Advanced Studies in International Economic Policy Research
Kiel Institute for the World Economy
Düsternbrooker Weg 120
D-24105 Kiel/Germany

Working Paper No. 451

On the Look-Out for the Bear: Predicting Stock Market Downturns in G7 Countries

by

Christian Friedrich and Melanie Klein

May 2009

On the Look-Out for the Bear: Predicting Stock Market Downturns in G7 Countries*

Christian Friedrich[†], Melanie Klein[‡]

01.05.2009

Abstract

The paper examines the informational content of a series of macroeconomic indicator variables with the intention to predict stock market downturns - colloquially also referred to as 'bear markets' - for G7 countries. The sample consists of monthly stock market indices and a set of exogenous indicator variables that are subject to examination, ranging from January 1970 to September 2008. The methodical approach is twofold. In the first step, a modified version of the Bry-Boschan business cycle dating algorithm is used to identify bull and bear markets from the data by creating dummy variable series. In the second step, a substantial number of probit estimations is carried out, by regressing the newly identified dummy variable series on different specifications of indicator variables. By applying widely used in- and out-of-sample measures, the specifications are evaluated and the forecasting performance of the indicators is assessed. The results are mixed. While industrial production, and money stock measures seem to have no predictive power, short and long term interest rates, term spreads as well as unemployment rate exhibit some. Here, it is clearly possible to extract some informational content even three months in advance and so to beat the predictions made by a recursively estimated constant.

JEL Classification:

Keywords: Bear Market Predictions, Bry-Boschan, Probit Model

*The authors would like to thank Jonas Dovern for helpful and precious comments.

†Kiel Institute for the World Economy. Email: cfriedrich1@t-online.de

‡Kiel Institute for the World Economy. Email: Klein.Mella@googlemail.com

1 Introduction

The recent financial turmoil has caused a full-blown economic crisis with severe output losses and alongside resulted in extensive stock market downturns for G7 countries. The Dow Jones Industrial Average for instance fell by 35 percent during January 2008 and January 2009. The experiences for other advanced economies are similar. At a point like this, routinely the question is asked, whether a stock market downturn could have been avoided or at least foreseen in an earlier stadium. Interest in answering this question is spread over various professions and interest groups - the most prominent ones include: Scholars that are active in the fields of forecasting and finance, policy makers that could ease monetary or fiscal policy early enough to ensure that their impact facilitates the downturn in time and last but not least it also is in the interest of investors to reallocate their portfolios some time in advance - either to another country or to another asset class.

Hence, the question this paper addresses is whether it is possible to predict stock market downturns - colloquially also referred to as 'bear markets' - or at least to identify some hints to be more attentive than usual. Although predicting stock market returns is a difficult business[1], and being overly successful in it would result in everything but an academic piece of work, it could nevertheless be possible to identify some potential harbingers - such as early reactions by the monetary authority or actions taken by companies in financial distress - that frequently precede a stock market downturn.

The methodical approach we apply consists of two steps. In the first step, a modified version of the Bry-Boschan business cycle dating algorithm is used to identify bull and bear markets from the data by creating dummy variable series.[2] This approach was first applied to the stock market by Pagan and Sossounov [2003] in 2002 and subsequently adopted by Biscarri and Perez de Gracia [2002], Gonzalez et al. [2005] and Cunado et al. [2008]. In the second step, we carry out a substantial number of probit estimations, by regressing the newly identified dummy variable series on different specifications of macroeconomic indicator variables. When applying widely used in- and out-of-sample measures, we are then able to evaluate the specifications and to assess the forecasting performance of the selected indicators. We use G7 market data since the recent crisis has emerged in the industrialized world and the high volumes of stocks traded within these countries make this sample more relevant to investors. Furthermore, a good data quality and broad availability enables us to use a sample that for most countries is dating back to the beginning of the 70ies.

[1]However, there is a wide range of empirical literature e.g. Campbell [1985], Chen et al. [1986], Lewellen [2004], Rapach et al. [2005], Campbell and Yogo [2006], Bekaert and Ang [2007].

[2]A different approach to identify the bear markets is the application of regime switching models, e.g. used by Maheu and McCurdy [2000], Ang and Bekaert [2002], Chen [2009]. Prejudicially for our investigation would be that these models do not date the phase of the cycles.

Even though the objective of this paper is not the prediction of the stock returns but the prediction of the bear markets, both approaches are similar as well as the selection of the independent variables. So far, the following macro variables have been investigated and are also partly included in our model. Having a look first on the money stock, Keran [1971] pointes out that monetary actions, measured by changes in the money stock, have no direct impact on stock prices. Instead, it occurs through their effect on inflation and corporate earnings expectations (For further readings see Fama [1981], Pearce and Roley [1983], Kaul [1987], Thorbecke [1997] among others). Regarding the evaluation of the interest rates, Keran [1971] also emphasizes that this measure is a reasonably good empirical explanation of stock price movements. Chen et al. [1986], who investigate nine macro variables regarding their impact on pricing of stocks, conclude that interest rates - the authors include the term structure as difference between the long-term government bonds and the treasury bill - as well as industrial production and changes in the risk premium are significant in explaining expected stock returns. Rapach et al. [2005] who tested for 12 industrialized countries the predictability of stock returns showed also a high forecast ability of interest rates across countries. But in contrary to Chen et al. [1986], who focused on the US market, their results concerning the industrial production and the unemployment rate are not so clear, and though depending on the country (See also for readings on industrial production and unemployment rate: Balvers et al. [1990], Boyd et al. [2001]).

The approach to identify and analyse recessions in the first part of this paper is adopted from Biscarri and Perez de Gracia [2002] and Cunado et al. [2008]. We expand their examination by increasing the number of countries and by carrying out in-sample and out-of-sample probit regressions with the objective to forecast bear markets, whereby the selection of the indicators was influenced by the results of the investigations on stock return predictability. Hence, our overall approach is similar to Chen [2009] who used in contrast to our paper a Markov-switching model for identifing the recession periods in the stock market and who just focused on the US market. His main findings are that term spread and inflation are the most useful predictors of downturns in the US stock market. Furthermore, Chen [2009] concludes when comparing the forecast ability of macroeconomic variables that they have higher explanatory power regarding the prediction of bear markets than of stock returns.

The remainder of this paper is organized as follows. Section 2 is primarily focusing on the modification of the Bry-Boschan business cycle dating algorithm to financial market environment and its application to the stock market index series. As a second task, the hereby obtained identification results undergo a first analysis and thirdly, a selection of stylized facts on bear markets is presented. Section 3 represents the core of the paper and contains the theoretical background for the probit estimations, the estimation results for the in- and out-of-sample case as well as a number of evaluation measures to identify the most promising specifications. Section 4 concludes.

2 Identifying Bear Markets

2.1 Empirical Methodology: The Bry-Boschan Algorithm

Before turning to a detailed description of the dating algorithm applied, some definitions are made. As stated in the introduction, we define bear (bull) markets as those phases of the stock market index series that exhibit a prolonged downward (upward) movement, however such parts can be interspersed with sporadic upward (downward) movements. Each phase is located between two turning points, whereas the starting point of a bear (bull) market is referred to as a peak (trough) and the ending point as a trough (peak). Finally, two phases in a row - i.e. the distance between two troughs or two peaks respectively - constitute a cycle.

The Bry-Boschan's business cycle dating is based on an algorithm created by Bry and Boschan [1971] that identifies the location of turning points in the logged GDP series. As stated above, in this paper, we apply a modified version developed by Pagan and Sossounov [2003] that adjusts the original algorithm to the needs of (logged) stock market index series. This modified algorithm divides each country's stock market index in bear and bull market phases by creating a dummy variable series that takes on the value 1 in case of a bear market and 0 in case of a bull market. In the remainder of this text, we refer to this series as the 'bear market series'.

The modified algorithm consists of several sub-procedures that are applied to the stock market series. The most important ones are presented in the following:

- The main procedure is based on a symmetric sixteen months (i.e. eight months in both directions) moving average filter that is used to derive an upper and a lower bound series to the observed stock market index series. Hereafter, the points of tangency of the stock market index series and the upper bound series become the initial location for peaks, whereas the points of tangency with the lower bound series become the algorithm's first guess for troughs. The corresponding filter procedure in the original algorithm generated symmetric one year (i.e. six months in both directions) moving averages and is part of a series of filter procedures that smooth the observed data several times. To avoid losing the informational content incorporated in outliers, which is of high relevance in the context of stock market data, the modified version only contains the above mentioned sixteen months filter procedure.

- After the initial location of peaks and troughs, a procedure called 'alternate' is launched to ensure that peaks and troughs occur consecutively. In the beginning, a step function that increases by a constant margin when a new peak is reached is generated. Thereafter, the lowest trough within each step (i.e. between two peaks) is identified and kept, while all other troughs in this interval are deleted from the set of potential troughs. By applying an analogous approach to the troughs and creating a second step function, the highest peaks within two troughs are kept and all others are eliminated. As stated above, this procedure ensures that peaks and troughs alternate and is therefore repeated after every change made to the series.

- The next two procedures are concerned with the enforcement of minimum distance requirements for the location of turning points. The so called 'mincycle' procedure ensures that the minimum stock market cycle length is at least sixteen months. This is only a slight modification of the 15-months minimum length criterion in the original algorithm for business cycles. This procedure works as follows: First, all turning points are extracted in a separate series. When all adjacent peaks (or troughs respectively) exhibit a higher distance than the minimum cycle length, both peaks (troughs) are kept and the program continues. In case this rule is violated, the mincycle procedure deletes the lowest of the two problematic peaks (or the highest of the problematic troughs respectively) before moving on.

- The second distance requirement procedure - the so called 'minphase' procedure - is targeted at the enforcement of the minimum phase length. Latter one is set to four months in the stock market context, which is one month less than in the original algorithm. Another modification is given by the introduction of a '20 percent criterion': Since according to Pagan and Soussonov, a significantly lower minimum phase requirement would produce spurious cycles, the authors include the criterion that sharp market movements in either direction overrule the minphase requirement. The minphase procedure works similar to the mincycle routine depicted above:

  1. The turning points are extracted in a separate series again.
  2. The distance between two adjacent turning points (note the difference to the mincylce procedure) is compared with the minimum requirement. If the latter one is not violated, all turning points are kept and the overall program continues.
  3. In any other case, the problematic location is tested for eligibility of the 20 percent criterion. The latter one is defined as a cumulated change in logged stock prices of more than 20 percent in less than four months.
  4. When the location is eligible for the 20 percent criterion, both turning points are kept and the program continues, whereas in any other case, the program deletes the first turning point before moving on.

The modified algorithm itself produces two dummy variable output series. A first series takes on the value of one when a peak is identified and generates a value of zero elsewhere. A second series indicates the locations of all troughs in an analogous manner. Furthermore, there are additional procedures that are used to refine the endpoints after the moving average filter has been applied, to display the results after the identification has been carried out, and finally a procedure that converts the generated (peak- and trough-) dummy variable series in the above defined 'bear market series'. While the former dummy variable series took on the value of one only on their respective turning points, the latter one - the bear market series - takes on the value of one during the bear market phases (and zero elsewhere).

