0.458	0.458	0.458
	QPS_{ITA}	0.482	0.483	0.478	0.479
	psR^2	$0.084^{\dagger}$	$0.071^{\ddagger}$	$0.092^{\dagger}$	$0.077^{\ddagger}$
	$adj.psR^2$	$0.062^{\dagger}$	$0.045^{\ddagger}$	$0.066^{\dagger}$	$0.049^{\ddagger}$
	$SR_{U.S.}$	0.638	0.654	0.638	0.638
	SR_{ITA}	0.565	0.554	0.589	0.581
	$AUC_{U.S.}$	0.620	0.623	0.618	0.620
	AUC_{ITA}	0.597	0.598	0.615	0.617
	$PT_{U.S.}$	8.441***	9.352***	10.276***	8.379***
	PT_{ITA}	5.776**	3.865**	11.139***	8.783

Notes: See the notes to Table 5.

Table 14: In-sample estimation results for bivariate Models 1–4 for the U.S. and Japan

Dep.	Exp.	Model 1	Model 2	Model 3	Model 4
$RMI_{U.S.}$	$CONST$	0.600***	0.633***	0.495**	0.568**
		(0.214)	(0.203)	(0.251)	(0.255)
	$DY_{U.S.,t-1}$	0.449***	0.410***	0.437***	0.462***
		(0.120)	(0.117)	(0.124)	(0.123)
	$RM_{U.S.,t-1}$	0.049**	0.040*	0.055**	0.053**
		(0.024)	(0.024)	(0.022)	(0.026)
	$RMI_{U.S.,t-1}$	-0.320	-0.278	-0.323*	-0.332
		(0.220)	(0.212)	(0.213)	(0.221)
	$IP_{U.S.,t-1}$	0.161	0.114	0.199**	0.183
		(0.106)	(0.106)	(0.097)	(0.121)
	$10Y_{U.S.,t-1}$	-0.206***	-0.196***	-0.188***	-0.206***
		(0.050)	(0.050)	(0.060)	(0.060)
RMI_{JPN}	$CONST$	0.056	0.053	0-0.071	-0.057
		(0.066)	(0.063)	(0.101)	(0.101)
	$RM_{JPN,t-1}$	0.030**	0.033*	0.028**	0.032*
		(0.013)	(0.019)	(0.014)	(0.018)
	ρ		0.411***		0.402***
			(0.013)		(0.015)
	c			0.512*	0.445
				(0.288)	(0.296)
	$logL$	-493.174	-480.713	-491.014	-479.069
	AIC	501.174	489.713	500.014	489.069
	$QPS_{U.S.}$	0.458	0.458	0.459	0.458
	QPS_{JPN}	0.490	0.490	0.484	0.485
	psR^2	$0.067^{\dagger}$	$0.066^{\ddagger}$	$0.078^{\dagger}$	$0.074^{\ddagger}$
	$adj.psR^2$	$0.046^{\dagger}$	$0.042^{\ddagger}$	$0.055^{\dagger}$	$0.048^{\ddagger}$
	$SR_{U.S.}$	0.638	0.649	0.630	0.632
	SR_{JPN}	0.543	0.546	0.559	0.541
	$AUC_{U.S.}$	0.620	0.622	0.616	0.618
	AUC_{JPN}	0.576	0.576	0.596	0.591
	$PT_{U.S.}$	8.441***	10.902***	6.897***	7.054***
	PT_{JPN}	0.276**	0.397	2.311	0.422

Notes: See the notes to Table 5.

