Stockholm School of Economics
Department of Economics
Master's Thesis
September 2006

Cognitive Biases in Macroeconomic Forecasts

– A Pooled Post-Mortem Study of OECD Forecasts

Abstract. This thesis has two purposes: 1) To evaluate the forecast quality of the OECD and 2) to analyze whether the OECD forecasts are afflicted by cognitive bias. Cognitive bias refers to a wide range of basic judgmental errors that are common to humans. Specifically we examine if the OECD forecasters 1) are too optimistic or pessimistic, 2) become more pessimistic the closer they are to a target year, 3) anchor too much on certain reference points, or that they 4) make forecasts that are too positive or negative after consistent change. We choose to look at the forecasters at the OECD during the period 1985 - 2004. Forecasts of five different horizons for GDP and its final domestic demand components represent the database for 23 developed countries. We find support for too optimistic projections of government consumption, while long-term forecasts are more optimistic than short-term forecasts for GDP, government consumption and fixed investment. Furthermore, forecasters make too negative forecasts after positive trends and vice versa. Regarding forecast quality, we find that OECD is accurate, which supports past findings. However, in contrast to past studies, we observe more irrationality in the forecasters.

Authors:	Olivia Fleischmann	(80277@student.hhs.se)
	Axel Svedbom	(19225@student.hhs.se)
Tutor:	Dr. Patric Andersson	
Examiner:	Ass. Prof. David Domeij	
Discussant:	Nadia Martin	
Presentation:	10:15-12:00 September 8, 2006	

Acknowledgments

We would like to thank our excellent tutor Dr. Patric Andersson for his invaluable help and support. We are also grateful to Mr. Colin Ash, University of Reading and Prof. Stephan Ekkehard, University of Cologne for supplying research material. Additionally, appreciation is also extended to Prof. Henry Montgomery, University of Stockholm, Mr. Gustav Törngren, University of Stockholm, and Mr. Karl Wennberg, Stockholm School of Economics, for taking time to discuss the subject of this thesis. Naturally, any errors or inaccuracies in the thesis are the sole responsibility of the authors and does not imply concurrence with any of the above-mentioned persons.

Table of contents

1. Introduction

"One of the soundest rules I try to remember when making forecasts in the field of economics is that whatever is to happen is happening already" (Sylvia Field Porter, 1994).

Sylvia F. Porter was a syndicated financial columnist who wrote for both the *New York Post* and the *New York Daily News* and she, like so many others in economics and business, searched for the key to forecasting success. This search continues to levels where the models and the technology used is highly advanced (Webb, 1999). Even though the models may change and the technology will become even more advanced, there is one aspect to forecasting that will always play a role: Judgment. When making projections for future periods, forecasters must decide what aspects will play the largest role in the future, how these aspects fit their model, and if the future is rosier than today or if it will become just that much darker.

This thesis focuses on one subsection of the enormous amount of economic forecasts that are produced - namely macroeconomic forecasts. More specifically, we explore the world of forecast evaluation from both the surface and below surface i.e. on the psychological level. The forecasting institution we examine is the Organization of Economic Co-operation and Development (OECD) that is based in Paris. Bi-annually, the OECD publishes the *Economic Outlook* that includes past outcomes and future projections of the economic climates for its 30 member countries.

Though the *Economic Outlook* has been published since 1967, prior to 1985 detailed forecasts were published for the G7 countries only. This study focuses on the period 1985-2004 to achieve the largest range of countries that remains stable at all the forecasting horizons. This particular dataset is one of the characteristics that set this study apart from others: To the best of our knowledge, no other pooled study has considered such a large sample of countries. Another defining characteristic of this study is the wide span of forecasting horizons, which appears, to the best of our knowledge, to be the broadest considered thus far.

The analysis focus around the macroeconomic variables of Gross Domestic Product (GDP) and its major expenditure components: Private consumption, government consumption and fixed investment. Together these components make up around 100% of GDP (where the remainder mainly consists of net exports and change in stockbuilding). These are major indicators of the economic climate of a country and together sum to the final domestic demand. Previous research on macroeconomic forecasts has seldom been concerned with these components and here we attempt to add to past findings, by possibly shedding light on an area that is sparse on research.

As human judgment plays a crucial role in forecasting, cognitive biases may affect the forecasts in many ways (Bolger & Harvey, 1998). However, little research has been conducted to address the question of cognitive biases in macroeconomic forecasts. The four cognitive biases considered in this post-mortem study are the *optimism and pessimism* bias, the *temporal differences in optimism* bias, the *anchoring and adjustment* bias, and the *hot hand* and *gambler's fallacy* bias. First, the optimism and pessimism bias relates to the possibility that the OECD has a systematic tendency to over and/or under shoot its projections. Second, temporal differences in optimism relates to possible differences in optimism for longer versus shorter horizon forecasts. Third, the anchoring and adjustment bias concerns forecasters' propensity to rely on a particular reference point and then make insufficient adjustments from this point. Fourth, the hot hand and gambler's fallacy bias concerns the notion that forecasts may be based on the belief that trends will continue (hot hand bias) or break (gambler's fallacy bias) more than is actually warranted. To ease readability, the optimism/pessimism bias is referred to as the optimism bias, the temporal differences in optimism bias is referred to as the temporal bias, the anchoring and adjustment bias is referred to as the anchoring bias, and the hot hand/gambler's fallacy bias is referred to as the representativeness bias.[1] If identification

[1] The name representativeness bias stems from a hypothesized origin of the hot hand and the gambler's fallacy biases. It has been argued that the hot hand and the gambler's fallacy biases comes from a heuristic,

of any of the above biases occurs in the dataset, it would appear that the results might help in future forecasting endeavors.

This thesis seeks to examine two main objectives. Our first objective is to evaluate the quality of the OECD's forecasts for GDP and its expenditure components. We have chosen the OECD for this study because of its renowned reputation for forecasting, in addition to its readily available data. Our second objective is to determine if the forecasts exhibit signs of being afflicted with systematic errors that are caused by cognitive biases. In the past, little research has been conducted on macroeconomic forecasts to investigate the biases we have chosen to analyze.

This thesis is organized as follows. Section II discusses the evaluation of forecasts. Section III describes the link between psychology and forecasting and provides a detailed explanation of each bias. Section IV specifies the dataset and the OECD's forecasting process. Section V explains the methodology used for the forecast evaluation and presents the results. Section VI explains the methodology used to analyze the cognitive biases and presents the results. Section VII provides a conclusion with a discussion of the results and suggestions for further research. Abbreviations used throughout the thesis are presented in Table 1.

---

or cognitive rule of thumb, called the representativeness heuristics (cf. Gilovich, Griffin & Kahneman, 2002).

Table 1. Abbreviations and guide to error calculation

Name	*Abbreviation*	*Publish Date and Calculation*
Estimate of the current year	ForEst	Dec Year t
0.5-year forecast	For0.5Y	May Year t
1.0-year forecast	For1.0Y	Dec Year t-1
1.5-year forecast	For1.5Y	May Year t-1
2.0-year forecast	For2.0Y	Dec Year t-2
Time	t	na
Forecast horizon	h	na
Variable	v	na

2. Evaluation of Macroeconomic Forecasts

This section begins with a general introduction to macroeconomic forecast evaluation. Subsequently, a review of previous studies is presented, followed by a summary of their findings. Finally, we discuss how this thesis links to previous research.

2.1 Introduction

There is a vast amount of literature that explores macroeconomic forecasting performance. Fildes and Stekler (2002) provide an extensive overview of this field. The two authors are concerned mainly with short-term forecasts (forecasts for one year or less) for GDP and inflation in both the UK and the US. They present four dimensions that most macroeconomic forecast evaluations use in their analysis.

- *Quantitative accuracy* measures forecast accuracy through describing the size of the forecast errors. Examples of quantitative accuracy measures are the *mean error (ME),* the *mean absolute error (MAE)*, the *mean square error (MSE)*, and the *root mean square error (RMSE).*[2] The formulas for calculating the measures can be found in Appendix I.

- *Directional accuracy* focuses on the forecasters' ability to correctly predict the direction of change in economic variables. There are several aspects of directional accuracy. It can describe forecasters' record of predicting turning points in the economy but it can also refer to the association between forecast directional change and actual directional change.

- *Naïve/no-change comparisons* assess whether expert forecasts are valuable to their users by comparing the forecasts to naïve benchmarks. Several models have been proposed to generate naïve benchmarks. The most commonly used model is one that generates forecasts that assume no change from the preceding period.

- *Rationality of forecasts* examines if the forecasts meet the criterion of a rationality definition. A frequently used definition of forecast rationality, which is exercised in this thesis, is that rational forecasts do not exhibit systematic errors. Wärneryd (2001) noted that the rationale of this definition is that economic agents have incentives to eliminate systematic errors up to the point where the cost of doing so becomes too large. Since the information of prior forecasts and actual outcomes are cheap, systematic errors should be eliminated.

Nearly all macroeconomic forecast evaluations investigate one or more of the above dimensions when analyzing forecasts. All forecast evaluations referred to below use at least one of those dimensions in their analysis.

[2] It is important to note that the error measures are not equivalent to the forecast error, which is defined as the actual outcome less the corresponding forecast.

2.2 Previous Research

Generally, the studies that are surveyed here have relied on OECD and/or IMF forecast data. The reason that so many studies use data from those institutions is that the two agencies have produced publicly available forecasts for numerous countries for over 20 years. Ash Smyth and Heravi (1990, 1998), Holden and Peel (1987) and OECD (2000) have conducted the most extensive evaluations of the OECD forecasts, while IMF (1988; 1996; 1993) have performed the most comprehensive evaluations of IMF forecasts. In addition, others have evaluated forecasting performance of the two institutions in comparative studies. Examples are Batchelor (2001), Pons (2000), Öller and Barot (2000) and Sveriges Riksbank (2000).

OECD forecast evaluations. Ash et al. (1990) conducted the most comprehensive study of OECD forecasts. Their study evaluated the OECD's forecasting performance from 1967 to 1987 for the G7 countries with respect to twelve macroeconomic variables, including those examined in the present study. The forecast horizons considered in their study were 0.5-year forecasts, 1.0-year forecasts and 1.5-year forecasts.

Ash et al. (1998) used the same data set as Ash et al. (1990), but the two studies are distinct in that the 1990 study conducted a quantitative analysis, while the 1998 study performed a qualitative, or a directional, analysis.

Another extensive review of the OECD's forecasting record was conducted by Holden and Peel (1987). Their study focused on 1.0-year forecasts for 24 countries of eight variables (including the variables examined in this thesis) published 1976 to 1984.

OECD (2000) used a sample set that varied somewhat depending on country and variable. The study generally used a data set that stretched from the early 1970's to the late 1990's. The study evaluated the OECD's 0.5-year and 1.0-year forecasts of GDP, inflation and current account balances for the G7 countries.

Vuchelen and Gutierrez (2005) studied the 2.0-year forecasts of GDP for the G7 countries for the period 1987 to 2002. This has been the only study that analyzed the 2.0-year forecasts.

IMF forecast evaluations. IMF (1988; 1993; 1996) was concerned with the IMF's forecasting record from the early 1970's onwards. The main interest was the short-term forecasting record (forecasts with horizons of less than a year) for the G7 countries for GDP, inflation and balance of payments. IMF (1996) also studied the forecasting record for developing countries and found that these forecasts were worse than the forecasts made for the G7 countries.

Comparative forecast evaluations. Some studies, which did not primarily focus on evaluating an institution per se, have provided extensive forecast evaluations. A prime example is Batchelor (2001) who compared consensus forecasts to those of the OECD and the IMF. The sample period was relatively short (1990 to 1996) and the analysis concentrated on forecasts of GDP, private consumption and fixed investment for the G7 countries. Öller and Barot (2000) conducted a study where 1.0-year forecasts of GDP for 13 individual countries were analyzed. Here, the OECD, as well as national forecasting institutions, were reviewed. Pons (2000) compared the IMF to the OECD using 0.5-year and 1.0-year GDP forecasts for the G7 countries for the period 1971 to 1995.

2.3 Findings from Previous Research

The main findings from the aforementioned studies are largely the same: (i) Forecast accuracy decreases with the length of the forecast horizon, (ii) forecasts issued by experts are generally better than forecasts generated from a model that predicts no change from the last period and (iii) forecast directional change is associated with the actual directional change. However, there are somewhat disparate findings on the rationality of forecasts. Holden and Peel (1987) found little evidence of irrationality, whereas OECD (2000) found some evidence. Moreover, Ash et al. (1990) found that 50 percent of the forecast series under examination did not pass all of the rationality tests that were used in their study.

2.4 Our Study in Relation to Previous Research

This thesis differs from earlier research in three dimensions. First, we evaluate across more forecast horizons than other studies have done in the past. The past studies, that we are aware of, investigated at most three forecasting horizons whereas we look at five horizons. Furthermore, the 2.0-year forecasts for private consumption, government consumption and fixed investment have not been evaluated before (cf. Vuchelen & Gutierrez, 2005). Additionally, the studies that have evaluated such variables (Holden & Peel, 1987 & Ash et al., 1990) did not analyze the near estimate horizon. The present study evaluates both 2.0-year forecasts and estimates for private consumption, government consumption and fixed investment. Second, this thesis covers more countries than any previous study that has used pooled data. Those, with the exception of Öller and Barot (2001), focused solely on the G7 countries. Öller and Barot (2001) were only concerned with 1.0-year forecasts of GDP. Third, the data set used in the present study is recent compared to similar studies that evaluated both GDP and the final domestic demand components. Those studies, Holden and Peel (1987) and Ash, et al. (1990), used datasets that ended in 1984 and 1987 respectively while our data set ends in 2004.

