Sondre Andersen

Forecasting Economic Downturns in The Scandinavian Countries using The Yield Curve

Exploring statistical relationships and out-of-sample performance using traditional binary response models, and support vector machine models from the field of machine learning.

Master's Thesis Spring 2020
Oslo Business School
Oslo Metropolitan University
Master's Program in Business Administration

Preface

This thesis is the final part of my Master's degree in Business and Administration with a specialization within Economics and Finance at OsloMet. The work with this thesis has given valuable insights into the field of economics and research in general. To be able to do the analysis that I have done in this thesis quite a bit of programming had to be learned. I use STATA, Python and R extensively in the analysis section. It has been very interesting to learn these programming tools/languages and also to see how useful they are in practice and how much they can speed up work that would otherwise take a long time to do in Excel. It has also been interesting to learn more about econometrics and machine learning and how new methods are being applied to solve economic problems.

I would also like to take this opportunity to thank my supervisor Fenella Carpena for good advice on this thesis and for being an inspirational lecturer in the Econometrics course. I would also thank my fellow students for interesting and funny conversations while working on this thesis. Also, a huge thanks to all my good friends and family for support.

Abstract

Economic downturns (in this thesis defined as recessions and negative output gaps) are costly and it is in the interest of both the Government and private agents to precisely predict them, either for the government to take actions to avoid them, or for private agents to prepare/adjust. Many advanced econometric models are developed for this exact purpose and experts regularly express their opinion about the future of the economy in the financial press. Yet, U.S. data have shown that a very simple probit model taking variables from the yield curve as explanatory variables has successfully predicted recessions in the past. The aim of this thesis is twofold. First, I test whether the established relationship between economic downturns and yield curve variables in the U.S. also hold in the Scandinavian countries. This is done by estimating a static probit model with the yield spread as the independent variable. Second, I test whether estimating a model from the field for machine learning, called Support Vector Machine (SVM), can improve on the forecasts made by the traditional probit models. To be able to compare forecasts I find a probability threshold, W, that produces binary forecasts from the probabilistic forecasts made by the probit models and calculate several performance measures based on pseudo out-of-sample forecasts. The SVM model directly output a binary forecast so the same performance measures can be calculated directly.

First, this thesis find that the coefficients for lags of the yield spread (long rate minus short rate) is significant a 5% percent level for all forecasting horizons tested (from one to eighteen months) when recession in Sweden is the dependent variable. For Denmark only shorter time horizons give significant results, while for Norway very few forecasting horizons prove significant. Models estimated with negative output gap, defined using the Hodrick-Prescott-filter, as the dependent variable yields statistically significant coefficients of the spread at almost all lags for Norway and almost non for Sweden and Denmark. I find no evidence that including more than one lag of the spread is useful.

Second, the pseudo out-of-sample tests show that economic downturns are better predicted by the yield curve than by lagged returns of a national stock index. This result holds independent of forecasting horizon, lag length and country. I also find that SVM models that take 10 year bond rates and 3 month T-bill rates as input variables in almost all cases outperform the binary forecasts from the probit model. With respect to which country can benefit the most from the models estimated here, the pattern from the statistical tests are repeated, meaning the out-of-sample results are best for Sweden and Denmark when forecasting recessions, and best for Norway when forecasting negative output gaps.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In late 2019 Senior Contributor to major news site Forbes, Chuck Jones, warned that a recession could hit the U.S. economy in 2020 (Jones 2019). His argument was that that in late May of 2019 the yield on the 10-year government bond dropped below the yield on the 3-month Treasury bill. Such an event, where yields on bonds with short maturities are higher than the yield on bonds with longer time to maturity, is referred to as an *inverted yield curve*. A chart displaying the market yield rates on government bonds on the y-axis and time to maturity on the x-axis would show a downward sloping curve in the scenario described above. Jones, in his article, went on to explain that the yield curve had been inverted every day from May 23rd until October 10th, except for one single day in July. To support his statement Jones take to the history books, in particular the most popular historical record among economists, namely economic time series data. He shows that an inversion of the yield curve has occurred between 8 and 13 months before all three recessions in the U.S. since 1990. So, if history is to repeat itself we should expect a recession somewhere between January and November 2020, he writes. Jones, was by no means the first to discuss the *yield curve inversion* – throughout late summer and fall of 2019 this event was discussed quite frequently in the financial press (for example by The Economist (2019)).

A major part of the discussion on the yield curve and its relation to economic downturns in popular media, financial press and the academic literature alike are focused on the USA. This might leave non-U.S. citizens wondering whether the results also hold for their own county. This thesis attempts to answer this question for the people of Scandinavia.

But first, let us take a step back and ask: why would it be useful to be able to forecast a recession or negative trends in the economy in general? Who could benefit from such knowledge?

Recessions are very costly to an economy and the people who are a part of it, on many levels. Dao and Loungani (2010) reports that by late 2010 global unemployment had increased by 30 million since 2007, that is, before the financial crisis that started in the American subprime mortgage market. The consequences for the people that loose their job during a recession have, historically, proved to be long lasting according to the same report by researchers at IMF. The unemployed could expect to have losses in earnings lasting 15-20 years after the recession, and even have their life expectancy reduced by 1 to 1.5 years. And the consequences does not stop there, even the children of these individuals are expected to have lower academic achievement and earnings.

The cost to individuals close to their retirement age can also be particularly high. During the 2008 financial crisis many seniors took such a hit (Mercado 2018). Younger peoples' savings are usually also negatively affected during a recession, at least for those heavily invested in stock markets. However, recessions and the following (or sometimes preceding) stock market crashes are particularly bad for seniors that cannot afford to wait to withdraw their money from the markets until after the bull markets that often follow recessions.

To the other question, who could benefit from knowledge about a coming recession? The individuals close to retirement discussed above could benefit a great deal by knowing a recession was on the horizon, and thus shift their savings from stocks to less volatile and risky assets such as bonds. Households could start to save more to smooth their income over time, to avoid the recession hitting as hard as it otherwise could. A group that could potentially have an even larger benefit of knowing a recession or economic downturn is on the horizon are central banks and the government. As the main mandate of central banks is to stabilize the economy whether through inflation targets or some other measures, having a good recession/negative output gap indicator is invaluable to such an institution. Investors, from small individual investors to managers of pension funds also have the obvious benefit of being able to look into the future of the economy.

The discussion in the previous paragraph have probably left some readers thinking of a possible problem here. Since central banks in most countries can heavily influence the domestic short term rate and thus also the slope of the yield curve, will the information provided by the yield curve simply reflect the consequences of the central banks' actions on the economy? This and other issues are discussed in chapter 7.

As stated earlier, most of the research on whether the information provided by the yield curve can predict future economic activity has focused on data from the USA. The goal of this thesis is to explore whether the yield curve can give useful information also for the Scandinavian countries. Most authors have applied more or less complicated statistical models to explore these relationships. One of the most popular estimation algorithms has been the probit model. This model gives a probability of a binary variable being equal to one given some value(s) of some independent variable(s), in this case lagged values of the difference between the yield on long and short term government bonds, the *yield spread*, is used. In resent years however, the field of machine learning (ML) has taken great leaps, primarily due to increases in affordable and great computational power. The methods have been applied increasingly in the field of economics, however, it is important to know their appropriate use cases. As Professor of economics Sendhill Mullainathan and PhD candidate in economics Jann Spiess put it; "Machine learning (or rather "supervised" machine learning [...]) revolves around the problem of *prediction*: produce predictions of y from x" (Mullainathan and Spiess 2017). However, the problem of forecasting economic activity *is* a problem of prediction, so analysis of how well such models work for this particular purpose is, in my opinion, a valuable contribution to the literature.

The research question this thesis attempts to answer is as follows:

> *Can lagged values of variables extracted from the yield curve help in forecasting economic downturns (in this thesis defined as recessions and negative output gaps) in the Scandinavian countries? And can techniques from the field of machine learning improve forecasts over traditional methods?*

So, in short, the main contributions this thesis makes to the literature is to test the forecasting performance of the yield curve in the Scandinavian countries, and to test whether methods from the field of Machine Learning can improve on the forecasts made by traditional methods.

Chapter 2

Literature Review

There is a sizable and still growing literature on the subject of the yield curve's (also sometimes referred to as the *term structure*) ability to forecast economic variables. The literature stretches all the way back to when Harvey (1986) found evidence that the "yield spread had some ability to predict future consumption growth both within-sample and out-of-sample." Harvey used two datasets in his thesis, the first was of quarterly frequency from 1953:Q2 to 1985:Q3 and the second was of annual frequency from 1872 to 1984. Using these datasets he estimated both linar and non-linear models. His results show that the yield spread performs better at predicting consumption growth than both lagged consumption and real stock returns.

Harvey (1988) found evidence that the real term structure of interest rates was a good predictor of future consumption growth. He also found that the information provided by the yield curve outperforms both lagged consumption growth and lagged stock returns when forecasting recessions, both in-sample and out-of-sample. Harvey also found that the yield curve model shows "promise" in out-of-sample forecasting when compared to more advanced econometric models. In the same decade, Harvey also found evidence to support the idea that the yield curve contains information that is useful in predicting future interest rates (Campbell and Shiller 1987).

It appears that one of the first to point out that there is a relation between the yield curve and real economic activity (usually measured by GDP) was Fama (1986).[1] However, he only suggests that it looks like there is a relation between the two, he does not provide statistical evidence.

Estrella and Hardouvelis (1991) test this idea of whether the yield curve can be used as a predictor of real economic activity, and they point out that there has, previously, been done little empirical research on the yield curve's forecasting performance with respect to changes in real economic activity. They also point out that since the real GDP series (they are looking at data from the USA) seems to follow a "near-random-walk"-behavior, finding that the yield curve can predict changes in this series would be "impressive". Estrella and Hardouvelis (1991) use quarterly data from the second quarter of 1995 until the final quarter of 1988. They define 10-year government bonds as the long rate(R^L) and 3-month T-bills as

[1]His paper was released in the same year as Harvey's PhD dissertation. Since Harvey looked at explaining consumption growth and since consumption growth is closely linked to GDP growth Fama was clearly not the only one to think of this relation. Harvey mentions Fama in his PhD as an important supervisor, so these ideas where almost certainly discussed at the time between the two.

the short rate(R^S). The slope of the yield curve is simply defined as the difference between the two rates ($SPREAD_t = R_t^L - R_t^S$). The hypothesis that a steeper (more positive) slope precedes faster growth in GDP, and vice versa, is supported in the data. However, in the regression both the constant term and the slope coefficient is positive which means, as they point out; "that a negative slope does not necessarily predict negative future real GNP growth." This means that the popular "inversion rule" (that a negative slope of the yield curve predicts a decrease in production) is not confirmed by the authors. Their regression results show that the relation between the slope of the yield curve and future real economic activity is highest when the forecasting horizon is 5 to 7 quarters ahead.

The same paper by Estrella and Hardouvelis (1991) also explores the possibility that the yield curve can predict a recession. This is done by using the probit method, which gives a probability of a recession in t quarters. By taking a different approach to exploiting the information in the yield curve, the authors have changed the problem from a continuous prediction problem to a binary classification problem. The definition they use for a recession is, according to the authors, the same as the one The National Bureau of Economic Research (NBER) uses, which is essentially defined as "two consecutive quarters of negative real GNP growth", but the final decision is subjective and made by a committee. They find that "the relation between the probability of a recession and the spread is statistically significant."

Estrella and Mishkin (1996) compares the predictive power of the yield curve to other financial and macroeconomic variables that are commonly used to predict recessions. They use a probit model to estimate the probability of a recession t quarters ahead using the difference between the ten-year Treasury note and the three-month Treasury bill. The performance of this model is compared to the forecasting performance of The New York Stock Exchange (NYSE) stock price index, the Commerce Department's index of leading economic indicators, and The Stock-Watson index. The main findings in the paper are, first, that all the variables assessed have some forecasting performance when the forecasting horizon is one quarter ahead, but the Stock-Watson index makes the best forecasts. Second, when the forecasting horizon increases to two or more quarters the yield curve clearly outperforms the other variables, and the advantage to the yield curve only grows as the forecasting horizon is increased. Estrella and Mishkin (1996) concludes that their results "suggests that the yield curve spread can have a useful role in macroeconomic prediction, particularly with longer lead times." They also emphasize that the relatively strong predictive power over longer horizons is particularly useful to policymakers as their actions usually take quite a while before one can see the entire effect on the economy, which means that policymakers would have to look quite far into the future when making decisions.

When forecasting an event such as a recession autocorrelation can be a problem, since recession months always occur in a row, so the i.i.d. assumption of many standard regression models could be violated. Ratcliff (2013) reports different ways of dealing with this problem, and the two most common ways are 1) Using heteroscedastic and autocorrelation robust standard errors when estimating the probit model, and 2) Estimate a dynamic probit model. The first method is probably the most common. The second method is used by, among others, Dueker (1997) and is a probit model that includes a lagged value of the recession variable. He argues that the reason why he uses a binary recession variable and not a continuous GDP growth variable, is that with a continuous variable the model might show a good fit, without performing well at predicting economic downturns, as the model might only work well in "normal" times. He continues; "The recession dummy variable, in contrast, isolates the accuracy with which one can predict the date of the onset and the expected length of recessions." The results of Dueker's probit model confirms the finding from previous papers that the yield curve works well for predicting recessions, but also in a dynamic framework.

In a large portion of the literature on the yield curve and recessions the focus have been on finding a statistically significant relationship between the slope of the yield curve and a binary recession variable. Ratcliff (2013) builds on the findings in these early studies by assessing the performance of the probit models in recession forecasting beyond simply looking at statistically significant relationships. Ratcliff claims to introduce techniques commonly used in the areas of statistics and meteorology, but not so much in economics, to evaluate probability forecasts. First, he looks at the probit models performance in giving non-probabilistic forecast of recessions, and second, looks at how useful the probabilities themselves are. Ratcliff finds that the probit models work well in the first case, when a forecasting horizon of 12 months is used. The non-probabilistic binary forecast is obtained form the probabilities estimated by the probit model by finding an optimal probability threshold that separates yes and no predictions. A natural threshold is often 50%, however, since recessions is such a rear event this threshold rule may not work particularly well. Ratcliff instead uses statistical methods for finding an optimal threshold. Intuitively, these methods aims to balance the number of correctly predicted recessions with the number of false alarms. He finds that a for a static model (only using the 12th lag of the term spread as explanatory variable) a optimal theshold rule lies between 20% and 30%. For a dynamic model (a model that also includes the first lag of the recession variable) a threshold between 20% and 75% is optimal. It turns out that using these decision rules, the probit models are useful in forecasting recessions. However, when looking more closely at the probabilities themselves, asking a question like; "does a steeper negative slope increase the probability of a recession", the model does not perform particularly well. The author looks at the frequency of recessions compared to the probability of a recession estimated by the probit models. He finds that when the probability of a recession (in the static model) rises above 35% the frequency of recessions does not increase with higher estimated probabilities. Ratcliff reports that; "[...] the estimated probabilities from the static model do not match the historical conditional probability of recession [...]", which means that a higher estimated probability above some threshold, or, equivalently, a more negative slope of the yield curve, does not actually result in a higher risk of a recession.

Since the early and fundamental papers found a clear relation between the slope of the yield curve and both economic activity and recessions, researchers have attempted to apply other methods than the simple logit and probit models to improve performance. For exmaple, Ozturk and Pereira (2014) used panel data methodology to be able to also estimate models for countries that have short times series on long and short term interest rates. Their dataset contained data from thirty-two OECD countries with quarterly data from the first quarter of 1990 until the first quarter of 2011. The definition of a recession that is used is two consecutive quarters of negative GDP growth and they create two models, one in which the first quarter of this event is also defined as a recession and one in which it is not. They find that the yield curve is a useful predictor of recessions for OECD countries, even for countries with short time series. However, they also find that their results are to some extent sensitive to the definition of a recession.

Gogas et al. (2015) contributes to the literature in two relatively new ways. First, they examine models that contain more than just two interest rates from the yield curve. This is done in order to include the arc information of the yield curve. Second, in addition to utilizing the commonly used probit model, they also investigate whether the information contained in the yield curve can be better modeled to predict output gaps using a method from the field of machine learning called Support Vector Machines (SVM). In their study Gogas et al. uses data from the US from the third quarter of 1976 to the final quarter of 2011. The goal of their study is to forecast positive and negative output gaps in the US, that is, deviations from the long-run trend of US GDP. When looking at the standard explanatory variable, which is one long and one short term interest rate, the authors find that the SVM method works better at predicting output gaps than probit models with the same input variables. When using the SVM method (with only two interest rates) they find that the best out-of-sample accuracy (forecasting both positive and negative output gaps) was reached when they used 3-month interest rates as the short term rate, and 10-year interst rates as the long term rate. When including one more interest rate in the analysis, the authors find that the best model is a SVM model that takes 3-month, 2-year and 3-year interest rates as input. Their results show that this particular model outperforms all the models they tested with only two interest rates as inputs. This result suggest that additional useful information can be extracted from the yield curve when information on the yield curve's arc is included.

Chapter 3

Theory and Definitions

This chapter goes through theory on the slope of the yield curve and argues why it could be able to predict recessions and output gaps based on economic theory. The chapter then goes on to define two key events, namely *recessions* and *output gaps*.

3.1 The Yield Curve

The yield curve, also referred to as the *term structure of interest rates*, is typically represented as a graph. The graph shows the relationship between interest rates on similar bonds with different times to maturity. When referring to the yield curve in this thesis I always refer to interest rates on government bonds/bills. There are three main theories that describe the slope of the yield curve(Brealey et al. 2017, page 59):

- **Expectations Theory of the Term Structure:** According to this theory investors are risk neutral and only care about returns. This means that investors will only be willing to hold short term bonds with lower interest rates than longer term bonds if they expect the interest rates on short term bonds to increase over the next few years. This means that the overall return from either investing only in a long term bond, or rolling over an investment in a short term bond over the same period should yield the same return. This theory suggest that the yield curve is upward sloping if market participants expect short term rates to increase, and that it is downward sloping if market participants expect short term rates to decrease.

- **Expectations Theory with Risk:** The price of long term bonds will be more volatile than the price of short term bonds(Brealey et al. 2017, page 60). For investors that have an investment horizon that is shorter than the time to maturity of a long term bond this higher risk will make short term bonds more attractive, all else being equal. This means that if these investors are to be indifferent between long term and short term bonds, then there must be a *risk premium* on the long term bond to compensate for the additional risk. This risk element will contribute to a more upward sloping yield curve.

- **Inflation and the Term structure:** Uncertainty about future inflation is another source of risk associated with long term bonds. If investors have a long investment horizon but are afraid inflation will increase over the period making their real return zero, or even negative, they would require a risk premium to go for these long bonds instead of rolling over short term bonds that are likely to adapt to changes in inflation. The element of inflation risk, will contribute to the slope of the yield curve in the same direction as the real interest rate risk discussed in the previous paragraph.

3.1.1 Why Consider The Yield Curve as a Predictor of Economic Downturns?

Apart from the fact that numerous studies have found a statistical relationship between past values of the *term spread* and both economic activity in general and economic downturns in particular, there are some coherent and theoretical explanations for this relationship that will be presented in this section.

First, it is important to note that both the demand side and the supply side of an economy are affected by real interest rates. Low interest rates will make borrowing cheap and saving less attractive, so consumers will spend more on consumption goods and firms will invest more. The reverse is also true, high interest rates will make borrowing more expensive and saving more attractive, so consumption and investment will drop.

Ozturk and Pereira (2014) starts their article by emphasizing the interconnectivity in macroeconomic data. They state that short term interest rates, which also affect long term rates according to the expectations theory of the term structure, in large is dependent on macroeconomic variables. They then go on to mention the following three arguments for the relation between the yield curve and real economic activity:

1. The first argument is grounded in the *Expectations Theory of the Term Structure*, which is explained above. This theory effectively states that, in equilibrium, long term rates are the weighted average of expected future short term rates. So, if long rates decrease, that effectively means that the expectation among investors is that future short term rates will decrease as well. Since short term interest rates normally decrease in periods of recessions due to the actions of central banks to stimulate the economy, decreased expected future short term rates can reflect an expectation of a future recession, and this expectation could be reflected in lower long-term rates some time in advance.

2. The second argument is similar to the argument by Estrella and Mishkin (1996). The argument is based how the yield curve is affected by monetary policy. Estrella and Mishkin (1996) points out that the term spread is highly influenced by the prevailing monetary policy. It is particularly the short term interest rate that is affected by monetary policy. Central banks and governments tend to increase the short term interest rate to avoid excessive real growth which is commonly associated with increasing inflation rates. Such actions also lead to a flatter yield curve and normally reduces real economic growth in the future. This reduced growth can then increase the probability of a recession as sectors that are particularly sensitive to macroeconomic variables might reduce their spending (Ozturk and Pereira 2014).

3. The third and final argument is based on Harvey (1988) and how consumers maximize their utility from consumption over time, and how they prefer a stable consumption path.[1] The argument follows logically from the following maximization problem: if consumers expect a downturn in the economy, they start to save in advance in order to avoid having to reduce their consumption considerably in the future expected economic downturn. In order to save to be able to smooth consumption over time, consumers start to invest in long term bonds. Since the price of a zero coupon bond with maturity in n years is: $P_0 = \text{Principal}/(1 + r)^n$, increased demand for long term bonds will increase the price, P_0, which implies that the interest rate, r, has to decrease. The reverse is also true, when consumers believe the future is bright and that their real income will increase, they start to spend more today, to smooth their consumption. This behavior reduces saving and demand for long term bonds, reducing the price and increasing the implied interest rate.

To add to the third argument; when consumers start to save more because they expect a recession in the future, they will naturally have less money to consume in the present. Reduced spending by consumers will then decrease the demand side of the economy, reducing growth. This means that just the *expectation* of a future recession and the associated reduced spending, does itself increase the probability of a future recession. This means that the expectation about the future of the economy, which can be reflected in the slope of the yield curve, can be a self fulfilling prophecy.

This means that most of the information provided by the yield curve reflect market participant expectations about future economic activity. This means that it is not the slope of the yield curve itself that *cause* recessions. Rather, the yield curve compresses expectations about the future into a single number; the spread.

In their article "Does the Yield Curve Really Forecast Recession?" Spewak and Andolfatto (2018) discuss the rational behind the yield curve inversion and recessions. The authors argue that in an economy that grows unevenly over time, a negative shock is more likely to push the economy into a recession if the economy is already in a low-growth state. Further, they argue that a flattening yield curve only predicts lower consumption (particularly due to reason number three above), leading to lower growth, but not necessarily a recession. However, a recession is much more likely to occur if a shock hits when the economy is already growing slowly. They conclude by saying that the exact date that a negative shock occurs is unpredictable. But that it is more likely that such a shock leads to a recession if the economy is already in a low-growth state. It might be this low-growth state that the yield curve excels at predicting.

3.2 Recession and output gaps

When researching whether the information provided by the yield curve can forecast economic downturns the definitions of these events are crucial. Therefore I will use this section to go through the definitions and estimation techniques used to define these events in this thesis.

[1]The assumption of stable consumption over time comes from Friedman's permanent income hypothesis (Spewak and Andolfatto 2018).

3.2.1 Estimating Historical Output Gaps

The output gap of an economy is a theoretical concept that attempts to capture the difference between the current state of the economy and the potential state of the economy, and this state is usually measured by GDP (Hagelund et al. 2018). While there is widely available data on countries' current and historical GDP measured on a quarterly basis, data on potential GDP is harder to come by. The reason probably is that potential GDP cannot be observed directly and there is no widely agreed upon consensus on a definition. As Gavin (2012) points out; "Potential gross domestic product (GDP) is a theoretical concept that means different things to different people."

The strict theoretical definition of potential GDP is that it reflects the GDP in an economist's perfect world. In this perfect world all resources are utilized in it's optimal way; every employee is matched to an employer where her abilities are perfectly inline with the employer's needs, all good ideas are put into practice, all markets have perfect competition, there are no taxes, and so on and so forth (Gavin 2012).

The above definition of potential GDP is, however, not very useful in the real world. Since policymakers typically use the output gap as an important tool when deciding on whether the economy needs more or less stimuli through either monetary policy, fiscal policy or both, they need a more practical definition.

According to Gavin (2012) monetary policymakers use econometric methods to estimate a trend from historical GDP data. These methods normally attempt to extract a trend component from the GDP series by removing business cycle fluctuations. The author then goes on to claim that an estimate of historical potential GDP, and by extension an estimate of historic output gap, is relatively unproblematic to estimate using one of several trend extracting methods, while a real time measure of potential GDP is a more complicated problem. For the purpose of this thesis precise real time estimation of potential GDP is not important as the analysis is based on historical data.

Hagelund et al. (2018) claims that simple univariate methods, also referred to as statistical filters, are the most commonly applied methods used to estimate potential GDP. He points out that these methods are normally quite simple to use in practice, and mentions the Hodrick-Prescott filter (HP filter) as an example of one such method. Jahan and Mahmud (2013) also mentions the Hodrick-Prescott filter as a popular approach. Hjelm and Jönsson (2010) discusses more advanced methods such as Structural Vector AutoRegression, Unobserved Components and Multivariate HP-filters. Hagelund et al. (2018) reports that the Norwegian bank uses multivariate models that not only utilize GDP data, but also data on variables such as inflation, investment and unemployment among others.

My impression is that the more advanced methods are more useful for accurate real time estimates of output gaps, and for forecast. However, as stated above the simpler methods that quite accurately estimates historical output gaps such as the HP-filter is more appropriate in this thesis.

Poloni and Sbrana (2017), among others such as Ravn and Uhlig (2002), reports that, despite criticism from Hamilton (2018) and others, the HP-filter is one of the most common techniques used for trend-cycle extraction in the macroeconomic literature. Gogas et al. (2015), which is referred to multiple times in this thesis are among the authors that uses this approach. Due to the reasons mentioned above I have chosen to use the HP-filter to find historic output gaps. The paragraphs below describe the technicalities of how the filter works.

The Hodrick-Prescott filter

Although Hodrick and Prescott (1997) where not the first to propose this method, they where the ones to popularize it in the field of macroeconomics. The method strives to decompose a time series into a trend component and a cyclical component. The method has been applied to multiple macroeconomic time series, but in the case of the GDP series the trend component is interpreted as potential GDP. So the output gap can simply be calculated as actual GDP minus this estimate of potential GDP.

The HP-filter assumes that the underlying data generating process of a time series y_t consists of a trend component τ_t and a cycle component c_t:

$$y_t = \tau_t + c_t, \ \ t = 1, \dots, T.$$

If the parameter λ is positive a $\{\tau_t\}$ exists that solves:

$$\min_{\{\tau_t\}_{t=-1}^{T}} \left\{ \sum_{t=1}^{T} (y_t - \tau_t)^2 + \lambda \sum_{t=1}^{T} [(\tau_t - \tau_{t-1}) - (\tau_{t-1} - \tau_{t-2})]^2 \right\}$$

The first term is the sum of squared deviations from the trend τ_t, or the cyclical component, c_t. It is referred to in the literature as the business cycle component of the series. This term of the HP-filter makes sure the cyclical component does not get to big. According to Hodrick and Prescott (1997) the second term, except λ, is a measure of the smoothness of the trend, $\{\tau_t\}$. The higher the parameter λ is the smoother the series $\{\tau_t\}$ will be.