Pagan and Sossounov mostly reason out the adjustments they made by citing from the so called 'Dow Theory', which was propagated by Charles Dow in the early 20th century. According to the authors, this strand of literature introduces the notion of 'bull and bear markets' and defines them formally with a set of general figures based on the market commentators experience at that time.

2.2 Data

Using MSCI stock price indices, we apply the previous algorithm and the subsequently described probit estimations to the G7 countries, which are: Germany, France, Italy, US, UK, Canada and Japan. Since the MSCI index, which is available from 1969 onwards, is published on daily basis, we extracted the last day of the months as proxy for the monthly price index over a period from January 1970 to October 2008. In addition, the natural logarithm of the stock prices is used for the identification of the starting and ending points of the bull and bear markets.

For further investigations on the predictability of bear markets, we decided to choose eight macroeconomic variables in monthly frequency based on previous results in the literatur. In effort to have better time-series properties, the first four of the following variables are included in the model as growth rates. The indicator variables we consider are:

- industrial production growth (the first difference in the log-levels of the industrial production index, PROD);
- inflation rate (the first difference in the log-levels of the consumer price index, CPI);
- narrow money growth (the first difference in the log-levels of the money stock M1);
- broad money growth (the first difference in the log-levels of the money stock M2);
- 3-month interbank rate (IB);
- 10-years government bond yield (GB);

- term spread (the difference between the 3-month interbank rate and the 10-years government bond yield, SPREAD);
- unemployment rate as percentage of the civilian labour force (UNEMP).

The sample was intended to cover the time from January 1970 to September 2008. However, for some variables particular periods were not available:

1. UNEMP for Italy (only available from November 1979 to July 2008), France (January 1978) and the UK (Dezember 1970);
2. the money stock M1 for Italy (first available in January 1980) and France (December 1977);
3. the money stock M2 for the UK (July 1982), France and Italy (January 1980)

The data is obtained via Datastream.

2.3 Identification Results

2.3.1 Turning Point Analysis

The following chapter analyses the results obtained through the application of the Bry-Boschan business dating algorithm to the sample countries. Over the time period 1970 - 2008, we obtained a large number of turning points (around 20 per country). In the following, we enlarge on the most important bear markets (peak $\longrightarrow$ trough) in recent history, and furthermore explain their impacts on the seven countries in greater detail.

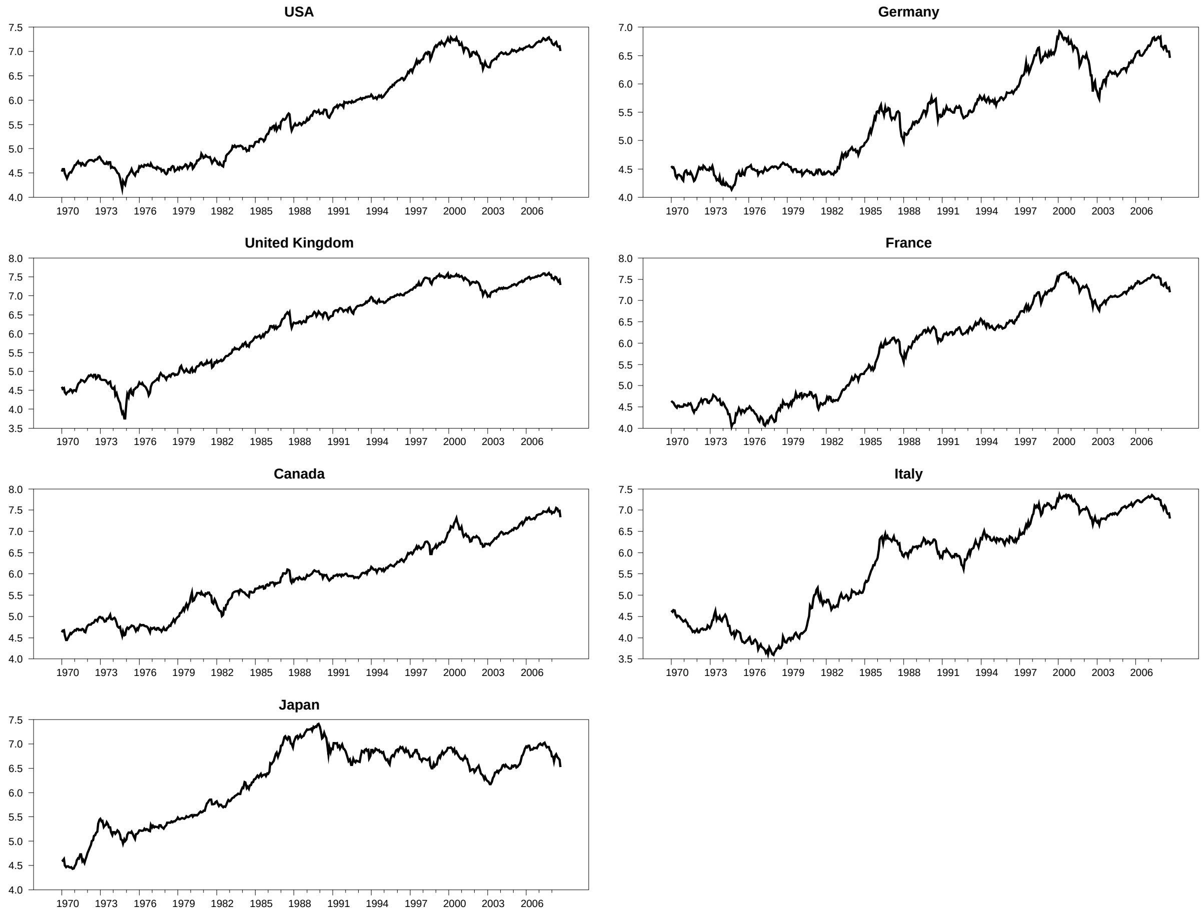
Table 1: Correlation matrix for the MSCI series of all countries

	Germany	France	Canada	Japan	US	UK	Italy
Germany	1						
France	0.98	1					
Canada	0.95	0.91	1				
Japan	0.48	0.61	0.5	1			
US	0.98	0.97	0.93	0.54	1		
UK	0.98	0.97	0.95	0.54	0.99	1	
Italy	0.97	0.97	0.93	0.59	0.98	0.99	1

Before looking at the graphs of figure 1, it is interesting to see the correlations between the seven MSCI series (see table 1).[3] The most striking feature of this table is that the correlations between all countries are very high with the exception of Japan which exhibits strong lower values. Furthermore, Canada shows slightly less correlations to the other European countries and to the US.

[3]Mean and standard deviations of the series can be found in the appendix (table 10).

Figure 1: Log stock prices 1970/01-2008/10 (MSCI series are in the domestic currency)



In the following we concentrate on three key occasions between 1970 and 2008 - for two of them the US market was the trigger. The *first* phase is the time period of the seventies. The origin for the bear market - it began first in Germany (1972:07) and Great Britain (1972:8) followed by the US in December 1972, Japan (1973:01), France in April, Italy in June and Canada in October - was the first oil crisis in 1973 and the dissolving of Bretton Woods between 1971 and 1973. The former was caused by an oil embargo of the OPEC as protest against the US support of Israel during the Yim-Kuppor War. Even though only France, the Netherlands and the US were affected directly by the embargo, the other countries were afflicted with the extreme rising oil price. During the war in 1973 the oil price increased from \$18 to approximately \$41 per barrel and rose to the pike of \$106 in April 1979 (Datastream). Compared to the first oil crisis, the second one in 1979 had not such strong negative implications on the stock prices. In contrast, Canada's and Italy's stock prices increased since 1978 and found their peaks in 1981, the highest values since 1970. Observing the seventies, Canada, Japan and Germany suffered less compared to the rest and underwent only a 'mild' downturn as reflected by their losses during the bear market of the first oil crisis. Germany's index value decreased by 10.1 percent during its downturn, going from 1972:07 to 1974:09, followed by Canada and Japan which lost 10.8 and 10.3 percent respectively, even though Canada underwent the shortest contraction phase between the seven countries with 11 month (from 1973:10 to 1974:09).[4] An explanation for this finding could be that Canada as a net exporter of oil is largely independent of the world oil supply. To find conclusive reasons for Japan's low sensitivity during this time, it is more concealed. A feasible explanation is its relative isolation from Europe, in 2007 only 14.8 percent of Japan's exports and 10.5 percent of Japan's imports were traded with the EU-27. Its main trading partners are the US and Asian countries. The second one could be its strengthened export sector which turned the trade balance in a surplus since 1965 and thus made the Japanese economy more robust.[5] Approximately in the beginning of the eighties, the stock prices of all countries recovered their initial level of the time before the first oil crisis, likewise Italy which has spent 54 month - the longest contraction phase within the whole analysis - in this bear market (1973:06 - 1977:12).

The *second* noticeable trough is identifiable through the strong fall in the stock index in the last two month of 1987 (see figure 1). Starting on October 19^{th} 1987 in the US, known as the 'Black Monday', the crash spread fast to the UK, Canada, Japan, followed by Germany, France and Italy. During this day in October, the Dow Jones Industrial Average fell by 508 points and caused the biggest crash on a single day in the history of the NYSE.[6] The distinctive characteristics of this bear market are: 1) it was short and 2) the level of per month losses is significantly higher than during the other bear markets, especially in the US and UK. In this second contraction phase two groups emerged, excluding Japan. The first one with the US, UK and Canada show a very short bear market (four, three, and five months) with similar losses. The stock market index decreased by 6.7 percent in the US,UK and 5.5 percent in Canada. On the contrary within the second group including Germany, France and Italy the markets are longer (21, 17, and 21 months) and the losses are higher (Germany 13.1 percent, France 10.3 percent, Italy 9.2 percent). Examining the per month losses the picture is different. The US with 1.7 percent and the UK with 2.2 percent show the highest monthly loss between 1970 and 2008.

The *third* downturn was initiated by the bust of the dot-com bubble whereas in the US the crisis were precipitated. A special characteristic for this downturn are a uniform phase length and a low level of monthly losses across all countries. Certainly, the incident on September 11^{th} 2001

[4]The residual losses per country are (the length of the bear market is in bracket): US 15.9 percent (21 months), France 18.6 percent (17 months), UK 31.2 percent (27 months), Italy 29 percent (54 months).

[5]The exports grew from 1965 to 1975 by 81.7 percent (Datastream).