The past research referred to above has been robust in both size and value and it has been the guiding force to the design of this thesis. Below is a general list of issues that were important to the design of this study and has to be considered when forecast evaluations are compared.

- Country aggregation: Results obtained using pooled data should be compared to previous research conducted on pooled data of the same or similar countries. The reason for this is that macroeconomic forecast errors may be specific to each country (cf. Pons, 2000; Ash, et al., 1990; Holden & Peel, 1987). Consequently, idiosyncratic country factors make it difficult to compare individual country forecasts to pooled forecasts. Most macroeconomic forecast evaluations have put forth results on an individual country basis.

- Forecast horizon: Matching forecasting horizons are preferable otherwise a direct comparison may be unbalanced. For example, it is easier to make accurate short-term forecasts than long-term forecasts.

- Vintage of outcome data: The time passed between the outcome date and the published outcome (the vintage) should be similar when two studies are compared, otherwise a direct comparison may be misleading. Later vintages often contain statistical revisions that are not included in the outcome data published earlier. An example is found in Fildes and Stekler (2002), who showed that the average forecast accuracy for 1.0-year forecasts of US GDP for the period 1971 to 1976 differed with more than 0.3 percentage points depending on the vintage of the outcome measure.

- Statistical methods: Statistical techniques should be similar or results may not be comparable. For example, the mean error and the mean absolute error are two measures that are commonly used to evaluate quantitative accuracy. However, they cannot be compared as they analyze two different aspects of accuracy.[3]

All studies presented in sections 2.2 use similar vintages for the outcome data. Furthermore, they use similar forecast horizons and the statistical methods that are applied are largely the same. However, relatively few of the studies analyzed pooled data. Consequently, it is not always possible to compare the results of this thesis directly to previous studies.

[3] The mean of a group of numbers and the mean of the absolute values of the same group of numbers are potentially very dissimilar. For example, the mean of 100 and -100 is zero whereas the mean of the absolutes of 100 and -100 is 100.

3. Psychology of Forecasting

This section begins with a discussion of common cognitive biases in forecasting. Subsequently, the biases analyzed in this thesis are elaborated upon – these are: Optimism bias, temporal bias, anchoring bias, and representativeness bias. The section concludes with a description of how the forecasts may come to be influenced by cognitive biases. To clarify the descriptions, an example forecaster, Sylvia Porter, is used for illustration.

3.1 Common Forecasting Biases

A large body of research has shown that individuals are not rational in the traditional *Homo Economicus* sense. Often when solving problems humans do not analyze all available information and are affected by the framing of the problems (Gilovich, Griffin & Kahneman, 2002). Instead, due to cognitive limitations, individuals often rely on heuristics (rules of thumb), to construct judgments and make decisions (Simon, 1955). Heuristics can be unreliable and may produce systematically biased judgments (Gilovich et al. 2002). Naturally, this is also true for forecasting. Numerous studies have shown that forecasts exhibit systematic biases (Bolger & Harvey, 1998).

Many studies concerned with systematic biases in forecasting have been based on laboratory experiments that analyzed the extrapolation of time series (cf. Bolger & Harvery, 1998). Most of the non-laboratory studies have been conducted on the financial markets, investigating implicitly or explicitly the biases of financial market forecasts (e.g. Abarbanell & Bernard, 1992; Choi & Ziebart, 2004). In a similar manner, the behavior of the entire financial market has also been studied (Bernard & Thomas 1989; DeBondt & Thaler 1985; 1987).

A handful of studies also examined real world forecasts. Bromiley (1987) analyzed the existence of an anchoring bias in forecasts produced for internal organizational use. The study focused on four types of forecasts: A plant-to-headquarters' operational reports, corporate financial planning, corporate strategic planning, and government budget forecasts. The analysis revealed that only the plant's operational reporting to headquarters exhibited significant anchoring bias.

3.2 Macroeconomic Forecasting Biases

Montgomery (2000) and Montgomery and Törngren (2003) analyzed biases in institutional macroeconomic forecasts. Montgomery (2000) analyzed Swedish governmental GDP forecasts from 1970 to 1998. The forecasts were for horizons of one year and two years. The study found indications of optimism bias, anchoring bias and availability bias.[4] However, the findings were not tested statistically. Montgomery and Törngren (2003) analyzed the existence of psychological biases in GDP forecasts produced by the OECD. The biases they tested for were: (i) Anchoring bias, (ii) optimism bias, (iii) availability bias (too strong impact of current growth data), (iv) base rate neglect (neglect of historical growth data for particular countries), and (iv) regression fallacy (too wide range between highest and lowest forecasts). Support was found for availability bias, base-rate neglect, and regression fallacy.

Studies focusing on cognitive biases, particularly those who have looked at macroeconomic forecasting, have led to the design of the part of the study that relates to its second purpose, namely to determine if macroeconomic forecasts show tendencies of being afflicted by cognitive biases. There are four main differences between this study and earlier research on cognitive bias in macroeconomic forecasts. Firstly, not only do we study GDP, but we also examine three of its expenditure components. Secondly, the dataset of 23 countries is larger than in previous studies. Montgomery (2003) studied a single country while Montgomery and Törngren (2003) looked at thirteen countries. The advantage with a larger data set is that is leads to more statistically secure findings (cf. Rosenthal & Dimatteo, 2001). Thirdly, the present study investigates five forecast

[4] Montgomery(2000) defined availability bias in exactly the same way as we defined anchoring on the last known actual outcome (cf. Section 3.5). Montgomery (2000) noted that the two interpretations are equally valid. He defines anchoring bias as insufficient adjustment from a typical value of growth for a country. We do not use that definition since it is not clear how that typical value should be estimated.

horizons while Montgomery (2000) analyzed two horizons and Törngren and Montgomery (2003) three horizon. Fourthly, the representativeness bias and the temporal differences bias is investigated in the present study. Neither Montogmery (2000) nor Montgomery and Törngren (2003) investigated those biases.

This study considers four cognitive biases: Optimism bias, temporal bias, anchoring bias, and representativeness bias. A major reason for choosing these particular biases is that their traces are detectable in pooled post-mortem data. From the biases analyzed by Montgomery (2000) and Törngren and Montgomery (2003), we have chosen not to examine availability bias, base rate neglect and regression fallacy. Availability bias was excluded mainly due to its likeness to the anchoring bias. Regression fallacy and base rate neglect were also excluded, mainly due to the need for limiting the scope of the thesis.

3.3 Optimism Bias

Many studies have shown that individuals are systematically too optimistic or too pessimistic. For example, in an experiment relevant to this thesis, Buehler, Griffin, and Ross (1994) showed that people often believed that it will take less time for them to complete projects than it actually does. Morrison, Ager and Wilock (1999) presented the results of two surveys on the perception of the likelihood of malaria infection. The study was conducted in villages of the African country Malawi and showed that participants were systematically too pessimistic when they considered their personal likelihood of infection.

Within the economic forecasting arena, there has been considerable research conducted on systematic optimism and pessimism. The behavior of security analysts is one area that has been thoroughly investigated. On the one hand, De Bondt and Thaler (1990) found that security analysts' earnings forecasts were systematically too optimistic; other studies confirm this with similar conclusions (e.g. Abarbanell & Bernard (1992); Malloy, 2005). Hong and Kubik (2003) found that a potential reason for such overestimation was that, analysts who, controlling for accuracy, were optimistic relative to the consensus were more likely to be promoted. On the other hand, there is some support for a pessimism bias as well. Brown (1996) found that analysts were too pessimistic in the booming years running up to 1996. Also, Brown (2001) observed that some forecasters were systematically too pessimistic. Regarding the cognitive bias studies, Montgomery (2000) noted that there might be an optimist bias in Swedish GDP forecasts. However, Törngren and Montogomery (2003) found no such bias. Most studies defined optimism bias as a systematic over/under prediction of an outcome; this definition is adopted here as well. Note that for private consumption, fixed investment and GDP, exceedingly high forecasts are a sign of an optimism bias. The reverse, however, is true for government consumption, as the general view within economics teaches us that government spending in most industrialized countries is generally too high.

Evaluations of macroeconomic forecasts often test for overall bias. In such evaluations, bias is defined as systematic over or under estimation of a variable. In other words, the forecasts are said to suffer from a bias if the mean forecast error is significantly different from zero. This definition corresponds to the definition of optimism and pessimism bias used in this study and is generally tested using a Student's t-test (Fildes & Stekler, 2002). None of the forecast evaluations referred to in Section 2.2 found a systematic bias in the GDP forecasts that could not be attributed to specific events. The same was true for private consumption and fixed investment in Batchelor (2001) and Ash et al. (1990). Ash et al. (1990), however, concluded that the government consumption forecasts were biased, as the forecasts were systematically too low.

As there is no consensus on the existence of systematic optimism or pessimism for either financial or macroeconomic forecasts, we do not form any hypotheses about optimism bias. Instead, we choose to describe and analyze the optimism bias in macroeconomic forecasts.

3.4 Temporal Bias

An important aspect of optimism is that it varies over time. For example, Andersson (2005) presented a study in which students were overly optimistic about which grade they would receive in the beginning of the course, but became overly pessimistic by the end of the course.

The same tendency has also been found in economic forecasting. Choi and Ziebart (2004) observed that the degree of optimism in management earnings forecasts varied over time. Forecasts for longer horizons were generally more optimistic than forecasts for shorter horizons. We are not aware of any studies on macroeconomic forecasts that have tested this tendency. Nevertheless, since financial forecasters were more optimistic when the forecast horizons were longer, we hypothesize that macroeconomic forecasters tend to behave similarly. For private consumption, fixed investment and gross domestic product a temporal change in optimism implies that long-term predictions are forecast higher than short-term prediction. Since low government spending is generally viewed as beneficial, the opposite relationship is hypothesized for government consumption, i.e. long-term predictions tend to be forecast lower than short-term predictions. Consequently, we formulate the following hypotheses

H1: For private consumption, forecasts with long horizons are higher than forecasts with short horizons.

H2: For government consumption, forecasts with long horizons are lower than forecasts with short horizons.

H3: For fixed investment, forecasts with long horizons are higher than forecasts with short horizons.

H4: For gross domestic product, forecasts with long horizons are higher than forecasts with short horizons.

3.5 Anchoring Bias

Several studies have shown that individuals analyze situations from a particular reference point. For example, Davis, Hoch and Ragsdalse (1986) studied 220 husband and wives to probe their thoughts on their spouse's preferences. The study showed that both husband and wives leaned heavily towards their own preferences when they predicted their spouse's preferences. Tversky and Kahneman (1974) labeled reliance on reference points and insufficient adjustments from those points *anchoring and adjustment bias*. A starting point, or an anchoring point, which influences behavior, need not be relevant to the problem at hand. An illustrative experiment on the power of anchoring was presented in Tversky and Kahneman (1974). Participants were told a random number between 1 and 100. Thereafter, they were asked to estimate the percentage of African countries belonging to the United Nations. On average, those who were given the number ten answered 25%, while those who were presented with the number 65 answered 65%.

Anchoring has also been shown to influence economic forecasting. For example, Northcraft and Neale (1987) conducted an experiment in which real estate agents were asked to estimate the value of a house. The true value was $74,900. When the real estate agents were subjected to four price anchors, ranging from $65,900 to $83,900, their valuations were significantly affected by each anchor. In another study, Russo and Shoemaker (1989) conducted an experiment where financial professionals where asked to forecast upcoming interest rates. When they were provided with an anchor, the estimates were significantly affected. Yet, some studies have not found support for an anchoring bias in economic forecasting. Lawrence and O'Connor (1995) asked participants to extrapolate real business related time-series. Instead of anchoring, they found the opposite of anchoring, subjects made adjustments that were too large.

It is puzzling that studies have found contradicting evidence on anchoring bias. One possible explanation for this is that most studies have not been real world experiments, but have been conducted in a laboratory. It seems that the studies with more real world context among the laboratory experiments exhibited more anchoring bias. This is far from definitive as Bromiley (1987), who conducted a real world study concluded that there was only weak evidence of anchoring. Though little has been done to evaluate anchoring and adjustment in macroeconomic forecasts, what has been done on the bias has provided contradictory evidence. As noted above, Montgomery (2000) suggested that an anchoring bias might exist in Swedish GDP forecasts, whereas Montgomery and Törngren (2003) did not find any anchoring effects in OECD forecasts. In the studies above, anchoring and adjustment is generally defined as insufficient adjustments from an anchor point. We also adopt this definition.