The value of λ has been discussed in the literature, and the most common value to use for quarterly data appears to be 1600, which is what Hodrick and Prescott (1997) chose in their early paper. Ravn and Uhlig (2002) conducts an analytical study of how λ should be adjusted to accommodate different frequencies of observations, and use $\lambda = 1600$ for quarterly data and recommends, based on their results, to change λ using the following formula;

$$\lambda = 1600 * \eta^4$$

where η is the change in frequency of the time series' observations from quarterly. This means that they recommend 6.25 ($= 1600*(1/4)^4$) for annual data, and 129 600 ($= 1600 * 3^4$) for monthly data. This thesis estimates output gaps from quarterly data so the λ used is 1600.

3.2.2 Recessions

As were the case with defining output gaps, neither recessions has an unambiguous definition in the literature. In their early paper on recession prediction, using the information provided by the yield curve, Estrella and Hardouvelis (1991) used data from the National Bureau of Economic Research (NBER). According to Gaski (2012) the NBER recessions are decided upon by NBER's Business Cycle Dating

Committee. And they loosely define a recession in the following way; "A recession is a significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in production, employment, real income, and other indicators" (NBER 2008). Gaski (2012) goes on to criticize NBER for this definition as it is not even remotely precise and thus the NBER recessions can be subject to the inevitable subjectivity of the members of the committee. He instead argues that a very precise, yet very simple definition should be used. The definition he proposes the NBER starts to use, which is in no way new, is two consecutive quarters of negative output growth. This definition is also used quite frequently in the literature (see for example Ozturk and Pereira (2014)).

The definition that I have chosen to use is two consecutive quarters of negative GDP growth. In addition to the fact that this definition is commonly used in the literature another advantage is that this definition is very easily applicable and transparent and can therefore easily be applied to different countries in a consistent way.

Chapter 4

Data

This section describes the data used to estimate all the models in this thesis. Three datasets are used in this thesis, one for each of the three Scandinavian countries. The time series for the different variables are of varying length, however, all datasets are limited to be of the same length, this is done in order to make comparisons fair across countries. The length of the dataset is described in following table: The

	Norway, Sweden and Denmark
`Dataset`	1987-01 until 2019-06

Table 4.1: *Length of dataset.*

relevant variables included in the dataset is listed below, in table 4.2.[1]

The rest of the sections in this chapter explains each of the variables in greater detail with respect to its source and, if relevant, how it was calculated/estimated.

4.1 GDP Growth

GDP, or Gross Domestic Product, is a monetary measure of the value created in a specific country in a specific period of time, usually during a quarter or a year (Holden 2016). GDP is usually considered one of the most important indicators in terms of how a country is doing economically. According to Okun's Law GDP is also closely tied to unemployment. This relationship has also been confirmed lately for the USA (Ball et al. 2017).

The GDP growth series for each of the three Scandinavian countries is not used directly in any of the models, but rather, to estimate the two binary indicator variables, *recession* and *output gap*.

All the data on GDP is downloaded from OECD's website and has a quarterly frequency.[2] The OECD data on GDP growth is adjusted for inflation, which means

[1]I say *relevant variables* because some other variables are included as well but only for the purposes of calculating others. For example, the dataset contain the trend and cycle variables returned by the HP-filter method used to calculate the recession variable, but these variables are not included in the list above because they were not directly used in the analysis part of this thesis.

[2]This is the URL to OECD's website: https://data.oecd.org/gdp/quarterly-gdp.htm#indicator-chart

Dataset
Date
10y interest rate
3m interest rate
Spread variable (10y-3m)
GDP growth rate
Output gap indicator
Recession indicator
Stock market index

Table 4.2: *Variables included in the dataset used, for each of the three countries.*

that we are dealing the real GDP growth. The GDP growth variable is also seasonally adjusted by the OECD[3].

Since the data downloaded from the OECD's website is a growth variable I have also calculated a GDP index, in order to be able to apply the HP-filter discussed in section 3.2.1.

4.2 Interest rates

The interest rates used in this thesis is also downloaded from the OECD's database, but have are of a monthly frequency. What is referred to as long term interest rates is government bonds with a 10 year maturity[4], while the short term interest rates are based on 3 month money market rates.[5] These 3 month bonds are typically referred to as "money market rates" and "treasury bill rates". The `spread` variable is then simply calculated as `10y interest rates` minus `3m interest rates`.

4.3 Recessions

The recession variable is estimated based on the GDP data downloaded from OECD's website. Recessions are estimated using the definition described in section 3.2.2 on the quarterly GDP index series. The `Python` programming language is used for the calculations.

It is important to note that the GDP data is of a quarterly frequency while the interest rate data and the other variables are of a monthly frequency. When estimat-

[3]Gogas et al. (2015) also used seasonally adjusted GDP.

[4]Downloaded from this URL: https://data.oecd.org/interest/long-term-interest-rates.htm#indicator-chart

[5]Downloaded from this URL: https://data.oecd.org/interest/short-term-interest-rates.htm#indicator-chart

ing statistical models and machine learning models the number of observations, and thus the amount of information, is critical. That is why this thesis uses a method called upsampling[6] on the recession variable in order to be able to use the higher resolution of the interest rate data. This essentially means that the datasets are of a monthly frequency even though the GDP data is of a quarterly frequency. The practical implication is that all months within recession quarters are also defined as recessions, the same goes for non recession months. The assumption that months within a recession quarter is also defined as a recession is in my opinion not a controversial[7] decision, but it helps take advantage of the higher frequency of the interest rate variable, and also all the other variables, when estimating models. Table 4.3 shows how many recessions are in the dataset for each country.

	Norway	**Sweden**	**Denmark**
`Dataset`	30/390 (7.7%)	57/389 (14.7%)	54/390 (13.8%)

Table 4.3: *Number of recessions over total number of observations. Percentage of recession months are in parenthesis. Also, note that Sweden has one less observation than the other two countries due to short term interest data is missing for november 2001 in the dataset from OECD.*

Note that table 4.3 describes the entire dataset, while some of the models estimated in chapter 6 have left out the first λ_{max} observations, where λ_{max} is the lag used in the model with the longest forecasting horizon. This is done in order to include the same number of observations in all models. This had to be done since different models uses different lags and if I had not limited the observations used, then the models with longer forecasting horizons would be estimated on fewer observations. Of course, limiting the dataset has the obvious drawback of a smaller dataset (which leads to more uncertainty in the models), however, it gives the benefit of making comparisons between models fair. The limitations are between 12 and 18 observations depending on the analysis. The probit models, in which the aim is to assess the statistical relationship between the yield curve and recessions/output gaps, are limited by 18 observations. All the out-of-sample analysis are limited by 12 observations, since the maximum forecasting horizon used iss 12 months.

4.4 Output gaps

Output gaps are also estimated using the GDP data downloaded from OECD's website. The estimation process is a bit more advanced for output gaps than for recessions. In section 3.2.1 the technicalities of the HP-filter, and why I chose to use it to define output gaps in this thesis, is discussed. I have used the Python package `Statsmodels` (Skipper and Perktold 2010) to implement the HP-filter on the GDP

[6]This is the specific python fuction used (which is a part of the Pandas data analysis library): `df.resample('MS').ffill()`

[7]It can however, lead to a problem of autocorrelated errors. This problem is discussed in length in section 5.1.2.

time series[8]. The python function then returns a trend variable and a cycle variable, which is interpreted as a potential GDP component and a business cycle component. Next, the output gap is calculated by taking the difference between the GDP index variable and the trend variable, and looking at the sign of the resulting variable. If the sign is positive, then GDP is above the trend and we have positive output gap. The reverse is also true; if the sign is negative then the economy is in a negative output gap.

The way I have chosen to define the output gap variable is to set it equal to one if the output gap is negative, since predicting future downturn is of primary interest. And, if the economy has a positive output gap the output gap indicator variable is equal to zero. The output gap data is also upsampled in the same way that the recession variable is. Table 4.4 displays the number of ouptgaps in each dataset for each country.

	Norway	**Sweden**	**Denmark**
`Dataset`	198/390 (50.8%)	179/389 (46.0%)	216/390 (55.4%)

Table 4.4: *Number of negative output gap over number total number of observations. Percentage of negative output gaps are in parenthesis.*

Note that the same limitations of the dataset holds for the output gap data as for the recession data discussed in the previous section (section 4.3).

4.5 Stock Market Index

The value of a share is the net present value of expected future profit from a company divided by the number of shares. A stock market index is the average of many shares.[9] This means that a stock market index for the top n companies in a country such as the OBX index in Norway, which tracks the development of the 25 most liquid stock on Oslo Stock Exchange, can be considered an expectation of how all these future companies will develop in the future. Since the effect of diversification removes most of the unsystematic risk in such a large portfolio, changes in a large stock market index can contain information about investors' expectations of future economic developments in the country. It follows from this argument that a change in the stock index for the top 20-30 stocks in small countries such as the Scandinavian countries, could potentially help predict future economic activity.

A number of the studies that look at the yield curve's ability to forecast future economic activity, also tested whether it performed better than lagged changes in a large stock index (see for example Harvey (1986) and Estrella and Mishkin (1996)). This is also done in this thesis but for the Scandinavian countries.

[8]The specific method used is `statsmodels.tsa.filters.hp_filter.hpfilter(*time series array*, lamb=1600)`. See the `Statsmodels` package's documentation for more information about this function: https://www.statsmodels.org/stable/generated/statsmodels.tsa.filters.hp_filter.hpfilter.html

[9]For most stock indexes each stock is weighted by its market capitalization divided by the total market capitalization for all the stocks included in the index (Banton 2020).

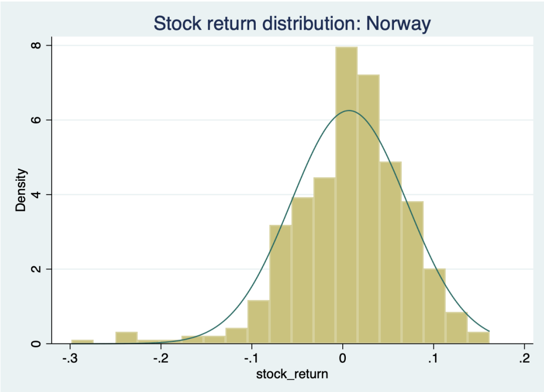
When choosing the indexes to include in the models, it is important to use an index that represents the economy in question quite well. It is, of course, impossible to find an index that completely represents the theoretical *market portfolio* that is used a lot in finance, however, it is important to try to find something that is similar. Another factor to consider is how long time series are available on different indexes. Also, it is important that the indexes are comparable across countries and time. Morgan Stanley Capital International or MSCI is a research firm that provides a large selection of indexes. The MSCI indexes that covers Norway, Sweden and Denmark are "designed to measure the performance of the large and mid cap segments of the [Norwegian, Swedish, Danish] market".[10] I have used the Thomson Reuters Eikon Time Series Request function in Excel to download these datasets. Thomson Reuters returns indexes that are based on values of the companies in local currencies (in contrast to what the PDFs linked to in the footnotes says). The final version of these datasets where downloaded and matched to the other datasets on 15th of May 2020. Table 4.5 gives more details on the indexes used for each country.

Country	**Stock index**	**Details**
Norway	MSCI Norway	10 companies
Sweden	MSCI Sweden	32 companies
Denmark	MSCI Denmark	17 companies

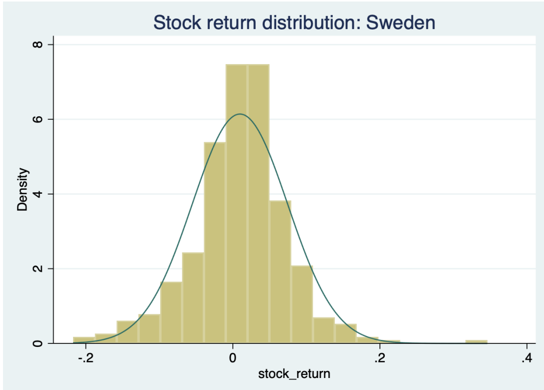
Table 4.5: *The stock indexes used in this thesis.*

Because the stock market data is only included in the comparison sections of this thesis, and not given the same space in the results/analysis chapter, some basic descriptions of the dataset are given here. Figure 4.1 gives indications that the stock return variables are not completely normally distributed. It it not possible to say this for sure, as I only have a sample of each country's stock return distribution, but one can get some indications by looking at histograms over such long time periods, as it is unlikely that the real return generating process deviates a lot from this, as the sample is quite large. It is, however, also possible that the underlying distribution has changed over time. This is often the case with stocks and stock indexes as the volatility often cluster around certain periods, for example Predescu and Stancu (2011) found, by using ARCH and GARCH models, that the volatility of some major stock indexes was higher around the financial crisis of 2008. I will not, however, go into more details on this in this thesis, as the main variables of interest here are yield curve variables. The distribution seems to have excess kurtosis, with more observations centered around the mean. The distributions also appear to exhibit heavy tails, which means that extreme, positive and negative outcomes appear more likely than what the normal distribution would predict.

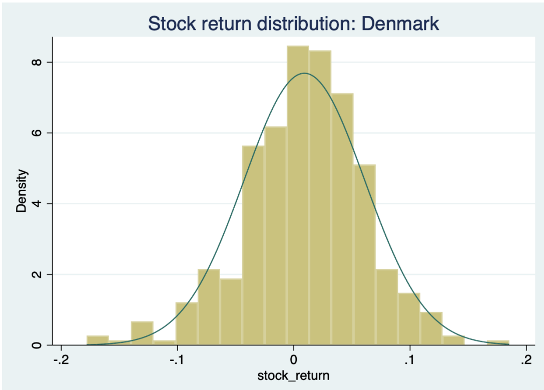
[10] Links to the PDFs that describe each index in more detail: https://www.msci.com/documents/10199/9d0f5852-2652-4307-9f60-9fe2724c6e22, https://www.msci.com/documents/10199/5b5d91b7-505a-4d4d-b060-51a3af6be160 and https://www.msci.com/documents/10199/5db4fa3f-1775-4d39-8838-e260a97d2b94.



(a) Norway



(b) Sweden



(c) Denmark

Figure 4.1: *Plots showing the distribution of stock returns from 1987m1 - 2019m6. The green curve is the normal distribution with mean and standard deviation equal to that of each stock return distribution.*

4.6 Estimation and Testing Data

In the analysis part of this thesis pseudo out-of-sample forecasting will be used to assess the performance of the models that are estimated. For the purposes of running tests on the datasets described above they will have to be divided into two, one part for estimation and another for testing. This can be done in multiple ways. One way is to use a 80/20 split where, the first 80% of observations are used for estimating the model and the final 20% is used for testing the model. In the case of recessions, which is a rare event, this might not work so well. If we look at the most recent 20% of the dataset, Norway and Sweden have only a single recession each and Denmark has none (see figure 6.1). As this is the case, I believe that randomly splitting the dataset might be a better solution, since this increases the chance of having recession periods in both the testing set and the estimation set.[11]

Since this thesis is all about forecasting and using lagged variables as explanatory variables, the lagged variables must be created *before* the split. The reason is that

[11]I say that a random split increases the chances of having recession periods in the testing set because the most resent periods in the dataset contains "unusually few" recessions, which can be seen from figure 6.1.

because the split is random, it does not makes sense to lag variables afterwards, as no observations will be in the original and correct order. The `shift()` function from the `Pandas Python` package is used for this purpose.

The procedure for dividing the dataset involves using `sklearn` in `Python`. This `Python` package has a function called `train_test_split` and takes the two arguments, `test_size` and `random_state`. The `test_size` is chosen to be 30%. This number is a little larger than the usual 20% used. The reason for having a little larger testing set is that recessions is such a rare event. So, in order to have a chance of including some months of recessions in the testing set also, I believe that a larger test set is appropriate. Although output gaps are not a rare event, the same estimation to test ratio is used.

The second argument that is used in the `train_test_split` function is the `random_state` parameter. This is effectively a *seed* for the (pseudo) random number generator used by the `sklearn` package. This parameter could be left empty and then be chosen randomly each time the script is run, but in order to be able to run the `Python` script multiple times to make sure it works, I chose to manually set the seed. However, it is important to note that the `numpy` library in `Python` was used to determine this number[12], before it was hard coded into the script. Using an explicit randomly chosen seed also makes it easier for others to replicate my results. The seed that is used is: 3411197133. Table 4.6 give more details on how the dataset is divided.

`Dataset`

	Norway	**Sweden**	**Denmark**
Recessions	20 / 10	40 / 17	43 / 11
Negative output gaps	143 / 55	130 / 49	150 / 66
Total nr of observations	273 / 117	272 / 117	273 / 117

Table 4.6: *Details on estimation and testing data for the dataset used in this thesis. On the left of each column are numbers describing the estimation dataset, and on the right are numbers describing the testing dataset.*

Again, table 4.6 shows details on the entire dataset, while many of the models estimated exclude some of the first observations in each dataset. See section 4.3 for more details.

[12]The `Numpy` function `np.random.uniform(low=low, high=high)` where used. The `low` parameter was set to zero, and the `high` parameter was set to $2^{32} - 1$, which is the range of numbers allowed by the `train_test_split` function. The number returned by the function was then rounded using the built in `round` function in `Python`.

Chapter 5

Methodology

In this thesis two fundamentally different estimation techniques will be used in order to attempt to forecast recessions/output gaps.[1] First, the tradition in the relevant literature on forecasting recessions has been to use a type of a binary response model such as logistic regression or the very similar probit regression. Estrella and Mishkin (1996), Estrella and Trubin (2006) and Evgenidis and Siriopoulos (2014) are among the authors that take this approach. These models fall under the category of a parametric method.[2] The second estimation technique is a type of non-parametric model. When using one of the models within this category, the researcher does not assume a functional form of the underlying population model of the relation between the variables in question, rather, the model itself involves testing many different functional forms and, with the use of sophisticated methods, choose the "best one" based on the data.

5.1 Probit regression

As stated above, binary response models has been used a lot in the literature, and in particular the probit model. The probit model yields very similar estimates and statistics to logistic regression models (logit). The reason one of these models are chosen over linear regression (in this context often referred to as the linear probability model), which is widely used in economics and finance, is that the dependent variable is a binary variable. When we are dealing with binary dependent variables the results from a regression model is commonly interpreted as the probability of the dependent variable being equal to one. The problem with linear regression is that it does not constrain its estimates and so it can yield a probability both higher than one and lower than zero, which is nonsensical. Probit and logit models are a type of non linear model that deals with the problem that the regular linear regression model presents. The only key difference between the two non linear models is that probit uses the cumulative standard normal distribution function, denoted Φ and the logit

[1] The `Python` and `R` code written to estimate and test the non-probabilistic models and the SVM models can be found on my GitHub: https://github.com/sondreandersen96/masters_thesis_code

[2] This is by far the most common approach to statistical modelling in economics, and it involves a two-step process; first, the researcher must specify a functional form of the relation between the dependent and the independent variable(s) and then use data to fit the specified model James et al. (2017).

model uses the cumulative standard logistic distribution function Stock and Watson (2015, p. 442).

As the two models yields similar results and the choice between the models mainly comes down to personal preferences, I have chosen to use the probit model as I find it a little easier to convert the z-values it outputs to probabilities, as one can simply use the standard normal cumulative density function.[3]

Generally, with n independent variables the probit model can be written as follows:

$$Pr(Y = 1|X_1, ..., X_n) = \Phi(\beta_0 + \beta_1 X_1 + ... + \beta_n X_n)$$

If we calculate $\Phi(z)$ for specific values of $X_1, ..., X_n$, where $z = (\beta_0 + \beta_1 X_1 + ... + \beta_n X_n)$, we get an estimate of the conditional probability that the dependent variable is equal to one given some values of the independent variables.

Since probit is a non linear model Ordinary Least Squares (OLS) can not be used to estimate coefficients. According to Stock and Watson (2015) most regression software uses an algorithm that maximizes a likelihood function when estimating probit models.[4]

5.1.1 Metrics for evaluating the probit models

When testing whether the probit models estimated in this thesis are useful tools in forecasting economic downturns some measures of performance needs to be decided upon. In this section I will go through these measures.

P-value

The probit model outputs p-values for all its coefficients, in the same way that linear regression does. The p-value, when $H_0 = 0$, is the estimated probability of observing a coefficient that is at least as far away from zero (H_0) as the coefficient estimated, assuming the coefficients are normally distributed when choosing a random sample from the population and that the true population coefficient is actually zero. So if the p-value is less than some significance level, α, the conclusion is that the observed outcome is so unlikely to happen if the true coefficient is zero, that we believe it is not. But of course, the p-value is the probability that we wrongly reject the null hypothesis.

In this thesis I will follow the norm in the field of economics and use an α of 5%. With a value of 0.05 I get a decent balance between type 1 and type 2 errors. I will use the p-values to asses if there is a statistically significant relationship between the dependent and independent variables.

Pseudo R^2

A commonly used measure of fit is R^2. R^2 does, however, not work well with models that have a binary dependent variable (Stock and Watson 2015, page 447), which is the case in this thesis. Stock and Watson (2015) instead suggest two other measures

[3]In STATA 16 I use the function `normal(z)`. Also, when using the `predict` command, STATA 16 automatically calculates probabilities unless something else is specified.

[4]It is my understanding that STATA 16 is using this estimation method.

of fit. The first measure is called *fraction correctly predicted*. This measure counts the fraction of correctly predicted observations by using a threshold of 50%. In section 5.1.2 I explain why a threshold of 50% might not be wise to use in this thesis.

The second measure is called *pseudo* R^2. According to Stock and Watson (2015, page 448) pseudo R^2 uses the likelihood function that was used to estimate the model in the first place to measure the fit of the model. The pseudo R^2 then compares the value of the likelihood function in the model estimated with the value of the likelihood function when no regressors are included. This measure then explain how much better the model that is assessed performs than a model that have no explanatory variables. It is the pseudo R^2 that will be used to assess the in-sample performance of these models. A modified version of the *fraction correctly predicted* will also be used, this method is explained in the section below on non-probabilistic forecasts.

BIC

An important question when specifying the functional form of a time series regression is how many lags, p, to include of the independent variable(s). The number of lags to include is a trade-off between bias and estimation error. If too few lags are estimated while the true population regression (unobservable) contains more lags, the regression will contain omitted variable bias, and valuable information in other lags are lost. On the other hand, if too many lags are included the estimation errors of the regression increase (Stock and Watson 2015, page 593).

A common way to find an *optimal* number of lags, p is by minimizing an *information criterion*. Two commonly used information criterion are BIC and AIC. Since BIC is the only one of the two that gives consistent estimates of p, this thesis will only use BIC (Stock and Watson 2015, page 594).

BIC, or Bayes information criterion, is calculated in the following way when dealing with models like the probit model that is estimated by maximizing a likelihood function:

$$BIC(p) = -2lnL + k * lnN$$

where lnL is the maximized log-likelihood function of the estimated model, k is the number of parameters estimated and N is the number of observations. The first term can only decrease or stay the same when more independent variables are added to the regression. This term is likely to decrease the BIC value as more lags are added. To compensate for this, the second term increases the BIC score as more lags are added, to balance the two effects described above.

The BIC score can be used to compare models in general, not only models with different lag lengths. However, it is very important that all models are estimated using the same number of observations (Stock and Watson 2015, page 595). In order to be able to do this, regressions might have to be restricted to a shorter time period, depending on the length of the longest lag variable, than what the datasets described in table 4.1 suggests.[5]

[5] In `STATA` a command of the following syntax could be used: `probit recession L6.spread if tin(1900m1, 2010m1)`.

5.1.2 Autocorrelated Regression Errors

An important assumption in linear regression as well as in probit regression is the assumption of independenty and identically distributed (i.i.d.) observations. This effectively means that all observations are drawn from the same underlying distribution and that knowing the value of one observation does not give information about the value of other observations, such as the next or previous observation. The dependent variables in this thesis, recessions and output gaps, could potentially contribute to violating this assumption, and population regression errors, ϵ_t, might be autocorrelated. This is probably true for several reasons: 1) Ressesions are defined as two or more consequitve quarters of negative GDP growth, this means that one recession never occur alone but is always followed or preceded by at least one other recession. 2) Both the recession variable and the output gap variable are upsampled[6], this also naturally induced autocorrelation in the dependent variable which might also result in autocorrelated population regression errors. The problem of having autocorrelated errors in the population is that the standard errors will not be consistent.

Due to the nature of only working with samples of data drawn from a population, one can never be certain that population regression errors are correlated (or not) but it is possible to do an hypothesis test, much in the same way p-values are used to test if $\beta_x = 0$. This means that if the *estimated* regression errors, $\hat{\epsilon}_t$, also know as the regression residuals, are autocorrelated one have reason to believe that the true regression errors, ϵ_t, are also correlated. In the next section such a test is conducted.

In section 5.2 it is stated that estimating unbiased estimators is *not* a very important issue in this thesis since the purpose is *forecasting* and not *parameter estimation*. Although the problem of bias is not as important in a forecasting setting, consistent standard errors are. The reason is that if the standard errors are not consistent then the inference statistics such as p-values, which are vital for the hypothesis test that are conducted in the analysis section, are invalid. In the section below I test some of the most central models in this thesis to check if it is necessary to correct for autocorrelation.

Are Regression Errors Autocorrelated?

In this thesis a large number of models are tested to see which lag lengths and number of lags are most appropriate in the probit models. This section tests for autocorrelated regression errors in some of the models used in this thesis to find out whether or not to correct for this problem in the analysis section.

There are many ways in which to check for autocorrelated regression errors, but since the true regression errors, ϵ_t, are unobservable the regression residuals, $\hat{\epsilon}_t$, are used as estimates. Listed below are the approaches taken in this thesis:

1. Visualization of residuals

2. AR(1) model for residuals

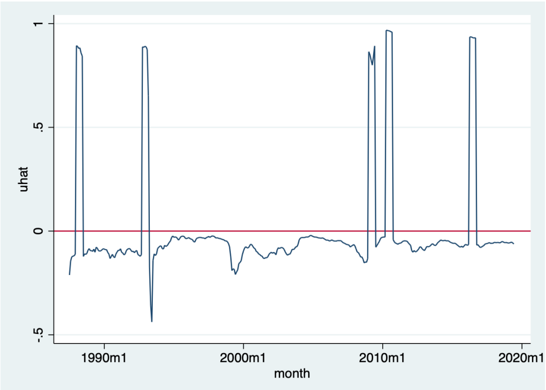
Autocorrelation is in this case pretty clear for the simple recession forecasting model by just looking at a plot of the residuals from a regression. Figure 5.1 displays

[6] See section 4.3 for explaination on the upsampling method.