[6]This crash was generated on the one hand through economic factors and on the other hand through technical and systemic problems. The former included among others a weak dollar, fear of inflation and a rapid growing of the American trade deficit. The latter problem was dominated by the former stock trading system. The trading pensum in these days was overwhelmed and could not be executed, in particular without internet and with the technology (e.g. computers) at that time. For further readings: Samuelson [2007].

was a strengthened effect which extended the bear market phase. Starting around the turn of the millennium and ending in early 2003, the average duration was 32 months. The losses per month varied between 0.2 and 0.6 percent (US: 0.33, UK: 0.2, Canada: 0.4, Japan: 0.3, Germany: 0.6, France: 0.4, Italy: 0.4). Another time Japan is a special case. It also underwent this bear market with approximately 37 months and a per month loss of 0.3 percent but did not experience the consistently increasing trend during the nineties compared to the other six countries. Slumping in its so far biggest economic crisis in 1990, whereby the downtrend of the stock prices began, the Japanese MSCI index has not recovered its value of 1990 up to now.

2.3.2 Stylized Facts on Bear Markets

After the peaks and troughs of the seven countries have been identified, we analyze the main characteristics of their bear markets. For this purpose we consider five properties:

1. The average duration of the bear markets (Duration).
2. The average amplitude of the markets for each country (Amplitude). This measure is defined as the percentage change of the log stock prices from one turning point to the following.
3. The approximation for the cumulative movements of the stock prices within the phase from peak to trough, relatively to the previous peak (Cum. movements).
4. The shape of the bear market, in particular the divergence from a triangle. Comparable to Pagan and Sossounov [2003], we will refer to this as the 'excess' index.
5. The last value is the percentaged fraction of all contractions, that exhibited a reduction of more than 20 percent between two turning points (Change $\geq$ 20 %).

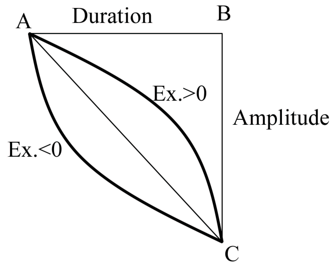
The initial point for these calculations is the computation of a dummy variable B_t, which turns one when the stock market is in a recession and takes on the value zero when the stock market follows an expansion phase. Having this, the total time spent in an bearish market is $\sum_{t=1}^{T} B_t$. Calculating the number of troughs:

$$NT = \sum_{t=1}^{T} B_t(1 - B_{t+1}) + 1 \tag{1}$$

and dividing this value by the time being in a bear market, yields the average duration of the contraction phases per country.

For a better understanding, figure 2 plots a stylized bear market with point A being a peak and C symbolizing a trough. If indeed the shape of a bear market follows a triangular, the data generating process is a random walk. Thus the best prediction for the next period would be the

Figure 2: Stylized Bear Market



Source: Hading/Pagan [2002], p.370

current stock price, which implies that in this case a forecast by exogenous variables should show worse results.

Based on the fact that the height of the triangle is the amplitude and the duration is part of the hypotenuse, the area of the triangle and so an approximation to the cumulative movements in the stock prices within the phase peak-trough must be:

$$Cum_i = 0.5(Dur_i \cdot Amp_i) \tag{2}$$

Because this triangle just drafts the real aspect of a bear market and it might be that the Cum_i differs from the actual cumulative movements, C_i. Thus the ‘excess’ index is needed, since the actual path through the phase may not be well approximated by a triangle, which is:

$$E_i = (C_i - 0.5Amp_i - Cum_i) \cdot Dur_i \tag{3}$$

To obtain the figure C_i per country, the numbers in the series of the log stock prices for the bear markets (the prices during the bull markets are set to zero) are added one after the other. Furthermore, we divided the sum of the resulting series by the number of troughs.[7]

Table 1 gives the results of these measures for the G7 countries. The first four rowes are targeted on evaluating the shape of the bear markets, while the last two values are measures for stock market volatility. By looking at the average duration of the bear markets, it is striking that in five cases - with the exception of Canada and Italy - the corresponding values lie between 14 and 19 months. This result is quite close to the claim of Hamilton in 1921, who assumed that over a period of 25 years the bear markets last on average 17 months (Rhea [1932] p37). Italy exhibits the largest bear markets with a duration of 22 months and an average capital loss of 54 percent. As a next measure, the cumulated movement values are calculated. While the values themselves have no interpretation,

[7]For a more detailed description and the underlying formulae see Pagan and Sossounov [2003] and Biscarri and Perez de Gracia [2002].

they are needed to derive the excess index, which in turn contains some interesting information. Based on the excess index values in table 1, the bear markets of Japan, Italy and the UK differ significantly from the triangular shape. Thus it might be possible to predict their recessions using macroeconomic indicator variables. The trend of the deviation from the stylized bear market is also noted by Pagan and Sossounov [2003] and Harding and Pagan [2002]. The former emphasize that for their sample period from 1945 to 1997 this behaviour became more emphatic over time for the US stock market. The last variable in the table proves that the decision to include the 20 percent criterion during the Bry and Boschan identification process was needed. All countries show a high number whereas the UK outperforms the rest.

Table 2: Bear market characteristics for all countries

	Germany	France	Canada	Japan	US	UK	Italy
Duration	15	14	11	19	15	17	22
Amplitude (in percent)	-30	-45	-31	-41	-36	-55	-54
Cum. movements	-2.11	-3.09	-2.05	-4.80	-3.09	-4.00	-7.48
Excess index(in percent)	1.22	2.1	-1.19	-3.39	-1.7	5.01	-5.38
Volatility	0.53	0.72	0.45	0.83	0.68	0.86	1.13
Change (in percent) $\geq$ 20 %	50	63	55	60	50	100	80

Finally, compared to Biscarri and Perez de Gracia [2002], our index for volatility turns out to be much higher. In their analysis, Italy exhibits a value of only 0.85 for the period from 1957 to 1998. Although this difference is significant, Italy is the most volatile market in their sample as well. Furthermore, as indicated by table 1 and noted by Biscarri and Perez de Gracia [2002] as well, the longer the duration and the higher the amplitude are, the higher is the volatility during the bear markets.

3 Predicting Bear Markets

3.1 Empirical Methodology: The Probit Model

A probit model is a regression model that contains a dependent variable which lies strictly between 0 and 1 and hence can be interpreted as a probability. The population probit model is given by

$$Pr(Y = 1|X_1, X_2, ..., X_k) = \Phi(\beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + ...\beta_k \cdot X_k) \tag{4}$$

where X_k are the regressors, β_k are the corresponding probit coefficients, and Φ is the cumulative standard normal distribution.[8] Since the probit coefficients are difficult to interpret, usually the

[8]See any standard textbook for econometrics: e.g. Stock and Watson [2008] p. 392.

predicted probabilities - i.e. the fitted values of the estimated model - are calculated based on the earlier obtained coefficients.

In our analysis, the dependent variable is given by the bear markets series - i.e. the series that was obtained previously by applying the modified Bry-Boschan business cycle dating algorithm to the stock market index series - and the exogenous variables are represented by indicator variables that are assessed for their forecasting power. Since forecasting stock market returns is a tedious business, we apply a huge variety of specifications. To structure our approach in the most traceable way, we first create a unified framework that contains a large number of specifications and modify it several times. The structure of this framework is the following:

In the first specification for each country, we include the first 12 lags of all indicator variables.[9] We do not include any contemporary values, since they are not available at the time the forecast is made. In the second specification, we again include the same number of lags for all indicator variables. But now, we apply a selection procedure to the regression output that eliminates all insignificant regressors and re-estimates the model until all regressors are significant at the 10 percent level at least. The following 16 specifications are based on an analogous framework but on a rather disaggregated level. The bear market series is now regressed on the first 12 lags of each of the eight indicator variables. Hereafter, the aforementioned selection procedure is applied again and the bear market series is regressed only on the significant lags of each of the indicator variables.

Before turning to the variations of the above depicted framework, a brief introduction into the most prominent evaluation concepts of forecasting models is presented: The 'in-sample' and the 'out-of-sample' analysis. The in-sample analysis serves as a preliminary assessment of the forecasting performance of indicator variables. The defining characteristic of the in-sample analysis is that each specification is estimated over the whole sample range. Since all available observations are used to estimate the model, no data points are left to evaluate the predictions made. This drawback is resolved by conducting an out-of-sample analysis. Here, the sample is divided into two parts. Whereas the first part of the sample (initial estimation sample) is used to estimate the model from the data on the exogenous variables and the second part (out-of-sample period) is used to evaluate the models predictions against the actual realizations of the dependent variable - in our case, the bear markets series. Since the exogenous variables are unkown at the time the forecast is made, the out-of-sample analysis should be consulted to evaluate the forecasting performance in the first place (Ashley et al. [1980], Meese and Rogoff [1983]). In setting up the out-of-sample analysis, it is crucial to decide on how the out-of-sample values are generated. They can be obtained by either a rolling window scheme that contains a fixed number of observations and is moved forward over time

[9] In the first two specifications, we exclude the SPREAD variable. Since including it together with the utilization of its defining interest rates would result in perfect multicollinearity.

or by a recursive estimation scheme, where the initial estimation sample is increased by a specific number of observations each period. In this paper, we apply the latter approach with one step ahead forcasts each period.

Having this in mind, we continue with a description of the cases in which the specification framework or its modifications are used: Firstly, the framework is applied as described above and serves as the baseline set-up for the in-sample analysis. Secondly, when turning to the out-of-sample analysis the lag length is increased to three which serves as the new baseline. Finally, by modifying the starting point of the evaluation period - i.e. the point where the out-of-sample forecast starts - a third and fourth modification is generated.

To systematically evaluate the large number of estimation results, we employ measures that are frequently used in the forecasting literature to assess the prediction power of indicator variables. But before turning to the description of more sophisticated measures, a fairly simple method of examination should be mentioned. As a first approach, the predicted probabilities are plotted against the time. When, in addition, the earlier obtained bear market series is added to the graph, a first visual impression of the models' forecasting power is gained. Especially models that perform particularly poor in this exercise can be identified at an early stage.

Although it only can be applied to the in-sample case, the so called pseudo R^2 is a second measure to evaluate probability forecasts (Estrella and Mishkin [1998]). It has to be calculated seperately for each specification via the following formula:

$$psR^2 = 1 - (\frac{LL_u}{LL_c})^{-2/T \cdot LL_c} \tag{5}$$

where LL_u denotes the log-likelihood of the unconstrained model and LL_c denotes the log-likelihood of the constrained model (i.e. the constant only). The values of the pseudo R^2 range between 0 and 1, where a higher value indicates a better fit.

The evaluation measures number three and four are the Quadratic Probability Scores (QPS) as suggested by Brier [1950] and the Logarithmic Probability Scores (LPS), both calculated on basis of the predicted probabilities and can be used for assessing the in-sample and out-of-sample results. The QPS is obtained by calculating two times the sum of the squared differences between the predicted probabilities and the realizations of the bear market series and dividing the result by the sample size. The entire formula is given by:

$$QPS = \frac{1}{T} \cdot \sum_{T}^{t-1} 2 \cdot (P_t - R_t)^2 \tag{6}$$

where P_t denotes the probability forecast for time t and R_t denotes its ex-post realization. The fit of the model is high, when the QPS takes on a small value. In similar fashion, the LPS can be computed:

$$LPS = -\frac{1}{T} \cdot \sum_{T}^{t-1} (1 - R_t) \cdot ln(1 - P) + R_t \cdot ln(P_t) \quad (7)$$

where P_t denotes the probability forecast made for time t and R_t is the realization of the corresponding sample point. The lower bound of both measures is 0 and indicates the highest model fit.