To test practically for anchoring in OECD forecasts, plausible anchors must be identified. Since the anchors do not need to relate to the problem at hand, there is theoretically a large number of anchors forecasters may affected by. However, due to the amount of countries, periods and horizons that are evaluated, an anchor that is not directly linked to the forecasting process would be highly unlikely. Richardson (1998) stated that the OECD's INTERLINK model "…has tended to rely on adaptive rather than forward-looking mechanisms." Consequently, the initial forecast for a given year (the 2.0-year forecast) might serve as an anchor. However, we are not aware of any studies that have tested for such an anchor empirically. Another possible anchor candidate is the last known outcome at the time the forecast was made, as often the past is the best predictor of the future. In contrast to the initial forecast, the last known actual outcome has been used as an anchor in previous studies (e.g. Bromiley, 1987; Lawrence and O'Connor, 1995).

As a consensus on the anchoring bias is lacking, we abstain from forming hypothesis on the bias. Instead, we choose to describe and analyze the data with respect to anchoring in order to assess whether it exists in macroeconomic forecasting.

3.6 Representativeness Bias

Hot hand bias and gambler's fallacy bias are two cognitive biases that have been discussed in behavioral decision-making research (e.g. Tversky & Kahnemann, 1971; Ayton & Fisher 2004). Hot hand bias occurs when observed trends are projected to continue more than is really warranted (Sundali & Croson, 2006) [5] An example of the hot hand bias is the belief that a basketball player who has scored several shots in a row is more likely to score on his next shot. Gilovich, Vallone and Tversky (1985) showed that there is no sequential dependence between successive scoring attempts for basketball players. The fact that a player has shot a successful string of baskets does not increase the probability that the player will score with his next shot. In contrast to hot hand bias, gambler's fallacy bias occurs when turnarounds of trends are projected more frequently than is really warranted (Sundali & Croson, 2006). An investor who buys stock in the belief that the stock's trend will turn upwards only because the stock has had a long spell of declines is an example of someone who is subject to the gambler's fallacy bias (Johnsson, Tellis & Macinnis 2005).

There has been some research examining the hot hand bias and gambler's fallacy bias in the context of economic forecasting. Decisions to buy or sell securities imply a belief about the future, which is based on a forecast. Consequently, through the decisions taken on markets, forecasts can be inferred. In a laboratory experiment, Johnson, et al. (2005) found some evidence of the existence of a hot hand bias when participants sold stocks. Using non-laboratory data, Baquero and Verbeek (2005) also showed that sophisticated investors exhibited a hot hand bias. Sirri and Tufano (1998) found increased flows into mutual funds where there was exceptional but statistically short-lived performance. Such behavior is consistent with the hot hand bias. In addition, De Bondt (1991) analyzed data

[5] Sundali and Croson (2006) make an interesting distinction with regards to the hot hand bias. They divide hot hand beliefs into hot hand bias and hot outcome bias. Hot hand bias refers to a person's belief about individuals capabilities based on recent performance, whereas hot outcome refers to the general belief that a series is more positively autocorrelated than it actually is. Following this definition, we evaluate hot outcome bias. Nevertheless, the term hot hand bias is used as it is much more frequently used in the literature.

for about 5400 individual forecasts of the S&P index. He found that forecasters often overreacted to movements in stock prices which is a behavior akin to a hot hand bias.

There is also some support for a gambler's fallacy bias in investor behavior. Shefrin and Statman (1985) found that consumers held on to losing stocks too long and sold winning stocks too early, which is a type of behavior that is consistent with gambler's fallacy bias.

In line with previous studies, we define hot hand bias as a tendency to project continuation of trends into the future more than is really warranted. We define gambler's fallacy bias as a tendency to project turnarounds of trends more than is really warranted. For example, the existence of a hot hand bias in private consumption would lead to too positive forecasts of private consumption after a period of accelerating growth in that variable. Accordingly, the gambler's fallacy bias implies a pattern that is the exact opposite: On average, forecasts of private consumption would be too negative after a period of accelerating growth in the variable in question.

To the best of our knowledge, no studies have been conducted on the representativeness bias in macroeconomic forecasting. Montgomery (2000) and Törngren and Montgomery (2003) did not evaluate the existence of the representativeness bias. A possible reason for this is that the data sets of these studies were relatively small, leading to few observed trends, which makes inferences about the existence of a representativeness bias highly uncertain.

As noted above, the previous research conducted on representativeness bias in economic forecasting disagrees on the existence of a representativeness bias. Consequently, we do not form any hypotheses about it. Instead, we choose to describe and analyze the possible existence of a hot hand or gambler's fallacy bias based the results we find.

3.7 Possible Bias Sources

Armstrong (1985) divided the forecasting process into five stages: (i) Formulation of the forecasting problem, (ii) choice of method, (iii) application of method, (iv) comparison and combination of forecasts, and (v) judgmental adjustment of forecasts. In each stage of the process, judgment is employed. As human judgment is often fallible, particularly when using heuristics, it may produce systematic bias. Bolger and Harvey (1998) elaborated on how the use of human judgment in each of these stages can lead to systematically biased forecasts. The following examples are inferred from Bolger and Harvey (1998).

The formulation phase. A common pitfall that might lead to systematic errors in the first stage of the forecasting process is misattribution of the problem. Misattribution of the problem means that forecasters systematically frame the problem in a certain way, while neglecting other important aspects. If our example forecaster Ms. Sylvia Porter believes that government consumption is something controlled solely by the government, she would miss other important aspects of government spending, such as the fact that civil servants affect government spending through wage negotiations. Such a misattribution of the problem may lead Sylvia to issue forecasts with a systematic optimism bias.

The choice of method. The second stage of the forecasting process may also introduce systematic biases in the forecast. Perhaps Sylvia, in constructing her forecasts, adopts a forecasting method that uses autoregressive forecasts, i.e. a method which takes its starting point in the actual outcome of previous period. If such a model places a too heavy emphasis on the last actual outcome, the forecasts may not be adjusted sufficiently from that value. Consequently, the resulting forecasts may show an anchoring bias.

The application of method. A problem that can arise in the third stage of the forecasting process stems from how information search and collection is conducted. If Sylvia consistently overlooks consumer confidence indices, she may be missing key facts. Since consumer consensus indices are early indicators of turning points in the economy, by neglecting those, she might too often predict that trends will continue, i.e. her forecasts could fall prey to a hot hand bias.

The comparison and combination of forecasts. The fourth stage of the forecasting process can also introduce systematic biases. Professional forecasters such as Sylvia often use several models in order to produce forecasts. The results from these models are then combined to generate a final forecast. If in one model, in which Sylvia particularly trusts, is systematically too negative, she could give this model too much weight in her final forecast. Consequently, her forecast could be subject to pessimism bias.

The judgmental adjustment of forecasts. Forecasts generated by econometrical and mathematical models are only valid as long as tomorrow is similar to today. Models cannot predict alterations of how the economy functions. This can only be done by human judgment. Since the world is constantly changing, human judgment can help in guiding these technical models. If Sylvia's own mental model of how the world works is somewhat flawed, then this type of judgmental adjustment may lead to biased forecasts. Perhaps Sylvia believes that the economy is more likely experience turnarounds than it actually is. Such a belief would introduce a gambler's fallacy bias to the forecasts.

The five stages above highlight the fact that there are several possible sources of bias in the forecasting process. The examples are simply there to provide an illustration of the general idea, giving the reader more concrete applications. Note that these sources may produce biases that counteract each other. The examples above show that the third stage might introduce gambler's fallacy bias while the fifth stage may introduce hot hand bias. In addition to this, different countries may be analyzed individually and subsequently be subject to different biases that might counteract each other in a pooled study. Thus, the question of bias existence and their origin is complicated by the way the forecasts are formed.

4. Data

This section describes the variables that are analyzed in this study and the overall structure of the OECD. Issues that have arisen in terms of stability and validity of the dataset are also put forth. Additionally, we recount the technicalities that underlie each issue of the *Economic Outlook* and discuss the usage of pooled data. Furthermore, we explain our interpretation of the different forecasting horizons that are formed with each publication.

4.1 Variable Description

The OECD defines the variables in accordance to the United Nations System of National Accounts. *Private consumption* is defined as: "[C]onsumption expenditure of households and private non-profit institutions serving households (SNA Para.1.49)." *Government consumption* "…consists of expenditure, including imputed expenditure, incurred by general government on both individual consumption goods and services and collective consumption services (SNA Para. 9.94)." *Gross domestic product* is "…an aggregate measure of production equal to the sum of the gross values added of all resident institutional units engaged in production (plus any taxes, and minus any subsidies, on products not included in the value of their outputs). The sum of the final uses of goods and services (all uses except intermediate consumption) measured in purchasers' prices, less the value of imports of goods and services, or the sum of primary incomes distributed by resident producer units (SNA Para. 1.128 and Paras. 2.173-2.174)." *Gross fixed capital formation* "…is measured by the total value of a producer's acquisitions, less disposals, of fixed assets during the accounting period plus certain additions to the value of non-produced assets (such as subsoil assets or major improvements in the quantity, quality or productivity of land) realized by the productive activity of institutional units (SNA Para.10.33 and Paras. 10.51 [10.26])." *Gross fixed investment* is "often described as…gross fixed capital formation (SNA Paras. 1.48 -1.57), and is generally used to analyze investment trends when using time series data.

4.2 OECD *Economic Outlook*

The OECD was founded in 1948 by 20 countries to help reconstruct the imbalances following World War II. Currently, the OECD consists of 30 members. Since July of 1967, the organization has been publishing its *Economic Outlook* bi-annually in May/June and in November/December.[6] The *Economic Outlook*s prior to 1985 delved into the issues plaguing the international markets, analyzing each upswing or downswing and their possible sources, while at the same time issuing predictions for where the global economy would be turning. Also in these editions were detailed statistics for the G7 countries: Canada, France, Germany, Italy, Japan, the UK, and the US. Substantially less detailed forecasts were also published for other member countries, but in June 1985, the forecasts of those other member countries became more detailed.

The forecast data for the period 1985 to 2004 for 23 of the OECD member countries is the focal point of this study. Data prior to 1998 was obtained from the only medium available at that time: In-print editions. Subsequent editions have the *Economic Outlook* available in an electronic format. Thus, the data has been manually transformed into excel format. In sum, 11040 data points (20 years × 23 countries × 4 variables × 6 horizons) has been used. To our knowledge, this dataset is unique; however, the OECD could probably replicate it easily. All measures to ensure a valid dataset have been undertaken throughout the data entry process. Also, in order to achieve a representative sample, some adjustments have had to be made to the dataset. In 1985, 24 member countries were included with detailed statistics, but Iceland was excluded from the study due to missing forecasts for various years and at numerous horizons. In addition, in order to guarantee a stable time-series study, all variables were examined to be made as stable as possible, and as a result, two corrections had to be made. First, before June 1996 the OECD examined aggregate private fixed investment for the US. All public fixed investment was included in the measure of government consumption. Thereafter, the focus switched to gross fixed investment, the standard measure for the other G7 countries. To correct for this, a measure of public gross fixed investment was subtracted from the aggregate gross fixed

[6] The month in which the Economic Outlook is published varies from year to year, but it is most often published in May and December.

investment, providing a measure of private fixed investment. For the other G7 countries the focus has been on developments in gross fixed investment and for the other 16 countries gross fixed capital formation has been used. Section 4.1 shows that the difference between the two measures of investment is small, which is why we have analyzed them as one variable named fixed investment. The second correction concerns Germany and its reunification in October of 1990. Prior to 1990, the OECD made projections for West Germany. After the reunification, these projections were for the united East and West Germany. This had an effect on the benchmark actual outcomes that are published in the May/June issue of the year directly following that of the forecast. Actual outcomes for 1990 and 1991 published were for West Germany, but the forecasts made for the two years after December 1990 were made for East and West Germany combined. It is unreasonable to assess the performance of the OECD forecasts when the benchmark, the actual outcome, is not the valid outcome for the variable and country they were forecasting. For this, we have chosen to exclude the forecasts and actual outcomes for Germany in 1990 and 1991 in the analysis.

4.3 Underlying Technical Specifications

Prior to 1996, the OECD described its sources and methods in the Technical Annex of the *Economic Outlook*. After June 1996, the "Sources and Methods" have been published as working papers. These more recent versions include a description of the OECD's INTERLINK system and the technical assumptions under which the projections are produced. INTERLINK is the multi-country model that is used for projections and is based on the following "exogenous" assumptions:

- Nominal exchange rates are set at a fixed level from a pre-specified cut-off date across the projection periods (this does not include those countries that have stated or effective policies that adjust their exchange rates).

- Fiscal policy assumptions are based on the officially declared policies, i.e. budgetary statements and stated practices for public sector spending and revenue estimates.

- Monetary policy assumptions and the resulting interest rates are based on the announced targets for the monetary aggregates and inflation.

- The price of crude oil is generally assumed to increase with inflation in the long-term. The OECD also seeks advice from the International Energy Agency regarding the short-term forecasts. The same method is applied to other commodity prices.

The INTERLINK system is one part of the methodology behind the OECD forecasts another part is the explicit use of judgment (OECD 1999). (OECD, 1999) stated that judgment is applied "…first, in assessing initial individual country conditions and, second, in evaluating the results generated by the model (p. 2)."