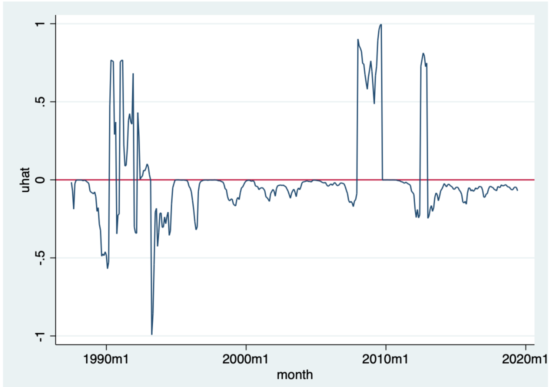
the residuals obtained when running the following regression for each of the three countries: [7]

$$recession_t = \beta_0 + \beta_1 * spread_{t-6} + \epsilon_t$$

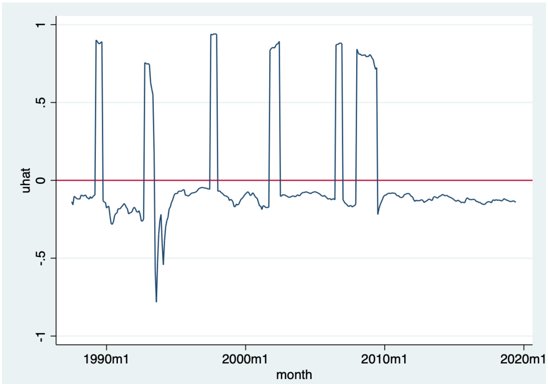
Figure 5.1 gives strong indications that this regression contains autocorrelated er-



(a) Norway



(b) Sweden



(c) Denmark

Figure 5.1: *Graphs showing plotted residuals, $\hat{u}$, obtained when estimating the equation above with a probit model, with* **`recession`** *as the dependent variable.*

rors. It is also clear by analyzing the residuals for all the countries that the mean of $\hat{u}$ is very close to zero, which is desirable in terms of meeting the assumptions underlying the model. Using other lag lengths than six, which was chosen above, yields almost identical results. By regressing the residuals on the first lag of the residuals the the hypothesis of uncorrelated residuals are rejected at a one percent significance level for all countries, results are shown in appendix A.1.1 on page 69. This is evidence of *first-order autocorrelation*. The results from both the visual approach and the AR(1) model on residuals indicate that the population regression contain autocorrelated errors, how to deal with this problem is discussed in the next section.

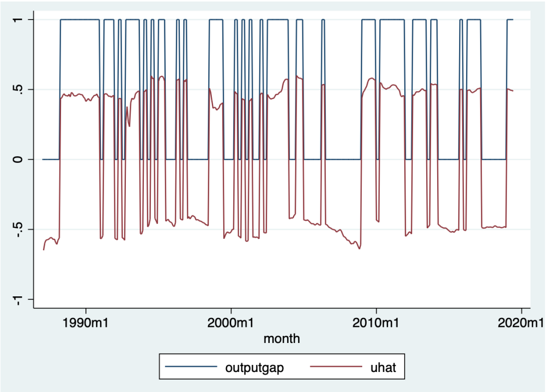
Next is an examination of autocorrelation in the regression with `output gap` as the dependent variable. Again, only a simple model that uses the sixth lag of the

[7] Residuals are calculated as $Y - \hat{Y}$, where Y is the data point for observation t which is either 0 or 1, and $\hat{Y}$ is the predicted probability (obtained by using the `Stata` function `predict`).

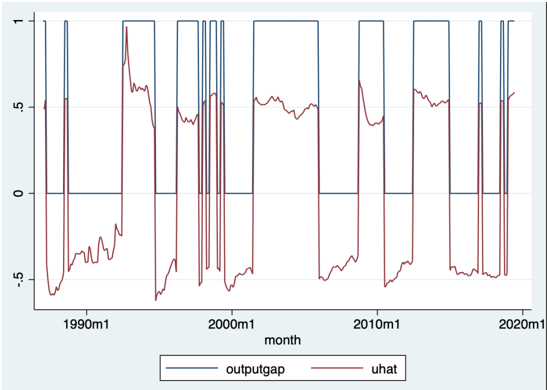
term spread as the independent variable is shown explicitly here, the results are also double checked by running regressions with other lag lengths. Figure 5.2 shows the residuals from the following probit regression:

$$outputgap_t = \beta_0 + \beta_1 * spread_{t-6} + \epsilon_t$$

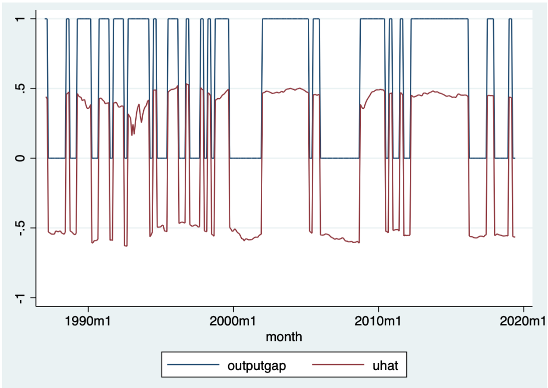
Looking at figure 5.2 it seems clear that the results from regressions with `recession`



(a) Norway



(b) Sweden



(c) Denmark

Figure 5.2: *Graphs showing plotted residuals, $\hat{u}$, obtained when estimating the equation above with a probit model, with* ***`output gap`*** *as the dependent variable.*

as the independent variable also holds with `output gap` as the independent variable. It is also quite clear that the residuals cluster around the dependent variable, which is natural in the case of probit models. Again, the indication of autocorrelation from the plots are confirmed by an AR(1) model on residuals, see appendix A.1.2 on page 70. Using other lag lengths yields very similar results, which means that autocorrelated errors is likely to be a problem in this model as well.

Other test of autocorrelation also exist, such as the Durbin-Watson statistic, which is a formal statistical test of first-order autocorrelation. However, since the hypthesis of no autocorrelation was rejected at alpha one percent level for all models, such a test does not seem necessary.

Dealing with Autocorrelated Errors

Since autocorrelated errors is likely to be a problem, both when `recession` and `outputgap` is the dependent variable, this thesis estimates heteroscedastic and autocorrelation consistent (HAC) standard errors for all the probit models.

HAC standard errors can be obtained in multiple ways. One way is to compute Newey-West standard errors. The Newey-West SEs can be applied to OLS regressions to correct for autocorrelation and heteroscedastic errors, however, this thesis used the non-linear probit models, so the Newey-West method may not be the most appropriate approach.

Instead, clustered standard errors are estimated in this thesis. The reason this method is chosen is to correct for the consequences of the upsampling method that was applied to the `recession` and `outputgap` variables (see section 4.3 and 4.4). The upsampling of these two variables lead to autocorrelation within quarters, since the quarterly variables were essentially copied three times in order to get data with a monthly frequency. In Stata the `cluster(quarter)` option is used, where the quarter variable is of a `YearQuarter` format such as "2010q1".

Non-probabilistic forecasts

In his paper, Ratcliff (2013) found that the probit models work best when assigning non-probabilistic forecast to future recessions (for a short summary of his paper see section 2). So beyond assessing the p-values of the probit coefficients I will focus on these non probabilistic forecasts.

Ratcliff (2013) finds a probability threshold, W, that triggers a yes/no forecast on future recession. Many different methods can be applied to get such a binary estimate. A common threshold is 50%, however, in this case recession is a relatively rare event compared to non-recessions so the threshold of 50% might be too high. In this thesis I will follow Ratcliff (2013) and select a threshold based maximizing the difference between the hit rate and the false alarm rate. However, Ratcliff (2013) did not only maximize this difference, rather, he used a combination of of minimizing in-sample bias (the difference between the number of predicted recessions/positive or negative output gaps, and the observed number of such events) and maximizing the difference between *hit rate* and the *false alarm rate* . Section 5.3 goes through all these metrics in details. Ratcliff (2013) also maximized an Equitable Threat Score, which essentially is a metric that is suited for evaluating models that predicts infrequent events. Instead of using this metric to assess the models performance in forecasting infrequent events this thesis uses the *precision* metric described in section 5.3. As is explained in section 5.3 this metric is also well suited when the event that one wants to classify is rare.

As stated above, Ratcliff (2013) used a combination of three metrics to determine the optimal threshold, W. The reason this thesis only maximizes the difference between hit rate and false alarm rate is in order to have a criterion that is as objective as possible, to make comparisons between countries and lags fair. Still, some of the other metrics mentioned here will also be discussed.

5.2 Support Vector Machine

In addition to the probit models explained in the previous section this thesis will also explore how well a type of supervised machine learning (ML) technique[8] called Support Vector Machine (SVM) works in predicting downturns in the economy. Gogas et al. (2015) have done this for US data, see the literature review in section 2.

Recently, several economists and finance researchers have been experimenting with ML algorithms for forecasting purposes. Both Ince and Trafalis (2006) and Brandl et al. (2009) has successfully used SVM to forecast exchange rate movements. Gogas et al. (2015, page 212) mentions additional examples in their paper.

Logistic regression, probit regression and SVM can all be used for classification. The advantage that the SVM can have over the other methods, with regards to forecasting, is that it is a type of non-parametric method, which means that the researcher does not have to specify a functional form of the relationship between the dependent and the independent variable(s). In their paper on Machine Learning in econometrics Mullainathan and Spiess (2017) claims that ML excels at *prediction*, that is $\hat{y}$ problems. The reason why this is it, is, according to the authors, that ML algorithms in many settings are good at finding generalizable patterns. Further, they state that the success of machine learning algorithms are "largely due to its ability to discover complex structure that was not specified in advance", hence the non-parametric nature of the methods.

A very common task in the fields of economics and finance is *parameter estimation*. When the goal is to estimate causal relations between variables, parameter estimation is very important. This means that the researcher wants to find the correct values of the betas in the underlying process that generated the observed data. OLS regression is commonly used for this purpose. It is important to stress that unbiased parameter estimation is not the goal of this thesis, rather, forecasting is. Mullainathan and Spiess (2017) stress this distinction when they say that ML algorithms rarely produce consistent estimators. The authors summarize the the difference in use cases quite nicely:

> *Put succinctly, machine learning belongs in the part of the toolbox marked $\hat{y}$ rather than in the more familiar $\hat{\beta}$ compartment.*

Part of the reason ML algorithmns generally and SVM in particular are appropriate when the goal is forecasting is their ability to find generalizable patterns in data and thus perform good out-of-sample. When trying to predict or forecast the value of a variable based on explanatory variables the goal is always to predict well out-of-sample. In out-of-sample prediction a central problem is the *bias-variance trade-off*. According to James et al. (2017, page 36) good out-of-sample performance requires both low variance and low bias. In the field of ML and Statistical

[8]Machine Learning algorithms are generally separated into two distinct categories; 1) Supervised ML and 2) Unsupervised ML. The former refer to algorithms that can be trained on existing data where the response variable is known and the problem revolves around fitting a model to this data. The latter referes to algorihms that attempt to find patterns where the response variable is unknown, one example where unsupervised learning could excel is at customer segmentation, where the segments are unknown but a lot of characteristics are known about each customer. *Clustering* is one example of such an algorithm.

Learning *variance* refers to the sample sensitivity of the estimator. This means that an estimator has high variance if the estimated function f changes a lot when a different sample is used in the estimation process. *Bias* on the other hand, is error that occurs due to how we attempt to approximate a relationship between variables by using a simple model, while the true, underlying process is much more complicated. This means that bias will not go away even if we increased the sample size, but estimated the same model. As James et al. (2017) points out it is easy to achieve the two extreme cases, either low variance by just drawing a horizontal line through the data, for example by simply using the average, but this approach has the drawback of having very large bias due to oversimplification. In the same way low bias can be achieved by simply connecting all observations by a straight line, however this model would have very high variance due to the problem of *overfitting*. Overfitting is problematic because it can give very good in-sample performance, but very poor out-of-sample performance. The SVM uses *kernels* and a form of resampling method called *cross-validation* to deal with the bias-variance trade-off.

In the rest of this section I will explain the concept of a Support Vector Machine and then go on to explain how the method is applied in this thesis.

5.2.1 How does Support Vector Machines Work?

In the following sections I will explain how SVMs work and why they have the potential to outperform the probit models explained in section 5.1. Since SVMs are not as commonly used in economics and finance yet, I will go into some detail about how the algorithm works and how it can be optimized, however, I will not go into the mathematical details of how it works, others have done so nicely already.[9]

The support vector machine (SVM) is an extension to the simpler support vector classifier (SVC). The support vector classifier only allow for linear boundaries in the classification of the dependent variable(James et al. 2017). Since the SVM builds on top of the SVC I will start by explaining how the SVC works.

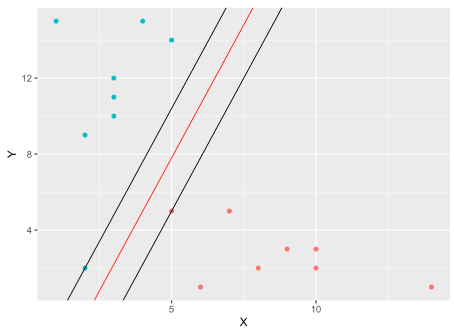
5.2.2 Support Vector Classifier

The basic idea behind the SVC is to find a way of separating observations according to their binary dependent variable. In this thesis I will only look at cases in which the dependent variable is binary but the method can be extended to handle non-binary variables. The way an SVC separates the observations is by finding an *optimal separating hyperplane*. A hyperplane is a kind of generalization of a flat surface in different dimensions. In a p-dimensional space this surface has $p - 1$ dimensions. This means that in a two dimensional space (think of a two dimensional scatter plot) the hyperplane will be a straight line. In a three dimensional space the hyperplane will be a two dimensional plane. The figures below should illustrate this.

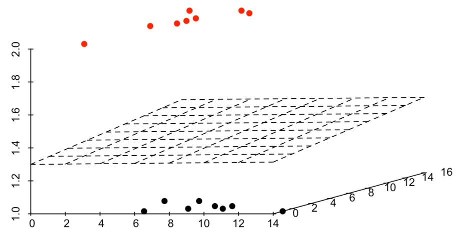
Hyperplanes do also work in higher dimensions than three, however, it is hard to visualize what that would look like. In a p-dimensional case the mathematical formula describing a hyperplane would look like this (James et al. 2017):

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p = 0$$

[9]See for example James et al. (2017, page 337-358) for a great introduction, or, for a more thorough explanation, see Hastie et al. (2017, page 417).



(a) Two dimensional



(b) Three dimensional

Figure 5.3: *Illustrations of a hyperplane in two- and three dimensional space. On the left: the red line is the separating hyperplane, and the black lines represent the margin. The figures are created in* `R` *and is just for illustrative purposes; no SVC has been used.*

The SVC classifies observations based on which side of this hyperplane the observation lies. That is, if the expression on the left hand side is greater than or less than zero. The further from zero the expression is when we insert values for X the more certain we can be that the classification is correct.

So how does the SVC find this separating hyperplane? One way of solving this problem is by using what is called the *maximal margin classifier*. This algorithm finds the hyperplane that creates the largest possible distance, often called *margin*, between the observations closest to the separating hyperplane in each of the two classes. Unfortunatly this method has some drawbacks. First, it only works if all observations can be perfectly separated by a hyperplane, which were the case in figure 5.3, but in many empirical situations this condition is not satisfied. Second, the maximal margin classifier can be very sensitive to new observations, that is, the estimator can have high variance.

A better solution to the problem of finding an optimal hyperplane is to use the SVC. This method is sometimes also referred to as a *soft margin classifier*. The reason the method is called a *soft* classifier is that it, in contrast to the maximal margin classifier, allows some observations to be misclassified. This can be both useful and sometimes necessary to be able to use this method at all. Think of figure 5.3 (a), if the blue point in the bottom left corner was not there the slope of the hyperplane could just as well be much flatter, but due to this one observation the hyperplane is drastically different. This means that the maximal margin classifier can be highly sensitive to outliers.

The SVC finds the optimal hyperplane by solving the following set of equations(James et al. 2017):

$$\max_{\beta_0,\beta_1,\ldots\beta_p,\epsilon_1,\ldots,\epsilon_n,M} M$$

$$\text{s.t.} \sum_{j=1}^{p} \beta_j^2 = 1,$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_p x_{ip}) \geq M(1-\epsilon_i),$$

$$\epsilon_i \geq 0, \sum_{i=1}^{n} \epsilon_i \leq C$$

The parameters in the formula has the following interpretation:

- C: Tuning parameter. C must be greater than or equal to zero.
- M: The width of the margin. That is, the distance between the red and the black line in figure 5.3 (a). The goal is to make this margin as large as we can while not violating the other conditions.
- $\epsilon_1, ..., \epsilon_n$: These are so called *slack variables*. It is these slack variables that allow for some observations to be misclassified.
- $x_{i1}, ..., x_{ip}$: This is the data that is used to fit the model.
- β_0, β_p: Are the coefficients describing the hyperplane.

The most important parameter to a researcher using a SVC is probably C. It is C that allows for adjusting the bias-variance trade-off. Since the formula above shows that C, often referred to as a *cost* parameter, limits the sum of the slack variables, C is often though of as a *budget* for how much the margins can be violated (James et al. 2017, page 347). This means that in the special case that $C = 0$ the SVC will be equivalent to the maximal margin classifier. The higher C is, and the more misclassifications are allowed, the lower the variance of the hyperplane will be, because single observations will not have as high influence. The bias, on the other hand, will be higher.

It follows from the discussion above that the tuning of the parameter C is vital for how well the SVC performs in out-of-sample testing. The value of C is usually decided using a technique called *cross-validation*. I explain this concept in section 5.2.3. But before I go into more details about cross-validation I will dive a little deeper into why the cost parameter is so important and how changing it will affect the resulting hyperplane and, by extension, the performance of the model.

According to James et al. (2017) the SVC has an interesting property that is worth mentioning to better understand why C is a central component in dealing with the bias-variance trade-off problem. The property that is referred to is that when solving the optimization problem above, it can be shown that observations that lies on the correct side of the margin and not on the margin, does not affect the support vector classifier at all. It is only the observations that lie on or on the wrong side of the margin that affects the classifier, and these observations are called *support vectors*. This means that the value of C affects the number of support vectors. If C is small, a small number of observations violate, or lie on the margin and thus there are few support vectors, and a narrow margin. If C is large, many observations will violate or lie on the margin, and thus the number of support vectors will be higher. This intuition makes it clearer why a large C will have many observations to determine the hyperplane and thus be less sensitive to new observations, i.e. it has lower variance. Unfortunatly, a high C will induce a potentially higher bias. If we instead choose a small value for C, the reverse will be true because as we get a large number of observations, that is when the sample size approached the population size, the model will improve dramatically towards the true underlying data generating process, i.e. have low bias. But if we have few observations the estimator is very sensitive to new data, i.e. has high variance.

In the field of machine learning over- and under-fitting is an important topic. It turns out that the choice of C have great influence on this topic. As stated before, a high C results in a wide margin and many support vectors, this means that the higher C is, the higher the chance of overfitting the data. The reverse is true for low values of C. The problem of overfitting is getting seemingly great performance in-sample but often quite bad performance out of sample. The solution to this problem is to use the method of *cross-validation* to determine the C parameter.

5.2.3 Cross-validation

Cross-validation is a type of *resampling* method. Resampling methods can be used primarily for two different purposes; 1) To estimate the uncertainty in the model, 2) to tune a model/compare different models. In this thesis it is used for tuning the parameters C, d and γ. In short cross-validation is a method that works by using a subsample of the dataset to estimate a model and then use the remaining observations to test the model, this is done multiple times, to obtain a *test error rate*.

Cross validation can be applied in different ways. One of the more popular approaches are called *k-folds cross validation*. This resampling method randomly splits the dataset k times, and then use $k-1$ of the splits to estimate a model, and then test the model using the final split, this is repeated k times. When this is complete we get an estimate of how well the model, on average, predicted the correct outcome.

A special case of the *k-folds cross validation* resampling method is the *leave-one-out cross validation (LOOCV)*. With LOOCV $k = n$. One advantage with LOOCV is that the performance estimate is the same every time we run the algorithm, this will not be the case when $k < n$, since the (pseudo-)random nature of the algorithm could yield different results each time. Further, James et al. (2017), reports that there is a bias-variance tradeoff also when selecting k. LOOCV will yield almost *unbiased* estimates of the test error, while the variance of the LOOCV estimate will be higher since the samples are more similar than when $k < n$, and therefore also more correlated and higher variance. To balance this trade-off James et al. (2017) concludes that $k = 5$ or $k = 10$ will yield the best results. In this thesis I will use what seems like the norm within the Machine Learning community which is *10-folds cross validation*.

As stated above this thesis uses *10-folds cross validation* to in order to tune the parameters C, d and γ.[10] The way in which this is done is by estimating a model, $\hat{f}$, based on $k-1$ of the folds of the training part of the dataset, and measuring the *test error rate* bases on the final fold:(James et al. 2017, page 37):

$$\frac{1}{n}\sum_{i=1}^{n} I(y_t \neq \hat{y_t})$$

where $\hat{y_t}$ is prediction made in period t by using the estimated model, $\hat{f}$. y_i is the actual outcome of the binary variable y. $I(y_t \neq \hat{y_t})$ is an *indicator variable* that

[10] The γ parameter is discussed in the following section since it is a part of the radial kernel, which have yet to be discussed at this point.

equals one if the expression inside the parenthesis is true, that is, if the prediction made where wrong. This is repeated ten times, and an average of the test error rates is calculated. So, the *test error rate* is the share of observations that were misclassified. Naturally the goal is to minimize the test error rate, so the SVM chooses the relevant parameters that results in the lowest average test error rate across all of the k estimated models for each of the combinations of the relevant parameters.

5.2.4 Kernels

So far we have looked at the the basic idea behind the SVM, SVC in more detail and the cross validation technique for calibrating the algorithm. Now we will take the final leap to understanding the SVM. This final component is called *kernels*. Several kernel functions exist, but the task of the kernel is to choose the optimal *functional form* of the model based on the input data. In linear regression, a common problem is whether to include higher dimensional terms, the same is true for SVM; the decision boundry (i.e. the hyperplane) could be linear, quadratic, cubic or have some other polynomial degree, or some other form of non-linearity.

The machine learning jargon for the job of the kernal is that it *enlarges the feature space*. Which is nothing else than finding the best way of transforming the input data to achieve the highest predictive performance. However, what is called the *kernal trick* is a computationally efficient way of implementing this idea. The mathematical details are somewhat complex and I refer to Hastie et al. (2017, chapter 12.3) for a more through explanation.

As stated above, different kernel functions exist. Three of the most popular kernels are:

- Linear Kernel
- Polynomial Kernel of degree d
- Radial Kernel with a parameter γ

The linear kernel is equivalent to the SVC that was discussed above. This kernel does not enlarge the feature space into higher dimensions and therefore yields linear hyperplanes.

Polynomial kernels yields hyperplanes that are of some polynomial degree, d. The parameter d can be chosen using the cross validation technique described in section 5.2.3. That is, we chose the degree d which minimizes the test error rate.

Radial kernels can yield even more complex functional forms of the hyperplane in the original feature space. According to James et al. (2017) the feature space of the radial kernel is *implicit* and infinite-dimensional. This sounds strage, however, the clever way that the kernel function works allows us to work with infinite dimensional data without actually transforming our data into such dimensions. The parameter, γ, is a parameter that affects the influence of a single observation on the decision boundary. The greater the γ the "shorter" the influence of a single observation reach, the reverse is true for low gammas. The consequence of having a γ that is too high is that overfitting is very likely, despite using a low value for C. On the other hand, the consequences of using a low γ is that the model will not be able to capture the shape of the data (scikit learn 2019). As with the parameter d and C, a good value for γ can be found using cross validation.

5.2.5 Practical Implementation of SVM models

In this thesis, all of the SVM models are implemented using the `R` package `e1071`. The package was developed by a group at the Department of Statistics at Vienna University of Technology. The name of the package stems from the name that was assigned to this group by the university.

When estimating all SVM models in this thesis, optimal values of of the parameters C and γ are found using 10-fold cross validation. Below is an example of the `R` code that is used to to this.

```
tune.out=tune(svm, recession~., data=*estimation_dataset*, kernel="radial",
                 ranges=list(cost=c(1:100, 200, 300, 400, 500, 600, 700, 800, 900, 1000,
                                    1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900,
                                    2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000,
                                    6500, 7000, 7500, 8000, 8500, 9000, 9500, 10000),
                             gamma=c(0.01, 0.025, 0.05, 0.75, 0.1, 0.25, 0.5, 0.75,1,2)))
summary(tune.out)
```

Figure 5.4: *Sample code for how to find optimal values of the parameters required by the SVM model. These models can take quite a while to estimate as the number of combinations of the cost and gamma parameters is quite large.*

Section 5.2.4 discussed different types of kernels. In this thesis only the Radial Kernel will be used, as this kernel allows for many complex functional forms. Other kernels have been tested but seemed to perform worse.

5.3 Performance measures

The performance measures used to asses both the probit models and the SVM models and to compare the two, are inspired by the metrics used in the relevant literature, especially by Gogas et al. (2015) and Ratcliff (2013). In the empirical part of this thesis, where the performance of the different models are tested, the events of particular interest are recessions and output gaps. These events are referred to as *positive* events, not in a normative sense of the word, but in the sense that these events are of particular interest. Below follows the definition of the key metrics used to assess the performance of the models estimated in this thesis:

$$\text{Hit rate} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}$$

$$\text{False alarm rate} = \frac{\text{False positives}}{\text{False positives} + \text{True negatives}}$$

$$\text{Bias} = \frac{\text{True positives} + \text{False positives}}{\text{Number of recessions}}$$

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}$$

$$\text{Accuracy} = \frac{\text{True positives} + \text{True negatives}}{\text{True positives} + \text{False positives} + \text{True negatives} + \text{False negatives}}$$

The following list goes through the *intuition* behind each metric to give a more intuitive understanding of the definitions:

- **Hit rate:** The number of correctly predicted recessions, divided by the total number of recessions. This metric tells us how often we predicted recessions. If hit rate = 0.90, we correctly predicted 90 percent of recessions. Hit rate can also be referred to as *recall ratio*.
- **False alarm rate:** The number of incorrectly predicted recessions divided by the number of non-recessions. This metric tells us how often the model wrongly predicted a recession.
- **Bias:** Number of predicted recessions (correctly and incorrectly) divided by the actual number of recessions. This metric tells us if we overall predicted the same number of recessions as what actually occurred.
- **Precision:** The portion of predicted recessions that where correct. This metric can be particularly useful when the event we want to classify is relatively rare, such as a disease in a population or a recession in the economy.
- **Accuracy:** Refers to the total number of correctly predicted observations against the total number of predictions. Thus this metric focuses on the forecast accuracy of both positive and negative cases.

Chapter 6

Analysis and Results

In the analysis chapter of this thesis the relationship between the yield curve and economic downturns will be assessed in several different ways for the three Scandinavian countries. The thesis evaluates both the yield curves ability to predict future recessions and negative economic output gaps. This chapter is divided into four parts. The first two parts are about recession forecasting and the last two parts are about output gap forecasting. In both cases the first part (part 1 and 3), will assess statistical relationships using a variety of probit models. The second part (part 2 and 4) looks at the out-of-sample forecasting performance in both the recession case and the output gap case. In these two sections the out-of-sample forecasts will be made using non-probabilistic forecasts obtained from probit models, and forecasts made using SVM models.

In the sections on non probabilistic forecasts and SVM models, only 3, 6 and 12 months lag of the explanatory variable will be tested. There are three reasons for doing this. Reason number one is that these are the most common forecasting horizons in the literature. Reason number two is that testing all lag lengths between one month and eighteen months for all models would take up a lot of space while not adding much value to the thesis. The final reason is that the 10-folds cross validation method is computationally intensive, so estimating many more models would take a very long time.