Table 15: In-sample estimation results for bivariate Models 1–4 for the U.S. and the Netherlands

Dep.	Exp.	Model 1	Model 2	Model 3	Model 4
$RMI_{U.S.}$	$CONST$	0.600***	0.554*	0.576**	0.562
		(0.214)	(0.307)	(0.231)	(0.362)
	$DY_{U.S.,t-1}$	0.449***	0.436***	0.427***	0.442**
		(0.120)	(0.162)	(0.145)	(0.215)
	$RM_{U.S.,t-1}$	0.049**	0.026	0.056*	0.028
		(0.024)	(0.028)	(0.029)	(0.046)
	$RMI_{U.S.,t-1}$	-0.320	-0.202	-0.348	-0.210
		(0.220)	(0.233)	(0.241)	(0.313)
	$IP_{U.S.,t-1}$	0.161	0.094	0.187	0.099
		(0.106)	(0.106)	(0.126)	(0.150)
	$10Y_{U.S.,t-1}$	-0.206***	-0.202***	-0.193***	-0.204**
		(0.050)	(0.069)	(0.067)	(0.097)
RMI_{NED}	$CONST$	0.175	0.098	0.123	0.094
		(0.166)	(0.195)	(0.151)	(0.208)
	$TB_{NED,t-1}$	-0.079**	-0.074*	-0.063	-0.072
		(0.033)	(0.038)	(0.043)	(0.045)
	$DY_{NED,t-1}$	0.122**	0.133*	0.097	0.130
		(0.060)	(0.075)	(0.073)	(0.084)
	$OIL_{NED,t-1}$	-0.015*	-0.009	-0.014*	-0.009
		(0.008)	(0.012)	(0.008)	(0.012)
	ρ		0.770***		0.770***
			(0.015)		(0.016)
	c			0.280	0.032
				(0.414)	(0.564)
	$logL$	-484.313	-428.768	-483.859	-428.763
	AIC	494.313	439.768	494.859	440.763
	$QPS_{U.S.}$	0.458	0.459	0.458	0.459
	QPS_{NED}	0.467	0.468	0.465	0.468
	psR^2	$0.079^{\dagger}$	$0.071^{\ddagger}$	$0.081^{\dagger}$	$0.071^{\ddagger}$
	$adj.psR^2$	$0.053^{\dagger}$	$0.042^{\ddagger}$	$0.053^{\dagger}$	$0.039^{\ddagger}$
	$SR_{U.S.}$	0.638	0.643	0.646	0.643
	SR_{NED}	0.619	0.616	0.624	0.614
	$AUC_{U.S.}$	0.620	0.624	0.619	0.624
	AUC_{NED}	0.611	0.604	0.619	0.605
	$PT_{U.S.}$	8.441***	4.506**	13.253***	4.506**
	PT_{NED}	3.641*	1.185	6.715***	1.212

Notes: See the notes to Table 5.

Table 16: In-sample estimation results for bivariate Models 1–4 for the U.S. and Switzerland

Dep.	Exp.	Model 1	Model 2	Model 3	Model 4
$RMI_{U.S.}$	$CONST$	0.600***	0.533**	0.613***	0.579***
		(0.214)	(0.233)	(0.228)	(0.266)
	$DY_{U.S.,t-1}$	0.449***	0.435***	0.443***	0.468***
		(0.120)	(0.124)	(0.124)	(0.134)
	$RM_{U.S.,t-1}$	0.049**	0.039	0.051**	0.045
		(0.024)	(0.034)	(0.024)	(0.035)
	$RMI_{U.S.,t-1}$	-0.320	-0.243	-0.336	-0.283
		(0.220)	(0.225)	(0.231)	(0.252)
	$IP_{U.S.,t-1}$	0.161	0.139	0.164	0.155
		(0.106)	(0.101)	(0.115)	(0.118)
	$10Y_{U.S.,t-1}$	-0.206***	-0.196***	-0.204***	-0.213***
		(0.050)	(0.057)	(0.051)	(0.062)
RMI_{SUI}	$CONST$	0.349***	0.371***	0.305**	0.324**
		(0.125)	(0.136)	(0.134)	(0.162)
	$RM_{SUI,t-1}$	0.068***	0.067	0.066***	0.065
		(0.021)	(0.090)	(0.021)	(0.080)
	$RMI_{SUI,t-1}$	-0.379**	-0.377	-0.371*	-0.373
		(0.190)	(0.405)	(0.192)	(0.371)
	$TS_{SUI,t-1}$	0.135***	0.107	0.128**	0.102
		(0.051)	(0.124)	(0.050)	(0.116)
	$OIL_{SUI,t-1}$	-0.013	-0.009	-0.012	-0.009
		(0.009)	(0.038)	(0.009)	(0.034)
	ρ		0.657***		0.658***
			(0.060)		(0.051)
	c			0.180	0.200
				(0.227)	(0.354)
	$logL$	-479.037	-442.724	-478.742	-442.341
	AIC	490.037	454.724	490.742	455.341
	$QPS_{U.S.}$	0.458	0.458	0.458	0.458
	QPS_{SUI}	0.454	0.455	0.453	0.455
	psR^2	$0.103^{\dagger}$	$0.094^{\ddagger}$	$0.105^{\dagger}$	$0.096^{\ddagger}$
	$adj.psR^2$	$0.076^{\dagger}$	$0.064^{\ddagger}$	$0.075^{\dagger}$	$0.063^{\ddagger}$
	$SR_{U.S.}$	0.638	0.643	0.641	0.635
	SR_{SUI}	0.627	0.595	0.624	0.603
	$AUC_{U.S.}$	0.620	0.623	0.619	0.622
	AUC_{SUI}	0.651	0.650	0.649	0.646
	$PT_{U.S.}$	8.441***	8.456***	9.997***	7.554***
	PT_{SUI}	5.052**	0.315	4.781**	0.778