All forecasts and outcomes are measured as percentage changes. This is the standard for forecast evaluations (Fildes & Stekler, 2002). Zarnowitz (1979) stated four reasons for this practice: (i) The percentage changes are less dependent on levels than absolute changes are, (ii) forecasts of percentage changes are less inclined to be affected by data revisions, (iii) widely used quantitative measures of performance are more appropriate to use on percentage changes than on level changes, and (iv) it is usually the growth rate in economic aggregates that are of the main interest to policy makers and the public.

4.4 Forecast Horizon Formation

As the *Economic Outlook* is published twice a year, each forecast is produced in terms of a half-year interval. In May/June, the OECD prints projections for the current and following year. The organization also publishes figures for the outcome of the previous year. Thus, we are able to extract from each May/June issue a 0.5-year forecast and a 1.5-year forecast, as well as the actual realized outcome the previous year. In each November/December edition, the OECD puts forth an estimate of that year and a projection for next year and two years ahead. Extracted then from the November/December editions are the estimate, the 1.0-year, and the 2.0-year forecast. Errors are estimated by subtracting the forecast from the actual outcome. Thus, a too high forecast leads to a negative value and a too low forecast to a positive value.

We have chosen the growth figure that is published in the May/June edition of the previous year as the benchmark for the forecasts. Most studies use this benchmark (Batchelor, 2001). Furthermore, this is the vintage that the OECD forecasters themselves are most likely aiming to predict (Ash et al., 1990).

4.5 Pooled Data – A Clarification

We use pooled data in the present study. By using pooled data, the country aspects of the forecasts are not the focus, but instead we concentrate on combinations of forecast horizons and variables. If each country were studied in isolation, there would have been 20 observations for each combination of country, forecast horizon and variable. By pooling the forecasts we get 460 (20×23) observations instead. There is a case for using pooled data in that pooling the data makes common traits among the countries more easily discernible. Another advantage is that the overall sample is made larger and therefore increases the power of the statistical tests (Gujarati, 1995). The drawback is that country specific issues disappear.

One way to think of pooled data is to look at the forecasting record of a particular country as an individual study of general macroeconomic forecasting. From that point of view, the pooling of countries can be seen as a meta-analysis (a combination of the results from studies that address the same hypotheses) where each country is considered as a separate study of the forecasting record (cf. Rosenthal & DiMatteo, 2001). However, instead of conducting the study on one country at a time and subsequently aggregating the results, which is the usual way to conduct meta-analysis, the countries are aggregated before the statistical analysis is carried out. By pooling the countries and in that sense conducting a meta-analysis, the large picture becomes clear. Consequently, by conducting such an investigation, we perform a study on the macroeconomic forecasting record in general rather than the forecasting record of individual countries. In other words, this allows us to see the forest, albeit at the expense of the trees. Furthermore, common traits among countries that are not statistically significant on the individual country level can become significant when the countries are pooled. Rosenthal and DiMatteo (2001) exemplified this by stating that ten results at ($p=0.10$), of which none is sufficient evidence against a null-hypothesis at the five percent level by itself, provides very strong evidence against a null-hypothesis when they are combined ($p=0.000025$).

5. Methodology and Analysis: Forecast Evaluation

The first purpose of the thesis is to evaluate the quality of OECD forecasts for GDP and its expenditure components. This section describes the methodology used to fulfill this purpose and presents the results of the corresponding analysis. Following the dimensions used by Fildes and Stekler (2002) (cf. Section 2.1) the text is divided into four sub-segments: Quantitative accuracy, directional accuracy, naïve/no-change comparisons, and rationality of forecasts. Methods and results are separated in all sub-segments so as to distinguish clearly between the two concepts. Table 2 provides a concise overview of the measurements and tests used in this section.

Table 2. Overview of measurements and tests used in Section 5

Name	*Purpose*	*Test statistic/ Measure*
Quantitative error measures	Measure the degree of quantitative errors using mean absolute error	Mean absolute error
Directional accuracy	Measure the directional accuracy	Chi-Square
Naïve/No change comparison	Analyze if OECD forecasts are better than naïve forecasts	Theil's U
Rationality tests	Test the rationality, i.e. if the forecasts are free from systematic errors	Regression, t-test

5.1 Quantitative Measures

Method. Several measures that describe the accuracy of forecasts (cf. section 2.1). We choose to present the results for the *mean absolute error (MAE*). *MAE* is the mean absolute difference of the actual outcomes and forecasts averaged over the period:

$$MAE=\frac{1}{n}\sum\left|A_t - P_{h,t}\right| \qquad (1)$$

In Equation 1, $P_{h,t}$ is the forecast and A_t is the corresponding actual outcome. The formula for the measure is presented in Appendix I. The *MAE* is put forth because it correlates with two other important measures: The *mean squared error* and the *root mean squared error*.[7] Furthermore, it is probably the most easily understood of the different error measures.

Results. Table 3 shows the *MAEs* and their corresponding standard deviations. Fixed investment exhibits the largest *MAEs* and standard deviations, whereas the other variables have relatively similar *MAEs* for the same forecast horizons. Table 3 also shows that *MAE* increases monotonically with the length of the forecast horizon for all variables. The results are highlighted in Figure 1, which is based on the same data as Table 3, but presents the results graphically. The *MAEs* for individual countries can be found in Appendix II.

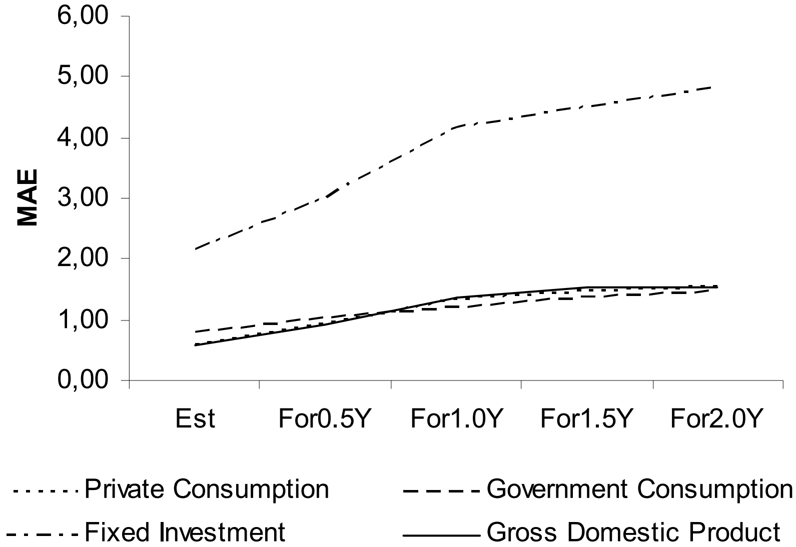
Table 3. Mean absolute error for all combinations of forecast horizons and variables

	ForEst	*For0.5Y*	*For1.0Y*	*For1.5Y*	*For2.0Y*
Private Consumption	0.59	0.91	1.33	1.47	1.53
	(0.85)	(1.10)	(1.52)	(1.54)	(1.56)
Government Consumption	0.78	1.01	1.19	1.37	1.46
	(0.85)	(1.15)	(1.34)	(1.43)	(1.43)
Fixed Investment	2.15	3.00	4.17	4.49	4.81
	(3.39)	(2.71)	(4.32)	(4.27)	(4.72)
Gross Domestic Product	0.57	0.93	1.35	1.53	1.55
	(0.75)	(0.99)	(1.35)	(1.43)	(1.48)

Note: The standard deviations of the mean absolute errors are presented within brackets.

[7] The other error measures referred to in Section 2.1 have been computed as well. The range of correlation between the error measures are 0.92-0.98 for private consumption, 0.96-0.99 for government consumption, 0.93-0.99 for GDP, but only 0.39-0-89 for fixed investment. They, as well as error measures for each country, can be obtained from the authors.

Figure 1. Mean absolute error for all combinations of forecast horizons and variables



5.2 Directional Analysis

Method. We mirror the approach taken by Pons (2000) and define a forecast as being directionally correct if the forecast direction, relative to the last known actual outcome at the time of the forecast, is correct. If it is not, then the forecast is deemed as directionally incorrect. Consequently, there are four possible combinations of forecasts and outcomes: Correct positive forecasts, correct negative forecasts, false positive forecasts, and false negative forecasts. Those four combinations can be presented in a contingency table such as Table 4. Pons (2000) used a chi-squared test of independence to test for association between the forecast and actual directional change. The null-hypothesis is that there is no association between the forecast and actual directional change. If the observed chi-squared value is larger than the critical chi-squared value (one degree of freedom), the directional forecasts contain information about the future actual direction of change.[8] The observed chi-squared values are obtained from Equation 2. Definitions of the variables in Equation 2 are explained by Table 4.

[8] It must be stressed that the cells of the leading diagonal (n_{00} and n_{11}) must have relatively more entries than the other diagonal. If the other diagonal have significantly more entries, the null-hypothesis of no association is rejected, not because the forecast and actual change are positively associated but because they are negatively associated.

Thus, if the observed chi-squared values obtained are significant, there is an association between forecast and actual directional change.

Table 4. Example contingency table for directional analysis

	Forecast		
	F>0	F<0	Sub Total
Outcome			
R>0	n_{00}	n_{01}	n_0
R<0	n_{10}	n_{11}	n_1
Subtotal	n_0	n_1	N

Note: The rows represent realized directional outcome and the columns represent forecast directional change.

$$\hat{\chi}^2 = \sum_{i=0}^{1}\sum_{j=0}^{1}\frac{\left[n_{ij} - n_i n_j / N\right]^2}{n_i n_j} \qquad (2)$$

Results. Table 5 shows that the observed values of the chi-squared tests are all significant at ($p<0.001$). Consequently, there is a highly significant association between the forecast and actual directional change for all combinations of variables and forecast horizons. The observed chi-squared values in Table 5 generally decrease with the length of the forecast horizon. This suggests that the association between forecast directional change and actual directional diminishes as the length of the forecast horizon increases.

Table 5. Results from the chi-squared test for directional accuracy and the corresponding significance for all combinations of forecast horizons and variables

	ForEst	*For0.5Y*	*For1.0Y*	*For1.5Y*	*For2.0Y*
Private Consumption	273.45^c	140.34^c	103.02^c	68.71^c	102.09^c
Government Consumption	148.81^c	89.55^c	142.91^c	81.63^c	55.34^c
Fixed Investment	237.84^c	134.69^c	140.73^c	88.37^c	108.35^c
Gross Domestic Product	287.21^c	171.82^c	96.28^c	44.51^c	79.98^c

Note: The values in the table show the observed chi-squared values obtained from Equation 1. Superscript c indicates (p<0.001).

5.3 Naïve/No-change Comparison

Method. The most common way to make a naïve/no-change comparison is to calculate the Theil's U coefficient (e.g. Fildes & Stekler, 2002; IMF, 1996; Ash et al., 1990). The coefficient compares the RMSE of the expert forecasts to the RMSE of a model that assumes the same outcome as the preceding period, i.e. it is naive (Ash et al., 1990). If the coefficient is less than one, then, on average the forecaster performs better than the naïve model. If, however, the coefficient is greater than one, then, on average, the forecaster performs worse than the naïve model. The formula for the calculation is provided in Appendix I.

One drawback to the Theil's U coefficient is that there is no test for its statistical significance. Consequently, if the coefficient is smaller than one, it is not possible to state that the expert forecasts are significantly better than the naïve forecasts (Fildes & Stekler, 2002). Another critique of the coefficient comes from Batchelor (2001), who claimed that no-change models are inappropriate benchmarks as they do not provide credible predictions for most macroeconomic variables. Nevertheless, we calculate the coefficient, both for the pooled data and for the median of individual country coefficients. The median value is presented in order to simplify comparisons with other studies. The values for the U-coefficient for individual countries are presented in Appendix III.

Results. The values for Theil's U, obtained using pooled data, implies that the OECD performs better than a naïve model for all combinations of forecast horizons and variables. The values are shown in Table 6. However, the medians of the country specific U-coefficients are substantially larger. Nevertheless, except for 1.5-year and 2.0-year forecasts of fixed investment, the coefficients are still considerably smaller than one. Furthermore, Table 6 shows that the median U-coefficients increase with the length of the forecast horizon. This implies that the superiority of the OECD forecasts to naïve forecasts becomes less pronounced as the forecast horizon lengthens.

Table 6. Pooled and median Theil's U coefficient for all combinations of variables and forecast horizons

	U-Est	*U-0.5Y*	*U-1.0Y*	*U-1.5Y*	*U-2.0Y*
Private	0.02	0.03	0.03	0.03	0.03
Consumption	[0.27]	[0.37]	[0.57]	[0.63]	[0.65]
Government	0.02	0.03	0.03	0.03	0.03
Consumption	[0.40]	[0.53]	[0.60]	[0.64]	[0.69]
Fixed	0.03	0.03	0.04	0.04	0.04
Investment	[0.38]	[0.52]	[0.77]	[0.89]	[0.96]
Gross Domestic	0.01	0.02	0.03	0.03	0.03
Product	[0.22]	[0.34]	[0.50]	[0.60]	[0.65]

Note: The table shows pooled U values. The values within square brackets show the median U values.