6.1 Recession Forecasting - Probit significance testing

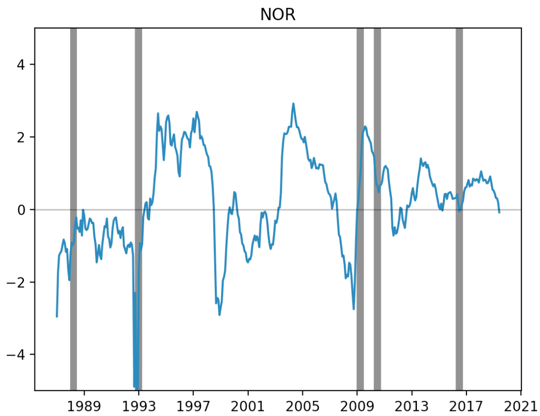
This section is in some ways a replication of the second part of Estrella and Hardouvelis (1991), but for the Scandinavian countries. This part of Estrella and Hardouvelis explores whether there is a statistically significant relationship between the yield spread and the recession indicator variable four quarters later. They do find such a relationship to be significant at the 5% level. Estrella and Hardouvelis (1991) and Estrella and Mishkin (1996) uses a static probit model to assess this relationship, this thesis does the same thing. Section 5.1 explains why the probit model is appropriate.

This section is structured in the following way. First, a simple illustration of the relationship between the slope of the yield curve and recessions are presented.

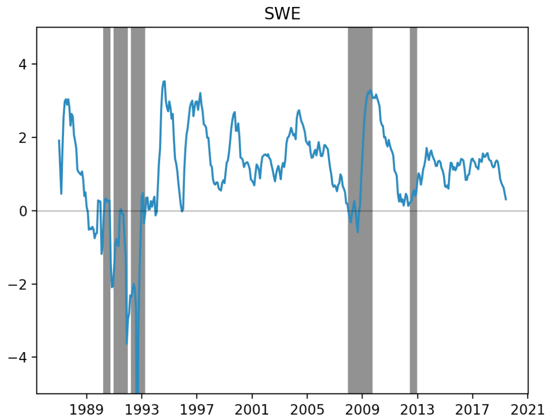
Second, I test the standard yield curve model with different lag lengths to find out which forecasting horizons give the strongest statistical relationship. Finally, I use the BIC statistic to find out whether it is worth it to include more than one lag of the spread variable in the models.

6.1.1 Yield Spread Model

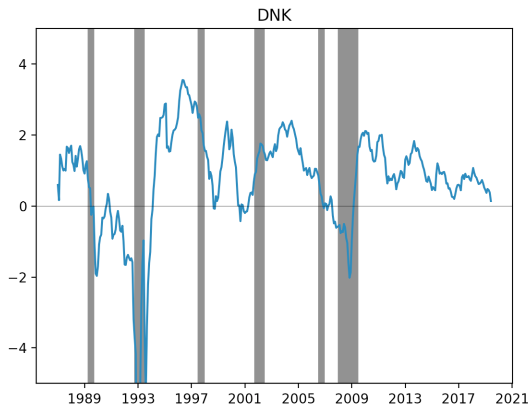
This section is in large a replication of the second part of Estrella and Hardouvelis (1991), where the relationship between the yield spread and future recessions are examined for the Scandinavian countries. Looking at figure 6.1 it seems like most recessions are preceded by a lower than usual level of the spread variable in all countries. However, exceptions do definitely exist and not nearly all periods with low or negative spread predicts a recession, and not all periods of recessions are preceded by low or negative spreads. This section examines the *statistical* relation between the two variables with different lag lengths and number of lags.



(a) Norway



(b) Sweden



(c) Denmark

Figure 6.1: *Graphs showing yield spread (10y - 3m) for all the Scandinavian countries, in blue. The shaded grey areas indicates periods of recession.*

Testing different lag lengths

At what lag lengths are the statistical relationship between the yield spread and recessions statistically significant, and at what lag length is the relation the strongest? Are there differences between the three countries? This sections attempts to answer these questions. To answer these questions the following regressions are estimated using the probit model with clustered standard errors at year-quarter level:

$$recession_t = \beta_0 + \beta_1 * spread_{t-\lambda} + \epsilon_t$$

where λ is the number of lags, and ϵ_t is an error term.

Figure B.1 in appendix B.1 shows quite poor results for Norway. Although all spread coefficients have negative signs, which is as expected when a downward sloping yield cruve is considered a bad sign, few coefficients are significantly different from zero. When testing lagged values of the spread, only lags of 17 and 18 months are statistically significant at 5% level, while lags of 2, 3 and 16 months are significant only at 10% level. This means that for most lag lengths it is not unlikely that the negative signs are just a random result of the sample, and not necessarily true for the population. Of course, one can never be certain that a regression coefficient is different from zero in the population, but in this case, for most lags, the probability of drawing a sample in which the coefficients are as negative as observed here is *not so* unlikely, even if the true coefficient is zero, that the null hypothesis is rejected.

Looking at figure B.2 in appendix B.1 the results are a lot more promising for Sweden, with statically significant spread coefficients at 1% level for all lags except 7, 8, 10 and 11 months. Coefficients for the spread with these lags are still significant at 5% level. It is also the case for Sweden that all coefficients are negative, as expected.

The results for the Danish data lies somewhere in between Norway and Sweden, with statistically significant coefficients for lags of the spread variable up til 4 months at 1% level, at 5% for 5 and 6 months and at 10% level for 7, 8 and 9 months.

Estrella and Hardouvelis (1991) found a statistically significant relationship between the 4th lag of spread (10y bond - 3month bill) at 5%. Since Estrella and Hardouvelis used data at quarter level, this result is equivalent to a one year lag of the spread. This means that their results only carries over in the Swedish case.

When it comes how well the models estimated explains the variance in the recession variable it seems, by looking at pseudo R^2 that, again, the results are a lot better for Sweden. For Sweden pseudo R^2 ranges between 0.1 and 0.374, with the highest pseduo R^2 being reach when a 6 month lag of the spread variable is used. The other two countries lies far behind as pseudo R^2 in the highest cases do not even reach 0.1.

Overall, the results seems to indicate that there is in fact a negative relationship between past values of the spread and the probability of a recession. And that the more negative the spread is, the more likely a recession is in the coming months. However, the strength of this relationship varies a lot depending on lag length and country, with Sweden showing the strongest relationship. In the case of Sweden and with some of the lagged spreads in the case of the other two countries (where the coefficients were significant at 5% level) one might say that the spread *Granger causes* recessions. It is important to note that this does not mean that certain values of the spread *cause* a recessions some months a head, it only means that the spread seems to be able to *predict* the probability of recessions. This difference is highly important.

Testing different number of lags

In the previous section only one lag of the spread was tested. This section tests whether including more lags, thus providing the models with more information (useful or not), is useful. The number of models that could, theoretically, be tested when there are 18 lags available and when the first lag to be included could be all of the lags is very large. Instead only models starting at lags 3 and 6 going up to 12 lags are tested, this reduces the number of models that needs to be estimated to 51.

BIC – Starting at 3 month lag

# of lags	Norway	Sweden	Denmark
3	**179**	237	**292**
4	184	225	298
5	190	**223**	303
6	196	**223**	309
7	201	227	314
8	207	232	318
9	213	235	325
10	219	238	330
11	225	242	336
12	230	247	342

Table 6.1: *Table showing BIC scores from estimating probit models with different lag lengths, starting at 3 months. Standard errors are clustered at year-quarter level. The minimum values are highlighted in bold font.*

Using the BIC criterion described in section 5.1.1 adding more lags seems for the most part to add more estimation error than valuable information. The exception is the the Swedish data, starting with a 3 month lag. In this case including the 4th, 5th and maybe 6th lag, proves valuable information according to this criterion. Results are shown in tables 6.1 and 6.2. These results are inline with a large portion of the literature as most authors only use one lag of the spread variable, although what lag is used varies (see for example Estrella and Hardouvelis (1991) and Ratcliff (2013)).

BIC – Starting at 6 month lag			
# of lags	**Norway**	**Sweden**	**Denmark**
6	**185**	**211**	**303**
7	189	213	307
8	195	218	313
9	200	220	319
10	207	222	324
11	212	226	329
12	218	231	335

Table 6.2: *Table showing BIC scores from estimating probit models with different lag lengths, starting at 6 months. Standard errors are clustered at year-quarter level. The minimum values are highlighted in bold font.*

6.2 Out-of-sample recession forecasting

This section focuses on out-of-sample recession forecasting using a variety of different models. First some of the probit models explored in the previous section will be used to create non-probabilistic forecasts which Ratcliff (2013) reports works better than using the same models to assign probabilities to such a rare event, which recessions are after all.[1] Second, another type of estimation technique called Support Vector Machine (SVM) is used to make the same out-of-sample forecasts as were made using the non-probabilistic model based on the forecasts from models similar to the probit models in the first part of this section (except estimated only on the `estimation` dataset). Finally all the models are compared to find out which gives the most accurate forecasts of recessions 3, 6 and 12 months ahead.

6.2.1 Non probabilistic forecasts

The non-probabilistic forecasts are obtained by first estimating the following equation using the `estimation` dataset, for each country, using a probit model with clustered standard errors at year-quarter level (equal to the model in the previous section (6.1)):

$$recession_t = \beta_0 + \beta_1 * spread_{t-\lambda} + \epsilon_t$$

where $\lambda \in \{3, 6, 12\}$. The coefficients from the resulting nine models are then used to make in-sample probabilistic predictions for $recession_t$. Then these probabilistic predictions are converted into binary (non probabilisitc) forcasts by using a threshold

[1] Read more about Ratcliff (2013) in the literature review in section 2.

W that takes the following values:

$$W = \frac{i}{200}$$

where $i \in [1, 200]$. This effectively means that the threshold, W, starts at 0.005 and goes up to 1 with increments of 0.005. Then the performance measures from section 5.3 is calculated. The estimated coefficients are shown in table 6.3.

	Norway			**Sweden**			**Denmark**		
Lag	**3**	**6**	**12**	**3**	**6**	**12**	**3**	**6**	**12**
`spread`	-0.191	-0.154	-0.086	-0.632	-0.876	-0.399	-0.224	-0.179	-0.110
`constant`	-1.502	-1.500	-1.500	-0.569	-0.506	-0.756	-0.870	-0.882	-0.903

Table 6.3: *Coefficients used in the non-probabilistic forecasts. Remember that these coefficients when multiplied with the independent variables needs to be converted into probabilities using the normal cumulative density function. These models are estimated on the* `estimation` *dataset.*

As stated in section 5.1.2 maximizing the difference between the hit rate and false alarm rate is how the threshold, W, is determined. Figure 6.2 shows in-sample performance for different values of W for the three Scandinavian countries at a 6 month horizon. Figures B.4 and B.5 displays the same charts for 3 and 12 month horizons and can be found in appendix B.2 on page 75. Table 6.4 shows in-sample results for all countries at the "optimal" W.

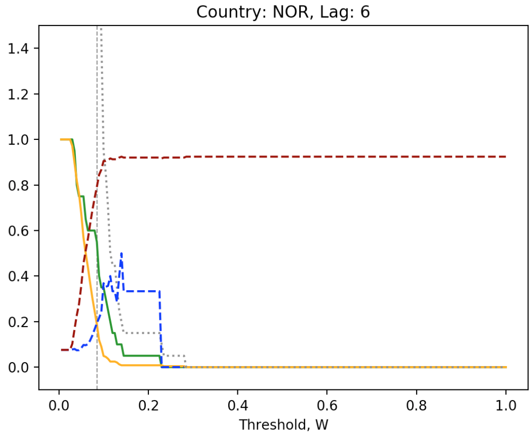
For all countries W is well below the traditional 50% threshold normally used for events such as elections.[2] For Norway, and to some extent Denmark, the difference between the hit rate and false alarm rate is quite low, although the numbers are a little higher for Denmark due to slightly higher hit rates, but also mostly higher false alarm rates.

The rest of the metrics further illustrate how bad this models works for Norway. The bias is very high, which means that the total number of recessions predicted is between 2.5 and 4.7 times higher than the actual number of recessions. Even with so many excess recession forecasts the model still only manage hit rates between 0.5 and 0.6. Which means that when this model predict a recession, the probability is very low that a recession actually will occur, this is reflected in the low precision scores. Overall accuracy is also very low for Norway.

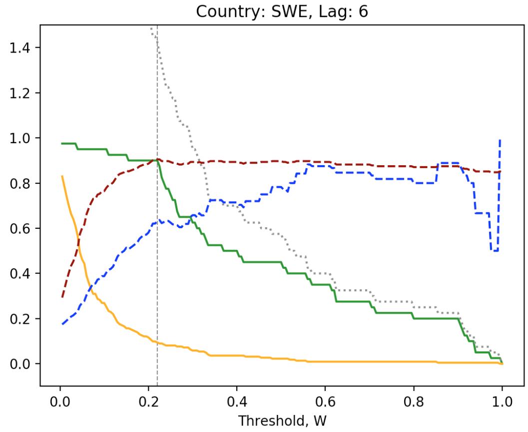
The metrics are overall quite good in the case of Sweden, compared to the other two countries, but the precision and bias scores are still not impressive as this is, after all, an in-sample analysis. The results for Denmark lies somewhere in between the results for the other two countries.

Figure 6.2 and the similar figures for 3 and 12 month horizon in appendix B.2 show that higher precision scores could be achieved in-sample if higher thresholds

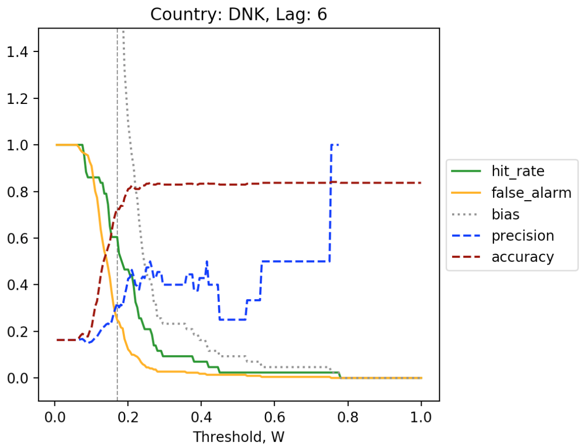
[2]If a candidate (in the US) has chance of winning of $p_{win} > 50\%$ and the candidate wins the election forecasts are normally considered to be "correct", even though the candidate would be expected to loose $1 - p_{win}$ of the time.



(a) Norway



(b) Sweden



(c) Denmark

Figure 6.2: *In-sample metrics for non-probabilistic recession forecasts at a 6 month horizon. The vertical grey dotted line indicates the W at which the difference between the hit rate and false alarm rate is the highest.*

	Norway			Sweden			Denmark		
Lag	**3**	**6**	**12**	**3**	**6**	**12**	**3**	**6**	**12**
W	0.090	0.085	0.070	0.200	0.220	0.150	0.220	0.170	0.175
Difference	0.332	0.362	0.260	0.724	0.806	0.703	0.357	0.356	0.437
Hit rate	0.500	0.550	0.600	0.850	0.900	0.900	0.465	0.605	0.651
False alarm	0.168	0.189	0.340	0.127	0.094	0.197	0.109	0.250	0.213
Bias	2.550	2.850	4.750	1.550	1.425	2.000	1.023	1.884	1.744
Precision	0.196	0.193	0.126	0.548	0.632	0.450	0.455	0.321	0.373
Accuracy	0.807	0.792	0.655	0.871	0.905	0.818	0.822	0.727	0.765

Table 6.4: *In-sample analysis of non-probabilistic recession forecasts: Table showing performance metrics for each country and lag for the values of W that maximize the difference between hit rate and false alarm rate.*

were chosen, this is also the case with the SVM models if the C parameter is increased. However, in contrast to the SVM model shown in figure 6.3 on page 46, in these non-probabilistic probit forecast models the other performance measures, except accuracy, goes down as W is increased. This means that increasing W does not lead to overfitting, which higher values of C does. Instead it mostly just leads to overall worse in-sample performance.

The in-sample analysis was not particularly impressive, the following paragraphs look at the out-of-sample performance of this non-probabilistic model. For all the three countries the out-of-sample metrics (see table 6.5) are quite similar to the in-sample metrics described above (and in table 6.4). On the bright side, this at least means that the previous model was not overfitted to the estimation dataset. However, it also means that the model does not work well in a Norwegian and, to some extent Danish, setting. Again, as were the case with the in-sample results, the models perform quite well for Sweden, especially when considering hit rates and false alarm rates. The precision metric is still quite low even for Sweden. And part of the reason the hit rate is as high as it is, could be due to the excessive number of recession forecasts issued by the model, shown by the bias score that ranges between 1.43 and 2.00.

	Norway			**Sweden**			**Denmark**		
Lag	**3**	**6**	**12**	**3**	**6**	**12**	**3**	**6**	**12**
W	0.090	0.085	0.070	0.200	0.220	0.150	0.220	0.170	0.175
Hit rate	0.500	0.600	0.600	0.824	0.647	1.000	0.455	0.546	0.636
False alarm	0.173	0.135	0.298	0.134	0.113	0.196	0.068	0.233	0.223
Bias	2.300	2.000	3.700	1.588	1.294	2.117	1.091	2.727	2.727
Precision	0.217	0.300	0.162	0.519	0.500	0.472	0.417	0.200	0.233
Accuracy	0.798	0.842	0.693	0.859	0.851	0.833	0.886	0.746	0.763

Table 6.5: *Out-of-sample analysis of non-probabilistic recession forecasts: Table showing performance metrics for each country and lag for the optimal values of W in the hold-out dataset. These tests are conducted using the* `Python` *programming language.*

6.2.2 Support Vector Machine Models

This section estimates models using the same variables as in section 6.2.1, but using the non-parametric SVM model from the field of machine learning to see if such a model can find a functional form that performs better at out-of-sample forecasting. This section follows the same order as the previous section.

Yield Spread SVM Model

The yield spread model is a very simple SVM model that only takes the yield spread as an independent variable. The results from estimating this model using the `estimation` dataset are shown in tables B.1, B.2 and B.3 in the appendix on page 75. For Norway and Denmark the results are underwhelming, with a hit rate of 0% independent of the lag length. The results are quite a lot better for Sweden with hit rates of 11%, 29% and 24% for three, six and twelve months, respectively. The precision metric could not be calculated for Norway and Denmark since no recessions were predicted at all, however, when estimating the SVM on the Swedish data a precision of 100%, 62.5% and 100% was achieved. This means that this model could potentially prove useful in a Swedish setting, since in the relatively rare instances, the model predicts a recession, the probability, based on this testing sample, is quite high that a recession will occur in the following months. Overall accuracy were high for all countries and lags with scores ranging between 86.8% and 91.2%, despite high numbers these are not particularly impressive. The reason such high overall accuracy can be achieved despite very low hit rate and precision is due to the fact that recession is a rare event, and relatively high accuracy can be achieved by simply always predicting the next month to be a non-recession month.

It seems that since recession is such a rare event, and the models are only given one explanatory variable, the optimal thing to do to get a low *error rate*, is simply to almost always just predict non-recessions.

Long and Short Interest Rate SVM Models

The models estimated in this section gives more flexibility to the SVM by not only providing the difference between long and short interest rates, but using both as input variables. When, in the previous section, the term spread was used as an explanatory variable, some information was lost. This section tests whether extracting more information from the yield curve by using the raw interest data as explanatory variables can improve the models.

Long and short interest rate SVM model – Norway

Lag	Hit rate	False alarm	Bias	Precision	Accuracy
3	60.00%	0.00%	0.60	100.00%	96.49%
6	10.00%	3.85%	0.50	20.00%	88.60%
12	20.00%	4.81%	0,70	28.57%	88.60%

Table 6.6: *Pseudo out-of-sample forecast using the SVM model on Norwegian data. The metrics shown here are calculated based on table B.4 in the appendix on page 77.*

Tables 6.6, 6.7 and 6.8 (see tables B.4, B.5 and B.4 in the appendix for details on the cost parameters) show that the cost parameter generally is higher for the models where interest where included separately, when compared to the spread models. Before discussing the results from this model, the next paragraph discusses the choice of C.

How the cost parameter is chosen using 10-folds cross validation was discussed in section 5.2. Figures 6.3[3] a) b) and c) show how the performance measures change

[3] All the models are estimated using `R`, the results are then written to a CSV file and graphs are created using the `Matplotlib` package in `Python`.

Long and short interest rate SVM model – Sweden

Lag	Hit rate	False alarm	Bias	Precision	Accuracy
3	64.71%	7.22%	1.06	61.11%	88.60%
6	52.94%	4.12%	0.76	69.23%	89.47%
12	88.24%	3.09%	1.06	83.33%	95.61%

Table 6.7: *Pseudo out-of-sample forecast using the SVM model on Swedish data. The metrics shown here are calculated based on table B.5 in the appendix on page 78.*

Long and short interest rate SVM model – Denmark

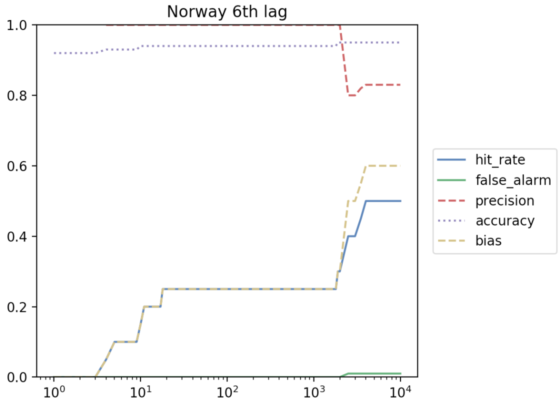
Lag	Hit rate	False alarm	Bias	Precision	Accuracy
3	72.73%	3.88%	1.09	66.67%	93.86%
6	81.82%	4.85%	1.27	64.29%	93.86%
12	72.73%	7.77%	1.45	50.00%	90.35%

Table 6.8: *Pseudo out-of-sample forecast using the SVM model on Danish data. The metrics shown here are calculated based on table B.6 in the appendix on page 78.*

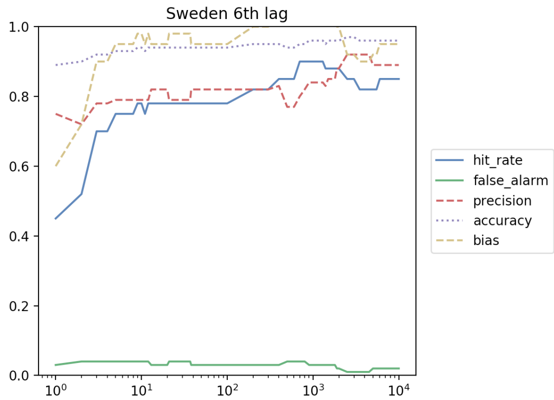
in-sample when the lag length is six[4], as the cost parameter is adjusted. Generally, when C is increased most performance measures increase. This is a result of the wider margin used by the SVM when C is high. The wider margin means more support vectors are used to estimate the hyperplane. A wider margin usually results in better in-sample performance, however, this is often due to overfitting. This means that a higher C is not always better, there is a trade-off between estimating a model that is well fit to the data, but that still manages to find *general patterns* in the observations used for estimation, that carries over to the testing dataset. Using the 10-folds cross validation method, the optimal C is found by estimating models on a subsection of the estimation dataset and and testing its performance on the rest of the estimation set, several times. Note that since the nature of the 10-folds cross validation builds on randomly splitting the dataset, each time such an algorithm is run the resulting "optimal" parameters might differ. The rest of this section discusses the results from these models.

Tables 6.6, 6.7 and 6.8 shows that this model that uses long and short interest rates independently drastically improves the performance of the models for all countries. This is especially true for Norway and Denmark whose results where terrible

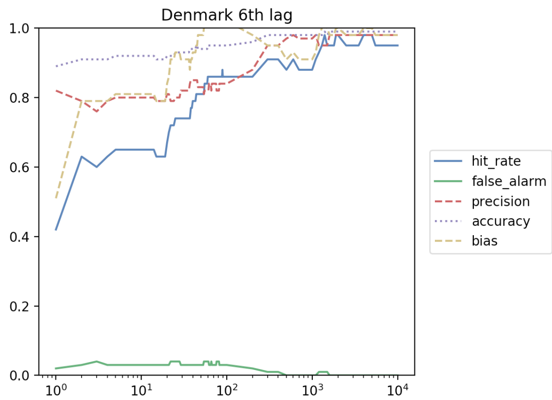
[4]A lag length of six was not chosen for any particular reason. However, I decided it would not be valuable to include graphs like these for all lag lengths, as that would just take up unnessesary space, without adding much value as the general pattern seen when increasing C is the same.



(a) Norway



(b) Sweden



(c) Denmark

Figure 6.3: *Graphs showing the in-sample performance of SVM when the cost parameter changes. The gamma parameter is constant at the optimal level for a six month lag length (see tables 6.6, 6.7 and 6.8). Obs: x-axis has a logarithmic scale. The optimal cost parameters chosen by the cross validation method is 6000, 3500 and 3500 for Norway, Sweden and Denmark, respectively.*

when the spread was the only explanatory variable.

For Sweden the precision metric has decreased for lags of three and six months, and false alarm rates have increased. However, the model now manages to predict more recessions, i.e. the hit rate is higher. The number of recessions predicted overall is also a lot closer to the actual number of recessions, with a bias metric around one.

The overall results, looking at all metrics together, show that the models works best out-of-sample for Sweden and Denmark, while the predictions for Norway seems useful only with a three month horizon. For Sweden and Denmark the results are quite promising, and show that the information provided by the yield curve could prove useful as a tool in recession forecasting. It can be speculated that, if the models were estimated on all the data available, not just 70% (an increase in the estimation set of 43%), the performance might be even better. Testing this hypothesis at this moment in time is, naturally, impossible.

6.2.3 Comparison

This section compares all the out-of-sample forecasts made in the section 6.2 to see which model works best.[5] First, I go through all results for each country individually to find the most effective models, and finally I compare the best results from each country to find out which country can benefit the most from studying models that build on information provided by the yield curve. The tables referred to in this section can be found in appendix B.2 on page 75.

I start by comparing the results of the out-of-sample recession forecasts for Norway, which are listed in table B.7. First of all it is clear that the non probabilistic models performs the most consistent across lags and input variables. The SVM models seems to only perform when the long and short interest rates are used as input variables. Actually, I would argue that, when looking at all metrics combined, the best performing model for Norway is the SVM model with long and short interest rate and a 3 month horizon. However, it is important to remember that these differences could be a characteristic of the sample and not the underlying data generating process.

The SVM models have overall higher accuracy, but this could be due to the fact the these models in general predict a low number of recessions (see the low bias scores), and since most periods are non-recession periods, a high accuracy score could be obtained by only predicting non-recessions. The non probabilistic models, on the other hand, have a relatively high bias, so the higher hit rates observed would also be expected. This also means that the SVM models tend to have lower false alarm rates.

Almost none of these models deliver forecasts with a decent level of precision, except the SVM model discussed above and the non probabilistic model with MSCI as input variable at a three month horizon.

Overall, neither of the models deliver particularly impressive results. However, the models that perform best is the SVM with long- and short interest rates at a three month horizon and the non probabilistic models with a 6 month horizon.

When it comes to the usefulness of the yield curve as a recession forecasting variable table B.7 shows that the spread delivers superior results when compared to models of the same lag length and estimation technique that take the returns of the stock indexes as input variables. This result suggests that the yield curve provide more information about medium to long term future economic conditions than lagged values of the monthly return of the MSCI stock index for Norway.

Table B.8 shows that overall, using the explanatory variables assessed in this thesis, recessions have been easier to forecast in the past for Sweden than for Norway. For all of the three forecasting horizons tested on Swedish data, the SVM models taking long and short interest rates as input variables perform best in pseudo out-of-sample tests, with the non probabilistic spread model coming in at second place. These two models have performed at at level where hit rates are high enough, false alarm rates are low enough and the precision is high enough to be useful in practice.