To effectively evaluate a specification of interest with the last two measures, a benchmark model is derived. For both the in-sample and the out-of-sample case, such a benchmark is given by a constant-only-specification that is evaluated by QPS and LPS as well. In the in-sample case, the constant is represented by the average value of the dependent variable calculated over the whole sample range. In the out-of-sample case, the constant is obtained by a stepwise determined dependent variable average over the estimation period - starting with the initial estimation sample average and increasing its range by one oberservation in each of the following periods. As a rule of assessment, it can be stated that as long as the specification of interest produces a lower QPS or LPS than the benchmark model, the indicator variables of the specification exhibit some forecasting power and thus, the model is helpful in predicting bear markets.

To formally assess the significance of the difference between the proposed forecasting model and the benchmark constants, a series of Diebold-Mariano tests is applied (Diebold and Mariano [1994]). These tests are conducted in two versions. Firstly, the null hypothesis that the foreceasting model is as good as the constant is tested against the alternative hypothesis that the forecasting model performs signifcantly better. Secondly, the opposite case is considered as well with the null stating that both models are equal and the alternative that the constant model performs better. Since both cases usually yield the same results, we only report the more interesting first case. The Diebold-Mariano test is evaluated by its p-value, indicating a rejection of the null - i.e. stating that the forecasting model performs significantly better - by a p-value smaller than five percent.

3.2 Estimation Results

The presentation of the results is divided into three parts. Firstly, we examine the fit via an in-sample analysis. Secondly, we assess the prediction power of the indicator variables by running the out-of-sample analysis. Thirdly, we carry out a large number of robustness checks for the out-of-sample case.

3.2.1 In-Sample Analysis

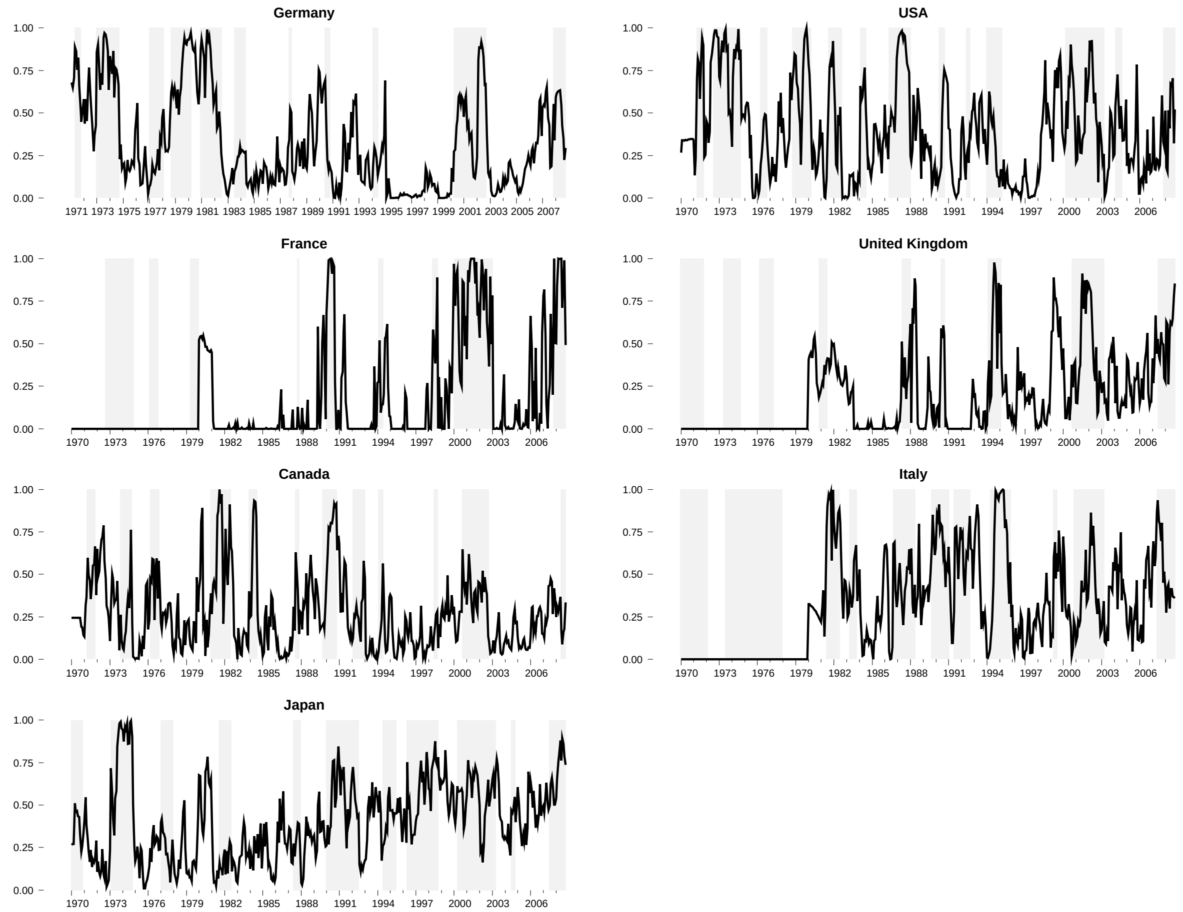
As stated above, the in-sample analysis serves as a preliminary assessment of the indicator variables' forecasting performance and is entirely based on an ex-post perspective. Hence, the specifications are estimated over the maximum sample range and the minimum lag length is chosen abitrarily. For Canada, Germany, Japan, and the US where data is available over the whole sample, the estimation is carried out from January 1970 to September 2008. Since the data on indicator variables for France, Italy, and the UK is not entirely available from January 1970 onwards and for each of these three countries at least one indicator series starts not before the late 70ies, we set their in-sample start to January 1980. Furthermore, the sample for Italy ends due to a lack of recent unemployment data already in June 2008.

As already mentioned in subsection 3.1., we first estimate the baseline set-up for each country, starting with a minimum lag length of one. After having obtained the coefficients, the predicted probabilities can be calculated. Plotted against the time and supplemented by the bear market series, these predicted probabilities generate a first impression of the model fit (see figure 3). At a first glance a number of bear markets were predicted fairly well (e.g. 1973 in Japan, 1986/7 in the US, 1994 in Italy and the UK and after 2000 in Germany and France). By contrast, e.g. the predicted probabilities in the case of Canada from 1991 onwards, rather seem to fluctuate independently of the earlier identified bear markets series.

As a typical in-sample measure, the pseudo R^2 is applied next to evaluate the different specifications (see table 3). The results indicate that the specifications with the highest pseudo R^2 and hence, the highest prediction power relatively to the constant-only-case, are obtained when all variables are included. The resulting pseudo R^2 values are different across countries and range from a low 0.26 for Canada to a high 0.65 for France. When in addition, the lag selection procedure is applied to the all-variable model for each country, slightly smaller values are obtained. This stems from the fact that the pseudo R^2 does not include a correction term for the number of regressors included in the model and hence, the model with the higher number of regressors is always found to be better.

In a next step, the forecasting performance of the single variable predictions is examined. The values of the pseudo R^2 are mostly located in the range between zero and small double digit values which is significantly lower than in the all-variable case. This finding indicates that no single variable has the prediction power to compete with the whole set of variables. However, it rather can be infered that PROD and M1 have a low informational content, especially when the lag selection procedure is applied and basically for most countries all lags of these two variables are thrown out. A similar case is stated for the pseudo R^2 values of M2 and CPI, but their after-lag selection

Figure 3: In-sample recession probabilities - all variables - 1 lag



Note: shaded part = actual bear markets; plotted line = fitted values of the probit estimation. The time period for the in-sample analysis of France, the United Kingdom and Italy starts in January 1980.

procedure specification includes a positive number of regressors. Better prediction qualities seem to be incorporated in the interest rate variables, such as GB, IB and SPREAD. Here, the pseudo R^2 values are frequently located in the lower two-digit range, such as 0.1 to 0.2, without any notable reduction when the lag selection procedure is applied.

In general, when the Quadratic Probability Scores (QPS) and Logarithmic Probability Scores (LPS) are examined a similar picture emerges (see table 4 and 5). Here, the figures from the first 18 columns are evaluated against the scores of the constant in column 19. As soon as the QPS or LPS values of an indicator variable specification is smaller than the QPS or LPS values of the constant, the respective specification can be seen as useful for prediction. The all-variable specification again has the highest prediction power across all countries - with all indicator values clearly being lower than the values of the constants.

Extending the country analysis to the single indicator variable case, the best predictions can be made for Japanese stock market downturns - here, basically every variable beats the constant. The

Table 3: In-sample results - pseudo R^2 - 1 lag

	All variables		PROD		CPI		M1		M2	
	N	S	N	S	N	S	N	S	N	S
Germany	0.38	0.30	0.03	0.00	0.12	0.10	0.04	0.00	0.02	0.00
France	0.65	0.54	0.05	0.00	0.10	0.09	0.02	0.00	0.05	0.03
Canada	0.26	0.16	0.02	0.00	0.03	0.02	0.04	0.03	0.04	0.03
Japan	0.28	0.19	0.09	0.03	0.08	0.06	0.10	0.10	0.12	0.09
USA	0.36	0.31	0.05	0.03	0.07	0.07	0.08	0.07	0.06	0.05
UK	0.48	0.31	0.07	0.00	0.08	0.02	0.07	0.00	0.17	0.15
Italy	0.34	0.23	0.05	0.00	0.07	0.00	0.07	0.00	0.05	0.00

	IB		GB		SPREAD		UNEMP	
	N	S	N	S	N	S	N	S
Germany	0.15	0.14	0.15	0.14	0.11	0.07	0.14	0.12
France	0.12	0.11	0.18	0.18	0.04	0.02	0.19	0.18
Canada	0.06	0.04	0.09	0.08	0.06	0.05	0.05	0.02
Japan	0.14	0.13	0.10	0.09	0.09	0.08	0.07	0.01
USA	0.05	0.03	0.05	0.02	0.06	0.02	0.05	0.05
UK	0.11	0.07	0.13	0.12	0.12	0.09	0.10	0.07
Italy	0.09	0.08	0.16	0.16	0.11	0.10	0.06	0.01

Note: 'N', 'S' are non-selected and selected. In the selected estimation are the significant parameters of the 12 lags non-selected estimation included.