Notes: See the notes to Table 5.

Research Papers 2013

CREATES
Center for Research in Econometric Analysis of Time Series

2015-03: Christian M. Hafner, Sebastien Laurent and Francesco Violante: Weak diffusion limits of dynamic conditional correlation models

2015-04: Maria Eugenia Sanin, Maria Mansanet-Bataller and Francesco Violante: Understanding volatility dynamics in the EU-ETS market

2015-05: Peter Christoffersen and Xuhui (Nick) Pan: Equity Portfolio Management Using Option Price Information

2015-06: Peter Christoffersen and Xuhui (Nick) Pan: Oil Volatility Risk and Expected Stock Returns

2015-07: Peter Christoffersen, Bruno Feunou and Yoontae Jeon: Option Valuation with Observable Volatility and Jump Dynamics

2015-08: Alfonso Irarrazabal and Juan Carlos Parra-Alvarez: Time-varying disaster risk models: An empirical assessment of the Rietz-Barro hypothesis

2015-09: Daniela Osterrieder, Daniel Ventosa-Santaulària and Eduardo Vera-Valdés: Unbalanced Regressions and the Predictive Equation

2015-10: Laurent Callot, Mehmet Caner, Anders Bredahl Kock and Juan Andres Riquelme: Sharp Threshold Detection Based on Sup-norm Error rates in High-dimensional Models

2015-11: Arianna Agosto, Giuseppe Cavaliere, Dennis Kristensen and Anders Rahbek: Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX)

2015-12: Tommaso Proietti, Martyna Marczak and Gianluigi Mazzi: EuroMInd-D: A Density Estimate of Monthly Gross Domestic Product for the Euro Area

2015-13: Michel van der Wel, Sait R. Ozturk and Dick van Dijk: Dynamic Factor Models for the Volatility Surface

2015-14: Tim Bollerslev, Andrew J. Patton and Rogier Quaedvlieg: Exploiting the Errors: A Simple Approach for Improved Volatility Forecasting

2015-15: Hossein Asgharian, Charlotte Christiansen and Ai Jun Hou: Effects of Macroeconomic Uncertainty upon the Stock and Bond Markets

2015-16: Markku Lanne, Mika Meitz and Pentti Saikkonen: Identification and estimation of non-Gaussian structural vector autoregressions

2015-17: Nicholas M. Kiefer and C. Erik Larson: Counting Processes for Retail Default Modeling

2015-18: Peter Reinhard Hansen: A Martingale Decomposition of Discrete Markov Chains

2015-19: Peter Reinhard Hansen, Guillaume Horel, Asger Lunde and Ilya Archakov: A Markov Chain Estimator of Multivariate Volatility from High Frequency Data

2015-20: Henri Nyberg and Harri Pönkä: International Sign Predictability of Stock Returns: The Role of the United States