Again, all the models that take the variables from the yield curve as input outperform the models that take the MSCI variable. This means that also the Swedish

[5] See the attached Excel file for interactive graphs for the performance metrics for all the models tested in this thesis.

data suggest that the yield curve is a much better recession predictor than monthly lagged returns of a stock index.

Table B.9 shows that the overall trend seems to carry over to the Danish data was well, with the SVM models that take long and short interest rate as the input variable performing best across all forecasting horizons and the non probabilistic spread model performing second best. However, when estimated on the Danish data the above mentioned SVM model clearly performes best in this out-of-sample test. The best results by this model is achieved with a 3 and 6 month forecast horizon.

Comparing all models across all countries and forecasting horizons the SVM model with long and short interest rate as input variable at a 12 month horizon, estimated on the Swedish data achieves the best performance metrics when tested on the hold-out sample. However, this model works well for all horizons for Sweden and Denmark, but only gives acceptable results for Norway when the forecast horizon is 3 months.

Overall, these results suggest that the yield curve, when estimated with the right model, provides useful information to recession forecasting in Sweden and Denmark. It appears that there is some information about future recessions in Norway as well in the yield curve, however the models estimated in this thesis does not seem to be very useful in practice as hit rates are overall quite low and so is precision as well. Still, for all countries the yield curve (either in the form of the spread variable or as long and short interest rates individually) clearly outperformed lagged returns of the stock market index.[6]

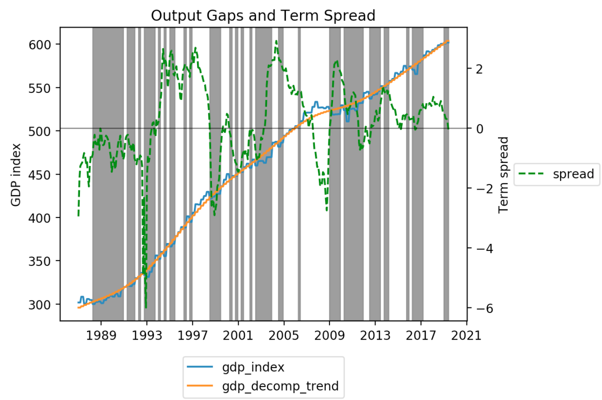
6.3 Output Gap Forecasting - Probit significance testing

This and the next section explores the yield curve's ability to forecast output gaps. Figure 6.4 shows the relation between the term spread and output gaps. This section estimates the same models as section 6.1 but with output gap as the explanatory variable.

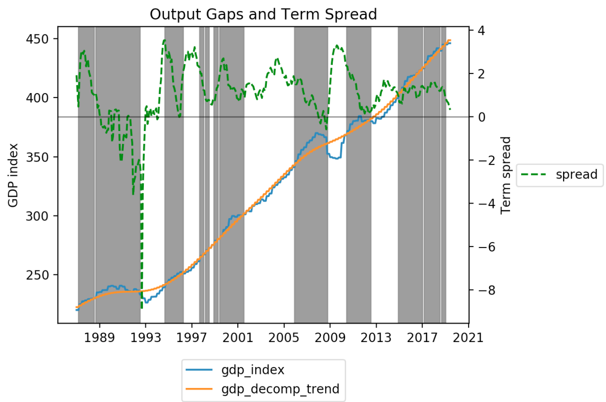
6.3.1 Yield Spread Model

Looking at figure 6.4 there are some periods in which the slope of the yield curve seems to be flatter in periods of negative output gaps, however, the relation seems much weaker than the relation between the yield curve and recessions (see figure 6.1 on page 37). The following two sections examines the statistical relationship between the yield curve and output gaps.

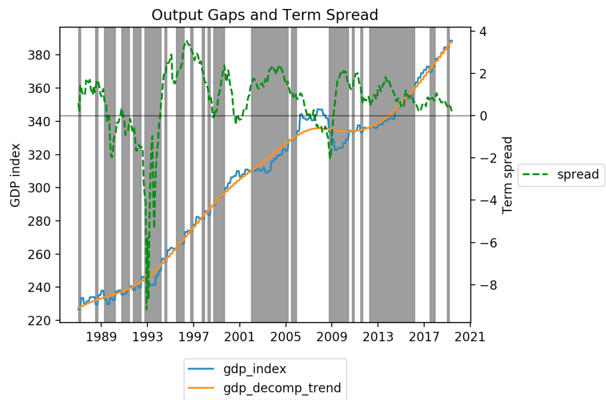
[6] This is true in all cases except maybe for the non probabilistic forecasts with a 3 month horizon for Norway, in this case the model outperforms the other models with a 6 month lag on all performance metrics except hit rate. However, this result does not change the overall concultion.



(a) Norway



(b) Sweden



(c) Denmark

Figure 6.4: *Graphs showing term spread and output gap throughout the dataset used for analysis. The shaded areas indicate negative output gaps.*

Testing different lag lengths

The following regression is estimated with clustered errors at year-quarter level in order to find out what lags of the term spread is better at predicting recessions.

$$outputgap_t = \beta_0 + \beta_1 * spread_{t-\lambda} + \epsilon_t$$

where λ is the number of lags and ϵ_t is an error term. Figures B.6, B.7 and B.8 in appendix B.3 on page 72-74 show the results discussed in the following paragraphs.

For Norway and Denmark the spread coefficient has a negative sign for all lags. This is consistent with the hypothesis that a flatter yield curve is associated with economic downturns in the near future. In the Norwegian setting all lags of spread are significantly different from zero at the 5% level, except the first lag which is only significant at 10% level.

For Sweden and Denmark the spread coefficient is significant (at 10%, 5% and and 1% level) for only a very few number of lag lengths. The results are the least promising for Sweden as some coefficients are even positive which is in conflict with the hypothesis discussed in the previous paragraph.

Overall, the measure of fit metric, pseudo R^2, is a lot lower when output gap is the explanatory variable, compared to when recession is. This is to be expected as the variance of the output gap variable is a lot higher than the variance of the recession variable for all the countries.[7]

Testing different number of lags

Although the statistical relationship between the yield spread and output gaps for Sweden and Denmark where questionable at best, this section will examine whether including more lags of the spread variable can improve the models estimated above.

Tables B.10 and B.11 in appendix B.3 on page 86 show the results from calculating BIC scores for multiple number of lags. However, the minimum values of the BIC score is always obtained when only one lag is included in the model, which suggest that the extra information contained in more lags of the variable does not outweigh the drawback of more estimation error that estimating more coefficients cause. So, similar to the recession BIC analysis, this thesis does not find evidence to support estimating models with more than one lag. However, the results presented here does not rule out that more than one lag of the spread variable could be appropriate if the model's first lag was different from 3 or 6.

6.4 Out-of-sample Output Gap Forecasting

The previous sections (1 and 2) assessed the yield curves ability to predict recessions, this section explores the yield curves ability to predict negative output gaps, out-of-sample. An important distinction between the two dependent variables is that while recession was a rare event, positive and negative output gaps happens roughly the same number of times over a longer period. This means that the requirements for good predictions are raised considerably, since merely flipping a coin would result in expected hit rates and false alarm rates around 50%, expected bias of 1 and expected precision and accuracy also around 50%.[8] To be precise, the precision metric is a

[7]The values of the regression constants are not particularly interest in this setting, since the point of this section is to measure the relationship between the spread variable and output gaps. For that particular reason they are not discussed here.

[8]Figure B.9 on page 89 in appendix B.4 shows results of simulating 10 000 random forecasts (the random forecast variable is uniformly distributed and can only take the two values 0 and 1) of the `outputgap` variable for Norway. These results support the statement about expected outcomes

little more affected by the relative number of negative output gaps in the sample. Sweden have a slightly lower share of negative output gaps (see table 4.4) and thus a slightly lower expected precision score, the reverse is true for Denmark. Looking at table 4.4 and comparing it to the simulation results, it appears that the expected value of the precision metric is very close or equal to the share of negative output gaps in the sample.

6.4.1 Non probabilistic forecasts

The analysis in this section is very similar to the one in section 6.2.1 except that the dependent variable is now the `output gap` variable. This means that for this analysis non-probabilistic forecasts are obtained by estimating the following regression for each country:

$$outputgap_t = \beta_0 + \beta_1 * spread_{t-\lambda} + \epsilon_t$$

This regression was estimated using a probit model with clustered errors at year-quarter level. How the predictions from the models are converted into binary forecasts is explained in section 6.2.1. Table 6.9 shows the coefficients used to create the binary forecasts. From section 6.3 it is already clear that when estimating these

	Norway			**Sweden**			**Denmark**		
Lag	**3**	**6**	**12**	**3**	**6**	**12**	**3**	**6**	**12**
`spread`	-0.195	-0.223	-0.119	0.155	-0.041	-0.153	-0.109	-0.115	-0.139
`constant`	0.164	0.170	0.134	-0.189	0.025	0.1478	0.249	0.259	0.285

Table 6.9: *Coefficients used in the non-probabilistic output gap forecasts. Remember that these coefficients when multiplied with the independent variable needs to be converted into probabilities using the normal cumulative density function. These models are estimated using the* ***`estimation`*** *dataset.*

coefficients on the entire dataset, most of these relationships are not significantly different from zero. This is not a promising start for the non probabilistic forecasts.

Looking at the in-sample analysis charts (see figures B.10, B.11 and B.12 in appendix B.4) the threshold parameter, W, is much closer to the commonly used 50%. This is likely due to the fact that, in contrast to recessions, negative output gaps happens more or less 50% of the time, so the probit model gets more observations of the event of interest. Further, the figures show a clear trade-off between either high hit rates with the following high false alarm rate and high bias or, low false alarm rates but with the negative side effect of low hit rates and bias way below one.

The charts also show that high precision can be achieved in most cases however, at the cost of very low hit rates and biases way below one, which means that the model would only issue a very small number of negative output gap forecasts and thus not be very useful even if most of these negative output gap predictions turned out be be correct.

by flipping a coin to make these forecasts. The simulations were completed using `Python`.

Using the maximal difference between the hit rate and the false alarm rate gives thresholds, Ws, that seems to balance the trade-offs discussed above quite well in most cases.[9] Next, the "optimal" values of W is used to make pseudo out-of-sample predictions. The results of the pseudo out-of-sample test is shown in table 6.10.

	Norway			**Sweden**			**Denmark**		
Lag	**3**	**6**	**12**	**3**	**6**	**12**	**3**	**6**	**12**
W	0.515	0.545	0.550	0.500	0.495	0.530	0.635	0.615	0.620
Hit rate	0.746	0.673	0.655	0.604	0.375	0.313	0.092	0.154	0.246
False alarm	0.458	0.305	0.322	0.455	0.349	0.182	0.061	0.184	0.102
Bias	1.236	1.000	1.000	1.229	0.854	0.563	0.139	0.292	0.323
Precision	0.603	0.673	0.655	0.492	0.439	0.556	0.667	0.526	0.762
Accuracy	0.640	0.684	0.667	0.570	0.535	0.605	0.456	0.439	0.526

Table 6.10: *Out-of-sample analysis of non-probabilistic negative output gap forecasts. This table shows performance metrics for each country and lag for the optimal values of W in the hold-out dataset. These out-of-sample test are done using the* `Python` *programming language.*

When interpreting these results I believe it is important to bear in mind the results of the random output gap forecasts (hereby referred to as ROF) of the `output gap` variable shown in figure B.9, as the standards for what are "good" values on these metrics are drastically different from the recession analysis in previous sections. In the following few paragraphs I discuss the results in table 6.10.

First, the hit rate varies dramatically between countries and lags. The hit rates are particularly low for Denmark, which probably is cause by the relatively high Ws. For Norway the hit rates are well above what is expected relative to ROFs. While in the Swedish case the hit rates are a little better than ROF when the forecast horizon is 3 months, otherwise the forecasts are very poor with respect to hit rates.

Second, the false alarm rates seems to follow the hit rates. When hit rates are high false alarm rates are also high, and vice versa. This is due to the trade-off discussed above, where a high W can give low false alarm rates and low hit rates, while the opposite is also true. Increasing or decreasing W can have the effect of skewing the forecasts towards either binary value that the `outputgap` variable can take. This skewness can be seen by studying the bias metric discussed in the following paragraph.

Third, the bias metric, shows the best results for Norway. Clearly, the number of predicted negative output gaps are quite close to the actual true number, which means that the relatively high hit rates are not caused by simply issuing mostly negative output gap forecasts, but by actually modelling the data quite well. For Sweden and Denmark, however, the bias is way below one (except for the third lag in Sweden) which means that the models predict too few negative output gaps

[9]One clear exception is in the case of Denmark with a 3 month horizon (see figure B.10 (c)) where the maxmial difference between the hit rate and false alarm rate happens for a relatively high W, which results in a model with high precision and low false alarm rate, but with the negative side effects of very few negative output gap forecasts relative to the total number of negative output gaps (see the low values for the bias metric) and low hit rates.

overall. Especially for Denmark, this is reflected in the low hit rates and low false alarm rates.

Fourth, the precision scores indicate that the forecasts for Sweden are just as good as ROFs. While for the Norwegian and Danish data the forecasts are clearly more precise than ROFs. This result is not particularly impressive for Denmark, since the threshold for issuing a negative output gap is relatively high, so the probabilities of a negative output gap outputted by the underlying probit model is quite high, and so the precision is also high, with the negative side effect of missing most negative output gaps (this is reflected in the low hit rates). Norway achieves precision scores that range between 0.6 and 0.67. This combined with bias near one and relatively good hit rates is a good sign that the model is quite useful in practice.

Finally, the accuracy, which measures the overall performance of the forecasts are not impressive, even in the Norwegian case where the other metrics showed quite good results. For Denmark, and to some extent Sweden the accuracy is close to the accuracy expected by ROFs.

Overall, this particular model clearly performed best when applied to the Norwegian data. In this case it gave somewhat useful results, however, the false alarm rate is still very high, while the hit rates are quite low even for Norway. In the case of the other two countries the model does not seem to provide any useful insight at all.

6.4.2 Support Vector Machine Models

This section tests the same models that were tested in section 6.2.2 but with `outputgap` as the dependent variable.

Yield Spread SVM Model

The section on out-of-sample output gap forecasting using support vector machines starts by examining models that take the yield spread as the explanatory variable.

Tables 6.11, 6.12 and 6.13 shows the performance metrics that results from es-

Term spread - Output Gap SVM model – Norway

Lag	Hit rate	False alarm	Bias	Precision	Accuracy
3	78.18%	54.24%	1.36	57.33%	61.40%
6	89.09%	69.49%	1.64	54.44%	58.77%
12	65.45%	40.68%	1.09	60.00%	62.28%

Table 6.11: *Pseudo out-of-sample forecast of output gaps in Norway using the SVM model, with the spread as the only explanatory variable. Calculations are based on table B.12 in appendix B.4.*

timating SVM models with `spread` as the explanatory variable. The results are particularly bad in the case of Sweden. With accuracy scores barely better than

Term spread - Output Gap SVM model – Sweden

Lag	Hit rate	False alarm	Bias	Precision	Accuracy
3	43.75%	40.91%	1.00	43.75%	52.63%
6	39.58%	33.33%	0.85	46.34%	55.26%
12	45.83%	25.76%	0.81	56.41%	62.28%

Table 6.12: *Pseudo out-of-sample forecast of output gaps in Sweden using the SVM model, with the spread as the only explanatory variable. Calculations are based on table B.13 in appendix B.4.*

Term spread - Output Gap SVM model – Denmark

Lag	Hit rate	False alarm	Bias	Precision	Accuracy
3	81.54%	55.10%	1.23	66.25%	65.79%
6	89.23%	77.55%	1.48	60.42%	60.53%
12	66.15%	46.94%	1.02	65.15%	60.53%

Table 6.13: *Pseudo out-of-sample forecast of output gaps in Denmark using the SVM model, with the spread as the only explanatory variable. Calculations are based on table B.14 in appendix B.4.*

than what one would expect from forecasts made by flipping a coin (equal to the ROF discussed earlier). The same goes for the precision score for lags 3 and 6 months, and a slight improvement above random forecasts is seen when the lag is 12 months. Hit rates for Sweden are well below the random benchmark forecasts, ranging between 39.6% and 45.9%. On the bright side, false alarm rates are a little lower than what is expected from random forecasts.

At first glance, the results seems more promising for Norway and Denmark, with hit rates well above 50% for all lags, and especially high hit rates for 3 and 6 month horizons. Unfortunately, the high hit rates are probably cased by to some extent biasing forecasts towards negative output gap forecasts – this is seen from the higher false alarm rates that follows the high hit rates. This suspicion is to a large extent confirmed by studying the bias metric, as it is well above the "optimal" value of one in the cases where hit rates are also high. This is quite clear evidence that the forecasts are skewed towards mostly predicting negative output gaps. With precision scores only around 60% and the high false alarm rates, these forecasts are not worth much in practice as they are only marginally better than simply guessing.

Long and Short Interest Rate SVM Models

In section 6.2.2 it was shown that out-of-sample forecasts were drastically improved when including the long and short interest pairs separately, rather than just using the difference between the two. This section explores whether that is also the case when `output gap` is the dependent variable. Results are shown in tables 6.14, 6.15 and 6.16.

Long and short interest rate SVM model – Norway

Lag	Hit rate	False alarm	Bias	Precision	Accuracy
3	72.73%	33.90%	1.09	66.67%	69.30%
6	81.82%	42.37%	1.27	64.29%	69.30%
12	78.18%	22.03%	1.02	76.79%	78.07%

Table 6.14: *Pseudo out-of-sample forecast using the SVM model to predict output gaps in Norway. The metrics shown here are calculated based on table B.15 in the appendix on page 97.*

Long and short interest rate SVM model – Sweden

Lag	Hit rate	False alarm	Bias	Precision	Accuracy
3	75.00%	18.18%	1.00	75.00%	86.67%
6	70.83%	13.64%	0.90	79.07%	74.56%
12	77.08%	24.24%	1.10	69.81%	80.70%

Table 6.15: *Pseudo out-of-sample forecast using the SVM model to predict output gaps in Sweden. The metrics shown here are calculated based on table B.16 in the appendix on page 97.*

To illustrate how the SVM models work, figures B.13, B.14 and B.15 show what the decision boundaries look like. The same charts could also be made for the long and short interest models in section 6.2.2.[10] The charts show that the SVM model is able to create very complex decision boundaries. This is one of the advantages of using SVM models with *k-folds cross validation*; the researcher does not have to assume a functional form in advance. For most countries and lags, it seems that the algorithm generally classifies observations where the short rate is higher than the

[10]These figures could be made for all non-probabilistic binary classification models that take two variables as input (since the chart is two dimensional). However, the chart is only included here as the number of figures and charts in this thesis is quite high, adding similar charts to the recession analysis would not add much value, in my opinion.

Long and short interest rate SVM model – Denmark					
Lag	**Hit rate**	**False alarm**	**Bias**	**Precision**	**Accuracy**
3	80.00%	16.33%	0.92	86.67%	81.58%
6	75.38%	26.53%	0.95	79.03%	74.56%
12	87.69%	28.57%	1.09	80.28%	80.70%

Table 6.16: *Pseudo out-of-sample forecast using the SVM model to predict output gaps in Denmark. The metrics shown here are calculated based on table B.17 in the appendix on page 97.*

long rate as negative output gaps, while it classifies observations in which the long rate is a little higher than the short rate and when they are about equal, as positive output gaps. Most of the models also seem to classify events in which the long rate is much higher than the short rate as negative output gaps. The first two attributes is pretty much inline with the literature and popular intuition; that a negatively sloped or flat yield curve indicate negative output gaps. The third attribute is in strong contrast to the literature and popular intuition and is probably cased by not having data for such observations in the `estimation` sample.

Looking at the results of the out-of-sample forecasts in tables 6.14, 6.15 and 6.16, again, it seems to be true that the SVM models work better when estimated using long and short interest rates as input variables rather than just the difference between the two. Most performance metrics are more stable across countries and lags and in most cases also better. The improvements are particularly clear in the case of Sweden with drastically lower false alarm rates and higher hit rates. The biases in each model is also generally closer to one. Precision and accuracy scores are also higher across the line.

This time it seems like the models perform best for Sweden and Denmark, as almost all performance measures are better than for Norway. This seems to be a pattern throughout all of the analysis conducted in this thesis. The next section directly compares the results from the different models and attempts to find a model that works best for each country.

6.4.3 Comparison

This section compares the out-of-sample results from the models tested in section 6.4. The section follows the same outline as comparison section 6.2.3 with first some comments on which models work best within each country and then I compare which country can benefit the most from using the yield curve as an output gap forecasting tool. When comparing results from different models it is still important to remember the benchmark ROF shown in figure B.9.

Table B.18 shows performance metrics for all models estimated on the Norwegian dataset.[11] First of all, it is clear that models that take lagged stock returns as input

[11] Remember that the attached Excel file provides graphs of these results. The graphs might

variables perform way worse than the ROFs, except for the probit model at a 12 month horizon which perform only marginally better than random. The rest of the probit models with lagged stock returns as the input variable either drastically underestimate the number of negative output gaps (3 month model) or only predict negative output gaps (6 month model). The SVM models with lagged stock returns as input variable simply always predict a negative output gap, which means that hit rates and false alarm rates are 100%.

The SVM model that take the long and short interest rates as input variables and the non probabilistic model that take the spread as input variable overall achieve performance metrics that are better than what is expected from ROFs (in the Norwegian setting). The SVM model marginally outperforms the non probabilistic model at all horizons. At what horizon these models perform best depends on which metrics are considered most important, however, the absolute best model seems to be the SVM model with long and short interest rate at a 12 month horizon (however, it could simply be due to random characteristics of this sample that makes this model marginally better than the other models discussed in this paragraph). The SVM model taking the spread as the input variable performs somewhere between the stock models and the models discussed in this paragraph.

Of the models estimated on the Swedish data (see table B.19), only the SVM model that takes long and short rates as input variables perform better than ROFs, this result holds across all forecasting horizons. Despite all the other models returning more or less useless results the SVM mentioned above performs well with average hit rate of 74.30%, average false alarm rate of 18.69%, average bias of 1.00, average precision of 74.60% and average accuracy of 78.36% across all three forecasting horizons. The deviations from these averages are also quite low.

As the previous paragraph suggests, the models that use lagged stock index returns does not give good forecasts for future negative output gaps in Sweden. This, again, goes to show that the information provided by the yield curve is more useful in forecasting negative output gaps.

Table B.20 shows that the results for the stock return models are just as bad for Denmark as it is for the other two countries. The forecasts range from being about as good as ROFs to much worse, this result hold across lags and estimation algorithms. Looking at the results from the rest of the models it is apparent that only the SVM with the interest rate pair as input variables perform well across lags. This model achieves average hit rates of 74.3%, average false alarm rates of 18.69%, average bias of 1.00, average precision of 74.60% and average accuracy of 78.36%.

Some general patterns emerge throughout this analysis. First, all models that take lagged stock returns as input variables performs poorly compared to both the best models that take yield curve variables as input but also compared to the expected ROFs. Second, SVM models that take the yield curve interest rate pair as input variables seems to consistently outperform other models, although the non probabilistic models that take the spread as input variable also perform nearly as good in most pseudo out-of-sample tests. These results show that out-of-sample negative output gap forecasting could be improved by using estimation techniques (SVM models) that are relatively new to the field of economics. Finally, no particular country stand out when it comes to the results discussed in this section. Using

make comparisons easier to understand as the tables can be a bit overwhelming.

the best performing models from estimated in this thesis, some valuable information could be extracted from these models when it comes to the future state of the economy. However, even if the models have proven to perform in random pseudo out-of-sample tests in the past, that is no guarantee that it will continue to do so in the future.

Chapter 7

Discussion

7.1 Discussion On Results and Their Usefulness

This thesis has estimated several different models trying to predict future recessions. In section 6.2.3 I found that the best models (usually the SVM model with long and short interest rate) worked quite well for predicting recessions in Sweden and Denmark, but not so well for Norway. The weaker relationship between recessions and the yield spread in Norway was apparent already in section 6.1, so expectations for the out-of-sample models were not particularly high. One possible explanation to this finding is that the Norwegian economy is highly dependent on the oil industry. This in turn means that the Norwegian economy can experience major negative shocks when the price of oil decreases significantly. Figure 6.1 shows that Norway was the only country out of the three Scandinavian countries that went through a recession in 2016 which corresponds to roughly the same time as the oil crisis hit the economy. Recessions that happen in Norway due to low oil prices may not be predicted as well by the yield curve due to the unpredictable nature of oil prices and the large impact it has on the Norwegian economy, compared to other economic shocks (to which the Norwegian economy has been remarkably robust).

Out of the three forecasting horizons tested, none stand out as "better than others". Although results differ across models, horizons and countries, I find no general pattern that the performance of the best models varies with horizon. This might seem surprising at first but may be due to fact that recessions usually are caused by some unpredictable event (i.e. a sudden drop in the oil price, collapse of the housing market, stock market crash etc.). As discussed in section 3.1.1, the flattening yield curve might, more than anything else, indicate that the economy is in, or is headed towards, a period of low growth and the economy is more likely to go into a recession when it is hit by a shock while being in this low growth state. So, the fact that I do not find evidence that the performance of the models generally varies with longer/shorter forecasting horizons *might* be because the yield curve, more than anything else, is predicting low growth periods in which negative macroeconomic shocks are more likely to trigger a recession.

The second part of this thesis tested how well output gaps could be predicted by using the same explanatory variables as when predicting recessions. It is hard to say whether or not the models work better for recession forecasting than output gap forecasting due to the fact that recession is a rare event while negative output gap is a quite common event. Still, the fact that only the SVM models that take the interest rate pair as input variables perform decent suggests that the yield curve is better at predicting recessions than output gaps. The fact that output gaps might be harder to predict, using the yield curve, was indicated already from the regression results in section 6.3 were only the results from Norway showed a significant relationship at most lags. The reason output gaps are seemingly harder to predict could be due to the fact that in relatively calm periods GDP varies over and under the long-term trend (estimated using the HP-filter) more or less randomly. While the periods of low growth that makes recessions more likely are more predictable and grounded in term structure theory of reduced consumption and increased savings to obtain a stable consumption path (see chapter 3.1.1), and should therefore be easier to predict using the yield curve as the predictor.

This thesis has found that the yield curve can prove useful in forecasting economic downturns, and that the best models clearly outperform models that use lagged stock returns as the explanatory variable. However, one might ask how useful these forecasts are in practice? The following three paragraphs discuss this question from multiple perspectives.