Table 4: In-sample results - QPS - 1 lag

	All variables		PROD		CPI		M1		M2	
	N	S	N	S	N	S	N	S	N	S
Germany	0.29	0.32	0.43	0.43	0.39	0.39	0.43	0.43	0.43	0.43
France	0.13	0.15	0.32	0.32	0.31	0.31	0.32	0.32	0.31	0.32
Canada	0.30	0.33	0.39	0.39	0.39	0.39	0.38	0.39	0.38	0.39
Japan	0.39	0.42	0.47	0.48	0.48	0.48	0.46	0.47	0.46	0.46
USA	0.33	0.35	0.46	0.47	0.45	0.45	0.44	0.45	0.46	0.46
UK	0.28	0.28	0.36	0.36	0.34	0.35	0.35	0.36	0.35	0.36
Italy	0.36	0.40	0.47	0.49	0.47	0.49	0.47	0.49	0.48	0.49

	IB		GB		SPREAD		UNEMP		Const
	N	S	N	S	N	S	N	S	
Germany	0.38	0.38	0.37	0.38	0.39	0.40	0.39	0.39	0.43
France	0.30	0.30	0.28	0.28	0.31	0.32	0.25	0.25	0.32
Canada	0.37	0.37	0.36	0.36	0.37	0.37	0.38	0.38	0.39
Japan	0.45	0.45	0.47	0.47	0.47	0.47	0.48	0.48	0.49
USA	0.46	0.46	0.46	0.47	0.45	0.46	0.46	0.46	0.47
UK	0.34	0.34	0.33	0.33	0.33	0.33	0.34	0.34	0.36
Italy	0.46	0.47	0.43	0.43	0.45	0.46	0.48	0.48	0.49

Note: 'N', 'S' are non-selected and selected. In the selected estimation are the significant parameters of the 12 lags non-selected estimation included. 'Const' is the Constant which is used as a benchmark for evaluating the forecasting abilities of the exogenous variables.

lowest number of indicator variables that exhibit prediction power are found in the case of Canada and to some lower extent also for Germany and the UK. But nevertheless, even for Canada, more than half of the indicator variables have a QPS and LPS that beat the corresponding measures of the constant.

Table 5: In-sample results - LPS - 1 lag

	All variables		PROD		CPI		M1		M2	
	N	S	N	S	N	S	N	S	N	S
Germany	0.44	0.48	0.63	0.62	0.57	0.58	0.62	0.62	0.62	0.62
France	0.20	0.26	0.49	0.50	0.46	0.47	0.50	0.50	0.48	0.50
Canada	0.46	0.51	0.57	0.58	0.57	0.58	0.57	0.57	0.57	0.57
Japan	0.57	0.61	0.66	0.67	0.67	0.68	0.65	0.66	0.65	0.65
USA	0.49	0.52	0.65	0.66	0.64	0.65	0.64	0.64	0.65	0.65
UK	0.41	0.43	0.55	0.54	0.52	0.54	0.53	0.54	0.53	0.54
Italy	0.52	0.58	0.67	0.68	0.67	0.68	0.66	0.68	0.68	0.68

	IB		GB		SPREAD		UNEMP		Const
	N	S	N	S	N	S	N	S	
Germany	0.56	0.56	0.56	0.56	0.58	0.59	0.57	0.58	0.62
France	0.45	0.45	0.42	0.42	0.49	0.49	0.42	0.42	0.50
Canada	0.56	0.56	0.54	0.55	0.56	0.56	0.56	0.57	0.58
Japan	0.64	0.64	0.66	0.66	0.66	0.67	0.67	0.67	0.68
USA	0.65	0.66	0.65	0.66	0.65	0.66	0.65	0.66	0.67
UK	0.51	0.52	0.50	0.51	0.51	0.51	0.51	0.52	0.54
Italy	0.65	0.66	0.62	0.62	0.64	0.65	0.67	0.68	0.68

Note: 'N', 'S' are non-selected and selected. In the selected estimation are the significant parameters of the 12 lags non-selected estimation included. 'Const' is the Constant which is used as a benchmark for evaluating the forecasting abilities of the exogenous variables.

Table 6: Diebold Mariano Test of the in-sample results - 1 lag (p-values)

	All variables		PROD		CPI		M1		M2	
	N	S	N	S	N	S	N	S	N	S
Germany	**0.00**	**0.00**	0.36	0.50	**0.00**	**0.00**	0.08	0.50	0.39	0.50
France	**0.00**	**0.00**	0.86	0.90	0.80	0.47	0.97	0.89	0.42	0.74
Canada	**0.00**	**0.00**	0.11	0.50	0.12	0.17	0.04	0.13	0.07	0.11
Japan	**0.00**	**0.00**	**0.03**	0.21	0.08	0.38	**0.01**	**0.01**	**0.00**	**0.01**
USA	**0.00**	**0.00**	**0.02**	0.07	**0.00**	**0.01**	**0.00**	**0.00**	**0.02**	**0.02**
UK	**0.00**	**0.00**	0.58	0.51	0.88	0.82	0.26	0.14	0.38	0.88
Italy	**0.00**	**0.00**	0.09	0.92	0.10	0.27	0.07	0.17	0.46	0.92

	IB		GB		SPREAD		UNEMP	
	N	S	N	S	N	S	N	S
Germany	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
France	0.05	**0.04**	**0.01**	**0.01**	0.39	0.44	**0.00**	**0.00**
Canada	**0.01**	**0.01**	**0.00**	**0.00**	**0.01**	**0.02**	**0.04**	0.10
Japan	**0.00**	**0.00**	**0.02**	**0.04**	**0.03**	0.06	0.11	0.17
USA	**0.02**	**0.04**	**0.03**	0.13	**0.01**	0.05	0.07	0.06
UK	**0.02**	0.05	**0.01**	**0.01**	**0.02**	**0.04**	**0.01**	**0.00**
Italy	**0.01**	**0.01**	**0.00**	**0.00**	**0.00**	**0.00**	0.12	0.25

Note: 'N', 'S' are non-selected and selected. H0: constant = forecasting model; H1: forecasting model is better than the constant; The null hypothesis is rejected by a p-value smaller than five percent.

When instead of the countries the indicator variables are evaluated most of them are found to have at least some informational content. With the interest rate measures leading the way, also UNEMP and CPI have significantly higher QPS and LPS than the corresponding constants. The forecasting performance of PROD, M1 and M2 is significantly lower but in more than half of the cases still higher than the constant.

The results of the Diebold-Mariano tests broadly support these findings (see table 6). The all-variable cases exhibit for all countries a p-value of zero indicating that the null of forecasting model and constant being equal is clearly rejected. Remaining in the country perspective but turning to the single variable cases, it emerges that the results here differ somewhat from the previous measures. The Diebold-Mariano framework indicates the best predictability for Germany, USA and UK which all have the double digit number of below-5-percent p-values. This indeed differs from the QPS and LPS measures since their results suggested that Japan was the country with the best and among others Germany and UK, the countries with the least predictable stock markets. When taking on the perspective of the single indicator variables the results are more similar. It can be seen that short (IB) and long term (GB) interest rates as well as their difference, the SPREAD variable, exhibit the lowest p-values among all variables for most of the countries. Hence, this indicates that the models with the best forecasting performances are based on interest rates measures.

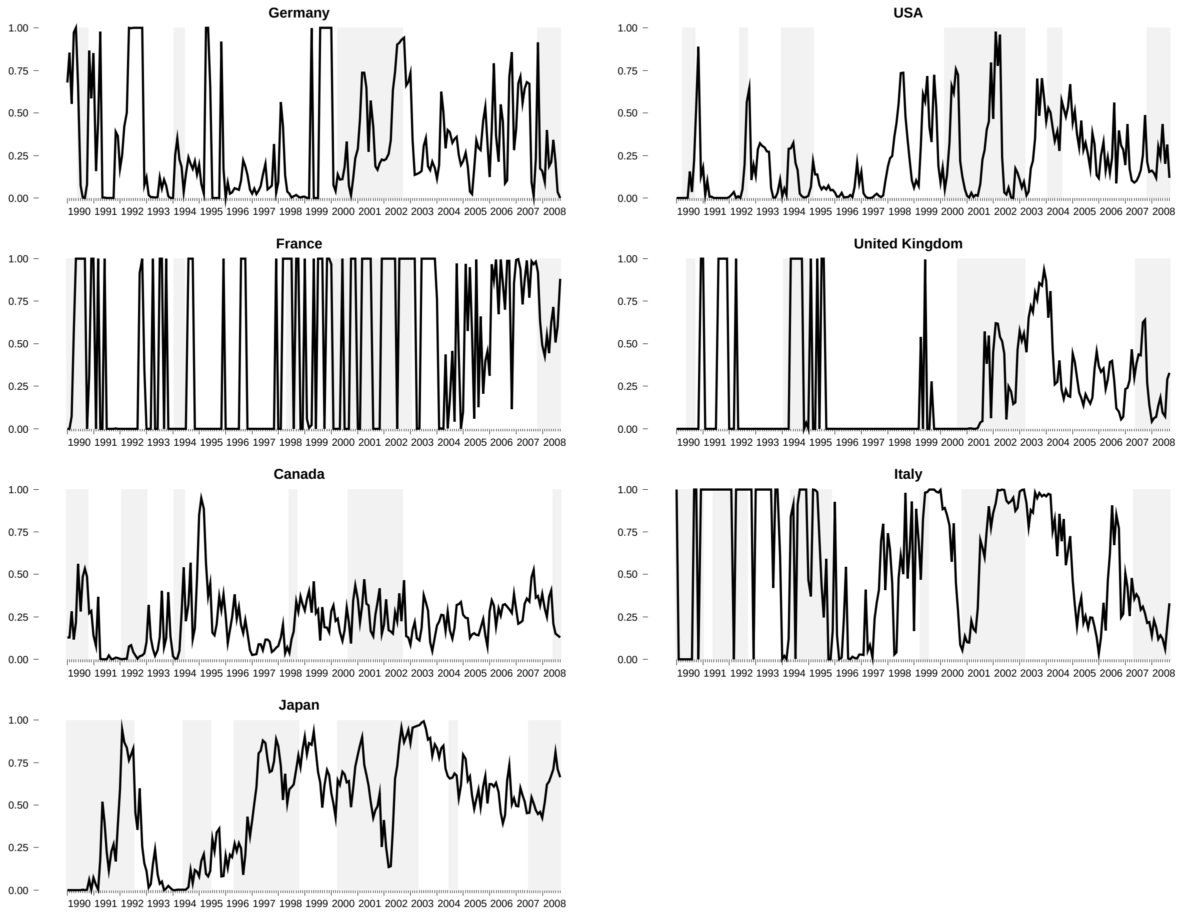
3.2.2 Out-of-Sample Analysis

To evaluate the prediction power of the indicator variables under more realistic conditions, we employ an out-of-sample analysis. From now on, the sample is divided in two parts: An estimation period, where the coefficient estimates are obtained and an out-of-sample period, where the predicted probabilities are computed and evaluated against the realized data points. In the baseline set-up January 1990 was selected as the date that separates both periods. This point in time is changed in the robustness section to 1995 and 2000 with the intention to cope with the trade-off between a longer estimation period, which delivers more reliable coefficients and a longer evaluation sample that facilitates the assessment of the forecasting power. To furthermore construct realistic forecasting conditions, we set the minimum lag length to three lags, since all indicator variables used in this analysis are publicly available to forecasters within a three month range.