5.4 Rationality Tests

Method. Recall that we define rational forecasts as forecasts that do not exhibit systematic errors. Forecast evaluations often contain tests of rationality (e.g. Fildes & Stekler, 1990; Batchelor 2001; Ash et al., 1990). The rationality test is based on IMF (1993) and uses two regressions, one that tests for correlation between forecasts and forecast errors and a second that tests for correlation between the forecast errors in subsequent periods (Pons, 2000).[9]

$$ForecastError_{h,t,v} = \alpha_{h,v} + \rho ForecastError_{h,t-1,v} + u_{h,t,v} \qquad (3)$$

$$ForecastError_{h,t,v} = \alpha_{h,v} + \beta Forecast_{h,t,v} + u_{h,t,v} \qquad (4)$$

Pons (2000) noted that the equations provide an interesting dichotomy between different forms of irrationality in forecasting. The coefficient ρ measures the correlation between consecutive forecast errors, while β measures the correlation between forecasts and their corresponding errors. If β is zero and ρ is significantly different from zero, then past errors are repeated. If ρ is zero and β is significantly different from zero, then the forecasts correlate with the forecast errors. If both β and ρ are significantly different from zero, then errors are repeated and correlate with the forecasts. It is sufficient that either ρ or β is significant to reject the notion of rationality.

[9] There are many other potential ways to analyze if the forecasts exhibit systematic errors. One way is to conduct a cognitive bias analysis as we do in Section 6. Nevertheless, the most common way is to perform a t-test.

Results. The results of Equations 3 and 4 are presented in Table 7. The table shows that only the 0.5-year forecast for private consumption passes the rationality test as neither of the null-hypotheses can be rejected. All other combinations of forecast horizons and variables fail at least one of the two tests. Furthermore, all forecast horizons of more than half a year fail the rationality test for correlation between forecast and forecast errors. Generally, the p-values decrease with the length of the forecast horizons. This is highlighted by the 2.0-year forecast horizons, where all coefficients are highly significant ($p<0.001$).

Table 7. Results of the rationality tests and the corresponding significance for all combinations of forecast horizons and variables

	ForEst	*For0.5Y*	*For1.0Y*	*For1.5Y*	*For2.0Y*
Private Consumption	-0.056^{a} [-0.044]	0.069 [0.079]	-0.276^{c} [0.074]	-0.325^{c} $[0.242]^{c}$	-0.424^{c} $[0.322]^{c}$
Government Consumption	-0.063^{a} [0.028]	-0.053 $[-0.106]^{a}$	-0.216^{c} $[-0.094]^{a}$	-0.246^{c} [0.038]	-0.310^{c} $[0.163]^{c}$
Gross Fixed Investment	-0.131^{c} $[-0.124]^{b}$	0.118^{b} $[0.114]^{a}$	-0.278^{c} [-0.022]	-0.198^{a} $[0.204]^{c}$	-0.455^{c} $[0.204]^{c}$
Gross Domestic Product	-0.002 $[0.128]^{c}$	0.114^{b} [0.079]	-0.175^{b} [0.025]	-0.343^{c} $[0.214]^{c}$	-0.436^{c} $[0.295]^{c}$

Note: The values are estimates of ρ obtained from Equation 3. The values within square brackets are estimates of β obtained from Equation 4. Superscript a indicates (p<0.05), b indicates (p<0.01), and c indicates (p<0.001).

6. Methodology and Analysis: Cognitive Biases

The second purpose of the thesis is to determine if the forecasts show signs of being afflicted by cognitive biases. This section describes the methodology used to fulfill this purpose and presents the results of the corresponding analysis. The section is divided into four sub-segments, one for each bias that is analyzed: Optimism bias, temporal bias, anchoring bias and representativeness bias. Table 8 provides a quick guide to the measurements and tests that are discussed in this section.

Table 8. Overview of tests used in Section 6

Bias Name	*Purpose*	*Test statistic*
General optimism pessimism	Analyze the existence of optimism bias	t-test
Temporal difference in Optimism	Analyze the existence of temporal bias	t-test
Unfiltered Anchoring	Analyze the existence of anchoring bias using all forecasts	Regression, t-test
Filtered anchoring	Analyze the existence of anchoring bias disregarding directional misses	Regression, t-test
Representativeness bias	Analyze the existence of representativeness bias	Binomial test

6.1 Optimism Bias

Method. Recall that optimism is defined as systematic under over estimation. If the mean forecast error for a combination of forecast horizon and variable is significantly different from zero, then systematic over/under estimation is deemed to exist for that combination.

The forecast error is defined as the actual outcome minus the corresponding forecast. If the difference is smaller than zero, then the forecast is too high; if it is greater than zero then the forecast is too low. Ordinary Student's t-tests (cf. Newbold, Carson, and Thorn, 2002) are used to analyze if the mean forecast errors are significantly different from zero, the test statistic is provided in Appendix I. For private consumption, fixed investment and GDP negative mean errors imply an optimism bias. The reverse is true for government consumption, where negative mean errors imply a pessimism bias. The reasoning is that high private consumption, fixed investment and GDP are perceived to be positive, whereas high government consumption is generally deemed negative.

If mean errors across all forecast horizons exhibit the same sign and are significant, the variable in question is subject to optimism or pessimism bias. For example, if the mean error across all forecast horizons for GDP are significantly greater than zero, then GDP is subject to a pessimism bias.

Results. Table 9 shows that government consumption is systematically underestimated, as forecasts across all horizons are significantly smaller than zero ($p<0.001$). However, none of the other variables exhibits a similar pattern, as none have more than two forecast horizons with mean errors significantly different from zero. Furthermore, all other variables have mean errors on both sides of zero. Thus, the forecasts of government consumption are systematically too optimistic, whereas there is no systematic tendency of optimism or pessimism for the other variables.

Table 9. Mean errors and the corresponding significance levels for all combinations of forecast horizons and variables.

	ForEst	*For0.5Y*	*For1.0Y*	*For1.5Y*	*For2.0Y*
Private	0.10[a]	0.16[a]	0.14	-0.02	0.02
Consumption	(1.03)	(1.42)	(2.02)	(2.13)	(2.19)
Government	0.18[c]	0.42[c]	0.43[c]	0.58[c]	0.55[c]
Consumption	(1.14)	(1.47)	(1.74)	(1.90)	(1.97)
Fixed	0.04	0.08	-0.28	-0.70[a]	-0.91b
Investments	(4.02)	(4.04)	(5.99)	(6.16)	(6.68)
Gross Domestic	0.02	0.14[a]	0.01	-0.15	-0.16
Product	(0.94)	(1.35)	(1.91)	(2.09)	(2.14)

Note: The table shows the mean errors. The standard deviations are shown within brackets. Superscript a indicates ($p<0.05$); b indicates ($p<0.01$); and c indicates ($p<0.001$).

6.2 Temporal Bias

Method. Hypotheses H1, H3 and H4 (cf. Section 3.4) are tested using a one-sided two-sample t-test under the null-hypothesis that the 2.0-year forecasts of a variable are greater than or equal to the estimates of that variable. If a null-hypothesis is rejected, then the 2.0-year forecasts for the variable in question are deemed more optimistic than the corresponding estimates and consequently, the matching hypothesis is supported. H2 is tested in the same way, but as the hypothesis regards optimism in government consumption, the null-hypothesis states that the 2.0-year forecasts are smaller than the estimates. The test statistic for the two sample t-test is found in Appendix I.

The rationale for comparing 2.0-year forecasts to estimates is that the 2.0-year forecasts are the first published predictions while the corresponding estimates are the last predictions.

Thus, if optimism decreases over time, the difference between those horizons should be the largest. Consequently, this difference should be the easiest to detect. Formally, the hypotheses used for the t-tests are shown below. For government consumption, the signs of the hypotheses are reversed.

H_0: Difference between estimates and 2.0-year forecasts ≥0

H_1: Difference between estimates and 2.0-year forecasts <0

Results. The differences between the 2.0-year forecasts and the estimates have the expected signs, i.e. 2.0-year forecasts are more optimistic than the estimates. The difference is significant at the five percent level for all variables except private consumption. For government consumption the difference is 0.38 percentage points (ppts) ($p<0.001$); for fixed investment it is 0.93 ppts ($p<0.001$) and for GDP it is 0.18 ppts ($p<0.025$). For private consumption, the difference is 0.08 ppts, which is not significant at the five percent level.

6.3 Anchoring Bias

Method. To analyze and describe anchoring bias, we rely on a test developed by Lawrence and O'Connor (1995) who used the following approach to analyze the existence of anchoring bias. Adjustment is defined as the difference between forecasts and the anchor.

$$Adjustment = Forecast - Anchor \qquad (5)$$

If the data exhibits anchoring, then adjustments from the anchor are insufficient. To analyze the existence of a potential anchoring bias the following equation is considered:

$$ActualOutcome = \lambda\, Anchor + \gamma\, Adjustment \qquad (6)$$

The anchoring bias is verified if λ is close to unity and γ is significantly greater than unity. To simplify the interpretation of Equation 6 somewhat, the following definition is made:

$$\rho = \gamma - \lambda \qquad (7)$$

Obviously, Equation 7 in itself does not simplify the interpretation of Equation 6. However, if Equation 5 and 6 are combined and 7 is used, we arrive at Equation 8, which is written in the form of a regression.

$$Actual\,Outcomev_{v,t} = \alpha + \beta\, Forecast_{h,t,v} + \rho\, Adjustment_{h,t,v} + u_{h,v} \quad (8)$$

If the anchor affects the forecasts, the adjustments have explanatory power for the actual outcome, controlling for the forecast. In that case, ρ, which measures the correlation between the actual outcome and the adjustment, controlling for the forecast, is significantly greater than zero. This reasoning yields the hypothesis-pair below, where the null-hypothesis implies no anchoring bias.

$$H_0 : \rho = 0 \qquad H_1 : \rho > 0$$

A potential problem arises when measuring the anchoring bias from a regression such as Equation 8. Some forecasts are directionally incorrect and have the wrong sign, i.e. the forecasts predict a downturn, but the actual outcome is an upturn. Even though the direction of the forecast is wrong, the forecast may still be affected by the anchor. Directional misses will result in a downward bias of ρ in equation 8. For this reason, Bromiley (1987) filtered out all directional misses in his analysis of the anchoring bias. Lawrence and O'Connor (1995) argued that eliminating directional misses produces results that are suboptimal and possibly skewed. Their argument is built around random walk time series and optimal forecasts. However, Batchelor (2001) stated that most macroeconomic variables are not characterized by random walk-series. Nevertheless, we consider both approaches in this study. Thus, the test is run twice: Once including all data points and once with directional misses filtered out.

The test is conducted for two anchors; those are the initial forecast and the last known actual outcome. The initial forecast is simply the 2.0-year forecast whereas the last known actual outcome is the last known actual outcome at the time the forecast was made. For example, the 1.0-year forecast for 1995 was published in December 1994. In December 1994, the last known actual outcome was published in June 1994 and considered 1993. Consequently the anchor point for the 1.0-year forecast for 1995 would be the actual outcome for 1993.

Results. The results from the anchoring test that includes directional misses are that only one combination of forecast horizon and variable, the 0.5-year forecast of private consumption which is anchored on the last known actual outcome, shows significant anchoring bias. None of the other combination of forecast horizon and variable exhibits significant anchoring on any of the anchors.

Table 10 presents the results from the anchoring tests with the directional misses filtered out. The numbers without brackets are estimates of ρ from Equation 8 using the last known actual outcome as the anchor. The numbers within square brackets also show estimates of ρ from Equation 8, but with the initial forecast used as anchor.

The anchoring effect on the last known actual outcome is prevalent, especially for the 0.5-year forecasts where all variables exhibit significant anchoring. Among the variables, fixed investment and private consumption are most affected, with four out of five forecast horizons showing significant anchoring.

Thus, concerning anchoring on the initial forecast, all forecast horizons of government consumption are significantly anchored. Furthermore, all 1.0- and 1.5-year forecasts except those of private consumption are significantly anchored.

Table 10. Results from filtered anchoring test and the corresponding significance levels for all combinations of forecast horizons and variables

	Est	*For0.5Y*	*For1.0Y*	*For1.5Y*	*For2.0Y*
Private Consumption	0.11^{c}	0.20^{c}	0.14^{c}	0.10^{b}	-0.01
	[-0.07]	[-0.08]	$[0.34]^{c}$	[0.27]	
Government Consumption	0.05^{a}	0.23^{c}	0.02	0.02	-0.08
	$[0.11]^{a}$	$[0.17]^{b}$	$[0.21]^{b}$	$[0.37]^{a}$	
Fixed Investment	0.02	0.07^{a}	0.11^{b}	0.13^{b}	0.13^{b}
	[0.02]	[0.06]	$[0.42]^{b}$	$[0.97]^{c}$	
Gross Domestic Product	0.02	0.16^{c}	0.14^{c}	0.17^{c}	0.09
	[0.05]	$[0.26]^{c}$	$[0.50]^{c}$	$[1.23]^{c}$	

Note: The table shows the estimated values of ρ from Equation 8 using the last known actual outcome as anchor. The figures within square brackets are estimated values of ρ from Equation 8 using the initial forecast as anchor. Superscript a indicates (p<0.05); b indicates (p<0.01); and c indicates (p<0.001). The For2.0Y column is blank for values of anchoring on the initial forecast. The reason is that the initial forecast (For2.0Y) cannot anchor on itself.