First, Estrella and Hardouvelis (1991) ask whether the slope of the yield curve provide information about future economic development over the information that it provides about current monetary policy. This is an interesting question because if the only reason the yield curve provide information about future economic conditions are caused by the fact that current monetary policy affects the short rates of the yield curve and economic development in exact opposite ways, central banks would effectively have no use of the information provided by the yield curve as it simply reflects their current actions. To test this, they simply run the same probit regressions with `recession` as the dependent variable and `spread` as independent variable but add the `short term interest rate` as another independent variable. This is done in order to see if the spread coefficients remain statistically significant also after adding the short term interest rates. The short term interest rates are thus viewed as a kind of proxy for the monetary policy set by the central bank. Estrella and Hardouvelis (1991) found that the predictive power remained almost intact, which means that the yield curve continues to have predictive power even when controlling for the actions of the central bank. Since this question is quite vital in terms of how useful the results found in section 6 are to the central banks of the Scandinavian countries I have done a quick analysis similar to that of Estrella and Hardouvelis (1991), but leave it to future research to examine the question in more detail (such as doing the same regressions for all lag lengths and also with output gap as the dependent variable). Figure C.2 show that for Norway the spread coefficient still remain at least as significant[1] when including the short term interest rate in the model. The same holds true for Sweden as well, except when the forecasting horizon is 12 months, in which case the spread coefficient goes from being significant at the 1% level to the 10% level. For Denmark on the other hand the

[1]This is at first a surprising results as when one adds a variable that is correlated to a variable already in the regression one would expect the coefficient of the first variable to at least loose some explanatory power (which could be seen by decreasing p-values). This would also have been the result if I did not cluster at year-quarter level. However, due to this clustering of the standard errors this unintuitive result emerge.

spread becomes insignificant when including the short term rate separately.

The fact that the significance level of the spread coefficient mostly remain intact for Norway and Sweden suggest that the information about future economic downturns provided by the yield curve could be useful not only to investors but also to the central banks of the two countries. For Denmark, however, it is unclear whether the yield curve can provide information useful to the domestic central bank, as I do not find evidence that the yield curve Granger causes recessions when controlling for the level of the short term interest rate. One possible explanation to the observed difference between Denmark and the other two countries could be caused by the fact that Denmark does not have a sovereign monetary policy due to having their exchange rate pegged to the Euro.[2]

Estrella and Hardouvelis (1991) points out another potential issue. If central banks adopt the yield curve as an indicator for future economic downturns will the yield curve then keep its forecasting power? They argue that the yield curve can only continue to be a useful forecasting tool to central banks if the historical statistical relations “reflect “deep” parameters in the optimal plans of private agents” and “monetary policy is neutral with respect to real output“. For a more through discussion on this issue see Estrella and Hardouvelis (1991).

Although this thesis have found many instances in which the information provided by the yield curve can give quite good forecasts of economic downturns, the best models for forecasting economic downturns surely includes other variables as well. The aim of this thesis was not to find the best forecasting model there are, but rather to see if the information provided by the yield curve, if modeled in a good way, can be useful as a complementary indicator in addition to other models or useful as a component of a larger model containing more variables. I would also stress that forecasting any future economic variable is a very hard thing to do as the underlying factors that affect the economy change all the time, and their relative importance can change as well.[3] This all means that the forecasts made by any forecasting model, no matter how sophisticated, should always be complimented with expert judgement, and even then the future is always going to be uncertain.

7.2 Critic and Suggestions On Future Research

This section discusses subjects that can potentially be a source of error in the thesis and things that could potentially be done better/in more detail in future studies.

Chapter 6 found that, in general, the best SVM models tend to outperform the best non probabilistic models. Of course, this could be due to the fact that SVMs are better at modelling the complex relationship between the yield curve and recessions/negative output gaps. However, one potential weakness of the non-probabilistic models is the way the threshold, W, was chosen. W was chosen by maximizing the difference between the hit rate and false alarm rate, however multiple

[2]It is not possible to have a sovereign monetary policy and a fixed exchange rate while keeping free capital flow, this trilemma is usually illustrated by the triangle referred to as the impossible trinity.

[3]For example, Döpke et al. (2017) found that the relative importance of the yield curve and stock market indexes has increased over the years, and that the effect of the short term rate has declined, with respect to recession forecasting.

other methods could be used. Ratcliff (2013, page 313) discusses some other methods such as minimizing bias (choosing a threshold that result in bias as close to one as possible) or maximizing the Equitable Threat Score. I would also suggest that future research within this field could attempt to use the k-folds cross validation technique that the SVM models used, in order to find the optimal threshold, W. The k-folds cross validation method could be implemented with different targets as well, not only the error rate which was used for the SVM model in this thesis, i.e. some of the methods discussed above could be implemented into such a resampling method. This could, in principle be done with both the SVM model and the non-probabilistic probit model. However, as far as I am concerned, such specialized models would probably have to be implemented by coding a significant part of the algorithm as not many econometric packages provide this functionality. Testing more customized models and resampling techniques could result in better choice of both W and C (in each model respectively), which might ultimately result in better out-of-sample forecasts.

In this thesis only the SVM models were tested with the long and short interest rates as separate input variables. However, this could also be done with the non probabilistic probit models. However, for such models to be able to model anything near the complex functional forms that the SVMs could, interaction terms of these two variables would probably have to be considered, since the probability of a recession/negative output gap probably is dependent on more than just the level of each variable. Adding interaction terms would allow the probit models to also take into account what happens if one rate is high and the other is low, or both are low etc., at the same time. This is another suggestion of how future research could potentially improve the out-of-sample performance of the non-probabilistic probit models.

This thesis divided the datasets into estimation and testing datasets by randomly splitting the data into two groups. This was, as is discussed in section 4.6 done in order to be able to have some recession periods in both the testing sample and the estimation sample for all the countries. The alternative would have been to use the first, lets say, 70% of the dataset to estimate a model and then the last 30% to test it. However, due to the problems discussed in section 4.6, this was not chosen as the method for splitting the data. There is one potential problem with the method used in this thesis, in my view, when considering the validity of results obtained from the SVM models. As figures B.13, B.14 and B.15 shows, the SVM algorithm effectively finds some areas in two-dimensional space (in the case of the long/short-interest rate model) that classify observations as either recessions/negative output gaps or as non-recessions/positive output gaps. And as figure C.1 shows, the long- and short interest rate time series are not stationary but appear to have negative trend throughout the sample. This means that if the SVM had only been estimated using, lets say, the first half of the dataset, almost all observations would be located in the upper right corner of the figures B.13-B.15, and so the SVM model might perform very poorly when classifying observations in the second half of the dataset as all observations have moved to the lower left of these diagrams (i.e. the interest rates tend to decrease with time). But since the datasets are split at random, chances are some observations from a large portion of recessions are included both in the estimation dataset and the testing dataset which means that the model could potentially, in some cases, have been estimated on the same recessions periods as it was tested on. Although this might not be a big issue, it is a drawback of the random sampling method used to split the datasets, and it could lead to better out-of-sample forecasts than what would be obtained in real-time forecasts in practice. The spread models are, however, to a greater extent resistant to this since the difference between these non-stationary time series seems to be stationary.[4] However, the problem of having potentially estimated the model on the same recession periods as the model is tested on, still persist.

The results in this thesis do to some extent provide evidence that better pseudo out-of-sample results could be obtained by using the SVM model from the field of Machine Learning.[5] However, as Gogas et al. (2015) reports, many previous studies have had success by implementing other machine learning methods when forecasting economic variables. One such algorithm is called Neural Networks and is an algorithm that attempts to mimic how the human brain learns in order to find general patterns in data and in order to make good predictions. I leave it to future research to explore whether such models can perform even better than the SVM models tested in this thesis.

In this thesis the yield curve was represented by 10-year government bonds and 3-month treasury bills. However, bonds with other maturities also exits, although the selection is a bit larger for US bonds than for Scandinavian bonds. Gogas et al. (2015) have departed from the standard in the literature, which is to only use a pair of interest rates, and had success in out-of-sample forecasting when using 3-month treasury bill rates, 2- and 3-year government bond rates on U.S. data. They argue that in a stable and developed economy market participants do not care much about short term fluctuations when considering long-term economic development, and therefore the long-term rates (such as 10-year government bonds) do not provide much information about future economic downturns. These two hypothesis could also be tested for the Scandinavian countries, that is, including more interest rates in the model to capture information about what is referred to as the arc of the yield curve, and to use bonds with less time to maturity than what was used in this thesis.

I would also suggest that future researchers within this sub-field of macroeconomic consider studying whether changes in the slope of the yield curve can help with predicting recessions. It could be that instead of the steepness of the slope, changes to the slope itself can indicate a coming recession due to investors suddenly moving large portions of their resources to long term bonds instead of short term bonds and stocks, thus lowering the yield on these bonds. Such an event does not necessary have to lead to a very flat yield curve, but could still change the slope considerably. For now, this is only speculations.

[4] I have not conducted a statistical test, such as the Dickey-Fuller test, to see if this series is stationary, however, looking at figure 6.1 the series seems to vary around a stable mean, however the variance of the series might be changing over time, so this is an argument against stationarity. But anyway, this series does not have the same problem of a negative trend which the long- and short interest rates have.

[5] The arguments against this viewpoint is discussed in the previous three paragraphs.

Chapter 8

Conclusion

This thesis has explored the insight that the yield curve can provide with respect to forecasting economic downturns. Specifically, it has examined the yield curve's ability to predict recessions and negative output gaps for the three Scandinavian countries. Both statistical tests and out-of-sample tests have been conducted in order to answer these questions.

In conducting the analysis in this thesis I have used a method called upsampling in order to take advantage of the fact that the interest rate time series can be obtained with a monthly frequency while the GDP data only exist at a quarterly frequency and an annual frequency. Applying this method to the dataset means that I effectively increase its resolution, and the resulting increased sample size is almost always desirable in statistical analysis. However, the process of effectively copying the GDP data three times for each observation leads to correlation within quarters, which in turn could lead to autocorrelated errors in the population regressions estimated in the analysis section of the thesis. To deal with this potential issue I also propose a method for dealing with the potential problem of auto correlated errors that might occur due to applying this resampling method. The method proposed and used is to cluster errors at year-quarter level to control for the correlation within quarters.

The statistical tests showed that the relationship between the spread and future recessions was strong for Sweden at forecasting horizons of one through eighteen months. For Denmark the relationship was only significant for forecasting horizons up until 6 months, and for Norway only the seventeenth and eighteenth lag of the spread variable were significantly different from zero at a 5% level. When looking at similar tests of the relationship between the yield spread and negative output gaps the results are almost opposite compared to the recession analysis, as the relationship is strong for Norway at all lags and only at very few lags for the other two countries. Further analysis also showed that including more lags in these models rarely is particularly useful.

The other main part of this thesis tested several models for forecasting recessions and output gaps out-of-sample using the yield curve. In chapter 7 I discussed some potential problems with the methods used and some ways in which the models could be made even better. However, the results showed clearly across all estimation techniques, countries and lags in both recession forecasting and negative output gap forecasting that these two types of economic downturns are better predicted by the information provided by the yield curve than by lagged stock returns. This result is consistent with a large part of the existing yield curve literature based on U.S. data.

I also find that the yield curve models yield more useful results for recession forecasting in Sweden and Denmark than it does for Norway. Further I speculate that this might be due to the fact that the Norwegian economy is highly dependent on oil prices. The results are, as were the case with the statistical tests, almost opposite in the negative output gap forecasting analysis. This means that the yield curve seems to provide useful information to a larger extent for Norway than for the other two countries in terms of forecasting negative output gaps and vice versa for recession forecasting.

In addition to contributing to the literature in terms of doing similar analysis for the Scandinavian countries that have already been done for the U.S. on economic downturn forecasting using the yield curve, I also test whether methods from the field of machine learning can contribute to improving forecasts.[1] I find that in almost all cases the Support Vector Machine model that take the long- and short-term interest rate pair as input variables outperform the non-probabilistic models based on the familiar probit model. This result suggests that advanced models that could be generated by machine learning algorithms can be very useful in forecasting economic downturns in the future.

[1]As far as I am concerned only Gogas et al. (2015) have done a similar analysis, but on U.S. data.

Bibliography

Ball, L., Leigh, D., & Loungani, P. (2017). Okun's law: Fit at 50? *Journal of Money, Credit and Banking*, *49*(7), 1413–1441.

Banton, C. (2020). *An introduction to u.s. stock market indexes*. Retrieved April 17, 2020, from https://www.investopedia.com/insights/introduction-to-stock-market-indices/

Brandl, B., Leopold-Wildburger, U., & Pickl, S. (2009). Increasing the fitness of fundamental exchange rate forecast models. *International Journal of Contemporary Mathematical Sciences*, *4*(16), 779–798.

Brealey, Myers, & Allen. (2017). *Principles of corporate finance* (12th ed.). McGraw-Hill.

Campbell, J. Y., & Shiller, R. J. (1987). Cointegration and tests of present value models. *Journal of Political Economy*, *95*(5), 1062–1088.

Dao, M., & Loungani, P. (2010). The human cost of recessions: Assessing it, reducing it. *IMF Staff Position Note*. http://ssrn.com/abstract=1711933

Döpke, J., Fritsche, U., & Pierdzioch, C. (2017). Predicting recessions with boosted regression trees. *International Journal of Forecasting*, *33*(4), 745–759.

Dueker, M. (1997). Strengthening the case for the yield curve as a predictor of US recessions. *Federal Reserve Bank of St. Louis Review*, *79*, 41–51.

Economist, T. (2019). Yield curves help predict economic growth across the rich world. *The Economist*. https://www.economist.com/graphic-detail/2019/07/27/yield-curves-help-predict-economic-growth-across-the-rich-world

Estrella, A., & Hardouvelis, G. A. (1991). The term structure as a predictor of real economic activity. *The Journal of Finance*, *46*(2), 555–576. https://doi.org/10.2307/2328836

Estrella, A., & Mishkin, F. S. (1996). The yield curve as a predictor of u.s. recessions. *Current Issues in Economics and Finance*, *2*(7).

Estrella, A., & Trubin, M. R. (2006). The yield curve as a leading indicator: Some practical issues. *Current Issues in Economics and Finance*, *12*(5).

Evgenidis, A., & Siriopoulos, C. (2014). Does the yield spread retain its forecasting ability during the 2007 recession? a comparative analysis. *Applied Economics Letters*, *21*(12), 817–822.

Fama, E. F. (1986). Term premiums and default premiums in money markets. *Journal of Financial Economics*, *17*(1), 175–196.

Gaski, J. F. (2012). On the competing definitions of recession. *Society*, *49*(2), 118–121.

Gavin, W. T. (2012). What is potential GDP and why does it matter? *Economic Research Federal Reserve Bank of St. Louis*. https://research.stlouisfed.org/

publications/economic-synopses/2012/04/20/what-is-potential-gdp-and-why-does-it-matter/
Gogas, P., Papadimitriou, T., & Chrysanthidou, E. (2015). Yield curve point triplets in recession forecasting. *International Finance*, *18*(2), 207–226. https://doi.org/https://doi.org/10.1111/infi.12067
Hagelund, K., Hansen, F., & Robstad, Ø. (2018). Model estimates of the output gap. *Norges Bank Staff Memo*, (4). https://static.norges-bank.no/contentassets/5ac97ee98d0443ab906c4af6a7578750/staff_memo_4_2018_en.pdf?v=05/07/2018125331&ft=.pdf
Hamilton, J. (2018). Why you should never use the hodrick-prescott filter. *Review of Economics and Statistics*, *100*(5), 831–843.
Harvey, C. R. (1986). *Recovering expectations of consumption growth from an equilibrium model of the term structure of interest rates* (Doctoral dissertation). University of Chicago. https://faculty.fuqua.duke.edu/~charvey/Research/Thesis/Thesis.htm
Harvey, C. R. (1988). The real term structure and consumption growth. *Journal of Financial Economics*, *22*(2), 305–333.
Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The elements of statistical learning* (12th). Springer.
Hjelm, G., & Jönsson, K. (2010). In search of a method for measuring the output gap of the swedish economy. *The National Institue of Economic Research (NIER)*, *115*. https://core.ac.uk/download/pdf/6370452.pdf
Hodrick, R. J., & Prescott, E. C. (1997). Postwar u.s. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, *29*(1), 1–16.
Holden, S. (2016). *Makroøkonomi* (1st ed.). Cappelen Damm.
Ince, H., & Trafalis, T. B. (2006). A hybrid model for exchange rate prediction. *Decision Support Systems*, *42*, 1054–1062.
Jahan, S., & Mahmud, A. S. (2013). What is the output gap? *Finance and Development*, *50*(3), 38–39. https://www.imf.org/external/pubs/ft/fandd/2013/09/pdf/basics.pdf
James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An introduction to statistical learning* (Vol. 8).
Jones, C. (2019). 2019's yield curve inversion means a recession could hit in 2020. *Forbes*. Retrieved May 13, 2020, from https://www.forbes.com/sites/chuckjones/2020/12/31/2019s-yield-curve-inversion-means-a-recession-could-hit-in-2020/#7c3f07d54229
Mercado, D. (2018). These 401(k) funds took a beating in 2008 - and it could happen again. *CNBC*. https://www.cnbc.com/2018/09/13/these-retirement-funds-took-a-beating-in-2008-it-could-happen-again.html
Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106.
NBER. (2008). Determination of the december 2007 peak in economic activity. Retrieved February 3, 2020, from https://www.nber.org/cycles/dec2008.html
Ozturk, H., & Pereira, L. F. V. (2014). Yield curve as a predictor of recessions: Evidence from panel data. *Emerging Markets Finance and Trade*, *49*, 194–212. https://doi.org/https://doi.org/10.2753/REE1540-496X4905S512

Poloni, F., & Sbrana, G. (2017). Multivariate trend-cycle extraction with the hodrick-prescott filter. *Macroeconomic Dynamics*, *21*(6), 1336–1360.

Predescu, O. M., & Stancu, S. (2011). Portfolio risk analysis using ARCH and GARCH models in the context of the financial crisis, *18*(2), 75–88.

Ratcliff, R. (2013). The 'probability of recession': Evaluating probabilistic and non-probabilistic forecasts from probit models of u.s. recessions. *Economic Letters*, *121*(2), 311–315.

Ravn, M. O., & Uhlig, H. (2002). On adjusting the hodrick-prescott filter for the frequency of observations. *Review of Economics and Statistics*, *84*(2), 371–376.

scikit learn. (2019). *RBF SVM parameters*. https://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html

Skipper, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modelling with python. *Proceeding of the 9th Python in Science Conference*.

Spewak, A., & Andolfatto. (2018). *Does the yield curve really forecast recession?* https://research.stlouisfed.org/publications/economic-synopses/2018/11/30/does-the-yield-curve-really-forecast-recession/

Stock, J. H., & Watson, M. W. (2015). *Introduction to econometrics* (Updated Third Edition - Global Edition). Person Education Limited.

Appendix A

Methodology Appendix

A.1 Testing for Autocorrelated Errors

A.1.1 Recession

	Norway	**Sweden**	**Denmark**
`L6.uhat`	0.81***	0.84***	0.85***
	(0.03)	(0.03)	(0.03)
`Constant`	0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)

Note: * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$. Standard errors are in parenthesis below the coefficients.

Table A.1: *Regression results of estimating:* $\hat{u}_t = \beta_0 + \beta_1 * \hat{u}_{t-1} + \epsilon_t$*, on the residuals obtained when estimating:* $recession_t = \beta_0 + \beta_1 * spread_{t-6} + \epsilon_t$*. Residuals are calculated as* $\hat{y}_t - y_t$ *where* $\hat{y}_t$ *is the estimated probability of a recession.*

A.1.2 Output gap

	Norway	**Sweden**	**Denmark**
`L6.uhat`	0.75***	0.87***	0.80***
	(0.03)	(0.03)	(0.03)
`Constant`	0.00	-0.00	-0.00
	(0.02)	(0.01)	(0.02)

Note: * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$. Standard errors are in parenthesis below the coefficients.

Table A.2: *Regression results of estimating:* $\hat{u}_t = \beta_0 + \beta_1 * \hat{u}_{t-1} + \epsilon_t$*, on the residuals obtained when estimating:* $outputgap_t = \beta_0 + \beta_1 * spread_{t-6} + \epsilon_t$*. Residuals are calculated as* $\hat{y}_t - y_t$ *where* $\hat{y}_t$ *is the estimated probability of negative output gap.*

Appendix B

Results Appendix

B.1 Yield curve and recessions - Statistical relationships

Testing different lags -- Norway

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession
recession																		
L.spread	-0.193																	
	(0.124)																	
L2.spread		-0.220^{*}																
		(0.125)																
L3.spread			-0.231^{*}															
			(0.131)															
L4.spread				-0.199														
				(0.135)														
L5.spread					-0.172													
					(0.134)													
L6.spread						-0.165												
						(0.134)												
L7.spread							-0.124											
							(0.135)											
L8.spread								-0.111										
								(0.137)										
L9.spread									-0.0962									
									(0.136)									
L10.spread										-0.0886								
										(0.136)								
L11.spread											-0.0852							
											(0.134)							
L12.spread												-0.0950						
												(0.129)						
L13.spread													-0.112					
													(0.124)					
L14.spread														-0.124				
														(0.114)				
L15.spread															-0.139			
															(0.100)			
L16.spread																-0.160^{*}		
																(0.0849)		
L17.spread																	-0.183^{**}	
																	(0.0776)	
L18.spread																		-0.192^{**}
																		(0.0821)
_cons	-1.505^{***}	-1.511^{***}	-1.515^{***}	-1.510^{***}	-1.506^{***}	-1.506^{***}	-1.503^{***}	-1.503^{***}	-1.504^{***}	-1.504^{***}	-1.505^{***}	-1.505^{***}	-1.506^{***}	-1.508^{***}	-1.510^{***}	-1.515^{***}	-1.522^{***}	-1.527^{***}
	(0.180)	(0.182)	(0.184)	(0.181)	(0.179)	(0.179)	(0.177)	(0.177)	(0.177)	(0.177)	(0.177)	(0.177)	(0.177)	(0.177)	(0.177)	(0.177)	(0.176)	(0.177)
N	372	372	372	372	372	372	372	372	372	372	372	372	372	372	372	372	372	372
pseudo R^2	0.041	0.053	0.060	0.044	0.032	0.030	0.016	0.013	0.010	0.008	0.008	0.009	0.013	0.015	0.019	0.024	0.031	0.035

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure B.1: *The table shows results of estimating:* $recession_t = \beta_0 + \beta_1 * spread_{t-\lambda} + \epsilon_t$*, where* $\lambda \in [1, 18]$ *for Norway. The standard errors are estimated by clustering at year-quarter level. The regression was estimated on data from 1988m7 - 2019m6, lags where calculated using data up to 18 months before 1988m7, this means that all regressions are estimated on the same number of observations,* N*.*

Testing different lags -- Sweden

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession
recession																		
L.spread	-0.486^{***}																	
	(0.165)																	
L2.spread		-0.581^{***}																
		(0.183)																
L3.spread			-0.679^{***}															
			(0.201)															
L4.spread				-0.817^{***}														
				(0.237)														
L5.spread					-0.887^{***}													
					(0.235)													
L6.spread						-0.874^{***}												
						(0.220)												
L7.spread							-0.555^{**}											
							(0.225)											
L8.spread								-0.538^{**}										
								(0.220)										
L9.spread									-0.529^{***}									
									(0.203)									
L10.spread										-0.525^{**}								
										(0.205)								
L11.spread											-0.518^{**}							
											(0.207)							
L12.spread												-0.486^{***}						
												(0.180)						
L13.spread													-0.448^{***}					
													(0.166)					
L14.spread														-0.407^{***}				
														(0.152)				
L15.spread															-0.378^{***}			
															(0.135)			
L16.spread																-0.352^{***}		
																(0.130)		
L17.spread																	-0.341^{***}	
																	(0.129)	
L18.spread																		-0.324^{***}
																		(0.119)
_cons	-0.655^{***}	-0.603^{***}	-0.551^{***}	-0.474^{***}	-0.450^{**}	-0.479^{**}	-0.667^{***}	-0.676^{***}	-0.681^{***}	-0.683^{***}	-0.684^{***}	-0.695^{***}	-0.708^{***}	-0.725^{***}	-0.738^{***}	-0.751^{***}	-0.755^{***}	-0.763^{***}
	(0.174)	(0.176)	(0.175)	(0.175)	(0.184)	(0.205)	(0.219)	(0.217)	(0.210)	(0.211)	(0.212)	(0.203)	(0.199)	(0.196)	(0.190)	(0.188)	(0.188)	(0.183)
N	370	370	370	370	370	370	370	370	370	370	370	370	370	370	370	370	370	370
pseudo R^2	0.183	0.238	0.291	0.346	0.371	0.374	0.266	0.251	0.244	0.242	0.238	0.213	0.184	0.154	0.133	0.117	0.110	0.100

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure B.2: *The table shows results of estimating:* $recession_t = \beta_0 + \beta_1 * spread_{t-\lambda} + \epsilon_t$*, where* $\lambda \in [1, 18]$ *for Sweden. The standard errors are estimated by clustering at year-quarter level. The regression was estimated on data from 1988m7 - 2019m6, lags where calculated using data up to 18 months before 1988m7, this means that all regressions are estimated on the same number of observations,* N*.*

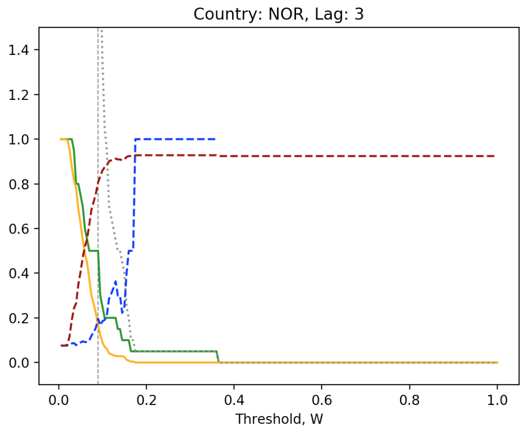
Testing different lags -- Denmark

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession	recession
recession																		
L.spread	-0.274***																	
	(0.0898)																	
L2.spread		-0.272***																
		(0.0910)																
L3.spread			-0.263***															
			(0.0922)															
L4.spread				-0.250***														
				(0.0898)														
L5.spread					-0.217**													
					(0.0901)													
L6.spread						-0.198**												
						(0.0911)												
L7.spread							-0.164*											
							(0.0902)											
L8.spread								-0.150*										
								(0.0843)										
L9.spread									-0.139*									
									(0.0833)									
L10.spread										-0.128								
										(0.0841)								
L11.spread											-0.117							
											(0.0810)							
L12.spread												-0.112						
												(0.0813)						
L13.spread													-0.109					
													(0.0829)					
L14.spread														-0.102				
														(0.0804)				
L15.spread															-0.0882			
															(0.0799)			
L16.spread																-0.0708		
																(0.0793)		
L17.spread																	-0.0549	
																	(0.0766)	
L18.spread																		-0.0364
																		(0.0751)
_cons	-0.929***	-0.928***	-0.931***	-0.936***	-0.951***	-0.957***	-0.970***	-0.975***	-0.979***	-0.983***	-0.987***	-0.990***	-0.991***	-0.994***	-1.001***	-1.010***	-1.020***	-1.032***
	(0.147)	(0.147)	(0.147)	(0.148)	(0.149)	(0.150)	(0.151)	(0.150)	(0.150)	(0.151)	(0.150)	(0.150)	(0.150)	(0.150)	(0.150)	(0.150)	(0.150)	(0.150)
N	372	372	372	372	372	372	372	372	372	372	372	372	372	372	372	372	372	372
pseudo R^2	0.097	0.097	0.092	0.084	0.067	0.056	0.039	0.032	0.027	0.023	0.019	0.017	0.016	0.014	0.010	0.006	0.004	0.002