For the preliminary analysis, again the predicted probabilities are graphed against the time (figure 4). At a first glance, it seems that the fitted values meet the bear market series less accurately than in the in-sample case. As an example the case of Canada might serve, where the probabilities drift persistently to the right without any change during recession times. Furthermore, the 0-1-pattern in France and Italy as well as to some extent in the UK is striking. This is due to the fact that the missing indicator variables in the beginning of the estimation sample are still feeding through after some time.

Since in the out-of-sample analysis, the pseudo R^2 ist not applied, we immediately turn to

Figure 4: Out-of-sample recession probabilities - all variables - 3 lags



Note: shaded part = actual bear markets; plotted line = fitted values of the probit estimation

the QPS and LPS analysis. In the all-variable case for all countries, the corresponding QPS and LPS measures are consequently lower than the constant indicating that there is no improvement by using all indicator variables for forecasting (see table 7 and 7). Closest to the constant are the variable values of Canada (0.46 (QPS), 0.79 (LPS)) with constants of 0.39 and 0.58 respectively. In contrast, the lowest prediction power is exhibited in the all-variable specifications for France and Italy, with exemplary values for the French measures of 0.76 (QPS) and 12.15 (LPS) and corresponding constants of 0.43 and 0.63.

When examining the country perspective for the single variable case, the following picture emerges. According to QPS and LPS, the best predictions can be made for the stock markets downturns in the UK, Italy and Japan - here, a large number of indicator variables have a lower QPS and LPS than the constant. In constrast to these findings, the US market shows no signs of predictability at all, with only one indicator variable being useful (M2 in the LPS results). For the markets of Germany, France and Canada the results are ambiguous, since in each case at least two indicator variables have more predictive power than the constant. But the downturns are less predictable than in the UK and Japan.

Table 7: Out-of-sample results 1990 - QPS - 3 lags

	All variables		PROD		CPI		M1		M2	
	N	S	N	S	N	S	N	S	N	S
Germany	0.57	0.42	0.39	0.38	0.43	0.46	0.43	0.43	0.51	0.48
France	0.76	0.60	0.47	0.47	0.45	0.44	0.46	0.43	0.49	0.45
Canada	0.46	0.42	0.40	0.39	0.41	0.41	0.40	0.39	0.39	0.39
Japan	0.70	0.67	0.63	0.64	0.66	0.64	0.62	0.60	0.63	0.61
USA	0.58	0.53	0.51	0.50	0.47	0.47	0.49	0.49	0.48	0.46
UK	0.54	**0.37**	0.41	0.41	0.42	0.42	0.41	0.41	0.49	0.48
Italy	0.77	0.61	0.50	0.50	**0.49**	**0.49**	0.51	0.51	0.55	0.52

	IB		GB		SPREAD		UNEMP		Const
	N	S	N	S	N	S	N	S	
Germany	0.38	0.38	0.39	0.38	0.39	0.38	**0.37**	**0.36**	0.38
France	0.48	0.46	0.47	0.47	0.45	0.44	**0.40**	**0.39**	0.43
Canada	0.40	0.39	0.40	0.40	**0.38**	**0.38**	0.42	0.41	0.39
Japan	**0.37**	**0.55**	0.59	**0.58**	**0.58**	**0.58**	0.64	0.62	0.59
USA	0.48	0.48	0.49	0.47	0.47	0.47	0.51	0.49	0.46
UK	**0.40**	**0.40**	**0.38**	**0.38**	**0.40**	**0.40**	**0.40**	**0.39**	0.41
Italy	0.50	**0.49**	**0.48**	**0.47**	0.51	0.50	0.58	0.53	0.50

Note: 'N', 'S' are non-selected and selected. In the selected estimation are the significant parameters of the 12 lags non-selected estimation included. 'Const' is the Constant which is used as a benchmark for evaluating the forecasting abilities of the exogenous variables (bold figures = smaller than the constant).

Table 8: Out-of-sample results 1990 - LPS - 3 lags

	All variables		PROD		CPI		M1		M2	
	N	S	N	S	N	S	N	S	N	S
Germany	1.39	1.09	0.58	0.57	0.66	0.72	**0.54**	**0.55**	1.22	1.28
France	12.15	1.82	0.70	0.70	0.66	0.65	0.67	0.63	0.80	0.66
Canada	0.79	0.67	0.59	0.58	0.61	0.61	0.59	0.58	0.59	0.58
Japan	1.72	1.17	0.85	0.85	0.88	0.85	0.84	0.82	0.85	0.83
USA	1.14	0.91	0.70	0.69	0.67	0.67	0.69	0.71	0.67	**0.65**
UK	6.65	1.88	0.60	0.60	0.62	0.61	0.60	0.60	0.82	0.85
Italy	4.99	7.17	0.69	0.69	**0.68**	**0.68**	0.70	0.71	0.74	0.71

	IB		GB		SPREAD		UNEMP		Const
	N	S	N	S	N	S	N	S	
Germany	0.57	**0.56**	0.58	0.57	0.58	0.57	0.58	**0.56**	0.57
France	0.69	0.67	0.67	0.67	0.66	0.65	0.64	**0.61**	0.63
Canada	0.60	0.58	0.60	0.59	**0.57**	**0.57**	0.61	0.60	0.58
Japan	**0.75**	**0.75**	0.79	**0.78**	0.79	**0.78**	0.88	0.84	0.79
USA	0.67	0.67	0.68	0.66	0.66	0.66	0.70	0.69	0.66
UK	**0.58**	**0.58**	**0.56**	**0.56**	**0.59**	**0.59**	**0.59**	**0.58**	0.60
Italy	0.69	**0.68**	**0.68**	**0.66**	0.71	0.69	0.80	0.75	0.69

Note: 'N', 'S' are non-selected and selected. In the selected estimation are the significant parameters of the 12 lags non-selected estimation included. 'Const' is the Constant which is used as a benchmark for evaluating the forecasting abilities of the exogenous variables (bold figures = smaller than the constant).

Changing the perspective and proceeding with an assessment of the different indicator variables, it can be clearly stated that the variables with the highest prediction power are among the interest rate measures. UNEMP, however, shows in most cases a significantly high prediction power as well.

Table 9: Diebold Mariano Test of the out-of-sample results 1990 - 3 lags (p-values)

	All variables		PROD		CPI		M1		M2	
	N	S	N	S	N	S	N	S	N	S
Germany	1.00	0.91	0.99	0.48	0.99	1.00	0.97	0.98	1.00	1.00
France	1.00	1.00	1.00	1.00	0.96	0.86	1.00	**0.02**	1.00	0.97
Canada	1.00	0.98	1.00	0.30	1.00	1.00	0.78	0.46	0.62	0.44
Japan	1.00	0.99	1.00	1.00	1.00	1.00	0.98	0.83	0.99	0.93
USA	1.00	0.99	1.00	1.00	0.90	0.96	0.87	0.93	0.97	0.37
UK	1.00	0.05	0.86	0.75	0.93	0.86	0.86	0.75	1.00	1.00
Italy	1.00	0.99	0.60	0.60	0.13	0.08	0.77	0.97	1.00	0.99

	IB		GB		SPREAD		UNEMP	
	N	S	N	S	N	S	N	S
Germany	0.65	0.64	0.69	0.59	0.84	0.59	0.38	0.19
France	0.95	0.90	0.88	0.86	0.95	0.74	**0.02**	**0.01**
Canada	0.67	0.40	0.86	0.81	0.15	0.11	1.00	0.99
Japan	**0.04**	**0.02**	0.39	0.27	0.25	0.18	0.97	0.89
USA	0.99	0.96	0.97	0.79	0.89	0.86	1.00	1.00
UK	0.12	0.11	**0.00**	**0.00**	0.21	0.17	0.25	0.07
Italy	0.38	**0.03**	0.15	**0.01**	0.82	0.46	1.00	0.95

Note: 'N', 'S' are non-selected and selected. H0: constant = forecasting model; H1: forecasting model is better than the constant; The null hypothesis is rejected by a p-value smaller than five percent.

At least for half of the countries, most QPS and LPS values in these categories, are smaller than those of the constant. The other indicator variables exhibit fairly less informational content and therefore have in most cases a higher QPS and LPS than the corresponding constant. The only notable exception is Italy, where according to all measures CPI indicates good prediciton qualities.

Turning here as well to the results of the Diebold-Mariano test, it is striking that the number of forecasting models that are significantly better than the constant is somewhat lower than indicated by the other measures (see table 9). Promising forecasts however can be made for France with the UNEMP, for Japan with IB, and for the UK with GB. However, the high number of positive QPS and LPS results for the UK seems to be largely insignificant. Furthermore, in the case of Italy, the selection procedure in both interest rate cases delivers a good forecasting model.

3.2.3 Robustness Check

In addition to the estimations undertaken in the last two subsections, we also check for the sensitivity of our results to a later starting point in the out-of-sample analysis as well as for variations in the identification procedure for bull and bear markets.

We first set the starting point for the out-of-sample analysis to January 1995 - see table 11-13 in the appendix - and later on to January 2000 (table 14-16). The findings organized by countries can be summerized as follows: Based on the already high prediction power of the 1990-sample, the QPS and LPS values for the indicator variables of Japan and Italy are often significantly higher than the corresponding values of the constant. On the other side of the spectrum, the US are the country whose stock market downturns are least predictable. In basically all cases, the US indicator variables were not able to beat the predictions made by the constant. A similar point can be stated for Canada, yet with a reservation on the LPS case for 1995, where a larger number of inidicator variables contains some prediction power. However, the difference to the LPS in the 1990-sample is very small for most variables of Canada. While the UK seems to have a high informational content in the 1990-sample, the number of cases, where the QPS and LPS of the indicator variables are smaller than the ones for the constant, decreases over time. Finally, in the 2000-sample for the UK only a few cases with good prediction power remained. Germany and France undergo development in the opposite direction. Both countries show a rather low prediction power in the 1990-sample with only two values of the QPS beating the constant. For 1995 in the case of Germany and 2000 in the one of France, both countries show clear signs of bear market predictability. Nevertheless, the high forecasting power for Germany is reduced somewhat in the 2000-sample.

When the presentation of the out-of-sample results is organzied by indicator variables, the above derived pattern, with interest rate variables and UNEMP being the best predictors for stock market downturns is largely conmfired in the 1995- and the 2000-sample. For the 1995-sample, GB seems to have a slightly exposed position as the very best predictor just like IB in the 2000-sample. In contrast, the other indicator variables barely beat the constant. Exemptions are again Italy, whose CPI value constantly indicates high predictive power and the two money supplies - predominantly M1 - for Germany and Japan which also exhibit some informational content.