6.4 Representativeness bias

Method. We develop a test to describe and analyze the presence of the representativeness bias. Compared to a rational forecaster, a forecaster who is affected by hot hand bias is more likely to predict continuation of trends more than is really warranted. Similarly, a forecaster who is affected by gambler's fallacy is more likely to predict discontinuation of trends more than is really warranted. In order to identify trends in the data, trends are defined as runs of acceleration or deceleration in the actual outcomes. Four types of trends are considered.

Positive runs of order two: $Actual\ Outcome_{t} > Actual\ Outcome_{t-1} > Actual\ Outcome_{t-2}$

Positive runs of order three: $Actual\ Outcome_{t} > Actual\ Outcome_{t-1} > Actual\ Outcome_{t-2} > Actual_{t-3}$

Negative runs of order two: $Actual\ Outcome_{t} < Actual\ Outcome_{t-1} < Actual\ Outcome_{t-2}$

Negative runs of order three: $Actual\ Outcome_{t} < Actual\ Outcome_{t-1} < Actual\ Outcome_{t-2} < Actual\ Outcome_{t-3}$

In order for a trend to affect the forecaster, the trend has to be known at the time of the forecast. For example, for a run of order three to affect the 1.5-year forecast of private consumption for 1989 the run must precede June 1988, which is when the forecast was made. In June 1988, the last known actual outcome is for 1987. Consequently, for a run to exist and affect the 1.5-year forecast of government consumption for 1989, the actual outcomes for private consumption must have accelerated or decelerated continuously from 1984 to 1987.

If a forecaster suffers from hot hand bias, negative runs are more likely to be followed by too negative rather than too positive forecasts. Likewise, positive runs are more likely to be followed by too positive rather than too negative forecasts. The opposite relationship would be expected to hold if the forecaster was subject to a gambler's fallacy bias, i.e. negative runs are more likely to be followed by too positive rather than too negative forecasts. Likewise, positive runs are more likely to be followed by too negative rather than too positive forecasts.

To analyze whether the forecasters are subject to gambler's fallacy or hot hand bias, we identify the runs in the data. We count the number of times a run is followed by an error, which implies that the forecaster is affected by hot hand bias and the number of the times a run is followed by an error that implies that the forecaster is affected by gambler's fallacy. Subsequently, we compare the number of gambler's fallacy errors to the number of hot hand errors. Using a binomial distribution, we test if one type of error is significantly more prevalent than the other. We run the test for all combinations of forecast horizons, variables and runs. The formal null-hypotheses for any combination of forecast horizon, variable and order of run is that there are equally many hot hand mistakes as gambler's fallacy mistakes. The alternative hypothesis is that the number of hot hand mistakes and gambler's fallacy mistakes are significantly different from each other. In order to assert whether there is a representative bias, there need to be systematic differences with respect to hot hand bias and gambler's fallacy bias.

H_0: *Number of hot hand errors = Number of Gambler's fallacy errors*

H_1: *Number of hot hand errors ≠ Number of Gambler's fallacy errors*

Results. In Table 11, the ratios without brackets show the number of gambler's fallacy errors to the number of hot hand errors after trends of order two, the ratios within square brackets show the same thing, but after trends of order three instead of two.

All 40 combinations of trends, forecast horizons and variables exhibit more gambler's fallacy errors than hot hand mistakes.[10]

After a trend of order two, all forecasts with a horizon of more than half a year exhibit significantly more gambler's fallacy mistakes than hot hand mistakes. The difference is highly significant ($p<0.001$) for seven of 14 forecast horizons that have significant differences.

Considering gambler's fallacy after runs of order three, all the forecasts with horizons of more than a year (except for the 1.5-year forecast of government consumption) exhibit a significantly larger number of gambler's fallacy mistakes than hot hand mistakes.

Table 11. Results of test for representativeness bias and the corresponding significance for all combinations of forecast horizons and variables

	ForEst	*For0.5Y*	*For1.0Y*	*For1.5Y*	*For2.0Y*
Private	69/73	73/86	60/89^{b}	43/102^{c}	46/91^{c}
Consumption	[25/29]	[23/37]	[20/36]^{a}	[17/42]^{c}	[11/45]^{c}
Government	57/73	61/75	50/79^{b}	46/86^{c}	56/74^{a}
Consumption	[23/29]	[22/32]	[21/30]	[22/29]	[17/31]^{a}
Fixed	83/95	71/115^{c}	61/120^{c}	57/125^{c}	72/98^{a}
Investment	[39/40]	[35/47]	[26/54]^{c}	[21/60]^{c}	[21/56]^{c}
Gross Domestic	63/76	63/88^{a}	53/93^{c}	49/100^{c}	59/80^{a}
Product	[27/28]	[28/30]	[27/32]	[24/36]^{a}	[18/37]^{b}

The table shows the number of gambler's fallacy mistakes over the number of hot hand mistakes after a trend of order two. The number of gambler's fallacy mistakes over the number of hot hand mistakes after a trend of order three are presented within brackets. Superscript a indicates ($p<0.05$); b indicates ($p<0.01$);

[10] The sum of gambler's fallacy and hot hand errors is not equal for all forecast horizons that share the same trend and variable. The main reason is that sometimes the forecasts are exactly correct and thus, forecasts can neither be classed as gambler's fallacy or hot hand mistakes. As the occurrence of exactly correct forecasts varies between the forecasts horizons, the sum of hot hand and gambler's fallacy mistakes differs between them.

7. Discussion

This section begins with a discussion of the results of the forecast evaluation. We focus this discussion on how the results of this thesis compares to those of previous forecast evaluations. Subsequently, we discuss the results relating to the cognitive biases analysis. Thereafter, discussion points are raised which focus on reliability and validity of the thesis. Suggestions for further research, as well as some concluding remarks follow.

7.1 Forecasting Accuracy Evaluation

The four most relevant findings from the evaluation of the OECD forecasts are:

- Fixed investment forecasts exhibit the largest errors;
- there is a clear association between forecast and actual directional change;
- the forecasts are better than naïve models;
- the forecasts are generally not rational since there are systematic forecast errors

Descriptive Error Measures. The descriptive error measure (*MAE)* showed that fixed investment had the largest errors. Within the same forecast horizons, the *MAEs* for private consumption, government consumption and GDP were all of similar size. The *MAEs* for all variables increased monotonically with the forecast horizon. Thus, the length of the forecast horizons is important for precision; the longer the forecasting horizon the less accurate the forecasts are.

Our results are largely in line with previous studies. Ash et al. (1990) presented quantitative accuracy measures for all variables we analyzed in this thesis. As for the forecast horizons, Ash et al. (1990) investigated (0.5-year, 1.0-year and 1.5-year forecasts), the *MAEs* presented in this thesis are similar to the values that they presented. The exception is government consumption. The *MAEs* for government consumption presented in the present study were 1.01 (For0.5Y), 1.19 (For1.0Y) and 1.37 (For 1.5Y). The corresponding values for Ash et al. (1990) were 1.41, 1.95 and 2.13. The reason for the differences in government consumption is not clear as Ash et al. (1990) analyzed the OECD forecasts using the same forecast horizons and vintage of data as we did, hence such factors do not help in explaining the difference in the results. One possibly important difference is that Ash et al. (1990) used data for the G7 countries only, whereas this thesis used the data for 23 countries. However, there is no intuitively appealing reason for why government consumption for the G7 countries should be particularly hard to forecast. Perhaps an explanation lies in the stricter fiscal regimes that were implemented during the 1990's given the requirements to join the EMU.

Directional Error Measure. Our study showed that there was a highly significant association between forecast directional change and actual directional change. This was true for all combinations of variables and forecast horizons. Consequently, the forecast directional change for all combinations of variables and forecasts were relatively reliable.

We compare our findings to those of Pons (2000). Pons (2000) analyzed GDP forecasts with 0.5- and 1.0-year horizons for the G7 countries. Unfortunately, the study did not pool the data. Nonetheless, the study showed that forecast directional change was significantly associated to actual directional change, except for Italy. As almost all forecast horizons were significant, pooling them would have led to highly significant results (cf. section 4.5). The results for 0.5-year and 1.0-year forecasts of GDP are also highly significant in our study. Consequently, the results of the present study are likely to be in line with previous research.

Naïve/No Change comparison. Using pooled data, the results from the comparison of the OECD forecasts to those generated from a naïve model indicated that the OECD forecasts were markedly better than the naïve model. This was also the case when median values of the 23 countries were analyzed, though the difference was not as large as for the pooled study. Also, the superiority of long-term forecasts (1.5- and 2.0-year forecasts) for OECD forecasts of fixed investments was very slight. Consequently, with the exception of long-term forecasts of fixed investment, on average, the OECD forecasts were substantially superior to forecasts generated from a naïve model.

Earlier evaluations of economic forecasting did not study the Theil's U coefficient using pooled data. Consequently, the results from the pooled analysis are not directly comparable to findings presented in previous studies. The U-coefficients for individual countries are similar to those presented in Holden and Peel (1987) and Ash et al. (1990). The same holds for the medians of the individual country U-coefficients. For example, the median U-coefficient for the 1.0-year forecast of government consumption presented in Holden and Peel (1987) was 0.71 while the value in the present study was 0.60. The corresponding values for GDP were 0.70 and 0.54. The fact that both values from the present study were lower might depend on that the present study has more observations per country 20 as compared to twelve in Holden and Peel (1987).

Rationality of Forecasts. Forecasts are considered as rational when they do not exhibit systematic forecast errors. Only one combination of variable and forecast horizon, the half year forecast for private consumption, passed both rationality tests. All other combinations of forecast horizons and variables failed at least one of the two tests, i.e. either forecast errors were repeated or the forecasts errors correlated significantly with the forecasts. Many failed both tests. Consequently, the OECD forecasts were generally not rational. This finding contrasts the findings of other studies. Ash et al. (1990) found that around 50 per cent of 252 combinations of variables, countries and forecast horizons failed to pass at least one of their tests of rationality. The present study rejects the notion of rationality for 19 of 20 combinations of forecast horizons and variables. The difference may be due to the fact that we used pooled data, whereas Ash et al. (1990) used data for individual country data. Pooling the data yields more observations which increases the statistical power of tests, making the tests more sensitive to small but consistent deviations from null-hypotheses (cf. Section 4.5).

7.2 Cognitive Bias Evaluation

The three most important findings from the cognitive bias investigation were:

- Government consumption forecasts were too low across all forecast horizons;
- two year forecasts were generally more optimistic than estimates;
- OECD forecasters overall exhibited signs of gambler's fallacy.

Optimism bias. The forecasts for government consumption were consistently too low across all forecast horizons. Thus government consumption exhibited optimism bias. The other components did not show a systematic optimism or pessimism bias.

These results are in line with previous research. Ash et al. (1990) also found a systematic tendency of too optimistic forecasts for government consumption. However, as the other variables did not exhibit a systematic bias, it is may not be a cognitive bias that causes the systematic optimism. Instead, a potential explanation for the systematic underestimation of government consumption might be derived from public choice theory. Mueller (2003) noted that politicians have an incentive to increase government spending continually. The reasoning is that the electorate perceives the cost of government spending to be less than it actually is. If the forecasters fail to consider this incentive, but trust budgeted spending, their forecasts might be to low. Consequently, a potential explanation is that the OECD forecasters do not take this tendency into account when forecasting government consumption. Another possible explanation might be that of time inconsistency problems in tax policies as presented by Kydland and Precott (1977). They showed that a government could pledge low taxes for certain kinds of activities, e.g. investment, but after the investments have been made, the government can renege on the promised tax-cuts. The revenues for the increased tax cuts may go towards increased government spending. If the forecasters fail to account for this problem, then the forecasts of government spending might be too low. As the fiscal policy assumptions in the forecasts are based on the officially declared policies, time-inconsistency on the part of politicians is a viable explanation for the systematic underestimation.

Temporal difference in optimism. Hypotheses H1 to H4 (cf. sec 3.4 p. 17) state that forecasts with longer horizons are more optimistic than forecasts with shorter horizons. The hypotheses were supported in that there was a significant difference in optimism between two year forecasts and estimates for government consumption, fixed investment and gross domestic product. However, there was no significant difference in private consumption. Hence, hypothesis H2, H3 and H4 were supported.

We are not aware of any study that has conducted a similar test on macroeconomic forecasts. Thus, we cannot compare our findings directly to other studies. Nevertheless, the hypotheses were mainly inspired by Andersson (2006) and Choi and Ziebart (2004), where this tendency had been noted in students and financial analysts. Thus, the results were not without precedence for forecasting in general.

Anchoring bias. Two anchor points were hypothesized, the initial forecast and the last known actual outcome. The results of the tests were mixed, given that including the directional misses led to a lack of evidence in favour of anchoring. When the directional misses were excluded, there was evidence for anchoring on both anchor points.