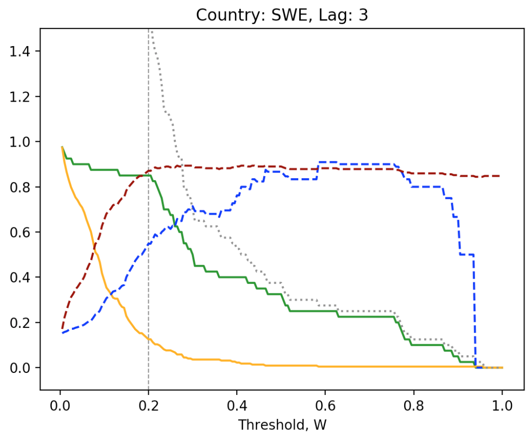
Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure B.3: *The table shows results of estimating:* $recession_t = \beta_0 + \beta_1 * spread_{t-\lambda} + \epsilon_t$*, where* $\lambda \in [1, 18]$ *for Denmark. The standard errors are estimated by clustering at year-quarter level. The regression was estimated on data from 1988m7 - 2019m6, lags where calculated using data up to 18 months before 1988m7, this means that all regressions are estimated on the same number of observations,* N*.*

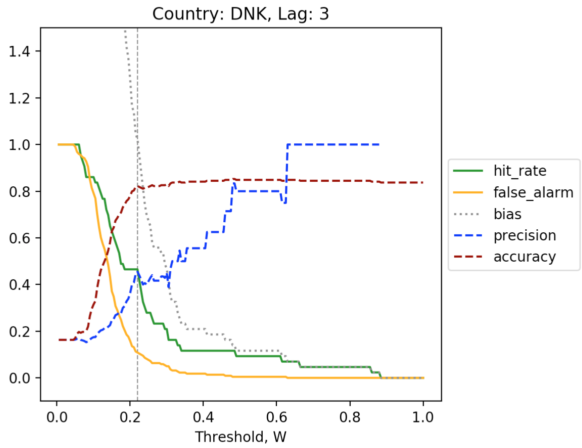
B.2 Out-of-sample recession forecasting



(a) Norway

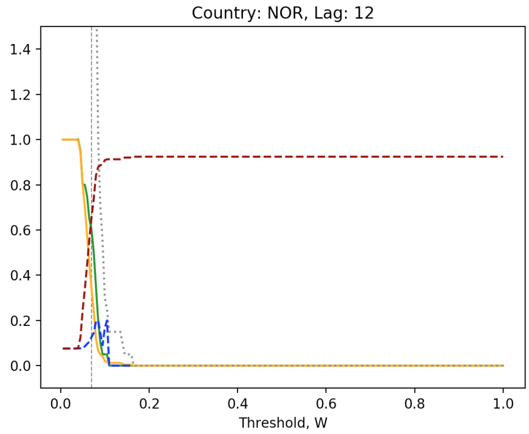


(b) Sweden

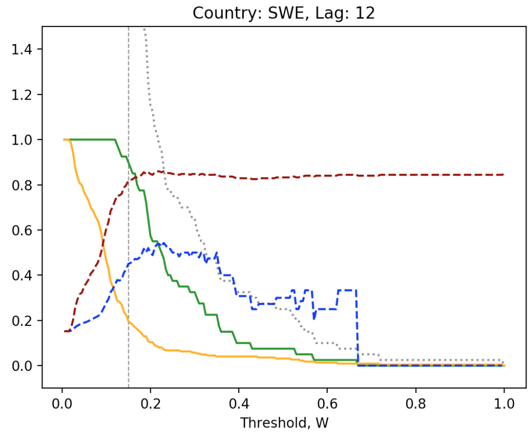


(c) Denmark

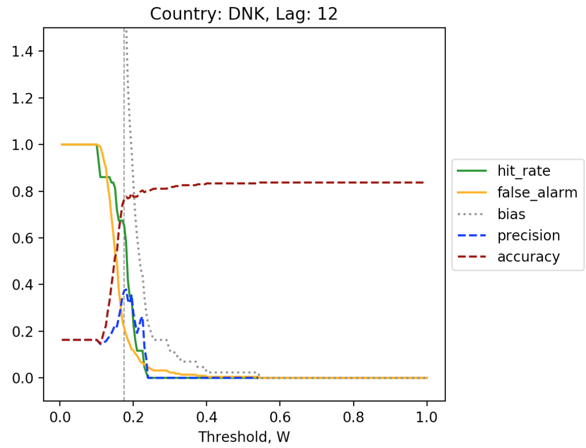
Figure B.4: *In-sample metrics for non-probabilistic recession forecasts at a 3 month horizon. The vertical grey dotted line indicates the W at which the difference between the hit rate and false alarm rate is the highest.*



(a) Norway



(b) Sweden



(c) Denmark

Figure B.5: *In-sample metrics for non-probabilistic recession forecasts at a 12 month horizon. The vertical grey dotted line indicates the W at which the difference between the hit rate and false alarm rate is the highest.*

Term spread SVM model – Norway

Lag	C / γ	**True negative**	**False negative**	**False positive**	**True positive**
3	1/0.01	104	10	0	0
6	1/0.01	104	10	0	0
12	1/0.01	104	10	0	0

Table B.1: *Term spread SVM model for Norway. The model was estimated on data between 1988-06 and 2019-6. Only the* `spread` *variable was used as explanatory variable. The dataset was limited a little to make sure models with different lag lengths where estimated using the same amount of data, to make the comparison fair.*

Term spread SVM model – Sweden

Lag	C / γ	True negative	False negative	False positive	True positive
3	4000/0.01	97	15	0	2
6	4500/0.025	94	12	3	5
12	5000/0.25	97	13	0	4

Table B.2: *Term spread SVM model for Sweden. The model was estimated on data between 1988-06 and 2019-6. Only the* `spread` *variable was used as explanatory variable. The dataset was limited a little to make sure models with different lag lengths where estimated using the same amount of data, to make the comparison fair.*

Term spread SVM model – Denmark

Lag	C / γ	True negative	False negative	False positive	True positive
3	15/0.01	103	11	0	0
6	1/0.01	103	11	0	0
12	2500/1.00	100	11	3	0

Table B.3: *Term spread SVM model for Denmark. The model was estimated on data between 1988-06 and 2019-6. Only the* `spread` *variable was used as explanatory variable. The dataset was limited a little to make sure models with different lag lengths where estimated using the same amount of data, to make the comparison fair.*

Long and short interest rate SVM model – Norway

Lag	C / γ	True negative	False negative	False positive	True positive
3	70/0.75	104	4	0	6
6	6000/0.1	100	9	4	1
12	9000/0.1	99	8	5	2

Table B.4: *Long and short interest rate SVM model for Norway. The model was estimated on data between 1988-06 and 2019-6. The* `long rate` *and* `short rate` *variables were used as explanatory variables. The dataset was limited a little to make sure models with different lag lengths where estimated using the same amount of data, to make the comparison fair.*

Long and short interest rate SVM model – Sweden

Lag	C / γ	True negative	False negative	False positive	True positive
3	9500/0.75	90	6	7	11
6	3500/0.05	93	8	4	9
12	1500/2.0	94	2	3	15

Table B.5: *Long and short interest rate SVM model for Sweden. The model was estimated on data between 1988-06 and 2019-6. The* `long rate` *and* `short rate` *variables were used as explanatory variables. The dataset was limited a little to make sure models with different lag lengths where estimated using the same amount of data, to make the comparison fair.*

Long and short interest rate SVM model – Denmark

Lag	C / γ	True negative	False negative	False positive	True positive
3	9500/0.75	99	3	4	8
6	3500/0.05	98	2	5	9
12	1500/2.0	95	3	8	8

Table B.6: *Long and short interest rate SVM model for Denmark. The model was estimated on data between 1988-06 and 2019-6. The* `long rate` *and* `short rate` *variables were used as explanatory variables. The dataset was limited a little to make sure models with different lag lengths where estimated using the same amount of data, to make the comparison fair.*

Norway – Recession

			Performance measures				
Lag	**Algorithm**	**Input**	**Hit rate**	**False alarm**	**Bias**	**Precision**	**Accuracy**
3	**Non prob**	MSCI	20.00%	17.31%	2.000	10.00%	77.19%
		Spread	50.00%	17.30%	2.300	21.70%	79.80%
	SVM	MSCI	0.00%	0.00%	0.000	-	91.23%
		Spread	0.00%	0.00%	0.000	-	91.20%
		Long/short	60.00%	0.00%	0.600	100.00%	96.50%
6	**Non prob**	MSCI	40.00%	2.88%	0.700	57.14%	92.11%
		Spread	60.00%	13.50%	2.000	30.00%	84.20%
	SVM	MSCI	0.00%	0.00%	0.000	-	91.23%
		Spread	0.00%	0.00%	0.000	-	91.20%
		Long/short	10.00%	3.60%	0.500	20.00%	88.60%
12	**Non prob**	MSCI	20.00%	18.27%	2.100	9.52%	76.32%
		Spread	60.00%	29.80%	3.700	16.20%	69.30%
	SVM	MSCI	0.00%	0.00%	0.000	-	91.23%
		Spread	0.00%	0.00%	0.000	-	91.20%
		Long/short	20.00%	4.80%	0.700	28.60%	88.60%

Table B.7: *The figure shows performance metrics for all models estimated for recessions in Norway. The algorithm column refers to the estimation technique used to estimate the model, the input column refers to the input variable(s) used in the model. MSCI refers to the models that uses lagged values of the monthly return of the stock indexes from MSCI as input variable, spread refers to the models that take the term spread as input variable and long/short refers to the models that take the long and short interest rates from the yield curve as input variables.*

Sweden – Recession

			Performance measures				
Lag	**Algorithm**	**Input**	**Hit rate**	**False alarm**	**Bias**	**Precision**	**Accuracy**
3	**Non prob**	MSCI	23.53%	10.31%	0.824	28.57%	79.82%
		Spread	82.40%	13.40%	1.588	51.90%	85.90%
	SVM	MSCI	0.00%	0.00%	0.000	-	85.09%
		Spread	11.80%	0.00%	0.118	100.00%	86.80%
		Long/short	64.70%	7.22%	1.060	61.11%	88.60%
6	**Non prob**	MSCI	11.76%	14.43%	0.941	12.50%	74.56%
		Spread	64.70%	11.30%	1.294	50.00%	85.10%
	SVM	MSCI	0.00%	0.00%	0.000	-	85.09%
		Spread	29.40%	3.10%	0.471	62.50%	86.80%
		Long/short	52.94%	4.12%	0.760	69.23%	89.47%
12	**Non prob**	MSCI	35.29%	24.74%	1.765	20.00%	69.30%
		Spread	100.00%	19.60%	2.117	47.20%	83.30%
	SVM	MSCI	0.00%	0.00%	0.000	-	85.09%
		Spread	23.50%	0.00%	0.235	100.00%	86.80%
		Long/short	88.24%	3.09%	1.060	83.33%	95.61%

Table B.8: *The figure shows performance metrics for all models estimated for recessions in Sweden. The algorithm column refers to the estimation technique used to estimate the model, the input column refers to the input variable(s) used in the model. MSCI refers to the models that uses lagged values of the monthly return of the stock indexes from MSCI as input variable, spread refers to the models that take the term spread as input variable and long/short refers to the models that take the long and short interest rates from the yield curve as input variables.*

Denmark – Recession							
			Performance measures				
Lag	**Algorithm**	**Input**	**Hit rate**	**False alarm**	**Bias**	**Precision**	**Accuracy**
3	**Non prob**	MSCI	18.18%	8.74%	1.000	18.18%	84.21%
		Spread	45.50%	6.80%	1.091	41.70%	88.60%
	SVM	MSCI	0.00%	0.00%	0.000	-	90.35%
		Spread	0.00%	0.00%	0.000	-	90.40%
		Long/short	72.73%	3.88%	1.090	66.67%	93.86%
6	**Non prob**	MSCI	36.36%	10.68%	1.364	26.67%	84.21%
		Spread	54.60%	23.30%	2.727	20.00%	74.60%
	SVM	MSCI	0.00%	0.00%	0.000	-	90.35%
		Spread	0.00%	0.00%	0.000	-	90.40%
		Long/short	81.82%	4.85%	1.270	64.29%	93.86%
12	**Non prob**	MSCI	0.00%	9.71%	0.909	0.00%	81.58%
		Spread	63.60%	22.30%	2.727	23.30%	76.30%
	SVM	MSCI	0.00%	0.00%	0.000	-	90.35%
		Spread	0.00%	2.90%	0.273	0.00%	87.70%
		Long/short	72.73%	7.77%	1.450	50.00%	90.35%

Table B.9: *The figure shows performance metrics for all models estimated for recessions in Denmark. The algorithm column refers to the estimation technique used to estimate the model, the input column refers to the input variable(s) used in the model. MSCI refers to the models that uses lagged values of the monthly return of the stock indexes from MSCI as input variable, spread refers to the models that take the term spread as input variable and long/short refers to the models that take the long and short interest rates from the yield curve as input variables.*

B.3 Yield Curve and Output Gaps - Statistical Relationships

Testing different lags -- Norway

	(1) outputgap	(2) outputgap	(3) outputgap	(4) outputgap	(5) outputgap	(6) outputgap	(7) outputgap	(8) outputgap	(9) outputgap	(10) outputgap	(11) outputgap	(12) outputgap	(13) outputgap	(14) outputgap	(15) outputgap	(16) outputgap	(17) outputgap	(18) outputgap
outputgap																		
L.spread	-0.166^{*}																	
	(0.0874)																	
L2.spread		-0.183^{**}																
		(0.0881)																
L3.spread			-0.207^{**}															
			(0.0884)															
L4.spread				-0.234^{***}														
				(0.0894)														
L5.spread					-0.244^{***}													
					(0.0905)													
L6.spread						-0.244^{***}												
						(0.0913)												
L7.spread							-0.232^{**}											
							(0.0914)											
L8.spread								-0.230^{**}										
								(0.0911)										
L9.spread									-0.233^{**}									
									(0.0912)									
L10.spread										-0.213^{**}								
										(0.0920)								
L11.spread											-0.202^{**}							
											(0.0959)							
L12.spread												-0.198^{**}						
												(0.0955)						
L13.spread													-0.199^{**}					
													(0.0941)					
L14.spread														-0.216^{**}				
														(0.0908)				
L15.spread															-0.227^{**}			
															(0.0904)			
L16.spread																-0.247^{***}		
																(0.0936)		
L17.spread																	-0.249^{**}	
																	(0.1000)	
L18.spread																		-0.253^{**}
																		(0.100)
_cons	0.115	0.121	0.129	0.137	0.139	0.138	0.132	0.130	0.129	0.120	0.115	0.113	0.113	0.118	0.120	0.124	0.122	0.121
	(0.117)	(0.118)	(0.118)	(0.119)	(0.119)	(0.120)	(0.119)	(0.119)	(0.119)	(0.119)	(0.119)	(0.119)	(0.119)	(0.118)	(0.118)	(0.119)	(0.120)	(0.119)
N	372	372	372	372	372	372	372	372	372	372	372	372	372	372	372	372	372	372
pseudo R^2	0.020	0.025	0.031	0.039	0.043	0.043	0.040	0.039	0.041	0.036	0.033	0.032	0.032	0.036	0.040	0.048	0.050	0.052

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure B.6: *The figure shows the results of estimating:* $outputgap_t = \beta_0 + \beta_1 * spread_{t-\lambda} + \epsilon_t$ *where* $\lambda \in [1, 18]$ *for Norway. The standard errors are clustered at year-quarter level. The regression was estimated on data from 1988m7 - 2019m6, lags were calculated using data up to 18 months before 1988m7, this means that all regressions are estimated on the same number of observations.*

Testing different lags -- Sweden

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap
outputgap																		
L.spread	0.214*																	
	(0.117)																	
L2.spread		0.171																
		(0.108)																
L3.spread			0.126															
			(0.0973)															
L4.spread				0.0807														
				(0.0916)														
L5.spread					0.0368													
					(0.0877)													
L6.spread						-0.00835												
						(0.0842)												
L7.spread							-0.0488											
							(0.0832)											
L8.spread								-0.0711										
								(0.0834)										
L9.spread									-0.0890									
									(0.0834)									
L10.spread										-0.111								
										(0.0838)								
L11.spread											-0.132							
											(0.0850)							
L12.spread												-0.149*						
												(0.0859)						
L13.spread													-0.169*					
													(0.0865)					
L14.spread														-0.189**				
														(0.0885)				
L15.spread															-0.209**			
															(0.0899)			
L16.spread																-0.230**		
																(0.0905)		
L17.spread																	-0.242***	
																	(0.0929)	
L18.spread																		-0.236**
																		(0.0923)
_cons	-0.294	-0.248	-0.201	-0.153	-0.107	-0.0589	-0.0154	0.00893	0.0286	0.0532	0.0770	0.0962	0.119	0.142	0.164	0.187	0.200	0.195
	(0.179)	(0.169)	(0.158)	(0.152)	(0.148)	(0.144)	(0.143)	(0.143)	(0.143)	(0.144)	(0.146)	(0.147)	(0.148)	(0.151)	(0.153)	(0.154)	(0.156)	(0.154)
N	370	370	370	370	370	370	370	370	370	370	370	370	370	370	370	370	370	370
pseudo R^2	0.035	0.023	0.012	0.005	0.001	0.000	0.002	0.004	0.006	0.009	0.013	0.016	0.020	0.025	0.030	0.036	0.039	0.037

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure B.7: *The figure shows the results of estimating:* $outputgap_t = \beta_0 + \beta_1 * spread_{t-\lambda} + \epsilon_t$ *where* $\lambda \in [1, 18]$ *for Sweden. The standard errors are clustered at year-quarter level. The regression was estimated on data from 1988m7 - 2019m6, lags were calculated using data up to 18 months before 1988m7, this means that all regressions are estimated on the same number of observations.*

Testing different lags -- Denmark

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap	outputgap
outputgap																		
L.spread	-0.0702																	
	(0.0673)																	
L2.spread		-0.0766																
		(0.0675)																
L3.spread			-0.0812															
			(0.0681)															
L4.spread				-0.0824														
				(0.0687)														
L5.spread					-0.0821													
					(0.0694)													
L6.spread						-0.0803												
						(0.0697)												
L7.spread							-0.0851											
							(0.0702)											
L8.spread								-0.0882										
								(0.0726)										
L9.spread									-0.101									
									(0.0751)									
L10.spread										-0.124^{*}								
										(0.0746)								
L11.spread											-0.145^{**}							
											(0.0729)							
L12.spread												-0.143^{**}						
												(0.0728)						
L13.spread													-0.112					
													(0.0742)					
L14.spread														-0.0694				
														(0.0772)				
L15.spread															-0.0480			
															(0.0777)			
L16.spread																-0.0309		
																(0.0775)		
L17.spread																	-0.0404	
																	(0.0758)	
L18.spread																		-0.0434
																		(0.0755)
_cons	0.236^{*}	0.241^{*}	0.245^{**}	0.246^{**}	0.246^{**}	0.245^{*}	0.249^{**}	0.251^{**}	0.263^{**}	0.282^{**}	0.301^{**}	0.299^{**}	0.272^{**}	0.237^{*}	0.220^{*}	0.207	0.214^{*}	0.216^{*}
	(0.123)	(0.124)	(0.124)	(0.124)	(0.125)	(0.125)	(0.126)	(0.127)	(0.129)	(0.129)	(0.128)	(0.127)	(0.128)	(0.130)	(0.130)	(0.129)	(0.128)	(0.128)
N	372	372	372	372	372	372	372	372	372	372	372	372	372	372	372	372	372	372
pseudo R^2	0.005	0.006	0.006	0.007	0.007	0.006	0.007	0.008	0.010	0.015	0.019	0.018	0.012	0.005	0.003	0.001	0.002	0.002

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure B.8: *The figure shows the results of estimating:* $outputgap_t = \beta_0 + \beta_1 * spread_{t-\lambda} + \epsilon_t$ *where* $\lambda \in [1, 18]$ *for Denmark. The standard errors are clustered at year-quarter level. The regression was estimated on data from 1988m7 - 2019m6, lags were calculated using data up to 18 months before 1988m7, this means that all regressions are estimated on the same number of observations.*

BIC – Starting at 3 month lag			
# of lags	**Norway**	**Sweden**	**Denmark**
3	**510.5**	**517.4**	**516.4**
4	511.4	519.6	522.2
5	515.7	520.2	528.0
6	521.1	519.4	533.9
7	527.0	518.3	539.4
8	532.0	521.2	544.9
9	536.9	524.2	549.4
10	542.8	526.0	552.0
11	548.6	528.3	555.3
12	554.1	530.5	561.2

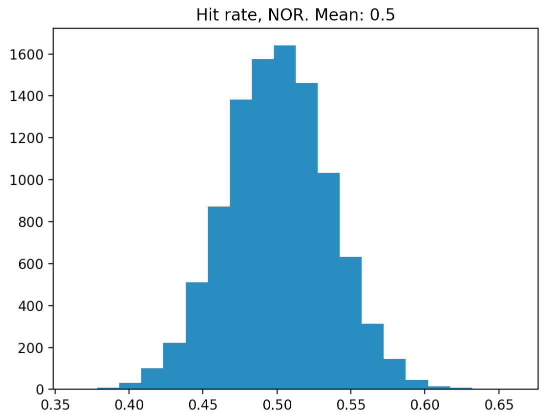
Table B.10: *Table showing BIC scores from estimating probit models with different lag lengths, starting at 3 months. Standard errors are clustered at year-quarter level. The minimum values are highlighted in bold font. The scores are estimated using data from 1988m7 to 2019m6, and lags are calculated on values prior to this date in order to have the same number of observations for all models.*

BIC – Starting at 6 month lag

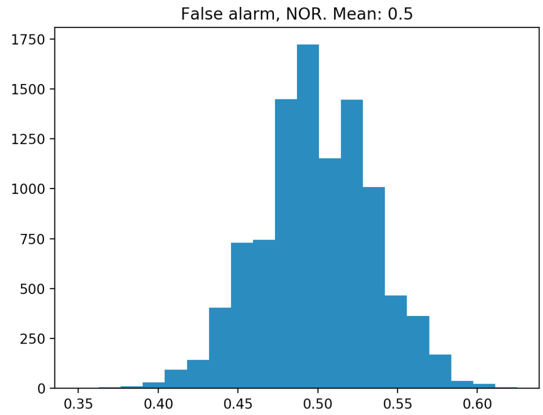
# of lags	Norway	Sweden	Denmark
6	504.4	523.6	516.5
7	510.3	524.0	521.9
8	515.7	526.7	527.6
9	520.7	529.9	532.0
10	526.6	532.0	534.5
11	532.3	534.1	537.8
12	537.9	536.7	543.7

Table B.11: *Table showing BIC scores from estimating probit models with different lag lengths, starting at 6 months. Standard errors are clustered at year-quarter level. The minimum values are highlighted in bold font. The scores are estimated using data from 1988m7 to 2019m6, and lags are calculated on values prior to this date in order to have the same number of observations for all models.*

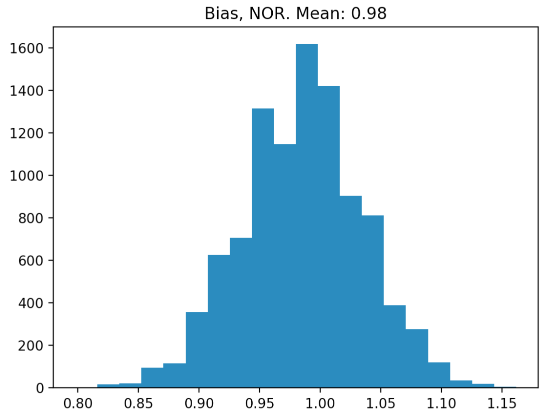
B.4 Out-of-sample Output Gap Forecasting



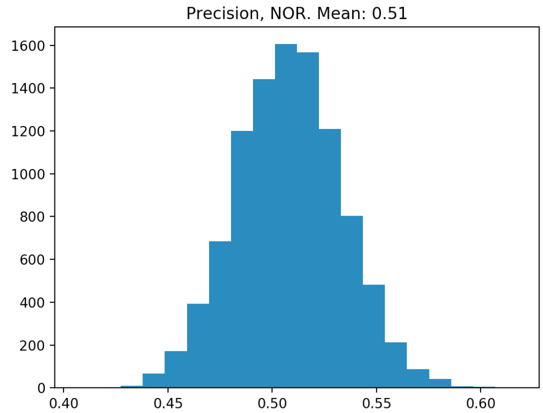
(a) Hit rate



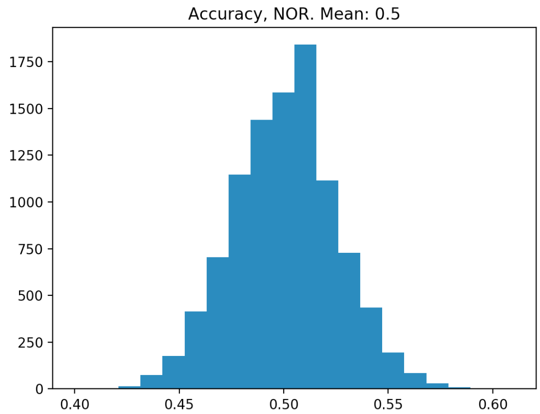
(b) False alarm



(c) Bias

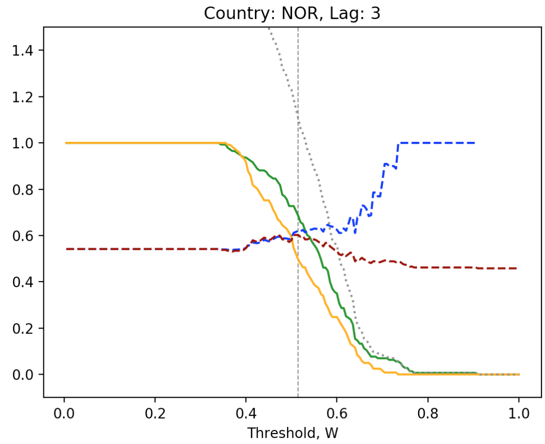


(d) Precision

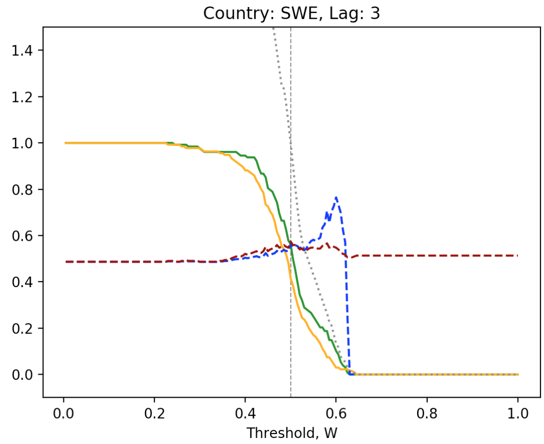


(e) Accuracy

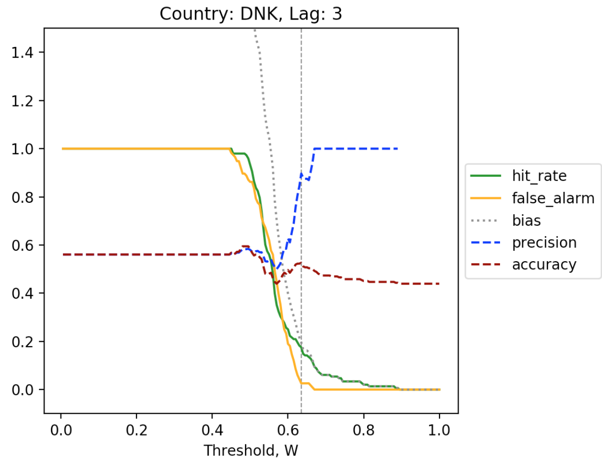
Figure B.9: *The histograms show results of 10 000 simulations of a random forecast of the output gap variable for Norway (results looks very similar for the other two countries, although the precision metric is a generally a little higher for countries that have a higher share of negative output gaps). These results stand as a benchmark of how well the actual forecast models should perform in order to provide any valuable information. The simulations are created using* `Python`.