The Diebold-Mariano results generally support the above mentioned findings. The highest number of models that perform significantly better than the constant is found for Germany, France, and Italy under both out-of-sample starting points. Countries such as the USA, UK and Canada, as well as to some extent Japan have much higher p-values than the 5-percent level, implying that the forecasting models cannot beat the model with only a constant. Taking on the indicator variable perspective, it is shown again that interest rate variables and UNEMP have the highest chances to beat the constant. However, when applying the Diebold-Mariano test to both out-of-sample starting points, it turns out that SPREAD and UNEMP seem to be the only variables that beat the constant in larger number of countries. These findings differ to some extent from the results of previous measures such as QPS and LPS, which assigned a high prediction power to all kinds of interest rate variables.

To check for robustness against the criteria of the bear market identification process, we change the minimum phase length of a bull or bear market from four to eight months. As expected, this step reduces the number of bear markets somewhat and increases the period between two turning points.

The in-sample results for the one lag length under the increased minimum phase length are broadly similar to the former results.[10] For the pseudo R^2, QPS and LPS most values are on largely similar levels and in some cases did not change at all. Interestingly, as QPS and LPS values suggest, a considerable number of values even seem to be consistently smaller by a certain margin. Hence, this shows that an increase in the minimum phase length improves the forecasting performance of the indicator variables instead of decreasing it. Therefore, keeping the minimum phase length at the originally selected four month is a rather cautious approach.

The out-of-sample results confirm these findings as well, when the longer minimum phase length requirement is applied to the out-of-sample specifications. The QPS values are again largely in line with the original results and whenever a difference ocurrs, the prediction power has increased, rather then diminished. For the LPS, the number of specifications with high informational content is lower and a few earlier values have changed. However, since countries whose stock market downturns were forecastable best and worst in the 1990-sample have remained the same. In place of the indicator variables, interest rates and UNEMP still exhibit the highest prediction power. The four months phase length seems thoroughly feasible.

3.3 Conclusion

The goal of this paper was to evaluate a set of macroeconomic indicator variables with respect to its prediction power on stock market downturns in G7 countries. The applied methodical approach consisted of two steps. At first, a modified version of the Bry-Boschan business cycle dating algorithm was applied to identify bull and bear markets from the data by creating dummy variable series. In a second step, under an in-sample and out-of-sample framework, probit estimations were carried out by regressing the newly identified dummy variable series on different specifications of indicator variables. To assess the hereby generated estimation results, typical forecasting evaluation measures were employed. Especially three findings have emerged from this analysis:

Firstly, in the in-sample analysis, the all-variable specifications, i.e. those including all indicator variables of the set, show the highest prediction power across all countries. In the out-of-sample analysis, this is not longer the case, since predictions made by single variables and its lagged values yield significantly better forecasts.

Secondly, when seen from the countries' perspective, stock market downturns can be predicted best for Japan, Italy, and to a lower extent for France and Germany. The countries with the least

[10] Tables are not reported but available on request.

predictable downturns were the US, Canada and the UK in more recent times. As an explanation for these findings could serve the fact that Anglo-American countries - who are all in the group with the less predictive marekts - have adopted maket based systems to conduct financial intermediation. As a result, the stock markets of these countries are large and deep. Indeed, when comparing the market capitalization of listed companies in percent of GDP for 2006, the UK (160 percent), the USA (148), and Canada (134) show significantly higher ratios than Japan and France (each 108) as well as Germany (57) and Italy (55).[11] Hereby, the market size could make the stock markets more informationally efficient - e.g. by creating a larger base of informed investors, stronger disclosure standards, or broader media coverage - and hence reduce the possibility to obtain good forecasts. In contrast, the banking system orientated European countries, such as Germany, Italy, France, and Japan might have relatively less developed stock markets, a lower degree of market efficiency and hence, better predictable stock prices.

Thirdly, macroeconomic indicator variables with the best prediction power in in- and out-of-sample are especially the interest rates and their linear combination - the term spread - and the unemployment rate. Prices follow thereafter. Finally, the least prediction power is found in the two money stock series. A possible reason for this observation could be that monetary authorities usually act not till the stock market downturn becomes severe and hence, the money stock variables only contain little informational content. However, the companies might act somewhat earlier and could sporadically adjust production and lay off workers before the stock market downturn sets in. In accordance with the literature, the interest rate based variables exhibit the highest prediction power of all variables in our analysis as well. A possible explanation for this result could be given by the fact that interest rates contain a risk premium, which doublessly increases, when a crisis approaches.

Referring to the purposes that this study should fulfill, besides its academic contribution, one can exploit its value for the two other target groups: For policy makers, the recommendation might be that an analysis of interest rates is a good point to start with, when the risk of a stock market downturn is assessed. For investors, the answer is somewhat more complex. According to our findings investors are advised to carefully examine interest rates. Furthermore, investments in the Japanese and Italian markets seem to yield the most non-random returns. However it should be noted that when investors decide to reallocate their investments, a certain ammount of transaction costs occurs. To examine the practical value of the here derived results, assumptions about these transaction costs and the investment's opportunity costs have to be made in the first place. Moreover, a personal threshold depending on each investor's preferences that indicates when the investor should exit the

[11] World Development Indicators

market has to be determined as well.

Possible lines of future research may also include: A comparison of different identification procedures for the bear market series - such as in Chen [2009], who identified the turning points via a Markov-switching model. Furthermore, the lag structure of the best specifications could be extracted and subjected to further tests. Finally, it would be interesting to see how other indicator variables, such as business or consumer sentiment indices, developments in commodity prices, and typical financial risk measures score in this assessment.

References

A. Ang and G. Bekaert. International asset allocation with regime shifts. *Review of Financial Studies*, 15:1137–1187, 2002.

R. A. Ashley, C. W. J. Granger, and R. Schmalensee. Advertising and aggregate consumption: An analysis of causality. *Econometrica*, 48:1149–1167, 1980.

R. Balvers, T. Cosimano, and T. McDonald. Predicting stock returns in an efficient market. *Journal of Finance*, 45:1109–1118, 1990.

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

J. G. Biscarri and F. Perez de Gracia. Bulls and bears: Lessons from some european countries. Working paper, Royal Economic Society, 2002.

J.H. Boyd, R. Jaganathan, and J. Hu. The stock market's reaction to unemployment rate: Why bad news is usually good for stock. (8092), 2001.

G.W. Brier. Verification of forecasts expressed in terms of probabilities. *Monthly Weather Review*, 75:491–504, 1950.

G. Bry and C. Boschan. *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*. UMI, 1971.

J. Y. Campbell. Stock returns and the term structure. Working Paper 1626, NBER, 1985.

J. Y. Campbell and M. Yogo. Efficient tests of stock return predictability. *Journal of Financial Economics*, 81:27–60, 2006.

N.-F. Chen, R. Roll, and S. A. Ross. Economic forces and the stock market. *The Journal of Business*, 59(1):383–403, 1986.

S.-S. Chen. Predicting the bear stock market: Macroeconomic variables as leading indicators. *Journal of Banking and Finance*, 33:211–223, 2009.

J. Cunado, L. A. Gil-Alana, and F. Perez de Gracia. Stock market volatility in us bull and bear markets. *Journal of Money, Investment and Banking*, 1:24–32, 2008.

F. X. Diebold and R. S. Mariano. Comparing predictive accuracy. Technical Working Paper 169, NBER, 1994.

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

E. F. Fama. Stock returns, real activity, inflation and money. *Journal of Financial Economics*, 71: 545–565, 1981.

L. Gonzalez, J. Powell, J. Shi, and A. Wilson. Stock market cycles, financial liberalization and volatiliy. *International Review of Economics and Finance*, 14:469–486, 2005.

D. Harding and A. R. Pagan. Dissecting the cycle: a methodological investigation. *Journal of Monetary Economics*, 49:365–381, 2002.

G. Kaul. Stock returns and inflation: The role of the monetary sector. *Journal of Financial Economics*, 18:253–276, 1987.

M. W. Keran. Expectations, money and the stock market. *Federal Reserve Bank of St.Louis Review*, 53(1):16–31, 1971.

J. Lewellen. Predicting returns with financial ratios. *Journal of Financial Economics*, 74(2):209–235, 2004.

J. M. Maheu and T. H. McCurdy. Identifying bull and bear markets in stock returns. *Journal of Business and Economic Statistics*, 18:100–112, 2000.

R. A. Meese and K. Rogoff. Empirical exchange rate models of the seventies. do they fit out of sample? *Journal of International Economics*, 14:3–24, 1983.

A. R. Pagan and K. A. Sossounov. A simple framework for analysing bull and bear markets. *Journal of Applied Econometrics*, 18(1):23–46, 2003.

D. K. Pearce and V. V. Roley. The reaction of stock prices to unanticipated changes in money. *Journal of Finance*, 38:1323–1333, 1983.

D. E. Rapach, M. E. Wohar, and J. Rangvid. Macro variables and international stock return predictability. *International Journal of Forecasting*, 21(1):137–166, 2005.

R. Rhea. *The Dow theory: an explanation of its development and an attempt to definite its usefulness as an aid in speculation*. Barron, 1932.

R. J. Samuelson. Lessons from the 1987 crash. *Newsweek*, October 2007.

J. H. Stock and M. W. Watson. *Introduction to Econometrics*. Pearson/Addison-Wesley, 2008.

W. Thorbecke. On stock market returns and monetary policy. *Journal of Finance*, 52:635–654, 1997.

Appendices

Table 10: Mean and standard deviation of the log-MSCI series

	Mean	Standard Deviation
Germany	5.75	0.99
France	6.17	1.1
Canada	5.91	0.86
Japan	6.23	0.77
US	5.4	0.84
UK	5.88	1.1
Italy	5.71	1.2

Table 11: Out-of-sample results 1995 - QPS - 3 lags

	All variables		PROD		CPI		M1		M2	
	N	S	N	S	N	S	N	S	N	S
Germany	0.52	**0.34**	0.40	0.39	**0.38**	0.41	0.46	0.46	0.49	0.46
France	0.78	0.59	0.49	0.49	0.46	0.46	0.48	0.46	0.50	**0.45**
Canada	0.37	0.32	0.34	0.33	0.34	0.34	0.34	0.33	0.33	0.33
Japan	0.63	0.60	0.58	0.59	0.59	0.58	**0.55**	**0.53**	0.57	**0.55**
USA	0.55	0.50	0.53	0.51	0.48	0.47	0.48	0.48	0.48	0.46
UK	0.53	**0.40**	0.42	0.42	0.43	0.42	0.42	0.42	0.47	0.47
Italy	0.73	0.57	0.50	0.50	**0.48**	**0.48**	0.50	0.52	0.54	0.50

	IB		GB		SPREAD		UNEMP		Const
	N	S	N	S	N	S	N	S	
Germany	**0.34**	**0.34**	**0.38**	**0.38**	**0.34**	**0.34**	**0.36**	**0.36**	0.39
France	0.51	0.50	0.50	0.49	0.46	**0.45**	**0.39**	**0.39**	0.46
Canada	0.34	0.33	0.33	0.33	**0.32**	**0.32**	0.35	0.34	0.33
Japan	**0.51**	**0.51**	**0.53**	**0.53**	0.56	0.56	0.61	0.59	0.56
USA	0.49	0.49	0.50	0.49	0.47	0.47	0.54	0.52	0.46
UK	0.43	0.43	**0.41**	**0.41**	0.43	0.43	**0.40**	**0.39**	0.42
Italy	0.50	0.50	**0.48**	**0.47**	**0.47**	**0.47**	0.53	**0.49**	0.50

Note: 'N', 'S' are non-selected and selected. In the selected estimation are the significant parameters of the 12 lags non-selected estimation included. 'Const' is the Constant which is used as a benchmark for evaluating the forecasting abilities of the exogenous variables (bold figures = smaller than the constant).