Directional misses may have a strong effect on the tests due to adjustments being made in the wrong direction. Lawrence and O'Connor (1995) argued that it is wrong to eliminate directional misses from a test of anchoring. Their argument was that such elimination would take out a significant part of optimal forecasts of a random walk series, yielding biased and possibly skewed results. The argument for taking out the directional misses is simply that those misses are affected by anchoring as well. However, including them in the analysis will lead to downward biased estimates of anchoring (Bromiley, 1987). It is notable that the lack of rationality in the forecasts implies that the forecasts were not optimal. In addition to this, Batchelor (2001) argued that most macroeconomic forecasts are not characterized by a random walk. Consequently, the reasoning presented by Lawrence and O'Connor (1995) is not entirely convincing. Nevertheless, we do not feel that we can discard their objections completely and we choose to neither confirm, nor dismiss the existence of an anchoring bias in macroeconomic forecasts.

Representativeness bias. For all combinations of variables, forecast horizons and trends, there were more gambler's fallacy mistakes than hot hand mistakes. If the probability of the occurrence of gambler's fallacy mistakes and hot hand mistakes were equal, then the probability of obtaining the outcome in this study is lower than 0.0000000001 per cent. To put it in perspective, the chance for a random person to be hit by the lightning in USA any given year is 1000 times larger than this probability (cf. Lopés & Holle 1987). Consequently, there is strong evidence that macroeconomic forecasters are subject to gambler's fallacy.

However, when specific combinations of forecast horizons, variables, and trends were studied, the evidence for the existence of gambler's fallacy was somewhat weaker. Nonetheless, all forecasts that were made after a trend of order two, with horizons of more than a half a year exhibited significantly more gambler's fallacy mistakes than hot hand mistakes. With the exception of government consumption, the same was true for forecast horizons with horizons of more than a year that were preceded by a trend of order three. It is notable that the forecasts seemed to be increasingly affected by gambler's fallacy as the forecast horizons lengthened. For the shortest forecast horizons (the estimates), no variable exhibited a significantly higher number of gambler's fallacy mistakes than hot hand mistakes; whereas for the longest forecast horizon (the 2-year forecasts) all variables exhibited a significantly higher number of gambler's fallacy mistakes than hot hand mistakes.

7.3 Discussion Points

Possible criticisms of the thesis rest in theoretical definitions of reliability and validity. Reliability refers to the possibility of replicating the results of the study. Validity refers to whether the tests measure what they intend to.

Reliability. The study is reliable in the sense that it is easy to replicate, given that both forecasts and outcome data are readily available from numerous forecasting institutions. Though it might be laborious to construct a similar dataset, it is definitely possible. The forecast evaluation results are also likely to be similar as our results are closely aligned to results obtained in previous studies. The results of a prospective bias analysis are also likely to be similar as it should not only be OECD forecasters for the period in question that are subject to the investigated cognitive biases. Nevertheless, two weaknesses of the thesis are that we have only studied the OECD forecasts and that it is far from ideal to study a period of only 20 years, as many important economic developments, such as the oil shocks of the 1970's took place outside the sampling period. Ultimately, these two issues might affect the reliability negatively.

Validity. One issue that concerns validity is related to the naïve/no-change comparison. Batchelor (2001) stated that comparing macroeconomic forecasts to models implying no change from the last period are rather meaningless. According to him, such a comparison is suited for variables that are characterized by random walks which is a property that most macroeconomic variables lack. The rationality tests can also be questioned on grounds of validity as well. The results from the rationality tests rest on the definition of rational forecasts. As there are many definitions of rational forecasts, perhaps another definition of rationality would have yielded other results.

Another issue is the validity of the findings from the cognitive bias analysis. It could be argued that the systematic underestimation of government consumption could be due to economic shocks, rather than a consistent cognitive bias. The difference in optimism of forecasts with long horizons vs. short horizons is tentative. The decrease in optimism with the shortening of the forecast horizons is not monotonic. Thus, the tests of the four hypotheses are not necessarily representative for all forecast horizons. Another weakness is that the hot hand and gambler's fallacy test does not estimate the size of the bias. Thus, the multitude of gambler's fallacy errors might be offset by the magnitude of the hot hand errors.

7.4 Suggestions for Further Research

Future research ideas abound within the aspects discussed in this thesis. The other expenditure variables of GDP, like imports, exports and change in stock building along with inflation, could be analyzed using similar methods. Moreover, it is possible to conduct a similar analysis on a country level rather than on pooled data. Perhaps such a study could answer the question of whether the cognitive biases are spread equally across countries or if they only affect specific countries. Furthermore, other biases noted in cognitive research, e.g. base rate fallacy and availability bias, could be tested for. It may also prove interesting to study cognitive biases on other forecasters for other time-periods.

7.4 Concluding Remarks

Our study had two purposes. (1) To evaluate the quality of the OECD's forecasts for GDP and its expenditure components and (2) to determine if the OECD's forecasts are afflicted by cognitive biases. We believe that both purposes have been fulfilled.

Our results from the accuracy evaluation of the macroeconomic forecasts are in line with previous research. However, conclusions about rationality of the forecasts differed from previous studies as we observed more irrationality. The difference may be due to the fact that the data in the present study used pooled data, resulting in higher statistical power than in earlier examinations of rationality of forecasts.

The main results from the analysis of cognitive biases were threefold. First, government consumption was substantially under estimated across all forecast horizons, showing a consistent optimism bias. Second, 2.0-year forecasts were generally more optimistic than 1.0-year forecasts, implying that there is a systematic temporal difference in optimism. Third, the forecasters were subject to the gambler's fallacy, especially for longer forecast horizons.

Often when results show a systematic tendency for an organization or an individual to err, the questions arises: Can this mistake be fixed? If so, how? As we have chosen to use pooled data, it is not advisable to correct the forecasts using the findings directly. Instead, each country should be evaluated in isolation in order to see how the forecasts could be improved upon by removing the influence of the cognitive bias in question. Nevertheless, we have shown that cognitive biases exist in macroeconomic forecasting. If those cognitive biases were corrected for, the forecasting quality may be improved. However, before problems can be corrected, the existence of the problems must be known. A person who does not know she is ill will not cure herself. The same is true for macroeconomic forecasting. In order to correct for cognitive biases it must be known that they exist and substantially affect the forecasts.

References

Abarbanell, J., & Bernard, V. (1992). Test of analysts' overreaction/underreaction to earnings information as an explanation of anomalous stock price behavior. *The Journal of Finance, 47,* 1181-1207.

Andersson, P. (2005). Overconfident but yet well-calibrated and underconfident: A research note on judgmental miscalibration and flawed self-assessment. University of Mannheim, *Working paper, 37*.

Armstrong, J. (1985). *Long range forecasting: From crystal ball to computer*. New York: Wiley

Ash, J., Smyth, D., & Heravi, S. (1990). The accuracy of OECD forecasts of the international economy. *International Journal of Forecasting, 6,* 379-392.

Ash, J., Smyth, D., & Heravi, S. (1998). Are OECD forecasts rational and useful? A directional analysis. *International Journal of Forecasting, 14,* 381 – 391.

Ayton, P., & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory and Cognition, 32,* 1369-1378.

Baquero, G., & Verbeek, M. (2005). Do sophisticated investors believe in the law of small numbers? Erasmus Univeristy Rotterdam, *Working Paper.*

Batchelor, R. (2001). How useful are the forecasts of intergovernmental agencies? The IMF and OECD versus the consensus. *Applied Economics, 33,* 225-235.

Bernard, V., & Thomas, J. (1989). Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research,27,* 1-36.

Bolger, F., & Harvey, N. (1998). Heuristics and biases in judgmental forecasting. In Wright G., & Goodwin, P. (Eds.), *Forecasting with judgment* (pp.113-137). Chichester: Wiley.

Bromiley, P. (1987). Do forecasts produced by organizations reflect anchoring and adjustment? *Journal of Forecasting, 6,* 201-210.

Brown, L. (1996). Analyst forecasting errors and their implications for security analysis: An alternative perspective. *Financial Analysts Journal, 52 (1)*, 40-47.

Brown, L. (2001). Temporal analysis of earnings surprises: Profit versus losses. *Journal of Accounting Research, 39,* 221-241.

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the planning fallacy: Why people underestimate their task completion times. *Journal of Personality and Social Psychology, 67,* 366-381.

Choi, J., & Ziebart, D. (2004). Management earnings forecasts and the market's reaction to the bias in the forecasts. *Asia-Pacific Journal of Accounting and Economics, 11,* 167-192.

Davis, H., Hoch, S., & Ragsdale, E. (1986). An anchoring and adjustment model of spousal prediction. *Journal of Consumer Research, 13,* 25-37.

De Bondt, W. (1991). What do economists know about the stock market? *The Journal of Portfolio Management, 17 (2), 85-91.*

DeBondt, W., & Thaler, R. (1985). Does the stock market overreact? *Journal of Finance, 40,* 793-805.

DeBondt, W., & Thaler, R. (1987). Further evidence of investor overreaction and stock market seasonality, *Journal of Finance, 42,* 557-582.

De Bondt, W., & Thaler, R. (1990). Do security analysts overreact? *The American Economic Review, 80* (2), 52-57.

Fildes, R., & Stekler, H. (2002). The state of macroeconomic forecasting. *Journal of Macroeconomics, 24,* 435-468.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment.* New York: Cambridge University Press.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17,* 295-314.

Gujarati, D. (2002). *Basic econometrics*. (4^{th} ed.). New York: McGraw-Hill.

Holden, K., & Peel, D. (1987). The accuracy of OECD forecasts. *Empirical Economics, 12,* 175-183.

Hong, H., & Kubik, J. (2003). Analyzing the analysts: Career concerns and biased earnings forecasts. *Journal of Finance, 58,* 313-351.

Johnson, J., Tellis, G., & Macinnis, D. (2005). Losers, winners and biased trades. *Journal of Consumer Research, 32,* 324-329.

International Monetary Fund. (1988). *How accurate is the World Economic Outlook? A post mortem on short-term forecasting at the International Monetary Fund*. Washington D.C: Artis, M..

International Monetary Fund. (1993). *How accurate are the World Economic Outlook projections?* Washington, D.C.: Barrionuevo, J..

International Monetary Fund. (1996). How accurate are the IMF's short-term forecasts? Another examination of the World Economic Outlook. Washington, D.C.: Artis, M..

Kydland, F., & Prescott, E. (1977). Rules rather than discretion: The inconsistency of optimal plans. *Journal of Political Economy, 85, 473-492.*

Lawrence, M., & O'Connor, M. (1995). The anchor and adjustment heuristic in time series forecasting. *Journal of Forecasting, 14,* 443-451.

Lopéz, R., & Holle, R. (1997). Changes in the number of lightning deaths in the United States during the twentieth century. *Journal of Climate, 11,* 2070-2077.

Malloy, C. (2005). The geography of equity analysis. *Journal of Finance, 60*, 719-755.

Montgomery, H. (2000). The psychology of economic forecasting: A possibility for Co-operation between JDM and NDM theories? In Mongtomery, H., Lipshitz, R., & Brehmer, B. (Eds.), *How professionals make decisions (pp. 125-135).* Mawah: Lawrence Erlbaum

Montgomery, H., & Törgren, G. (2003). The psychology of forecasting: What can be seen in public data? Presented at the Nineteenth Biannual Conference on Subjective Probability, Utility and Decision Making, August 25-27, Zurich: Switzerland.

Morrison, W., Ager, A., & Willock, J. (1999). Perceived risk of tropical diseases in Malawi: Evidence of unrealistic pessimism and the irrelevance of beliefs of personal control? *Psychology, Health and Medicine, 4,* 361-368.

Mueller, D. (2003). *Public Choice III*. Cambridge, UK: Cambridge University Press.

Newbold, P., Carlson, W., & Thorne, B. (2002*). Statistics for business and economics*. (5th ed.). New Jersey: Prentice-Hall.

Northcraft, G., & Neale, M. (1987). Expert, amateurs, and real estate: An anchoring and adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes, 39,* 84-97.

Organization of Economic Co-operation and Development. (1988). *The structure and simulation properties of the OECD's INTERLINK model.* Paris, France: Richardson, P..

Organization for Economic Co-operation and Development. (1999). *Policy Brief: OECD Economic Projections.* Paris, France: OECD Public Affairs Division and Communications Directive.

Organization of Economic Co-operation and Development. (2000). *A post-mortem on Economic Outlook projections.* Paris, France: Koutsogeorgopoulou, V..

Pons, J. (2000). The accuracy of IMF and OECD forecasts for G7 countries. *Journal of Forecasting, 19,* 53-63.

Rosenthal, R., & DiMatteo, M. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology,* 52, 59-82.

Russo, E., & Shoemaker, P. (1989). *Decision Traps*. New York: Simon and Schuster.

Shefrin, H., & Statman, M. (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *The Journal of Finance, 40,* 777-790.

Simon, H. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics, 69,* 99-118.

Sirri, E., & Tufano, P. (1998). Costly search and mutual fund flows. *Journal of Finance, 53,* 1589-1622.

Sundali, J., & Croson, R. (2006). Biases in casino betting: The hot hand and the gambler's fallacy. *Judgment and Decision Making, 1,* 1-12.