(a) Norway 3 month lag

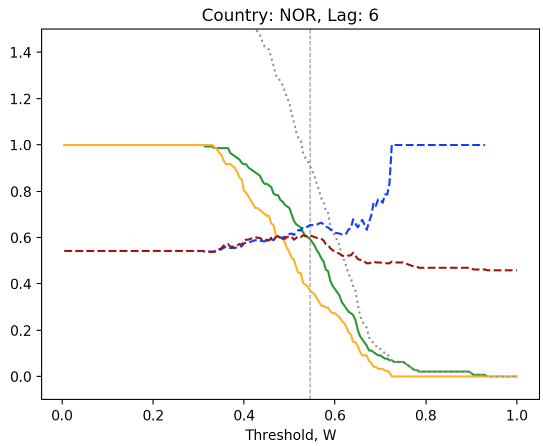


(b) Sweden 3 month lag

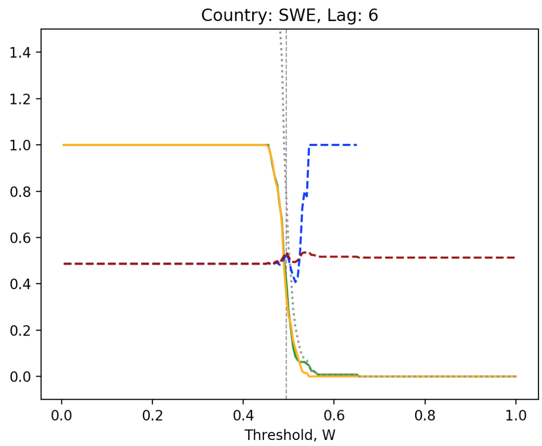


(c) Denmark 3 month lag

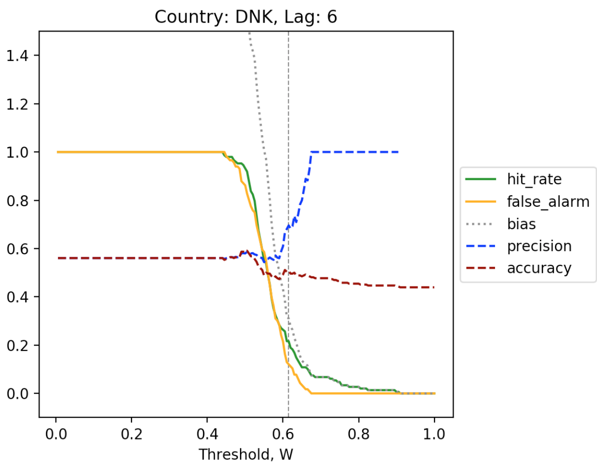
Figure B.10: *Figures indicate how the performance metrics change when W varies between 0 and 1 in the case of the in-sample analysis of the non probabilistic output gap model with the 3 month lag of* **`spread`**. *These results are calculated based on the coefficients in table 6.9 on page 51.*



(a) Norway 6 month lag

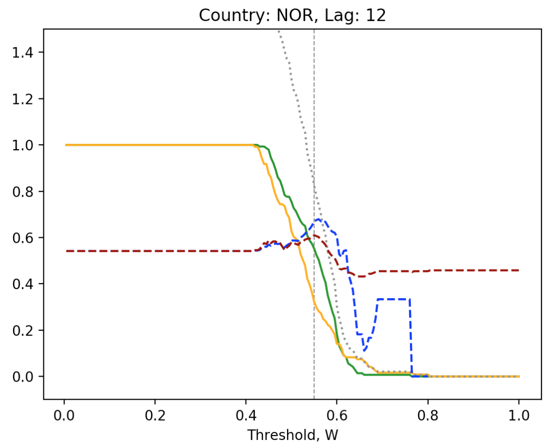


(b) Sweden 6 month lag

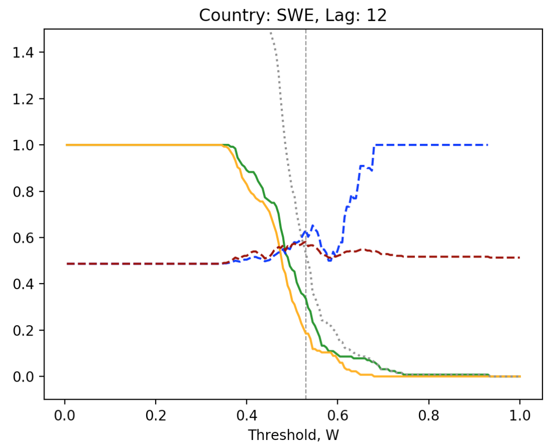


(c) Denmark 6 month lag

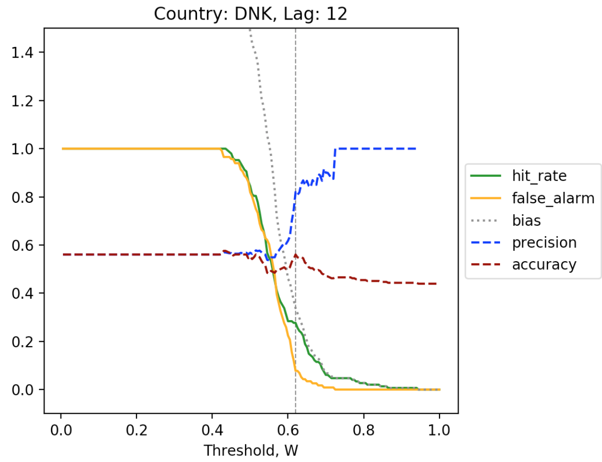
Figure B.11: *Figures indicate how the performance metrics change when W varies between 0 and 1 in the case of the in-sample analysis of the non probabilistic output gap model with the 6 month lag of* **`spread`**. *These results are calculated based on the coefficients in table 6.9 on page 51.*



(a) Norway 12 month lag



(b) Sweden 12 month lag



(c) Denmark 12 month lag

Figure B.12: *Figures indicate how the performance metrics change when W varies between 0 and 1 in the case of the in-sample analysis of the non probabilistic output gap model with the 12 month lag of* **`spread`**. *These results are calculated based on the coefficients in table 6.9 on page 51.*

Term spread SVM Model – Norway

Lag	C / γ	True negative	False negative	False positive	True positive
3	44/0.75	27	12	32	43
6	500/0.01	18	6	41	49
12	7500/2.00	35	19	24	36

Table B.12: *The model was estimated on data between 1988-06 and 2019-6 for Norway. The `spread` variable was used as the explanatory variable. The dataset was limited a little to make sure models with different lag lengths where estimated using the same amount of data, to make the comparison fair*

Term spread SVM Model – Sweden

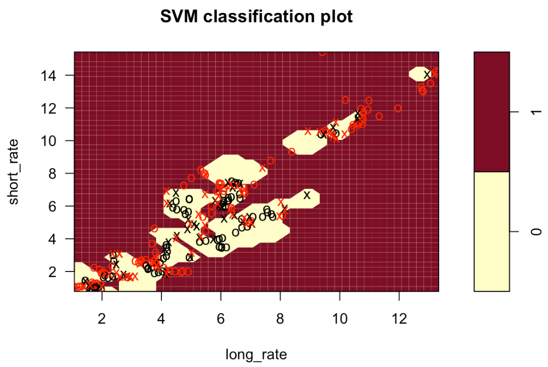
Lag	C / γ	True negative	False negative	False positive	True positive
3	6000/1.00	39	27	27	21
6	900/0.25	44	29	22	19
12	1600/2.00	49	26	17	22

Table B.13: *The model was estimated on data between 1988-06 and 2019-6 for Sweden. The `spread` variable was used as the explanatory variable. The dataset was limited a little to make sure models with different lag lengths where estimated using the same amount of data, to make the comparison fair*

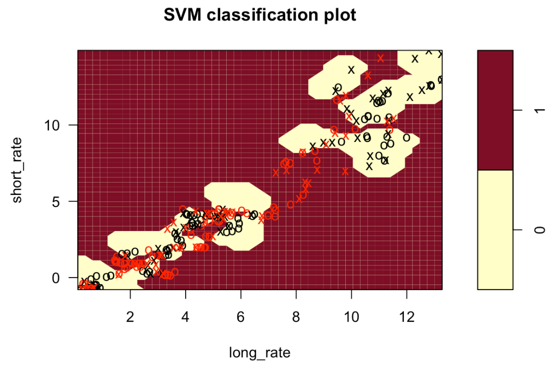
Term spread SVM Model – Denmark

Lag	C / γ	True negative	False negative	False positive	True positive
3	5000/2.00	22	12	27	53
6	37/2.00	11	7	38	58
12	8/2.00	26	22	23	43

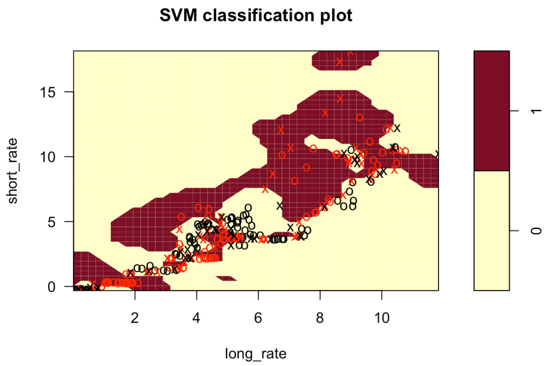
Table B.14: *The model was estimated on data between 1988-06 and 2019-6 for Denmark. The `spread` variable was used as the explanatory variable. The dataset was limited a little to make sure models with different lag lengths where estimated using the same amount of data, to make the comparison fair*



(a) Norway 3 month lag

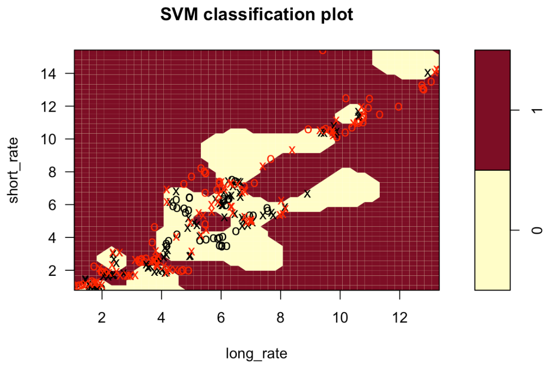


(b) Sweden 3 month lag

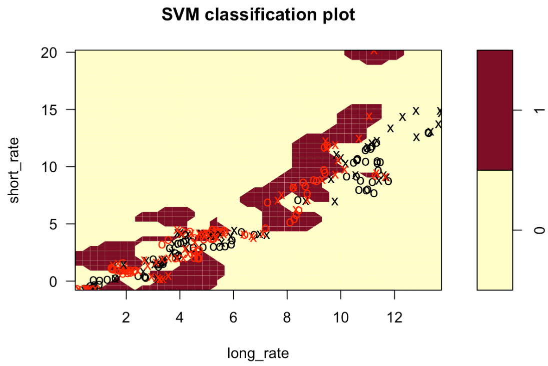


(c) Denmark 3 month lag

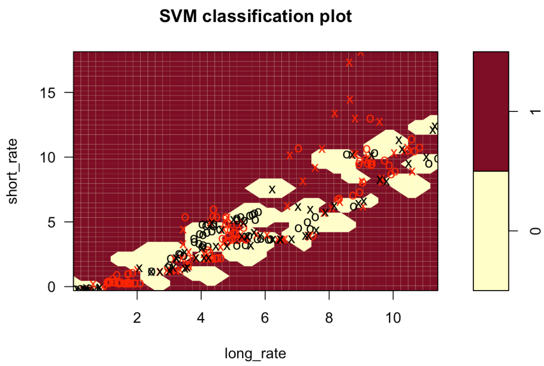
Figure B.13: *Figure showing in-sample results from estimating a SVM model with a 3 month lag of the independent variables, long and short interest rates.* ***NOTE:*** *The support vector observations are plotted as crosses and the non-support vector observations are circles James et al. (2017, page 360). The background color of the chart indicate the classification of an observation with a given coordinate that the SVM assigns based on the decision boundary calculated. Also, the decision boundary looks jagged in the plot however, according to this is just a property of the plotting library used, in reality the decision lines are smooth.*



(a) Norway 6 month lag

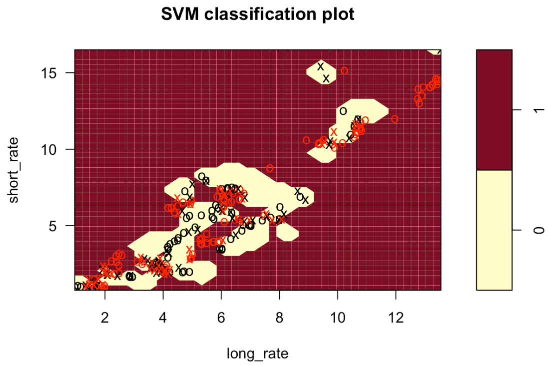


(b) Sweden 6 month lag

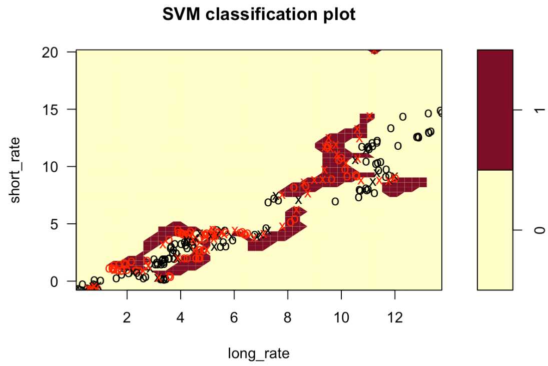


(c) Denmark 6 month lag

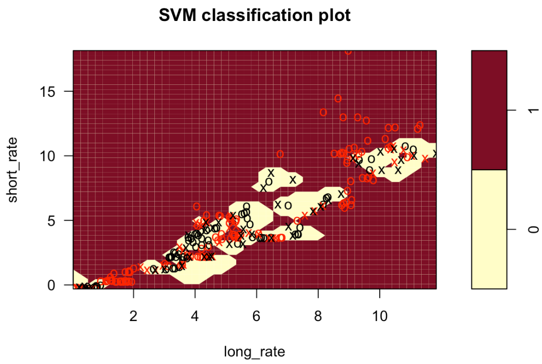
Figure B.14: *Figure showing in-sample results from estimating a SVM model with a 6 month lag of the independent variables, long and short interest rates.* ***NOTE:*** *The support vector observations are plotted as crosses and the non-support vector observations are circles James et al. (2017, page 360). The background color of the chart indicate the classification of an observation with a given coordinate that the SVM assigns based on the decision boundary calculated. Also, the decision boundary looks jagged in the plot however, according to this is just a property of the plotting library used, in reality the decision lines are smooth.*



(a) Norway 12 month lag



(b) Sweden 12 month lag



(c) Denmark 12 month lag

Figure B.15: *Figure showing in-sample results from estimating a SVM model with a 12 month lag of the independent variables, long and short interest rates.* ***NOTE:*** *The support vector observations are plotted as crosses and the non-support vector observations are circles James et al. (2017, page 360). The background color of the chart indicate the classification of an observation with a given coordinate that the SVM assigns based on the decision boundary calculated. Also, the decision boundary looks jagged in the plot however, according to this is just a property of the plotting library used, in reality the decision lines are smooth.*

Long and short interest rate SVM Model – Norway

Lag	C / γ	True negative	False negative	False positive	True positive
3	5500/2.00	39	15	20	40
6	1000/0.25	34	10	25	45
12	7500/2.00	46	12	13	43

Table B.15: *Results from SVM with interest rates as input variables and output gap as dependent variable. The model was estimated on data between 1988-06 and 2019-6 for Norway. The dataset was limited a little to make sure models with different lag lengths where estimated using the same amount of data, to make the comparison fair.*

Long and short interest rate SVM Model – Sweden

Lag	C / γ	True negative	False negative	False positive	True positive
3	1900/2.00	54	12	12	36
6	6000/2.00	57	14	9	34
12	1500/2.00	50	11	16	37

Table B.16: *Results from SVM with interest rates as input variables and output gap as dependent variable. The model was estimated on data between 1988-06 and 2019-6 for Sweden. The dataset was limited a little to make sure models with different lag lengths where estimated using the same amount of data, to make the comparison fair.*

Long and short interest rate SVM Model – Denmark

Lag	C / γ	True negative	False negative	False positive	True positive
3	6000/1.00	41	13	8	52
6	200/2.00	36	16	13	49
12	6000/2.00	35	8	14	57

Table B.17: *Results from SVM with interest rates as input variables and output gap as dependent variable. The model was estimated on data between 1988-06 and 2019-6 for Denmark. The dataset was limited a little to make sure models with different lag lengths where estimated using the same amount of data, to make the comparison fair.*

Norway

			Performance measures				
Lag	**Algorithm**	**Input**	**Hit rate**	**False alarm**	**Bias**	**Precision**	**Accuracy**
3	**Non prob**	MSCI	23.64%	6.78%	0.309	76.47%	59.65%
		Spread	74.55%	45.76%	1.236	60.29%	64.04%
	SVM	MSCI	100.00%	100.00%	2.073	48.25%	48.25%
		Spread	78.18%	54.24%	1.360	57.33%	61.40%
		Long/short	72.73%	33.90%	1.090	66.67%	69.30%
6	**Non prob**	MSCI	32.73%	15.25%	0.491	66.67%	59.65%
		Spread	67.27%	30.51%	1.000	67.27%	68.42%
	SVM	MSCI	100.00%	100.00%	2.073	48.25%	48.25%
		Spread	89.09%	69.49%	1.640	54.44%	58.77%
		Long/short	81.82%	42.37%	1.270	64.29%	69.30%
12	**Non prob**	MSCI	56.36%	42.37%	1.018	55.36%	57.02%
		Spread	65.45%	32.20%	1.000	65.45%	66.67%
	SVM	MSCI	100.00%	100.00%	2.073	48.25%	48.25%
		Spread	65.45%	40.68%	1.090	60.00%	62.28%
		Long/short	78.18%	22.03%	1.020	76.79%	78.07%

Table B.18: *The figure shows performance metrics for all models estimated for output gaps in Norway. The algorithm column refers to the estimation technique used to estimate the model, the input column refers to the input variable(s) used in the model. MSCI refers to the models that uses the stock indexes from MSCI as input variable, spread refers to the models that take the term spread as input variable and long/short refers to the models that take the lon and short interest rates from the yield curve as input variables.*

Sweden

			Performance measures				
Lag	**Algorithm**	**Input**	**Hit rate**	**False alarm**	**Bias**	**Precision**	**Accuracy**
3	**Non prob**	MSCI	6.25%	18.18%	0.313	20.00%	50.00%
		Spread	60.42%	45.45%	1.229	49.15%	57.02%
	SVM	MSCI	2.08%	4.55%	0.083	25.00%	56.14%
		Spread	43.75%	40.91%	1.000	43.75%	52.63%
		Long/short	75.00%	18.18%	1.000	75.00%	78.95%
6	**Non prob**	MSCI	56.25%	63.64%	1.438	39.13%	44.74%
		Spread	37.50%	34.85%	0.854	43.90%	53.51%
	SVM	MSCI	0.00%	0.00%	0.000	-	57.89%
		Spread	39.58%	33.33%	0.850	46.34%	55.26%
		Long/short	70.83%	13.64%	0.900	79.07%	79.82%
12	**Non prob**	MSCI	54.17%	57.58%	1.333	40.62%	47.37%
		Spread	31.25%	18.18%	0.563	55.56%	60.53%
	SVM	MSCI	16.67%	22.73%	0.479	34.78%	51.75%
		Spread	45.83%	25.76%	0.810	56.41%	62.28%
		Long/short	77.08%	24.24%	1.100	69.81%	76.32%

Table B.19: *The figure shows performance metrics for all models estimated for output gaps in Sweden. The algorithm column refers to the estimation technique used to estimate the model, the input column refers to the input variable(s) used in the model. MSCI refers to the models that uses the stock indexes from MSCI as input variable, spread refers to the models that take the term spread as input variable and long/short refers to the models that take the lon and short interest rates from the yield curve as input variables.*

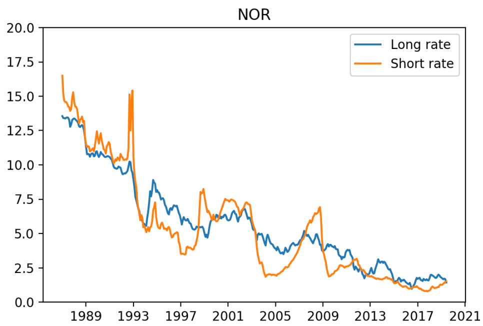
Denmark

			Performance measures				
Lag	**Algorithm**	**Input**	**Hit rate**	**False alarm**	**Bias**	**Precision**	**Accuracy**
3	**Non prob**	MSCI	47.69%	46.94%	0.831	57.41%	50.00%
		Spread	9.23%	6.12%	0.139	66.67%	45.61%
	SVM	MSCI	100.00%	100.00%	1.754	57.02%	57.02%
		Spread	81.54%	55.10%	1.230	66.25%	65.79%
		Long/short	80.00%	16.33%	0.920	86.67%	81.58%
6	**Non prob**	MSCI	49.23%	46.94%	0.846	58.18%	50.88%
		Spread	15.38%	18.37%	0.292	52.63%	43.86%
	SVM	MSCI	100.00%	100.00%	1.754	57.02%	57.02%
		Spread	89.23%	77.55%	1.480	60.42%	60.53%
		Long/short	75.38%	26.53%	0.950	79.03%	74.56%
12	**Non prob**	MSCI	67.69%	59.18%	1.123	60.27%	56.14%
		Spread	24.62%	10.20%	0.323	76.19%	52.63%
	SVM	MSCI	100.00%	100.00%	1.754	57.02%	57.02%
		Spread	66.15%	46.94%	1.020	65.15%	60.53%
		Long/short	87.69%	28.57%	1.090	80.28%	80.70%

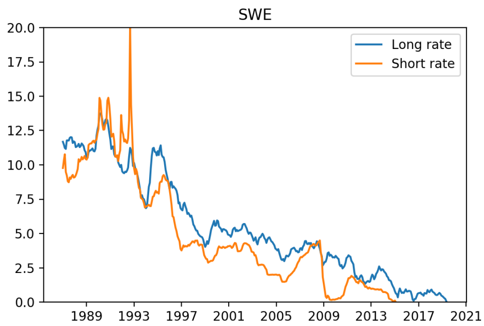
Table B.20: *The figure shows performance metrics for all models estimated for output gaps in Sweden. The algorithm column refers to the estimation technique used to estimate the model, the input column refers to the input variable(s) used in the model. MSCI refers to the models that uses the stock indexes from MSCI as input variable, spread refers to the models that take the term spread as input variable and long/short refers to the models that take the lon and short interest rates from the yield curve as input variables.*

Appendix C

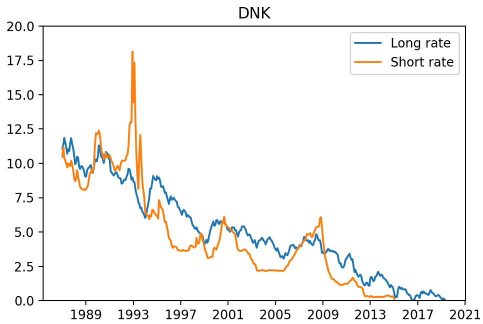
Discussion Appendix



(a) Norway



(b) Sweden



(c) Denmark

Figure C.1: *Figures show a time series plot of the interest rate data used in this thesis.*

Adding short term rate - NOR

	(1) recession	(2) recession
recession		
L3.spread	-0.231*	-0.432***
	(0.131)	(0.161)
L3.short_rate		-0.108*
		(0.0627)
_cons	-1.515***	-0.984***
	(0.184)	(0.361)
N	372	372
pseudo R^2	0.060	0.099

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Adding short term rate - NOR

	(1) recession	(2) recession
recession		
L6.spread	-0.165	-0.293*
	(0.134)	(0.167)
L6.short_rate		-0.0720
		(0.0633)
_cons	-1.506***	-1.130***
	(0.179)	(0.366)
N	372	372
pseudo R^2	0.030	0.050

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Adding short term rate - NOR

	(1) recession	(2) recession
recession		
L12.spread	-0.0950	-0.198
	(0.129)	(0.166)
L12.short_rate		-0.0583
		(0.0617)
_cons	-1.505***	-1.187***
	(0.177)	(0.369)
N	372	372
pseudo R^2	0.009	0.024

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Adding short term rate - SWE

	(1) recession	(2) recession
recession		
L3.spread	-0.679***	-0.665***
	(0.201)	(0.258)
L3.short_rate		0.00497
		(0.0437)
_cons	-0.551***	-0.584
	(0.175)	(0.356)
N	370	370
pseudo R^2	0.291	0.291

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Adding short term rate - SWE

	(1) recession	(2) recession
recession		
L6.spread	-0.874***	-0.945***
	(0.220)	(0.300)
L6.short_rate		-0.0215
		(0.0494)
_cons	-0.479**	-0.332
	(0.205)	(0.407)
N	370	370
pseudo R^2	0.374	0.375

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Adding short term rate - SWE

	(1) recession	(2) recession
recession		
L12.spread	-0.486***	-0.376*
	(0.180)	(0.196)
L12.short_rate		0.0511
		(0.0407)
_cons	-0.695***	-1.041***
	(0.203)	(0.378)
N	370	370
pseudo R^2	0.213	0.226

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Adding short term rate - DNK

	(1) recession	(2) recession
recession		
L3.spread	-0.263***	-0.183
	(0.0922)	(0.120)
L3.short_rate		0.0535
		(0.0406)
_cons	-0.931***	-1.220***
	(0.147)	(0.257)
N	372	372
pseudo R^2	0.092	0.102

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Adding short term rate - DNK

	(1) recession	(2) recession
recession		
L6.spread	-0.198**	-0.110
	(0.0911)	(0.113)
L6.short_rate		0.0607
		(0.0397)
_cons	-0.957***	-1.290***
	(0.150)	(0.261)
N	372	372
pseudo R^2	0.056	0.070

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Adding short term rate - DNK

	(1) recession	(2) recession
recession		
L12.spread	-0.112	-0.0286
	(0.0813)	(0.103)
L12.short_rate		0.0612
		(0.0400)
_cons	-0.990***	-1.335***
	(0.150)	(0.272)
N	372	372
pseudo R^2	0.017	0.033

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure C.2: *The figure show the results of adding the short term interest rate to the probit models estimated with recession as the dependent variable in chapter 6. The following regression was estimated with clustered errors at year-quarter level:* $recession_t = \beta_0 + \beta_1 * spread_{t-\lambda} + \beta_2 * short\_rate_{t-\lambda} + \epsilon_t$ *where* $\lambda \in \{3, 6, 12\}$.