Table 12: Out-of-sample results 1995 - LPS - 3 lags

	All variables		PROD		CPI		M1		M2	
	N	S	N	S	N	S	N	S	N	S
Germany	0.97	**0.52**	0.59	0.58	0.63	0.69	**0.54**	**0.55**	1.32	1.41
France	5.93	1.62	0.72	0.71	0.67	0.67	0.69	0.65	0.74	**0.64**
Canada	0.57	0.50	0.53	0.52	0.53	0.53	0.52	0.52	**0.51**	0.52
Japan	0.91	0.84	0.78	0.79	0.80	0.78	**0.74**	**0.72**	0.77	0.75
USA	0.94	0.82	0.73	0.71	0.67	0.67	0.68	0.70	0.68	0.65
UK	2.31	2.01	0.62	0.61	0.63	0.61	0.62	0.61	0.69	0.68
Italy	1.28	0.84	**0.69**	**0.69**	**0.67**	**0.68**	**0.69**	0.71	0.73	0.70

	IB		GB		SPREAD		UNEMP		Const
	N	S	N	S	N	S	N	S	
Germany	**0.52**	**0.52**	**0.57**	**0.56**	**0.52**	**0.52**	**0.55**	**0.54**	0.58
France	0.70	0.69	0.70	0.69	0.66	0.65	**0.58**	**0.56**	0.65
Canada	0.52	**0.51**	**0.51**	**0.51**	**0.50**	**0.50**	0.54	0.53	0.52
Japan	**0.71**	**0.70**	**0.73**	**0.73**	0.76	0.75	0.84	0.81	0.75
USA	0.69	0.69	0.70	0.68	0.66	0.66	0.73	0.72	0.65
UK	0.62	0.62	**0.60**	**0.60**	0.63	0.62	**0.59**	**0.57**	0.61
Italy	**0.69**	**0.69**	**0.67**	**0.66**	**0.67**	**0.66**	0.73	**0.68**	0.70

Note: 'N', 'S' are non-selected and selected. In the selected estimation are the significant parameters of the 12 lags non-selected estimation included. 'Const' is the Constant which is used as a benchmark for evaluating the forecasting abilities of the exogenous variables (bold figures = smaller than the constant).

Table 13: Diebold Mariano Test of the out-of-sample results 1995 - 3 lags (p-values)

	All variables		PROD		CPI		M1		M2	
	N	S	N	S	N	S	N	S	N	S
Germany	1.00	**0.02**	0.97	0.48	0.41	0.88	0.98	0.99	1.00	0.99
France	1.00	0.97	0.95	0.95	0.70	0.49	1.00	**0.02**	1.00	**0.02**
Canada	0.96	0.11	1.00	0.66	0.87	0.87	0.62	0.38	0.25	0.32
Japan	0.94	0.83	0.89	1.00	1.00	1.00	0.26	**0.01**	0.90	0.34
USA	0.99	0.89	1.00	1.00	0.92	0.95	0.73	0.77	0.98	0.42
UK	1.00	0.30	0.81	0.74	0.84	0.61	0.81	0.74	1.00	1.00
Italy	1.00	0.91	**0.01**	0.16	**0.03**	**0.03**	0.37	0.95	0.96	0.50

	IB		GB		SPREAD		UNEMP	
	N	S	N	S	N	S	N	S
Germany	**0.00**	**0.00**	0.33	0.29	**0.00**	**0.00**	0.14	0.09
France	0.94	0.90	0.90	0.86	0.52	0.37	**0.00**	**0.00**
Canada	0.60	0.23	0.35	0.29	0.07	0.05	0.99	0.98
Japan	0.07	**0.03**	0.20	0.21	0.50	0.46	0.95	0.87
USA	1.00	1.00	1.00	0.99	0.88	0.91	1.00	1.00
UK	0.86	0.79	0.25	0.16	0.95	0.83	0.06	**0.00**
Italy	0.08	0.07	0.06	**0.01**	**0.01**	**0.00**	0.96	0.09

Note: 'N', 'S' are non-selected and selected. H0: constant = forecasting model; H1: forecasting model is better than the constant; The null hypothesis is rejected by a p-value smaller than five percent.

Table 14: Out-of-sample results 2000 - QPS - 3 lags

	All variables		PROD		CPI		M1		M2	
	N	S	N	S	N	S	N	S	N	S
Germany	0.62	0.52	0.50	0.49	0.57	0.61	0.50	0.51	0.49	0.49
France	0.92	0.62	0.64	0.64	0.63	0.62	0.66	0.62	0.68	0.62
Canada	0.42	0.40	0.40	0.40	0.43	0.42	0.41	0.40	0.41	0.41
Japan	0.61	0.59	0.56	0.57	0.55	0.54	**0.52**	0.51	0.54	**0.52**
USA	0.73	0.66	0.59	0.57	0.59	0.58	0.68	0.68	0.57	**0.54**
UK	0.74	0.60	0.58	0.56	0.61	0.60	0.58	0.56	0.63	0.62
Italy	0.80	0.65	0.50	0.50	0.50	0.50	0.53	0.53	0.53	0.50

	IB		GB		SPREAD		UNEMP		Const
	N	S	N	S	N	S	N	S	
Germany	0.49	0.49	0.54	0.53	**0.45**	**0.45**	0.57	0.56	0.49
France	**0.58**	**0.58**	**0.57**	**0.58**	0.62	0.62	**0.53**	**0.52**	0.62
Canada	0.43	0.42	0.41	0.41	0.40	0.40	0.42	0.40	0.40
Japan	**0.52**	**0.52**	**0.52**	**0.52**	**0.51**	**0.51**	0.57	0.54	0.54
USA	0.60	0.59	0.60	0.58	0.56	0.57	0.58	0.57	0.55
UK	0.56	**0.55**	0.57	0.56	0.59	0.58	**0.54**	**0.51**	0.56
Italy	**0.49**	0.50	0.54	0.52	**0.45**	**0.45**	0.54	**0.49**	0.50

Note: 'N', 'S' are non-selected and selected. In the selected estimation are the significant parameters of the 12 lags non-selected estimation included. 'Const' is the Constant which is used as a benchmark for evaluating the forecasting abilities of the exogenous variables (bold figures = smaller than the constant).

Table 15: Out-of-sample results 2000 - LPS - 3 lags

	All variables		PROD		CPI		M1		M2	
	N	S	N	S	N	S	N	S	N	S
Germany	0.94	0.77	0.70	0.69	0.88	0.98	**0.68**	0.69	0.69	0.69
France	6.22	1.31	0.93	0.93	0.86	0.86	0.91	0.84	0.96	0.84
Canada	0.62	0.59	0.59	0.59	0.62	0.61	0.60	0.59	0.61	0.60
Japan	0.89	0.82	0.77	0.77	0.74	0.74	**0.71**	**0.70**	0.74	**0.71**
USA	1.25	1.10	0.79	0.77	0.79	0.78	0.94	0.96	0.77	**0.73**
UK	1.52	3.09	0.80	0.77	0.84	0.82	0.80	0.77	0.87	0.86
Italy	1.20	0.91	0.69	0.69	0.69	0.69	0.73	0.73	0.72	0.69

	IB		GB		SPREAD		UNEMP		Const
	N	S	N	S	N	S	N	S	
Germany	0.69	0.69	0.76	0.75	**0.64**	**0.64**	0.82	0.81	0.69
France	**0.78**	**0.77**	**0.77**	0.78	0.85	0.85	**0.73**	**0.72**	0.84
Canada	0.64	0.62	0.61	0.61	0.59	0.59	0.61	0.59	0.59
Japan	**0.72**	**0.71**	**0.71**	**0.71**	**0.70**	**0.71**	0.79	0.76	0.73
USA	0.81	0.79	0.80	0.77	0.76	0.77	0.78	0.77	0.75
UK	**0.76**	**0.75**	0.78	**0.76**	0.81	0.80	**0.73**	**0.71**	0.77
Italy	**0.68**	0.69	0.73	0.71	**0.64**	**0.65**	0.73	0.69	0.69

Note: 'N', 'S' are non-selected and selected. In the selected estimation are the significant parameters of the 12 lags non-selected estimation included. 'Const' is the Constant which is used as a benchmark for evaluating the forecasting abilities of the exogenous variables (bold figures = smaller than the constant).

Table 16: Diebold Mariano Test of the out-of-sample results 2000 - 3 lags (p-values)

	All variables		PROD		CPI		M1		M2	
	N	S	N	S	N	S	N	S	N	S
Germany	1.00	0.85	0.75	**0.02**	1.00	1.00	0.64	0.84	0.63	**0.02**
France	1.00	0.50	0.89	0.89	0.68	0.50	1.00	**0.02**	1.00	**0.02**
Canada	0.83	0.55	0.87	0.36	1.00	0.99	0.74	0.45	0.95	0.82
Japan	0.89	0.83	0.92	1.00	0.96	0.97	0.15	**0.04**	0.68	0.16
USA	1.00	0.99	1.00	0.96	0.99	1.00	1.00	1.00	0.89	0.12
UK	1.00	0.87	1.00	0.25	1.00	1.00	1.00	0.25	1.00	1.00
Italy	1.00	1.00	0.50	0.77	0.48	0.47	0.96	0.99	0.97	0.77

	IB		GB		SPREAD		UNEMP	
	N	S	N	S	N	S	N	S
Germany	0.43	0.44	0.98	0.98	**0.00**	**0.00**	0.99	0.99
France	0.23	0.20	0.18	0.23	0.54	0.48	**0.00**	**0.00**
Canada	0.97	0.88	0.83	0.81	0.41	0.52	0.92	0.67
Japan	0.36	0.33	0.35	0.31	0.05	0.11	0.76	0.58
USA	1.00	1.00	1.00	0.96	0.91	0.99	1.00	0.95
UK	0.46	0.21	0.63	0.37	1.00	0.98	**0.03**	**0.00**
Italy	0.18	0.63	0.99	0.94	**0.00**	**0.00**	1.00	0.39

Note: 'N', 'S' are non-selected and selected. H0: constant = forecasting model; H1: forecasting model is better than the constant; The null hypothesis is rejected by a p-value smaller than five percent.