Sveriges Riksbank. (2001). *How good is the forecasting performance of major institutions?* Sveriges Riksbank, Stockholm: Blix, M., Wadefjord, J., Wieneke, U., & Ådahl, M..

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76,* 105-110.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124-1130.

United Nations. (1994). *System of national accounts*. New York: United Nations Publications.

Vuchelen, J., & Gutierrez, M-I. (2005). Do the OECD 24 months horizon growth forecasts for the G7-countries contain information? *Applied Economics, 37,* 855–862.

Webb, R. (1999). Two approaches to macroeconomic forecasting. *Federal Reserve Bank of Economic Quarterly, 85* (3), 23-40.

Wärnreryd, K-E. (2001). *Stock Market Psychology.* Cheltenham: Edward Elgar.

Zarnowitz, V. (1979). An analysis of annual and multiperiod quarterly forecasts of aggregate income, output, and the price level. *Journal of Business, 52*, 1-33.

Öller, L., & Barot, B. (2000). Comparing the accuracy of European GDP forecasts. *International Journal of Forecasting, 16,* 293-315.

Appendix

Appendix I. Measures and Tests

Name	*Abbreviation*	*Formula*
Mean Error	ME	$\frac{1}{n}\sum\left(P_t - A_t\right)$
Mean Absolute Error	MAE	$\frac{1}{n}\sum\left\|P_t - A_t\right\|$
Mean Square Error	MSE	$\frac{1}{n}\sum\left(P_t - A_t\right)^2$
Root Mean Square Error	RMSE	$\sqrt{\frac{1}{n}\sum(P_t - A_t)^2}$
Theil's U Coefficient	U	$\frac{\sqrt{\frac{1}{n}\sum\left(P_t - A_t\right)^2}}{\sqrt{\frac{1}{n}\sum A_t^2}}$
Student's t-test	t	$\frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n-1}}}$
Two-sample t-test	t	$\frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\left[\frac{(n_1 - 1)\sigma_1{}^2 + (n_2 - 1)\sigma_2{}^2}{n_1 + n_2 - 2}\right]}}$

Note: Concerning the error measures, P stands for predicted, A for average, n for the number of observations Concerning the t-tests, subscripts (1) and (2) denotes groups, X-bar stands for the mean, σ^2 for the estimated variance and n for number of observations. The t-tests are from Newbold, Carlson and Thorne (2002) whereas the error measures can be found in Ash et al. (1990).

Appendix II. Mean Absolute Error for all combinations of countries, forecast horizons and variables

	Private Consumption					Government Consumption					Fixed Investment					Gross Domestic Product				
	MAE-Est	MAE-0.5Y	MAE-1.0Y	MAE-1.5Y	MAE-2.0Y	MAE-Est	MAE-0.5Y	MAE-1.0Y	MAE-1.5Y	MAE-2.0Y	MAE-Est	MAE-0.5Y	MAE-1.0Y	MAE-1.5Y	MAE-2.0Y	MAE-Est	MAE-0.5Y	MAE-1.0Y	MAE-1.5Y	MAE-2.0Y
Australia	0.56	0.68	1.12	1.03	1.12	1.26	1.37	1.57	1.88	1.58	2.18	3.33	4.97	4.43	5.03	0.85	1.05	1.50	1.24	1.08
Austria	0.60	0.64	0.88	0.89	1.09	0.55	0.77	0.68	0.70	0.79	1.25	1.96	2.36	2.61	2.60	0.37	0.71	1.02	1.25	1.30
Belgium	0.54	0.71	1.08	1.10	1.29	0.71	0.79	0.81	0.87	1.00	2.02	2.39	3.24	3.51	3.77	0.42	0.68	1.13	1.29	1.18
Canada	0.41	0.74	0.92	1.30	1.20	0.54	0.67	0.97	1.06	1.34	1.50	2.98	4.04	3.92	3.79	0.25	0.76	1.24	1.32	1.26
Denmark	0.48	1.11	1.50	1.55	1.48	0.67	0.81	0.93	1.05	1.26	1.84	3.29	4.18	4.59	5.23	0.36	0.59	0.91	0.94	1.17
Finland	0.66	1.14	1.62	1.96	1.90	0.69	1.03	1.22	1.52	1.61	2.31	3.05	5.33	5.53	6.10	0.61	1.31	2.08	2.30	2.45
France	0.31	0.51	0.70	0.93	0.97	0.47	0.59	0.70	0.86	0.95	0.74	1.38	2.29	2.69	3.25	0.28	0.51	0.82	1.05	1.04
Germany	0.45	0.59	0.99	1.17	1.30	0.37	0.66	0.82	1.00	1.03	1.03	2.09	2.89	3.50	3.75	0.22	0.59	0.96	1.26	1.27
Greece	0.38	0.53	0.45	0.64	0.77	1.30	1.61	2.09	2.20	2.22	2.04	2.05	2.44	2.95	3.21	1.27	1.61	2.09	2.20	2.34
Ireland	0.66	1.38	1.79	2.17	2.14	1.13	0.90	1.66	2.07	2.37	1.89	3.14	4.26	5.00	5.64	1.00	1.59	1.97	2.61	2.60
Italy	0.38	0.67	0.85	1.17	1.12	0.65	0.70	0.74	1.00	0.97	1.25	1.69	2.63	3.20	3.30	0.24	0.50	0.86	1.24	1.35
Japan	0.02	0.03	0.04	0.06	0.06	0.84	0.95	1.37	1.31	1.69	0.93	2.85	3.63	4.91	4.90	0.56	1.01	1.45	1.95	1.95
Luxembourg	0.49	0.76	1.01	1.06	1.45	0.85	1.15	1.45	1.41	1.51	1.34	2.08	2.56	2.83	3.00	0.68	1.18	1.43	1.76	1.71
Netherlands	0.42	0.72	1.14	1.53	1.45	0.70	0.79	0.90	1.02	1.01	1.67	2.37	2.76	2.86	2.74	0.27	0.75	0.97	1.24	1.15
New Zealand	0.95	1.13	1.83	1.79	1.76	1.20	1.25	1.40	2.16	2.08	4.47	5.89	6.77	6.83	6.79	1.14	1.00	1.37	1.32	1.31
Norway	0.81	0.90	1.28	1.68	1.74	0.87	0.82	1.04	0.98	1.21	7.81	5.08	8.03	7.73	10.17	0.87	0.82	1.04	0.98	1.21
Portugal	0.91	1.22	1.90	1.94	2.15	0.76	0.68	0.90	1.17	1.28	3.43	3.62	5.22	5.00	5.09	0.54	0.81	1.20	1.44	1.72
Spain	0.44	0.58	0.92	1.10	1.23	0.76	0.99	1.16	1.57	1.62	1.53	2.73	3.92	4.45	5.02	0.19	0.48	0.73	1.12	1.08
Sweden	0.72	0.93	1.23	1.60	1.73	0.42	0.76	0.79	0.78	0.99	2.78	3.16	4.14	4.88	5.69	0.36	0.55	0.89	0.97	1.23
Switzerland	0.34	0.46	0.70	0.86	0.87	0.36	0.62	0.63	0.66	0.87	0.82	2.21	3.15	3.83	4.10	0.45	0.75	1.19	1.40	1.36
Turkey	1.87	3.82	4.96	5.04	4.72	1.62	3.49	3.63	4.13	3.65	3.31	7.36	10.04	10.56	10.13	1.23	3.03	3.92	3.83	3.65
United Kingdom	0.46	0.66	1.28	1.21	1.41	0.63	0.98	1.07	1.07	1.11	1.39	1.94	3.10	3.20	3.50	0.37	0.42	0.87	1.03	1.08
United States	0.28	0.45	1.07	1.01	1.15	0.45	0.71	0.81	1.00	1.29	1.85	2.09	3.47	3.79	3.87	0.56	0.74	1.33	1.37	1.36

Appendix III. Theil's U Coefficient for all combinations of countries, forecast horizons and variables

	Private Consumption					Government Consumption					Fixed Investment					Gross Domestic Product				
	U-Est	U-0.5Y	U-1.0Y	U-1.5Y	U-2.0Y	U-Est	U-0.5Y	U-1.0Y	U-1.5Y	U-2.0Y	U-Est	U-0.5Y	U-1.0Y	U-1.5Y	U-2.0Y	U-Est	U-0.5Y	U-1.0Y	U-1.5Y	U-2.0Y
Australia	0.27	0.24	0.38	0.34	0.37	0.44	0.47	0.53	0.60	0.54	0.32	0.50	0.92	0.98	0.91	0.30	0.34	0.49	0.45	0.41
Austria	0.53	0.55	0.61	0.55	0.65	0.50	0.61	0.51	0.55	0.66	0.43	0.76	1.01	1.22	1.29	0.19	0.36	0.50	0.59	0.60
Belgium	0.45	0.38	0.66	0.63	0.71	0.54	0.59	0.63	0.64	0.72	0.67	0.99	1.13	1.36	1.38	0.31	0.34	0.56	0.63	0.66
Canada	0.18	0.29	0.34	0.49	0.45	0.34	0.44	0.62	0.70	0.79	0.49	0.62	0.35	0.37	0.36	0.10	0.28	0.46	0.52	0.49
Denmark	0.22	0.43	0.67	0.66	0.66	0.57	0.62	0.78	0.86	1.01	0.32	0.70	0.75	0.78	0.70	0.21	0.31	0.54	0.54	0.67
Finland	0.23	0.40	0.66	0.74	0.75	0.38	0.53	0.65	0.83	0.92	0.14	0.36	0.97	0.84	0.62	0.20	0.44	0.69	0.77	0.81
France	0.21	0.29	0.36	0.46	0.51	0.26	0.36	0.40	0.48	0.52	0.13	0.24	0.52	0.43	0.44	0.15	0.26	0.44	0.60	0.60
Germany	0.41	0.42	0.76	0.82	0.94	0.32	0.57	0.68	0.82	0.88	0.91	0.54	1.18	1.03	1.13	0.11	0.31	0.50	0.65	0.69
Greece	0.28	0.33	0.29	0.42	0.57	0.57	0.73	0.82	0.86	0.88	0.78	0.46	0.63	0.86	0.96	0.56	0.73	0.82	0.86	1.00
Ireland	0.18	0.36	0.45	0.60	0.58	0.40	0.34	0.57	0.64	0.74	0.21	0.30	0.61	1.03	1.06	0.22	0.31	0.37	0.50	0.51
Italy	0.21	0.37	0.44	0.63	0.68	0.49	0.52	0.50	0.71	0.69	0.08	0.16	0.27	0.32	0.31	0.14	0.31	0.47	0.69	0.75
Japan	0.27	0.30	0.57	0.54	0.58	0.49	0.53	0.82	0.65	0.76	0.22	0.56	0.84	1.04	0.98	0.25	0.38	0.57	0.73	0.76
Luxembourg	0.23	0.34	0.46	0.53	0.65	0.36	0.42	0.50	0.44	0.46	0.47	0.40	0.49	0.42	0.60	0.26	0.39	0.47	0.61	0.60
Netherlands	0.27	0.38	0.53	0.68	0.65	0.48	0.54	0.60	0.68	0.62	0.54	1.14	1.26	1.34	1.06	0.13	0.35	0.48	0.60	0.58
New Zealand	0.53	0.45	0.71	0.68	0.67	0.50	0.54	0.61	1.03	0.99	0.45	0.48	0.67	0.53	0.53	0.55	0.50	0.67	0.68	0.66
Norway	0.45	0.38	0.63	0.65	0.69	0.42	0.42	0.52	0.49	0.54	1.10	0.86	0.77	0.89	1.17	0.42	0.42	0.52	0.49	0.54
Portugal	0.43	0.40	0.61	0.63	0.70	0.40	0.35	0.45	0.55	0.63	0.53	1.29	1.37	1.69	1.77	0.35	0.32	0.53	0.58	0.70
Spain	0.23	0.22	0.40	0.48	0.54	0.27	0.35	0.43	0.54	0.59	0.33	0.50	0.65	0.79	0.84	0.09	0.22	0.31	0.45	0.46
Sweden	0.34	0.45	0.64	0.79	0.84	0.38	0.68	0.74	0.68	0.79	0.22	0.24	0.40	0.50	0.88	0.21	0.31	0.45	0.53	0.65
Switzerland	0.31	0.36	0.57	0.69	0.70	0.32	0.46	0.49	0.50	0.68	0.32	0.52	0.87	0.95	1.01	0.29	0.43	0.66	0.81	0.80
Turkey	0.34	0.62	0.82	0.81	0.78	0.36	0.71	0.79	0.82	0.76	0.37	0.63	0.79	0.88	0.83	0.26	0.55	0.79	0.76	0.75
United Kingdom	0.18	0.26	0.44	0.42	0.48	0.34	0.58	0.67	0.64	0.66	0.38	0.89	0.77	1.83	1.84	0.17	0.23	0.45	0.52	0.54
United States	0.16	0.18	0.39	0.39	0.44	0.37	0.36	0.37	0.49	0.65	0.38	0.10	0.96	1.46	1.56	0.42	0.36	0.58	0.57	0.53